An Introduction To The Finite Element Method

Solving PDEs the fancy way

Luca Solbiati

ZURU Tech Italy

May 19, 2023

Abstract

Partial differential equations have a vast amount of applications, for example they can be used to model physical phenomena, describe the value of a financial options, and to analyse the structural integrity of a building.
Even though they are powerful, their study is quite complicated, and it's often the case that no explicit formula for solving them exists.
For this reason, the field of numerical analysis has developed many numerical methods to find approximate solutions to PDEs. Among them, one of the most successful is known as the finite element method.
In this presentation we will give a brief overview of the mathematical and numerical tools behind this technique.

Contents

Introduction To Partial Differential Equations

What Are Partial Differential Equations

A partial differential equation (PDE) is a functional relation where the unknown is a function $u : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, which involves both the independent variables $x \in \Omega$, the dependent variable u and its partial derivatives:

$$F(x, u(x), \nabla u(x), ..., \nabla^{(k)} u(x)) = 0 \qquad \forall\, x \in \Omega \tag{1}$$

Where:

- $F : \Omega \times \mathbb{R}^n \times \mathbb{R}^n \times ... \times \mathbb{R}^{n^k} \to \mathbb{R}$ is the function that defines the equation
- k is called order of the equation

Elliptic PDEs: The Poisson Equation

A notable class of equations is the one of elliptic PDEs, which can be used to describe:

- The position at equilibrium of a membrane
- The temperature of a homogeneous isotropic body
- The chemical concentration of a substance

One important elliptic PDE is the Poisson equation:

$$-\Delta_x u(x) := -\mathrm{Div}_x(\nabla_x u(x)) = -(\partial^2_{x_1 x_1} u(x) + \ldots + \partial^2_{x_n x_n} u(x)) = f(x) \qquad \forall x \in \Omega \tag{2}$$

Usually accompanied by a boundary condition, i.e. a condition that the function u (or its derivatives) has to respect at the boundary of its domain ($\partial\Omega$). The most common one is the Dirichlet boundary condition:

$$u(x) = g(x) \qquad \forall x \in \partial\Omega \tag{3}$$

Strong Solutions

Consider the Poisson equation with homogeneous Dirichlet boundary conditions:

$$-\Delta u(x) = f(x) \qquad \forall\, x \in \Omega \tag{4}$$

$$u(x) = 0 \qquad \forall\, x \in \partial\Omega \tag{5}$$

For this equation, we would like, ideally, to find a classical solution that satisfies the relationship for each point.

Definition 1

$u \in \mathcal{C}^2(\Omega)$ is a classical (or strong) solution of the Poisson equation if it identically satisfies the PDE for each point $x \in \Omega$

Unfortunately, studying classical solution can be really hard, if not impossible! Indeed, there are some PDEs for which no classical solution even exists.

Weak Solutions

So, what do we do when no classical solutions to a PDE exist? Mathematicians invented the concept of weak solution.
Let us consider any $\phi \in \mathcal{C}_c^\infty(\Omega)$, also known as a *test function*. Then, if u was a solution of the Poisson equation it would hold that:

$$-\Delta u(x) = f(x) \implies -\Delta u(x) \cdot \phi(x) = f(x) \cdot \phi(x) \implies -\int_\Omega \Delta u(x) \cdot \phi(x) dx = \int_\Omega f(x) \cdot \phi(x) dx$$

$$\overset{\text{integration by parts}}{\implies} \int_\Omega \nabla u(x) \nabla \phi(x) dx - \int_{\partial\Omega} \phi(x) \nabla u \cdot \overrightarrow{n} \, d\sigma(x) = \int_\Omega f(x) \cdot \phi(x) dx$$

Since $\phi(x) = 0$ for all $x \in \partial\Omega$ we can conclude that for any classical solution it must hold:

$$\int_\Omega \nabla u(x) \nabla \phi(x) dx = \int_\Omega f(x) \cdot \phi(x) dx \qquad \forall \, \phi \in \mathcal{C}_c^\infty(\Omega) \tag{6}$$

Weak Solutions (Cont.)

With these computations we found a condition that every classical solution must satisfy. Now, an interesting question we could ask ourselves is the following: is it true that any function that satisfies condition (6) is also a classical solution? Under certain assumptions the answer is yes! Mainly, if u satisfies (6) and it's also a smooth function in $\mathcal{C}^2$, then it's also a classical solution.

Definition 2

We say that u is a weak solution of the Poisson equation if it satisfies (6)

Weak solutions are really useful, because they allow us to define a concept of solution even for interesting PDEs that don't have any strong solutions. They offer us the possibility to use more advanced and convenient mathematical tools to prove the existence and uniqueness of solutions of PDEs, while being consistent with the definition of classical solution when one such solution exists.

