FINITE ELEMENT METHOD

– Project –

for

Numerical Differential Equations

submitted by

Chirantan Mukherjee

January 14, 2021

Chirantan Mukherjee
c.mukherjee@student.uw.edu.pl
Student ID: K-12784

Preface

This note is part of my project on "Finite Element Method".

I have tried to provide a very "generalized" theory of the finite element method (for $1D$ case) and covered almost all background knowledge that would be required to understand them without any difficulty. The first chapter sets the general introduction and notation that we are going to use throughout the notes and deals with: *linear second order elliptic operators, the Poisson equation and the Laplace equation, Dirichlet and Neumann boundary conditions, strong and weak formulation of the one dimensional Poisson problem with homogeneous Dirichlet boundary conditions.*The second chapter deals with: *Galerkin approximation. Galerkin orthogonality and Cea's lemma.* The third chapter deals with: *a $1D$ example of finite element method: the basis of the Galerkin finite dimensional space, construction of a cardinal basis for piecewise polynomials of degree one and piecewise polynomials of degree two and base functions in the reference element, affine transformation of the reference element to a generic element of the mesh, local construction of the cardinal basis, hierarchical basis and computation of the stiffness matrix using the reference element.* The final chapter deals with: *interpolation error and approximation error, convergence of the finite element method and convergence rate.*
I have also tried to make this note error-free to the best of my ability. But if you notice any error or any inconsistency, please let me know.

Lastly, thank you for choosing to read my note. I hope you will enjoy it :)

Chirantan Mukherjee
c.mukherjee@student.uw.edu.pl

Contents

1 Second Order Linear Elliptic Operator **1**

2 Galerkin Method **5**

3 Finite Element Method **7**

3.1 FEM in one dimensional case: basis 8

3.2 FEM in one dimensional case: linear system 12

4 Error and Convergence rate **14**

1 Second Order Linear Elliptic Operator

Definition 1.1. *The general form of a* **second order linear differential equation** *is:*

$$L\omega = -\sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}[a_{ij}\frac{\partial \omega}{\partial x_j}] + \sum_{i=1}^{d} b_i \frac{\partial \omega}{\partial x_i} + a_0\omega$$
$$= div(A\nabla\omega) + \vec{\mathbf{b}} \cdot \nabla\omega + a_0\omega$$

where,

L *is the* **second order linear differential operator**.

$\vec{\mathbf{x}} \in \mathbb{R}^d$ *i.e.* $\vec{\mathbf{x}} = (x_1, x_2, \ldots, x_d)$. *In this note* $d = \{1, 2, 3\}$.

$A = A(x)$, *matrix with entries* $a_{ij} = a_{ij}(x)$.

$\vec{\mathbf{b}} = \vec{\mathbf{b}}(x)$, *vector with component* $b_i = b_i(x)$.

$div[\vec{\mathbf{F}}(\vec{\mathbf{x}})] = \sum_{i=1}^{d} \frac{\partial F_i}{\partial x_i} \in \mathbb{R}$.

$\vec{\mathbf{F}}(\vec{\mathbf{x}}) = \begin{pmatrix} F_1(\vec{\mathbf{x}}) \\ F_2(\vec{\mathbf{x}}) \\ \vdots \\ F_d(\vec{\mathbf{x}}) \end{pmatrix}$ *where,* $\vec{\mathbf{F}} \colon \vec{\mathbf{x}} \to \vec{\mathbf{F}}(\vec{\mathbf{x}}) \in \mathbb{R}^d$.

$\nabla\varphi(\vec{\mathbf{x}}) = \begin{pmatrix} \frac{\partial\varphi}{\partial x_1}(\vec{\mathbf{x}}) \\ \frac{\partial\varphi}{\partial x_2}(\vec{\mathbf{x}}) \\ \vdots \\ \frac{\partial\varphi}{\partial x_d}(\vec{\mathbf{x}}) \end{pmatrix}$

Definition 1.2. $L\omega = div(A\nabla\omega) + \vec{\mathbf{b}} \cdot \nabla\omega + a_0\omega$

where,

The first term $\mathbf{div(A\nabla}\omega)$ *is called the* **diffusion term**.

The second term $\tilde{\mathbf{b}} \cdot \nabla\omega$ *is called the* **transport term**.

The last term $\mathbf{a_0}\omega$ *is called the* **reaction term**.

Let, Ω be a bounded domain (open and connected) in $\mathbb{R}^d$.

Definition 1.3. *The differential operator* L *is said to be* **elliptic** *in* Ω *if there exists a constant* α_0 *such that*

$$\sum_{i,j=1}^{d} = a_{ij}(x)\xi_i\xi_j \geq \alpha_0|\xi|^2$$

for each $\xi \in \mathbb{R}^d$ *and almost every* $x \in \Omega$ *and* $|\xi|^2 = \sum_{i=1}^{d} \xi^2$.

Example 1 (Laplace Operator). $L\omega = -\Delta\omega$
Comparing with the general form, we see, $A = I$ *and* $div\nabla = \Delta$
$$\Delta\varphi(x) = \sum_{i=1}^{d} \frac{\partial^2 \varphi}{\partial x_i^2}$$

Definition 1.4 (Poisson Equation). $-\Delta u = f$, *where* f *is the given function.*

Definition 1.5 (Laplace Equation). *Is the homogeneous Poisson equation,* $-\Delta u = 0$.

For instance, the unknown u in the Poisson equation can represent the displacement of an elastic membrane ($d = 2$) due to the application of a vertical force of intensity f.

