Politecnico di Milano
School of Industrial and Information Engineering
Academic Year 2024-2025

# Numerical Analysis for Partial Differential Equations

Professor: Paola Antonietti

*Last updated: February 18, 2025*

**This document is intended for educational purposes only.**
**These are unreviewed notes and may contain errors.**

Made by Roberto Benatuil Valera

# Contents

Chapter 1

# Numerical methods for diffusion problems

---

## 1.1 Introduction: boundary value problems

Consider the following problem:

$$\begin{cases} \mathcal{L}u = f & \text{in } \Omega \\ \text{B.C.} & \text{on } \partial\Omega \end{cases} \tag{1.1}$$

where:

- $\Omega \subset \mathbb{R}^d$ is an open bounded domain, with $d = 2, 3$
- $\partial\Omega$ is the boundary of $\Omega$
- $f$ given function
- Boundary conditions (BC) to be prescribed according to $\mathcal{L}$
- $\mathcal{L}$ a 2nd order differential operator

**E.g.:** 1. $\mathcal{L}u = -\operatorname{div}(\mu\nabla u) + \mathbf{b} \cdot \nabla u + \sigma u$ (non-conservative form)

2. $\mathcal{L}u = -\operatorname{div}(\mu\nabla u) + \operatorname{div}(\mathbf{b}u) + \sigma u$ (conservative form)

with $\mu \in L^\infty(\Omega), \mu(x) \geq \mu_0 > 0$, $\mathbf{b} \in (L^\infty(\Omega))^d$, $\sigma, f \in L^2(\Omega)$.

We may propose the following boundary conditions:

$$\begin{cases} u = 0 & \text{on } \Gamma_D \\ \mu\nabla u \cdot \mathbf{n} = g & \text{on } \Gamma_N \end{cases} \tag{1.2}$$

Where $g \in L^2(\Gamma_D)$, $\partial\Omega = \Gamma_D \cup \Gamma_N$, $\Gamma_D \cap \Gamma_N = \emptyset$. We call $\Gamma_D$ the Dirichlet boundary and $\Gamma_N$ the Neumann boundary.

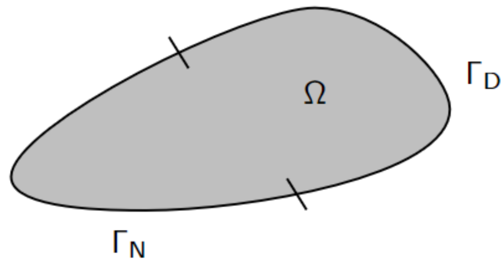


Figure 1.1: Representation of Dirichlet and Neumann boundary

So in general, let us focus on the following problem:

$$\begin{cases} -\operatorname{div}(\mu\nabla u) + \mathbf{b}\cdot\nabla u + \sigma u = f & \text{in } \Omega \\ u = 0 & \text{on } \Gamma_D \\ \mu\nabla u\cdot\mathbf{n} = g & \text{on } \Gamma_N \end{cases} \tag{1.3}$$

To solve this problem, it is useful to change the way it is formulated.

## 1.2 Weak formulation

The idea behind the weak formulation is as follows:

1. Multiply the equation by a test function $v \in V$ (we don't know $V$ yet)

$$(-\operatorname{div}(\mu\nabla u) + \mathbf{b}\cdot\nabla u + \sigma u)\, v = fv$$

2. Integrate over $\Omega$

$$\int_\Omega (-\operatorname{div}(\mu\nabla u) + \mathbf{b}\cdot\nabla u + \sigma u)\, v\, dx = \int_\Omega fv\, dx$$

3. Integrate by parts

$$\int_\Omega \mu\nabla u\cdot\nabla v\, dx - \int_{\partial\Omega} \mu\nabla u\cdot\mathbf{n}v\, ds + \int_\Omega \mathbf{b}\cdot\nabla uv\, dx + \int_\Omega \sigma uv\, dx = \int_\Omega fv\, dx$$

