The Finite Element Method – Lecture Notes

Per-Olof Persson
persson@berkeley.edu

March 10, 2022

1 Introduction to FEM

1.1 A simple example

Consider the model problem

$$-u''(x) = 1, \text{ for } x \in (0,1) \tag{1.1}$$

$$u(0) = u(1) = 0 \tag{1.2}$$

with exact solution $u(x) = x(1-x)/2$. Find an approximate solution of the form

$$\hat{u}(x) = A\sin(\pi x) = A\varphi(x) \tag{1.3}$$

Various ways to impose the equation:

Collocation : Impose $-\hat{u}''(x_c) = 1$ for some *collocation* point $x_c \Longrightarrow A\pi^2\sin(\pi x_c) = 1$

- $x_c = \frac{1}{2} \Longrightarrow A = \frac{1}{\pi^2} = 0.1013$
- $x_c = \frac{1}{4} \Longrightarrow A = \frac{\sqrt{2}}{\pi^2} = 0.1433$

Average : Impose the *average* of the equation over the interval:

$$\int_0^1 -\hat{u}''(x)\,dx = \int_0^1 1\,dx \tag{1.4}$$

$$\int_0^1 A\pi^2\sin(\pi x)\,dx = A\pi^2\frac{2}{\pi} = 2A\pi = 1 \tag{1.5}$$

$$A = \frac{1}{2\pi} = 0.159 \tag{1.6}$$

Galerkin : Impose the *weighted average* of the equation over the interval:

$$\int_0^1 -\hat{u}''(x)v(x)\,dx = \int_0^1 1\cdot v(x)\,dx \tag{1.7}$$

A Galerkin method using the weight functions $v(x)$ from the same space as the solution space, that is, $v(x) = \varphi(x)$. This gives

$$\int_0^1 A\pi^2\sin^2(\pi x)\,dx = \int_0^1 \sin(\pi x)\,dx \tag{1.8}$$

$$A\pi^2\cdot\frac{1}{2} = \frac{2}{\pi} \tag{1.9}$$

$$A = \frac{4}{\pi^3} = 0.129 \tag{1.10}$$

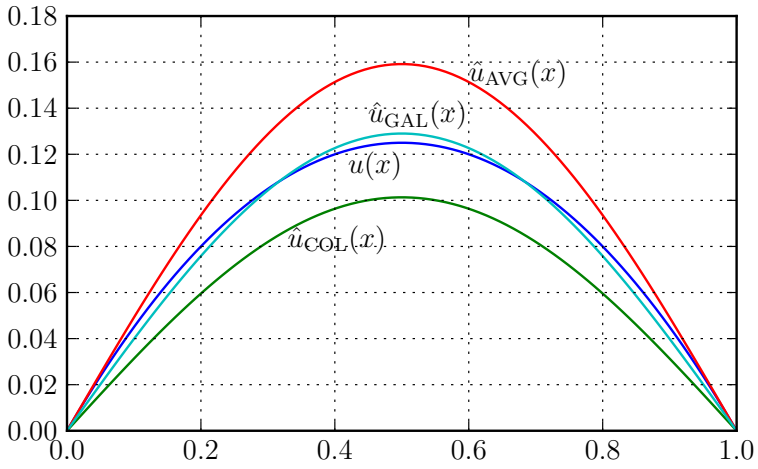


Figure 1: Solutions of the model problem (1.1)-(1.2) using collocation, average, and Galerkin.

The three solutions are shown in figure 1.1. The finite element method is based on the Galerkin formulation, which in this example clearly is superior to collocation or averaging.

1.2 Other function spaces

Use piecewise linear, continuous functions of the form $\hat{u}(x) = A\varphi(x)$ with

$$\varphi(x) = \begin{cases} 2x & x \leq \frac{1}{2} \\ 2 - 2x & x > \frac{1}{2} \end{cases} \tag{1.11}$$

Galerkin gives the FEM formulation

$$\int_0^1 -\hat{u}''(x)\varphi(x)\,dx = \int_0^1 \varphi(x)\,dx \tag{1.12}$$

Since $\hat{u}''$ is not bounded, integrate LHS by parts:

$$\int_0^1 \hat{u}'(x)\varphi'(x)\,dx - \left[\hat{u}'(x)\varphi(x)\right]_0^1 = \int_0^1 \hat{u}'(x)\varphi'(x)\,dx - \hat{u}'(1)\varphi(1) + \hat{u}'(0)\varphi(0) \tag{1.13}$$

$$= \int_0^1 \hat{u}'(x)\varphi'(x)\,dx, \tag{1.14}$$

since $\varphi(0) = \varphi(1) = 0$. This leads to the final formulation

$$\int_0^1 \hat{u}'(x)\varphi'(x)\,dx = \int_0^1 \varphi(x)\,dx. \tag{1.15}$$

To solve for $\hat{u}(x) = A\varphi(x)$, note that $\hat{u}'(x) = A\varphi'(x)$ and

$$\varphi'(x) = \begin{cases} 2 & x \leq \frac{1}{2} \\ -2 & x > \frac{1}{2} \end{cases} \tag{1.16}$$