For a given function f, the Poisson equation $-\Delta u = f$ can have **infinite solutions**. To fix a unique solution we add **boundary conditions** to it.

$$(\mathbf{D}) \begin{cases} -\Delta u = f & \text{in } \Omega \\ u = g & \text{on } \partial\Omega \end{cases}$$

where, g is a given function defined on $\partial\Omega$.

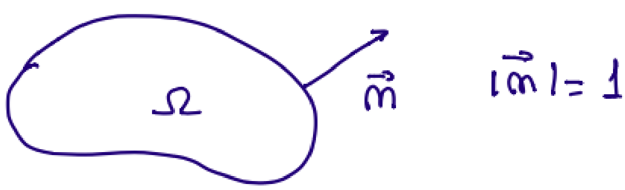
(D) $\equiv$ Poisson problem with **Dirichlet boundary condition** (non-homogeneous if $g \neq 0$).

Example 2. $\begin{cases} -\Delta u = f & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases}$
The membrane is clamped on $\partial\Omega$.

Another possibility is to impose the **Neumann boundary condition**:

$$(\mathbf{N}) \begin{cases} -\Delta u = f & \text{in } \Omega \\ \dfrac{\partial u}{\partial \eta} = \mu & \text{on } \partial\Omega \end{cases}$$

where, $\dfrac{\partial u}{\partial \eta} = \nabla \cdot \vec{\eta}$ and $\vec{\eta}$ is the outward normal unit vector to $\partial\Omega$.



μ is a given function defined on $\partial\Omega$.

If $\mu = 0$, **N** is called homogeneous Neumann boundary condition and if $\mu \neq 0$, **N** is called non-homogeneous Neumann boundary condition.

Example 3. *Poisson equation with homogeneous Dirichlet boundary condition:*

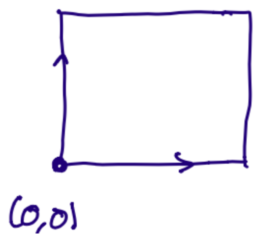
$$\begin{cases} -\Delta u = f & \textit{in } \Omega \\ u = 0 & \textit{on } \partial\Omega \end{cases}$$

Strong *or* **classical formulation** *of the problem.*

If $\Omega = (0,1) \times (0,1)$ *and* $f \equiv 1$

The boundary condition impose:

$$\frac{\partial^2 u}{\partial x_1^2}(0,0) = \frac{\partial^2 u}{\partial x_2^2}(0,0) = 0$$

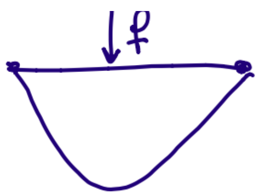


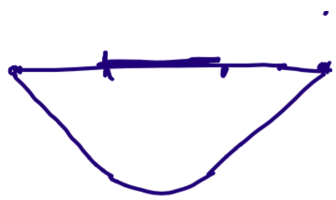
$-\Delta u(0,0) = 0 \neq 1$

We can look for $u \in \mathcal{C}^2(\Omega) \cup \mathcal{C}^0(\overline{\Omega})$, *but not* $u \in \mathcal{C}^2(\overline{\Omega})$.

Example 4. $\begin{cases} -u''(x) = f & \textit{if } x \in (0,1) \\ u(0) = 0 & u(1) = 1 \end{cases}$

An elastic string fixed on $x = 0$ *and* $x = 1$ *loaded with a transversal force* f *of intensity* f





$$f(x) = \begin{cases} 0 & \textit{if } x \in (0, 0.4) \\ -1 & \textit{if } x \in (0.4, 0.6) \\ 0 & \textit{if } x \in (0.6, 1) \end{cases}$$

$$u(x) = \begin{cases} -\dfrac{1}{10} & \textit{if } x \in (0, 0.4) \\ \dfrac{1}{2}x^2 - \dfrac{1}{2}x + \dfrac{2}{25} & \textit{if } x \in (0.4, 0.6) \\ -\dfrac{1}{10}(1-x) & \textit{if } x \in (0.6, 1) \end{cases}$$

This function has a physical meaning as a solution of the problem $u'' = f$.

However, $u \in \mathcal{C}^1(\Omega)$ *but* $u \notin \mathcal{C}^2(\Omega)$

In order to give sense to this kind of solutions we consider a **test function** v and multiply the equation with v, integrating over $(0, 1)$ and obtain,

$$\int_0^1 -u''(x)v(x)dx = \int_0^1 f(x)v(x)dx$$

Integrating the left hand side of the equation by parts, we obtain

$$\int_0^1 u'(x)v'(x)dx - u'(x)v(x) \mid_{x=0}^{x=1}$$

$$= \int_0^1 u'(x)v'(x)dx - [u'(1)v(1) - u'(0)v(0)]$$

Since, we are looking for $u\colon u(1) = 0 = u(1)$

It is "natural" to consider the test function with the same behaviour. This mean that the boundary term

$u'(1)v(1) - u'(0)v(0) = u'(1) \cdot 0 - u'(0) \cdot 0 = 0 - 0 = 0$

So, we look for $u\colon u(1) = 0 = u(o)$ and $\displaystyle\int_0^1 u'(x)v'(x)dx = \int_0^1 f(x)v(x)dx$ for a test function $v\colon v(1) = 0 = v(0)$

$u'v' \in L^1(\Omega) = \{v\colon \int_\Omega |v(x)|dx < +\infty\}$

$fv \in L^1(\Omega)$

We will look for $u \in \mathcal{H}^1(\Omega) = \{v \in L^2(\Omega)\colon v' \in L^2(\Omega)\}$

and we will consider test function:

$v \in \mathcal{H}^1(\Omega)$

2 Galerkin Method

In this chapter we describe the numerical solution of the elliptic boundary value problems by introducing the Galerkin method. We then illustrate the finite element method as a particular case.

