ICMC-USP

Lecture Notes
Graduate SME5785, Undergrad SME0254

Introduction to the Finite Element Method

Theory and Applications with the `FEniCS(x)` platform

Roberto F. Ausas

April 22, 2024

Instituto de Ciências Matemáticas e de Computação
University of São Paulo

Preface

These notes cover the basics of the finite element method (FEM), including the most relevant mathematical aspects as well as classical applications in the area of continuum mechanics. They are intended for undergrad and graduate students. For the former, some technical details and proofs can be ommited with almost no loss. This will be made clear throughout the lectures. Some classical coding aspects of the method that are commonly tought and exerciced in finite element courses will not be covered, since I have chosen the `FEniCS` project, which is a high-level library, for solving practical problems in solid and fluid mechanics, so many things remain hidden to the user. However, a few concepts will be mentioned so as to understand what is behind the scenes when using such libraries. Other platforms such as `Firedrake`, with a similar behavior can also be used and the examples provided easily adapted.

At the end of the course, I expect the student would have learned what it is behind the following typical piece of code:

```
import libraries
...
mesh = generate_mesh(domain,refinement)
x = SpatialCoordinate(mesh)
P = FiniteElement("Lagrange", "triangle", order)
V = FunctionSpace(mesh, P)
u = TrialFunction(V)
v = TestFunction(V)
a = inner(mu(x)*grad(u), grad(v))*dx
L = f(x)*v*dx - h(x)*v*ds(Neumann)
w = Function(V)
solve(a == L, w, DirichletBC(V, Constant(0), D_boundary))
...
```

Moreover, we will learn:

- How to approximate the solution of a PDE by the Finite Element Method;
- In which situations the plain vanilla FEM works and which ones fails, and in such a case, what are the possible remedies;
- How to solve practical problems (not merely academic) by means of modern FE open-source libraries as in the example above.

Bibliographical remarks

Many chapters were adapted from the Lecture Notes of Prof. G.C. Buscaglia [1], that have been used in the past years to teach this course at ICMC-USP. Throughout this document we also take ideas from other classical references on the subject, among which are [2–5]. However, some topics, notational aspects, numerical and implementation examples may differ greatly from all these sources.

Roberto F. Ausas

Contents

INTRODUCTION 1

1.1 Prelude

The finite element method (FEM) is by now one of the most popular methods for numerically solving Partial Differential Equations (PDEs) in science, engineering and applied mathematics. There are a number of good reasons for this:

- ☺ For **elliptic** and **parabolic** problems[1], the FEM provides very accurate solutions;
- ☺ It is **general**, not restricted to linear problems, or to isotropic problems, or to any subclass of mathematical problems;
- ☺ It is **geometrically flexible**, complex domains are quite easily treated, not requiring adaptations of the method itself;
- ☺ It is **easy to code**, and the coding is quite problem-independent. Boundary conditions are much easier to deal with than in other methods;
- ☺ It is **robust**, because in most cases the mathematical problem has an underlying variational structure (energy minimization, for example).

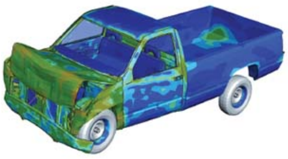
1.2 What This Course Covers

As a summary of the content of this document, let me enumerate the main topics to be covered:

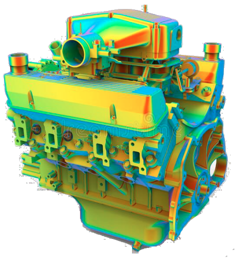
Chapter |01| Examples of PDEs in fluid and solid mechanics;
Chapter |02| Notions of functional analysis: Sobolev and Hilbert spaces;
Chapter |03| Abstract form of variational problems;
Chapter |04| Examples of elliptics PDEs in variational form;
Chapter |05| The Galerkin method;
Chapter |06| Construction of Finite element spaces;
Chapter |07| Interpolation and error estimates;
Chapter |08| `FEniCSx` behind the scenes: Implementation aspects;
Chapter |09| More details on Elliptic PDEs in variational form;
Chapter |10| Stokes equations - Mixed problems;

[1] **Elliptic problems**: Stationary diffusion, heat conduction, fully developed laminar flows in ducts, linear elasticity. **Parabolic problems**: Transient diffusion or heat conduction, chemical kinetics, fluid dynamics.

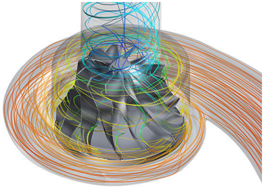
Solid mechanics



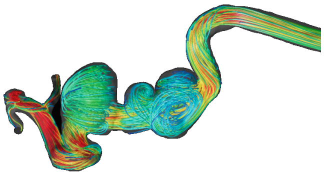
Heat transfer



Internal flows: Turbomachinery



Computational hemodynamics



External flows: Aerodynamics

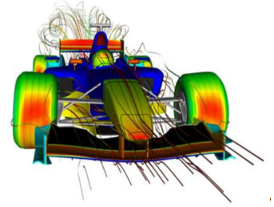


Figure 1.1: Examples solved by FEM.

1.3 Motivating examples in solid and fluid mechanics

We begin by recalling some prototypical examples of PDEs we aim to solve by the finite element method. We introduce thing with a relatively informal language a provide some physical interpretation of the different quantities and processes involved.

Figure 1.2: Siméon Denis Poisson (France, 1781–1840).

1.3.1 Poisson's equation

The simplest second order PDE we will consider in this course is Poisson's equations that models several physical phenomena, such as, heat conduction, mass transport by diffusion or even fully developed flows in ducts as we will see later on in this chapter. The problem reads: Given a region $\Omega \subset \mathbb{R}^d,\ \ d = 1, 2$ or 3 with boundary $\partial\Omega$, find u such as[2]

$$\left\{ \begin{array}{rcll} -\nabla \cdot (\mu(\mathbf{x})\nabla u(\mathbf{x})) & = & f(\mathbf{x}) & \mathbf{x} \in \Omega \\ u(\mathbf{x}) & = & g(\mathbf{x}) & \mathbf{x} \in \partial\Omega \end{array} \right. \tag{1.1}$$

where the source term $f : \Omega \longrightarrow \mathbb{R}$ is a given function and the boundary data $g : \partial\Omega \longrightarrow \mathbb{R}$ is also a given function. This problem falls into the cathegory of **elliptic** problems. The scalar field u represents different physical quantities depending on the problem. Notice that if μ is a constant, the left hand side becomes the Laplace equation, i.e.,

$$\nabla \cdot (\mu(\mathbf{x})\nabla u(\mathbf{x})) = \mu\, \nabla^2 u(\mathbf{x}) \tag{1.2}$$

Of importance to use is a variant of this problem which includes flux boundary conditions over all or in some part of $\partial\Omega$ (ommiting the $\mathbf{x}$ dependence for simplificity of notation)

Figure 1.3: Pierre-Simon Laplace (France, 1749–1827).

$$\left\{ \begin{array}{rcll} -\nabla \cdot (\mu \nabla u) & = & f & \text{in } \Omega \\ u & = & g & \text{on } \Gamma_D \\ -\mu\, \nabla u \cdot \check{\mathbf{n}} & = & h & \text{on } \Gamma_N \end{array} \right. \tag{1.3}$$

where $\partial\Omega = \Gamma_D \cup \Gamma_N$ and $\Gamma_D \cap \Gamma_N = \emptyset$. The so called mixed formulation introduces an additional field to represent the flux of the quantity of interest

$$\left\{ \begin{array}{rcll} \nabla \cdot \boldsymbol{\mathcal{J}} & = & f & \text{in } \Omega \\ \boldsymbol{\mathcal{J}} & = & -\mu\, \nabla u & \text{in } \Omega \\ u & = & g & \text{on } \Gamma_D \\ \boldsymbol{\mathcal{J}} \cdot \check{\mathbf{n}} & = & h & \text{on } \Gamma_N \end{array} \right. \tag{1.4}$$

Although both problems are equivalent in the exact setting we are focused now, in the discrete setting things may differ a lot, very different approaches being needed to deal with each formulation. The equation defining the relation between $\boldsymbol{\mathcal{J}}$ and u is what we call **constitutive law**.

[2] For $d = 2$, recall from Calculus, the *nabla* operator. In cartesian coordinates:

- The gradient of a scalar-valued function
$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right)$$
- The divergence of a vector-valued function
$$\nabla \cdot \mathbf{F} = \frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2}$$

Finally, it is instructive to interpret the different quantities according to the physical problem being solved, as summarized in table Table 1.1.

Table 1.1: Physical interpretation of quantities in Poisson's problem. Units in SI.

Physical problem	u	f	$\mathcal{J}$	μ
Heat conduction	Temperature [°K]	Heat source $[\frac{\mathrm{W}}{\mathrm{m}^3}]$	Heat flux $[\frac{\mathrm{W}}{\mathrm{m}^2}]$	Conductivity $[\frac{\mathrm{W}}{\mathrm{m}^\circ\mathrm{K}}]$
Mass diffusion	Mass $[\frac{\mathrm{mol}}{\mathrm{m}^3}]$	Mass source $[\frac{\mathrm{mol}}{\mathrm{m}^3\mathrm{s}}]$	Mass flux $[\frac{\mathrm{mol}}{\mathrm{m}^2\mathrm{s}}]$	Diffusivity $[\frac{\mathrm{m}^2}{\mathrm{s}}]$
Flow in ducts	Velocity $[\frac{\mathrm{m}}{\mathrm{s}}]$	Pressure gradient $[\frac{\mathrm{N}}{\mathrm{m}^3}]$	Shear stress [Pa]	Viscosity [Pa · s]

1.3.2 Transient heat conduction

The transient or unsteady version of the previous problems includes an additional term involving the rate of change of the physical quantity of interest. For the heat conduction problem the equation reads

$$\begin{cases} a(\mathbf{x},t)\,\dfrac{\partial u(\mathbf{x},t)}{\partial t} - \nabla\cdot(\mu(\mathbf{x},t)\nabla u(\mathbf{x},t)) & = & f(\mathbf{x},t) & \mathbf{x}\in\Omega,\ \ t\in[0,T] \\ u(\mathbf{x},t) & = & g(\mathbf{x},t) & \mathbf{x}\in\partial\Omega,\ \ t\in[0,T] \\ u(\mathbf{x},0) & = & u_0(\mathbf{x}) & \mathbf{x}\in\Omega \end{cases} \tag{1.5}$$

where now an initial condition for the scalar unknown (e.g. the temperature) has been provided. This problem falls into the cathegory of **parabolic** problems. The parameter a in front of the time derivative has different meanings depending on the problem at hand (see Table 1.2)

The different forms of this problem, such as the one with a flux boundary condition or the mixed formulation can be formulated in a similar way as in the stationary case.

Table 1.2: Physical interpretation of the a factor in front of $\partial_t u$ in the transient Poisson's problem. Units in SI.

Problem	Factor a	Description
Heat conduction	$\rho\, c$	$[\frac{\mathrm{Kg}}{\mathrm{m}^3}\ \frac{\mathrm{J}}{\mathrm{Kg}\,^\circ\mathrm{K}}]$
Mass diffusion	1	[-]
Flow in ducts	ρ	$[\frac{\mathrm{Kg}}{\mathrm{m}^3}]$

1.3.3 The equations of linear elasticity

This is the prototypical example in solid mechanics in which we describe the deformation of a solid domain. If we consider a domain $\Omega\subset\mathbb{R}^d$ its shape is defined by a map $\chi:\Omega\longrightarrow\mathbb{R}^d$, such that for any $\mathbf{x}\in\Omega$ we write

$$\chi(\mathbf{x},t)=\mathbf{x}+\mathbf{u}(\mathbf{x},t) \tag{1.6}$$

where the vector field $\mathbf{u}:\Omega\longrightarrow\mathbb{R}^d$ represents the displacement at the fixed location in space $\mathbf{x}$. The PDE governing the problem follows from

Newton's dynamical equilibrium equations:

$$\underbrace{\rho(\mathbf{x})\frac{\partial^2\mathbf{u}(\mathbf{x},t)}{\partial t^2}}_{\text{mass} \times \text{acceleration}} - \underbrace{\nabla\cdot\boldsymbol{\sigma}(\mathbf{x},t) = \mathbf{f}(\mathbf{x},t)}_{\text{Forces}} \tag{1.7}$$

where the second order time derivative in the left hand side corresponds to the acceleration and **f** in the right hand side to the body forces (e.g., gravity), $\boldsymbol{\sigma}(\mathbf{x},t)$ defines the stresses in the body at location **x**, which under the small deformation assumption is given by the **constitutive law**

$$\sigma = \underbrace{\mu\,(\nabla\mathbf{u} + \nabla^{\mathsf{T}}\mathbf{u})}_{2\mu\,\varepsilon(\mathbf{u})} + \lambda(\nabla\cdot\mathbf{u})\,\mathbf{I} \tag{1.8}$$

where λ and μ are known material parameters and **I** is the identity matrix of $d \times d$[3]. In the stationary case this is an elliptic problem in which the unknown field is a vector-valued function. In order to have a well-posed problem the displacement must be restricted in some part of the boundary (say, $\Gamma_{\mathbf{u}}$) so as to eliminate rigid body motions (i.e., translations and rotations). Also, a surface force distribution can be applied on the rest of the boundary $\Gamma_{\mathscr{F}} = \partial\Omega \setminus \Gamma_{\mathbf{u}}$.

$$\left\{\begin{array}{rcll} -\nabla\cdot(2\mu\varepsilon(\mathbf{u}) + \lambda\,(\nabla\cdot\mathbf{u})\,\mathbf{I}) & = & \mathbf{f} & \text{in}\,\Omega \\ \mathbf{u} & = & \mathbf{u}_D & \text{on}\,\Gamma_{\mathbf{u}} \\ (2\mu\varepsilon(\mathbf{u}) + \lambda\,(\nabla\cdot\mathbf{u})\,\mathbf{I})\cdot\check{\mathbf{n}} & = & \mathscr{F} & \text{on}\,\Gamma_{\mathscr{F}} \end{array}\right. \tag{1.9}$$

where $\mathbf{u}_D$ and $\mathscr{F}$ are given functions.

1.3.4 The incompressible Navier-Stokes problem

A natural extension of the previous problem are the Navier-Stokes equations. Now, we have two primary variables, namely, the velocity field $\mathbf{u}(\mathbf{x},t)$ and the pressure field $p(\mathbf{x},t)$. We consider an Eulerian formulatio, so, **u** is the velocity at fixed position **x** in space. Also, we restrict ourselves to the particular case of incompressible fields, i.e.,

$$\nabla\cdot\mathbf{u} = \frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3} = 0 \tag{1.10}$$

The stresses in the fluid are described by the Cauchy stress tensor

$$\sigma(\mathbf{x},t) = -p(\mathbf{x},t)\mathbf{I} + \sigma^*(\mathbf{x},t) \tag{1.11}$$

which is the sum of a volumetric part

$$-p(\mathbf{x},t)\mathbf{I} = \begin{bmatrix} -p & 0 & 0 \\ 0 & -p & 0 \\ 0 & 0 & -p \end{bmatrix} \tag{1.12}$$

and a deviatoric part

$$\sigma^*(\mathbf{x},t) = 2\,\mu\varepsilon(\mathbf{u}) = \mu\,(\nabla\mathbf{u} + \nabla^{\mathsf{T}}\mathbf{u}) \tag{1.13}$$

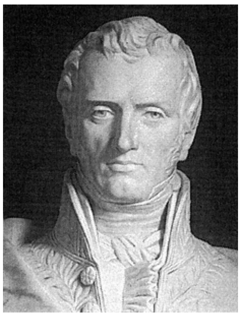


Figure 1.4: Claude Louis Marie Henri Navier (France, 1785–1836).

Figure 1.5: George Stokes (Ireland(1819)–England(1903)).

[3] For $d = 2$, in cartesian coordinates

- The stress tensor is the matrix

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

- The divergence of $\boldsymbol{\sigma}$ is the vector

$$\nabla\cdot\boldsymbol{\sigma} = \begin{bmatrix} \frac{\partial\sigma_{11}}{\partial x_1} + \frac{\partial\sigma_{12}}{\partial x_2} \\ \frac{\partial\sigma_{21}}{\partial x_1} + \frac{\partial\sigma_{22}}{\partial x_2} \end{bmatrix}$$

- The gradient of the vector field **u** is the matrix

$$\nabla\mathbf{u} = \begin{bmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} \end{bmatrix}$$

where μ is the viscosity of the fluid. Now, we write the momentum equation in the so called convective form

$$\rho \frac{D\mathbf{u}}{Dt} - \nabla \cdot \boldsymbol{\sigma} = \rho \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) - \nabla \cdot (2\,\mu \varepsilon(\mathbf{u})) + \nabla p = \mathbf{f} \tag{1.14}$$

where the first (non-linear) term introduces the convective acceleration, the second term the viscous effects and the third term the forces due to pressure gradients. As in the previous case, in the left hand side we have the body forces (e.g., $\mathbf{f} = -\rho\, g \check{\mathbf{e}}_3$). Given an initial velocity field $\mathbf{u}_0$ that satisfies the incompressibility constraint, the Navier-Stokes problem reads

$$\begin{cases} \rho \left(\dfrac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) - \nabla \cdot (2\,\mu \varepsilon(\mathbf{u})) + \nabla p = \mathbf{f} & \text{in } \Omega \\ \nabla \cdot \mathbf{u} = 0 & \text{in } \Omega \\ \mathbf{u} = \mathbf{u}_D & \text{on } \Gamma_D \\ [-p\mathbf{I} + 2\,\mu\, \varepsilon(\mathbf{u})] \cdot \check{\mathbf{n}} = \mathcal{F} & \text{on } \Gamma_N \end{cases} \tag{1.15}$$

where $\mathbf{u}_D$ is a given function on Γ_D and $\mathcal{F}$ is a given function on Γ_N.

Prior to jumping into our first Miniproject/Assignment, let solve a few exercices that will help us to get used with the previous problems:

Exercices

- Check the units of quantities in Table 1.1.
- Consider the Poisson's equation with $f = -6$ and $\mu = 1$ in the domain $\Omega = [0,1]^2$. Verify that

  $$u(x_1, x_2) = 1 + x_1^2 + 2\,x_2^2 \tag{1.16}$$

  is a solution. Which boundary conditions this solution satisfies?
- In the 2D case, consider a scalar function ψ such that

  $$u_1 = -\frac{\partial \psi}{\partial x_2}, \quad u_2 = \frac{\partial \psi}{\partial x_1} \tag{1.17}$$

  (ψ is called the stream function). Show that $\nabla \cdot \mathbf{u} = 0$.
- In the previous problem, compute the stress tensor $\boldsymbol{\sigma}$.
- In the Navier-Stokes problem in 3D, write *in extensum* the 3 momentum equations corresponding to each component.
- Consider the following **manufactured** solution in 2D for the Navier-Stokes equations:

  $$\begin{aligned} \mathbf{u}(x_1, x_2, t) &= [\sin(x_1)\sin(x_2 + t),\ \cos(x_1)\cos(x_2 + t)]^{\intercal}\,, \\ p(x_1, x_2, t) &= \cos(x_1)\sin(x_2 + t), \end{aligned}$$

Compute to which body force $\mathbf{f}$ and boundary conditions it corresponds.

- Write the nondimensional form of the Navier-Stokes equations in terms of nondimendional variables $\hat{\mathbf{u}}, \hat{\mathbf{x}}, \hat{p}, \hat{t}, \hat{\mathbf{f}}$, the parameter and Re= $\dfrac{\rho U L}{\mu}$ (a.k.a. the Reynolds number).

1.4 Assignment 1: Poisson's problem for fully developed flow

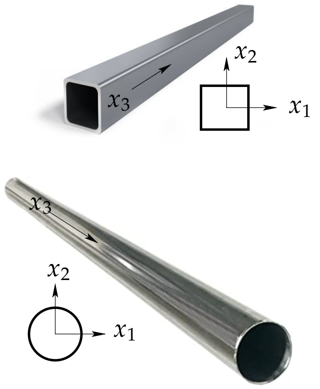


Figure 1.6: Domain for the fully developed flow.

Fully developed flow in ducts is a particular case of flow governed by the Navier-Stokes equations, which corresponds to the following situation (see Figure 1.6):

- Incompressible flow along a long cylinder of cross section $\Omega \subset \mathbb{R}^2$. The flow domain is $\mathcal{B} = \Omega \times (0, L)$.
- The flow is driven by a uniform pressure gradient

$$\mathcal{G}(t) = \frac{p(L,t) - p(0)}{L} \tag{1.18}$$

  The pressure field is thus linear as a function of x_3. Also, notice that when $\mathcal{G}(t) > 0$ we expect $u_3 < 0$ and viceversa.

- If L is sufficiently large, the entry and exit effects can be neglected and all cross sections are essentially identical, except for the pressure.
- We consider as solution, velocity fields of the form:

$$\mathbf{u} = [0, 0, u(x_1, x_2, t)]^\intercal \tag{1.19}$$

- For any differentiable u, the proposed velocity is divergence free, i.e.,

$$\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3} = \frac{\partial u}{\partial x_3} = 0 \tag{1.20}$$

- Decomposing the stress tensor in pressure and non-pressure components, we have

$$\boldsymbol{\sigma}(x_1, x_2, x_3, t) = -p(x_3, t)\,\mathbf{I} + \boldsymbol{\sigma}^*(x_1, x_2, t)\,. \tag{1.21}$$

  where, if the fluid is Newtonian ,

$$\boldsymbol{\sigma}^* = \mu \begin{pmatrix} 0 & 0 & u_{,1} \\ 0 & 0 & u_{,2} \\ u_{,1} & u_{,2} & 0 \end{pmatrix} \qquad \Rightarrow \qquad \boldsymbol{\tau} = \mu \nabla u\,, \tag{1.22}$$

  being $\boldsymbol{\tau}$ a vector having the shear stresses as its components.

- By inserting $\mathbf{u}$, p and $\boldsymbol{\sigma}^*$ into the momentum balance of Equation

1.15, we finally arrive

$$\rho\frac{\partial u}{\partial t} + \mathcal{G}(t) - \nabla \cdot (\mu \nabla u) = 0 \quad \forall \mathbf{x} \in \Omega \,, \tag{1.23}$$

- Finally, for the boundary conditions we assume the standard hypothesis of **non-slip**, which implies

  $$u = 0 \quad \forall \mathbf{x} \in \partial\Omega \tag{1.24}$$

  This means that the fluid is attached to the walls of the domain, which is a physical observation valid in most of the cases[4].

- We conclude by noticing that this problem is mathematically identical to the transient heat conduction problem presented above, and in the stationary case ($\partial_t u = 0$) to the Poisson's problem. Physically, the quantity being diffused is linear momentum.

1.4.1 The variational formulation

We will now proceed quite informally and postpone some details for later chapters. The idea first is to cast this problem into variational form, which amounts to multiply by a sufficiently regular **test** function v that satifies the boundary condition, i.e.,

$$v(\mathbf{x}) = 0 \quad \forall \mathbf{x} \in \partial\Omega \tag{1.25}$$

and integrate over Ω. Let us consider for simplicity the stationary case, and call $f = -\mathcal{G}$,

$$-\int_\Omega \nabla \cdot (\mu(\mathbf{x})\, \nabla u(\mathbf{x}))\; v(\mathbf{x})\, dx = \int_\Omega f\, v(\mathbf{x})\, dx \tag{1.26}$$

From our Calculus course we know that (ommiting the $\mathbf{x}$ dependence to simplify notation)

$$\nabla \cdot (\mu \nabla u\, v) = \nabla \cdot (\mu \nabla u)\; v + \mu \nabla u \cdot \nabla u$$

Figure 1.7: Carl Friedrich Gauss (Germany, 1777–1855).

We thus have

$$\int_\Omega \mu\, \nabla u \cdot \nabla v\, dx - \int_\Omega \nabla \cdot ((\mu \nabla u)\, v)\, dx = \int_\Omega f\, v\, dx \tag{1.27}$$

For the second term in the left hand side we apply Gauss theorem

$$\int_\Omega \nabla \cdot ((\mu \nabla u)\, v)\, dx = \int_{\partial\Omega} v\, (\mu \nabla u) \cdot \check{\mathbf{n}}\, ds \tag{1.28}$$

which is identically zero by Equation 1.25, yielding[5]

$$\int_\Omega \mu\, \nabla u \cdot \nabla v\, dx = \int_\Omega f\, v\, dx \tag{1.29}$$

Let us agree that this should be valid for any v belonging to some space of functions $v : \Omega \longrightarrow \mathbb{R}$ (say $V(\Omega)$) that satisfy the homogeneous boundary

---

[4] More general conditions, such as **slip** conditions observed in rarefied flows (typically, gas flows at very small scale) can be considered as well.

[5] What we have done is also called integration by parts. Also notice that the boundary of Ω must have some regularity for the Gauss theorem to be applicable.

condition at $\partial\Omega$, for which the integrals in the left and right hand sides at least don't blow up. The precise definitions and details will come soon. We may formally write what we call from now on the weak[6] form of the problem

Weak form of Poisson's problem

Find $u \in V(\Omega)$ such that

$$\begin{cases} \underbrace{\int_\Omega \mu(\mathbf{x})\,\nabla u(\mathbf{x}) \cdot \nabla v(\mathbf{x})\,dx}_{a(u,v)} = \underbrace{\int_\Omega f\,v(\mathbf{x})\,dx}_{\ell(v)} \\ \forall v \in V(\Omega). \end{cases} \tag{1.30}$$

What appears in the lhs is a bilinear form $a(\cdot,\cdot)$, whereas in the rhs we have a linear form $\ell(\cdot)$, two objects we will encounter frequently in the course.

Now, we will solve this problem numerically using the `FEniCS` platform. In the next section I will provide some guidelines for you to implement in few lines of code a Finite Element solver so as to familiarize with the main features of the project.

1.4.2 The `FEniCSx` implementation

You need first to install the library. I suggest to use the `Docker containers` as explained in `FEniCSx Tutorial`. Basically, you must run in a terminal (assuming the `Docker` is installed):

```
> docker run -ti dolfinx/dolfinx:latest  # To install the first time
> python3 -c "import dolfinx" # Check installation
> exit
> cd to_your_working_directory/
> docker run -ti -v $(pwd):/root/shared -w /root/shared  dolfinx/dolfinx
> python3 basic_poisson.py # Script provided
```

The implementation follows the next script :

- Import some libraries

```
from dolfinx import fem, io, mesh
import ufl
from ufl import dx, ds, grad, inner, FacetNormal
...
```

- Construct a partition of the domain Ω, in this case the rectangle $[0,1] \times [0,1]$, made of triangles. This is called a mesh

```
msh = mesh.create_rectangle(comm=MPI.COMM_WORLD,
                            points=((0.0, 0.0), (1.0, 1.0)),
                            n=(16, 16),
                            cell_type=mesh.CellType.triangle)
```

[6] The problem is said to be in **weak** form because in this formulation we require less differentiability from u in contrast to the PDEs in **strong** form for which we require u to have at least continuous second order derivatives if we pretend the equation to be satisfied pointwise.

- Associated to this mesh we construct a discrete space $V_d \subset V(\Omega)$, in this case, piecewise Lagrangian polynomials of degree 1 defined over the triangles of the mesh,

```
Vd = fem.FunctionSpace(msh, ("Lagrange", 1))
```

- Identify the facets that form $\partial\Omega$ and imposing a uniform Dirichlet boundary condition (this piece of code is perhaps the most difficult to understand)

```
tdim = msh.topology.dim
fdim = tdim - 1
msh.topology.create_connectivity(fdim, tdim)
facets=np.flatnonzero(mesh.compute_boundary_facets(msh.topology))
dofs = fem.locate_dofs_topological(Vd, fdim, facets)
u_boundary = fem.Constant(msh, ScalarType(0.0))
bc = fem.dirichletbc(u_boundary, dofs, Vd)
```

- Write the discrete version of Equation 1.30 by using the so called Galerkin method: Find $u_d \in V_d$, such that

$$\int_\Omega \mu \nabla u_d \cdot \nabla v \, dx = \int_\Omega f \, v \, dx \quad \forall v \in V_d. \tag{1.31}$$

We need to express this weak formulation in the `UFL` language

```
u = ufl.TrialFunction(Vd)
v = ufl.TestFunction(Vd)
f  = fem.Constant(msh, ScalarType(1.0))
mu = fem.Constant(msh, ScalarType(1.0))
a = inner(mu*grad(u), grad(v)) * dx
L = inner(f, v) * dx
```

- Solve the problem by assemblying an associated linear system and using a linear algebra package such as `Petsc`

```
opts = {"ksp_type": "preonly", "pc_type": "lu"}
prob = fem.petsc.LinearProblem(a, L, bcs=[bc], petsc_options=opts)
ud   = prob.solve()
```

- Save the function u_d to a file and visualize for instance in `Paraview` (see Figure 1.8).

Implement the following modifications to `basic_poisson.py`

- Inspect and run the script provided and make sure you understand its different sections.

- Implement a viscosity coefficient μ that depends on $\mathbf{x}$ according to

$$\mu(\mathbf{x}) = 0.01 + e^{-100\left[(x-\frac{1}{2})^2+(y-\frac{1}{2})^2\right]}$$

To retrieve the mesh coordinates you will need to add the following line to the script

```
x = ufl.SpatialCoordinate(msh)
```

Notice that the x and y coordinates corresponds respectively to `x[0]` and `x[1]`.

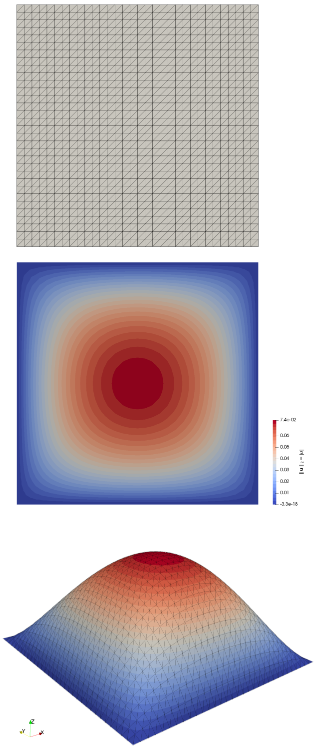


Figure 1.8: Contours of velocity magnitude $\|\mathbf{u}\|_2$ and finite element partition for the fully developed flow in a duct of square cross section.

- Solve the problem by considering an increasing number of elements in `mesh.create_rectangle()` so as to better capture the localized behavior of $\mu(\mathbf{x})$ (take e.g., 16×16, 32×32, . . . 128×128). Consider also the case of elements of quadrangular shape by setting `cell_type=mesh.CellType.quadrilateral`.

- By following the guidelines in FEniCS Tutorial solve the problem in domains Ω of circular and triangular shapes.

- (**Optional**) Compute the shear stresses $\boldsymbol{\tau} = \mu \nabla u$. Plot the norm $\|\boldsymbol{\tau}\|_2 = \sqrt{\boldsymbol{\tau} \cdot \boldsymbol{\tau}}$ as a scalar field. You will need to interpolate $\|\boldsymbol{\tau}\|_2$ into a space of elementwise constant functions.

  ```
  Vstress = fem.FunctionSpace(msh, ("DG", 0))
  ```

  Compute the wall shear stresses $\boldsymbol{\sigma}^* \cdot \check{\mathbf{n}} = (\boldsymbol{\tau} \cdot \check{\mathbf{n}})\check{\mathbf{e}}_3$, where $\check{\mathbf{e}}_3$ is the unit vector in the z–direction, by integrating over the boundary

  $$W_{ss} = \int_{\partial\Omega} \boldsymbol{\tau} \cdot \check{\mathbf{n}} \, ds$$

  ```
  tau = mu*ufl.grad(ud)
  n = FacetNormal(msh)
  Wss = fem.form(ufl.inner(tau,n)*ds)
  print(fem.assemble_scalar(Wss))
  ```

- Prepare a short presentation to report the results.

SOBOLEV SPACES 2

Prior to proceed with the theory and some mathematical aspects of the finite element method we need to recall a minimum set of tools of functional analysis, since we will be dealing with function spaces of infinite dimension. This is essential because the finite element methods is formulated in finite-dimensional subspaces of so called Sobolev spaces. A complete reference for the topics to be covered in this chapter can be found for instance in the Book by Reddy [4] (which is quite accessible and didactic) and the book by Brenner and Scott [2].

2.1 Function spaces


Figure 2.1: Sergei Sobolev (Russia, 1908–1989).

In this course we consider **vector** (or linear) spaces made up of functions. Recall, a vector space is said to be n–dimensional if we can find n linearly independent elements, but $n+1$ elements are linearly dependent. This is what we tipically study in linear algebra courses. Now, suppose that n linearly independent elements can be found in the space for every n, then, the space is said to be **infinite dimensional**.

Let us consider a space V over the field $\mathbb{R}$ and let $\Omega \subset \mathbb{R}^d$, $d = 1, 2$ or 3 be an open and bounded domain. If f, g are functions (or elements) of V, i.e., $f, g : \Omega \to \mathbb{R}$, then for any $x \in \Omega$, we have

- $f + g \in V$: $(f+g)(x) = f(x) + g(x)$
- $\alpha f \in V$: $(\alpha f)(x) = \alpha f(x)$ for any scalar $\alpha \in \mathbb{R}$.

In the sequel we deal with the concept of **continuity**, **differentiability** and **integrability** of functions. An important example is the space $C^k(\Omega)$ defined as the set of all real-valued functions that are continuous and have its partial derivatives up to order k also continuous on Ω, particular cases being

$$C^0(\Omega) = \{f : \Omega \to \mathbb{R} : f \text{ is continuous } \forall x \in \Omega\} \doteq C(\Omega) \tag{2.1}$$

and

$$C^\infty(\Omega) = \{f : \Omega \to \mathbb{R} : f \text{ has continuous derivatives of all orders } \forall x \in \Omega\} \tag{2.2}$$

Notice that $C^\infty(\Omega) \subset \cdots \subset C^k(\Omega) \subset \cdots \subset C^0(\Omega) = C(\Omega)$.

Another example that we will encounter frequently in this course is the space of square integrable functions

$$L^2(\Omega) = \{f : \Omega \to \mathbb{R} : \int_\Omega f^2\, dx < +\infty\} \tag{2.3}$$

and the space of square integrable functions whose derivatives are also square integrable

$$H^1(\Omega) = \{f : \Omega \to \mathbb{R} : f \in L^2(\Omega) \text{ and } \int_\Omega (\nabla f \cdot \nabla f)\, dx < +\infty\} \quad (2.4)$$

which is quite relevant to us [1].

As a final example consider a subspace of the previous, made up of functions that are equal to zero on the boundary $\Gamma = \partial\Omega$

$$H_0^1(\Omega) = \{f \in H^1(\Omega) : f|_\Gamma = 0\}, \quad (2.5)$$

This one is also of particular interest to us, since it incorporates a **boundary condition**[2]. We will encounter these spaces frequently throughout.

Now, we need to equippe these spaces with the notion of a norm to determine "how large is a given function f" or "how close are two functions f and g".

2.2 Norms and inner products

Definition 2.2.1 *(**Norm**): Given a vector space V over the field $\mathbb{R}$, a norm in V is a function $\| \cdot \|: V \to \mathbb{R}$ with the following properties*

1. $\| v \| \geq 0$, $\| v \| = 0 \Leftrightarrow v = 0$;
2. $\| \alpha v \| = |\alpha| \, \| v \|$ *for any scalar* $\alpha \in \mathbb{R}$;
3. $\| v + w \| \leq \| v \| + \| w \|$ *(**Triangle inequality**).*

Note that a norm induces a distance (or a **metric**) by the formula

$$d(v, w) = \|v - w\| \quad (2.6)$$

In finite–dimensional spaces all norms are equivalent, i.e., if $\| \cdot \|_a$ and $\| \cdot \|_b$ are two different norms over a linear space V, $\exists c_1, c_2$ such that

$$c_1 \|v\|_b \leq \|v\|_a \leq c_2 \|v\|_b, \forall v \in V \quad (2.7)$$

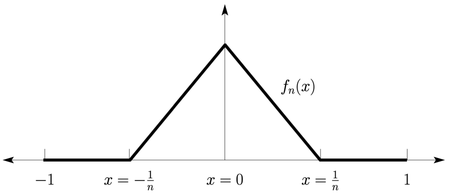


Figure 2.2: Example of a sequence of functions $f_n(x)$ to illustrate that the $L^\infty(\Omega)$ and $L^2(\Omega)$–norms are not equivalent.

this means that, if we have a sequence of functions $\{f_n\}_{n=1}^\infty$, if $f_n(x) \xrightarrow[\|\cdot\|_a]{} F \Rightarrow f_n \xrightarrow[\|\cdot\|_b]{} F$. However, with infinite dimensional spaces this does not necessarily hold. Consider for instance the space of continuous functions $C^0(\Omega)$ being $\Omega = [-1, 1]$, the sequence of functions[3] shown in figure Figure 2.2 and the norms

$$\begin{aligned} \| f_n \|_{L^\infty(\Omega)} &= \sup\{|f_n(x)| : x \in \Omega\} = 1 \;\forall n && (2.8) \\ \| f_n \|_{L^2(\Omega)} &= \left(\int_\Omega [f_n(x)]^2 \, dx\right)^{\frac{1}{2}} = \left(\frac{2}{3n}\right)^{\frac{1}{2}} \xrightarrow[n\to\infty]{} 0 && (2.9) \end{aligned}$$

Let us recall that continuous functions on a compact[4] domain are automatically bounded, therefore the supremum norm $\| \cdot \|_{L^\infty(\Omega)}$, is a natural

[1] In Equation 2.4, since $\nabla f : \Omega \to \mathbb{R}^d$ (is a vector–valued function), the symbol "$\cdot$" stands for the inner (or usual "dot") product of vectors, i.e., $\nabla f(x) \cdot \nabla f(x) = \sum_{i=1}^{d} (\nabla f)_i(x) \, (\nabla f)_i(x) = \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(x) \, \frac{\partial f}{\partial x_i}(x)$.

[2] We use the notation $f|_\Gamma$ to denote the restriction of a function f to the boundary Γ, i.e., the values of $f(x)$, $x \in \Gamma$. To do this properly, we must introduce the **trace** operator $\gamma : C(\bar{\Omega}) \to C(\Gamma)$, $\gamma(f) = f|_\Gamma$.

[3] The expression of f_n, $n = 1, \ldots,$ is

$$f_n(x) = \begin{cases} 1 - |x|\, n & \text{if } |x| < \frac{1}{n} \\ 0 & \text{otherwise} \end{cases}$$

[4] For any subset $\Omega \subset \mathbb{R}^d$, Ω is compact $\Leftrightarrow$ it is closed and bounded, i.e., it contains all its limit points.

norm to equippe $C^0(\Omega)$. We have also introduced above the $\|\cdot\|_{L^2(\Omega)}$ norm which is a natural choice for the space of square integrable functions $L^2(\Omega)$.

Definition 2.2.2 *(**Inner product**): Given a vector space V over the field $\mathbb{R}$, an inner product is a function that takes two arguments $(\cdot,\cdot) : V \times V \longrightarrow \mathbb{R}$ with the following properties*

1. $(v, w) = (w, v)$ *(**Symmetry**);*
2. $(\alpha v + \beta w, u) = \alpha(v, u) + \beta(w, u)$ *(**Linearity**);*
3. $(v, v) \geq 0$, $(v, v) = 0 \Leftrightarrow v = 0$ *(**Positivity**)*

Note that an inner product is a bilinear form. Also, an important fact is that an inner product induces a norm by the formula

$$\|v\| = \sqrt{(v, v)} \tag{2.10}$$

and, therefore a **metric**.

2.3 Banach, Hilbert and Lebesgue spaces

Figure 2.3: Stefan Banach (Kraków(1892)–Ukraine(1945)).

Figure 2.4: David Hilbert (Germany, 1862–1943).

Figure 2.5: Henri Lebesgue (France, 1875–1941).

In the theory of PDEs an important role is played by the Banach spaces and in particular by the Hilbert spaces which are central to this course.