Note that, for the boundary terms, we have:

$$\int_{\partial\Omega} \mu \nabla u \cdot \mathbf{n} v \, ds = \int_{\Gamma_D} \mu \nabla u \cdot \mathbf{n} v \, ds + \int_{\Gamma_N} \mu \nabla u \cdot \mathbf{n} v \, ds$$

and, if we impose that $v = 0$ on $\Gamma_D$, then the first term is zero. So we have:

$$\int_{\partial\Omega} \mu \nabla u \cdot \mathbf{n} v \, ds = \int_{\Gamma_N} g v \, ds$$

where $g = \mu \nabla u \cdot \mathbf{n}$, by the Neumann boundary condition. We rewrite the equation as:

$$\underbrace{\int_{\Omega} \mu \nabla u \cdot \nabla v \, dx + \int_{\Omega} \mathbf{b} \cdot \nabla u v \, dx + \int_{\Omega} \sigma u v \, dx}_{a(u,v)} = \underbrace{\int_{\Omega} f v \, dx + \int_{\Gamma_N} g v \, ds}_{F(v)}$$

$$\implies a(u, v) = F(v) \tag{1.4}$$

We still need to define the space $V$. We will define it as:

$$V = \{v \in H^1(\Omega) \mid v = 0 \text{ on } \Gamma_D\}$$

where $H^1(\Omega)$ is the Sobolev space, formally defined as:

**Definition 1.2.1.** Let $\Omega \subset \mathbb{R}^d$ be an open bounded domain. The **Sobolev space** $H^1(\Omega)$ is defined as:

$$H^1(\Omega) = \{v \in L^2(\Omega) \mid \nabla v \in (L^2(\Omega))^d\}$$

where $\nabla v = (\partial_1 v, \ldots, \partial_d v)$. The norm in $H^1(\Omega)$ is defined as:

$$\|v\|_{H^1(\Omega)} = \left(\|v\|^2_{L^2(\Omega)} + \|\nabla v\|^2_{L^2(\Omega)}\right)^{1/2}$$

Then, notice that $V$ is just a subspace of $H^1(\Omega)$, where we impose the Dirichlet boundary condition. We will use the following notation:

$$H^1_{\Gamma_D}(\Omega) = \{v \in H^1(\Omega) \mid v = 0 \text{ on } \Gamma_D\} \tag{1.5}$$

Thus, the weak formulation of the problem, given the space $V$, is:

**Definition 1.2.2** (Abstract weak formulation)**.** Find $u \in V$ such that:

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

where $a : V \times V \to \mathbb{R}$ is a bilinear form and $F : V \to \mathbb{R}$ is a linear form, both defined as (1.4)

How do we know that the problem has a solution? We need to prove the existence and uniqueness of the solution, by using the Lax-Milgram theorem.

### 1.2.1 Lax-Milgram theorem

**Theorem 1.2.1** (Lax-Milgram)**.** *Let $V$ be a Hilbert space, $a : V \times V \to \mathbb{R}$ a bilinear form, and $F : V \to \mathbb{R}$ a linear form. Assume that:*

*(i) $V$ is a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|_V$.*

*(ii) $F \in V'$, where $V'$ is the dual space of $V$ (i.e., $F$ is bounded).*

*(iii) $a$ is continuous, i.e., $\exists M > 0$ s.t.:*

$$|a(u, v)| \leq M \|u\|_V \|v\|_V \quad \forall u, v \in V$$

*(iv) $a$ is coercive, i.e., $\exists \alpha > 0$ s.t.:*

$$a(u, u) \geq \alpha \|u\|_V^2 \quad \forall u \in V$$

*Then, there exists a unique solution $u \in V$ to the weak formulation problem. Moreover, the solution satisfies:*

$$\|u\|_V \leq \frac{1}{\alpha} \|F\|_{V'}$$

**E.g.** (**Diffusion equation**)**:** Let us consider the following problem:

$$\begin{cases} -\operatorname{div}(\mu \nabla u) = f & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega = \Gamma_D \end{cases} \tag{1.6}$$