Definition 2.1. *The* **weak formulation** *of a generic elliptic problem set on* $\Omega \subset \mathbb{R}^d$ *for* $d = \{1, 2, 3\}$*, can be written in the following form way:*
find $u \in V : a(u, v) = F(v)\ \forall v \in V$
where, V *is a Hilbert subspace of the space* $\mathcal{H}^1(\Omega)$*,*
$a(\cdot, \cdot)$ *being a continuous and coercive bilinear form from* $V \times V$ *in* $\mathbb{R}$ *associated to the differential operator* L*,*
$F(\cdot)$ *being a continuous linear functional from* V *in* $\mathbb{R}$ $F(v) = \int_\Omega fv$.

Definition 2.2. $a(\cdot, \cdot) : V \times V \to \mathbb{R}$ *is* **continuous** *if there exists a positive constant* M *such that* $|a(w, v)| \leq M||w||_V ||v||_V\ \forall w, v \in V$.

Definition 2.3. $a(\cdot, \cdot) : V \times V \to \mathbb{R}$ *is* **coercive** *if there exists a positive constant* α *such that* $|a(v, v)| \geq \alpha||v||_V^2\ \forall v \in V$.

If $a(\cdot, \cdot)$ is coercive in V, then **Lax-Milgram lemma** implies the uniqueness and the existence of the "weak formulation".

For the **Galerkin Method** consider a finite dimensional subspace $V_h \subset V$ and the so called Galerkin problem (or discrete problem):

Definition 2.4 (Galerkin Problem). *Find* $u_h \in V_h$ *such that* $a(u_h, v_h) = F(v_h), \forall v_h \in V_h$.

Since V_h is closed, Lax Milgram lemma gives us the existence and the uniqueness of the solution of the Galerkin problem.

The **finite element method (FEM)** is an example of the Galerkin method.

NOTE: Galerkin method are interesting (from a numerical point of view) because given a basis $v_h \in V_h$, the solution of the Galerkin problem can be computed solving a linear system of equations.

Moreover we can have some information about the "error".
$u_h \in V_h \subset V$. So, it is natural to consider the error measured as $||u - u_h||_V$.

An important property of the Galerkin method is the **Galerkin orthogonality**:
$a(u, v) = F(v), \forall v \in V$

In particular, $a(u, v_h) = f(v_h), \forall v_h \in V_h \subset V$
On the other hand $a(u_h, v_h) = F(v_h), \forall v_h \in V_h$
$\implies a(u - u_h, v_h) = 0, \forall v_h \in V_h$
So, we have by coercivity

$$\begin{aligned}
\alpha||u - u_h||_V^2 &\leq a(u - u_h, u - u_h) \\
&= a(u - u_h, a - v_h + v_h - u_h), \forall v_h \in V_h \\
&= a(u - u_h, v - v_h) \text{ by the Galerkin orthoganility} \\
&\leq M||u - u_h||_V ||u - v_h||_V \text{ by continuity of } a(\cdot, \cdot)
\end{aligned}$$

Thus, $||u - u_h||_V \leq \dfrac{M}{\alpha}||u - v_h||_V, \forall v_h \in V_h.$

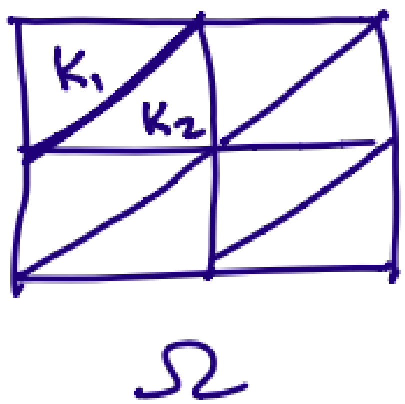
Theorem 1 (Cea's lemma). $||u - u_h||_V \leq \dfrac{M}{\alpha} inf_{v_h \in V_h} ||u - v_h||_V, \forall v_h \in V_h,$ *where, M and α are the continuity and coercivity constant of the bilinear form $a(\cdot, \cdot)$ respectively.*

3 Finite Element Method

We want to study when the solution of a particular Galerkin method converges to the solution of the weak problem when h tends to zero.

In FEM, the finite dimensional subspaces are spaces of "piecewise polynomial" functions, related with the decomposition of elements in Ω

Let us assume Ω is a polyhedral and that we know a family of d-simplices K_m,



$\overline{\Omega} = \bigcup_{m=1}^{M} K_m$ This means functions $v_h \in V_h$ such that
$v_h \mid_{K_m} \in \mathbb{P}_r \forall K_m$, where $\mathbb{P}_r$ is the space of polynomial of degree less than or equal to r.

We will assume that:

1. $K_m^{\circ} \neq \phi, \forall m = \{1, \ldots, M\}$
2. $K_m^{\circ} \cap K_n^{\circ} = \phi$ if $m \neq n$
3. If $K_m \cap K_n = F \neq \phi$ then F is a common subsimplex of both K_m and K_n

Let us denote by h_m, the diameter of K_m.
$h_m = max_{x,y \in K_m}|x - y|$ and $h = max_{1 \leq m \leq M} h_m$.

We call $z_h = \{K_m\}_{m=1}^{M}$ a **mesh** of Ω, i.e. a decomposition of $\overline{\Omega}$ into non-overlapping elements.