Definition 2.3.1 *(**Banach space**): A Banach space V is a vector space equipped with a norm $\|\cdot\|_V$, which is also complete with respect to that norm, i.e., all Cauchy sequences in V converge to an element $v \in V$.*

so, if $\{v_n\}_{n=1,2,\ldots}$ is a Cauchy sequence of functions $v_n \in V$, $\|v_n - v\|_V \xrightarrow[n\to\infty]{} 0$, $v \in V$, but, what is a Cauchy sequence?

Definition 2.3.2 *(**Cauchy sequence**): A sequence of elements $v_1, v_2, v_3, \ldots$ is called a Cauchy sequence (or a fundamental sequence) if for every $\epsilon > 0$, exists a positive integer N such that for all $m, n > N$, $\| v_n - v_m \| < \epsilon$.*

i.e., a sequence in which its elements get arbitrarily close to each other. It is instructive to consider the following example. Take the space $C^0(\Omega)$, $\Omega = [0, 1]$ equipped with the $L^2(\Omega)$-norm, i.e., $\|v\|_{L^2(\Omega)} = \sqrt{\int_0^1 v^2\,dx}$ and the Cauchy sequence of functions $\{v_n\}$ defined by (see Figure 2.6)

$$v_n = \begin{cases} 0 & \text{if } 0 < x < \frac{1}{2} \\ (x - \frac{1}{2})^{\frac{1}{n}} & \text{if } \frac{1}{2} \leq x \leq 1 \end{cases} \tag{2.11}$$

whose limit as $n \longrightarrow \infty$ is

$$v = \begin{cases} 0 & \text{if } 0 < x < \frac{1}{2} \\ 1 & \text{if } \frac{1}{2} \leq x \leq 1 \end{cases} \tag{2.12}$$

It can be seen that in the L^2-norm $v_n \xrightarrow[n\to\infty]{} v$ (check that), however notice that v is dicontinuous. We have found an example of a Cauchy sequence in $C^0(\Omega)$ converging to an element not belonging to $C^0(\Omega)$, so the space is not complete with respect to the $L^2(\Omega)$-norm.

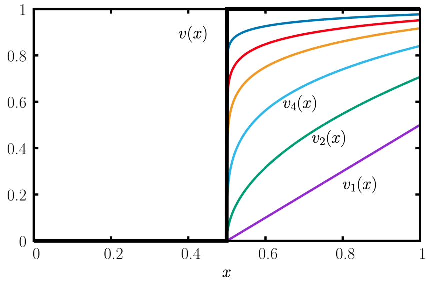


Figure 2.6: Example of the sequence of functions $v_n(x)$ tending towards the step function.

Let us now define an important class of spaces for this course:

Definition 2.3.3 *(**Hilbert space**): A Hilbert space is a vector space equipped with an inner product, that is also a complete metric space with respect to the norm induced by the inner product.*

so, a Hilbert space is also a Banach space. Important examples to us of inner products are the $L^2(\Omega)$-inner product

$$(v, w)_{L^2(\Omega)} = \int_\Omega v\, w\, dx \tag{2.13}$$

and the $H^1(\Omega)$-inner product

$$(v, w)_{H^1(\Omega)} = \int_\Omega (v\, w + \nabla v \cdot \nabla w)\, dx \tag{2.14}$$

which induces the previously mentioned norms. Notice also that if the functions are vector-valued, i.e., $\mathbf{v}, \mathbf{w} : \Omega \longrightarrow \mathbb{R}^d$, $d = 2$ or 3, we write[5]

$$(\mathbf{v}, \mathbf{w})_{H^1(\Omega)} = \int_\Omega (\mathbf{v} \cdot \mathbf{w} + \nabla \mathbf{v} : \nabla \mathbf{w})\, dx \tag{2.15}$$

Finally, let us consider the so called Lebesgue spaces:

Definition 2.3.4 *(**Lebesgue space**): The Lebesgue space of order $p \in [1, \infty)$ is defined as*

$$L^p(\Omega) = \{f : \Omega \longrightarrow \mathbb{R} : \int_\Omega |f|^p\, dx < +\infty\} \tag{2.16}$$

For $p = \infty$

$$L^\infty(\Omega) = \{f : \Omega \longrightarrow \mathbb{R} \ : \ \|f\|_{L^\infty(\Omega)} < +\infty\} \tag{2.17}$$

The natural norm for these spaces is the $L^p(\Omega)$–norm

$$\|f\|_{L^p(\Omega)} = \left(\int_\Omega |f|^p\, dx\right)^{\frac{1}{p}}. \tag{2.18}$$

A few comments are in order:

- $L^p(\Omega)$ are Banach spaces;
- $L^2(\Omega)$ (already introduced in Equation 2.3) is a Hilbert space;
- $L^p(\Omega) \subset L^q(\Omega)\ \forall q < p$ (e.g., $q = 1,\ p = 2$, of all functions that are integrable in Ω, just some of them are square integrable, and so on);

Figure 2.7: Bernhard Riemann (Hanover(1826)–Italy(1866)).

---

[5] In Equation 2.15, since $\nabla \mathbf{v} : \Omega \longrightarrow \mathbb{R}^{d \times d}$ (is a matrix–valued function), the symbol ":" stands for the inner product of tensors or the double contraction, i.e., $\nabla \mathbf{v} : \nabla \mathbf{w} = \sum_{i,j=1}^{d} (\nabla \mathbf{v})_{ij}\, (\nabla \mathbf{w})_{ij}$.

- For purely technical reasons, we need to introduce the notion of **Lebesgue integration**, which is a generalization of the **Riemann integration** we have learned in elementary courses. All the theory is rigorously built on top of that. Lebesgue integration is able to handle more general functions, but, it coincides with the Riemann integral, in case the latter exists;

- The spaces L^p define equivalence classes of functions. This means, two functions f and g are said to be equivalent if they differ only on a set of zero measure[6]. Then, if $f = g$ almost everywhere ($f \neq g$ only on a set of zero measure)

$$\int_\Omega |f - g|^p \, dx = 0 \tag{2.19}$$

We are not done yet, a fundamental concept is finally needed to complete the picture:

2.4 Weak derivatives and Sobolev spaces

Consider the one dimensional function $f(x) = 1 - |x|$ in the interval $[-1, 1]$. This is a continuous function whose first derivative is not defined in the classical sense in $x = 0$, however, we intuitively define its derivative as a function $g(x)$ given by

$$g(x) = \begin{cases} 1 & \text{if } x < 0 \\ -1 & \text{if } x > 0 \end{cases} \tag{2.20}$$

so, this function if **piecewise continuous differentiable**[7]. Also, note that $g \in L^2(-1, 1)$. Now, take a function $\phi(x)$, smooth everywhere in $\Omega = [-1, 1]$ and vanishing outside Ω, and proceed as follows

$$\int_{-1}^{1} f(x)\,\phi'(x)\,dx = \int_{-1}^{0^-} (1 + x)\,\phi'(x)\,dx + \int_{0^+}^{1} (1 - x)\,\phi'(x)\,dx =$$
$$= \underbrace{[(1 + x)\,\phi]_{-1}^{0^-} + [(1 - x)\,\phi]_{0^+}^{1}}_{=0} - \int_{-1}^{0^-} 1\,\phi(x)\,dx - \int_{0^+}^{1} (-1)\,\phi(x)\,dx =$$
$$= -\int_{-1}^{1} g(x)\,\phi(x)\,dx \tag{2.21}$$

where we have applied the integration by parts formula we have learned in calculus. The function g is called the weak derivative of f. In order to formalize this, let us first define the space

$$\mathcal{D}(\Omega) = \{\phi \in C^\infty(\Omega) : \phi \text{ has compact support in } \Omega\} \tag{2.22}$$

Compact support in Ω means that ϕ in particular, vanishes on the boundary of Ω, i.e., the set of points in which $\phi \neq 0$ is closed and bounded (**compact**), as shown in figure Figure 2.8.

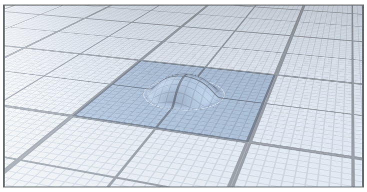


Figure 2.8: Example of a function with compact support (the Bump function). By JoshDif - Own work, CC BY-SA 4.0, `https://commons.wikimedia.org/w/index.php?curid=67554863`

Definition 2.4.1 *(**Weak derivative**): A given function $f : \Omega \longrightarrow \mathbb{R}$ is said to have weak derivative with respect to x_i, $D_w^i f$ if there exists a function g*

---

[6] 6: A set of zero measure means: a set of isolated points in 1D (no length), a line in 2D (no area) and a surface in 3D (no volume).

[7] 7: In the finite element method we approximate the solution of PDEs by using this type of functions most part of the time.

such that

$$\int_\Omega g(x)\,\phi(x)\,dx = -\int_\Omega f(x)\,\frac{\partial \phi}{\partial x_i}(x)\,dx \ \ \forall \phi \in \mathcal{D}(\Omega) \tag{2.23}$$

If such g exists, we define $D^i_w f = \frac{\partial f}{\partial x_i} = g$.

The new definition of derivative is the same as the classical $\lim\limits_{h\to 0} \frac{f(x+h)-f(x)}{h}$, if f is regular enough at x.

The definition can be extended to higher order derivatives. Given a function $f : \Omega \longrightarrow \mathbb{R}^n$ and a multi-index $\vec{\alpha} = (\alpha_1, \dots, \alpha_n)$, $|\vec{\alpha}| = \alpha_1 + \dots + \alpha_n$, $D^{\vec{\alpha}}_w f = \dfrac{\partial^{|\vec{\alpha}|} f}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}}$ is the weak derivative of f of order $|\vec{\alpha}|$, if there exists a function g such that

$$\int_\Omega g(x)\,\phi(x)\,dx = (-1)^{|\vec{\alpha}|} \int_\Omega f(x)\,\phi^{(\vec{\alpha})}(x)\,dx \ \ \forall \phi \in \mathcal{D}(\Omega) \tag{2.24}$$

We are now ready to define the so called Sobolev spaces:

Definition 2.4.2 *The Sobolev space $W^{m,p}(\Omega)$ is defined as*

$$W^{m,p}(\Omega) = \{f : \Omega \rightarrow \mathbb{R} : f \in L^p(\Omega),\ D^{\vec{\alpha}}_w f \in L^p(\Omega)\ \forall \vec{\alpha}, |\vec{\alpha}| \leq m\} \tag{2.25}$$

with the norm

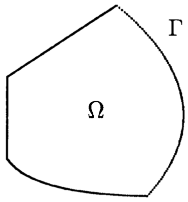
$$\|f\|_{W^{m,p}(\Omega)} = \left(\sum_{|\vec{\alpha}| \leq m} \|D^{\vec{\alpha}}_w f\|^p_{L^p(\Omega)} \right)^{\frac{1}{p}} \tag{2.26}$$

Notice that, previously mentioned spaces are especial cases[8]

- $W^{0,p}(\Omega) = L^p(\Omega)$ is the Lebesgue space of order p (e.g., $W^{0,2}(\Omega) = L^2(\Omega)$, the space of square integrable functions);
- $W^{m,2}(\Omega) = H^m(\Omega)$ is a Hilbert space (e.g., $W^{1,2}(\Omega) = H^1(\Omega)$, the space of square integrable functions whose derivatives are also square integrable).

There are also important properties of these spaces that are worth to know:

- It can be shown that $W^{m,p}(\Omega)$ is a Banach space (see [2]);
- $H^m(\Omega)$ is the completion of $C^\infty(\Omega)$ with respect to the norm $\|\cdot\|_{H^m(\Omega)}$. This means, for any $v \in H^m(\Omega)$ it is possible to find an infinitely differentiable function f that is arbitrarely close to v, i.e., $\|v - f\|_{H^m(\Omega)} < \epsilon$, $\forall \epsilon > 0$. We say that $C^\infty(\Omega)$ is dense in $H^m(\Omega)$[9];
- An alternative definition of the space $H^m(\Omega)$, instead of looking whether the weak derivatives belong or not to $L^2(\Omega)$ is by completion of the space $C^m(\Omega)$ with respect to the norm $\|\cdot\|_{H^m(\Omega)}$.



Γ Lipschitz

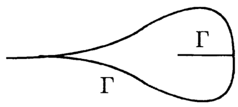


Figure 2.9: Examples of domains with Lipschitz boundary (top) and not Lipschitz boundary (bottom).

[8] Both, the $L^2(\Omega)$ and $H^1(\Omega)$ are the most relevant spaces in this course.

[9] This allows us to prove results involving the space $H^1(\Omega)$, by showing them for smooth functions. This is called a *density argument*.

In particular, we must state the following important theorem concerning Hilbert spaces H^m and its connection to the spaces C^k:

Theorem 2.4.1 (***Sobolev embedding***): *Given a bounded domain $\Omega \subset \mathbb{R}^d$ with Lipschitz boundary*[10]. *Let m and k such that $m - k > \frac{d}{2}$, then, every function of $H^m(\Omega)$ belongs to $C^k(\Omega)$. Furthermore, the inclusion $H^m(\Omega) \subset C^k(\Omega)$ is continuous.*

- Continuous inclusion or **continuous embedding** of $H^m(\Omega)$ in $C^k(\Omega)$, that is denoted sometimes $H^m(\Omega) \hookrightarrow C^k(\Omega)$, means[11]

$$\|v\|_{W^{k,\infty}(\Omega)} \leq c\|v\|_{H^m(\Omega)} \tag{2.27}$$

- In one spatial dimension $d = 1$, take $m = 1$ and $k = 0$, any function in $H^1(\Omega)$ is automatically continuous, i.e., $H^1(\Omega) \subset C^0(\Omega)$;
- In two and three spatial dimensions $d \geq 2$, take $m = 2$, $k = 0$, we can say that any function in $H^2(\Omega)$ is continuous, i.e., $H^2(\Omega) \subset C^0(\Omega)$;
- There are others Sobolev's embeddings for general $W^{m,p}(\Omega)$ spaces (see a more advanced reference).

2.5 Conclusion

To conclude this chapter, let us work out a few exercices which also serve to introduce some additional tools to be used later on:

Solve the following exercices

- **Cauchy-Schwarz inequality**. Consider a space V with inner product $(\cdot,\cdot)_V$ and induced norm $\|\cdot\|_V$. For any $v, w \in V$

$$|(v,w)_V| \leq \|v\|_V\|w\|_V \tag{2.28}$$

  Prove this important inequality. This inequality is a particular case of the more general **Hölder's inequality**: Let $p, q \in [1,\infty]$ such that $\frac{1}{p}+\frac{1}{q} = 1$, $v \in L^p(\Omega)$ and $w \in L^q(\Omega)$, then $v\,w \in L^1(\Omega)$ and

$$\|v\,w\|_{L^1(\Omega)} \leq \|v\|_{L^p(\Omega)}\|w\|_{L^q(\Omega)} \tag{2.29}$$

- **Triangle inequality**. Using the Cauchy-Schwarz inequality, prove that for any $v, w \in V$

$$\|v+w\|_V \leq \|v\|_V + \|w\|_V \tag{2.30}$$

- Prove that the sequence of Equation 2.11 is a Cauchy sequence.
- Consider the space $H^1(\Omega)$, $\Omega = (a,b)$, the Hilbert space of functions $v : (a,b) \longrightarrow \mathbb{R}$ and the corresponding induced norm of Equation 2.14, i.e., $\|v\|_{H^1(a,b)} = \left(\int_a^b (v^2 + v'^2)\,dx\right)^{\frac{1}{2}}$. Show

Figure 2.10: Augustin-Louis Cauchy (France, 1789–1857).

---

[10] 10: We need the shape of Ω to be "reasonable". Lipschitz boundary means that, everywhere the boundary is the graph of a Lipschitz function (i.e., $|f(\mathbf{x}) - f(\mathbf{y})| < L|\mathbf{x}-\mathbf{y}|$). Examples and counter examples are shown in Figure 2.9.

[11] 11: In Equation 2.27 we define $\|v\|_{W^{k,\infty}(\Omega)} = \max_{|\vec{\alpha}|\leq k} \|D_w^{\vec{\alpha}} v\|_{L^\infty(\Omega)}$. As a particular case, for $k = 0$ we have the $L^\infty(\Omega)$-norm defined in Equation 2.8, which is a natural norm for the space of continous functions.

that

$$\begin{aligned} |||v|||_{H^1(a,b)} &= \left(\|v\|^2_{L^2(\Omega)} + \|v'\|^2_{L^2(\Omega)}\right)^{\frac{1}{2}} \\ |||v|||_{H^1(a,b)} &= \|v\|_{L^2(\Omega)} + \|\kappa(x)v'\|_{L^2(\Omega)} \end{aligned}$$

where $0 \leq \kappa_{\min} < \kappa(x) \leq \kappa_{\max}$, is an equivalent norm to $\|\cdot\|_{H^1(a,b)}$.

- Let $f : \Omega \longrightarrow \mathbb{R}$, $\Omega \subset \mathbb{R}^3$. Write the partial derivative of f corresponding to $\vec{\alpha} = (2, 0, 0)$ and $\vec{\alpha} = (0, 1, 1)$.
- Let $f : \Omega \longrightarrow \mathbb{R}$, $\Omega \subset \mathbb{R}^2$. Write in detail the definition of the space $H^2(\Omega)$ and its corresponding norm (Equation 2.26).
- (**Optional**) Show that the space

$$V = \{w : (0,1) \longrightarrow \mathbb{R}, \int_0^1 w(x)^2\, dx < +\infty, \\ \int_0^1 w'(x)^2\, dx < +\infty\}$$

is contained in $C^0(0,1)$. What space is V?.

2.6 Assignment 2: Computing some norms in `FEniCSx`

Implement a script that

- Compute the $L^2(\Omega)$ and $H^1(\Omega)$-norms (see Equation 2.13 and Equation 2.14) of the solution u_d to Poisson's problem from the previous Assignment:
  - Program the calculation by using the `FEniCSx` keyword `inner`[12], i.e.,

```
L2norm = fem.assemble_scalar( ... )
gradterms = fem.form(inner( ... , ... )*dx )
H1norm = L2norm + fem.assemble_scalar( gradterms )
H1norm = np.sqrt(H1norm)
```

    and doing the componentwise multiplication of the derivatives, i.e.,

```
L2norm =  fem.assemble_scalar( ... )
gradterms = fem.form( (ud.dx(0)**2 + ... )*dx )
H1norm = L2norm + fem.assemble_scalar( gradterms )
H1norm = np.sqrt(H1norm)
```

  - Consider the case of the viscosity μ being constant or dependent on $\mathbf{x}$ as taken in the second item of that Assignment.
  - Build a table reporting the results as a function of the number of elements in the discretized domain.
- Define the vector–valued function

[12] In `FEniCSx`, for vector–valued functions, the `inner` and `dot` keywords are synonyms, and simply refer to the usual "·" product of vectors in $\mathbb{R}^d$.

$$\mathbf{u}(x_1, x_2, t) = [\sin(x_1)\sin(x_2),\ \cos(x_1)\cos(x_2)]^\intercal \tag{2.31}$$

Now, compute its $L^2(\Omega)$ and $H^1(\Omega)$-norms as done in the previous item, i.e., by using the `inner`[13] keyword and by doing "by hand" the componentwise calculation.

- Consider $\Omega = [-\pi, \pi]$ and the functions

$$u(x) = \sin(mx),\ \ v(x) = \sin(nx) \tag{2.32}$$

$m, n \in \mathbb{N}$. Compute by hand the $L^2(\Omega)$–inner product of u and v for different combinations of m and n (e.g., $n = m = 1$ and $n = 1,\ m = 2$). Then, write a `FEniCSx` script to compute the same products and compare the results to the exact solution as the number of elements in the mesh is increased.

[13] In `FEniCSx`, for matrix–valued functions, the `inner` keyword refers to the double contraction ":" of matrizes in $\mathbb{R}^{d\times d}$.

VARIATIONAL FORMULATIONS I: Abstract problems

3

3.1 Introduction

First, we recall the Poisson's problem with homegeneous Dirichlet data, we have presented in the first chapter. Let $\Omega \subset \mathbb{R}^d$ be an open domain with Lipschitz boundary $\partial\Omega$. Consider the source term f and coefficient μ. The problem reads: Find $u \in C^2(\Omega) \cap C(\bar{\Omega})$ that satisfy pointwise the partial differential equation:

$$(\mathbf{DP}): \quad \left\{ \begin{array}{rcll} -\nabla \cdot (\mu(\mathbf{x}) \nabla u(\mathbf{x})) & = & f(\mathbf{x}) & \mathbf{x} \in \Omega \\ u(\mathbf{x}) & = & 0 & \mathbf{x} \in \partial\Omega \end{array} \right. \tag{3.1}$$

and its variational (or **weak**) form we have used to solve the problem by the finite element method using the `FEniCS` platform (although we didn't provide too many details on what was behind such resolution):

Weak form of Poisson's problem

Find $u \in V(\Omega)$ such that

$$(\mathbf{VF}): \quad \left\{ \begin{array}{l} \underbrace{\int_\Omega \mu(\mathbf{x})\, \nabla u(\mathbf{x}) \cdot \nabla v(\mathbf{x})\, dx}_{a(u,v)} = \underbrace{\int_\Omega f(\mathbf{x})\, v(\mathbf{x})\, dx}_{\ell(v)} \\ \\ \forall v \in V(\Omega). \end{array} \right. \tag{3.2}$$

Notice that in Equation 3.2 we have:

- In the left hand side, a **bilinear form** $a(u,v)$, that is, a form that takes two functions as arguments and is linear in both arguments[1], i.e., $a(\cdot,\cdot): V \times V \longrightarrow \mathbb{R}$,

$$a(u,v) = \int_\Omega \mu\, \nabla u \cdot \nabla v\, dx \tag{3.3}$$

- In the right hand side, a **linear form** $\ell(v)$, a form that takes one function as its argument and is linear, i.e., $\ell(\cdot): V \longrightarrow \mathbb{R}$,

$$\ell(v) = \int_\Omega f\, v\, dx \tag{3.4}$$

In Chapter 1 no rigorous definition of the functional setting for this problem, the proper definition of the space V or any regularity assumption as to the data f and μ were provided. For instance, what happen if coefficient $\mu(\mathbf{x})$ is a rough function looking as it is shown in Figure 3.1?

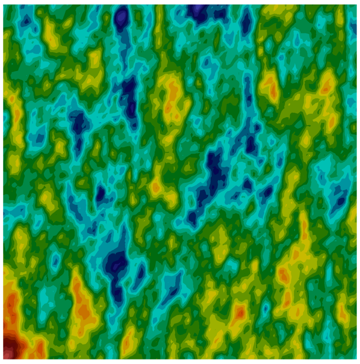


Figure 3.1: Example of a rough coefficient $\mu(\mathbf{x})$ in Equation 3.2.

[1] Linearity of a in both arguments means that forall $u, v, w \in V$, $\alpha, \beta \in \mathbb{R}$

$$a(\alpha\, u + \beta\, v, w) = \alpha\, a(u,w) + \beta\, a(v,w)$$
$$a(u, \alpha\, v + \beta\, w) = \alpha\, a(u,v) + \beta\, a(u,w)$$

We aim in this and following chapters to fill in this gap and introduce under which conditions the problem is mathematically well-posed[2].

Problem (**VF**) as well as others to be discussed later can be cast in a general abstract form:

Abstract variational problem

Find $u \in V$ such that

$$a(u, v) = \ell(v) \tag{3.5}$$

forall $v \in V$.

Figure 3.2: Jacques Salomon Hadamard (France, 1865–1963).

3.2 Linear and bilinear forms

It is clear that if we aim to establish whether a given variational problem is well posed or not, we have to inspect the properties of the bilinear and the linear forms that appear Equation 3.5, so before proceeding we introduce some definitions and a two very important theorems.

Definition 3.2.1 *(**Bounded linear functional**): Let V be a Hilbert space with norm $\|\cdot\|_V$. A linear functional $\ell : V \longrightarrow \mathbb{R}$ is said to be bounded if there exists a constant N_ℓ such that*

$$|\ell(v)| \ \leq \ N_\ell \, \|v\|_V \quad \forall v \in V \tag{3.6}$$

- If the linear functional ℓ is bounded, this also means that is continuous, roughly speaking, if we change v a little, $\ell(v)$ also changes little[3], thus, **continuous** and **bounded** linear functionals are used as synonysms;
- Equation 3.6 implies that $\ell \in V^*$, the dual of V which is by definition the set of all bounded linear functionals on V;
- The minimum value N_ℓ for which Equation 3.6 holds is called the norm of ℓ and we write

$$\|\ell\|_{V^*} = N_\ell \tag{3.7}$$

Definition 3.2.2 *(**Bounded bilinear form**): Let V be a Hilbert space with norm $\|\cdot\|_V$. A bilinear form $a : V \times V \longrightarrow \mathbb{R}$ is said to be bounded if there exists a constant N_a such that*

$$|a(u, v)| \ \leq \ N_a \, \|u\|_V \, \|v\|_V \quad \forall u, v \in V \tag{3.8}$$

- A bilinear form is bounded *iff* is continuous (you can fix one of its arguments and see it as a linear form;
- N_a is called the continuity constant).

Finally, we introduce the concept of Ellipticity or Coercivity of a bilinear form which is essential to show **Existence** and **Uniqueness** of solutions to the variational problem:

[2] **Well-posed problem** (Hadamard, 1932): A problem is well-posed if it admits one and only one solution which depends continuously on the forcing data.

[3] If ℓ is bounded, we can prove it is continuous by doing

$$|\ell(v+w)-\ell(v)| = |\ell(w)| \leq N_\ell \, \|w\|_V \xrightarrow[w\to 0]{} 0.$$

The converse is also true.

Definition 3.2.3 *(**Strongly coercive bilinear form**): Let V be a Hilbert space with norm $\|\cdot\|_V$. A bilinear form $a : V \times V \longrightarrow \mathbb{R}$ is said to be strongly coercive if there exists a constant $\alpha > 0$ such that*

$$a(u,u) \;\geq\; \alpha \, \|u\|_V^2 \quad \forall u \in V \tag{3.9}$$

3.3 Symmetric problems: Riesz representation

We begin by stating the Riez representation theorem which is a fundamental tool to deal with variational problems in which the bilinear form is symmetric. In principle, the theorem will seem a bit abstract but then we will explain how to use it to prove existence, uniqueness and stability of solutions.

Theorem 3.3.1 *(**Riesz representation theorem**): Let V be a Hilbert space with inner product $(\cdot,\cdot)_H$. For each continuous linear functional $\ell \in V^*$, there exists a **unique** $u \in V$ such that*

$$\begin{aligned} \|\ell\|_{V^*} &= \|u\|_H \\ \ell(v) &= (u,v)_H \quad \forall v \in V \end{aligned}$$

Figure 3.3: Frigyes Riesz (Hungary, 1880–1956).

*you may be asking yourself how this relates to the solution of (**VF**) ... so, let's go back to our variational problem*

$$\text{Find } u \in V \text{ s.t. } a(u,v) = \ell(v) \;\; \forall v \in V.$$

where V is a Hilbert space. Let consider a bilinear form a that is:

Symmetric: This implies that

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

Bounded and **Coercive**: These imply that

$$\sqrt{\alpha}\, \|u\|_V \leq \sqrt{a(u,u)} \leq \sqrt{N_a}\, \|u\|_V \quad \forall u \in V \tag{3.10}$$

Now, we reason as follows:

- Looking at Definition 2.2.2, we acknowledge that the bilinear form $a(\cdot,\cdot)$ defines an inner product[4].
- Recall that if V is a Hilbert space, it is an inner product space, that is complete with respect to the norm $\|\cdot\|_V$, i.e., for a sequence of functions $\{v_n\}$, $\|v_n - v\|_V \xrightarrow[n\to\infty]{} 0,\;\; v \in V$;
- Since by Equation 3.10, $\|\cdot\|_V$ and $\|\cdot\|_a$ are equivalent norms, then, $\|v_n - v\|_a \xrightarrow[n\to\infty]{} 0,\;\; v \in V$, so V is also complete with respect to the energy norm $\|\cdot\|_a$;
- Now, in Theorem 3.3.1 take as inner product on V the energy product, i.e., $(\cdot,\cdot)_H = (\cdot,\cdot)_a = a(\cdot,\cdot)$, so we can restate the Riesz representation theorem for our particular case as:

[4] This inner product is sometimes called the *energy inner product* and the induced norm is called the *energy norm* that is denoted by $\|u\|_a = \sqrt{a(u,u)}$

Theorem 3.3.2 (***Well-posedness of* (VF)**): *Let V be a Hilbert space, $a : V \times V \longrightarrow \mathbb{R}$ be a **symmetric**, **continuous** and **strongly coercive** bilinear form. For each continuous linear functional $\ell \in V^*$, there exists a **unique** $u \in V$ such that*

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

*Moreover, the solution u is **stable**, i.e., $\|u\|_V \leq \frac{1}{\alpha}\|\ell\|_{V^*}$*

A few comments are in order:

- The stability statement, means that the solution depends continuously on the "data". For instance, in Equation 3.4, if the forcing term f changes little, the solution also changes little;
- The inequality $\|u\|_V \leq \frac{1}{\alpha}\|\ell\|_{V^*}$ comes from the fact that, $\|\ell\|_a = \|u\|_a$[5], $\|u\|_a \geq \sqrt{\alpha}\|u\|_V$ and also $\|\ell\|_a \leq \frac{1}{\sqrt{\alpha}}\|\ell\|_{V}$;
- For each (**VF**) in particular, that $\ell : V \longrightarrow \mathbb{R}$ belongs to V^* has to be proven;
- The symmetry, continuity and coercivity of a has to be proven as well. This can be more or less difficult depending on each case.

In the next chapter we consider specific examples of variational problems and work on the two latter points.

3.4 Nonsymmetric problems: Lax-Milgram

The previous theorem is restricted to symmetric problems. However, there problems in which the bilinear form is nonsymmetric. The canonical example in fluid mechanics is the case of a **convection** (or advection) term in the Poisson's problem, namely, given a velocity field $\boldsymbol{\beta}$ we consider,

$$-\nabla \cdot (\mu \nabla u) + \underbrace{\boldsymbol{\beta} \cdot \nabla u}_{\text{convection term}} = f$$

As done before, we multiply this equation by a function v, the associated bilinear form becomes

$$a(u,v) = \int_\Omega \mu \, \nabla u \cdot \nabla v \, dx + \int_\Omega (\boldsymbol{\beta} \cdot \nabla u) \, v \, dx$$

which is clearly nonsymmetric, since by inspection, we can not exchange u by v (i.e, $a(u,v) \neq a(v,u)$). For cases in which the bilinear form is not necessarily symmetric, there exists the Lax-Milgram theorem that states exactly the same as Theorem 3.3.2 except that the form a need not to be an inner product:

Theorem 3.4.1 (***Lax-Milgram***): *Let V be a Hilbert space, $a : V \times V \longrightarrow \mathbb{R}$ be a **continuous** and **strongly coercive** bilinear form. Then, the problem:*

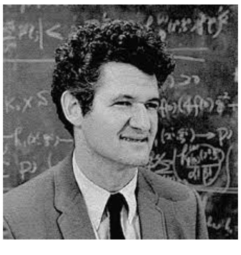


Figure 3.4: Peter D. Lax (Hungary, 1926).

[5] This is because we have chosen the $a-$ inner product in Theorem 3.3.1. Also, we are abusing notation by denoting $\|\ell\|_a$ the norm on V^* with respect to the energy norm in V.

Find $u \in V$ such that

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

is well posed for any continuous linear functional $\ell \in V^*$ and $\|u\|_V \leq \frac{1}{\alpha}\|\ell\|_{V^*}$.

3.5 Conclusion

To conclude this chapter, let us work out a few exercices:

Solve the following exercices

- Show that a scalar product $(\cdot,\cdot)$ is a continuous bilinear form.
- Complete all the details to show that a symmetric, continuous and coercive bilinear form satisfy the definition of inner product given in Definition 2.2.2. Also show that if for a sequence of functions $\{v_n\}$ in V, $\|v_n - v\|_V \xrightarrow[n\to\infty]{} 0, \quad v \in V \quad \Rightarrow \|v_n - v\|_a \xrightarrow[n\to\infty]{} 0.$

3.6 Assignment 2: Poisson's problem with inclusions

Let us continue with our exploration of Poisson's problem in the FEniCSx platform, as a model for instance of fully developed flow in ducts or heat conduction in solids. We haven't learned yet all the theoretical tools to fully understand what is behind the implementation, however, as we did in Chapter 1, we will proceed at a more or less informal level. The problem we aim to solve now is the following:

Given the computational domain $\Omega = [0, L_x] \times [0, L_y] \subset \mathbb{R}^2$, $\partial\Omega = \Gamma_{\text{left}} \cup \Gamma_{\text{ad}} \cup \Gamma_{\text{right}}$, source term f and thermal diffusivity μ, find temperature field u satisfying

$$\left\{\begin{array}{rcll} -\nabla\cdot(\mu(\mathbf{x})\,\nabla u(\mathbf{x})) & = & f(\mathbf{x}) & \mathbf{x}\in\Omega \\ u(\mathbf{x}) & = & u_\ell & \mathbf{x}\in\Gamma_{\text{left}} \\ u(\mathbf{x}) & = & u_r & \mathbf{x}\in\Gamma_{\text{right}} \\ -\mu(\mathbf{x})\,\nabla u(\mathbf{x})\cdot\check{\mathbf{n}}(\mathbf{x}) & = & 0 & \mathbf{x}\in\Gamma_{\text{ad}} \end{array}\right.$$

The thermal diffusivity μ is piecewise constant and defined by

$$\mu(\mathbf{x}) = \left\{\begin{array}{lll} \mu_B & \text{if} & \mathbf{x}\in\Omega_B = \bigcup_{i=1}^{n_b}\omega_i,\ \omega_i = \{\mathbf{x}\in\Omega, \|\,\mathbf{x}-\mathbf{x}_c^i\,\| < r_i\} \\ \mu_A & \text{if} & \mathbf{x}\in\Omega_A = \Omega\setminus\Omega_B \end{array}\right.$$

i.e., μ takes the value μ_A everywhere in Ω, except at some known distributed circular inclusions with centers and radii $\{\mathbf{x}_c^i, r_i\}_{i=1}^{n_b}$, in which it takes the value μ_B. Similarly, the source term f can be piecewise constant. The objetive is to study the influence of the ratio μ_B/μ_A on the solution for different settings to be precised later on. Typical results are shown in Figure 3.5. Notice that Γ_{ad} represents the top and bottom walls, in which a zero flux is being assummed (adiabatic walls). From the countour plots we observe indeed that the normal derivative of u is zero at those walls.

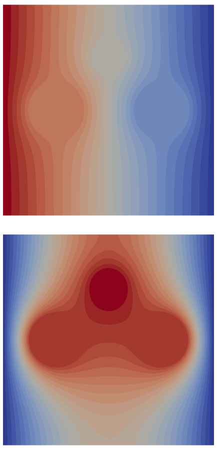


Figure 3.5: Two typical results considering homogenous source term (top) and homegenous Dirichlet conditions (bottom) taking $\mu_B \gg \mu_A$.

The discrete variational formulation

The finite element formulation follows by proceeding as in Chapter 1. The main difference now is that the test functions v must satisfy

$$v(\mathbf{x}) = 0 \quad \forall \mathbf{x} \in \Gamma_{\text{left}} \cup \Gamma_{\text{right}} \tag{3.11}$$

The temperature field is sought in a finite dimensional space V_h[6]. We mention in advance that this space V_h must be a subspace of the Hilbert space introduced in Chapter 2

$$H^1(\Omega) = \{v : \Omega \longrightarrow \mathbb{R}, \;\; v \in L^2(\Omega), \;\; \nabla v \in [L^2(\Omega)]^d\}$$

In fact, since we consider nonhomegeneous boundary conditions in $\Gamma_{\text{left}} \cup \Gamma_{\text{right}}$, we seek for solutions in the set

$$V_{hg} \subset V_g(\Omega) = \{v \in H^1(\Omega), \;\; v|_{\Gamma_{\text{left}}} = u_\ell, \;\; v|_{\Gamma_{\text{right}}} = u_r\}$$

In particular, we will consider polynomial spaces of degree k, e.g., linear or quadratic functions on the elements of a partition of Ω[7]. Let $\{\phi^1, \ldots, \phi^n\}$ be a basis of V_h and u_h the expansion

$$u_h(\mathbf{x}) = \sum_{i=1}^{n} U_i \, \phi^i(\mathbf{x}).$$

The variational formulation reads: Find $u_h \in V_{hg}$ such that

$$\int_\Omega \mu \nabla u_h \cdot \nabla v_h \, dx = \int_\Omega f \, v_h \, dx \;\; \forall v_h \in V_{h0}, \tag{3.12}$$

where V_{h0} is V_{hg} when $u_\ell = u_r = 0$.

Implement the following modifications to `inclusions_poisson.py`

- Inspect and run the script provided and make sure to understand its different sections.
- Run the script by setting $\mu_A = \mu_B = 1, f_A = f_B = 0$ and compare the solution with the exact one. Visualize the results in `Paraview`.

[6] Instead of using the subindex d to denote the discrete space, we adopt the subindex h, which is the standard notation in the FEM. The h refers to the mesh refinement, i.e., the typical size of elements in the partition of Ω.

[7] Further details on how to construct polynomial spaces on finite element meshes, will be given later on.

- Compute the error in the $L^2(\Omega)$ and $H^1(\Omega)$-norms, i.e.,

$$\|u - u_h\|_{L^2(\Omega)} = \sqrt{\int_\Omega (u - u_h)^2 \, dx},$$

$$\|u - u_h\|_{H^1(\Omega)} = \sqrt{\int_\Omega \left[(u - u_h)^2 + |\nabla(u - u_h)|^2\right] dx}$$

  where u is the exact solution you must first determine. Consider and complete the code:

```
ue = uleft - (uleft - uright)*x[0]/Lx
gradue = as_vector([-(uleft - uright)/Lx, 0.0])
errorL2 = assemble_scalar(form((uh - ue)**2*dx))
errorH1 = errorL2 + ...
print("    |-L2error=", np.sqrt(errorL2))
print("    |-H1error=", np.sqrt(errorH1))
```

  From now on, we consider two situations:

  - Homegeneous source term ($f = 0$) and nonhomegeneous Dirichlet conditions ($u_\ell = 100, \; u_r = 1$);
  - Non-homegeneous source term ($f = 100$) and homogenous Dirichlet conditions ($u_\ell = u_r = 0$);

- Run the script by setting different values for μ_B and visualize the solution;

- Compute the amount of heat entering or leaving the domain through the left and the right boundaries, i.e.

$$Q_{\text{in}} = \int_{\Gamma_{\text{left}}} \mu_A \nabla u_h \cdot \check{\mathbf{e}}_1 \, ds, \quad Q_{\text{out}} = \int_{\Gamma_{\text{right}}} -\mu_A \nabla u_h \cdot \check{\mathbf{e}}_1 \, ds$$

  To that end consider the code:

```
n = FacetNormal(mesh)
ds = Measure('ds', subdomain_data=ft)
Qleft =assemble_scalar(form(inner(mu*grad(uh),n)*ds(3)))
Qright= ...
```

- In the first situation, compute the effective thermal diffusivity, i.e.,

$$\mu_{\text{eff}} = \frac{|Q|/L_y}{|u_\ell - u_r|/L_x}$$

- In the second situation, compute the total amount of energy being produced in the domain, i.e.,

$$S = \int_\Omega f \, dx$$

  and compare to $Q_{\text{in}} + Q_{\text{out}}$.

- Let consider $\mu_B \gg \mu_A$, such that the temperature on each circular region is nearly uniform. Implement the computation of the average temperature on each inclusion, i.e.,

$$\bar{T}_i = \frac{1}{|\omega_i|} \int_{\omega_i} u(\mathbf{x})\, dx$$

in which $|\omega_i|$ stands for the the area of region ω_i. As a guideline, the code can be structured as shown below

```
dx = Measure('dx', subdomain_data=ct)
one = Constant(mesh, ScalarType(1))
Tincav = []
for k in range(ninclusions):
   area = assemble_scalar(form(one * dx(3+k)))
   Tinc = ...
   Tincav.append(Tinc/area)
```

- Prepare a short presentation to report the results.