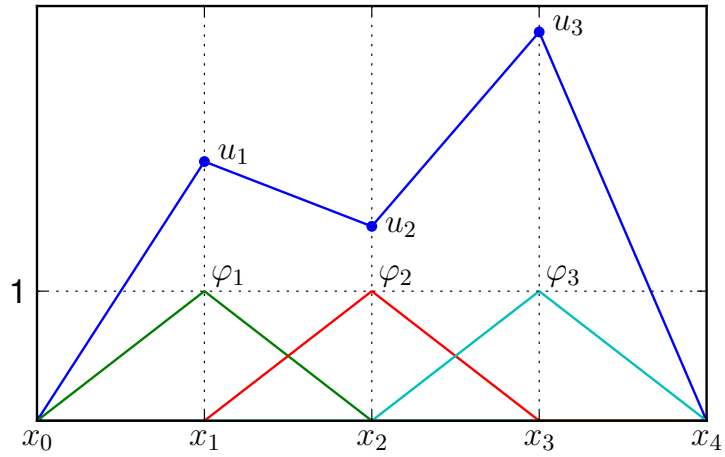


Figure 2: A refined piecewise linear function space, same solution and basis functions.

Equation (1.15) then becomes

$$\int_0^{1/2} (2A)(2)\,dx + \int_{1/2}^1 (-2A)(-2)\,dx = \int_0^{1/2} 2x\,dx + \int_{1/2}^1 (2-2x)\,dx \tag{1.17}$$

or $4A = 1/2$, that is, $A = 1/8$, and the FEM solution is $\hat{u}(x) = \varphi(x)/8$.

1.3 More basis functions

To refine the solution space, introduce a *triangulation* of the domain $\Omega = [0,1]$ into non-overlapping *elements*:

$$T_h = \{K_1, K_2, \ldots\} \tag{1.18}$$

such that $\Omega = \cup_{K\in T_h} K$. Now consider the space of continuous functions that are piecewise linear on the triangulation and zero at the end points:

$$V_h = \{v \in C^0([0,1]) : v|_K \in \mathbb{P}_1(K)\ \forall K \in T_h,\ v(0) = v(1) = 0\}. \tag{1.19}$$

Here $\mathbb{P}_p(K)$ is the space of polynomials on K of degree at most p. Define a basis $\{\varphi_i\}$ for V_h by the basis functions $\varphi_i \in V_h$ with $\varphi_i(x_j) = \delta_{ij}$, for $i, j = 1, \ldots, n$. Our approximate solution $u_h(x)$ can then be written in terms of its expansion coefficients and the basis functions as

$$u_h(x) = \sum_{i=1}^n u_i \varphi_i(x), \tag{1.20}$$

where we note that this particular basis has the convenient property that $u_h(x_j) = u_j$, for $j = 1, \ldots, n$.

A Galerkin formulation for (1.1)-(1.2) can now be stated as: Find $u_h \in V_h$ such that

$$\int_0^1 u_h'(x)v'(x)\,dx = \int_0^1 v(x)\,dx, \qquad \forall v \in V_h \tag{1.21}$$

In particular, (1.21) should be satisfied for $v = \varphi_i$, $i = 1, \ldots, n$, which leads to n equations of the form

$$\int_0^1 u_h'(x)\varphi_i'(x)\,dx = \int_0^1 \varphi_i(x)\,dx, \qquad \text{for } i = 1, \ldots, n \tag{1.22}$$

Insert the expression (1.20) for the approximate solution and its derivative, $u_h'(x) = \sum_{i=1}^n u_i\varphi_i'(x)$:

$$\int_0^1 \left(\sum_{j=1}^n u_j\varphi_j'(x)\right)\varphi_i'(x)\,dx = \int_0^1 \varphi_x(x)\,dx, \qquad i = 1, \ldots, n \tag{1.23}$$

Change order of integration/summation:

$$\sum_{j=1}^n u_j\left[\int_0^1 \varphi_i'(x)\varphi_j'(x)\,dx\right] = \int_0^1 \varphi_i(x)\,dx, \qquad i = 1, \ldots, n \tag{1.24}$$

This is a linear system of equations $A\boldsymbol{u} = \boldsymbol{f}$, with $A = [a_{ij}]$, $\boldsymbol{u} = [u_i]$, $\boldsymbol{f} = [f_i]$, for $i, j = 1, \ldots, n$, where

$$a_{ij} = \int_0^1 \varphi_i'(x)\varphi_j'(x)\,dx \tag{1.25}$$

$$f_i = \int_0^1 \varphi_i(x)\,dx \tag{1.26}$$

Example : Consider a triangulation of $[0, 1]$ into four elements of width $h = 1/4$ between the node points $x_i = ih$, $i = 0, \ldots, 4$. We then get a solution space of dimension $n = 3$, and basis functions $\varphi_1, \varphi_2, \varphi_3$. When calculating the entries of A, note that

- A is symmetric, that is, $a_{ij} = a_{ji}$
- A is tridiagonal, that is, $a_{ij} = 0$ whenever $|i - j| > 1$
- For our equidistant triangulation, $a_{ii} = a_{jj}$ and $a_{i,i+1} = a_{j,j+1}$