In FEM, $V_h = \{v_h \in V : v_h \mid_{K_m} \in \mathbb{P}_r, \forall K_m \in z_h\}$
where $\mathbb{P}_r$ is the space of polynomial of degree less than or equal to r.

Ω is assumed to be a polyhedral and the elements of z_h are the $d-$simplices, namely **triangles** if $d = 2$, **tetrahedra** if $d = 3$.

3.1 FEM in one dimensional case: basis

Let us assume Ω is an interval (a, b). In this case a mesh of the domain is a partition z_h of (a, b) in sub-intervals:
$K_m = [P_{m-1}, P_m], m = \{1, \ldots, M\}$
$P_0 = 1, P_M = b$ and $P_0 < P_1 < \ldots P_M$
diameter $h_m = P_m - P_{m-1}$

$\overline{\Omega} = [a, b] = \bigcup_{m=1}^{M} K_m$ such that:

1. $K_m^{\circ} \neq \phi, \forall m = \{1, \ldots, M\}$
2. $K_m^{\circ} \cap K_n^{\circ} = \phi$ if $m \neq n$
3. If $K_m \cap K_n = F \neq \phi$ then F is a common vertex of both K_m and K_n

We consider the following family of spaces
$X_h^r = \{v_h \in \mathcal{C}^0(\overline{\Omega}) \colon v_h \,|_{K_m} \in \mathbb{P}_r \forall K_m \in z_h\}$
where, $\mathbb{P}_r$ is the space of polynomial of degree less than or equal to r.

NOTE: The space X_h^r is a subspace of $\mathcal{H}^1(\Omega)$

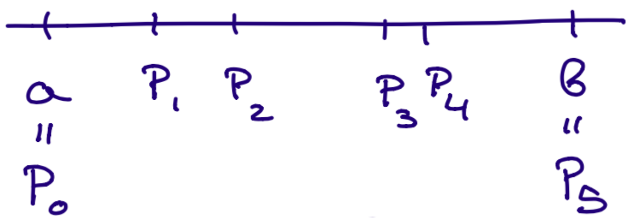
For all Galerkin problem we will consider $V_h = X_h^r \cup V$

We must now "identify" a basis of X_h^r in order to "construct" a linear system to solve and compute the solution of the Galerkin problem.

NOTE: If $\{\varphi_i\}_{i=1}^{N_h}$ is a basis of V_h, the Galerkin problem reads: find $u_h = \sum_{j=1}^{N_h} V_j \varphi_{hj}$
such that,
$\sum_{j=1}^{N_h} V_j a(\varphi_{hj}, \varphi_{hi}) = F(\varphi_{hi}) \; \forall i \in \{1, \ldots, N_h\}$
or equivalently,
find $V \in \mathbb{R}^{N_h}$ such that $SV = F$
where, $S \in \mathbb{R}^{N_h \times N_h}$, the matrix with entries s_{ij} and $F \in \mathbb{R}^{N_h}$, the vector with components F_i.
$V \in \mathbb{R}^{N_h}$ is the vector with components V_h.

In real application, N_h can be very large ($N_h \equiv 10^6$ for instance) so, it is important to use basis functions with small support.
In this way $a(\varphi_{hj}, \varphi_{hi} = 0)$ "often", for example when S is **sparse**.

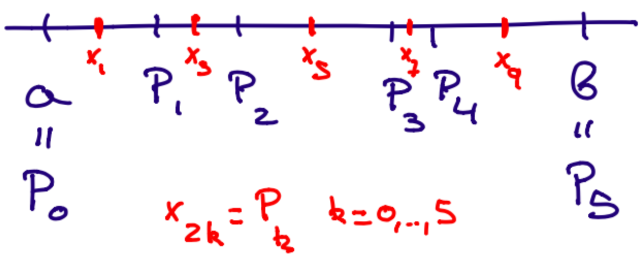
Example 5. *The space* X_h^1
$z_h = \{K_m\}_{m=1}^5$
diameter of $X_h^1 = 6$



$$\varphi_{hi}(x) = \begin{cases} \dfrac{x - P_{i-1}}{P_i - P_{i-1}} & \textit{if } x \in [P_{i-1}, P_i] \\ \dfrac{P_{i+1} - x}{P_{i+1} - P_i} & \textit{if } x \in [P_i, P_{i+1}] \\ 0 & \textit{otherwise} \end{cases} \qquad \forall i \in \{1, 2, 3, 4\}$$

$$\varphi_{h0}(x) = \begin{cases} \dfrac{P_1 - x}{P_1 - P_0)} & \textit{if } x \in [P_0, P_1] \\ 0 & \textit{otherwise} \end{cases} \quad \textit{and } \varphi_{h5}(x) = \begin{cases} \dfrac{x - P_5}{P_5 - P_4} & \textit{if } x \in [P_4, P_5] \\ 0 & \textit{otherwise} \end{cases}$$