VARIATIONAL FORMULATIONS II: Elliptic examples

4

In the previous chapter we considered abstract problems and at the end we solved one Poisson's equation, although without justification of some choices. In this chapter, we continue with concrete examples of elliptic problems, such as Poisson's and linear elasticity, and their variational formulations and provide additional details to as the choice of the appropriate functional setting.

4.1 Poisson's problem

4.1.1 Homegeneous Dirichlet BCs

Let consider once again the Poisson's equation on a domain Ω with **homegeneous Dirichlet** boundary conditions:

$$\begin{cases} -\nabla\cdot(\mu\nabla u) &= f \quad \text{in } \Omega \\ \\ u &= 0 \quad \text{on } \partial\Omega \end{cases} \tag{4.1}$$

and recall the variational (or **weak**) form: Find $u \in V$ such that

$$\underbrace{\int_\Omega \mu\,\nabla u\cdot\nabla v\,dx}_{a(u,v)} = \underbrace{\int_\Omega f\,v\,dx}_{\ell(v)} \quad \forall v \in V. \tag{4.2}$$

What can be said about the regularity of f and v and the appropriate choice for the space V?

- f and v should at least belong to $L^2(\Omega)$ if we pretend that $|\ell(v)| < +\infty$ (i.e., ℓ must belong to V^* for the Theorem 3.3.2 or Theorem 3.4.1 to be applicable);

- At the same time, the weak derivatives of u and v, must also belong to $L^2(\Omega)$ if we pretend that $|a(u,v)| < +\infty$, this clearly suggests that V has to be a subspace of

$$H^1(\Omega) = \{v : \Omega \to \mathbb{R} : \int_\Omega v^2\,dx < +\infty \text{ and } \int_\Omega (\nabla v\cdot\nabla v)\,dx < +\infty\}$$

- Finally, we must satisfy the boundary condition $u|_\Gamma = 0$ on $\partial\Omega$, so the space of choice will clearly be[1]

$$H^1_0(\Omega) = \{v \in H^1(\Omega) : v|_\Gamma = 0\}.$$

  This is a closed subspace[2];

- That $a(\cdot,\cdot)$ is strongly coercive over this space (i.e., $a(v,v) \geq \alpha\|v\|^2_{H^1_0(\Omega)}$ $\forall v \in H^1_0(\Omega)$) remains to be proved.

---

[1] For this reason, Dirichlet conditions are also named as **essential** boundary conditions, since, they are embedded into the definition of the space V.

[2] A closed subspace of a Hilbert space is also a Hilbert space, a fact that is very often invoked throughout. At this point, we recall that a subset Y of a complete normed space X is complete *iff* Y is closed in X.

Lastly, the problem can be written as

Weak form of Poisson's problem with $g = 0$

$$\text{Find } u \in H_0^1(\Omega) \text{ s.t. } \int_\Omega \mu \, \nabla u \cdot \nabla v \, dx = \int_\Omega f \, v \, dx \quad \forall v \in H_0^1(\Omega).$$

In what follows, we consider some variants of this problem.

4.1.2 Inhomegeneous Dirichlet BCs

The natural thing to do now is to consider a more general situation and admit an nonhomegeneous data on $\partial\Omega$. The (**DF**) thus reads

$$\begin{cases} -\nabla \cdot (\mu \nabla u) &= f \quad \text{in } \Omega \\ \\ u &= g \quad \text{on } \partial\Omega \end{cases} \tag{4.3}$$

where the function g is defined on $\partial\Omega$. This function must be regular enough such that there exits a function $u_g \in H^1(\Omega)$ that is equal to g forall $x \in \partial\Omega$. If you want, you can also think g as a function in $H^1(\Omega)$, defined over all Ω. In order to solve this problem we apply the following trick. The solution is written as:

$$u = u_0 + u_g \tag{4.4}$$

where $u_0 = 0 \ \forall x \in \partial\Omega$. Clearly, u satisfies the Dirichlet condition[3].

To derive the variational formulation we substitute into the PDE, multiply by a function $v \in H_0^1(\Omega)$ and integrate by parts

$$\int_\Omega \mu \, \nabla(u_0 + u_g) \cdot \nabla v \, dx = \int_\Omega f \, v \, dx \quad \forall v \in H_0^1(\Omega). \tag{4.5}$$

Now, we move to the right hand side the term that depends on the given data u_g and obtain

Weak form of Poisson's problem with $g \neq 0$

$$\text{Find } u_0 \in H_0^1(\Omega) \text{ s.t. } \int_\Omega \mu \, \nabla u_0 \cdot \nabla v \, dx = \int_\Omega f \, v \, dx - \int_\Omega \mu \, \nabla u_g \cdot \nabla v \, dx \quad \forall v \in H_0^1(\Omega).$$

and obtain u as a post-processing step by summing u_0 and u_g. Notice that we have a problem with the same bilinear form $a(\cdot, \cdot)$ and space V as in the previous case, except that the linear form has now changed to

$$\ell(v) = \int_\Omega f \, v \, dx - \underbrace{\int_\Omega \mu \, \nabla u_g \cdot \nabla v \, dx}_{a(u_g, v)} \tag{4.6}$$

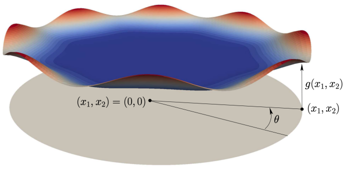


Figure 4.1: Example of a function u_g that corresponds to the function $g(x_1, x_2) = \sin^2(4\,\theta), \ \theta = \text{atan}(x_2/x_1)$.

---

[3] Note that the choice of u_g is quite flexible, as far as it has the required regularity and satisfies the boundary condition. For instance, it can be a function that is equal to g on $\partial\Omega$ and decays to zero or same value when we move far away from $\partial\Omega$, as illustrated in Figure 4.1.

Now, we embark in the task of showing that this problem is actually well posed.

4.1.3 Proof of well-posedness

Let us collect everything:

$$\left[\begin{array}{rl}\text{The space:} & V = H^1_0(\Omega)\\ \text{The inner product on } V\text{:} & (u,v)_V = \displaystyle\int_\Omega (u\,v + \nabla u\cdot\nabla v)\,dx = (u,v)_{H^1(\Omega)}\\ \text{The induced norm on } V\text{:} & \|u\|^2_V = \displaystyle\int_\Omega \left(u^2 + \nabla u\cdot\nabla u\right)dx = \|u\|^2_{H^1(\Omega)}\\ \text{The bilinear form:} & a(u,v) = \displaystyle\int_\Omega \mu\,\nabla u\cdot\nabla v\,dx\\ \text{The linear form:} & \ell(v) = \displaystyle\int_\Omega f\,v\,dx - a(u_g,v)\end{array}\right.$$

Continuity of *a*

We begin by assuming that $\exists\ \mu_m > 0$ and μ_M such that $\mu_m \le \mu(\mathbf{x}) \le \mu_M\ \forall\mathbf{x}\in\Omega$, which is "physically" quite reasonable. Let proceed by showing all the details:

$$\begin{aligned}
a(u,v) &= \int_\Omega \mu\,\nabla u\cdot\nabla v\,dx\\
&\le \mu_M\int_\Omega |\nabla u\cdot\nabla v|\,dx\\
&\le \mu_M\int_\Omega \|\nabla u\|_2\|\nabla v\|_2\,dx\ \ \text{(Cauchy-Schwarz in } \mathbb{R}^d\text{)}\\
&= \mu_M\int_\Omega \sqrt{\nabla u\cdot\nabla u}\sqrt{\nabla v\cdot\nabla v}\,dx\ \ \text{(Definition of Euclidean norm in } \mathbb{R}^d\text{)}\\
&\le \mu_M\left(\int_\Omega \nabla u\cdot\nabla u\,dx\right)^{\frac12}\left(\int_\Omega \nabla v\cdot\nabla v\,dx\right)^{\frac12}\ \ \text{(Cauchy-Schwarz in } L^2(\Omega)\text{)}\\
&\le \mu_M\left(\int_\Omega (u^2+\nabla u\cdot\nabla u)\,dx\right)^{\frac12}\left(\int_\Omega (v^2+\nabla v\cdot\nabla v)\,dx\right)^{\frac12}\ \ \text{(Summing } u^2 \text{ and } v^2\text{)}\\
&= N_a\,\|u\|_{H^1(\Omega)}\,\|v\|_{H^1(\Omega)}\ \ \text{(Definition of } H^1(\Omega)\text{-norm)}
\end{aligned}$$

where in the 3rd step we used the Cauchy-Schwarz inequality in $\mathbb{R}^d$[4] and took $N_a = \mu_M$. Summarizing, $a(\cdot,\cdot)$ is bounded:

$$a(u,v) \le \mu_M\,\|u\|_{H^1(\Omega)}\,\|v\|_{H^1(\Omega)}\ \forall u,v\in V \tag{4.7}$$

[4] For vector-valued functions $\mathbf{f},\mathbf{g}:\Omega\to\mathbb{R}^d$, $|\mathbf{f}\cdot\mathbf{g}| \le \|\mathbf{f}\|_2\,\|\mathbf{g}\|_2$ where $\|\cdot\|_2$ denotes the Euclidean norm (i.e., $\|\mathbf{f}\|_2 = \sqrt{\mathbf{f}\cdot\mathbf{f}}$).

Coercivity of *a*

This is a bit more difficult. We will need the Poincaré-Friedrichs inequality (*one more ...*)

Figure 4.2: Henri Poincaré (France, 1854–1912).

Theorem 4.1.1 ***(Poincaré-Friedrichs inequality)*** *Let* $\Omega \subset \mathbb{R}^d$ *be a bounded domain with Lipschitz boundary,* $\Gamma_D \subset \partial\Omega$, *meas*$(\Gamma_D) \neq 0$ *and the space*

$$H^1_{D0}(\Omega) = \{v \in H^1(\Omega) : v|_{\Gamma_D} = 0\},$$

Then, there exists C_{Ω,Γ_D} *(depending only on* Ω *and* Γ_D*) such that, for any* $u \in H^1_{D0}$

$$\int_\Omega \nabla u \cdot \nabla u \, dx \geq C_{\Omega,\Gamma_D} \int_\Omega u^2 \, dx \qquad (4.8)$$

Figure 4.3: Kurt Otto Friedrichs (Germany (1901)–USA(1982)).

In the case we have at hand, $\Gamma_D = \partial\Omega$.

Now,

$$\begin{aligned}
a(u,u) &= \int_\Omega \mu \nabla u \cdot \nabla u \, dx \\
&\geq \mu_m \int_\Omega \nabla u \cdot \nabla u \, dx \\
&= \frac{1}{2}\mu_m \int_\Omega \nabla u \cdot \nabla u \, dx + \frac{1}{2}\mu_m \int_\Omega \nabla u \cdot \nabla u \, dx \\
&\geq \frac{1}{2}\mu_m C_{\Omega,\Gamma_D} \int_\Omega u^2 \, dx + \frac{1}{2}\mu_m \int_\Omega \nabla u \cdot \nabla u \, dx \ \text{(Poincaré-Friedrichs)} \\
&\geq \min\left\{\frac{1}{2}\mu_m C_{\Omega,\Gamma_D}, \frac{1}{2}\mu_m\right\} \int_\Omega (u^2 + \nabla u \cdot \nabla u) \, dx \\
&= \alpha \, \|u\|^2_{H^1(\Omega)} \ \text{(Definition of } H^1(\Omega)\text{-norm)}
\end{aligned}$$

Summarizing, $a(\cdot,\cdot)$ is strongly coercive:

$$a(u,u) \geq \alpha \, \|u\|^2_{H^1(\Omega)} \ \forall \, u \in V \qquad (4.9)$$

Continuity of ℓ

We consider the case $g = 0$ for simplicity and left the more general case for a later moment[5].

$$\begin{aligned}
|\ell(v)| &= \left|\int_\Omega f \, v \, dx\right| \\
&\leq \left(\int_\Omega f^2 \, dx\right)^{\frac{1}{2}} \left(\int_\Omega v^2 \, dx\right)^{\frac{1}{2}} \ \text{(Cauchy-Schwarz in } L^2(\Omega)) \\
&\leq \|f\|_{L^2(\Omega)} \left(\int_\Omega (v^2 + \nabla v \cdot \nabla v) \, dx\right)^{\frac{1}{2}} \ \text{(Summing a positive term)} \\
&= N_\ell \, \|v\|_{H^1(\Omega)} \ \text{(Definition of } H^1(\Omega)\text{-norm)}
\end{aligned}$$

where N_ℓ is chosen as $\|f\|_{L^2(\Omega)}$. Summarizing, $\ell(\cdot)$ is continuous:

$$|\ell(v)| \leq N_\ell \, \|v\|_{H^1(\Omega)} \ \forall \, v \in V \qquad (4.10)$$

[5] 5: The problem is indeed well-posed under reasonable assumptions on g and $\partial\Omega$. To prove that we need the trace operator, to be introduce later on.

4.1.4 Mixed Dirichlet-Neumman BCs

We consider now the boundary $\partial\Omega$ to be the union of two disjoint parts Γ_D and Γ_N (i.e., $\Gamma_D \cup \Gamma_N = \partial\Omega$ and $\Gamma_D \cap \Gamma_N = \emptyset$)[6]

$$\begin{cases} -\nabla \cdot (\mu \nabla u) &= f \quad \text{in } \Omega \\ u &= g \quad \text{on } \Gamma_D \\ -\mu \nabla u \cdot \check{\mathbf{n}} &= h \quad \text{on } \Gamma_N \end{cases} \tag{4.11}$$

See Figure 4.4 for and illustration.

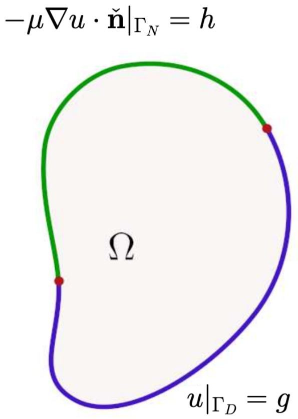


Figure 4.4: Domain Ω whose boundary is partitioned into a Dirichlet and a Neumann part.

First, we define the space

$$H^1_{D0}(\Omega) = \{v \in H^1(\Omega) : v|_{\Gamma_D} = 0\},$$

and integrating by parts, we retain the integral over Γ_N, i.e.,

$$\int_\Omega \mu \nabla u \cdot \nabla v \, dx - \int_{\Gamma_N \cup \Gamma_D} \mu \nabla u \cdot \check{\mathbf{n}} \, v \, ds = \int_\Omega \mu \nabla u \cdot \nabla v \, dx + \int_{\Gamma_N} h \, v \, ds \tag{4.12}$$

We again consider the split $u = u_0 + u_g$ giving

Weak form of Poisson's with Mixed data

Find $u_0 \in H^1_{D0}(\Omega)$ s.t. $\int_\Omega \mu \nabla u_0 \cdot \nabla v \, dx =$

$$\int_\Omega f \, v \, dx - \int_\Omega \mu \nabla u_g \cdot \nabla v \, dx - \int_{\Gamma_N} h \, v \, ds \quad \forall v \in H^1_{D0}(\Omega).$$

As before, the bilinear form remains the same, but the space changed and the linear form became[7]

$$\ell(v) = \int_\Omega f \, v \, dx - \int_\Omega \mu \nabla u_g \cdot \nabla v \, dx - \int_{\Gamma_N} h \, v \, ds \tag{4.13}$$

so, the difficult part is to show that $\ell(\cdot)$ is continuous. In chapter 9 we present a theorem that establishes under which conditions the general case, having $h \neq 0$ and $g \neq 0$ is well-posed. By now, we conclude the study of Poisson's problem by considering an important case:

4.1.5 The case of $\Gamma_D = \emptyset$

The case of **pure Neumann** boundary conditions deserves some attention

$$\begin{cases} -\nabla \cdot (\mu \nabla u) &= f \quad \text{in } \Omega \\ -\mu \nabla u \cdot \check{\mathbf{n}} &= h \quad \text{on } \partial\Omega \end{cases} \tag{4.14}$$

First, notice that if we hope to have a solution, the data f and h must satisfy a **compatibility** condition, which comes from integrating the

[6] In Equation 4.11, we define the Neumann BC in terms of the flux of quantity $-\mu\nabla u$, which is defined, as usual, with the $-$ sign in front of.

[7] The Neumann BCs are also named **natural** boundary conditions, since h appears naturally in the variational formulation when we integrate by part.

equation over Ω and applying Gauss' theorem

$$-\int_{\Omega} \nabla \cdot (\mu \, \nabla u) \, dx = -\int_{\partial\Omega} (\mu \, \nabla u) \cdot \check{\mathbf{n}} \, ds = \int_{\partial\Omega} h \, ds = \int_{\Omega} f \, dx \qquad (4.15)$$

Also, notice that if the problem admits a solution, this will be determined up to an additive constant c, i.e., if u satisfies Equation 4.14, so do $u + c$. We can not proceed as in the previous case and apply Poincaré-Friedrichs inequality, because meas$(\Gamma_D) = 0$, however, if we consider the space of functions with "zero mean",

$$V(\Omega) = \{v \in H^1(\Omega) : \int_{\Omega} v \, dx = 0\},$$

the problem is indeed well-posed. The proof relies on a similar inequality (The Poincaré-Neumann inequality), which is left aside now for the sake of brevity.

4.2 Linear Elasticity

As a final example of an elliptic problem we consider the classical linear isotropic elasticity widely used in solid mechanics. This is a problem involving a vector field as unknown, the displacement $\mathbf{u} : \Omega \longrightarrow \mathbf{R}^d$ which describes the deformation of a solid material with respect to a reference configuration. By now, we consider the simplest case involving homegeneous restrictions for the displacement on $\Gamma_{\mathbf{u}}$ on the Dirichlet boundary, (such that any rigid body motion is absent) and homegeneous forces on $\Gamma_{\mathcal{F}}$ (the Neumann boundary), $\Gamma_{\mathbf{u}} \cap \Gamma_{\mathcal{F}} = \emptyset$.

$$\begin{cases} -\nabla \cdot \left(2\,\mu \nabla^S \mathbf{u} + \lambda \, \nabla \cdot \mathbf{u} \, \mathbf{I}\right) & = & \mathbf{f} & \text{in } \Omega \\ \mathbf{u} & = & 0 & \text{on } \Gamma_{\mathbf{u}} \\ (2\,\mu \nabla^S \mathbf{u} + \lambda \, \nabla \cdot \mathbf{u}) \cdot \check{\mathbf{n}} & = & 0 & \text{on } \Gamma_{\mathcal{F}} \end{cases} \qquad (4.16)$$

where $\mathbf{I} \in \mathbb{R}^{d \times d}$, $\mu > 0$ and λ (> 0 in general) are known as the Lamé's parameters , which relate to the perhaps better known Young and Poisson modulus. Let define $\boldsymbol{\varepsilon}(\mathbf{u}) = \frac{1}{2}(\nabla \mathbf{u} + \nabla^{\intercal} \mathbf{u})$. Recall, the variational (or **weak**) formulation is obtained after multiplying by a test function $\mathbf{v}$ and integrating by parts: Find $\mathbf{u} \in V$ such that

$$\underbrace{\int_{\Omega} [2\,\mu \boldsymbol{\varepsilon}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v}) + \lambda \, (\nabla \cdot \mathbf{u}) \, (\nabla \cdot \mathbf{v})] \, dx}_{a(\mathbf{u},\mathbf{v})} = \underbrace{\int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, dx}_{\ell(\mathbf{v})} \quad \forall \mathbf{v} \in V. \qquad (4.17)$$

Figure 4.5: Gabriel Lamé (France, 1795-1870)

Now, we need a space for vector-valued functions. It turns out that the appropriate setting for this problem should be one in which $\mathbf{u}$ and its first weak derivatives, encoded in the $\boldsymbol{\varepsilon}(\mathbf{u})$ tensor, be square integrable, i.e.[8],

$$\left[H_0^1(\Omega)\right]^d = \{\mathbf{v} \in \left[H^1(\Omega)\right]^d : \mathbf{v}|_{\Gamma} = 0\}. \qquad (4.18)$$

[8] In Equation 4.18, the space $\left[H^1(\Omega)\right]^d$ is the cartesian product of $H^1(\Omega)$ d times of , e.g., for $d = 2$, $\left[H^1(\Omega)\right]^2 = H^1(\Omega) \times H^1(\Omega)$, so a function $\mathbf{v} \in \left[H^1(\Omega)\right]^2$, means $v_i \in H^1(\Omega), i = 1, 2$.

Collecting all the ingredients:

$$\left[\begin{array}{rl}
\text{The space:} & V = \left[H_0^1(\Omega)\right]^d \\
\text{The inner product on } V\text{:} & (\mathbf{u},\mathbf{v})_V = \displaystyle\int_\Omega (\mathbf{u}\cdot\mathbf{v} + \nabla\mathbf{u} : \nabla\mathbf{v})\, dx = (\mathbf{u},\mathbf{v})_{[H_0^1(\Omega)]^d} \\
\text{The induced norm on } V\text{:} & \|\mathbf{u}\|_V^2 = \displaystyle\int_\Omega (\mathbf{u}\cdot\mathbf{u} + \nabla\mathbf{u}\cdot\nabla\mathbf{u})\, dx = \|\mathbf{u}\|^2_{[H^1(\Omega)]^d} \\
\text{The bilinear form:} & a(\mathbf{u},\mathbf{v}) = \displaystyle\int_\Omega \left[2\,\mu\,\boldsymbol{\varepsilon}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v}) + \lambda\,(\nabla\cdot\mathbf{u})\,(\nabla\cdot\mathbf{v})\right]\, dx \\
\text{The linear form:} & \ell(\mathbf{v}) = \displaystyle\int_\Omega \mathbf{f}\cdot\mathbf{v}\, dx
\end{array}\right.$$

The well-posedness is shown by following the same script as in the previous problem. In order to apply the Riesz representation theorem, continuity of $\ell(\cdot)$ and $a(\cdot,\cdot)$ and its coercivity has to be proven. The more general case, with nonhomeogeneous conditions on $\Gamma_{\mathcal{F}}$ and the details of the proof are left for Chapter 9.

4.3 A different view point: Energy minimization

So far we have been dealing with the differential form (**DF**) and with the variational form (**VF**) of boundary value problems. Now, *in the case of $a(\cdot,\cdot)$ being a symmetric bilinear form*, we explore a different (though equivalent) way of looking at these problems, namely, the Extremal Formulation (**EF**). To that end consider the Poisson's problem and the functional $\mathcal{J} : V \longrightarrow \mathbb{R}$ given by

$$\mathcal{J}(v) = \frac{1}{2}\int_\Omega \mu\,\nabla v\cdot\nabla v\, dx - \int_\Omega f\, v\, dx \tag{4.19}$$

Notice that $\mathcal{J}(v) = \frac{1}{2}a(v,v) - \ell(v)$, which is called the energy of the system. Problem (**VF**):

$$\text{Find } u \in H_0^1(\Omega) \;\text{ s.t. }\; a(u,v) = \ell(v) \;\; \forall v \in H_0^1(\Omega). \tag{4.20}$$

can be recast as the following minimization problem:

Extremal form of Poisson's problem

Find u such that

$$(\mathbf{EF}):\quad u \;=\; \underset{v\in H_0^1(\Omega)}{\operatorname{argmin}}\, \mathcal{J}(v) \tag{4.21}$$

We can show that if u is the unique solution to (**VF**), then it is also the unique solution to (**EF**). To see that, we proceed as follows. Being $a(\cdot,\cdot)$ a **symmetric** and **strongly coercive** bilinear form, for all $v \in H_0^1(\Omega)$ we have that

$$
\begin{aligned}
\mathcal{J}(v)-\mathcal{J}(u) &= \left[\frac{1}{2}a(v,v)-\ell(v)\right]-\left[\frac{1}{2}a(u,u)-\ell(u)\right]= \\
&= \left[\frac{1}{2}a(v,v)-\frac{1}{2}a(u,u)\right]-\left[\ell(v)-\ell(u)\right]= \\
&= \left[\frac{1}{2}a(v,v)-\frac{1}{2}a(u,u)\right]-\ell(v-u)= \\
&= \left[\frac{1}{2}a(v,v)-\frac{1}{2}a(u,u)\right]-a(u,v-u)= \\
&= \frac{1}{2}\left[a(v,v)-a(u,u)\right]-a(u,v)+a(u,u)= \\
&= \frac{1}{2}\left[a(v,v)-2\,a(u,v)+a(u,u)\right]= \\
&= \frac{1}{2}\left[a(v,v)-a(u,v)-a(v,u)+a(u,u)\right]= \\
&= \frac{1}{2}\left[a(v-u,v)-a(v-u,u)\right]= \\
&= \frac{1}{2}\left[a(v-u,v-u)\right]\geq\frac{\alpha}{2}\|v-u\|^2_{H^1_0(\Omega)}
\end{aligned}
$$

this is, $\mathcal{J}(v)\geq\mathcal{J}(u)\forall v\in H^1_0(\Omega)$ and thus u minimizes $\mathcal{J}(\cdot)$ over $H^1_0(\Omega)$.

As a final comment, we can also define an equivalent minimization problem for the elasticity problem, in which we seek for a minimizer of the following potential energy

$$
\mathcal{J}(\mathbf{v})=\int_\Omega\left[\mu\boldsymbol{\varepsilon}(\mathbf{v}):\boldsymbol{\varepsilon}(\mathbf{v})+\frac{\lambda}{2}(\nabla\cdot\mathbf{v})^2\right]dx-\int_\Omega\mathbf{f}\cdot\mathbf{v}\,dx \tag{4.22}
$$

in this case, it is perhaps more natural to talk about an energy of the system: which is divided into a stored energy of the solid and a potential energy of the loading. This means that the solution $\mathbf{u}\in\left[H^1_0(\Omega)\right]^d$ of Equation 9.10, is the one that renders the minimum energy of the solid, a fact that is well known.

Finally, to conclude this part:

Solve the following exercices

- Write the variational formulation of the following problem that involves **Robin**[9] BCs:

$$
\begin{cases}
-\nabla\cdot(\mu\nabla u) = f & \text{in }\Omega \\
u = g & \text{on }\Gamma_D \\
\mu\nabla u\cdot\check{\mathbf{n}}+\gamma u = h & \text{on }\Gamma_R
\end{cases}
$$

  where $\gamma\in\mathbb{R}$ is a given parameter and $\Gamma_D\cup\Gamma_R=\partial\Omega$ (and $\Gamma_D\cap\Gamma_R=\emptyset$). In fluid mechanics, this condition serves for instance to model heat exchange between a solid wall and an ambient fluid by convection.

[9] Victor Gustave Robin (1855-1897) (a.k.a. Batman's best friend).

- Let $\Omega = (0,1)$. Let $\varphi(x) = x$. Think on a sequence $\{\varphi_n\} \subset V_{N0}(\Omega) = \{v \in H^1(\Omega),\ v'(0) = 0\}$. i.e, $\varphi_n'(0) = 0$ for all n and such that $\varphi_n \longrightarrow \varphi$ in the $H^1(\Omega)$-norm.
  Hint: For $1/n = \epsilon > 0$ consider the "trimmed" function
  $$T_{\epsilon(n)}\varphi(x) = \begin{cases} \varphi(\epsilon) & \text{if } x < \epsilon(n) \\ \varphi(x) & \text{if } x \geq \epsilon(n) \end{cases}$$
  This example shows that the set of functions that satisfy a certain Neumann boundary condition does not necessarily contains all its accumulation points (in this case, the $V_{N0}(\Omega)$ is not closed in $H^1(\Omega)$).
- Show that if u is the unique mimimizer of $\mathcal{J}$ (i.e., the solution to (**EF**) over $H_0^1(\Omega)$), then u is the unique solution of (**VF**) (Equation 4.20).
- **Elasticity problem:** Show that the bilinear form of Equation 9.10 is symmetric, so the bilinear form defines an inner product and Riesz representation theorem can be applied.

4.4 Assignment 3: Variants of Poisson's problem

We continue with the implementation of some variants of Poisson's problem into the FEniCSx platform. The problem we aim to solve now is the following:

Consider a rectangular domain with some holes, i.e., $\Omega = ([0, L_x] \times [0, L_y]) \setminus \Omega_{\text{holes}} \subset \mathbb{R}^2$, $\partial\Omega = \Gamma_{\text{left}} \cup \Gamma_{\text{bottom}} \cup \Gamma_{\text{right}} \cup \Gamma_{\text{top}} \cup \Gamma_{\text{holes}}$, source term f and thermal diffusivity μ. The problem reads: Find u such that

$$\begin{cases} -\nabla \cdot (\mu \nabla u) &= f & \mathbf{x} \in \Omega \\ u &= u_\ell & \mathbf{x} \in \Gamma_{\text{left}} \\ u &= u_r & \mathbf{x} \in \Gamma_{\text{right}} \\ \mu \nabla u \cdot \check{\mathbf{n}} + \gamma u &= h_t & \mathbf{x} \in \Gamma_{\text{top}} \\ \mu \nabla u \cdot \check{\mathbf{n}} + \gamma u &= h_b & \mathbf{x} \in \Gamma_{\text{bottom}} \\ u &= u_i & \mathbf{x} \in \Gamma_i \end{cases}$$

where Γ_i is the boundary of the i–th hole, with centers and radii $\{\mathbf{x}_c^i, r_i\}_{i=1}^{n_b}$, being $\Gamma_{\text{holes}} = \cup_{i=1}^{n_b} \Gamma_i$. A mesh and a typical result are shown in Figure 4.6.

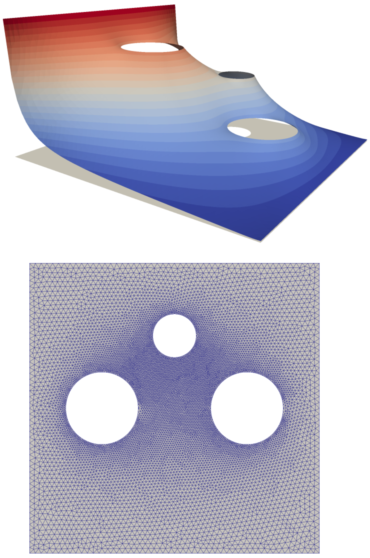


Figure 4.6: Typical result and mesh considering homogenous source term, Dirichlet conditions on the holes, left and right walls and Robin type conditions in the bottom and top walls.

The discrete variational formulation

As in the previous chapter we consider a discrete version of the variational formulation, similar to the one you should have obtained in the previous theoretical excercise: Find $u_h \in V_{hg}$ such that

$$\int_\Omega \mu \, \nabla u_h \cdot \nabla v_h \, dx + \int_{\Gamma_{\text{top}} \cup \Gamma_{\text{bottom}}} \gamma u_h \, v_h \, ds = \int_\Omega f \, v_h \, dx + \\ + \int_{\Gamma_{\text{top}}} h_t \, v_h \, ds + \int_{\Gamma_{\text{bottom}}} h_b \, v_h \, ds \quad \forall v_h \in V_{h0}, \tag{4.23}$$

where

$$V_{hg} \subset V_g(\Omega) = \{v \in H^1(\Omega), \; v|_{\Gamma_{\text{left}}} = u_\ell, \; v|_{\Gamma_{\text{right}}} = u_r, v|_{\Gamma_i} = u_i\}$$

Implement the following modifications to `variants_poisson.py`

- Inspect the script provided and make sure to understand its different sections.
- In the first place, consider $\gamma = 0$, so the Robin condition is nothing but a Neumann (or natural condition). Take $h_t = h_b = 1$, $\mu = 1$, $f = 0$, $u_\ell = 100$, $u_r = 1$ and u_i, $i = 1, 2, 3$ being the mean temperatures you obtained in the previous assignment in each inclusion when μ_B is set to a very large value. Compare the temperature field with that of the previous assignment.
- Now consider the same conditions of the previous item but $\gamma = 1$. Plot the results.
- Consider the coefficient γ assumes different values (e.g., 10^{-5}, $0.1, 1.0, 10.0, 100.0, 1000.0$). Take $f = 0, \mu = 1, u_\ell = u_r = 10$, the same u_i's of the previous item and $h_t = h_b = \gamma u_\ell$. Implement a code that makes continuation over γ and interpret the results by plotting the solution so as to appreciate what's going on as γ grows. Base on the following code to implement:

```
xdmfrob = XDMFFile(MPI.COMM_WORLD, "robin.xdmf", "w")
xdmfrob.write_mesh(mesh)
for gam in [1.0e-5, 0.1, 1.0, 10.0, 100.0, 1000.0]:
    gamma = Constant(mesh, ScalarType(gam))
    ht = ...
    hb = ...
    a = inner(mu*grad(u), grad(v))*dx + ...
    L = source * v * dx + ...
    .
    .
    xdmfrob.write_function(uh, gam)
```

- In the previous item, now let $u_r = 10$, but set $u_\ell = 100$. Run the code and visualize the results. Can you explain what is going on when γ is large?
- Lastly, take $f = 0$, $\mu = 1$, $\gamma = 0$ and consider boundary condi-

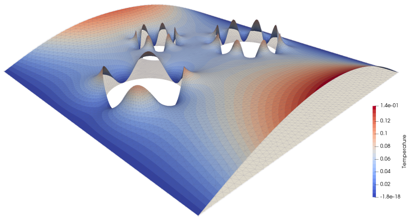


Figure 4.7: Result corresponding to a Dirichlet condition $u_i = 0.1 \sin^2[(i + 2)\theta_i]$, $\theta_i = \operatorname{atan}\left(\frac{x_2 - x^i_{c2}}{x_1 - x^i_{c1}}\right)$, $i = 1, 2, 3$ at holes' boundaries.

tions given as functions depending on **x** as follows:

- $u_r = u_\ell = 0$;
- $h_t = h_b = 0.2x_1\,(L_x - x_1)$;
- $u_i = 0.1\sin^2[(i+2)\theta_i]$, $\theta_i = \text{atan}\left(\frac{x_2 - x^i_{c2}}{x_1 - x^i_{c1}}\right)$, $i = 1, 2, 3$.

To implement the Dirichlet conditions consider the code:

```
def gfunc1(x):
    ...
    return ...
u_aux = Function(V)
u_aux.interpolate(gfunc1)
bc_hole1 = dirichletbc(u_aux, hole1_dofs)
.
.
```

► Prepare a short presentation to report the results.

VARIATIONAL FORMULATIONS III: Galerkin

5

5.1 The Galerkin method

In this chapter we introduce a key ingredient of the finite elememt method, namely, the Galerkin method (also referred to as Ritz-Galerkin) , which takes as starting point the variational formulation. Let recall the abstract variational problem: Find $u \in V$ such that

$$a(u,v) \quad = \quad \ell(v) \quad \forall v \in V \tag{5.1}$$

of which we have just seen several examples. As commented in the motivational example of Chapter 1, the idea behind the Galerkin method is simply to replace the space V by a finite dimensional space $V_h \subset V$. The space V_h can be spanned by a set of linearly independent functions in V, $\{\phi_1, \phi_2, \ldots, \phi_n\}$ [1], so, we write a discrete variational formulation (**VF**$_h$): Find $u_h \in V_h$ such that

$$a(u_h, v_h) \quad = \quad \ell(v_h) \quad \forall v_h \in V_h \tag{5.2}$$

The idea is that $u_h \in V_h$ defined by

$$u_h(\mathbf{x}) = \sum_{j=1}^{n} U_j \, \phi_j(\mathbf{x}) \tag{5.3}$$

will get closer and closer to the solution $u \in V$ as $n \longrightarrow \infty$.

Figure 5.1: Boris Galerkin (Russia, 1871-1945).

Assuming that problem given by Equation 5.1 is well posed, the logical question that emerges is if the new problem given in Equation 5.2, is also well posed ...

The answer is yes, because the discrete problem inherits the properties of the continuous one. First, observe that:

- Being V Hilbert, a finite dimensional subspace $V_h \subset V$ is also Hilbert[2];
- Since a is bounded in V (i.e., $|a(u,v)| \leq N_a \, \|u\|_V \, \|v\|_V \; \forall u, v \in V$), it is in particular bounded in V_h;
- Since a is strongly coercive (i.e., $a(u,u) \geq \alpha \, \|u\|_V^2 \; \forall u \in V$) it is in particular coercive in V_h;
- For the same reason, $\ell(v)$ is also continuous in V_h;

As a corollary of Theorem 3.3.2 in case $a(\cdot,\cdot)$ is symmetric or Theorem 3.4.1, if it is not necessarily symmetric, we have the:

---

[1] Recall, the subindex h is used to denote finite dimensional or discrete spaces. The h in the finite element method is related to the refinement of a partition of Ω (the mesh), such that $h \longrightarrow 0$ as $n \longrightarrow \infty$.

[2] A finite dimensional subspace is necessarily closed.

Corollary 5.1.1 *There exists a unique $u_h \in V_h$ that satisfies the problem ($\mathbf{VF}_h$): Find $u_h \in V_h$ such that*

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

The next question is if u_h is computable ...

The interesting thing about the Galerkin method is that it provides the **methodology** to compute u_h. If we choose a basis $\{\phi_1, \phi_2, \ldots, \phi_n\}$ of V_h, replace the Equation 5.3 and take $v_h = \phi_i$ in Equation 5.2, we obtain

$$a\left(\sum_{j=1}^{n} U_j \phi_j, \phi_i\right) = \sum_{j=1}^{n} \underbrace{a(\phi_j, \phi_i)}_{A_{ij}} U_j = \underbrace{\ell(\phi_i)}_{b_i}, \quad i = 1, \ldots, n \tag{5.4}$$

which is clearly a linear system of equations

$$\mathsf{A}\,\mathbf{U} = \mathsf{b} \tag{5.5}$$

where $\mathbf{U} \in \mathbb{R}^n$. $\mathsf{A} \in \mathbb{R}^{n\times n}$ and $\mathsf{b} \in \mathbb{R}^n$ are defined by

$$A_{ij} = a(\phi_j, \phi_i), \quad b_i = \ell(\phi_i). \tag{5.6}$$

About matrix A we can say:

- if $a(\phi_j, \phi_i) = a(\phi_i, \phi_j)$ (symmetry of a), clearly A is also symmetric;
- if $a(v_h, v_h) > 0$ (strong coercivity), taking any arbitrary nonzero $v_h = \sum_{j=1}^{n} V_j \phi_j$ (i.e., $\mathbf{V} \neq \mathbf{0}$), we get

  $$a\left(\sum_{j=1}^{n} V_j \phi_j, \sum_{i=1}^{n} V_i \phi_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} V_i A_{ij} V_j = \mathbf{V}^{\mathsf{T}}\mathsf{A}\mathbf{V} > 0, \quad \mathbf{V} \neq \mathbf{0} \tag{5.7}$$

  thus, A is positive definite and therefore **invertible**, so, we can safely solve the system and find the unique $\mathbf{U}$ that defines u_h.