We can write the weak formulation as:

$$a(u, v) = \int_\Omega \mu \nabla u \cdot \nabla v \, dx, \quad F(v) = \int_\Omega f v \, dx$$

with $V = H^1_{\Gamma_D}(\Omega)$. We need to prove that the Lax-Milgram theorem holds for this problem:

(i) $V$ is a Hilbert space, since $H^1_{\Gamma_D}(\Omega)$ is a closed subspace of $H^1(\Omega)$.

(ii) $F \in V'$, since:

$$|F(v)| \leq \|f\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} \leq \|f\|_{L^2(\Omega)} \|v\|_V$$

(iii) $a$ is continuous, since:

$$|a(u,v)| = \left| \int_\Omega \mu \nabla u \cdot \nabla v \, dx \right| \leq \|\mu\|_{L^\infty(\Omega)} \|\nabla u\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)} \leq M \|u\|_V \|v\|_V$$

with $M = \|\mu\|_{L^\infty(\Omega)}$.

(iv) $a$ is coercive, since:

$$a(u,u) = \int_\Omega \mu |\nabla u|^2 \, dx \geq \mu_0 \|u\|_L^2 (\Omega) \geq \alpha \|u\|_V^2 , \quad \alpha = \frac{\mu_0}{1 + C_p^2}$$

with $C_p$ the Poincaré constant. The last inequality is due to the Poincaré inequality, which we will state next.

**Theorem 1.2.2** (Poincaré inequality)**.** *If $\Gamma_D$ is a set of positive measure (in 1D, it is sufficient that is not empty), then:*

$$\exists C_p > 0 : \ \|v\|^2_{L^2(\Omega)} \leq C_p \|\nabla v\|^2_{L^2(\Omega)} \quad \forall v \in H^1_{\Gamma_D}(\Omega)$$

**Remark:** The Poincaré inequality let us prove that, for $V = H^1_{\Gamma_D}(\Omega)$:

$$\|v\|_V^2 = \|v\|^2_{L^2(\Omega)} + \|\nabla v\|^2_{L^2(\Omega)} \leq (1 + C_p^2) \|\nabla v\|^2_{L^2(\Omega)}$$

$$\implies \|\nabla v\|^2_{L^2(\Omega)} \geq \frac{1}{1 + C_p^2} \|v\|_V^2$$

**Exercise:** Show that the weak formulation of:

$$\begin{cases} -\operatorname{div}(\mu \nabla u) + \mathbf{b} \cdot \nabla u + \sigma u = f & \text{in } \Omega \\ u = 0 & \text{on } \Gamma_D \\ \mu \nabla u \cdot \mathbf{n} = g & \text{on } \Gamma_N \end{cases}$$

is well-posed (or give the conditions for it to be well-posed), by proving that the Lax-Milgram theorem holds for this problem. (See class notes P1 p.13)

## 1.3 Approximation: the Galerkin paradigm

The Galerkin paradigm is a method to approximate the solution of a problem, by using a finite-dimensional subspace of the original space $V$. The Galerkin method is based on the following idea:

**Definition 1.3.1** (Galerkin method)**.** Let $V_h \subset V$ be a finite-dimensional subspace of $V$. The Galerkin method consists in finding $u_h \in V_h$ such that:

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

where $a$ and $F$ are defined as in (1.4). Note that:

- $dim(V_h) = N_h < \infty$
- By notation, $N_h \to \infty$ as $h \to 0$

To prove well-posedness of the Galerkin method, we may use 2 approaches:

**Theorem 1.3.1** (Approach 1)**.** *The problem proposed by the Galerkin method is equivalent to the following linear system of equations:*

$$A\mathbf{u} = F$$

*where* $A \in \mathbb{R}^{N_h \times N_h}$*,* $\mathbf{u} \in \mathbb{R}^{N_h}$ *is the vector of unknowns, and* $F \in \mathbb{R}^{N_h}$