Example 6. *The space* X_h^2
$z_h = \{K_m\}_{m=1}^5$
diameter of $X_h^2 = 11$



$$\varphi_{2k}(x) = \begin{cases} \dfrac{(x - x_{2(k-1)})(x - x_{2k-1})}{(x_{2k} - x_{2(k-1)})(x_{2k} - x_{2k-1})} & \textit{if } x \in [P_{k-1}, P_k] \\ \dfrac{(x_{2k+1} - x)(x_{2(k+1)} - x}{(x_{2k+1} - x_{2k})(x_{2(k+1)} - x_{2k})} & \textit{if } x \in [P_k, P_{k+1}] \\ 0 & \textit{otherwise} \end{cases} \qquad \forall k \in \{1, 2, 3, 4\}$$

$$\varphi_0(x) = \begin{cases} \dfrac{(x_1 - x)(x_2 - x)}{(x_1 - x_0)(x_2 - x_0)} & \textit{if } x \in [P_0, P_1] \\ 0 & \textit{otherwise} \end{cases} \quad \textit{and}$$
$$\varphi_{10}(x) = \begin{cases} \dfrac{(x - x_8)(x - x_9)}{(x_{10} - x_8)(x_{10} - x_9)} & \textit{if } x \in [P_4, P_5] \\ 0 & \textit{otherwise} \end{cases}$$
$$\varphi_{2k-1}(x) = \begin{cases} \dfrac{(x - x_{2(k-1)})(x_{2k} - x)}{(x_{2k-1} - x_{2(k-1)})(x_{2k} - x_{2k-1})} & \textit{if } x \in [P_{k-1}, P_k] \\ 0 & \textit{otherwise} \end{cases} \qquad \forall k \in \{1, 2, 3, 4, 5\}$$

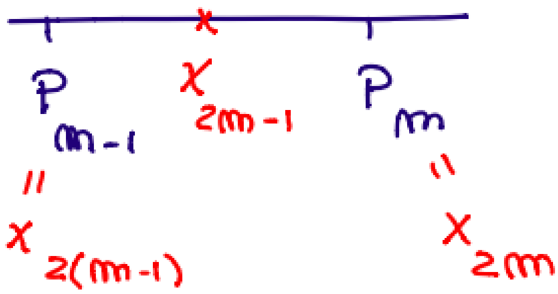
$\{\varphi_{hi}^{(1)}\}_{i=0}^{5}$ is the cardinal basis of X_h^1
$\{\varphi_{hi}^{(2)}\}_{i=0}^{10}$ is the cardinal basis of X_h^2

On each element K_m, there are only two functions of the cardinal basis of X_h^1 that are different from zero: $\varphi_{h(m-1)}^{(1)}$ and $\varphi_{hm}^{(1)}$.

The generic element of the mesh $K_m = [P_{m-1}, P_m]$ is an affine transformation of the reference element $\hat{K} = [0, 1]$
$T_{K_m} : \hat{K} \to K_m$
$\qquad \hat{x} \mapsto x = P_{m-1} + \hat{x}(P_m - P_{m-1})$

If we define the two base functions
$\hat{\varphi}_0^{(1)}(\hat{x}) = 1 - \hat{x}$ and $\hat{\varphi_1}^{(1)}(\hat{x}) = \hat{x}$
then for each $x \in K_m$ we have,
$\hat{\varphi}_{h(m-1)}^{(1)}(x) = \hat{\varphi}_0^{(1)}(T_{K_m}^{-1}(x))$ and $\hat{\varphi}_{h(m)}^{(1)}(x) = \hat{\varphi}_1^{(1)}(T_{K_m}^{-1}(x))$
The definition of the base functions locally (element by element) form the basis defined on a reference element and then transforming it on a specific element will be fundamental when considering a problem in several variables (for example $d = 2$ or $d = 3$).

Similarly on each element K_m there are only three elements of the cardinal basis of X_h^2 that are different from zero: $\varphi_{h2(m-1)}^{(2)}$, $\varphi_{h(2m-1)}^{(2)}$ and $\varphi_{h2m}^{(2)}$.



When we define these three functions on the reference element $\hat{K} = [0, 1]$
$\hat{\varphi}_0^{(2)}(\hat{x}) = 2(\frac{1}{2} - \hat{x})(1 - \hat{x})$,
$\hat{\varphi}_1^{(2)}(\hat{x}) = 4\hat{x}(1 - \hat{x})$ and
$\hat{\varphi}_2^{(2)}(\hat{x}) = 2\hat{x}(\hat{x} - \frac{1}{2})$
then for each element $x \in K_m = [P_{m-1}, P_m]$, we have
$\varphi_{h2(m-1)}^{(2)}(\hat{x}) = \hat{\varphi}_0^{(2)}(T_{K_m}^{-1}(x))$,
$\varphi_{h(2m-1)}^{(2)}(\hat{x}) = \hat{\varphi}_1^{(2)}(T_{K_m}^{-1}(x))$ and
$\varphi_{h2m}^{(2)}(\hat{x}) = \hat{\varphi}_2^{(2)}(T_{K_m}^{-1}(x))$

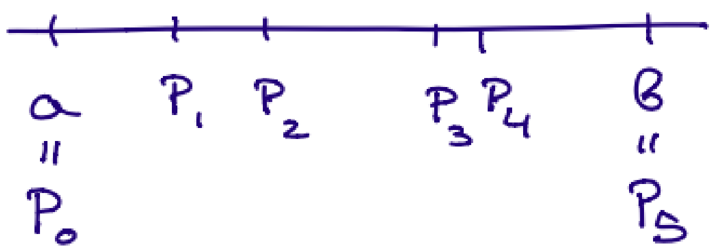
If $\{\varphi_{hi}\}_{i=1}^{N_h}$ is a cardinal basis of V_h and v_h is an element of V_h then, $v_h(x) = \sum_{i=1}^{N_h} v_h(x_i)\varphi_{hi}(x)$, which is called the **degrees of freedom**.
However, it could be convenient to consider other basis (i.e. different from the cardinal basis) for instance,
$\hat{\psi}_0(\hat{x}) = 1 - \hat{x}$,
$\hat{\psi}_1(\hat{x}) = \hat{x}$ and
$\hat{\psi}_2(\hat{x}) = \hat{x}(1 - \hat{x})$
and then for $x \in K_m$, we have
$\psi_{h2(m-1)}(\hat{x}) = \hat{\psi}_0(T_{K_m}^{-1}(x))$,
$\psi_{h(2m-1)}(\hat{x}) = \hat{\psi}_1(T_{K_m}^{-1}(x))$ and
$\psi_{h2m}(\hat{x}) = \hat{\psi}_2(T_{K_m}^{-1}(x))$
This way we are able to define another basis of X_h^2. It is a **hierarchical basis** because it is constructed by adding functions (one for each element) to a basis of X_h^1.