5.2 Galerkin orthogonality & Best approximation

An important property of the Galerkin solution u_h to ($\mathbf{VF}_h$) can be stated. Begin by saying that

$$\begin{aligned} a(u, v_h) &= \ell(v_h) \quad \forall v_h \in V_h \\ a(u_h, v_h) &= \ell(v_h) \quad \forall v_h \in V_h \end{aligned}$$

so, substracting

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

This fact, allows us to show the following important lemma:

Lemma 5.2.1 *(**J. Céa**) Let $a(\cdot,\cdot)$ and $\ell(\cdot)$ be continuous forms in V, with $a(\cdot,\cdot)$ strongly coercive, then*

$$\|u-u_h\|_V \le \frac{N_a}{\alpha}\|u-v_h\|_V \ \ \forall v_h \in V_h \tag{5.8}$$

This can be shown by using the trick:

$$\begin{aligned} a(u-u_h, u-u_h) &= a(u-u_h, u-u_h+v_h-v_h) = \\ &= \underbrace{a(u-u_h, v_h-u_h)}_{=0} + a(u-u_h, u-v_h) \\ &= a(u-u_h, u-v_h) \ \ \forall v_h \in V_h \end{aligned}$$

but, by the coercivity $a(u-u_h, u-u_h) \ge \alpha\|u-u_h\|_V^2$, then

$$\|u-u_h\|_V^2 \le \frac{1}{\alpha}a(u-u_h,u-u_h) = \frac{1}{\alpha}a(u-u_h,u-v_h) \ \ \forall v_h \in V_h \tag{5.9}$$

and by the continuity $a(u-u_h, u-v_h) \le N_a\|u-u_h\|_V\|u-v_h\|_V$, implying that

$$\|u-u_h\|_V^2 \le \frac{N_a}{\alpha}\|u-u_h\|_V\|u-v_h\|_V \ \ \forall v_h \in V_h. \tag{5.10}$$

This is equivalent to say that[3]:

$$\|u-u_h\|_V \le \frac{N_a}{\alpha}\min_{v_h\in V_h}\|u-v_h\|_V. \tag{5.11}$$

So far, it has not been assumed that $a(\cdot,\cdot)$ is symmetric. If a happens to be, so it defines an inner product over V, then:

- $a(u-u_h, v_h) = 0 \ \forall \ v_h \in V_h$ implies that the error $e_h = u-u_h$ is orthogonal to V_h, therefore, we say that the Galerkin approximation u_h is the **orthogonal projection** of u onto V_h (see Figure 5.2);
- The Galerkin approximation u_h is **optimal** in the energy norm, i.e.,

  $$\|u-u_h\|_a \le \|u-v_h\|_a \ \ \forall v_h \in V_h. \tag{5.12}$$

  since in this norm the coercivity and continuity constants are both equal to 1.

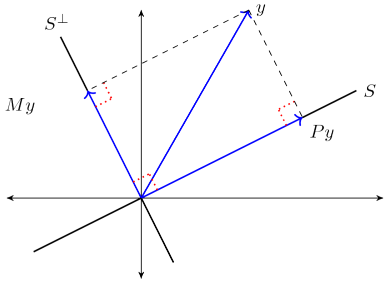


Figure 5.2: Classical picture of orthogonal projection in Euclidean spaces.

5.3 Distance to V_h

Let us briefly comment on an altenative way of thinking these issues. First, recall that given an inner product, we have a natural way of defining a **distance** $d(\cdot,\cdot)$ between two functions in the space V, i.e.,

$$d(f,g) = \|f-g\|_V = \sqrt{(f-g, f-g)_V}$$

---

[3] Note that we may run into trouble if $\frac{N_a}{\alpha} \gg 1$, because this does not provide a practical control on the error. Saying that $\|u-u_h\|_V$ is lower than a very large number is not that useful.

This can be the $L^2(\Omega)$ or the $H^1(\Omega)$-inner products. It can be shown that there exists one and only one element $u_h \in V_h$ such that

$$d(u, u_h) \leq d(u, v_h) \;\; \forall v_h \in V_h \tag{5.13}$$

and u_h is the orthogonal projection of u onto V_h. To construct such u_h, let us suppose that a u_h that satisfies Equation 5.13 exists. Then, given any $v_h \in V_h$ the functional

$$j(s) = d(u, u_h + s\, v_h)^2 = \|u - (u_h + s\, v_h)\|_V^2 \tag{5.14}$$

will have a minimum at $s = 0$, i.e., if u_h minimizes $d(u, u_h)^2$, then $j(0) \leq j(s) \;\forall s \in \mathbb{R}$. Using the linearity and symmetry of the inner product we show that

$$\begin{aligned} j(s) = d(u, u_h + s\, v_h)^2 &= (u - (u_h + s\, v_h), u - (u_h + s\, v_h))_V = \\ &= (u - u_h, u - u_h)_V - 2\, s\, (u - u_h, v_h)_V + s^2\, (v_h, v_h)_V \\ &= \|u - u_h\|_V^2 - 2\, s\, (u - u_h, v_h)_V + s^2\, \|v_h\|_V^2 \end{aligned}$$

Now, to have a minimum of this functional for all $v_h \in V_h$ when $s = 0$, its derivative with respect to s is necessarily equal to zero at $s = 0$, i.e.,

$$\frac{dj}{ds}(s = 0) = -2\, (u - u_h, v_h)_V = 0 \;\; \forall\, v_h \in V_h$$

which, again, means that the difference between u and u_h is orthogonal to all $v_h \in V_h$ and u_h satisfies[4]

$$(u_h, v_h)_V = (u, v_h)_V \;\; \forall v_h \in V_h$$

[4] Notice that this is a variational problem in which the bilinear form $a(\cdot, \cdot)$ and the linear form $\ell(\cdot)$ are given by the inner product:

$$\begin{cases} a(u_h, v_h) = (u_h, v_h)_V \\ \\ \ell(v_h) = (u, v_h)_V \end{cases}$$

5.4 Conclusion - Consequence of Céa's lemma

The distance from u to V_h estimates de error, so the idea is that if we build a sequence of discrete function spaces $\{V_{h,n}\}$ (as is done in the FEM) with $h \longrightarrow 0$ (as $n \longrightarrow \infty$), we pretend that

$$\lim_{h \to 0} \left(\min_{v_h \in V_h} \|u - v_h\| \right) = 0 \tag{5.15}$$

or in simple terms we aim

$$\lim_{h \to 0} u_h = u \tag{5.16}$$

In the next few chapters we dedicate to study how to construct discrete function spaces on partitions of Ω (**meshes**), in particular, piecewise polynomial spaces and which are their approximability and interpolation properties. Concerning what we have just presented, this is fundamental because, as we shall see

$$\min_{v_h \in V_h} \|u - v_h\|_V \leq \|u - \mathcal{I}_h u\|_V,$$

where $\mathcal{I}_h u$ stands for the interpolant of the exact solution u onto the space V_h.

So, the plan for the next part is to introduce:

- Polynomial spaces on finite element partitions of Ω;
- Interpolation operators;
- Local estimates of the interpolation error;
- Global error estimates;
- Some issues about regularity of meshes;

... but, before that, to fix some of the ideas of this chapter we must:

Solve the following exercices

- (**Linear Algebra**) Show that of if $\mathsf{A} \in \mathbb{R}^{n\times n}$ is positive definite ($\mathbf{V}^\intercal\mathsf{A}\mathbf{V} > 0\ \forall V \in \mathbb{R}^n$), then, A is invertible.

- In the optimallity of the Galerkin approximation u_h, it is used that in the energy norm $\|\cdot\|_a$, the coercivity and continuity constants are both equal to 1. Explain.

5.5 Assignment 4: Projection of functions

Consider the variational problem (orthogonal projection): Find $u_h \in V_h$ such that

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

where

$$a(u_h, v_h) = (u_h, v_h)_V, \quad \ell(v_h) = (u, v_h)_V$$

in which u is a given function and V denotes the $L^2(\Omega)$ inner product, i.e.,

$$(v, w)_{L^2(\Omega)} = \int_\Omega v\, w\, dx$$

or the $H^1(\Omega)$-inner product

$$(v, w)_{H^1(\Omega)} = \int_\Omega (v\, w + \nabla v \cdot \nabla w)\, dx$$

Notice that in this problem no boundary conditions are imposed.

- Implement a script that solves the projection problem;

- First, consider Ω to be the unit interval $[0, 1]$, partitioned into `nels` elements, i.e.

```
nels = ...
msh = mesh.create_interval(comm=MPI.COMM_WORLD,
                           points=((0.0), (1.0)),
                           nx=nels)
```

  and let u be the function given by

$$u(x) = \sin(2\pi\, x)$$

- Take V_h to be a space of piecewise polynomial functions of different degree, i.e., functions that are polynomials of degree k (constants, linears, quadratics, etc) on each element and discontinuous between elements[5]:

  ```
  degree = ...
  V = fem.FunctionSpace(msh, ("DG", degree))
  ```

  Compute u_h, the projection of u onto V_h with the $L^2(\Omega)$ inner product. Take different number of elements (`nels = 1, 2, 4, 8, 12, 16, 32`) and `degree = 0, 1, 2`.

- Repeat the previous item using the $H^1(\Omega)$ inner product, but now take a space of continuous functions between elements[6]:

  ```
  degree = ...
  V = fem.FunctionSpace(msh, ("CG", degree))
  ```

  with `degree = 1, 2, 3`. This is alse referred to as a conforming space.

- Plot the solution by reconstructing the corresponding polynomial on each element. To that end consider accessing the solution values of u_h from the array:

  ```
  .
  uh.x.array[:]
  .
  ```

  Interpret the number of components in the array and the values according to the case being considered (discontinuous or continuous functions and the degree k of the polynomial on each element of the partition). Then, use the function `numpy.polyfit` to build the polynomials on each element and plot using `matplotlib` (see the example in Figure 5.3).

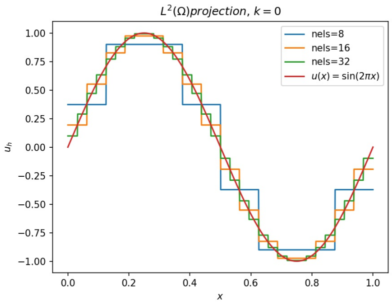


Figure 5.3: Example of the $L^2(\Omega)$ orthogonal projection of $u(x) = \sin(2\pi x)$ onto spaces of piecewise constant functions associated to partitions of the unit interval.

► Now, consider a 2D case, the unit square $\Omega = [0,1]^2$ and the function

$$u(x_1, x_2) = \sin(\pi\, x_1)\, \sin(\pi\, x_2)$$

Compute the $L^2(\Omega)$ projection onto a space of piecewise constant (discontinuous) functions and the $H^1(\Omega)$ projection onto a space of continuous piecewise linear functions. Visualize in `Paraview`.

► Prepare a short monograph to report the results;

[5] These spaces are named as Discontinuous Galerkin (`DG` in `FEniCS`)

[6] These spaces are named as Continuous Galerkin (`CG` or `Lagrange` in `FEniCS`). Details about how all these spaces are actually defined and built are explained in the next chapter.

FINITE ELEMENT SPACES 6

In the following sections the aim is to construct finite element spaces V_h to solve this problem. We begin with a few simple examples, introduce the concept of degrees of freedom and also some classical finite element basis.

6.1 Simple examples in 1D

6.1.1 A space of polynomial functions in (a, b)

Consider the interval (a, b). We define the space as

$$V_h = P_k(a,b) = \{v, v = \sum_{i=0}^{k} \alpha_k\, x^k\} \tag{6.1}$$

e.g. $k = 1$

$$P_1(a,b) = \{v, v = \alpha + \beta\, x\} \tag{6.2}$$

This space has dimension 2. Once we have defined the space, we proceed like this:

1. Define a set of degrees of freedom $\{\sigma_1, \sigma_2\}$ (i.e., a set of linear functionals of P_k in $\mathbb{R}$).
2. Define $\{\phi^1(x), \phi^2(x)\}$ by the relation: $\sigma_i(\phi^j) = \delta_{ij}$
   (this is the Kronecker delta property, i.e., $\delta_{ij} = 1$ if $i = j$ and 0 otherwise).

One choice is to consider as degrees of freedom the value of the function at the end points of the interval:

$$\sigma_1(v) = v(a) \tag{6.3}$$
$$\sigma_2(v) = v(b) \tag{6.4}$$

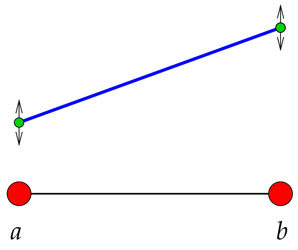


Figure 6.1: Space $P_1(a, b)$.

To compute the basis, consider functions $\phi^j(x) = \alpha_j + \beta_j\, x$. In order to find α_j and β_j, $j = 1, 2$ we have two 2×2 systems to solve:

$$\sigma_1(\phi^1) = \alpha_1 + \beta_1\, a = 1, \quad \sigma_1(\phi^2) = \alpha_2 + \beta_2\, a = 0$$
$$\sigma_2(\phi^1) = \alpha_1 + \beta_1\, b = 0, \quad \sigma_2(\phi^2) = \alpha_2 + \beta_2\, b = 1$$

Therefore, the basis is:

$$\phi^1(x) = \frac{b - x}{b - a}, \quad \phi^2(x) = \frac{x - a}{b - a} \tag{6.5}$$

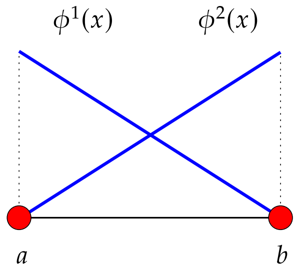


Figure 6.2: P_1 basis function for one single element.

The last choice seemed arbitrary, but it is a very practical one. If we want to describe a function in that space, the only thing we need is the function values at $x^1 = a$ and $x^2 = b$, since $\phi^i(x^j) = \delta_{ij}$, i.e.

$$w(x) = \sum_{i=1}^{2} W^j \, \phi^j(x) = \underbrace{W^1}_{w(a)} \, \phi^1(x) + \underbrace{W^2}_{w(b)} \, \phi^2(x) = w(a)\, \phi^1(x) + w(b)\, \phi^2(x)$$

A numerical method is said to be a Galerkin finite element method if:

- It is based on a variational formulation, i.e., Find $u_h \in V_h$ such that

  $$a(u_h, v_h) = \ell(v_h) \qquad \forall\, v_h \in V_h \tag{6.6}$$

  where $a(\cdot,\cdot) : V_h \times V_h \longrightarrow \mathbb{R}$ is a bilinear form and $\ell : V_h \longrightarrow \mathbb{R}$ is a linear form. Both, a and ℓ corresponds to the exact problem.

- The discrete space V_h is a **finite element space.**

In what follows we consider a prototypical 1D model problem:

1D Diffusion-Reaction problem

Determine $u_h \in V_h \subset H^1(0,1)$, such that $u_h(0) = 0$ and that

$$\int_0^1 (u_h' \, v_h' + \theta \, u_h \, v_h) \, dx = \int_0^1 f \, v_h \, dx \tag{6.7}$$

holds for all $v_h \in V_h$ satisfying $v_h(0) = 0$.

As an exercise, we aim to compute the matrix A for this problem, i.e.

$$A_{ij} = a(\phi^i, \phi^j), \;\; i, j = 1, 2$$

We write A as sum of two matrices

$$\mathsf{A} = \mathsf{K} + \theta \, \mathsf{M} \tag{6.8}$$

where the entries are computed according to:

$$K_{ij} = a_d(\phi_i, \phi_j) = \int_a^b (\phi^i)'(\phi^j)' \, dx, \quad M_{ij} = a_r(\phi_i, \phi_j) = \int_a^b \phi^i \, \phi^j \, dx$$

which by computing the integrals yield

$$K_{11} = a_d(\phi^1, \phi^1) = \int_a^b (\phi^1)'(\phi^1)' \, dx = \frac{1}{b-a},$$

$$K_{12} = K_{21} = a_d(\phi^2, \phi^1) = \int_a^b (\phi^2)'(\phi^1)' \, dx = -\frac{1}{b-a},$$

$$K_{22} = a_d(\phi^2, \phi^2) = \int_a^b (\phi^2)'(\phi^2)' \, dx = \frac{1}{b-a},$$

and

$$M_{11} = a_r(\phi^1, \phi^1) = \int_a^b \phi^1 \phi^1 \, dx = \frac{b-a}{3},$$

$$M_{12} = M_{21} = a_r(\phi^1, \phi^2) = \int_a^b \phi^1 \phi^2 \, dx = \frac{b-a}{6},$$

and so on ..., finally giving[1]

$$\mathsf{A} = \begin{bmatrix} \frac{1}{b-a} & -\frac{1}{b-a} \\ -\frac{1}{b-a} & \frac{1}{b-a} \end{bmatrix} + \theta \begin{bmatrix} \frac{b-a}{3} & \frac{b-a}{6} \\ \frac{b-a}{6} & \frac{b-a}{3} \end{bmatrix}$$

6.1.2 A polynomial space by parts

Consider the intervals (a, b) and (b, c) and define the space

$$V_h = P_k^{\text{disc}} = \{v, \; v|_{(a,b)} \in P_k(a,b), \;\; v|_{(b,c)} \in P_k(b,c)\}.$$

Consider again the case $k = 1$ for simplicity. The space has dimension 4. This is more or less evident if we notice that functions in this space are:

$$v = \begin{cases} \alpha + \beta \, x & \text{if } x \in (a,b) \\ \gamma + \epsilon \, x & \text{if } x \in (b,c) \end{cases}$$

Notice that such functions are not necessarily continuous at $x = b$ and therefore we have 4 degrees of freedom. For instance, choose for them:

$$\begin{aligned} \sigma_1(v) &= v(a) \\ \sigma_2(v) &= v(b^-) \\ \sigma_3(v) &= v(b^+) \\ \sigma_4(v) &= v(c) \end{aligned}$$

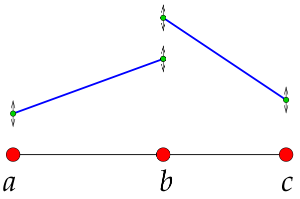


Figure 6.3: Space $P_1^{\text{disc}}(a, c)$.

Now, we can compute the basis by using the relation $\sigma_i(\phi^j) = \delta_{ij}$. We write the basis function as

$$\phi^j(x) = \begin{cases} \alpha_j + \beta_j \, x & \text{if } x \in (a,b) \\ \gamma_j + \epsilon_j \, x & \text{if } x \in (b,c) \end{cases}$$

For $j = 1, \ldots, 4$ we have

$$\begin{aligned} \sigma_1(\phi^j) &= \alpha_j + \beta_j \, a \;\; = \delta_{1j} \\ \sigma_2(\phi^j) &= \alpha_j + \beta_j \, b^- = \delta_{2j} \\ \sigma_3(\phi^j) &= \gamma_j + \; \epsilon_j \, b^+ = \delta_{3j} \\ \sigma_4(\phi^j) &= \gamma_j + \; \epsilon_j \, c \;\; = \delta_{4j} \end{aligned}$$

$$\begin{cases} \alpha_1 + \beta_1 \, a & = 1 \\ \alpha_1 + \beta_1 \, b & = 0 \\ \gamma_1 + \epsilon_1 \, b & = 0 \\ \gamma_1 + \epsilon_1 \, c & = 0 \end{cases}, \quad \begin{cases} \alpha_2 + \beta_2 \, a & = 0 \\ \alpha_2 + \beta_2 \, b & = 1 \\ \gamma_2 + \epsilon_2 \, b & = 0 \\ \gamma_2 + \epsilon_2 \, c & = 0 \end{cases}, \quad \begin{cases} \alpha_3 + \beta_3 \, a & = 0 \\ \alpha_3 + \beta_3 \, b & = 0 \\ \gamma_3 + \epsilon_3 \, b & = 1 \\ \gamma_3 + \epsilon_3 \, c & = 0 \end{cases}, \quad \begin{cases} \alpha_4 + \beta_4 \, a & = 0 \\ \alpha_4 + \beta_4 \, b & = 0 \\ \gamma_4 + \epsilon_4 \, b & = 0 \\ \gamma_4 + \epsilon_4 \, c & = 1 \end{cases},$$

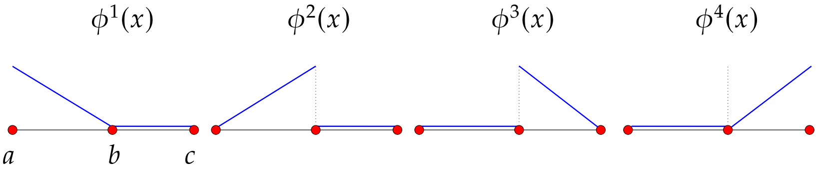
---

[1] All these computations at individual intervals or elements will be useful later on when we construct aproximations on spaces defined on a collection of such elements.

By inspection we find that the basis is:

$$\phi^1(x) = \begin{cases} \dfrac{b-x}{b-a} & \text{if } x \in (a,b) \\ \\ 0 & \text{if } x \in (b,c) \end{cases} \quad , \quad \phi^2(x) = \begin{cases} \dfrac{x-a}{b-a} & \text{if } x \in (a,b) \\ \\ 0 & \text{if } x \in (b,c) \end{cases}$$

$$\phi^3(x) = \begin{cases} 0 & \text{if } x \in (a,b) \\ \\ \dfrac{c-x}{c-b} & \text{if } x \in (b,c) \end{cases} \quad , \quad \phi^4(x) = \begin{cases} 0 & \text{if } x \in (a,b) \\ \\ \dfrac{x-b}{c-b} & \text{if } x \in (b,c) \end{cases}$$



Now, if we define

$$V_1 = \{v, v|_{(a,b)} \in P_1(a,b),\ v(x) = 0\ \forall x \notin (a,b)\}$$
$$V_2 = \{v, v|_{(b,c)} \in P_1(b,c),\ v(x) = 0\ \forall x \notin (b,c)\}$$

which are the extensions by zero of the space P_1 we have defined at the beginning of the section. This space can also be defined as

$$P_1^{\text{disc}} = V_1 \oplus V_2 = \{v,\ v = v_1 + v_2,\ v_i \in V_i\}$$

Now, we try to find an approximation u_h of u from this space to solve the model problem when $\theta = 0$, but, first of all, recall the exact problem we are dealing with: Determine $u \in V = H^1(0,1)$, such that $u(0) = 0$ and that

$$\int_0^1 u'\, v'\, dx = \int_0^1 f\, v\, dx$$

holds for all $v \in V$ satisfying $v(0) = 0$. In this case we are taking $a = 0,\ c = 1$. Notice that the integral above can be written as

$$\int_0^1 u'\, v'\, dx = \int_0^b u'\, v'\, dx + \int_b^1 u'\, v'\, dx$$

and similarly for the integral in the right hand side. Motivated by this, consider the following Galerkin formulation: Determine $u_h \in V_h = P_1^{\text{disc}}$, such that $u_h(0) = 0$ and that

$$\int_0^b u_h'\, v_h'\, dx + \int_b^1 u_h'\, v_h'\, dx = \int_0^b f\, v_h\, dx + \int_b^1 f\, v_h\, dx$$

holds for all $v_h \in V_h$ satisfying $v_h(0) = 0$.

The discrete solution $u_h \in V_h$ can be written as

$$u_h = \sum_{i=1}^{4} U_j \phi^j(x)$$

(i) First, we have to include the boundary condition in the definition of the space, for which we define

$$V_{h0} = \{v \in P_1^{\text{disc}}, v(0) = 0\}$$

Notice that this removes one degree of freedom, so this subspace has dimension 3 and is spanned by $\{\phi^2, \phi^3, \phi^4\}$. This is like taking $U_1 = 0$ above.

(ii) Second, we have to compute the coefficients $a_d(\phi^i, \phi^j)$ of the matrix $\mathsf{K} \in \mathbb{R}^{3\times 3}$ appearing in the linear system

$$\mathsf{K}\,\mathbf{U} = \mathbf{F}$$

Considering the basis of V_{h0} to be the set of functions $\{\psi^1, \psi^2, \psi^3\} = \{\phi^2, \phi^3, \phi^4\}$, we compute the matrix:

$$\mathsf{K} = \begin{bmatrix} a_d(\phi^2,\phi^2) & a_d(\phi^3,\phi^2) & a_d(\phi^4,\phi^2) \\ a_d(\phi^2,\phi^3) & a_d(\phi^3,\phi^3) & a_d(\phi^4,\phi^3) \\ a_d(\phi^2,\phi^4) & a_d(\phi^3,\phi^4) & a_d(\phi^4,\phi^4) \end{bmatrix} \tag{6.9}$$

and calculating the integrals we obtain

$$K_{11} = a_d(\phi^2,\phi^2) = \int_0^b (\phi^2)'(\phi^2)'\,dx + \int_b^1 (\phi^2)'(\phi^2)'\,dx = \int_0^b (\phi^2)'(\phi^2)'\,dx + 0 = \frac{1}{b},$$

$$K_{12} = K_{21} = a_d(\phi^3,\phi^2) = \int_0^b (\phi^3)'(\phi^2)'\,dx + \int_b^1 (\phi^3)'(\phi^2)'\,dx = \int_0^b 0\,(\phi^2)'\,dx + \int_b^1 (\phi^3)'\,0\,dx = 0,$$

$$K_{13} = K_{31} = a_d(\phi^4,\phi^2) = \int_0^b (\phi^4)'(\phi^2)'\,dx + \int_b^1 (\phi^4)'(\phi^2)'\,dx = \int_0^b 0\,(\phi^2)'\,dx + \int_b^1 (\phi^4)'\,0\,dx = 0,$$

and so on, giving

$$\mathsf{K} = \begin{bmatrix} \frac{1}{b} & 0 & 0 \\ 0 & \frac{1}{1-b} & -\frac{1}{1-b} \\ 0 & -\frac{1}{1-b} & \frac{1}{1-b} \end{bmatrix} \tag{6.10}$$

Notice that the term $K_{11} = a_d(\phi^2, \phi^2)$ is exactly what we had computed before when introducing the $P_1(a, b)$ space simply with $a = 0$, so, we could just have reused that result. Similarly for the second diagonal 2×2 block of (6.10),

$$\begin{bmatrix} \frac{1}{1-b} & -\frac{1}{1-b} \\ -\frac{1}{1-b} & \frac{1}{1-b} \end{bmatrix}$$

which is exactly the matrix we have computed before but in the interval (b, c) instead of (a, b) and taking $c = 1$.

Matrix K **is singular!**. Notice that the space V_h is not a subset of $H^1(0,1)$[2].

[2] This is related to Theorem 6.8.2 to be introduced at the end of this chapter.

In the previous example, since functions in the space are discontinuous, their derivatives appearing in the integrals are Dirac delta functions at $x = b$, so, the integrals are not defined, however, since we partitioned the integrals, we naively proceed with the calculations and obtained a singular matrix. Following with this naive approach, it is interesting also to perform the computation of the system matrix when $\theta = 1$ and see what happens, for which it only remains the computation of matrix M

$$M_{ij} = \int_0^1 \phi^i\, \phi^j\, dx = \int_0^b \phi^i\, \phi^j\, dx + \int_b^1 \phi^i\, \phi^j\, dx$$

Again, we can reuse the results already obtained when describing the space P_1 for a single interval. The final matrix will be the sum of the previously computed K and M.

$$\mathsf{A} = \mathsf{K} + \mathsf{M} = \begin{bmatrix} \frac{1}{b} + \frac{b}{3} & 0 & 0 \\ 0 & \frac{1}{1-b} + \frac{1-b}{3} & -\frac{1}{1-b} + \frac{1-b}{6} \\ 0 & -\frac{1}{1-b} + \frac{1-b}{6} & \frac{1}{1-b} + \frac{1-b}{3} \end{bmatrix} \tag{6.11}$$

In this case, the matrix is not singular. For instance, if we take the function in the right hand side of the variational formulation to be the constant function $f = 1$ and we calculate the coefficients $\ell(\phi^i)$ of vector $\underline{\mathsf{F}} \in \mathbb{R}^3$ we get

$$F_1 = \ell(\phi^2) = \int_0^1 \phi^2\, dx = \frac{b}{2}$$

$$F_2 = \ell(\phi^3) = \int_0^1 \phi^3\, dx = \frac{1-b}{2}$$

$$F_3 = \ell(\phi^4) = \int_0^1 \phi^4\, dx = \frac{1-b}{2}$$

Taking now e.g. $b = 0.5$ and solving, we finally obtain $\mathbf{U} = \begin{bmatrix}0.1154 & 1 & 1\end{bmatrix}^T \Rightarrow u_h = 0.1154\,\phi^2(x) + \phi^3(x) + \phi^4(x)$, which is plotted Figure 6.4 and compared with the exact solution for this problem. This solution does not make sense, as a consequence of the incorrect choice for V_h. In the next section we remedy this by defining a space of continuous functions.

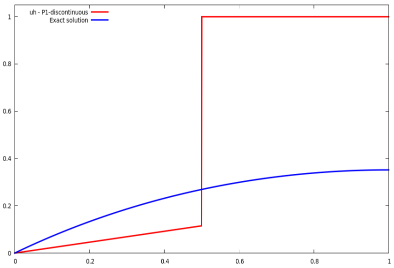


Figure 6.4: u_h vs exact solution when using a P_1 discontinous space.

6.1.3 A P_1 continuous space

Consider the intervals (a, b) and (b, c). In the previous example "glue" the degree of freedom at $x = b$, of the interval to the left and to the right of this point, by imposing the restriction $v(b^-) = v(b^+)$. In this case we only have three degrees of freedom:

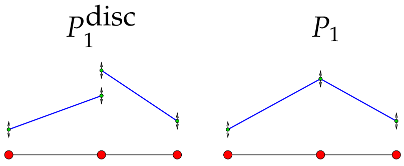


Figure 6.5: P_1 continuous vs discontinuous space.

$$\begin{aligned}\sigma_1(v) &= v(a)\\ \sigma_2(v) &= v(b)\\ \sigma_3(v) &= v(c)\end{aligned}$$

so, the space has dimension 3. This choice automatically leads to a space of continuous functions in (a,c) which we describe as

$$V_h = \{v,\ v|_{(a,b)} \in P_1(a,b),\ \ v|_{(b,c)} \in P_1(b,c)\}\ \cap\ C^0(a,c)$$

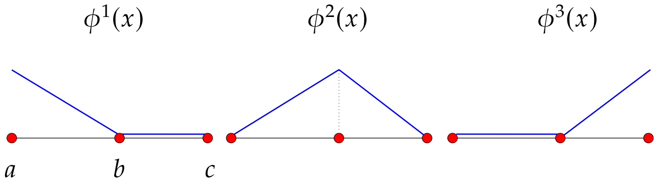
Again, considering

$$v(x) = \begin{cases}\alpha + \beta\, x & \text{if } x \in (a,b)\\ \gamma + \epsilon\, x & \text{if } x \in (b,c)\end{cases}$$

By inspection we find that the basis is:

$$\phi^1(x) = \begin{cases}\dfrac{b-x}{b-a} & \text{if } x \in (a,b)\\[2ex] 0 & \text{if } x \in (b,c)\end{cases}, \ \phi^2(x) = \begin{cases}\dfrac{x-a}{b-a} & \text{if } x \in (a,b)\\[2ex] \dfrac{c-x}{c-b} & \text{if } x \in (b,c)\end{cases}, \ \phi^3(x) = \begin{cases}0 & \text{if } x \in (a,b)\\[2ex] \dfrac{x-b}{c-b} & \text{if } x \in (b,c)\end{cases}$$

which clearly satisfies that $\sigma_i(\phi^j) = \delta_{ij}$. By computing matrix A for the model problem in this case and comparing with the one obtained using the P_1^{disc} space we can see how good the approximation is as illustrated below



Now, we generalize this to partitions of the interval with an increasing number of subintervals:

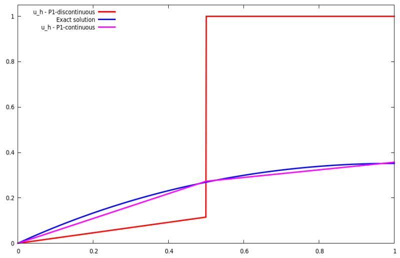


Figure 6.6: u_h vs exact solution when using a P_1 discontinous and continuous spaces.

6.2 1D finite element meshes

Let consider a partition $\mathscr{T}_h$ of $\Omega = [0,1]$, i.e., an indexed collection of intervals

$$\bar{\Omega} = \bigcup_{j=1}^{N} I_j$$

where $I_j = [x^j, x^{j+1}]$ and the $N_v = N+1$ **nodes** (arbitrarily numbered) are $0 = x^1 < x^2 < \cdots < x^{N+1} = 1$. Define $h_i = x^{i+1} - x^i$ and

$$h = \max_j h_j$$

which is a measure of how fine the partition is.

6.2.1 A $P_1^{\text{disc}}(\mathcal{T}_h)$ (totally discontinuous) space in 1D

With the partition of Ω just defined, we begin by defining the spaces:

$$V_i = \{v, v|_{I_i} \in P_1(I_i),\ v(x) = 0\ \forall x \notin I_i\}$$

where $P_1(I_i) = P_1(x^i, x^{i+1})$ is the space P_1 for an individual interval the we introduced before.

We define a totally discontinuous space associated to the partition $\mathcal{T}_h$ as the direct sum of these V_i's:

$$X_h(\mathcal{T}_h) = V_1 \oplus V_2 + \cdots \oplus V_N = \{v,\ v = v_1 + v_2 + \cdots + v_N,\ v_i \in V_i\}$$

This space has dimension equal to $N \times 2$, but it is not in $H^1(0,1)$.

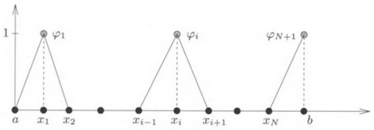


Fig. 1.1. One-dimensional hat functions.

Figure 6.7: Hat basis functions in 1D.

6.2.2 $P_1(\mathcal{T}_h)$ conforming space in 1D

Now, if we "glue" the local degrees of freedom of the individual intervals at the corresponding common nodes of $\mathcal{T}_h$, which is equivalent to choosing as degrees of freedom **the values of the function at these nodes**, we naturally define a space of continuous functions

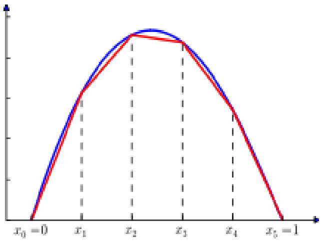


Figure 6.8: Continuous, piecewise linear function.

$$V_h = P_1(\mathcal{T}_h) = X_h(\mathcal{T}_h)\ \cap\ C^0(0,1)$$

and the basis functions will be

$$\phi^i(x) = \begin{cases} \dfrac{x - x^{i-1}}{h_{i-1}} & \text{if } x \in I_{i-1} \\[2ex] \dfrac{x^{i+1} - x}{h_i} & \text{if } x \in I_i \\[2ex] 0 & \text{otherwise} \end{cases}$$

- The dimension of V_h is equal to N_v;
- Since the degress of freedom are the values of the function at the nodes of $\mathcal{T}_h$ and the ϕ^i's are linearly independent, any function in V_h is uniquely determined precisely by these values, i.e.

$$w = \sum_{i=1}^{N+1} W^i\, \phi^i(x) = \sum_{i=1}^{N+1} w(x^i)\, \phi^i(x)$$

- These functions are linear on each interval (or element) and continuous, but their derivatives are not defined in the classical sense at all points. However, as theorem Theorem 6.8.2 states, $V_h \subset H^1(0,1)$.

We can use this space to solve example and study $\|u - u_h\|_{H^1(0,1)}$ as we refine the partition[3]

[3] It is intuitive that the Galerkin approximation u_h from this space will converge to the solution u as $h \to 0$, since any continuous function can be approximated by polygonals with an increasing number of nodes.

6.3 2D examples

6.3.1 P_1 element for a triangle

Consider a triangle K in $\mathbb{R}^2$ with vertices $(\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3)$. We want to find a basis for

$$V_h = P_1(K) = \{v : K \longrightarrow \mathbb{R},\ v = \alpha + \beta x + \gamma\, y\} \tag{6.12}$$

The space has dimension 3.

Again we start by defining the degrees of freedom. As done in previous examples we use the value of the function at a set of points, the vertices in this case

$$\sigma_i(v) = v(\mathbf{x}^i) \tag{6.13}$$

and the basis for $P_1(K)$ is defined by the relation $\sigma_i(\phi^j) = \delta_{ij}$. The coefficients of the basis functions are determined by solving the 3×3 systems:

$$\begin{aligned}
&\alpha_1 + \beta_1\, x^1 + \gamma_1\, y^1 = 1, \quad &&\alpha_2 + \beta_2\, x^1 + \gamma_2\, y^1 = 0, \quad &&\alpha_3 + \beta_3\, x^1 + \gamma_3\, y^1 = 0\\
&\alpha_1 + \beta_1\, x^2 + \gamma_1\, y^2 = 0, \quad &&\alpha_2 + \beta_2\, x^2 + \gamma_2\, y^2 = 1, \quad &&\alpha_3 + \beta_3\, x^2 + \gamma_3\, y^2 = 0\\
&\alpha_1 + \beta_1\, x^3 + \gamma_1\, y^3 = 0, \quad &&\alpha_2 + \beta_2\, x^3 + \gamma_2\, y^3 = 0, \quad &&\alpha_3 + \beta_3\, x^3 + \gamma_3\, y^3 = 1
\end{aligned}$$

Notice that for simplicity of notation above we have used (x^i, y^i) instead of (x^i_1, x^i_2).

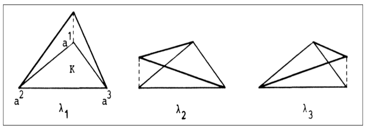


Figure 6.9: P_1 basis functions on a single triangle.

6.3.2 2D finite element meshes

Let consider a domain $\Omega \subset \mathbb{R}^2$ and for simplicity assume its boundary $\partial\Omega$ is a polygonal curve. Now, consider a partition $\mathcal{T}_h = \{K_i\}_{i=1}^N$ of Ω, such that

$$\bar{\Omega} = \bigcup_{i=1}^{N} \bar{K}_i$$

where $K_i \cap K_j = \emptyset$ if $i \neq j$. $\mathcal{T}_h$ is called a triangulation of Ω.

Given a triangulation like the one shown in Figure 6.10, what types of V_h's can be constructed?

- We can certainly construct spaces of **discontinuous functions** (see Figure 6.11). The partition having N_e triangular elements and taking P_1-triangles, for instance, we have 3 (**local**) degrees of freedom per single triangle. Then a space of totally discontinuous functions associated to $\mathcal{T}_h$ will be the direct sum of the (local) P_1 spaces $V_K = \{v : K \longrightarrow \mathbb{R}, v|_K \in P_1(K),\ v(\mathbf{x}) = 0\ \forall \mathbf{x} \notin K\},\ \ K = 1, \ldots, N_e$, i.e.,

$$X_h(\mathcal{T}_h) = V_1 \oplus V_2 + \cdots \oplus V_N = \{v,\ v = v_1 + v_2 + \cdots + v_N,\ v_i \in V_i\}$$

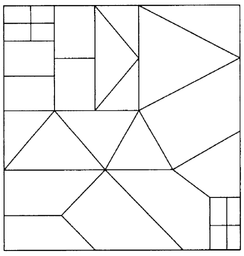


Figure 6.10: A general partition of Ω.

and its dimension will be $N_e \times 3$, however, recall that this space is not in $H^1(\Omega)$.

- We can also construct spaces of **continuous functions**. However, it turns out that this task may be not trivial (or even possible) in general meshes (as those shown in Figure 6.10). However, for the so called conforming meshes this is perfectly possible:

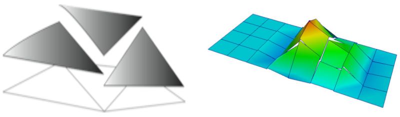


Figure 6.11: Examples of discontinuous functions.

Definition 6.3.1 *A partition $\mathcal{T}_h$ of a domain Ω is **conforming** if $\bar{K}_i \cap \bar{K}_j$ is either*

- *empty, or,*
- *a vertex, or*
- *a complete edge.*
- *a complete face (in 3D)*

*otherwise the partition is said to be **nonconforming**.*

6.3.3 $P_1(\mathcal{T}_h)$ conforming space in 2D

We proceed similarly to the 1D case. Given a **conforming triangulation** $\mathcal{T}_h$ of a polygonal domain we can build a space of continuous functions. Start with the space

$$X(\mathcal{T}_h) = \{v,\ v|_{K_i} \in P_1(K_i)\ \forall K_i \in \mathcal{T}_h\} \tag{6.14}$$

where $v|_K$ denotes the restriction of v to K and $P_1(K)$ is the space of polynomial functions of degree ≤ 1 on triangle K that we have already defined in subsection 6.3.1

We define as degrees of freedom the value of the function at the nodes of the triangulation.