*Proof.* Let $\{\varphi_1, \dots, \varphi_{N_h}\}$ be a basis of $V_h$. Then, we can write:

$$u_h = \sum_{j=1}^{N_h} u_j \varphi_j, \quad u_j \in \mathbb{R} \; \forall j = 1, ..., N_h$$

and if we take $v_h = \varphi_i$ in the Galerkin method, we have:

$$a\left(\sum_{j=1}^{N_h} u_j \varphi_j, \varphi_i\right) = F(\varphi_i) \quad \forall i = 1, ..., N_h$$

By linearity of $a$, we have:

$$\sum_{j=1}^{N_h} u_j a(\varphi_j, \varphi_i) = F(\varphi_i) \quad \forall i = 1, ..., N_h$$

which can be written as:

$$\sum_{j=1}^{N_h} A_{ij} u_j = F_i \quad \forall i = 1, ..., N_h$$

where $A_{ij} = a(\varphi_j, \varphi_i)$ and $F_i = F(\varphi_i)$. This is the linear system of equations that we wanted to prove.

■

**Remark:** By proving that the matrix $A$ is invertible, we can prove the well-posedness of the Galerkin method. We can prove that, for the diffusion equation, the matrix $A$ is s.d.p, and thus invertible.

The second theorem is more straightforward:

**Theorem 1.3.2** (Approach 2). *The Galerkin method is well-posed, if the Lax-Milgram theorem holds for $V$.*

*Proof.* The proof is easy. We just need to observe that $V_h$ is a finite-dimensional subspace of $V$, and thus the Lax-Milgram theorem holds for $V_h$ trivially. Thus, the Galerkin method is well-posed.

■

### 1.3.1 Analysis of the Galerkin method

There are some important properties of the Galerkin method that we need to analyze:

- **Existence and uniqueness of the solution:** The Galerkin method provides a unique solution to the problem, if the Lax-Milgram theorem holds for $V$.
- **Stability:** We have the same stability properties as the Lax-Milgram theorem, i.e., the solution satisfies:
$$\|u_h\|_V \leq \frac{1}{\alpha} \|F\|_{V'}$$
- **Consistency:** the error of the Galerking method satisfies the following:
$$a(u - u_h, v_h) = 0 \quad \forall v_h \in V_h$$
This is called the **Galerkin orthogonality**.
- **Convergence:** the solution of the Galerkin method satisfies the following lemma:

**Lemma 1.3.3** (Céa lemma)**.** *Let $u \in V$ be the solution of the weak formulation problem, and $u_h \in V_h$ the solution of the Galerkin method. Then, the error satisfies:*

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

*where $M$ is the continuity constant of $a$ and $\alpha$ is the coercivity constant of $a$.*

*Proof.* We have:

$$\begin{aligned}
\alpha \|u - u_h\|_V^2 &\leq a(u - u_h, u - u_h) \\
&= a(u - u_h, u - v_h) + \underbrace{a(u - u_h, v_h - u_h)}_{=0 \text{(Galerkin orth.)}} \\
&\leq M \|u - u_h\|_V \|u - v_h\|_V \ \ \forall v_h \in V_h
\end{aligned}$$

Then:

$$\begin{aligned}
&\|u - u_h\|_V \leq \frac{M}{\alpha} \|u - v_h\|_V \ \ \forall v_h \in V_h \\
\implies \ &\|u - u_h\|_V \leq \frac{M}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|_V
\end{aligned}$$

■

**Remark:** Notice that, under the assumption of space saturation:

$$\forall v \in V, \ \lim_{h \to 0} \inf_{v_h \in V_h} \|v - v_h\|_V = 0$$

we have that the Galerkin method converges to the solution of the weak formulation problem. Indeed, the Céa lemma along with this assumption implies that:

$$\lim_{h \to 0} \|u - u_h\|_V = 0$$