3.2 FEM in one dimensional case: linear system

Approximation with linear finite elements:

Example 7. $-u'' + u = f$ *when* $a < x < b$
$u(a) = 0$ *and* $u(b) = 0$

The weak formulation of the example reads: find $u \in \mathcal{H}_0^1(a,b)$ such that $\int_a^b u'v' + \int_a^b uv = \int_a^b fv \; \forall v \in H_0^1(a,b)$.
We introduce a decomposition z_h of (a,b) in M sub-intervals
$z_h = \{K_m\}_{m=1}^{M_h}$ and $K_m = [P_{m-1}, P_m]$
and consider the finite dimensional subspace of $\mathcal{H}_0^1(\Omega)$
$V_h = \{v_h \in X_h^1 \colon v_h(a) = v_h(b) = 0\}$



The corresponding FEM problem reads:
find $u_h \in V_h$ such that
$\int_a^b (u_h' v_h' + u_h v_h) = \int_a^b f v_h \; \forall v_h \in V_h$.
If $\{\varphi_{hi}\}_{i=0}^{M_h}$ is the basis associated to the mesh z_h then a basis of $V_h = X_h^1 \cap \mathcal{H}_0^1(\Omega)$ is given by $\{\varphi_{hi}\}_{i=0}^{M_h - 1}$.

In this example, the $dimV_h = N_h = M_h - 1$ where M_h is the number of elements (sub-intervals) of the mesh.

The FEM problem is equivalent to the linear system: $S\vec{\mathbf{V}} = \vec{\mathbf{f}}$
where, S is the matrix with entries s_{ij}
$s_{ij} = \int_a^b (\varphi_{hj}' \varphi_{hi}' + \varphi_{hj}\varphi_{hi})$
and $\vec{\mathbf{f}}$ with coefficients f_i
$f_i = \int_a^b f\varphi_{hi}$

If $\vec{\mathbf{V}} \subset \mathbb{R}^{N_h}$ is a solution of this linear system, then $u_h(x) = \sum_{i=1}^{N_h} V_i \varphi_{hi}(x)$ is the solution of the finite element problem.

Now we have to compute the entries of S and the coefficients of $\vec{\mathbf{f}}$.
If we consider the support of the base function we notice that S is not only sparse but also tridiagonal and symmetric.
A genereic non-null element of the stiffness matrix is given by,

$$a_{ij} = \int_{P_{i-1}}^{P_i} (\varphi_j'\varphi_i' + \varphi_j\varphi_i) + \int_{P_i}^{P_{i+1}} (\varphi_j'\varphi_i' + \varphi_j\varphi_i) \neq 0 \text{ iff } j \in \{i-1, i, i+1\}$$

Using the fact that $K_m = [P_{m-1}, P_m] = T_{K_m}\hat{K}$ the first addendum becomes

$$\int_{P_{i-1}}^{P_i} [\varphi_j'(x)\varphi_i'(x) + \varphi_j(x)\varphi_i(x)]dx$$

(recall, $x = P_{i-1} + \hat{x}(P_i - P_{i-1}) = T_{K_i}(x)$)

$$= \int_0^1 (P_i - P_{i-1})[\varphi_j'(T_{K_i}\hat{x})\varphi_i'(T_{K_i}\hat{x}) + \varphi_j(T_{K_i}\hat{x})\varphi_j(T_{K_i}\hat{x})]d\hat{x}$$

and we compute the second one similarly.

We recall that

$\varphi_{i-1}(T_{K_i}\hat{x}) = \hat{\varphi}_0(\hat{x})$ and $\varphi_i(T_{K_i}\hat{x}) = \hat{\varphi}_1(\hat{x})$

Concerning the derivatives, using chain rule, we obtain,

$$\begin{aligned}\hat{\varphi}_0'(\hat{x}) &= \frac{\partial}{\partial\hat{x}}\hat{\varphi}_0(\hat{x}) \\ &= \frac{\partial}{\partial\hat{x}}[\varphi_{i-1}(T_{K_i}(\hat{x}))] = \varphi_{i-1}'(T_{K_i}(\hat{x}))(P_i - P_{i-1})\end{aligned}$$

analogously,

$$\hat{\varphi}_1'(\hat{x}) = \frac{\partial}{\partial\hat{x}}\hat{\varphi}_1(\hat{x}) = \varphi_i'(T_{K_i}(\hat{x}))(P_i - P_{i-1})$$

Hence, $\varphi_{i-1}'(T_{K_i}(\hat{x})) = \dfrac{1}{P_i - P_{i-1}}\hat{\varphi}'_0(\hat{x})$

$$\varphi_i'(T_{K_i}(\hat{x})) = \frac{1}{P_i - P_{i-1}}\hat{\varphi}'_1(\hat{x})$$

and,

$$\begin{aligned}s_{i-1\ i} &= \int_0^1 (P_i - P_{i-1})[\frac{\hat{\varphi}_0'(\hat{x})}{P_i - P_{i-1}}\frac{\hat{\varphi}_1'(\hat{x})}{P_i - P_{i-1}} + \hat{\varphi}_0(\hat{x})\hat{\varphi}_1(\hat{x})]d\hat{x} \\ &= \frac{1}{P_i - P_{i-1}}\int_0^1 \hat{\varphi}_0'(\hat{x})\hat{\varphi}_1'(\hat{x})d\hat{x} + \int_0^1 \hat{\varphi}_0'(\hat{x}) + \hat{\varphi}_1'(\hat{x})d\hat{x}\end{aligned}$$

NOTE: This two integrals are computed in the reference element. In this example with constant coefficients, they do not depend on i.