Since we are assuming now that $\mathcal{T}_h$ is conforming, each vertex of any triangle can only be a vertex of other triangles an cannot be on an edge. Thus, we can "glue" the (local) degrees of freedom of the individual triangles. This naturally leads to the following description of the space we have constructed

$$V_h = X(\mathcal{T}_h)\ \cap\ C^0(\bar{\Omega}) = \{v \in C^0(\bar{\Omega}),\ v|_K \in P_1(K)\ \forall K \in \mathcal{T}_h\}$$

We can construct a basis for this space. Let us assume $\mathcal{T}_h$ has N_v vertices whose coordinates are $\{\mathbf{x}^i\}_{i=1}^{N_v}$. Let $\phi^i,\ i = 1, \ldots, N_v$ be the functions that satisfies

$$\phi^i(\mathbf{x}^j) = \delta_{ij} \tag{6.15}$$

whose restriction to element K having j as one of its vertices is the corresponding function in $P_1(K)$ and 0 otherwise. Any function $v = \sum_{i=1}^{N_v} v(\mathbf{x}^i)\, \phi^i(x) \in V_h$ is uniquely determined by the degrees of freedom

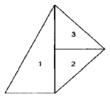


Figure 4.1. *Two examples of nonconforming triangulations. In both examples, the intersection of triangles 1 and 2 is a line segment that is not an edge of triangle 1.*

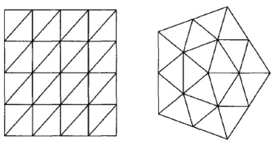


Figure 4.2. *Triangulations of two polygonal domains.*

Figure 6.12: Different conforming and nonconforming partitions of Ω.

that are precisely the values of the function at the N_v nodes of $\mathcal{T}_h$. Notice that

- $\{\phi^j\}_{j=1}^{N_v}$ are linearly independent;
- $V_h = \text{span}\{\phi^j\} = \{v_h, v_h = \sum_{j=1}^{N_v} a^j\, \phi^j\}$;
- $\dim(V_h) = N_v$.

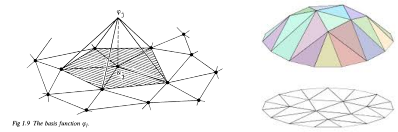


Figure 6.13: Hat function and an elementwise P_1 function on a triangular mesh.

6.4 More examples of finite elements and their associated global spaces

6.4.1 P_2 triangular element

Consider a triangle K in $\mathbb{R}^2$ with vertices $(\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3)$. We want to find a basis for

$$V_h = P_2(K) = \{v : K \to \mathbb{R},\ v = \alpha_0 + \alpha_1 x + \alpha_2\, y + \alpha_3\, x^2 + \alpha_4 y^2 + \alpha_5\, x\, y\}$$

The space has dimension 6 since an element of $P_2(K)$ is determined by six independent parameters. Any function is uniquely determined by its values at:

- the vertices of the triangle;
- the midpoints of the three edges.

Take two points $\mathbf{x}^i$ and $\mathbf{x}^j$. If a function v belongs to $P_2(K)$ then

$$v\left((1-s)\,\mathbf{x}^i + s\,\mathbf{x}^j\right) \in P_2(s) = \{w,\ w = \beta_0 + \beta_1\, s + \beta_2\, s^2\}$$

where $0 \le s \le 1$, this is, the function restricted to the straight segment joining $\mathbf{x}^i$ and $\mathbf{x}^j$ of the triangle, is a parabolic function, which is uniquely determined by its values at the three points. When considering the master triangle $\hat{K}$ used above we have:

$$\hat{\psi}^1 = (1-\hat{x}-\hat{y})\,(1-2\hat{x}-2\hat{y}),\quad \hat{\psi}^2 = \hat{x}\,(2\hat{x}-1),\quad \hat{\psi}^3 = \hat{y}\,(2\hat{y}-1)$$

$$\hat{\psi}^4 = 4\,\hat{x}\,\hat{y},\quad \hat{\psi}^5 = 4\,\hat{y}\,(1-\hat{x}-\hat{y}),\quad \hat{\psi}^6 = 4\,\hat{x}\,(1-\hat{x}-\hat{y}) \tag{6.16}$$

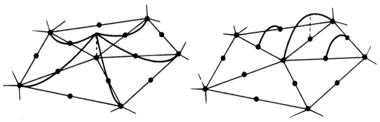


Figure 6.14: P_2 functions on a cluster of triangles.

These functions clearly satisfy $\phi^i(\vec{p}^j) = \delta_{ij}$, where $\vec{p}^j$ corresponds to the vertices for $j = 1,2,3$ and to the midpoints of sides for $j = 4,5,6$. Given a conforming triangulation $\mathcal{T}_h$, we want to construct a space of continuous functions as we did before, i.e., "gluing" together the degrees of freedom of all the $P_2(K)$-triangles in $\mathcal{T}_h$, that share a vertex or a midpoint. In order to do so, simply fix the value of the function at all vertices and at all midpoints (on edges shared by two triangles). The resulting function will be **continuous** and belong to the space:

$$V_h = P_2(\mathcal{T}_h) = \{v \in C^0(\bar{\Omega}),\ v|_K \in P_2(K)\ \forall K \in \mathcal{T}_h\} \tag{6.17}$$

6.4.2 Triangular elements of arbitrary degree

We can generalize the finite element spaces we have considered to construct space of continuous piecewise polynomial functions of arbitrary degree k on a triangle. The placement of the nodes on the triangle is determined by:

(i) On each edge there must by $k+1$ nodes since a one dimensional polynomial of degree k has $k+1$ degrees of freedom. Each edge has two vertices and the other $k-1$ nodes will be regularly spaces between them;

(ii) A polynomial of degree k in two variables is determined by

$$1+2+3+\cdots+(k+1)=\frac{(k+1)(k+2)}{2}$$

parameters. Therefore, the number of interior nodes will be $\frac{(k+1)(k+2)}{2}-3k$.

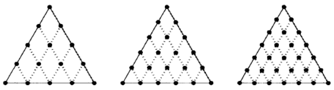


Figure 6.15: Lagrange elements of order 4, 5 and 6. The points indicate the location of the dofs.

When using high order Lagrange elements, the stiffness matrix K ($K_{ij}=a_d(\phi^i,\phi^j)$) may become ill-conditioned as the mesh is refined. This is a consequence of the chosen basis, and can be circunvented by choosing other basis functions.

Given a triangulation of a domain Ω it is interesting to know the relation between the number of vertices N_v, the number of edges N_{edges}, the number of elements N_e. This is given by the Euler relations:

Lemma 6.4.1 *(* ***Euler relations****) Let $\mathcal{T}_h$ be a conforming partition of a polygonal domain $\Omega\subset\mathbb{R}^2$, then*

$$\begin{aligned} N_e-N_{edges}+N_v &= 1-I\\ N_v^{\partial}-N_{edges}^{\partial} &= 0\end{aligned}$$

where I is the number of holes in Ω and the supraindex ∂ denotes de boundary. In particular, if the elements are polygons with v vertices

$$2N_{edges}+N_{edges}^{\partial}=v\,N_e$$

Figure 6.16: Leonhard Euler (1707(Switzerland)-1783(Russia)).

6.5 General definition of a Finite element

The formal definition of a finite element is given next. This was introduced in the book of [3].

[3]: Ciarlet (1991), *Finite Element Methods (Part 1)*

Definition 6.5.1 *(****Ciarlet****) A finite element in $\mathbb{R}^d$ (typically d = 1, 2 or 3) is a triplet (K,P_K,Σ_K) where*

(i) K is a closed bounded subset of $\mathbb{R}^d$ with a non-empty interior and Lipschitz boundary;

(ii) P_K *is a finite dimensional space of functions defined over K of dimension* n*;*

(iii) Σ_K *is a set of* n *linear functionals* $\{\sigma_i\}_{i=1,\dots,n}$ *such that for any real scalars* $\alpha_i,\ i = 1,\dots,n$ *there exist a unique function* $p \in P_K$ *that satisfies (**unisolvence**)*

$$\sigma_i(p) = \alpha_i \tag{6.18}$$

Proposition 6.5.1 *There exists a basis* $\{\psi^1, \psi^2, \dots, \psi^n\}$ *in* P_K *such that*

$$\sigma_i(\psi^j) = \delta_{ij}, \quad 1 \le i, j \le n \tag{6.19}$$

- The linear forms $\{\sigma_1, \sigma_2, \dots, \sigma_n\}$ are called the degrees of freedom;
- The functions $\{\psi^1, \psi^2, \dots, \psi^n\}$ are called the basis functions;
- Let $\{K, P_K, \Sigma_K\}$ be a finite element. If there is a set of points $\{\vec{a}^1, \dots, \vec{a}^n\}$ in K such that for all $p \in P_K, \sigma_i(p) = p(\vec{a}^i)\ i = 1, \dots, n$, $\{K, P_K, \Sigma_K\}$ is called a **Lagrange finite element**.

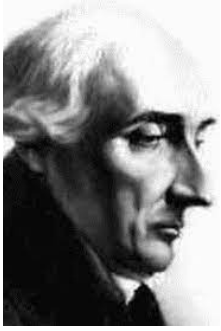


Figure 6.17: Joseph-Louis Lagrange (France, 1736-1813).

6.6 Affine family of finite elements

A central concept in the FEM is that of affine family of finite elements, that establishes the relation between reference and real (**physical**) elements[4](see Figure 6.18). Its formal definition is the following:

Definition 6.6.1 *(**Affine elements**) A family of finite elements is called an affine family if all its elements are affine equivalent to a single reference or **master** element.*

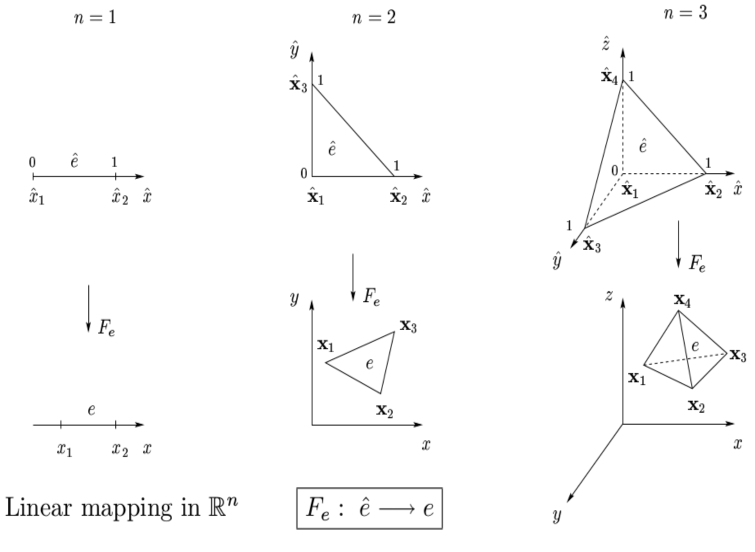


Figure 6.18: Affine mapping in 1, 2 and 3 dimensions

What is affine equivalence?

[4] This is important because the coefficients of linear systems are computed on the master element. Also, the interpolation theory, that is the basis of most convergence theorems is easier to develop.

An affine transformation $F_K : \hat{K} \longrightarrow K$ of the reference element $\hat{K}$ with vertices $\hat{\mathbf{x}}^i$ onto an element K with vertices $\mathbf{x}^i$ is defined by:

$$F_K(\hat{\mathbf{x}}) = B_K \cdot \hat{\mathbf{x}} + \mathbf{b}_K, \quad B_K \in \mathbb{R}^{d\times d}, \ \mathbf{b} \in \mathbb{R}^d \tag{6.20}$$

We have

$$d = 1, \qquad B_K = x^2 - x^1, \qquad b_K = x^1$$

$$d = 2, \qquad B_K = \begin{bmatrix} x^2 - x^1 & x^3 - x^1 \\ y^2 - y^1 & y^3 - y^1 \end{bmatrix}, \qquad b_K = \begin{bmatrix} x^1 \\ y^1 \end{bmatrix}$$

Note that in this case we have:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x^1 \\ y^1 \end{bmatrix} (1 - \hat{x} - \hat{y}) + \begin{bmatrix} x^2 \\ y^2 \end{bmatrix} \hat{x} + \begin{bmatrix} x^3 \\ y^3 \end{bmatrix} \hat{y} = \sum_{j=1}^{3} \mathbf{x}^j \, \hat{\psi}^j(\hat{\mathbf{x}})$$

Properties of the affine mapping

(a) Vertices are mapped onto vertices:

$$\mathbf{x}^i = F_K(\hat{\mathbf{x}}^i)$$

(b) Midpoints of sides are mapped onto midpoints of sides:

$$\mathbf{x}^{ij} = \frac{\mathbf{x}^i + \mathbf{x}^j}{2} = F_K\left(\frac{\hat{\mathbf{x}}^i + \hat{\mathbf{x}}^j}{2}\right) = F_K(\hat{\mathbf{x}}^{ij})$$

(c) Barycenters are mapped onto barycenters

$$\mathbf{x}^{ijk} = \frac{\mathbf{x}^i + \mathbf{x}^j + \mathbf{x}^k}{3} = F_K\left(\frac{\hat{\mathbf{x}}^i + \hat{\mathbf{x}}^j + \hat{\mathbf{x}}^k}{3}\right) = F_K(\hat{\mathbf{x}}^{ijk})$$

(d) For a function ψ defined on K, we define $\hat{\psi}$ on $\hat{K}$ by

$$\hat{\psi}(\hat{\mathbf{x}}) = \psi\left(F_K(\hat{\mathbf{x}})\right) = \psi(\mathbf{x})$$

Therefore, if function ψ is a polynomial of degree k on K, $\hat{\psi}$ is also a polynomial of degree k on $\hat{K}$.

(e) The derivatives of ψ and $\hat{\psi}$ are related by

$$\nabla\psi(\mathbf{x}) = B_K^{-T} \cdot \hat{\nabla}\hat{\psi}(\hat{\mathbf{x}}) \tag{6.21}$$

(f) $|\det B_K| = \dfrac{\text{meas}(K)}{\text{meas}(\hat{K})}$

It is instructive to prove these properties, in particular (e). First, note that:

$$\hat{\mathbf{x}} = B_K^{-1} \cdot \mathbf{x} + \tilde{\mathbf{b}}_K$$

where $\tilde{\mathbf{b}}_K = -B_K^{-1} \cdot \mathbf{b}_K$. Using index notation

$$\hat{x}_k = \left[B_K^{-1}\right]_{k\ell} x_\ell + \left[\tilde{\mathbf{b}}_K\right]_k \quad \Rightarrow \quad \frac{\partial \hat{x}_k}{\partial x_\ell} = \left[B_K^{-1}\right]_{k\ell}$$

Now, applying the chain rule

$$\frac{\partial\psi}{\partial x_k}(\mathbf{x}) = \frac{\partial\hat{\psi}}{\partial\hat{x}_\ell}\left(F_K^{-1}(\mathbf{x})\right)\frac{\partial\hat{x}_\ell}{\partial x_k} = \frac{\partial\hat{\psi}}{\partial\hat{x}_\ell}\left(F_K^{-1}(\mathbf{x})\right)\left[B_K^{-1}\right]_{\ell k} = \left[B_K^{-T}\right]_{k\ell}\frac{\partial\hat{\psi}}{\partial\hat{x}_\ell}\left(F_K^{-1}(\mathbf{x})\right)$$

or

$$\nabla\psi(\mathbf{x}) = B_K^{-T}\cdot\hat{\nabla}\hat{\psi}\left(F_K^{-1}(\mathbf{x})\right)$$

The case of P_1 linear elements is particularly simple

$$\hat{\psi}^1 = 1 - \hat{x} - \hat{y}, \quad \hat{\psi}^2 = \hat{x}, \quad \hat{\psi}^3 = \hat{y}$$

whose gradients are

$$\hat{\nabla}\hat{\psi}^1 = \begin{bmatrix}-1\\-1\end{bmatrix}, \quad \hat{\nabla}\hat{\psi}^2 = \begin{bmatrix}1\\0\end{bmatrix}, \quad \hat{\nabla}\hat{\psi}^3 = \begin{bmatrix}0\\1\end{bmatrix}$$

therefore, to compute $\nabla\psi^j,\ j = 1, 2, 3$, we transpose the inverse of B_K and multlying by those constant vectors.

Proposition 6.6.1 *If K and $\hat{K}$ are affine equivalent and if the triplet $(\hat{K}, \hat{P}, \hat{\Sigma}_K)$ is a finite element, then we can define (K, P_K, Σ_K) and it is a finite element.*

To see this, let $F_K : \hat{K} \longrightarrow K$ be the affine mapping. We have to show how to contruct P_K and Σ based on $\hat{P}$ and $\hat{\Sigma}$. We define for any $\hat{v} \in \hat{P}$ the function $v \in P_K$ by $v(\mathbf{x}) = \hat{v}\left(F_K^{-1}(\mathbf{x})\right)$

$$P_K = \{v : K \longrightarrow \mathbb{R},\ \hat{v} \in \hat{P}\} \tag{6.22}$$

and

$$\Sigma_K = \{\sigma : P_K \longrightarrow \mathbb{R}, \sigma(v) = \hat{\sigma}(\hat{v})\ \forall\ \hat{v} \in \hat{P} \text{ and } \hat{\sigma} \in \hat{\Sigma}\} \tag{6.23}$$

6.7 Hermitian finite elements ☺

Figure 6.19: Charles Hermite (France, 1822-1901).

When constructing the degrees of freedom as shown, the only ones that are preserved when passing from the master to the real element are those involving values of the function at a set of points, i.e., the **Lagrangian** dofs. The case of the so called **Hermitian** elements, involving derivatives of the function at a set of points as dofs, have to be considered differently.

We restrict ourselves to the **1D case** for simplicity. Let $\hat{P} = P_3(\hat{K})$, being the master element $\hat{K} = (0, 1)$ and consider the degrees of freedom

$$\hat{\sigma}_1(\hat{v}) = \hat{v}(0), \quad \hat{\sigma}_2(\hat{v}) = \hat{v}'(0)$$

$$\hat{\sigma}_3(\hat{v}) = \hat{v}(1), \quad \hat{\sigma}_4(\hat{v}) = \hat{v}'(1)$$

The local basis functions are thus given by

$$\hat{\psi}^1_{\hat{K}} = (2\,\hat{x} + 1)\,(\hat{x} - 1)^2, \quad \hat{\psi}^2_{\hat{K}} = \hat{x}\,(\hat{x} - 1)^2$$

$$\hat{\psi}^3_{\hat{K}} = (3 - 2\,\hat{x})\,\hat{x}^2, \quad \hat{\psi}^4_{\hat{K}} = (\hat{x} - 1)\,\hat{x}^2$$

This is called the Hermite element. You can check that $\hat{\sigma}_i(\hat{\psi}^j_K) = \delta_{ij}$

Remember how Σ_K is defined: we consider a basis of linear functionals $\hat{\sigma}_i \in \hat{\Sigma},\ i = 1, \dots, n$ and define $\sigma_i \in \Sigma_K$ as:

$$\sigma_i(v) = \hat{\sigma}_i(\hat{v}) \ \forall\, \hat{v} \in \hat{P}$$

We add a new ingredient at this point: scaling. This is, we define a set of coefficients $\alpha_i,\ i = 1, \dots, n$ and the degrees of freedom $\sigma_i \in P_k$ such that

$$\sigma_i(v) = \alpha_i\, \hat{\sigma}_i(\hat{v}) \ \forall\, \hat{v} \in \hat{P}$$

In this way, it can be shown that (K, P_K, Σ_K) will also be a finite element. Consider now a partition $\mathcal{T}_h$ of Ω, i.e., a 1D finite element mesh made of nonoverlapping intervals and choose as coefficients α_i on each element K:

$$\alpha_1 = \alpha_3 = 1,\ \ \alpha_2 = \alpha_4 = \frac{1}{h_K}$$

where h_K is the size of element K. The local basis functions defined on K are, for any $x \in K,\ x = F_K(\hat{x})$

$$\psi^1_K(x) = \hat{\psi}^1_{\hat{K}}(\hat{x}),\quad \psi^2_K(x) = h_K\, \hat{\psi}^2_{\hat{K}}(\hat{x})$$

$$\psi^3_K(x) = \hat{\psi}^3_{\hat{K}}(\hat{x}),\quad \psi^4_K(x) = h_K\, \hat{\psi}^4_{\hat{K}}(\hat{x})$$

so we define a space of functions associated to this partition which has $C^1(\Omega)$ continuity. Globally, the space is described by

$$V_h = \{v,\ v|_K \in P_3(K),\ \ \forall K \in \mathcal{T}_h\} \ \cap\ C^1(\Omega)$$

At node i, shared by elements K_l (left) and K_r (right), we define two basis functions

$$\phi^{i,0}(x) = \begin{cases} \psi^3_{K_l} & \text{if } x \in K_l \\ \psi^1_{K_r} & \text{if } x \in K_r \\ 0 & \text{otherwise} \end{cases} \quad , \quad \phi^{i,1}(x) = \begin{cases} \psi^4_{K_l} & \text{if } x \in K_l \\ \psi^2_{K_r} & \text{if } x \in K_r \\ 0 & \text{otherwise} \end{cases}$$

The set $\{\phi^{0,0}, \phi^{0,1}, \phi^{1,0}, \phi^{1,1}, \dots, \phi^{N_v,0}, \phi^{N_v,1}\}$, where N_v is the number of nodes, is a basis for V_h.

6.8 Conclusion

We finally state the following theorem which is extremely important to the FEM and justify the choice of piecewise polynomial spaces to approximate the weak solution of PDEs:

Theorem 6.8.1 *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain which can be partitioned into N_e Lipschitz subdomains K_j (i.e., $\overline{\Omega} = \bigcup_{j=1}^{N_e} \bar{K}_j$, $K_i \cap K_j =$*

$\emptyset\ \forall i \neq j$. Then, for every $k \geq 1$ and $1 \leq p \leq \infty$, the set

$$\left\{v \in C^{k-1}(\overline{\Omega}),\ v|_{K_j} \in C^k(\overline{K_j}),\ j = 1, \ldots, N_e\right\} \subset W^{k,p}(\Omega)$$

of interests to us is the next corollary

Corollary 6.8.2 *Let v be a* ***piecewise-polynomial*** *function on a partition of a domain Ω as in Theorem 6.8.1, then*

$$v \in H^1(\Omega) \iff v \in C^0(\overline{\Omega})$$

also

$$v \in H^2(\Omega) \iff v \in C^1(\overline{\Omega})$$

Just to fix ideas, consider the case of linear polynomials and 2D partitions made of triangles[5]. We aim to show that $V_h = P_1(\mathcal{T}_h) \subset H^1(\Omega)$, i.e., a continuous piecewise linear function belongs to $H^1(\Omega)$. Start by considering $v \in P_1(\mathcal{T}_h)$, whose restriction to element K is $v|_K = a_K + b_K\, x + c_K\, y$. Such functions belong to $L^2(\Omega)$, since

$$\int_\Omega v^2 = \sum_{K \in \mathcal{T}_h} \int_K v|_K^2 = \sum_{K \in \mathcal{T}_h} \int_K (a_K + b_K\, x + c_K\, y)^2 < \infty$$

For such v, denoting by int(K), the interior of triangle K, define

$$w_x = \begin{cases} b_K & \text{if } (x, y) \in \text{int}(K) \\ 0 & \text{otherwise} \end{cases} \tag{6.24}$$

and similar for w_y. As a global function defined on Ω, in general the derivatives are not continuous across edges shared by two triangles, so the classical derivatives of v are undefined at ∂K. We have to show that (Equation 6.24) defines the weak partial derivatives of v. Consider the derivative with respect to x (for y is analogous). For any $\phi \in \mathcal{D}(\Omega)$

$$\int_\Omega w_x\, \phi = \sum_{K \in \mathcal{T}_h} \int_K b_K\, \phi = \sum_{K \in \mathcal{T}_h} \int_K \frac{\partial (v|_K)}{\partial x}\, \phi = \sum_{K \in \mathcal{T}_h} \left[\int_{\partial K} v|_K\, (\phi \check{\mathbf{e}}_x) \cdot \check{\mathbf{n}}^K - \int_K v|_K\, \frac{\partial \phi}{\partial x}\right] \tag{6.25}$$

where $\check{\mathbf{n}}^K$ is the outward unit normal to ∂K. The integral over ∂K must be carried over all edges of triangle K. If edge e belongs to $\partial\Omega$

$$\int_e v\, (\phi \check{\mathbf{e}}_x) \cdot \check{\mathbf{n}}^K = \int_e v\, \phi \check{\mathbf{n}}_x^K = 0 \tag{6.26}$$

since ϕ has compact support in Ω. Otherwise, for any internal edge e shared by elements K and K', since v is continuous (i.e., $v|_K = v|_{K'}\ \forall \mathbf{x} \in e$) and $\check{\mathbf{n}}^K = -\check{\mathbf{n}}^{K'}$ on e, the sum over boundary terms vanishes. Then,

$$\int_\Omega w_x\, \phi = -\sum_{K \in \mathcal{T}_h} \int_K v|_K\, \frac{\partial \phi}{\partial x} = -\int_\Omega v\, \frac{\partial \phi}{\partial x} \quad \forall \phi \in \mathcal{D}(\Omega) \tag{6.27}$$

which is the definition of weak derivative. Since w_x belongs to $L^2(\Omega)$, we thus conclude that $v \in H^1(\Omega)$.

To conclude this chapter:

[5] The generalization to other polynomial degrees or functions is quite straightforward.

Solve the following exercices

- Consider the space $P_2(a,b) = \{v, v = \alpha_0 + \alpha_1\, x + \alpha_2\, x^2\}$. Compute the basis $\{\phi^1, \phi^2, \phi^3\}$ when we choose as degrees of freedom

$$\begin{aligned}\sigma_1(v) &= v(a)\\ \sigma_2(v) &= v\left(\frac{a+b}{2}\right)\\ \sigma_3(v) &= v(b)\end{aligned}$$

- Write the linear functions for the master triangle $\hat{K}$ having as vertices $((0,0),(1,0),(0,1))$.

- Considering the bilinear functions $v(x,y) = \alpha_0 + \alpha_1\, x + \alpha_2\, y + \alpha_3\, x\, y$ and and the master quadrangle $\hat{K} = [-1,1] \times [-1,1]$. Write the basis functions considering as dofs the function values at the vertices (Lagrange).

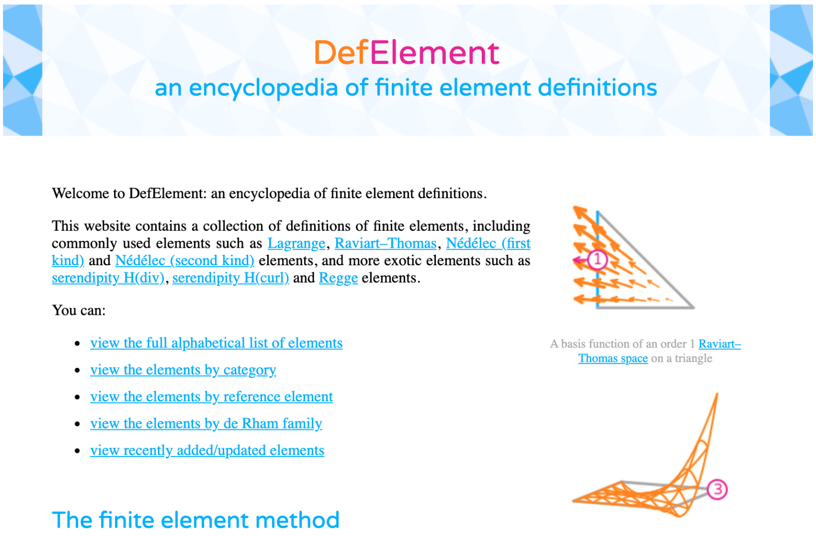
- Noticing that the support of function ϕ^j in Figure 6.13 are all the elements sharing node j, what are the consequences for the matrix A with $A_{ij} = a(\phi^i, \phi^j)$, when choosing such space to compute an approximation to u?

- Show that Property *(iii)* in the definition of a finite element is equivalent to:

$$\sigma_i(p) = 0 \Leftrightarrow p = 0, \quad i = 1, \ldots, n \tag{6.28}$$

- Which is the dimension of the space V_h given by Equation 6.17 associated to a mesh $\mathcal{T}_h$ made of triagles?

- Take a look at the finite element library that describes all (or at least most) of the known Finite Elements: `DefElement`



DefElement

an encyclopedia of finite element definitions

Welcome to DefElement: an encyclopedia of finite element definitions.

This website contains a collection of definitions of finite elements, including commonly used elements such as Lagrange, Raviart–Thomas, Nédélec (first kind) and Nédélec (second kind) elements, and more exotic elements such as serendipity H(div), serendipity H(curl) and Regge elements.

You can:

- view the full alphabetical list of elements
- view the elements by category
- view the elements by reference element
- view the elements by de Rham family
- view recently added/updated elements

A basis function of an order 1 Raviart–Thomas space on a triangle

The finite element method

6.9 Assignment 5: Poisson's problem - Convergence

Consider the unit square domain $\Omega = [0,1]^2$ and the usual Poisson's problem: Find u such that

$$\begin{cases} -\nabla \cdot (\mu(\mathbf{x})\, \nabla u(\mathbf{x})) &= f(\mathbf{x}) \quad \mathbf{x} \in \Omega \\ u(\mathbf{x}) &= u_{\mathcal{B}}(\mathbf{x}) \quad \mathbf{x} \in \partial\Omega \end{cases}$$

where f is the source term corresponding to the following manufactured solution:

$$\begin{aligned} u(\mathbf{x}) &= \sin(2\pi x_1)\, \sin(2\pi x_2) & \mathbf{x} \in \Omega \\ u_{\mathcal{B}}(\mathbf{x}) &= \sin(2\pi x_1)\, \sin(2\pi x_2) & \mathbf{x} \in \partial\Omega \\ \mu(\mathbf{x}) &= 1 + \cos(2\pi x_1)\, \cos(2\pi x_2) & \mathbf{x} \in \Omega \end{aligned}$$

The discrete variational formulation

As done many times before we consider a discrete variational formulation, i.e., a variational formulation posed on a space of finite dimension, as in the Galerkin method we have just presented: Find $u_h \in V(\mathcal{T}_h) \subset H_0^1(\Omega)$, such that

$$\int_\Omega \mu\, \nabla u_h \cdot \nabla v_h \, dx = \int_\Omega f\, v_h \, dx \quad \forall v_h \in V(\mathcal{T}_h) \tag{6.29}$$

where $\mathcal{T}_h$ denotes a partition of Ω into elements or cells (triangles or quadrilaterals) of characteristic size h, i.e., $\mathcal{T}_h = \{K_i\}_{i=1}^N$ such that

$$\bar{\Omega} = \bigcup_{i=1}^{N} \bar{K}_i$$

and the discrete space $V(\mathcal{T}_h)$ is defined as

$$V(\mathcal{T}_h) = \left\{ v \in H_0^1(\Omega), v|_K \in P_k(K) \;\; \forall K \in \mathcal{T}_h \right\} \tag{6.30}$$

where $P_k(K)$ stands for the space of polynomials of degree k in each element K (linears, quadratics, etc.).

The task is to study how the error of the finite element solution $e_h = u - u_h$ converges as h is refined and the degree k of polynomials is increased.

Implement the following modifications to `convergence_poisson.py`

- Implement a function that solves the Poisson's problem by encapsulating everything needed to compute the solution error in the $L^2(\Omega)$ and the $H^1(\Omega)$–norms as well as the error in the integral of u

$$I_u = \int_\Omega u \, dx$$

Take care to apply the exact solution u as a Dirichlet boundary condition. Also, notice that, you don't actually need to compute by hand the expression of the source term f, which may be tedious task, instead, define:

```
.
f = -div(mu(x)*grad(uex(x)))
.
```

which does the job for you. The implementation may follow the next guidelines:

```
def SolvePoisson(N=16, degree=1, celltype):

    msh = mesh.create_rectangle(comm=MPI.COMM_WORLD,
                                points=((0.0, 0.0),
                                        (1.0, 1.0)),
                                n=(N, N),
                                cell_type=celltype)

    V = fem.FunctionSpace(msh, ("Lagrange", degree))
    .
    .
    .
    eh = uh - uex(x)
    .
    .
    return EL2, EH1, EInt
```

- Perform a numerical assessment of the formulation by:
  - Refining the mesh;
  - Increasing the polynomial order;
  - Switching from triangular to quadrilateral cells;

You can perform a loop:

```
celltype = mesh.CellType.triangle
degree = 1
hh, errsL2, errsH1, errsInt = [], [], [], []
for N in [16, 32, 64, 128, 256]:
    EL2, EH1, EInt = SolvePoisson(N, degree, celltype)
    hh.append(1.0/N)
    errsL2.append(...


    .
    .
```

- Plot the error in loglog scale as a function of the mesh refinement h. Estimate the convergence rates, i.e., the powers p and q and r such that
  - $e_{L^2(\Omega)} = \|u - u_h\|_{L^2(\Omega)} \sim h^p$;
  - $e_{H^1(\Omega)} = \|u - u_h\|_{H^1(\Omega)} \sim h^q$;
  - $e_{\int_\Omega u} = \left|\int_\Omega u\, dx - \int_\Omega u_h\, dx\right| \sim h^r$

You should obtain something similar to what is shown in the figure. The theory behind these orders will appear in coming

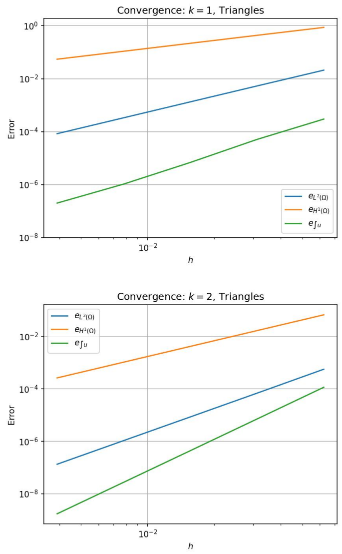


Figure 6.20: Errors as a function of h for the finite element formulation of Poisson's problem corresponding to the exact solution $u = \sin(2\pi x_1)\,\sin(2\pi x_2)$.

lectures.

- Prepare a short report.

INTERPOLATION AND ERROR ESTIMATES

7

Let us recall from chapter 5 that (Céa's Lemma 5.2.1)

$$\|u - u_h\|_V \leq \frac{N_a}{\alpha} \min_{v_h \in V_h} \|u - v_h\|_V ,$$

A particular member of V_h whose properties are well known is the so called interpolant $\mathcal{I}_h u$ (e.g., the Lagrange interpolant), where $\mathcal{I}_h : V \longrightarrow V_h$ is the interpolation operator. So, if we estimate the error of $u - \mathcal{I}_h u$, we can estimate the error $\|u - u_h\|_V$. This is the task in this chapter.

7.1 Interpolation operator

7.1.1 Local and Global interpolation

We consider a finite element (K, P_K, Σ_K), being $\dim(P_K) = n$, a space of functions $V(K)$ defined on K and introduce the interpolation operator:

Definition 7.1.1 ***(Local interpolation operator)*** *Is the operator* $\mathcal{I}_K : V(K) \longrightarrow P_K$, *such that, for a function* $v \in V(K)$ *it satisfies*

$$\sigma_i(\mathcal{I}_K v) = \sigma_i(v), \quad i = 1, \dots, n$$

If $\{\psi^1, \psi^2, \dots, \psi^n\}$ *is a nodal basis of* P_K, *the local interpolation takes the form:*

$$\mathcal{I}_K v = \sum_{i=1}^{n} \sigma_i(v)\, \psi^i \ \ \forall\, v \in V(K) \tag{7.1}$$

Notice that,

- The local interpolation coincides with the function being interpolated at the dofs;
- For the so called Lagrange interpolant, in which de dofs are the function values at a set of points, we have

$$\mathcal{I}_K v = \sum_{i=1}^{n} v(\mathbf{x}^i)\, \psi^i \ \ \forall\, v \in V(K)$$

- The previous definition is relevant to as interpolation on individual elements. The next step is to define a global interpolator:

Definition 7.1.2 ***(Global interpolation operator)*** *Consider a finite element function space* V_h, *i.e., a discrete space associated to a partition* $\mathcal{T}_h$ *of* Ω *equipped with finite elements,*

$$V_h = \{v \in V,\ v|_K \in P_K \ \forall K \in \mathcal{T}_h\}$$

The global interpolation operator $\mathcal{I}_h : V \longrightarrow V_h$ is defined, for any $u \in V$ such that

$$\mathcal{I}_h u|_K = \mathcal{I}_K u$$

- The global interpolation operator can also be defined by

$$\mathcal{I}_h u = \sum_{K \in \mathcal{T}_h} \sum_{i=1}^{n} \sigma_{K,i}(v|_K)\, \psi_K^i \tag{7.2}$$

- Depending on the choice of degrees of freedom, it may occur that $\mathcal{I}_h u$ is multiple-valued at element boundaries.

- The informal idea of "gluing" the degrees of freedom of neighboring elements to construct spaces of globally continuous functions $V_h = \text{span}\{\phi^1, \ldots, \phi^N\}$ as shown in the previous chapter, can be formalized by introducing a local-to-global-map, as follows:

- Consider a finite element mesh $\mathcal{T}_h$, each cell equipped with a finite element, $\{(K, P_K, \Sigma_K) : K \in \mathcal{T}_h\}$. The local to global mapping is defined by specifying how the local degrees of freedom $\sigma_{K,i}$ relates to global degrees of freedom. We must define a mapping

$$LG_K : \{1, \ldots, n(K)\} \longrightarrow \{1, \ldots, N\},\ \forall K$$

such that[1]

$$\sigma_{LG_K(i)}(v) = \sigma_{K,i}(v|_K),\ \ i = 1, \ldots, n(K)$$

7.2 Error estimates in the $L^\infty(\Omega)$-norm

7.2.1 Shape regular meshes

Prior to introducing classical error estimates, we need an additional concept. Consider an individual triangular element as shown in Figure 7.1. The shape of this element is characterized by its largest side h_K, called the **diameter** and by the size of the largest ball inscribed in K, ρ_K, called the inner diameter

Definition 7.2.1 *(**Shape-regular family of meshes**) A mesh $\mathcal{T}_h$, parameterized by the parameter $h = \max_{K \in \mathcal{T}_h} h_K$, is said to be shape-regular, if there exists $S \in \mathbb{R}$ such that*

$$\frac{h_K}{\rho_K} \leq S \quad \forall K \in \mathcal{T}_h$$

Although, we are interested in error estimates in the $L^2(\Omega)$ and $H^1(\Omega)$-norms, which are the relevant ones for the variational problems at hands, it is instructive to consider the following theorem regarding the *local* interpolation error in the $L^\infty(\Omega)$-norm.

For simplicity sake **we consider from now on P_1 triangular elements**. This is enough to introduce the main ingredients involved and several important results. The idea is to study at the element level the difference between u and its interpolant $\mathcal{I}_K u$ (see Figure 7.1).

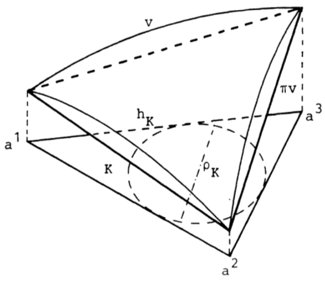


Figure 7.1: A triangular element, showing its diameter h_K, the inner diameter ρ_K and a P_1 interpolant of a function v.

[1] Notice that, if matching dofs of neighboring elements are all mapped to the same global dof, the global approximation ends up being **globally continuous**, as illustrated in Figure **??**.