Weak Formulation Through The Lens Of Functional Operators

Bilinear And Linear Forms For The Weak Formulation

Recall the weak formulation of the Poisson equation:

$$\int_\Omega \nabla u(x)\nabla\phi(x)dx = \int_\Omega f(x)\cdot\phi(x)dx \qquad \forall\ \phi \in \mathcal{C}_c^\infty(\Omega)$$

Let $V := \mathcal{C}_c^2(\Omega)$ (should actually be $V := H_0^1(\Omega)$, if you're familiar with Sobolev spaces), by defining the two functions:

$$a(u,v) := \int_\Omega \nabla u(x)\nabla v(x)dx \tag{7}$$

$$F(v) := \int_\Omega f(x)\cdot v(x)dx \tag{8}$$

we could rewrite the weak formulation as:

$$a(u,v) = F(v) \qquad \forall\ v \in V \tag{9}$$

Properties Of The Weak Formulation Forms

It's easy to prove that $a : V \times V \to R$ and $F : V \to R$ are linear in their arguments. In particular both functions are continuous, so there exist two constants $\lambda > 0$ and $\gamma > 0$ such that:

$$|a(u, v)| \leq \gamma ||u||_V ||v||_V \tag{10}$$

$$|F(u)| \leq \lambda ||u||_V \tag{11}$$

where the $|| \cdot ||_V$ is the norm of the functional space V :

$$||u||_V := \sqrt{\int_\Omega |u(x)|^2 + ||\nabla u(x)||^2 dx}$$

Most importantly, a is a symmetric coercive bilinear form, i.e. there exists $\alpha > 0$ such that:

$$a(u, u) \geq \alpha ||u||_V^2 \tag{12}$$

Lax-Milgram Theorem

By formulating our weak problem through functional forms, we are able to use the Lax-Milgram theorem to prove existence and uniqueness:

Theorem 3 (Lax-Milgram)

Given a continuous coercive bilinear form $a : V \times V \to R$, and a continuous linear form $F : V \to \mathbb{R}$ in V Hilbert space, there exists a unique $u \in V$ such that:

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

Moreover, the following condition holds:

$$||u||_V \leq \frac{\gamma}{\alpha}\lambda$$

This theorem also states something about the well posedeness of our problem. Indeed, it is telling us that the final solution depends at most linearly from the data expressed by F (i.e. f, for our Poisson equation). This means that if $f_1 \simeq f_2$ then the two associated solutions will be also similar $u_1 \simeq u_2$.

The Theory Of FEM

Approximating The Space Of Solutions (Galerkin Approximation)

From a practical standpoint, all this theory hasn't really brought us any closer to finding an actual solution to the Poisson equation.
The main issue of the weak formulation is that it's using an infinite functional space, and we can barely count to 100.
What if instead of considering the whole functional space V, we only considered a finite subspace $V_h \leq V$? Then the involved quantities would be finite, and we would be able to practically solve the problem.
We shall consider the following formulation:

$$a(u_h, v_h) = F(v_h) \qquad \forall\, v_h \in V_h \tag{13}$$

By the Lax-Milgram theorem this formulation also has a unique (well posed) solution u_h. What we hope for is that by choosing V_h in an appropriate way, then u_h would be a good approximation of the actual solution u.

Cea's Lemma On The Approximated Solution

Theorem 4 (Céa's lemma)

Let u be a solution of:

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

and u_h a solution of:

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

Then:

$$||u - u_h||_V \leq \inf_{v \in V_h} ||u - v_h||_V \cdot \frac{\gamma}{\alpha} \tag{14}$$

What Cea's lemma is stating is that, up to a constant factor, the approximated solution u_h is the best approximation of u that we could find in the space V_h

Cea's Lemma Proof

Proof.

For any $v_h \in V_h$ we note that:

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

and also:

$$F(u_h) - F(v_h) = a(u_h, u_h) - a(u_h, v_h) = a(u_h, u_h - v_h)$$

Hence:

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

Using this we get:

$$\begin{aligned}\alpha ||u - u_h||_V^2 &\overset{\text{coercivity}}{\leq} a(u - u_h, u - u_h) \overset{+0}{=} a(u - u_h, u - u_h) + a(u - u_h, u_h - v_h) \\ &= a(u - u_h, u - v_h) \overset{\text{continuity}}{\leq} \gamma ||u - u_h||_V ||u - v_h||\end{aligned}$$

By dividing at the start and at the end of the inequalities we conclude that:

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

□

Matrix Notation for FEM Formulation

Thanks to Cea's lemma we know that using the finite (Galerkin) approximation we get a good enough solution, as long as V_h approximates well the space V.
Now, we reformulate the finite formulation using matrices, so that we can numerically compute the solution.
Let $< v_1, ..., v_n >= V_h$ be a basis of V_h, and let us define the matrix:

$$A := (A_{i,j})_{i,j} = (a(v_j, v_i))_{i,j} \tag{15}$$

And the vector:

$$b := (b_i)_i = (F(v_i))_i \tag{16}$$

Then, if we define $u_h := \sum_{i=1}^{n} u_i v_i$, solving the finite weak formulation is equivalent to finding the vector $\overrightarrow{u} = (u_i)_i$ such that:

$$A\overrightarrow{u} = b \tag{17}$$

Building The Approximation Space V_h

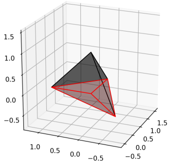
Choosing V_h Appropriately

How do we choose the approximation space in practice? Indeed, the choice of V_h influences the method in many ways. Ideally:

1. V_h should contain all the good approximations of the possible solution u (V_h should be "dense" in V)
2. V_h has a basis that allows us to compute $a(v_i, v_j)$ and $F(v_i)$ in a simple way
3. The basis of V_h should have mostly zero $a(v_i, v_j)$, so that the resulting matrix A is sparse and the linear system is easier to solve

Using Piecewise Polynomials

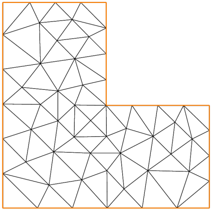
One way to satisfy the previous requirements is to use piecewise polynomials, each defined on little pieces (known as elements) of the domain Ω.
A common choice is to use piecewise linear polynomials on triangles, that have value 1 on one node, and zero on all the other nodes of the elements.



The support (i.e. the domain where the function is nonzero) of such a function is limited only to the elements surrounding the node. Hence, if we are to compute the form a against another function like this, the result is zero unless they share a common element node.

Building V_h By Meshing Ω

The main idea of the FEM is to divide the domain Ω, using a triangle mesh, in order to define a basis of these piecewise polynomials. There will be as many polynomials as there are nodes, so we can index the functions in the same way we index the nodes.



By choosing the basis functions to have value one only on their node, and zero on the others, we can directly infer that the value of $u_h = \sum_{i=1}^{n} u_i v_i$ on the i-th node is given by the $i - th$ coefficient that defines u_h, i.e. u_i

Rules For Building A Mesh

In order to build a mesh, we need to transform the domain Ω into a polygonal domain Ω_p. A triangulation on Ω_p it's a partition $\mathcal{T}_h(\Omega_p) := \{T_e \subseteq \Omega_p \mid e = 1, ..., M\}$ such that:

1. $\bigcup_{T \in \mathcal{T}_h} T = \Omega_p$
2. $T_i \bigcap T_j = \sigma_{i,j}$ is a either a point or a "face" (i.e. an edge for triangles) of the element for every $i \neq j$
3. $T_i^\circ \bigcap T_j^\circ = \emptyset$ for all $i \neq j$

It's not in the scope of this presentation, but usually when discussing the element of a FEM, it's best practice to formally define a basic element $K \subseteq \mathbb{R}^n$ and describe all other elements as affine transformations of K:

$$\phi_l : K \to T_l \tag{18}$$

This helps when computing formulas for the stiffness matrix, and when discussing formal properties of the method.

Building The Stiffness Matrix

Once we have meshed the domain, we need to build the so called stiffness matrix $A = (a(v_j, v_i))_{i,j}$.
Since by construction polynomials that don't share elements nodes have $a(v_j, v_i) = 0$, one clever way to go about this task is to start from a zero matrix, and progressively add the single contribution of each element.

As stated before, the computation of $a(v_i, v_j)$ is done on a basic element K where computations are easy. Then, for all other possible elements, the same computation is performed by using a change of variables.

Convergence Results

The following theorem states that, as long as the triangle mesh gets smaller (in a special way!), the interpolation of nodes using these piecewise polynomials will converge to the actual function.

Definition 5

$$|v|_{H^m(\Omega)} := \sqrt{\int_\Omega ||D^m v||^2} \tag{19}$$

Theorem 6

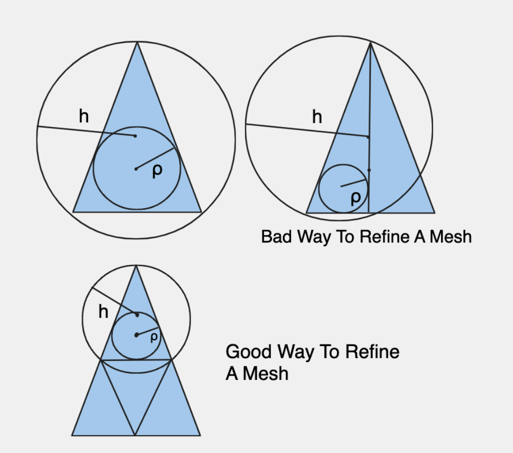
Let $\Pi_h : \mathcal{C}(\Omega) \to \mathbb{P}^1(K)$ be the operator that sends a continuous function to its interpolation (made using fixed nodes, meshing Ω with elements of shape K). Then there exists a constant $c > 0$ such that:

$$|v - \Pi_h(v)|_{H^m(\phi_l(K))} \leq c \frac{(h_{\phi_l(K)})^2}{(\rho_{\phi_l(K)})^m} |v|_{H^2(\phi_l(K))} \tag{20}$$

Where $h_{\phi_l(K)}$ and $\rho_{\phi_l(K)}$ are the radius of the excircle and incircle of $\phi_l(K)$

What The Convergence Result Is Telling Us

As long as we keep the ratio of the excircle and incircle the same, for smaller elements (i.e. for smaller h) the finite approximation space will become dense in the space of solutions of the PDE. Coupled with Cea's lemma, this ensures us that the FEM will converge to the actual solution.



Let's See A Coded Example!

Thank you for your attention!