4 Error and Convergence rate

Interpolation operator and interpolation error:

Let us set, $\Omega = (a, b)$, $z_h = \{K_m\}_{m=1}^{M_h}$, a mesh of Ω and X_h^1, the function element of piecewise polynomial function of degree 1.

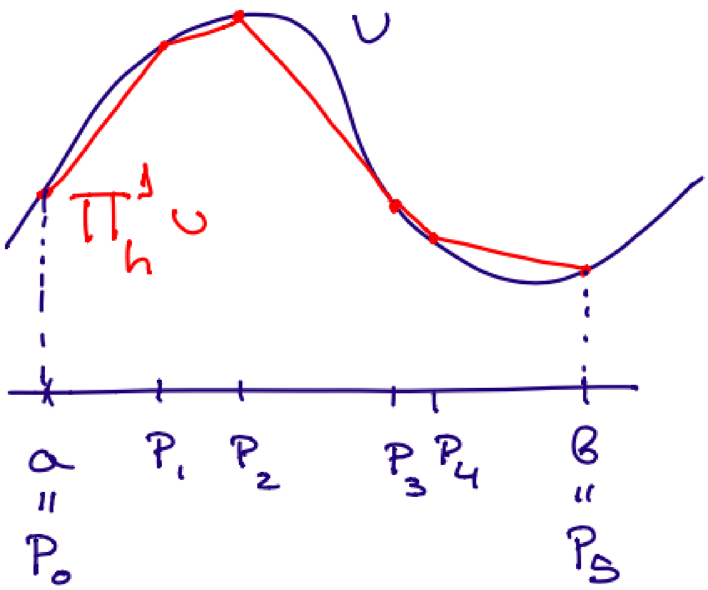
Given, $v \in \mathcal{C}^0(\overline{\Omega})$, we define the **interpolant** of v in X_h^1 the function $\Pi_h^1 v \in X_h^1$ such that,

$\Pi_h^1(v)(P_i) = v(P_i)\ \forall i \in \{0, \ldots, M_h\}$

By using the cardinal basis $\{\varphi_{hi}^{(1)}\}_{i=0}^{M_h}$ of X_h^1,

$$\Pi_h^1(v)(P_i) = \sum_{i=0}^{M_h} v(P_i)\varphi_{hi}(x)$$

Definition 4.1. *The operator* $\Pi_h^1\colon \mathcal{C}^0(\overline{\Omega}) \to X_h^1$ *is called the* **interpolation operator**.



Analogously, we can define the interpolation operator $\Pi_h^2\colon \mathcal{C}^0(\Omega) \to X_h^2$ given by $v \in \mathcal{C}^0(\Omega)$.

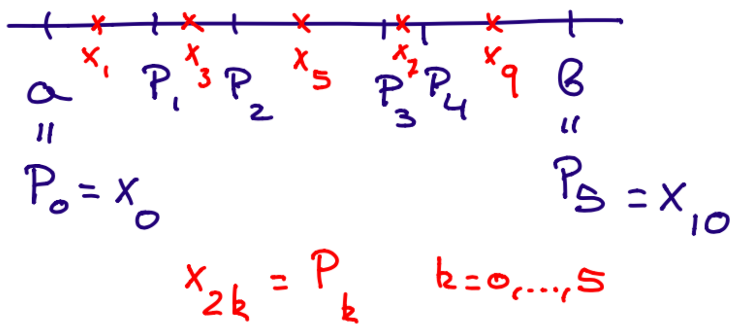
$\Pi_h^2 v \in X_h^2$ is the unique function of X_h^2 such that

$\Pi_h^2 v(x_i) = v(x_i)\ i \in \{0, \ldots, 2M_h\}$

Using the cardinal basis of $\{\varphi_{hi}^{(2)}\}_{i=0}^{2M_h}$ of X_h^2,

we have,

$$\Pi_h^2 v(x) = \sum_{i=0}^{2M_h} v(x_i)\varphi_{hi}^{(2)}(x)$$



Theorem 2. *Let, $v \in \mathcal{H}^{r+1}(a, b)$ and let $\Pi_h^r v \in X_h^r$ be its interpolating function $(r = 1, 2)$. The following estimates of the interpolation error holds,*

1. $||v - \Pi_h^r v||_{\mathcal{L}^2(a,b)} \leq C_0 h^{r+1} |v|_{\mathcal{H}^{r+1}(a,b)}$
2. $|v - \Pi_h^r v|_{\mathcal{H}^1(a,b)} \leq C_1 h^{r+1} |v|_{\mathcal{H}^{r+1}(a,b)}$

The constants C_0 and C_1 are independent of v and h.

Remember that, $||v||_{\mathcal{L}^2(a,b)} = \displaystyle\int_a^b v(x)^2 dx$

and, $|v|^2_{\mathcal{H}^s(a,b)} = \displaystyle\int_a^b [v^{(s)}(x)]^2 dx$

Using this result on the interpolation error and Cea's lemma, it is easy to obtain and estimate the **approximation error**.