Theorem 7.2.1 *Let K be a P_1-triangle, h_K its diameter and ρ_K its inner diameter. Then, for all $v \in C^\infty$,*

(a) $\|v - \mathcal{I}_K v\|_{L^\infty(K)} \leq 2\, h_K^2 \max\limits_{|\alpha|=2} \|D^\alpha v\|_{L^\infty(K)}$

(b) $\max\limits_{|\alpha|=1} \|D^\alpha (v - \mathcal{I}_K v)\|_{L^\infty(K)} \leq 6\, \dfrac{h_K^2}{\rho_K} \max\limits_{|\alpha|=2} \|D^\alpha v\|_{L^\infty(K)}$

The proof of this theorem is instructive and involves essentially tools from Calculus. Let $\{\psi^1, \psi^2, \psi^3\}$ be the basis for $P_1(K)$ and let $\mathbf{x}^j = (X_1^j, X_2^j)$ be the position of the j-th node of the element. The local interpolant is

$$\mathcal{I}_K v(\mathbf{x}) = \sum_{i=1}^{3} v(\mathbf{x}^j)\, \psi^j(\mathbf{x}), \quad x \in K \tag{7.3}$$

Now, perform a Taylor expansion around $\mathbf{x} \in K$

$$v(\mathbf{y}) = v(\mathbf{x}) + \sum_{k=1}^{2} \frac{\partial v}{\partial x_k}(\mathbf{x})\, (y_k - x_k) + R(\mathbf{x}, \mathbf{y}), \tag{7.4}$$

where the rest R is given by

$$R(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \sum_{k,\ell}^{2} \frac{\partial^2 v}{\partial x_k \partial x_\ell}(\boldsymbol{\xi})\, (y_k - x_k)\, (y_l - x_l) \tag{7.5}$$

and $\boldsymbol{\xi}$ is a point on the line segment between $\mathbf{x}$ and $\mathbf{y}$. Now, evaluate the expansion at $\mathbf{y} = \mathbf{x}^j$

$$v(\mathbf{x}^j) = v(\mathbf{x}) + p^j(\mathbf{x}) + R^j(\mathbf{x}), \tag{7.6}$$

where $p^j(\mathbf{x}) = \sum\limits_{k=1}^{2} \dfrac{\partial v}{\partial x_k}(\mathbf{x})\, (X_k^j - x_k)$ and $R^j(\mathbf{x}) = R(\mathbf{x}, \mathbf{x}^j)$. Since $\|X_i^j - x_i\| \leq h_K, \;\; j = 1, 2, 3, \;\; i = 1, 2,$ we can write

$$|R^j(\mathbf{x})| \leq 2\, h_K^2 \max_{|\alpha|=2} \|D^\alpha v\|_{L^\infty(K)} \tag{7.7}$$

and inserting $v(\mathbf{x}^j)$ into the definition of the interpolant:

$$\mathcal{I}_K v(\mathbf{x}) = \sum_{j=1}^{3} v(\mathbf{x})\, \psi^j(\mathbf{x}) + \sum_{j=1}^{3} p^j(\mathbf{x})\, \psi^j(\mathbf{x}) + \sum_{j=1}^{3} R^j(\mathbf{x})\, \psi^j(\mathbf{x}) \tag{7.8}$$

Let us consider each term separately:

$$\sum_{j=1}^{3} v(\mathbf{x})\, \psi^j(\mathbf{x}) = v(\mathbf{x}) \sum_{j=1}^{3} \psi^j(\mathbf{x}) = v(\mathbf{x}) \tag{7.9}$$

since $\sum\limits_{j=1}^{3} \psi^j(\mathbf{x}) = 1$. For the second term

$$\sum_{j=1}^{3} p^j(\mathbf{x})\, \psi^j(\mathbf{x}) = \sum_{j=1}^{3} \sum_{k=1}^{2} \frac{\partial v}{\partial x_k}(\mathbf{x})(X_k^j - x_k)\, \psi^j(\mathbf{x}) = \sum_{k=1}^{2} \frac{\partial v}{\partial x_k}(\mathbf{x}) \left\{ \sum_{j=1}^{3} X_k^j \psi^j(\mathbf{x}) - x_k \sum_{j=1}^{3} \psi^j(\mathbf{x}) \right\}$$

But, since $\sum_{j=1}^{3} X_k^j \psi^j(\mathbf{x}) = x_k$, the second term vanishes, yielding

$$\mathcal{I}_K v(\mathbf{x}) = v(\mathbf{x}) + \sum_{j=1}^{3} R^j(\mathbf{x})\,\psi^j(\mathbf{x})$$

and

$$|v(\mathbf{x}) - \mathcal{I} v_K(\mathbf{x})| \le \max_j |R^j(\mathbf{x})| \sum_{j=1}^{3} \psi^j(\mathbf{x}) = \max_j |R^j(\mathbf{x})| \le 2\, h_K^2 \max_{|\alpha|=2} \|D^\alpha v\|_{L^\infty(K)}$$

implying assertion **(a)**. Now, by differentiating $\mathcal{I}_K v(\mathbf{x}) = \sum_{i=1}^{3} v(\mathbf{x}^j)\,\psi^j(\mathbf{x})$ with respect to x_m and using the Taylor expansion again evaluated at $\mathbf{x}^j$ we obtain

$$\frac{\partial \mathcal{I}_K v}{\partial x_m}(x) = \sum_{j=1}^{3} v(x)\,\frac{\partial \psi^j}{\partial x_m}(x) + \sum_{j=1}^{3} p^j(x)\,\frac{\partial \psi^j}{\partial x_m}(x) + \sum_{j=1}^{3} R^j(x)\,\frac{\partial \psi^j}{\partial x_m}(x)$$

On the right-hand side, the first term vanishes. The second term is

$$\sum_{j=1}^{3} p^j(\mathbf{x})\,\frac{\partial \psi^j}{\partial x_m}(\mathbf{x}) = \sum_{j,k} \frac{\partial v}{\partial x_k}(\mathbf{x})\,(X_k^j - x_k)\,\frac{\partial \psi^j}{\partial x_m}(\mathbf{x}) = \sum_{k=1}^{2} \frac{\partial v}{\partial x_k}(\mathbf{x}) \left[\sum_{j=1}^{3} X_k^j \frac{\partial \psi^j}{\partial x_m}(\mathbf{x}) - x_k \sum_{j=1}^{3} \frac{\partial \psi^j}{\partial x_m}(\mathbf{x}) \right] =$$

$$= \sum_{k=1}^{2} \frac{\partial v}{\partial x_k}(\mathbf{x}) \frac{\partial}{\partial x_m} \sum_{j=1}^{3} X_k^j \psi^j(\mathbf{x}) = \sum_{k=1}^{2} \frac{\partial v}{\partial x_k}(\mathbf{x}) \frac{\partial}{\partial x_m} x_k = \sum_{k=1}^{2} \frac{\partial v}{\partial x_k}(\mathbf{x}) \delta_{km} = \frac{\partial v}{\partial x_m}(\mathbf{x})$$

Finally, taking the absolute value and using previous results for $|R^j(\mathbf{x})|$

$$\left| \frac{\partial v}{\partial x_m}(\mathbf{x}) - \frac{\partial \mathcal{I}_K v}{\partial x_m}(\mathbf{x}) \right| = \left| \sum_{j=1}^{3} R^j(\mathbf{x})\,\frac{\partial \psi^j}{\partial x_m}(\mathbf{x}) \right| \le \max_j |R^j(\mathbf{x})| \sum_{j=1}^{3} \left| \frac{\partial \psi^j}{\partial x_m} \right| \le 2\,\frac{h_K^2}{\rho_K} \max_{|\alpha|=2} \|D^\alpha v\|_{L^\infty(K)}$$

since $\left| \frac{\partial \psi^j}{\partial x_m} \right| \le \frac{1}{\rho_K}$ (by looking at the figure in the beginning of this section, the reader can convince himself that the derivative of a P_1 function which equals 1 at a given node and is zero on the oposite side can never be grater than $1/\rho_K$).

The results generalizes for dimension d as

$$\|v - \mathcal{I}_K v\|_{L^\infty(K)} \le \frac{d^2}{2}\, h_K^2 \max_{|\alpha|=2} \|D^\alpha v\|_{L^\infty(K)}$$

$$\max_{|\alpha|=1} \|D^\alpha (v - \mathcal{I}_K v)\|_{L^\infty(K)} \le \frac{(d+1)d^2}{2}\, \frac{h_K^2}{\rho_K} \max_{|\alpha|=2} \|D^\alpha v\|_{L^\infty(K)}$$

7.2.2 From local to global

By collecting the contributions from all the elements in the mesh we can build a global error estimate as follows:

$$\|u - \mathcal{I}_h u\|_{L^\infty(\Omega)} \le \max_{K \in \mathcal{T}_h} \|u - \mathcal{I}_K u\|_{L^\infty(K)} \le \frac{d^2}{2} \max_{K \in \mathcal{T}_h} \left\{ h_K^2\, \|D^\alpha v\|_{L^\infty(K)} \right\} \le \frac{d^2}{2}\, h^2 \|D^2 v\|_{L^\infty(\Omega)}$$

and similarly for the gradient

$$\|\nabla u - \nabla(\mathcal{I}_h u)\|_{L^\infty(\Omega)} \leq \frac{(d+1)\, d^2 S}{2}\, h^2 \|D^2 v\|_{L^\infty(\Omega)}$$

7.3 Error estimates in Sobolev norms for Lagrange elements

We begin by stating a local estimate in the following theorem:

Theorem 7.3.1 *Let (K, P_K, Σ_K) be a Lagrange finite element such that:*

- *P_K contains all polynomials of degree $\leq k$*
- *it is affine equivalent to the master element $(\hat{K}, \hat{P}, \hat{\Sigma})$*

Then, the Lagrange interpolant $\mathcal{I}_K u = \sum_{i=1}^{n} v(\mathbf{x}^i)\psi^i$ satifies the following error estimates: For all $\ell \leq k$

(a) $\|u - \mathcal{I}_K u\|_{L^2(K)} \leq C\, h_K^{\ell+1} \|D^{\ell+1} u\|_{L^2(K)}$

(b) $\|\nabla u - \nabla(\mathcal{I}_K u)\|_{L^2(K)} \leq C\, \dfrac{h_K^{\ell+1}}{\rho_K} \|D^{\ell+1} u\|_{L^2(K)}$

with C being independent on h_K and u.

- The $L^2(K)$-norm of the gradient is sometimes called the $H^1(K)$-seminorm[2], i.e.,

$$\|\nabla u - \nabla(\mathcal{I}_K u)\|_{L^2(K)} = |u - \mathcal{I}_K u|_{H^1(K)}$$

- Notice that in the estimate for this seminorm, we also have ρ_K dividing.

The proof can be found in [5]. We finally present a global error estimate for continuous Lagrange elements[3], which is of fundamental importance for this course:

Theorem 7.3.2 *Let V_h be a function space associated to a family of shape-regular meshes $\mathcal{T}_h,\ h > 0$ of a domain $\Omega \subset \mathbb{R}^d$, equipped with continuous Lagrange elements of order p. Let $u \in H^{p+1}(\Omega)$, and let $\mathcal{I}_h : H^{p+1}(\Omega) \to V_h$ be the global interpolation operator. Then, there exists a constant C independent of u and h, such that*

$$\|u - \mathcal{I}_h u\|_{H^1(\Omega)} \ \leq\ C\, h^p\, |u|_{H^{p+1}(\Omega)}$$

Let us comment about the previous theorem:

- First, the theorem says that the interpolation error when using continuous Lagrange elements of order p is bounded by the size (in the $L^2(\Omega)$-norm) of the $(p+1)$–th derivatives of the function being interpolated;

- This bound, also says that the error scales as $\mathcal{O}(h^p)$. This means, we have two paths to reduce the error: we can refine the mesh size h or we can increase the polynomial order p.

[2] It is called a seminorm because $|u|$ can be zero with u not being necessarely zero.

[5]: Ern et al. (2004), *Theory and Practice of Finite Elements*

[3] Recall, these are the most commonly used ones in finite elements. In particular, those of order 1 or 2.

- The latter is true, provided the solution is smooth enough, such that it belongs to H^{q+1}, $q > p$. For instance, let us suppose $u \in H^2(\Omega)$ but $u \notin H^3(\Omega)$. Then, using linear elements ($p = 1$) is fine, however, upgrading to quadratics may not improve the result[4].

- Notice that u must belong at least to $H^2(\Omega)$, so, we need to guarantee, that the solution of the continuous problem, that is posed in $H^1(\Omega)$, actually lives in $H^2(\Omega)$ at least [5];

- In Figure 7.2 we see the infamous example of a solution to Laplace's problem in the domain $[-1, 1] \times [-1, 1]$ with a re-entrant corner. The solution to this problem reads

  $$u = r^{\gamma} \sin(\gamma\theta), \quad \theta = \operatorname{atan}(y/x), \quad \gamma = \frac{\pi}{\omega}$$

  where ω is the angle of the re-entrant corner measured in counter-clockwise sense ($\gamma = \dfrac{\pi}{7\pi/4}$ in the example). The figure shows the interpolation $\mathcal{I}_h u$ onto a mesh made of triangles, locally refined near the corner. This is an example of a function belonging to $H^{1+\gamma-\varepsilon}(\Omega)\ \forall \varepsilon > 0$, so, it does not belong to $H^2(\Omega)$.

- Going back to Céa's Lemma 5.2.1, for the solution of our variational problem (be it Poisson's or the elasticity problem), using polynomials of degree p, provided the solution $u \in H^{p+1}(\Omega)$, we have the following error estimate[6].

  $$\|u - u_h\|_{H^1(\Omega)} \;\leq\; C\, \frac{N_a}{\alpha}\, |u|_{H^{p+1}(\Omega)}.$$

Solve the following exercices

- Read carefully the proof of Theorem 7.2.1;

- Show that the local interpolation operator is a projection, i.e.,

  $$\mathcal{I}_K p = p \qquad \text{for all } p \in P_K .$$

- Let F be the linear form defined by $F(v) = \int_{\Omega} f(x)\, v(x)\, dx$, where $f \in L^2(\Omega)$. For instance, if $f = 1$, $F(v)$ is nothing but the integral of v over Ω. Let us suppose we have $u_h \in V_h$, where V_h is a discrete space made up of piecewise polynomials of degree p (as in Theorem 7.3.2), such that $\|u - u_h\|_{H^1(\Omega)} \leq C h^p$. Let us consider the following PDE

  $$\begin{aligned} \mathcal{L}(w) &= f \text{ in } \Omega \\ w &= 0 \text{ on } \partial\Omega \end{aligned}$$

  Considering the domain is regular enough for the operator $\mathcal{L}$ to have a smoothing property, the solution w satisfies

  $$\|w\|_{H^2(\Omega)} \leq c\|f\|_{L^2(\Omega)}$$

  The operator $\mathcal{L}(\cdot)$ can be the Laplacian for instance. This prob-

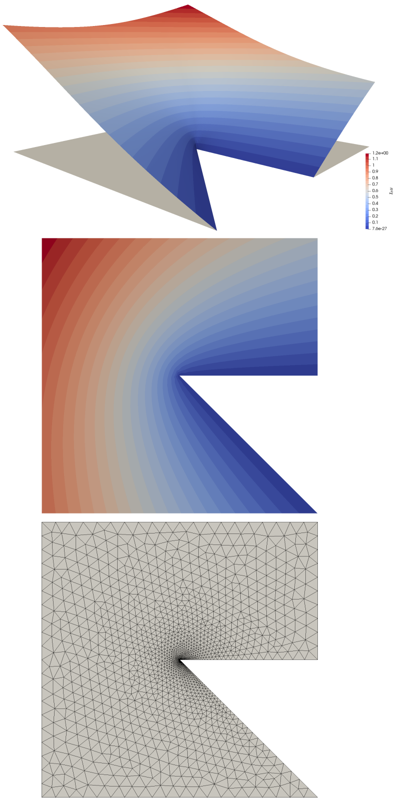


Figure 7.2: Example of the function $u = r^{\gamma} \sin(\gamma\theta) \notin H^2(\Omega)$, $\gamma = \frac{4}{7}$, solution of Laplace's problem on a domain with a re-entrant corner.

[4] This is called h–adaptivity and p–adaptivity in the context of finite elements. Also, hp–adaptivity can be done.

[5] For this, we need something called an elliptic regularity result, to be introduced in Chapter 9.

[6] Recall, we may run into trouble if $\frac{N_a}{\alpha} \gg 1$.

lem in weak form is written as:

$$a(w, v) = F(v) \;\; \forall v \in H_0^1(\Omega)$$

where the bilinear form $a(\cdot, \cdot)$ is bounded with continuity constant N_a. Using the following calculation

$$\begin{aligned} F(u - u_h) = a(w, u - u_h) &\overset{\text{why?}}{=} a(w - \mathcal{I}_h w, u - u_h) \\ &\leq N_a \, \|w - \mathcal{I}_h w\|_{H^1(\Omega)} \, \|u - u_h\|_{H^1(\Omega)} \end{aligned}$$

to prove that $|F(u) - F(u_h)| \leq \tilde{C} h^{2p}$. Is this consistent with the results we obtained in the last computational assignment?.

`FEniCSx` BEHIND THE SCENES

8

Now, we discuss practical aspects and introduce some "technology" needed for the actual computation of a finite element matrix as those previously introduced and considering some of the finite element spaces constructed. Here, we consider affine families of finite elements.

8.1 Finite element matrizes

Having selected a finite dimensional space and a finite element basis $\{\phi^1, \dots, \phi^n\}$ associated to a finite element partition $\mathcal{T}_h$ of Ω

$$\bar{\Omega} = \bigcup_{i=1}^{N} \bar{K}_i$$

computation of a finite element matrix tyically involves integrals of the type

$$A_{ij} = a(\phi^i, \phi^j) = \int_\Omega \left[\phi^i(\mathbf{x})\, \phi^j(\mathbf{x}) + \nabla\phi^i(\mathbf{x}) \cdot \nabla\phi^j(\mathbf{x})\right] d\Omega$$

Let us consider the first term in the integral above. We can compute then the integral summing over all the elements

$$M_{ij} = \int_\Omega \phi^i(\mathbf{x})\, \phi^j(\mathbf{x})\, d\Omega = \sum_{K_m \in \mathcal{T}_h} \int_{K_m} \phi^i(\mathbf{x})|_{K_m}\, \phi^j(\mathbf{x})|_{K_m}\, dK \tag{8.1}$$

The notation above is redundant, because we are integrating on K_m. Now, we make use of the affine mapping we have previously introduced. The idea is to transform the integral over K_m into an integral over $\hat{K}$ which is **easier** to handle. By doing the change of variables

$$\int_{K_m} \phi^i(\mathbf{x})|_{K_m}\, \phi^j(\mathbf{x})|_{K_m}\, dK = \int_{\hat{K}} \phi^i(F_K(\hat{\mathbf{x}}))\, \phi^j(F_K(\hat{\mathbf{x}}))\, |J_{K_m}|\, d\hat{K}$$

where

$$|J_{K_m}| = |\mathrm{det} B_{K_m}|$$

i.e., the determinant of the Jacobian of the affine transformation for element K. The idea is to use the basis functions defined on the master element and not the functions defined on the real element. As an example, consider the case of a triangular mesh $\mathcal{T}_h$ and P_1 linear elements. We have constructed basically two types of spaces:

- **Space of totally discontinous functions**

  $$X_h(\mathcal{T}_h) = P_1^{\text{disc}}(\mathcal{T}_h) = \{v,\, v|_{K_i} \in P_1(K_i),\, v(\mathbf{x}) = 0 \,\text{if}\, \mathbf{x} \notin K_i\, \forall K_i \in \mathcal{T}_h\}$$

  which is spanned by a set of $n = 3 \times N_e$ basis functions $\{\phi^1, \phi^2, \dots, \phi^n\} = \{\psi^1_{K_1}, \psi^2_{K_1}, \psi^3_{K_1}, \dots, \psi^1_{K_{N_e}}, \psi^2_{K_{N_e}}, \psi^3_{K_{N_e}}\}$, where there is a correspondence between the index of ϕ^I and the supraindex and subindex of $\psi^r_{K_m}$, say $I = LG_{K_m}(r)$ (we denote the local-to-global map sometimes by iglob(r,K_m)). Each set $\{\psi^1_{K_m}, \psi^2_{K_m}, \psi^3_{K_m}\}$ forms a set of local basis functions on K_m, for which we have the set $\{\hat{\psi}^1_{\hat{K}}, \hat{\psi}^2_{\hat{K}}, \hat{\psi}^3_{\hat{K}}\}$ of functions defined on the master element $\hat{K}$ because both are affine equivalent, i.e.

$$\psi^r_{K_m}(\mathbf{x}) = \psi^r_{K_m}(F_{K_m}(\hat{\mathbf{x}})) = \hat{\psi}^r_{\hat{K}}(\hat{\mathbf{x}}), \quad r = 1, 2, 3$$

Now, `A` will be constructed by summing over all the elements, however, in this case the support of any function is a single element, so, if supp (ϕ^i) = supp $(\phi^j) = K_m$, we have

$$A_{ij} = \int_{K_m} \phi^i|_{K_m}\, \phi^j|_{K_m}\, dK = \int_{K_m} \psi^r_{K_m}\, \psi^s_{K_m}\, dK = \int_{\hat{K}} \hat{\psi}^r_{\hat{K}}\, \hat{\psi}^s_{\hat{K}}\, |J_{K_m}|\, d\hat{K}$$

otherwise A_{ij} will be zero.

- **Space of continous functions**

$$V_h(\mathcal{T}_h) = P_1(\mathcal{T}_h) = X(\mathcal{T}_h) \cap C^0(\Omega)$$

which is spanned by a set of $n = N_v$ basis functions $\{\phi^1, \phi^2, \ldots, \phi^n\}$. Again, `A` will be constructed by summing over all the elements. In this case the support of basis function ϕ^i are all the triangles that share vertex i, so, coefficient A_{ij} will be

$$A_{ij} = \sum_{\substack{K_m \in \\ (\mathrm{supp}(\phi^i) \cap \\ \mathrm{supp}(\phi^j))}} \int_{K_m} \phi^i|_{K_m}\, \phi^j|_{K_m}\, dK$$

but $\phi^i|_{K_m} = \psi^r_{K_m}$ and $\phi^j|_{K_m} = \psi^s_{K_m}$ for some r and s, then

$$\int_{K_m} \phi^i|_{K_m}\, \phi^j|_{K_m}\, dK = \int_{K_m} \psi^r_{K_m}\, \psi^s_{K_m}\, dK = \int_{\hat{K}} \hat{\psi}^r_{\hat{K}}\, \hat{\psi}^s_{\hat{K}}\, |J_{K_m}|\, d\hat{K}$$

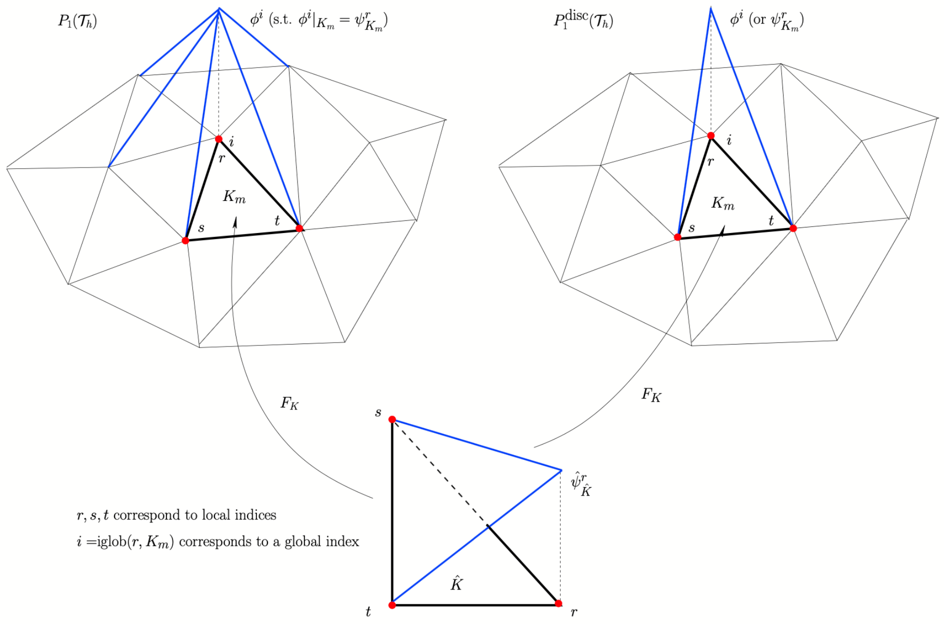
As seen, in either case, what just need to compute elemental contributions to matrix `A` *by integrating the basis functions defined on the master element* $\hat{K}$*.* Now consider the term involving the derivatives of the basis functions. We have

$$K_{ij} = \int_\Omega \nabla\phi^i(\mathbf{x}) \cdot \nabla\phi^j(\mathbf{x})\, d\Omega = \sum_{K_m \in \mathcal{T}_h} \int_{K_m} \nabla\phi^i(\mathbf{x})|_{K_m} \cdot \nabla\phi^j(\mathbf{x})|_{K_m}\, dK \tag{8.2}$$

Once again, we transform the integral over K_m into an integral over $\hat{K}$, for which we need the previous result obtained in 6.21,

$$\int_{K_m} \nabla\phi^i(\mathbf{x})|_{K_m} \cdot \nabla\phi^j(\mathbf{x})|_{K_m}\, dK = \int_{K_m} \nabla\psi^r_{K_m}(\mathbf{x}) \cdot \nabla\psi^s_{K_m}(\mathbf{x})\, dK =$$
$$= \int_{\hat{K}} [B^{-T}_{K_m} \cdot \hat{\nabla}\hat{\psi}^r_{\hat{K}}(\hat{\mathbf{x}})] \cdot [B^{-T}_{K_m} \cdot \hat{\nabla}\hat{\psi}^s_{\hat{K}}(\hat{\mathbf{x}})]\, |J_{K_m}|\, d\hat{K}$$

Again, we work with the local basis functions.



8.2 Numerical integration

Although, the master element $\hat{K}$ has a simpler shape, integrals are sometimes difficult to be perfomed exactly. Even when the coefficients of matrix A involve the integration of polynomial functions, the right hand side may involve any function f:

$$F_i = \int_K f(\mathbf{x})\,\phi^i(\mathbf{x})|_K\,dK = \int_{\hat{K}} f(F_K(\hat{\mathbf{x}}))\,\psi^r_K(F_K(\hat{\mathbf{x}}))\,|J_K|\,d\hat{K} = \int_{\hat{K}} \hat{f}(\hat{\mathbf{x}})\,\hat{\psi}^r_{\hat{K}}(\hat{\mathbf{x}})\,|J_K|\,d\hat{K}$$

In theses cases we use numerical integration.

Definition 8.2.1 *(**Quadrature rule**) Let $\hat{K}$ be non-empty compact connected subset. Let n_g be an integer. A quadrature on $\hat{K}$ with n_g points consists of:*

(i) A set of n_g real numbers $\{w_1, w_2, \ldots, w_{n_g}\}$ called quadrature weights;
(ii) A set of n_g points $\{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_{n_g}\}$ called Gauss points or quadrature nodes.

The largest integer such that

$$\forall \hat{p} \in \hat{P}_k \quad \int_{\hat{K}} \hat{p}(\hat{\mathbf{x}})\,d\hat{K} = \sum_{g=1}^{n_g} w_g\,\hat{p}(\hat{\mathbf{x}}_g)$$

is called the quadrature order and is denoted by r. It can be shown that

$$\frac{1}{meas(\hat{K})}\left|\int_{\hat{K}} f(\mathbf{x})\,d\hat{K} - \sum_{g=1}^{n_g} w_g\,\hat{f}(\hat{\mathbf{x}}_g)\right| \le c\,h_{\hat{K}}^{r+1} \sup_{\hat{\mathbf{x}}\in\hat{K},\ |\alpha|=r+1} |D^\alpha f(\hat{\mathbf{x}})|$$

*where $h_{\hat{K}}$ is the **diameter** of $\hat{K}$ (the largest side) and $c > 0$ is a constant.*


Figure 8.1: Adrien-Marie Legendre (France, 1752–1833).

We see again the practicity of working on the master element, since in this case we define the rules only once and for all.

These quadratures are "tabulated" (see Table 8.1). In 1D we have the so called Gauss-Legendre quadratures. Considering the master element being the interval $[-1,1]$ quadratures of order $r = 3$, 5 and 7 are displayed in the table below. The quadrature points are zeros of Legendre polynomials[1].

Table 8.1: 1D Gauss-Legendre quadrature rules (points in reference element $[-1,1]$ and weights. Note these quadratures are symmetric.

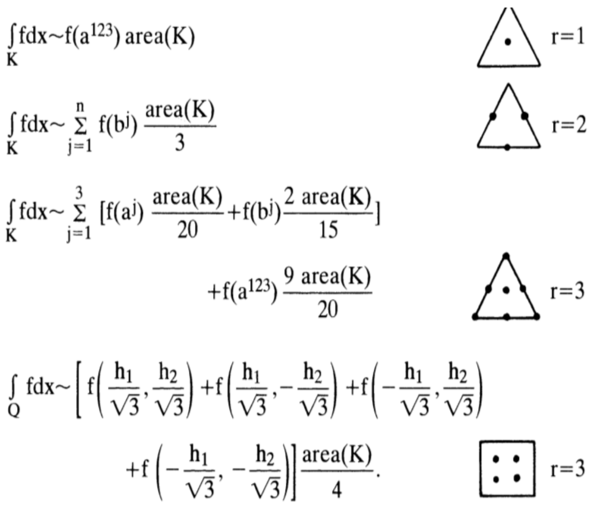
n	Points ξ_i	Weights w_i
2	0.5773502692 −0.5773502692	1.0000000000 1.0000000000
3	0.7745966692 0.0000000000 −0.7745966692	0.5555555556 0.8888888889 0.5555555556
4	0.8611363116 0.3399810436 −0.3399810436 −0.8611363116	0.3478548451 0.6521451549 0.6521451549 0.3478548451
5 .		

- These rules can be adapted to other intervals rather than the reference one by simple change of variables

$$\int_a^b f(x)\,dx = \frac{b-a}{2}\int_{-1}^{1} f\left(\frac{a+b}{2}+\frac{b-a}{2}\,\xi\right)d\xi$$

- The cartesian product of 1D quadratures can be used in 2D and 3D so as to construct quadratures on quadrilateral and hexahedral elements;

- For elements of other shapes, such as triangles, tetrahedra, etc, the quadratures can also be computed (see Figure 8.2).



$$\int_K f dx \sim f(a^{123})\,\text{area}(K)$$

$$\int_K f dx \sim \sum_{j=1}^{n} f(b^j)\,\frac{\text{area}(K)}{3}$$

$$\int_K f dx \sim \sum_{j=1}^{3}\left[f(a^j)\,\frac{\text{area}(K)}{20}+f(b^j)\frac{2\,\text{area}(K)}{15}\right] + f(a^{123})\,\frac{9\,\text{area}(K)}{20}$$

$$\int_Q f dx \sim \left[f\left(\frac{h_1}{\sqrt{3}},\frac{h_2}{\sqrt{3}}\right)+f\left(\frac{h_1}{\sqrt{3}},-\frac{h_2}{\sqrt{3}}\right)+f\left(-\frac{h_1}{\sqrt{3}},\frac{h_2}{\sqrt{3}}\right) + f\left(-\frac{h_1}{\sqrt{3}},-\frac{h_2}{\sqrt{3}}\right)\right]\frac{\text{area}(K)}{4}.$$

Figure 8.2: Some examples of quadrature rules on triangular and quadrilateral elements.

8.3 Programming the assembly of finite element matrices

In this section we explain the assembly process used to construct the linear system of equations associated to one of the variational problems described above.

Summarizing

1. A partition $\mathcal{T}_h$ of Ω made up of elements $K_m,\ m = 1,\dots,N_{el}$ that are affine equivalent to a master element $\hat{K}$.
2. A space of functions associated to the partition: $V_h(\mathcal{T}_h)$, $\dim V_h = N$ (it can be a space of totally discontinuous or continuous

---

[1] Note that, by the definition, these rules are defined such that they provide the exact result for polynomials up to the corresponding degree, otherwise they provide an approximate value. This is sometimes referred to as a **variational crime**

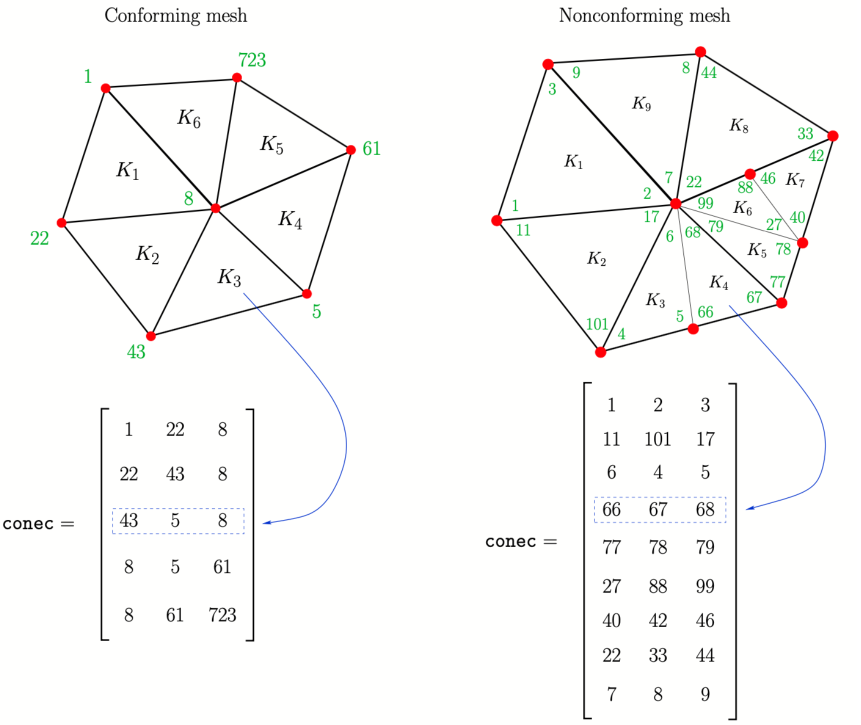
functions).

3. An incidence or **connectivity** matrix `conec` of dimension $N_{el} \times n_{loc}$ that describes the relation between the elements in $\mathcal{T}_h$ and the global unknowns (see figure below).
4. A quadrature rule on $\hat{K}$: $\{(w_g, \hat{\mathbf{x}}_g)\},\ g = 1, \ldots, n_g$.

As an example, let us consider the $H^1(\Omega)$-projection of a function u over the space $V(\mathcal{T}_h)$, i.e., Find $u_h \in V_h$ such that

$$\int_\Omega (u_h\, v_h + \nabla u_h \cdot \nabla v_h)\, dx = \int_\Omega (u\, v_h + \nabla u \cdot \nabla v_h)\, dx \ \ \forall v_h \in V_h$$

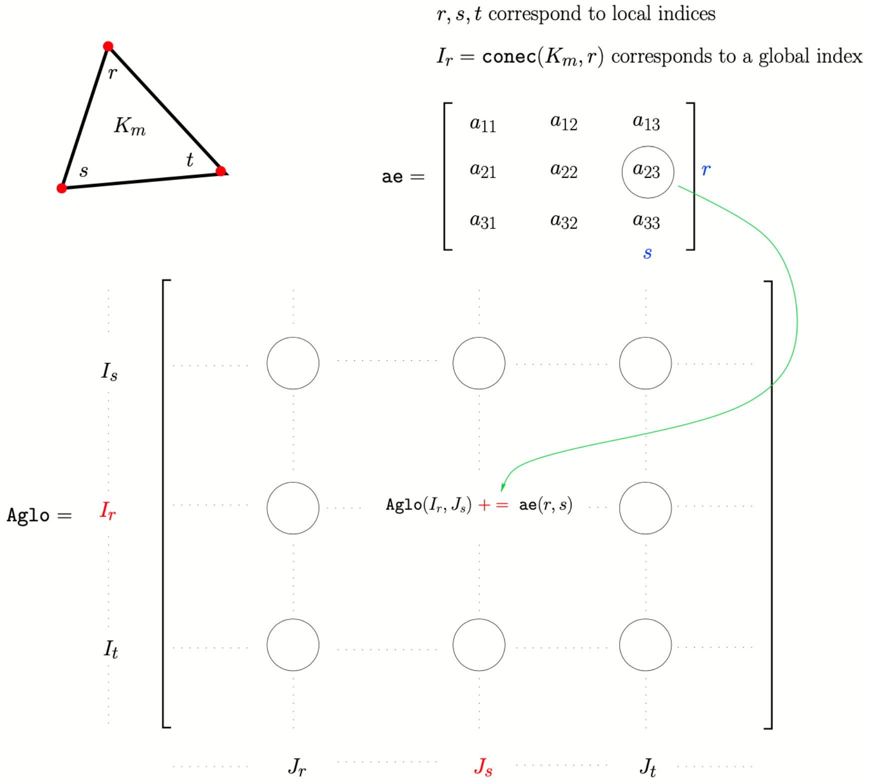
with these ingredients the global matrix A denoted by `Aglo` below and the global right hand side vector **F** denoted by `RHS` can be assembled as shown in the pseudo-code below.



$$\texttt{conec} = \begin{bmatrix} 1 & 22 & 8 \\ 22 & 43 & 8 \\ 43 & 5 & 8 \\ 8 & 5 & 61 \\ 8 & 61 & 723 \end{bmatrix}$$

$$\texttt{conec} = \begin{bmatrix} 1 & 2 & 3 \\ 11 & 101 & 17 \\ 6 & 4 & 5 \\ 66 & 67 & 68 \\ 77 & 78 & 79 \\ 27 & 88 & 99 \\ 40 & 42 & 46 \\ 22 & 33 & 44 \\ 7 & 8 & 9 \end{bmatrix}$$

```
function [Aglo   RHS] = Assembly( ... )
    for $g = 1, \ldots, n_g$ do                ▷ Basis functions and derivatives at Gauss points on $\hat{K}$
        Calculate $\hat{\psi}^r_{\hat{K}}(\hat{\mathbf{x}}_g)$ and $\hat{\nabla}\hat{\psi}^r_{\hat{K}}(\hat{\mathbf{x}}_g)$,  $r = 1, \ldots, n_{loc}$
    end for
    Initialize RHS and Aglo to zero
    for $m = 1, \ldots, N_{el}$ do                ▷ Loop over elements
        Calculate $|J_{K_m}|$ and $B^{-T}_{K_m}$
        Initialize rhse and ae to zero
        for $g = 1, \ldots, n_g$ do                ▷ Loop over Gauss points
            for $r = 1, \ldots, n_{loc}$ do
                $\mathtt{rhse}(r) = \mathtt{rhse}(r) + |J_{K_m}| * w_g * \left[u(F_K(\hat{\mathbf{x}}_g)) * \hat{\psi}^r_{\hat{K}}(\hat{\mathbf{x}}_g) + \nabla u(F_K(\hat{\mathbf{x}}_g)) \cdot B^{-T}_{K_m} \cdot \hat{\nabla}\hat{\psi}^r_{\hat{K}}(\hat{\mathbf{x}}_g)\right]$
                for $s = 1, \ldots, n_{loc}$ do
                    $\mathtt{ae}(r,s) = \mathtt{ae}(r,s) + |J_{K_m}| * w_g * \left[\hat{\psi}^r_{\hat{K}}(\hat{\mathbf{x}}_g) * \hat{\psi}^s_{\hat{K}}(\hat{\mathbf{x}}_g) + B^{-T}_{K_m} \cdot \hat{\nabla}\hat{\psi}^r_{\hat{K}}(\hat{\mathbf{x}}_g) \cdot B^{-T}_{K_m} \cdot \hat{\nabla}\hat{\psi}^s_{\hat{K}}(\hat{\mathbf{x}}_g)\right]$
                end for
            end for
        end for                ▷ End loop over Gauss points
        for $r = 1, \ldots, n_{loc}$ do                ▷ Assembly elementary matrix into global matrix
            $I = \mathtt{conec}(K_m, r)$
            $\mathtt{RHS}(I) = \mathtt{RHS}(I) + \mathtt{rhse}(r)$
            for $s = 1, \ldots, n_{loc}$ do
                $J = \mathtt{conec}(K_m, s)$
                $\mathtt{Aglo}(I,J) = \mathtt{Aglo}(I,J) + \mathtt{ae}(r,s)$
            end for
        end for
    end for                ▷ End loop over elements
end function
```



8.4 `FEniCSx` components

At this point, it becomes quite clear that many of the computations we have been presenting in the last chapters, such as, **construction of finite element spaces**, **assembly of finite element matrizes**, **numerical integration**, **imposition of boundary conditions**, **matrix solving**, and so on, that were done using the high-level `FEniCSx` platform, remains totally hidden to the user. All these calculations being possible, thanks to several components that form the `FEniCSx` platform. These are[2]

1. `UFL`: `Python` library for writing problems in variational form. It provides the syntaxis to define linear and bilinear forms, finite elements spaces, such that the PDEs can be written in weak form in a language that is close to the mathematical one;

2. `DOLFINx`: Dynamic Object Oriented Library for Finite Element Computation. It provides the computational environment of FEniCSx in C++ and Python and serves, among other things, to interface to parallel linear algebra routines, such as `Petsc`;

3. `FFCx`: FEniCS form compiler. From a high-level description of the form in the Unified Form Language (UFL), it generates efficient low-level C code that can be used to assemble the corresponding discrete operators. Also, the tutorial about *Just-in-time-compilation* reveals interesting information (see `JIT`);

4. `Basix`: Is a finite element definition and tabulation runtime library. Basix allows users e.g. to evaluate finite element basis functions, access geometric and topological information about reference cells, interpolate into a finite element space, among other things;

For further details, it is highly recommended to access the specific repositories in the green links above.