Let, $u \in V$ be the solution of the weak problem and $u_h \in V_h = V \cap X_h^r$, the solution of the finite element problem.

Cea's lemma: $||u - u_h||_V \leq \dfrac{M}{\alpha} inf_{v_h \in V_h} ||u - v_h||_V$

In the $1D$ example, we have $\Omega = (a, b)$, $u \in V \subset \mathcal{H}^1(\Omega) \subset \mathcal{C}^0(\overline{\Omega})$.
Hence, $\Pi_h^r u$ is well-defined and,

$$||u - u_h||_V \leq \frac{M}{\alpha} ||u - \Pi_h^r u||_V.$$

If $u \in \mathcal{H}^{r+1}(a, b)$ then,

$$\begin{aligned}
||u - \Pi_h^r u||_V^2 &\leq ||u - \Pi_h^r u||^2_{\mathcal{L}^2(a,b)} + |u - \Pi_h^r u|^2_{\mathcal{H}^1(a,b)} \\
&\leq (C_0 h^{r+1})^2 |u|^2_{\mathcal{H}^{r+1}(a,b)} + (C_1 h^r)^2 |u|^2_{\mathcal{H}^{r+1}(a,b)} \\
&= h^{2r}(C_0^2 h^2 + C_1^2)|u|^2_{\mathcal{H}^{r+1}(a,b)}
\end{aligned}$$

Taking for instance $h < 1$, we have

$$||u - \Pi_h^r u||_V \leq (C_0^2 + C_1^2)^{\frac{1}{2}} h^r |u|_{\mathcal{H}^{r+1}(a,b)}$$

Denoting $(C_0^2 + C_1^2)^{\frac{1}{2}}$ by $\hat{C}$, we get
$$||u - \Pi_h^r u||_V \leq \hat{C} h^r |u|_{\mathcal{H}^{r+1}(a,b)}$$
Hence,
$$||u - u_h||_V \leq \frac{M}{\alpha} \hat{C} h^r |u|_{\mathcal{H}^{r+1}(a,b)}$$
It follows that if $u \in \mathcal{H}^r(a,b)$ and u_h denotes the solution of the finite element Galerkin problem using piecewise polynomial function of degree r on a mesh $z_h = \{P_{m-1}, P_m\}_{m=1}^{M_h}$ of diameter h, then considering a family of meshes with $h \to 0$ and $r \geq 1$, then
$$||u - u_h||_V \leq \frac{M}{\alpha} \hat{C} h^r |u|_{\mathcal{H}^{r+1}(a,b)} \to 0.$$
In other words finite element approximation converges in V to the solution of weak problem u.

Notice, that we need $r \geq 1$ and $u \in \mathcal{H}^{r+1}(a,b)$ but in general we only know $u \in \mathcal{H}^1(a,b)$.

However since regular functions are dense in $\mathcal{H}^1(a,b)$ $\forall \epsilon > 0$ $\exists v_\epsilon \in \mathcal{H}^{r+1}(a,b)$ such that $|u - v_\epsilon|_V \leq \frac{\alpha}{M} \cdot \frac{\epsilon}{2}$.

So, we have,
$$\begin{aligned} ||u - u_h||_V &\leq \frac{M}{\alpha} inf_{v_h \in V_h} ||u - v_h|| \\ &\leq \frac{M}{\alpha} ||u - \Pi_h^r v_\epsilon||_V \\ &\leq \frac{M}{\alpha} [||u - v_\epsilon||_V + ||v_\epsilon - \Pi_h^r v_\epsilon||_V] \\ &\leq \frac{\epsilon}{2} + \hat{C} h^r |v_\epsilon|_{\mathcal{H}^{r+1}(a,b)} \end{aligned}$$

Taking $h < [\dfrac{\hat{C}\epsilon}{|v_\epsilon|_{\mathcal{H}^{r+1}(a,b)} 2}]$, we have $\hat{C} h^r |v_\epsilon|_{\mathcal{H}^{r+1}(a,b)} \leq \dfrac{\epsilon}{2}$

and,

$||u - u_h||_V \leq \epsilon$

In other words, if $u \in \mathcal{H}^1(a,b)$ then $||u - u_h||_V \to 0$ as $h \to 0$, being $u_h \in V \cap X_h^r$.

We know the **convergence rate**
$$||u - u_h||_V \leq C h^r$$
only if, $u \in \mathcal{H}^{r+1}(a,b)$

If $u \in \mathcal{H}^{p+1}(a,b)$ with $p \geq 1$ and $u_h \in V \cap X_h^r$ with $r \geq 1$, then,
$$||u - u_h||_V \leq C_* h^s |v|_{\mathcal{H}^{p+1}(a,b)}$$
where, $s = min\{p, r\}$.

In other words, if $u \in \mathcal{H}^{p+1}(a,b)$, the maximum value of r (polynomial degree) that makes sense to take is $r = p$.

NOTE: Values higher than p does not ensure a better rate of convergence.

References

[1] Alfio Quarteroni. *Numerical Models for Differential Problems (MS&A)*. Springer; 1st ed. 2009. Corr. 3rd printing 2012 edition (April 8, 2009). ISBN: 8847010705.

[2] Dietrich Braess. *Finite Elements Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press; 3rd edition (April 12, 2007). ISBN: 0721662439.

[3] Leszek Marcinkowski. *Solving numerically differential equations*. Lecture Notes (January 11, 2021).