[2] The descriptions of the different components are just a brief summary taken from their corresponding repository at `https://github.com/FEniCS`.

MORE ON ELLIPTIC PROBLEMS

9

9.1 Preliminaries

When discussing elliptic problems in Chapter 4, we have left several details apart. In this chapter we aim to fill in some of these gaps. We begin by stating an important theorem that says about how to properly define the restriction of a functions that belongs to $H^1(\Omega)$ to the boundary of the domain, which is of fundamental importance when dealing with PDEs and boundary value problems.

Boundary terms and *traces*

Theorem 9.1.1 *(**The trace theorem**) Let $\Omega \subset \mathbb{R}^d$ be an open bounded domain with Lipschitz boundary Γ. Then, there is a unique bounded linear map*

$$\gamma : H^1(\Omega) \to L^2(\Gamma)$$

and a constant $C > 0$ such that

$$\|\gamma(u)\|_{L^2(\Gamma)} \leq C\|u\|_{H^1(\Omega)} \quad \forall\, u \in H^1(\Omega)$$

with the property that if $u \in C^1(\Omega)$, then $\gamma(u) = u|_\Gamma$ in the conventional sense.

- The operator γ maps u to its values on Γ;
- Note that we usually ommit in the notation the trace operator γ and simply write $u|_\Gamma$. However, this should always be understood in the *sense of traces*, which as mentioned, coincides with the conventional one for C^1 functions;
- This theorem is important, since it provides the rigorous way of defining (unambigously) boundary values of function belonging to H^1.

9.2 Convection-Diffusion-Reaction problem

In this section we consider a more general version of Poisson's problem, namely the Convection-Diffusion-Reaction (CDR) equation. We assume $\Omega \subset \mathbb{R}^d$ and boundary $\partial\Omega$. We assume mixed boundary data, with Dirichlet and Neumann parts (as usual, $\partial\Omega = \Gamma_D \cup \Gamma_N$ and $\Gamma_D \cap \Gamma_N = \emptyset$).

The problem reads

$$\begin{cases} \overbrace{-\nabla\cdot(\boldsymbol{\kappa}\cdot\nabla u)}^{\text{diffusion}} + \overbrace{\boldsymbol{\beta}\cdot\nabla u}^{\text{convection}} + \overbrace{\sigma\, u}^{\text{reaction}} = f & \text{in } \Omega \\ u = g & \text{on } \Gamma_D \\ -(\boldsymbol{\kappa}\cdot\nabla u)\cdot\check{\mathbf{n}} = H & \text{on } \Gamma_N \end{cases} \tag{9.1}$$

Also, we admit the diffusivity $\boldsymbol{\kappa}$ to be a $d\times d$ symmetric positive definite matrix, however, notice that the simple scalar case corresponds to

$$\boldsymbol{\kappa} = \kappa\,\mathbf{I}_{d\times d} \quad\Rightarrow\quad \nabla\cdot(\boldsymbol{\kappa}\cdot\nabla u) = \nabla\cdot(\kappa\nabla u)$$

Finally, the convection velocity $\boldsymbol{\beta}$ is a d-dimensional vector and the reaction coefficient σ is scalar-valued. The variational formulation follows after multiplying by a sufficiently regular test function and integrating by parts. As done in Chapter 4, we consider an additive decomposition for u such that we write

$$u = u_0 + u_g$$

where u_0 satisfies homogeneous Dirichlet conditions on Γ_D and u_g is a lifting function satisfying the nonhomogeneous data g on Γ_D (see Figure 4.1)[1]

By defining the bilinear form

$$a(w,v) = \int_\Omega (\boldsymbol{\kappa}\cdot\nabla w)\cdot\nabla v\,dx + \int_\Omega \left(\boldsymbol{\beta}\cdot\nabla w\right)\,v\,dx + \int_\Omega \sigma w\,v\,dx$$

the variational problem follows:

Weak form of the CDR problem

Given $u_g \in H^1(\Omega)$, $u_g = g$ on Γ_D, find $u = u_0 + u_g$, such that $u_0 \in H^1_{D0}(\Omega)$ satisfies

$$a(u_0, v) = \int_\Omega f\,v\,dx - \int_{\Gamma_N} H\,v\,ds - a(u_g, v) \quad \forall v \in H^1_{D0}(\Omega) \tag{9.2}$$

9.2.1 Well-posedness of the continous CDR problem

Now, we embark in the task of showing the following theorem that establishes under which conditions the problem is well posed:

Theorem 9.2.1 *Let consider the weak form of the CDR problem (Equation 9.2) and assume:*

- $\boldsymbol{\kappa} \in [L^\infty(\Omega)]^{d\times d}$, $\boldsymbol{\beta} \in [L^\infty(\Omega)]^d$ *and* $\sigma \in L^\infty(\Omega)$*;*
- f, g, H, Γ_N *and* Γ_D *are regular enough for the right-hand-side of Equation 9.2 to be a bounded linear operator on* $H^1_{D0}(\Omega)$*;*

[1] The data g and Γ_D must be regular enough for a function $u_g \in H^1(\Omega)$, that satisfies $u_g = g$ on Γ_D (in the sense of traces) to exist. If such function exists, we say, it belongs to a trace space.

- *Assume further that*

$$\nabla \cdot \boldsymbol{\beta} \in L^\infty(\Omega), \qquad \boldsymbol{\beta}(\mathbf{x}) \cdot \check{\mathbf{n}}(\mathbf{x}) > 0 \quad \text{a.e. on } \Gamma_N \tag{9.3}$$

$$\boldsymbol{\xi} \cdot (\boldsymbol{\kappa}(\mathbf{x})\boldsymbol{\xi}) \geq \kappa_0 \, |\boldsymbol{\xi}|^2 \qquad \forall \boldsymbol{\xi} \in \mathbb{R}^d; \text{ a.e. in } \Omega \tag{9.4}$$

$$\sigma(\mathbf{x}) - \frac{1}{2}\nabla \cdot \boldsymbol{\beta}(\mathbf{x}) \geq s_{\min} \qquad \text{a.e. in } \Omega \tag{9.5}$$

where κ_0 and $s_{\min}$ are strictly positive constants.

Then, the CDR problem is well-posed.

The proof involves several steps, namely, we need to prove that the billinear form $a(\cdot,\cdot)$ is continuous and strongly coercive, the linear form in the right-hand-side is continuous, and the space $H^1_{D0}(\Omega)$ is in fact a closed subspace of $H^1(\Omega)$, and thus, Hilbert.

Continuity and coercivity of the bilinear form

For the continuity of $a(\cdot,\cdot)$, we first define

$$c_\kappa = \|\boldsymbol{\kappa}\|_{L^\infty(\Omega)}, \quad c_\beta = \|\boldsymbol{\beta}\|_{L^\infty(\Omega)}, \quad c_\sigma = \|\sigma\|_{L^\infty(\Omega)}$$

and the relations

$$\|\boldsymbol{\kappa}(\mathbf{x})\|_2 \leq c_\kappa \, d, \quad \|\boldsymbol{\beta}(\mathbf{x})\|_2 \leq c_\beta \sqrt{d} \tag{9.6}$$

where d is the number of spatial dimensions[2]. Proceeding term by term we obtain

$$\begin{aligned}
\int_\Omega (\boldsymbol{\kappa} \cdot \nabla u) \cdot \nabla v \, dx &\leq \int_\Omega |(\boldsymbol{\kappa} \cdot \nabla u) \cdot \nabla v| \, dx \\
&\leq \int_\Omega \|\boldsymbol{\kappa} \cdot \nabla u\|_2 \|\nabla v\|_2 \, dx \\
&\leq \int_\Omega \|\boldsymbol{\kappa}\|_2 \|\nabla u\|_2 \|\nabla v\|_2 \, dx \\
&\leq c_\kappa \, d \int_\Omega \|\nabla u\|_2 \|\nabla v\|_2 \, dx \\
&\leq c_\kappa \, d \, \|\nabla u\|_{L^2(\Omega)} \, \|\nabla v\|_{L^2(\Omega)} \\
&\leq c_\kappa \, d \, \|u\|_{H^1(\Omega)} \, \|v\|_{H^1(\Omega)}
\end{aligned}$$

$$\begin{aligned}
\int_\Omega \left(\boldsymbol{\beta} \cdot \nabla u\right) \, v \, dx &\leq \int_\Omega | \left(\boldsymbol{\beta} \cdot \nabla u\right) \, v| \, dx \\
&\leq \int_\Omega \|\boldsymbol{\beta}\|_2 \|\nabla u\|_2 \, |v| \, dx \\
&\leq c_\beta \sqrt{d} \int_\Omega \|\nabla u\|_2 \, |v| \, dx \\
&\leq c_\beta \sqrt{d} \, \|\nabla u\|_{L^2(\Omega)} \, \|v\|_{L^2(\Omega)} \\
&\leq c_\beta \sqrt{d} \, \|u\|_{H^1(\Omega)} \, \|v\|_{H^1(\Omega)}
\end{aligned}$$

[2] The relations given in Equation 9.6 yield from:

$$\|\boldsymbol{\kappa}(\mathbf{x})\|_2 = \sup_{\mathbf{w} \in \mathbb{R}^d} \frac{\|\boldsymbol{\kappa}(\mathbf{x}) \cdot \mathbf{w}\|_2}{\|\mathbf{w}\|_2}$$

but,

$$\|\boldsymbol{\kappa}(\mathbf{x}) \cdot \mathbf{w}\|_2 = \left[\sum_{i=1}^{d} \left(\sum_{j=1}^{d} \boldsymbol{\kappa}_{ij}(\mathbf{x}) w_j\right)^2\right]^{\frac{1}{2}} \leq c_\kappa \, d \, \|\mathbf{w}\|_2$$

Similarly

$$\|\boldsymbol{\beta}(\mathbf{x})\|_2 = \left[\sum_{i=1}^{d} \boldsymbol{\beta}_i(\mathbf{x})^2\right]^{\frac{1}{2}} \leq c_\beta \sqrt{d}$$

$$
\begin{aligned}
\int_\Omega \sigma\, u\, v\, dx &\leq \int_\Omega |\sigma\, u\, v|\, dx \\
&\leq \int_\Omega |\sigma|\,|u|\,|v|\, dx \\
&\leq c_\sigma \int_\Omega |u|\,|v|\, dx \\
&\leq c_\sigma\, \|u\|_{L^2(\Omega)}\, \|v\|_{L^2(\Omega)} \\
&\leq c_\sigma\, \|u\|_{H^1(\Omega)}\, \|v\|_{H^1(\Omega)}
\end{aligned}
$$

thus giving

$$
|a(u,v)| \quad \leq \quad \underbrace{\left(c_\kappa\, d + c_\beta \sqrt{d} + c_\sigma\right)}_{N_a} \|u\|_{H^1(\Omega)}\, \|v\|_{H^1(\Omega)}
$$

As for the coercivity of $a(\cdot,\cdot)$, let us first consider the following trick

$$
\frac{1}{2}\nabla \cdot \left(u^2 \boldsymbol{\beta}\right) = \frac{1}{2}\, u^2 \left(\nabla \cdot \boldsymbol{\beta}\right) + \left(\boldsymbol{\beta} \cdot \nabla u\right)\, u
$$

and consider $u \in H^1_{D0}$, such that $u|_{\Gamma_D} = 0$,

$$
\begin{aligned}
a(u,u) &= \int_\Omega \left[(\boldsymbol{\kappa} \cdot \nabla u) \cdot \nabla u + \left(\boldsymbol{\beta} \cdot \nabla u\right)\, u + \sigma\, u^2\right] dx \\
&\geq \int_\Omega \left[\kappa_0 \|\nabla u\|_2^2 + \left(\sigma - \frac{1}{2}\nabla \cdot \boldsymbol{\beta}\right) u^2 + \frac{1}{2}\nabla \cdot \left(u^2 \boldsymbol{\beta}\right)\right] dx \\
&\geq \int_\Omega \left[\kappa_0 \|\nabla u\|_2^2 + s_{\min}\, u^2\right] dx + \frac{1}{2}\int_\Omega \nabla \cdot \left(u^2 \boldsymbol{\beta}\right) dx \\
&\geq \min\{\kappa_0, s_{\min}\} \int_\Omega \left[\|\nabla u\|_2^2 + u^2\right] dx + \frac{1}{2}\int_{\Gamma_N} u^2 \boldsymbol{\beta} \cdot \check{\mathbf{n}}\, ds \\
&\geq \min\{\kappa_0, s_{\min}\} \int_\Omega \left(u^2 + \nabla u \cdot \nabla u\right) dx \\
&= \alpha\, \|u\|^2_{H^1(\Omega)}
\end{aligned}
$$

where $\alpha = \min\{\kappa_0, s_{\min}\}$[3].

Linear form

Lemma 9.2.2 ***(Boundedness of linear form)*** *Let $\ell : H^1_{D0}(\Omega) \longrightarrow \mathbb{R}$ be defined by*

$$
\ell(v) = \int_\Omega f\, v\, dx - \int_{\Gamma_N} H\, v\, ds - a(u_g, v) \qquad (9.7)
$$

Under reasonable regularity assumptions on Γ_D, if $H \in L^2(\Gamma_N)$ and $g \in H^{\frac{1}{2}}(\Gamma_D) \subset L^2(\Gamma_D)$, then $\ell(v)$ is bounded in H^1_{D0}.

[3] Notice that, thanks to the reaction term and Equation 9.5 ($s_{\min} > 0$), we didn't need to use the Poincaré-Friedrichs inequality, as in Chapter 4. However, if $s_{\min}$ happens to be zero, the proof of well-posedness can be done by invoking the inequality Theorem 4.1.1.

Characterization of H^1_{D0}

The task is to characterize $H^1_{D0}(\Omega) = \{v \in H^1(\Omega),\ v|_{\Gamma_D} = 0\}$. The first thing to notice, is that, although we already have a trace operator, say

$$\gamma : H^1(\Omega) \longrightarrow L^2(\partial\Omega)$$

we also need to introduce a restriction operator

$$r_D : L^2(\partial\Omega) \longrightarrow L^2(\Gamma_D)$$

such that, the value of any function $v \in H^1(\Omega)$ on Γ_D is thus given by[4] $j(v) = r_D(\gamma(v))$. The space H^1_{D0} can now be defined as

$$H^1_{D0}(\Omega) = \{v \in H^1(\Omega),\ j(v) = 0 \text{ on } \Gamma_D\}$$

We have to see that $H^1_{D0}(\Omega)$ is closed. Let $D = \{0\}$ (a closed set) and consider a sequence $v_n \longrightarrow v$, in which $v_n \in j^{-1}(D)$, and we will show that also $v \in j^{-1}(D)$. Since j is continuous, we have a convergent sequence

$$\lim_{n\to\infty} j(v_n) = j(\lim_{n\to\infty} v_n) = j(v) = 0$$

But we know 0 is in the range of j (by definition of range), hence, v is in the domain of j (the preimage). So the preimage, $H^1_{D0}(\Omega)$, is also closed, since it contains all its limit points.

9.2.2 Ritz-Galerkin aproximation of the CDR problem

To conclude this section, let us consider the Galerkin treatment of the CDR problem. Take a discrete space $V_h \subset H^1(\Omega)$ associated to a partition $\mathcal{T}_h$ of Ω, equipped with finite elements. Consider the linear set

$$V_{hg} = \{v \in H^1(\Omega),\ v|_{\Gamma_D} = g\}$$

Noticing that V_{h0} is a subspace of V_h. The discrete variational formulation is:

Discrete (Ritz-Galerking) form of the CDR problem

Find $u_h \in V_{hg}$ such that

$$a(u_h, v_h) = \int_\Omega f\, v_h\, dx - \int_{\Gamma_N} H\, v_h\, ds \tag{9.8}$$

$\forall v_h \in V_{h0}$.

Two comments are in order here:

- In practice, we do not build a function u_g to deal with the Dirichlet boundary conditions, but, we set to zero all the degrees of freedom associated to points located on Γ_D;
- The function u_h actually coincides with some interpolation $\mathcal{I}_h g$ of the boundary data g on Γ_D.

---

[4] The linear operator $j : H^1(\Omega) \to L^2(\Gamma_D)$, being the composition of continuous linear operators, is also continuous.

9.3 Assignment 6: Boundary layers in the CDR problem

We consider the general CDR problem with Dirichlet and Neumann data on a rectangular domain $[0, L] \times [0, 1]$

$$\begin{cases} -\nabla \cdot (\kappa \nabla u) + \boldsymbol{\beta} \cdot \nabla u + \sigma\, u &= f \quad \text{in } \Omega \\ u &= g \quad \text{on } \Gamma_D \\ -\kappa \nabla u \cdot \check{\mathbf{n}} &= 0 \quad \text{on } \Gamma_N \end{cases}$$

Depending on problem at hand (choice of BCs, parameters, etc.), the solution to the CDR equation may exhibit *boundary layers*, that are regions of the domain in which the solution u changes abruptly. These layer can appear at internal interfaces, as is the case of a reaction boundary layer, or near the boundaries of the domain, as we will see in the convection dominated case.

Reaction boundary layer at internal interface

Consider the domain $\Omega = [0, L] \times [0, 1]$ with a vertical internal interface Γ located at $x = L/2$, which divides the domain into subdomains Ω_1 and Ω_2. Consider the Diffusion-Reaction equation

$$\begin{aligned} -\kappa\, \nabla^2 u + \sigma\, u &= f \quad \text{in } \Omega \\ \nabla u \cdot \check{\mathbf{n}} &= 0 \quad \text{in } \partial\Omega \end{aligned}$$

where κ is the scalar diffusion coefficient and the function f in the right hand side is given by

$$f = \begin{cases} 0 & \text{if } x \in \Omega_1 \\ \sigma & \text{if } x \in \Omega_2 \end{cases}$$

A boundary layer is observed in this problem around the interface Γ, its characteristic length scale being

$$\delta_r \sim \sqrt{\frac{\kappa}{\sigma}}$$

This can be seen by solving a simple one dimensional problem whose solution is

$$u(x) = \begin{cases} 1 - \dfrac{1}{2} e^{-\dfrac{x-1}{\delta_r}} & \text{if } x \geq 1 \\ \dfrac{1}{2} e^{\dfrac{x-1}{\delta_r}} & \text{if } x < 1 \end{cases}$$

Notice that the solution has derivatives that actually go to zero only for $x \longrightarrow \pm\infty$. This case is illustrated in Figure 9.1.

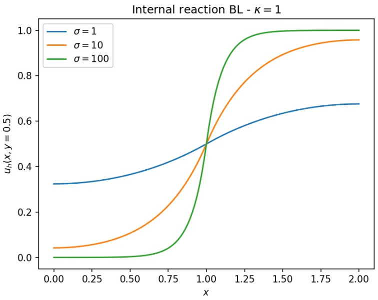


Figure 9.1: Solution of the Diffusion-Reaction problem along the x axis for different values of σ.

Advection boundary layer at exit

Consider the computational domain $\Omega = [0, L] \times [0, 1]$ and the Diffusion-Convection equation with homogeneous right hand side

$$
\begin{aligned}
-\kappa \nabla^2 u + \boldsymbol{\beta} \cdot \nabla u &= 0 \quad \text{in } \Omega \\
\frac{\partial u}{\partial y}(x, 0) &= 0, \\
\frac{\partial u}{\partial y}(x, 1) &= 0, \\
u(0, y) &= 0, \\
u(L, y) &= 1
\end{aligned}
$$

A boundary layer is observed in this problem near the exit wall, its characteristic length scale being now

$$\delta_c \sim \frac{\kappa}{\|\boldsymbol{\beta}\|}$$

Again, this can be seen by solving the corresponding one dimensional problem with $\boldsymbol{\beta} = [1, 0]^\intercal$ whose solution is

$$u(x) = \frac{e^{Pe \frac{x}{L}} - 1}{e^{Pe} - 1}$$

which depends on the Peclet number that is defined as $Pe = \|\boldsymbol{\beta}\| L/\kappa$. This case is illustrated in Figure 9.2.

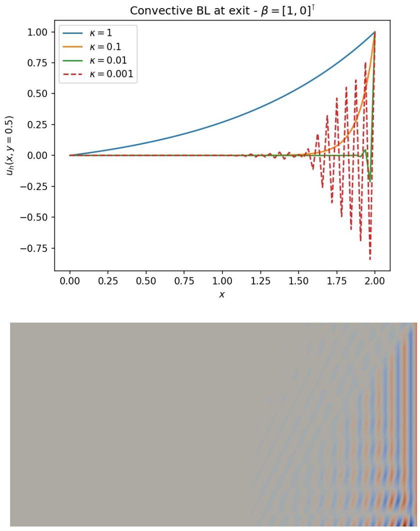


Figure 9.2: Solution of the Diffusion-Convection problem along the x axis for different values of κ (top) and 2D solution corresponding to $\kappa = 0.001$ (bottom).

9.3.1 Petrov-Galerkin (SUPG) Stabilization: A Remedy

Consider the following stabilized (consistent) formulation:

$$a(u_h, v_h) + r(u_h, v_h) = \ell(v_h)$$

where $a(\cdot, \cdot)$ and $\ell(\cdot)$ are the bilinear and linear forms respectively of the original problem, and the perturbation term r is given by

$$r(u_h, v_h) = \sum_{K \in \mathcal{T}_h} \int_K \mathcal{P}(v_h)\, \tau_K\, \mathcal{R}(u_h)\, dx$$

where $\mathcal{R}(u) = -\nabla \cdot (\kappa \nabla u) + \boldsymbol{\beta} \cdot \nabla u + \sigma\, u - f$ is the residual of the original differential equation, the elementwise stabilization parameter τ_K is taken as

$$\tau_K = \left[\frac{4\,\kappa}{h_K^2} + \frac{2\,\|\boldsymbol{\beta}\|}{h_K} + \sigma \right]^{-1}$$

and the perturbed test function $\mathcal{P}(v)$ is

$$\mathcal{P}(v) = \boldsymbol{\beta} \cdot \nabla v$$

Other choices for $\mathcal{P}(v)$ lead to different methods, such as the Galerkin Least Square (GLS) or the Algebraic Subgrid Scale (ASGS) methods (see Codina, 1997, CMAME).

Consider the script `solve_CDR.py`

1. **Internal reaction Layer**
   - (a) The first task is to solve the diffusion-reaction equation setting $L = 2$, $\kappa = 1$ and values of σ ranging from 1 to 10000. Use a structured type grid (`flag_unstructured = False`).
   - (b) Compare the solution to the analytical one for different grid sizes h and the values of σ of the previous item.
   - (c) Consider unstructured grids (`flag_unstructured = True`). Implement the necessary modifications so as to identify the internal interface located at $x = L/2$ and perform a grid refinement around it (similar to what is implemented under `flag_refine_exit`.
2. **Convection layer at exit**
   - (a) Consider the velocity field $\boldsymbol{\beta} = [1, 0]^\intercal$ and a mesh with characteristic size of elements equal to $1/32$. Compute the solution with values of κ ranging from 1 to 0.001. For each case, plot the solution and observe the boundary layer.
   - (b) Setting $\kappa = 10^{-2}$, perform a **uniform mesh refinement** and observe the numerical solution improving.
   - (c) Setting $\kappa = 10^{-2}$ perform a **local mesh refinement near the exit**, so as to better capture the boundary layer.
3. **Stabilization**
   - (a) Complete the implementation of the Petrov-Galerkin stabilization;
   - (b) Recompute the previous solutions and compare accuracy of results and convergence rates with the unstabilized case.
4. **Optional: Tensorial diffusion coefficient**
   Introduce the necessary lines to consider the case of a tensorial diffusion coefficient. Consider the domain $\Omega = [0, 1]^2$, $\boldsymbol{\beta} = [0, 0]^\intercal$ and $\sigma = 0$. In order to build an analytical solution we use the following tensor

$$\boldsymbol{\kappa} = \begin{bmatrix} (x+1)^2 + y^2 & \sin(x\,y) \\ \sin(x\,y) & (x+1)^2 \end{bmatrix}$$

   and build a source term $f(x, y)$ such that the solution is

$$u(x, y) = x^3\,y^4 + x^2 + \sin(x\,y)\,\cos(y)$$

   To solve this problem, impose the exact solution as Dirichlet condition on the whole boundary.

9.4 Linear elasticity problem: Elastostatic

Let us go back to the prototypical example of an elliptic problem with vector unknown, namely, the case of linear elasticity, whose differential form is:

$$\begin{cases} -\nabla\cdot\sigma(\mathbf{u}) & = & \mathbf{f} & \text{in } \Omega \\ \mathbf{u} & = & \mathbf{g} & \text{on } \Gamma_{\mathbf{u}} \\ \sigma\cdot\check{\mathbf{n}} & = & \mathscr{F} & \text{on } \Gamma_{\mathscr{F}} \end{cases} \tag{9.9}$$

where $\boldsymbol{\sigma} = 2\,\mu\boldsymbol{\varepsilon}(\mathbf{u}) + \lambda\,(\nabla\cdot\mathbf{u})\,\mathbf{I}$[5]. The variational formulation can be obtained from this PDE, multiplying by a vector test function $\mathbf{v}$ and integrating by parts. However, it can also be established as a fundamental principle known as **Principle of Virtual Work**

Principle of Virtual Work

Consider a deformable body subject to external body forces $\mathbf{f}$ and surface force distributions $\mathscr{F}$. Then, *the virtual work* of internal stresses equals *the virtual work* of the applied forces. This must hold for all virtual vector fields that are kinematically admissible variations of the body motion, i.e.,

$$\text{Total Internal Virtual Work} \triangleq \int_\Omega \sigma(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v})\,dx = \int_\Omega \mathbf{f}\cdot\mathbf{v}\,dx + \int_{\Gamma_{\mathscr{F}}} \mathscr{F}\cdot\mathbf{v}\,ds \triangleq \text{Total External Virtual Work}$$

By defining the linear set of kinematically admissible motions

$$V_{Dg} = \{\mathbf{v}\in\left[H^1(\Omega)\right]^d,\ \mathbf{v}=\mathbf{g} \text{ on } \Gamma_D\}$$

The variational problem reads: Find $\mathbf{u}\in V_{Dg}$ such that

$$\underbrace{\int_\Omega \left[2\,\mu\boldsymbol{\varepsilon}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v}) + \lambda\,(\nabla\cdot\mathbf{u})\,(\nabla\cdot\mathbf{v})\right]\,dx}_{a(\mathbf{u},\mathbf{v})} = \underbrace{\int_\Omega \mathbf{f}\cdot\mathbf{v}\,dx + \int_{\Gamma_{\mathscr{F}}} \mathscr{F}\cdot\mathbf{v}\,ds}_{\ell(\mathbf{v})} \tag{9.10}$$

$\forall\mathbf{v}\in V_{D0}$.

Theorem 9.4.1 ***(Well-posedness of the Elastostatic problem)*** *Let $\Omega\subset\mathbb{R}^d$ be a regular domain with meas$(\Gamma_D)>0$. Assume further that:*

- $\mathbf{f}\in\left[L^2(\Omega)\right]^d$
- $\mathscr{F}\in\left[L^2(\Gamma_{\mathscr{F}})\right]^d$
- $0\le\mu_{\min}<\mu<\mu_{\max}$ *and* $\lambda>-2\mu$

Then, there exists a unique $\mathbf{u}$ *that satisfies Equation 9.10 and it depends continuously on the data, i.e.,*

$$\|\mathbf{u}\|_{H^1(\Omega)} \le C\,\left(\|\mathbf{f}\|_{L^2(\Omega)} + \|\mathscr{F}\|_{L^2(\Gamma_{\mathscr{F}})}\right)$$

[5] It is worth also to recall the equations using index notation, which may ease some calculations:

$$\sigma_{ij} = 2\mu\,\varepsilon_{ij} + \delta_{ij}\,u_{k,k}$$

where $\varepsilon_{ij} = \frac{1}{2}(u_{i,j}+u_{j,i})$. The equilibrium equation becomes:

$$\sum_{j=1}^d \frac{\partial\sigma_{ij}}{\partial x_j} = \mathbf{f}_i$$

and the boundary conditions $u_i = g_i$ on $\Gamma_{\mathbf{u}}$ and

$$\sum_{j=1}^d \sigma_{ij}\check{\mathbf{n}}_j = \mathscr{F}_i$$

on $\Gamma_{\mathscr{F}}$. Recall, that using Einstein's convention, the sum signs in front of the equations can be omitted, implying summation over repeated indexes.

As in the previous case, the proof involves showing that the bilinear form is bounded and coercive. For the former we have[6]

$$\begin{aligned}
a(\mathbf{u},\mathbf{v}) &= \int_\Omega [2\,\mu\boldsymbol{\varepsilon}(\mathbf{u}):\boldsymbol{\varepsilon}(\mathbf{v}) + \lambda\,(\nabla\cdot\mathbf{u})\,(\nabla\cdot\mathbf{v})]\,dx \\
&\leq \lambda_{\max}\,(\nabla\cdot\mathbf{u},\nabla\cdot\mathbf{v})_{L^2(\Omega)} + 2\mu_{\max}\,(\boldsymbol{\varepsilon}(\mathbf{u}),\boldsymbol{\varepsilon}(\mathbf{v}))_{L^2(\Omega)} \\
&\leq \lambda_{\max}\,\|\nabla\cdot\mathbf{u}\|_{L^2(\Omega)}\|\nabla\cdot\mathbf{v}\|_{L^2(\Omega)} + 2\mu_{\max}\|\boldsymbol{\varepsilon}(\mathbf{u})\|_{L^2(\Omega)}\|\boldsymbol{\varepsilon}(\mathbf{v})\|_{L^2(\Omega)} \\
&\leq C\,\max\{\lambda_{\max},\mu_{\max}\}\|\nabla\mathbf{u}\|_{L^2(\Omega)}\|\nabla\mathbf{v}\|_{L^2(\Omega)} \\
&\leq C\,\max\{\lambda_{\max},\mu_{\max}\}\|\mathbf{u}\|_{H^1(\Omega)}\|\mathbf{v}\|_{H^1(\Omega)}
\end{aligned}$$

As for coercivity, the situation is different. We need an inequality similar to Poincaré-Friedrichs (see Theorem 4.1.1), but one that involves all first order derivatives of the vector field that appear in the definition of $\boldsymbol{\varepsilon}(\mathbf{u})$. This is given by:

Theorem 9.4.2 *(**Korn's Inequality**) Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary, $\Gamma_D \subset \partial\Omega$, meas$(\Gamma_D) \neq 0$. Then, there exists $C_K > 0$ (depending only on Ω and Γ_D) such that, for any $\mathbf{u} \in \left[H^1_{D0}\right]^d$*

$$\int_\Omega \boldsymbol{\varepsilon}(\mathbf{u}):\boldsymbol{\varepsilon}(\mathbf{u})\,dx \geq C_K\,\|\mathbf{u}\|^2_{H^1(\Omega)} \tag{9.11}$$

In such case, we have:

$$\begin{aligned}
a(\mathbf{u},\mathbf{u}) &= \int_\Omega \left[2\,\mu\boldsymbol{\varepsilon}(\mathbf{u}):\boldsymbol{\varepsilon}(\mathbf{u}) + \lambda\,(\nabla\cdot\mathbf{u})^2\right]dx \\
&\geq \int_\Omega 2\,\mu\boldsymbol{\varepsilon}(\mathbf{u}):\boldsymbol{\varepsilon}(\mathbf{u})\,dx \\
&\geq 2\mu_{\min}C_K\|\mathbf{u}\|^2_{H^1(\Omega)}
\end{aligned}$$

It remains to apply Lax-Milgram Theorem 3.4.1.

9.4.1 Finite element spaces for vector-valued functions

We have to extend the ideas of finite element spaces we have been using for scalar problems to the case of vector-valued functions, as required for the elasticity problem and the (Navier-)Stokes problem. The simplest thing to do is to use the canonical basis of $\mathbb{R}^d$. To that end we define as degrees of freedom the values of the x_1 and x_2-components of the displacement field $\mathbf{u}$. For triangular linear elements, e.g.,

$$\begin{aligned}
\sigma_1(\mathbf{u}) &= u_1(\mathbf{x}^1) \\
\sigma_2(\mathbf{u}) &= u_2(\mathbf{x}^1) \\
\sigma_3(\mathbf{u}) &= u_1(\mathbf{x}^2) \\
\sigma_4(\mathbf{u}) &= u_2(\mathbf{x}^2) \\
\sigma_5(\mathbf{u}) &= u_1(\mathbf{x}^3) \\
\sigma_6(\mathbf{u}) &= u_2(\mathbf{x}^3)
\end{aligned}$$

[6] From the second to the third step, we have used the fact that the L^2 norms are taken over various combinations of first order derivatives.

As we have done many times before, in order to build the basis functions, we write $\sigma_i(\mathbf{N}^j) = \delta_{ij}$ and end up with:

$$\mathbf{N}^1 = \psi^1(\mathbf{x}) \begin{pmatrix}1\\0\end{pmatrix},\ \mathbf{N}^2 = \psi^1(\mathbf{x}) \begin{pmatrix}0\\1\end{pmatrix},\ \mathbf{N}^3 = \psi^2(\mathbf{x}) \begin{pmatrix}1\\0\end{pmatrix},\ \mathbf{N}^4 = \psi^2(\mathbf{x}) \begin{pmatrix}0\\1\end{pmatrix},\ \ldots$$

i.e.,

$$\mathbf{N}^1 = \psi^1(\mathbf{x})\,\check{\mathbf{e}}^1,\ \ \mathbf{N}^2 = \psi^1(\mathbf{x})\,\check{\mathbf{e}}^2,\ \ \mathbf{N}^3 = \psi^2(\mathbf{x})\,\check{\mathbf{e}}^1,\ \ \mathbf{N}^4 = \psi^2(\mathbf{x})\,\check{\mathbf{e}}^2,\ \ \ldots$$

However, notice that this is not the only possible choice. We can also use different degrees of freedom. Typically, if the nodes are over a boundary where the normal and tangential components of the displacement field are to be specified, we may use as degrees of freedom the normal and tangential components of the displacement (u_n, u_t). This facilitates the imposition of the boundary conditions on the global system. Moreover, if u_n component is the only one that is specified while the tangential component is left free, this approach allows to handle this situation. In this **oblique** case we define the following degrees of freedom instead

$$\begin{aligned}
\sigma_1(\mathbf{u}) &= \mathbf{u}(\mathbf{x}^1)\cdot\check{\mathbf{n}}^1\\
\sigma_2(\mathbf{u}) &= \mathbf{u}(\mathbf{x}^1)\cdot\check{\boldsymbol{\tau}}^1\\
\sigma_3(\mathbf{u}) &= \mathbf{u}(\mathbf{x}^2)\cdot\check{\mathbf{n}}^2\\
\sigma_4(\mathbf{u}) &= \mathbf{u}(\mathbf{x}^2)\cdot\check{\boldsymbol{\tau}}^2\\
\sigma_5(\mathbf{u}) &= \mathbf{u}(\mathbf{x}^3)\cdot\check{\mathbf{n}}^3\\
\sigma_6(\mathbf{u}) &= \mathbf{u}(\mathbf{x}^3)\cdot\check{\boldsymbol{\tau}}^3
\end{aligned}$$

for given vectors $\check{\mathbf{n}}^r$ and $\check{\boldsymbol{\tau}}^r$ on each node. The corresponding basis is

$$\mathbf{N}^1 = \psi^1(\mathbf{x})\,\check{\mathbf{n}}^1,\ \ \mathbf{N}^2 = \psi^1(\mathbf{x})\,\check{\boldsymbol{\tau}}^1,\ \ \mathbf{N}^3 = \psi^2(\mathbf{x})\,\check{\mathbf{n}}^2,\ \ \mathbf{N}^4 = \psi^2(\mathbf{x})\,\check{\boldsymbol{\tau}}^2,\ \ \ldots$$

Now, we use these basis functions to construct the discrete system $\mathsf{A}\,\mathbf{U} = \mathbf{F}$. At the element level we need to compute the 6×6 matrix

$$A^K_{ij} = a(\mathbf{N}^j, \mathbf{N}^i),\ \ F^K_i = \ell(\mathbf{N}^i),\ \ i, j = 1, 2, \ldots, 6$$

and assembly it into the global system as we have learned. This matrix is written as

$$A^K = \begin{pmatrix}
a^K_{(1,1)} & a^K_{(1,2)} & a^K_{(1,3)}\\
a^K_{(2,1)} & a^K_{(2,2)} & a^K_{(2,3)}\\
a^K_{(3,1)} & a^K_{(3,2)} & a^K_{(3,3)}
\end{pmatrix}$$

where the blocks $a^K_{(r,s)}$, $r, s = 1, \ldots, 3$ are 2×2 matrizes corresponding to nodes (r, s) and similarly for the elemental right hand side.

An alternative equivalent approach is using the canonical basis given above and prior to final assembly, transform them using the change of basis matrix , i.e., for each 2×2 block $a^K_{(r,s)}$ and 2×1 block $f^K_{(r)}$ corresponding to nodes (r, s) we have:

$$\widetilde{a}^K_{(r,s)} = Q^T_{(r)}\,a^K_{(r,s)}\,Q_{(s)},\ \ \widetilde{f}^K_{(r)} = Q^T_{(r)}\,f^K_{(r)}$$

where $Q_{(r)} = \left[\check{\mathbf{n}}^r \mid \check{\tau}^r \right], r = 1, \dots, 3$ is the change of basis matrix corresponding to node r. Unfortunately, this can not be easily implemented in the `FEniCS` platform. To deal with oblique basis instead other approaches such as Nitsche's method [**nitsche**], in which Dirichlet conditions are enforced weakly.

nitsche

9.5 Assignment 7: 2D and 3D Elastostatic

Consider the computational domain $\Omega \subset \mathbb{R}^2$ shown in the figure.

The elastostatic problem to be solved on this domain reads:

$$\begin{cases} \sigma &=& \lambda \nabla \cdot \mathbf{u}\, \mathbf{I} + \mu (\nabla \mathbf{u} + \nabla \mathbf{u}^\intercal) \\ -\nabla \cdot \sigma &=& \mathbf{f} \ \text{ in } \Omega \\ \mathbf{u} &=& \mathbf{u}_{\text{bottom}} \ \text{ in } \Gamma_{\text{bottom}} = \{(x_1, x_2) \in \partial\Omega,\ x_2 = 0\} \\ \mathbf{u} &=& \mathbf{u}_{\text{top}} \ \text{ in } \Gamma_{\text{top}} = \{(x_1, x_2) \in \partial\Omega,\ x_2 = H\} \\ \sigma \cdot \check{\mathbf{n}} &=& \mathcal{F} \ \text{ in } \Gamma_N = \partial\Omega \setminus (\Gamma_{\text{bottom}} \cup \Gamma_{\text{top}}) \end{cases}$$

where

$$\mu = \frac{E}{2(1+\nu)}, \quad \lambda = \frac{E\nu}{(1+\nu)(1-2\nu)}$$

and

- $E = 10$ (Young's modulus), $\nu = 0.3$ (Poisson's ratio)
- $\mathbf{f} = (0, 0)^\intercal$
- $\mathcal{F} = (0, 0)^\intercal$
- $\mathbf{u}_{\text{bottom}} = (0, 0)^\intercal$
- $\mathbf{u}_{\text{top}} = (0, 0.1)^\intercal$

- Write the continuous variational formulation. Define the linear and bilinear forms and the trial and test spaces to be considered.
- Write the discrete variational formulation. Considering continuous P_1 elements, define the discrete spaces to be used.
- Implement a FEniCSx script to solve the problem above. To that end, construct a 2D finite element triangulation of the domain as shown in the figure. Notice that we need to declare now a space of vector-valued functions

```
.
V = fem.VectorFunctionSpace(msh, ("CG", 1))
.
```

- Given the deviatoric stresses

$$s = \sigma - \frac{\mathrm{tr}(\sigma)\boldsymbol{I}}{3}$$

Compute the scalar quantity known as the Von Mises stress

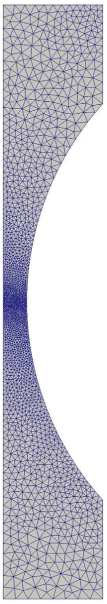


Figure 9.3: Computational domain to solve the elastostatic problem in 2D. The total height is 2.4 units, the width and height of the bottom and top parts is 0.4 and the radius of the central arc (the neck) is 1.2.

defined as the second invariant of the deviatoric stresses[7]:

$$\sigma_V = \sqrt{\frac{3}{2}\boldsymbol{s} : \boldsymbol{s}}$$

Implement in Fenics. For visualization of results, project σ_V onto a space of elementwise constant functions.

- Repeat the previous points but now assuming the only component to be specified on Γ_{top} is the vertical one, meaning that $\mathbf{u}_1$ is free (no restrictions) and $\mathbf{u}_2 = 0.1$. This can be done by adding the suffix `sub(k)` to `V` as follows:

```
.
msh = CreateMesh( ... )
dim = msh.topology.dim-1
bfacets = locate_entities_boundary(msh, dim, tag_top)
dofsy = locate_dofs_topological(V.sub(1),dim, bfacets)
bcy = dirichletbc(ScalarType(0.1), dofsy, V.sub(1))
.
```

- Implement the 3D version of the previous problem. The geometry is shown in the figure. Consider the following data:
  - $E = 10$ (Young's modulus), $\nu = 0.3$ (Poisson's ratio)
  - $\mathbf{f} = (0, 0, 0)^\intercal$
  - $\mathcal{F} = (0, 0, 0)^\intercal$
  - $\mathbf{u}_{\text{bottom}} = (0, 0, 0)^\intercal$
  - $\mathbf{u}_{\text{top}} = (0, 0.1, 0)^\intercal$

- In problems with symmetry of revolution (or axial symmetry), as in the previous example, it is convenient (cheaper) to solve the equations in the axisymmetric form. A three dimensional vector field $\mathbf{v}$ in such case has the form $(v_r,\ v_y,\ 0)^\intercal$, where r stands for the radial component (the distance to the symmetry axis) and y the vertical component[8]. It is assumed that both v_r and v_y are only functions of (r, y) and independent of the angular coordinate $\theta \in [0, 2\pi)$. Hence,

$$\nabla\mathbf{v} = \left[\begin{array}{cc|c} \multicolumn{2}{c|}{\nabla\mathbf{v}_{\text{2D}}} & \begin{array}{c}0\\0\end{array} \\ \hline 0 & 0 & \dfrac{v_r}{r} \end{array}\right], \quad \nabla\mathbf{v}_{\text{2D}} = \begin{bmatrix} v_{r,r} & v_{r,y} \\ v_{y,r} & v_{y,y} \end{bmatrix}.$$

[7] In engineering the criterion is that yielding of a ductile material begins when σ_V reaches a critical value.

Richard v. Mises

Figure 9.4: Richard Edler von Mises (Ukraine(1883)–USA(1953)).

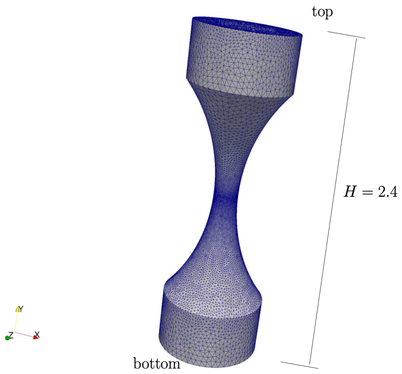


Figure 9.5: Computational domain to solve the elastostatic problem in 3D.

[8] Notice that in `FEniCSx` the radial component r (the perpendicular distance to the axis) will correspond in our problem to `x[0]` and the axial or vertical component y to `x[1]`.

Denoting by $\mathbf{v}_{\text{2D}} = (v_r,\ v_y)^\intercal$, we have[9]

$$\int_{\Omega_{\text{3D}}} 2\mu\, \boldsymbol{\varepsilon}(\mathbf{u}) : \nabla \mathbf{v}\, dx_{\text{3D}} = \\ = 2\,\pi \int_{\Omega_{\text{2D}}} 2\mu \left(\boldsymbol{\varepsilon}_{\text{2D}}(\mathbf{u}) : \nabla \mathbf{v}_{\text{2D}} + \frac{u_r\, v_r}{r^2} \right) r\, dx_{\text{2D}},$$

$$\int_{\Omega_{\text{3D}}} \lambda\, (\nabla \cdot \mathbf{u})(\nabla \cdot \mathbf{v})\, dx_{\text{3D}} = \\ = 2\,\pi \int_{\Omega_{\text{2D}}} \lambda \left(\nabla \cdot \mathbf{u}_{\text{2D}} + \frac{u_r}{r} \right) \left(\nabla \cdot \mathbf{v}_{\text{2D}} + \frac{v_r}{r} \right) r\, dx_{\text{2D}}$$

where we have used that $dx_{\text{3D}} = 2\pi r dx_{\text{2D}}$. Implement the axisymmetric formulation and compare results to the 3D full case. Take special care with the boundary conditions on the symmetry axis such that the two problems are equivalent.

[9] Note that $\boldsymbol{\varepsilon}(\mathbf{u}) : \nabla \mathbf{v} = \boldsymbol{\varepsilon}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v})$, because $\boldsymbol{\varepsilon}$ is a symmetric tensor.

STOKES PROBLEM 10

10.1 Preliminaries

So far we have been dealing with strongly coercive problems. According to Lax-Milgram theorem, strong coercivity is a *sufficient condition* for well-posedness. In the discrete case, this condition amounts to positive definiteness of the system matrix arising from the Galerkin formulation, which again, is a sufficient condition for its invertibility. The Galerkin formulation, thus inherits the well-posedness of the continuous or infinite dimensional problem. There are problems that are not coercive, but still well-posed. The necessary and sufficient conditions for well-posedness are known as the Babuška-Brezzi conditions, that we introduce for further use.

Definition 10.1.1 *(**Weak coercivity**): A bilinear form $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ is said to be weakly coercive if there exists a constant $\beta \in \mathbb{R}$ such that*

$$\inf_{0\neq u\in V} \sup_{0\neq v\in V} \frac{a(u,v)}{\|u\|_V \|v\|_V} \geq \beta > 0 \tag{10.1}$$

The following comments are in order:

- The Condition given in Equation 10.1 is also called an inf-sup condition;

- Strong coercivity (see Definition 3.2.3) implies weak coercivity, indeed,

  $$a(u,u) \geq \alpha \|u\|_V^2 \quad \forall u \in V$$

  Taking any $u \neq 0$, and dividing by $\|u\|_V$ we have

  $$\sup_{0\neq v\in V} \frac{a(u,v)}{\|v\|_V} \geq \frac{a(u,u)}{\|u\|_V} \geq \alpha \|u\|_V$$

  and taking the *infimum* over $u \in V$ we conclude

  $$\inf_{0\neq u\in V} \sup_{0\neq v\in V} \frac{a(u,v)}{\|u\|_V \|v\|_V} \geq \alpha > 0$$

- If $a(\cdot,\cdot)$ is strongly coercive over V, then

  $$\inf_{0\neq v\in V} \frac{a(v,v)}{\|v\|_V^2} = \alpha > 0 \tag{10.2}$$

  If $V_h \subset V$, then $a(\cdot,\cdot)$ will continue to be coercive[1], because the *infimum* is being taken over **a smaller set**. However, with weak coercivity this is no longer the case, in general, i.e., the following is not **necessarily** true!

  $$\inf_{0\neq u_h\in V_h} \sup_{0\neq v_h\in V_h} \frac{a(u_h,v_h)}{\|u_h\|_V \|v_h\|_V} \geq \beta > 0 \tag{10.3}$$

[1] This has already be mentioned 5.1.

because, the *supremum* is being taken over **a smaller set**[2].

- If V is finite dimensional, the bilinear form $a(\cdot,\cdot)$ is weakly coercive *iff* the associated matrix[3] A is invertible.

We state a generalization of the Lax-Milgram theorem to as necessary and sufficient conditions for well posedness of a variational problem[4]:

Theorem 10.1.1 *(Babuška theorem) Let* V *be a Hilbert space and* $a(\cdot,\cdot) : V \times V \longrightarrow \mathbb{R}$ *a bilinear form. Then, forall* $\ell \in V^*$*, the problem: Find* $u \in V$ *such that*

$$a(u,v) = \ell(v) \quad \forall v \in V \tag{10.4}$$

is well posed if and only if the following conditions hold:

$$|a(u,v)| \le N_a \, \|u\|_V \, \|v\|_V \quad \textit{(continuity)}$$

$$\inf_{0\neq u\in V} \sup_{0\neq v\in V} \frac{a(u,v)}{\|u\|_V \|v\|_V} \ge \beta > 0 \quad \textit{(weak coercivity)}$$

Now, consider the Galerkin approximation: Find $u_h \in V_h$ such that

$$a(u_h,v_h) = \ell(v_h) \quad \forall v_h \in V_h \tag{10.5}$$

and assume that the bilinear form $a(\cdot,\cdot)$ is continuous and weakly coercive in V_h, i.e., $\exists \beta_h > 0$ such that

$$\inf_{0\neq u_h\in V_h} \sup_{0\neq v_h\in V_h} \frac{a(u_h,v_h)}{\|u_h\|_V \|v_h\|_V} \ge \beta_h > 0 \tag{10.6}$$

For weakly coercive problems we have a theorem analogous to Céa's lemma on the optimality of the Galerkin approximation:

Theorem 10.1.2 *(**Optimality of the Galerkin approximation for Weakly coercive problems**): If* $a(\cdot,\cdot)$ *is continuous and weakly coercive in* V_h *with coercivity constant* β_h*, then the Galerkin approximation satisfies the following estimate*

$$\|u - u_h\|_V \le \left(1 + \frac{N_a}{\beta_h}\right) \inf_{v_h\in V_h} \|u - v_h\|_V$$

The proof is as follows:

$$\begin{aligned}
\beta_h \, \|v_h - u_h\|_V &\le \sup_{0\neq w_h\in V_h} \frac{a(v_h - u_h, w_h)}{\|w_h\|_V} \\
&= \sup_{0\neq w_h\in V_h} \frac{a(u - u_h, w_h) + a(v_h - u, w_h)}{\|w_h\|_V} \\
&= \sup_{0\neq w_h\in V_h} \frac{a(v_h - u, w_h)}{\|w_h\|_V} \\
&\le \sup_{0\neq w_h\in V_h} \frac{N_a \|v_h - u\|_V \|w_h\|_V}{\|w_h\|_V} \\
&= N_a \, \|v_h - u\|_V
\end{aligned}$$

[2] In this case, the weak coercivity of the discrete problem must be proven independently, it is not inherited from the weak coercivity of the continuous problem.

[3] If V is finite dimensional and $\{\phi^1, \dots, \phi^n\}$ is a base, recall $a_{ij} = a(\phi^j, \phi^i)$.

[4] The theorem is actually more general and allows the test and trial spaces to be different, i.e., $a(\cdot,\cdot) : V_1 \times V_2 \longrightarrow \mathbb{R}$. This can be important in several cases. One case we have seen is the Petrov-Galerkin (SUPG) stabilized formulation in the previous chapter.

but, by the triangle inequality (see Definition 2.2.1)

$$\begin{aligned} \|u-u_h\|_V &= \|u-v_h+v_h-u_h\|_V \\ &\leq \|u-v_h\|_V+\|v_h-u_h\|_V \\ &\leq \|v_h-u\|_V+\frac{N_a}{\beta_h}\|v_h-u\|_V \\ &= \left(1+\frac{N_a}{\beta_h}\right)\|u-v_h\|_V \qquad (10.7) \end{aligned}$$

A very important point to as approximability is that if we aim that

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

there must exist a constant $\beta_0 > 0$, such that $\beta_h > \beta_0$ for all h. This holds if the family of discrete space $\{V_h\}_{h>0} \subset V$ satisfies the **approximability property**:

$$\lim_{h\to 0} \inf_{v_h\in V_h} \|u-v_h\|_V = 0.$$

10.2 Motivation: Incompressible elasticity - Mixed problems

We have been considering continuous and discrete problems in the so called **primal form**, where only one unknown field is sought. In this section we consider the so called **mixed form**, where two or more unknown fields must be determined simultaneously.

First, let recall the extremal formulation for the elasticity problem, namely

Extremal form of the Elasticity problem

Find $\mathbf{u}$ such that

$$(\mathbf{EF}): \quad \mathbf{u} = \operatorname*{argmin}_{\mathbf{v}\in V_{Dg}} \mathcal{J}(\mathbf{v}) \qquad (10.8)$$

where

$$\mathcal{J}(\mathbf{v}) = \int_\Omega \left[\frac{\lambda}{2}(\nabla\cdot\mathbf{v})^2+\mu\|\boldsymbol{\varepsilon}(\mathbf{v})\|_2^2\right]dx - \int_\Omega \mathbf{f}\cdot\mathbf{v}\,dx - \int_{\Gamma_{\mathcal{F}}} \boldsymbol{\mathcal{F}}\cdot\mathbf{v}\,ds \quad (10.9)$$

As with Poisson's problem, being the bilinear form **symmetric** and **strongly coercive**, we can show that the unique solution $\mathbf{u}$ of the variational problem (**VP**): Find $\mathbf{u} \in V_{Dg}$ such that

$$a(\mathbf{u},\mathbf{v}) \triangleq \int_\Omega [\lambda\,(\nabla\cdot\mathbf{u})\,(\nabla\cdot\mathbf{v})+2\,\mu\boldsymbol{\varepsilon}(\mathbf{u}):\boldsymbol{\varepsilon}(\mathbf{v})]\,dx = \int_\Omega \mathbf{f}\cdot\mathbf{v}\,dx+\int_{\Gamma_{\mathcal{F}}}\boldsymbol{\mathcal{F}}\cdot\mathbf{v}\,ds \triangleq \ell(\mathbf{v})$$

$\forall\mathbf{v} \in V_{D0}$, is also the unique minimizer of the extremal problem (**EF**), and viceversa (see Chapter 4).

- There exist certain materials (such as rubber) that behave as incompressible, so they preserve the volume, i.e., the deformation field is divergence free[5]

$$\nabla \cdot \mathbf{u} = 0$$

- Notice that the first term of Equation 10.9 acts as a penalization of the incompressibility constraint as $\lambda \longrightarrow +\infty$ (or is very large compared to μ).

One may attempt to solve the problem of incompressible elasticity by incorporating the divergence-free constraint into the definition of the space, as we do with the Dirichlet conditions. Then, we may define the space

$$Z_{Dg} = \{\mathbf{v} \in V_{Dg},\ \nabla \cdot \mathbf{v} = 0\} \tag{10.10}$$

so, the extremal formulation consists in finding the mininum over Z_{Dg} of the following functional

$$\tilde{\mathcal{J}}(\mathbf{v}) = \int_\Omega \mu \|\boldsymbol{\varepsilon}(\mathbf{v})\|_2^2 \, dx - \int_\Omega \mathbf{f} \cdot \mathbf{v} \, dx - \int_{\Gamma_{\mathcal{F}}} \boldsymbol{\mathcal{F}} \cdot \mathbf{v} \, ds \tag{10.11}$$

since the term involving λ becomes irrelevant. The associated variational problem being: Find $\mathbf{u} \in Z_{Dg}$ such that

$$\tilde{a}(\mathbf{u}, \mathbf{v}) \triangleq \int_\Omega 2\,\mu \boldsymbol{\varepsilon}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v}) \, dx = \int_\Omega \mathbf{f} \cdot \mathbf{v} \, dx + \int_{\Gamma_{\mathcal{F}}} \boldsymbol{\mathcal{F}} \cdot \mathbf{v} \, ds \triangleq \ell(\mathbf{v}) \tag{10.12}$$

$\forall \mathbf{v} \in Z_{D0}$. This approach indeed leads to a well-posed problem, however, its practical implementation, which requires to construct finite element spaces that respect the zero-divergence constraint, is very difficult.

Instead, we introduce a new extremal formulation as:

Extremal form of the Incompressible elasticity problem

Find **u** such that

$$(\mathbf{EF}): \left\{ \begin{array}{rcl} \mathbf{u} & = & \underset{\mathbf{v} \in V_{Dg}}{\operatorname{argmin}} \tilde{\mathcal{J}}(\mathbf{v}) \\ \text{subject to} & & \nabla \cdot \mathbf{v} = 0 \end{array} \right. \tag{10.13}$$

We can reformulate this problem, introducing a Lagrange multiplier $p \in L^2(\Omega)$ and defining the Lagrangian

$$\mathcal{L}(\mathbf{u}, p) = \tilde{\mathcal{J}}(\mathbf{u}) - \underbrace{\int_\Omega p\, \nabla \cdot \mathbf{u} \, dx}_{b(\mathbf{u},p)} = \frac{1}{2}\tilde{a}(\mathbf{u}, \mathbf{u}) - \ell(\mathbf{u}) - b(\mathbf{u}, p) \tag{10.14}$$

from which a new **Mixed variational problem** emerges:

Variational form of the Incompressible elasticity problem

[5] A free divergence field indeed, implies volume preservation

$$\int_\Omega (\nabla \cdot \mathbf{u}) \, dx = \int_{\partial\Omega} \mathbf{u} \cdot \check{\mathbf{n}} \, ds = 0$$

Find $(\mathbf{u}, p) \in V_{Dg} \times L^2(\Omega)$ such that

$$\int_\Omega 2\,\mu\boldsymbol{\varepsilon}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v})\,dx - \int_\Omega p\,\nabla\cdot\mathbf{v}\,dx \;=\; \int_\Omega \mathbf{f}\cdot\mathbf{v}\,dx + \int_{\Gamma_{\mathscr{F}}} \mathscr{F}\cdot\mathbf{v}\,ds \tag{10.15}$$

$$\int_\Omega q\,\nabla\cdot\mathbf{u}\,dx \;=\; 0 \tag{10.16}$$

$\forall(\mathbf{v}, q) \in V_{D0} \times L^2(\Omega)$.

or in abstract form:

$$\begin{cases} \tilde{a}(\mathbf{u},\mathbf{v}) - b(\mathbf{v},p) = \ell(\mathbf{v}) & \forall\mathbf{v} \in V_{D0} \\ b(\mathbf{u},q) = 0 & \forall q \in L^2(\Omega) \end{cases}$$

This variational formulation is obtained by considering the optimality conditions

$$\mathscr{L}_{\mathbf{u}}(\mathbf{u},p;\mathbf{v}) = \lim_{\epsilon\to 0} \frac{\mathscr{L}(\mathbf{u}+\epsilon\,\mathbf{v},p) - \mathscr{L}(\mathbf{u},p)}{\epsilon} = 0 \quad \forall\mathbf{v} \in V_{D0} \tag{10.17}$$

$$\mathscr{L}_p(\mathbf{u},p;q) = \lim_{\epsilon\to 0} \frac{\mathscr{L}(\mathbf{u},p+\epsilon q) - \mathscr{L}(\mathbf{u},p)}{\epsilon} = 0 \quad \forall q \in L^2(\Omega) \tag{10.18}$$

This is left as an exercice.

10.2.1 Other problems in mixed form: Stokes and Darcy flows

The abstract mixed formulation (10.15)-(10.16) is the prototypical form of other mixed problems we have seen in Chapter 1, namely, the Stokes problem of an incompressible Newtonian viscous flow and the Poisson's problem in mixed form (a.k.a. Darcy's problem in the context of flow in porous media). The former corresponds to the Navier-Stokes equations Equation 1.15, but neglecting inertial terms[6]. This becomes more evident if we integrate by parts and write the differential form of the problem:

$$\begin{cases} -\nabla\cdot(2\,\mu\boldsymbol{\varepsilon}(\mathbf{u})) + \nabla p = \mathbf{f} & \text{in } \Omega \\ \nabla\cdot\mathbf{u} = 0 & \text{in } \Omega \\ \mathbf{u} = \mathbf{u}_D & \text{on } \Gamma_D \\ [-p\mathbf{I} + 2\,\mu\,\boldsymbol{\varepsilon}(\mathbf{u})]\cdot\check{\mathbf{n}} = \mathscr{F} & \text{on } \Gamma_N \end{cases} \tag{10.19}$$

Notice that the incompressibility constraint *materializes* into the momentum equation as the gradient of the **unknown** Lagrange multiplier p. Also, if μ is constant, in the incompressible case, the momentum equation simplifies to[7]

$$-\nabla\cdot(2\,\mu\boldsymbol{\varepsilon}(\mathbf{u})) + \nabla p = -\mu\nabla^2\mathbf{u} + \nabla p$$

where the Laplacian of $\mathbf{u}$ is a vector-valued function whose components are the Laplacian of each component of $\mathbf{u}$.

[6] Notice, the problem is formally identical to that of incompressible elasticity, except that the vector field $\mathbf{u}$ corresponds to a velocity instead of a displacement field and μ corresponds to the fluid viscosity.

[7] This is easily shown working with index notation:

$$[\nabla\cdot(2\boldsymbol{\varepsilon}(\mathbf{u}))]_i = [\nabla\cdot(\nabla\mathbf{u} + \nabla^{\intercal}\mathbf{u})]_i = \frac{\partial}{\partial x_j}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) =$$

$$= \frac{\partial^2 u_i}{\partial x_j \partial x_j} + \frac{\partial}{\partial x_i}\left(\frac{\partial u_j}{\partial x_j}\right) = \nabla^2 u_i + \frac{\partial}{\partial x_i}(\nabla\cdot\mathbf{u}) = \nabla^2 u_i$$

For Darcy's flow, we take as the flux variable $\boldsymbol{J}$ and the scalar variable p (the pressure), that satisfy

$$\left\{\begin{array}{rcll} \nabla\cdot\boldsymbol{J} & = & f & \text{in } \Omega \\ \boldsymbol{J} & = & -\boldsymbol{\kappa}\cdot\nabla p & \text{in } \Omega \\ p & = & p_D & \text{on } \Gamma_D \\ \boldsymbol{J}\cdot\check{\mathbf{n}} & = & H & \text{on } \Gamma_N \end{array}\right. \tag{10.20}$$

By rewritting the second equation as

$$\boldsymbol{\kappa}^{-1}\cdot\boldsymbol{J} + \nabla p = 0$$

This leads us to the following variational problem[8]: Find $(\boldsymbol{J}, p) \in V_H \times L^2(\Omega)$ such that

$$\int_\Omega \boldsymbol{\kappa}^{-1}\cdot\boldsymbol{J}\cdot\mathbf{v}\,dx - \int_\Omega p\,\nabla\cdot\mathbf{v}\,dx = -\int_{\Gamma_D} p_D\,(\mathbf{v}\cdot\check{\mathbf{n}})\,ds \tag{10.21}$$

$$\int_\Omega q\,\nabla\cdot\boldsymbol{J}\,dx = \int_\Omega f\,q\,dx \tag{10.22}$$

$\forall(\mathbf{v}, q) \in V_0 \times L^2(\Omega)$.

Before proceeding we write the general abstract form of a mixed problem:

Abstract variational mixed problem

Find $(\mathbf{u}, p) \in V \times Q$ such that

$$\begin{array}{rcl} a(\mathbf{u},\mathbf{v}) - b(\mathbf{v},p) & = & \ell(\mathbf{v}) \\ b(\mathbf{u},q) & = & g(q) \end{array} \tag{10.23}$$

$\forall(\mathbf{v}, q) \in V \times Q$.

where the spaces V and Q and the bilinear and linear forms have to be interpreted accordingly in each case.

10.3 Well-posedness of the mixed variational problem

First, notice that alternatively, the problem can be written by defining a unique bilinear and linear form by substracting the two equations: Find $(\mathbf{u}, p) \in V \times Q$ such that

$$\underbrace{a(\mathbf{u},\mathbf{v}) - b(\mathbf{v},p) - b(\mathbf{u},q)}_{B((\mathbf{u},p),(\mathbf{v},q))} = \underbrace{\ell(\mathbf{v}) - g(q)}_{L((\mathbf{v},q))} \tag{10.24}$$

$\forall(\mathbf{v}, q) \in V \times Q$, so the problem is defined in the product space $W = V \times Q$ as

[8] In contrast to the primal form of Poisson's problem we have been dealing until now, in this mixed form, the Neumann condition H on Γ_N appears (**strongly**) as a restriction on the trial set V_H for $\boldsymbol{J}$ and the Dirichlet condition p_D on Γ_D appears (**naturally**) in the variational problem

Abstract variational mixed problem

Find $(\mathbf{u}, p) \in W$ such that

$$B\left((\mathbf{u}, p), (\mathbf{v}, q)\right) = L\left((\mathbf{v}, q)\right) \quad \forall (\mathbf{v}, q) \in W \tag{10.25}$$

The converse is also true, if $(\mathbf{u}, p)$ satisfies (10.25), it also satisfies (10.23). The details are left as an exercise. To prove well-posedness of this abstract problem we need the bilinear form to be bounded and weakly coercive, such that we can apply Theorem 10.1.1. This is summarized in the following theorem:

Theorem 10.3.1 *(**Well–posedness of Abstract mixed problem**): Let the bilinear form $B(\cdot,\cdot)$ be defined as in (10.24). Assume that the bilinear forms $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$ are continuous, the bilinear form $a(\cdot,\cdot)$ is strongly coercive in $Z_0 = \{\mathbf{v} \in V,\ b(\mathbf{v}, q) = 0\ \forall q \in Q\}$*

$$a(\mathbf{u}, \mathbf{u}) > \alpha \|\mathbf{u}\|_V^2$$

and the bilinear form $b(\cdot,\cdot)$ satisfies an inf–sup *condition*

$$\inf_{0 \neq q \in Q} \sup_{0 \neq \mathbf{v} \in V} \frac{b(\mathbf{u}, q)}{\|q\|_Q \|\mathbf{v}\|_V} \geq \beta > 0$$

Then, the abstract mixed problem 10.23 is well–posed.

First, we prove that B is bounded. Since, both bilinear forms a and b are continuous, i.e.,

$$|a(\mathbf{u}, \mathbf{v})| \leq N_a \|\mathbf{u}\|_V \|\mathbf{v}\|_V \quad \forall \mathbf{u}, \mathbf{v} \in V$$

$$|b(\mathbf{u}, q)| \leq N_b \|\mathbf{u}\|_V \|q\|_Q \quad \forall \mathbf{u} \in V, q \in Q$$

If we define the norm on the product space W as

$$\|(\mathbf{v}, q)\|_W = \|\mathbf{v}\|_V + \|q\|_Q \tag{10.26}$$

and proceed as usual

$$\begin{aligned}
|B((\mathbf{u}, p), (\mathbf{v}, q))| &\leq |a(\mathbf{u}, \mathbf{v})| + |b(\mathbf{u}, q)| + |b(\mathbf{v}, p)| \\
&\leq N_a \|\mathbf{u}\|_V \|\mathbf{v}\|_V + N_b \left(\|\mathbf{u}\|_V \|q\|_Q + \|\mathbf{v}\|_V \|p\|_Q\right) \\
&\leq N_a \left(\|\mathbf{u}\|_V + \|p\|_Q\right)\left(\|\mathbf{v}\|_V + \|q\|_Q\right) + N_b \left(\|\mathbf{u}\|_V \|q\|_Q + \|\mathbf{v}\|_V \|p\|_Q + \|\mathbf{u}\|_V \|\mathbf{v}\|_V + \|p\|_Q \|q\|_Q\right) \\
&\leq (N_a + N_b) \|(\mathbf{u}, p)\|_W \|(\mathbf{v}, q)\|_W
\end{aligned}$$

It remains to show that B is weakly coercive such that we can apply Theorem 10.1.1 to prove well-posedness. We will ommit the details of the proof for brevity.

10.4 Approximation of the mixed variational problem

In the previous sections we have been dealing with the continuous problem. Now, we have to tackle the problem in a discrete setting. We consider discrete spaces $V_h \subset V$ and $Q_h \subset Q$. Define $W = V \times Q$ and $W_h = V_h \times Q_h$.

Theorem 10.4.1 *(Well–posedness of Discrete abstract mixed problem): Let* $(\mathbf{u}, p) \in W$ *satisfy*

$$B\left((\mathbf{u}, p), (\mathbf{v}_h, q_h)\right) = L\left((\mathbf{v}_h, q_h)\right) \quad \forall (\mathbf{v}_h, q_h) \in W_h$$

Assume further that for all $\mathbf{v} \in Z_{h0} = \{\mathbf{v}_h \in V_h,\ b(\mathbf{v}_h, q_h) = 0\, \forall q_h \in Q_h\}$*:*

$$a(\mathbf{v}_h, \mathbf{v}_h) > \alpha \|\mathbf{v}_h\|_V^2$$

and the bilinear form $b(\cdot, \cdot)$ *satisfies an* inf–sup *condition*

$$\inf_{0 \neq q_h \in Q_h} \sup_{0 \neq \mathbf{v}_h \in V_h} \frac{b(\mathbf{u}_h, q_h)}{\|q_h\|_Q \|\mathbf{v}_h\|_V} \geq \gamma_h > 0$$

or, in a more general setting:

$$\inf_{(\mathbf{u}_h, p_h) \in W_h} \sup_{(\mathbf{v}_h, q_h) \in W_h} \frac{B((\mathbf{u}_h, p_h), (\mathbf{v}_h, q_h))}{\|(\mathbf{u}_h, p_h)\|_W \|(\mathbf{v}_h, q_h)\|_W} \geq \gamma_h > 0$$

Then, the discrete solution $(\mathbf{u}_h, p_h) \in W_h$ *of*

$$B\left((\mathbf{u}_h, p_h), (\mathbf{v}_h, q_h)\right) = L\left((\mathbf{v}_h, q_h)\right) \quad \forall (\mathbf{v}_h, q_h) \in W_h$$

exists and is unique. Further, the following estimate holds

$$\|\mathbf{u} - \mathbf{u}_h\|_V + \|p - p_h\|_Q \leq C \left(\inf_{\mathbf{v}_h \in V_h} \|\mathbf{u} - \mathbf{v}_h\|_V + \inf_{q_h \in Q_h} \|p - q_h\| \right)$$

where C depends on $N_a, N_b, \alpha, \gamma_h$.

The following comments are in order:

- The proof of the theorem is done by invoking previous theorems;
- For the particular case of the Stokes problem, the inf–sup condition takes the form:

$$\inf_{0 \neq q_h \in Q_h} \sup_{0 \neq \mathbf{v}_h \in V_h} \frac{\int_\Omega q_h \nabla \cdot \mathbf{v}_h \, dx}{\|\mathbf{v}\|_{H^1(\Omega)} \|q_h\|_{L^2(\Omega)}} > 0$$

- The spaces V_h and Q_h for velocity and pressure need to be chosen such that the inf–sup condition above is satisfied. These combinations are called **stable mixed elements**;
- To show that a given pair V_h-Q_h is stable can be a cumbersome task. However, there are several combinations of velocity and pressure

spaces that render the formulation stable;

- Equal order formulations are not stable (e.g., continuous $[P_1]^d$ elements for velocity and continuous P_1 elements for pressure). Q_h must be poorer than V_h, for instance continuous $[P_2]^d$ elements for velocity and P_1 elements for pressure, work[9];
- The alternative (more general) hypothesis allows us to perturb the variational formulation, e.g., by adding stabilization terms, that leads to non-Galerkin formulations, and also enables to use equal order formulations for velocity and pressure [10];
- If the formulation is not stable, we will observe something called *spurious pressure modes*, that typically manifests as spurious (non-physical) oscillations of the solution;

To fill in some of the gaps troughout the chapter:

Solve the following exercices

- Complete the details to obtain the abstract mixed variational problem (10.15)-(10.16) from the optimality conditions (10.17)-(10.18).
- Complete the details to obtain the strong form of Stokes problem (10.19) by integrating by parts (10.15)-(10.16).
- Complete all the details to obtain Eqs. (10.21)-(10.22) from (Equation 10.20) by multiplying the first equation by a test function q and the second equation by a test function $\mathbf{v}$. Write also the corresponding Lagrangian function.
- Show the details to prove that $(\mathbf{u}, p)$ is a solution of (10.23) $\Leftrightarrow$ is a solution of (10.25).

10.5 Assignment 8: Incompressibility

Consider the unit square domain $\Omega = [0,1]^2$ and the Navier-Poisson equation

$$-\nabla\cdot\left(\lambda\,\nabla\cdot\mathbf{u}\,\mathbf{I}+2\mu\nabla^{S}\mathbf{u}\right)=\mathbf{f},\quad \text{in}\quad \Omega$$

with $\mathbf{f} = 0$ and boundary conditions on the bottom and top walls:

$$\begin{cases} (\mathbf{u}_1,\mathbf{u}_2) & = & (0,0), & x_1\in[0,1],\ x_2=0 \\ \mathscr{F}=\boldsymbol{\sigma}\cdot\check{\mathbf{n}} & = & \left(0,-e^{-20\,(x_1-0.5)^2}\right)^{\mathsf{T}}, & x_1\in[0,1],\ x_2=1 \end{cases}$$

The remaining part of the boundary is free.

---

[9] This pair is known as the Taylor-Hood element.

[10] This will be explored in the computational assignment at the end of the chapter.

Fixing $\mu = 1$, solve the problem for $\lambda = 10^r$, $r = -2, \ldots, 6$.

1. Plot the results and compare the displacement field.

2. Obtain the deformed configuration of the domain by moving the nodes of the mesh according to the obtained displacement **u** and compute the area of the resulting domain as a function of λ.

3. Compute the quantity

$$p_\lambda = -\lambda \, \nabla \cdot \mathbf{u}$$

as a function of λ. Project p_λ onto a space of elementwise constant functions for visualization. This is by definition the pressure.

4. Introduce an additional field p to implement the resolution of the incompressible Stokes problem (consider the script `solve_stokes.py`):

$$\begin{cases} -\nabla \cdot (2\,\mu \nabla^S \mathbf{u}) + \nabla p = \mathbf{f} & \text{in } \Omega \\ \nabla \cdot \mathbf{u} = 0 & \text{in } \Omega \\ + \text{B.C.} & \text{on } \partial\Omega \end{cases}$$

whose discrete variational formulation reads: Find $\mathbf{u}_h \in V_h$ and $p_h \in Q_h$, such that

$$\int_\Omega \mu(\nabla \mathbf{u}_h + \nabla \mathbf{u}_h^\intercal) : \nabla \mathbf{v}_h \, dx - \int_\Omega p_h \, \nabla \cdot \mathbf{v}_h \, dx + \int_\Omega q_h \, \nabla \cdot \mathbf{u}_h \, dx =$$

$$= \int_\Omega \mathbf{f} \cdot \mathbf{v}_h \, dx + \int_\Omega \boldsymbol{\mathcal{F}} \cdot \mathbf{v}_h \, ds$$

$\forall (\mathbf{v}_h, q_h) \in V_{h0} \times Q_h$.

Consider the same boundary conditions as in the elasticity problem and use P_k spaces ($k = 1, 2$) for velocity and P_1 for pressure. Plot the pressure field and compare the results when using the pair P_1–P_1 and P_2–P_1. The product function space is declared in `FEniCS` as

```
P2 = ufl.VectorElement("Lagrange", msh.ufl_cell(), 2)
P1 = ufl.FiniteElement("Lagrange", msh.ufl_cell(), 1)
TH = P2 * P1
W = FunctionSpace(msh, TH)
```

5. Implement the stabilized formulation by adding the term:

$$\ldots + \sum_{K \in \mathcal{T}_h} \int_K \tau_K \, \mathcal{R}(\mathbf{u}_h, p_h) \cdot \mathcal{P}(\mathbf{v}_h, p_h) \, dx = \ldots$$

where

$$\mathcal{R}(\mathbf{u}_h, p_h) = -\nabla \cdot (2\,\mu \nabla^S \mathbf{u}_h) + \nabla p_h - \mathbf{f},$$

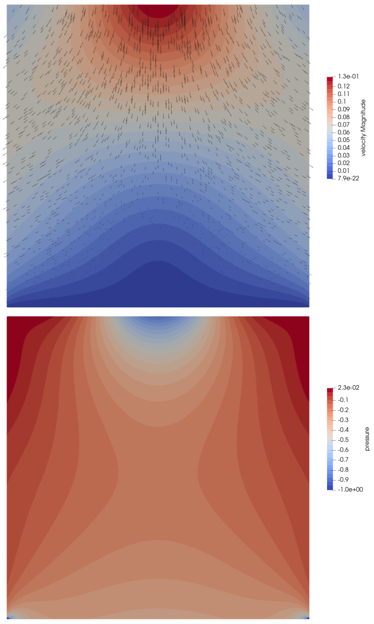


Figure 10.1: Velocity and pressure when using the Taylor-Hood (P_2–P_1) element.

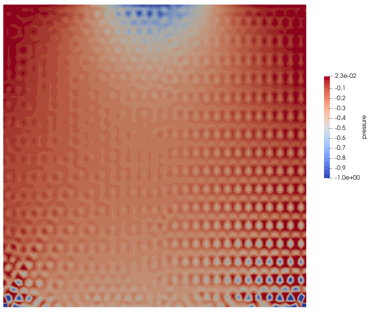


Figure 10.2: Typical spurious pressure mode when using the P_1–P_1 (unstable) element.

$$\mathscr{P}(\mathbf{v}_h, q_h) = \pm\nabla \cdot (2\,\mu\nabla^S \mathbf{v}_h) + \nabla q_h$$

and the stabilization parameter $\tau_K = \dfrac{h_K^2}{4^k \mu}$. Notice that this is not a Galerkin formulation, but is allows to circumvent the discrete inf–sup condition also referred to as the Babuška-Brezzi condition.

Bibliography

Here are the references in citation order.

[1] Gustavo C. Buscaglia ICMC-USP. *Lecture notes: Introduction to Finite Element Method*. 2020 (cited on page ii).

[2] S.C. Brenner and L.R. Scott. *The Mathematical Theory of Finite Element Methods, 3rt ed.* Ed. by Springer. Vol. 15. 2008 (cited on pages ii, 11, 16).

[3] P. Ciarlet. *Finite Element Methods (Part 1)*. Ed. by Elsevier by P. Ciarlet J.L. Lions. Vol. II. 1991 (cited on pages ii, 56).

[4] B.D. Reddy. *Introductory Functional Analysis*. Ed. by Springer. Vol. 27. 1998 (cited on pages ii, 11).

[5] A. Ern and J.L. Guermond. *Theory and Practice of Finite Elements*. Ed. by Springer. Vol. 159. 2004 (cited on pages ii, 70).

Alphabetical Index