This gives

$$a_{11} = 4 \cdot 4 \cdot \frac{1}{4} + (-4)(-4)\frac{1}{4} = 8 \tag{1.27}$$

$$a_{12} = (-4) \cdot 4 \cdot \frac{1}{4} = -4 \tag{1.28}$$

$$a_{22} = a_{33} = a_{11} = 8 \tag{1.29}$$

$$a_{21} = a_{12} = a_{23} = a_{32} = -4 \tag{1.30}$$

$$a_{13} = a_{31} = 0 \tag{1.31}$$

and

$$f_1 = f_2 = f_3 = \int_0^1 \varphi_1(x)\,dx = \frac{1}{4} \tag{1.32}$$

and the linear system becomes

$$\begin{bmatrix} 8 & -4 & 0 \\ -4 & 8 & -4 \\ 0 & -4 & 8 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} \tag{1.33}$$

with solution

$$u = (u_1, u_2, u_3)^T = (3/32, 1/8, 3/32)^T \tag{1.34}$$

Note that the discretization exactly matches the one obtained with finite differences and the 2nd order 3-point stencil:

$$\frac{1}{(1/4)^2}\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \tag{1.35}$$

1.4 Neumann boundary conditions

The Dirichlet conditions $u(0) = u(1) = 0$ where enforced directly into the approximation space V_h. In the finite element method, a Neumann condition (or *natural* condition) is instead implemented by modifying the variational formulation. Consider the model problem

$$-u''(x) = f(x) \text{ for } x \in (0,1) \tag{1.36}$$

$$u(0) = 0 \tag{1.37}$$

$$u'(1) = g \tag{1.38}$$

The function space is only enforcing the Dirichlet condition at the left end point:

$$V_h = \{v \in C^0([0,1]) : v|_K \in \mathbb{P}_1(K)\ \forall K \in T_h,\ v(0) = 0\}, \tag{1.39}$$

which in our example gives an additional degree of freedom. The Neumann condition appears in the formulation after integration by parts:

$$\int_0^1 \hat{u}'(x)v'(x)\,dx - \left[\hat{u}'(x)v(x)\right]_0^1 = \int_0^1 \hat{u}'(x)v'(x)\,dx - \hat{u}'(1)v(1) + \hat{u}'(0)v(0) \tag{1.40}$$

$$= \int_0^1 \hat{u}'(x)v'(x)\,dx - gv(1) \tag{1.41}$$

since $v(0) = 0$ and $u_h'(1) = g$. This leads to the final formulation: Find $u_h \in V_h$ such that

$$\int_0^1 u_h'(x)v'(x)\,dx = \int_0^1 f(x)v(x)\,dx + gv(1), \qquad \forall v \in V_h \tag{1.42}$$

Example : With our previous triangulation, we now get another basis function φ_4. For simplicity, set $f = g = 1$. All the matrix entries and right-hand side values are then identical, and we only calculate the new values:

$$a_{44} = (4)^2\frac{1}{4} = 4 \tag{1.43}$$

$$a_{34} = a_{43} = (-4)\cdot 4 \cdot \frac{1}{4} = -4 \tag{1.44}$$

$$f_4 = \frac{1}{8} + g \cdot 1 = \frac{9}{8} \tag{1.45}$$

The linear system then becomes

$$\begin{bmatrix} 8 & -4 & 0 & 0 \\ -4 & 8 & -4 & 0 \\ 0 & -4 & 8 & -4 \\ 0 & 0 & -4 & 4 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 9/8 \end{bmatrix} \tag{1.46}$$

with solution

$$u = (u_1, u_2, u_3, u_4)^T = (15/32, 7/8, 39/32, 3/2)^T. \tag{1.47}$$

1.5 Inhomogeneous Dirichlet problems

Use two spaces: V_h, which enforces the inhomogeneous Dirichlet condition, for the solution u_h, and V_h^0, which enforces homogeneous Dirichlet conditions, for the test function v. Consider for example the model problem

$$-u''(x) = f, \text{ for } x \in (0,1) \tag{1.48}$$
$$u(0) = 0 \tag{1.49}$$
$$u(1) = 1 \tag{1.50}$$

with exact solution $u(x) = x(3-x)/2$. Use the spaces

$$V_h = \{v \in C^0([0,1]) : v|_K \in \mathbb{P}_1(K)\ \forall K \in T_h,\ v(0) = 0, v(1) = 1\}, \tag{1.51}$$
$$V_h^0 = \{v \in C^0([0,1]) : v|_K \in \mathbb{P}_1(K)\ \forall K \in T_h,\ v(0) = 0, v(1) = 0\}. \tag{1.52}$$

The FEM formulation becomes: Find $u_h \in V_h$ such that

$$\int_0^1 u_h' v'\, dx = \int_0^1 f v\, dx, \qquad \forall v \in V_h^0. \tag{1.53}$$

Note that the function space V_h is not a linear vector space, due to the inhomogeneous constraint. In practice, Dirichlet conditions are implemented by first considering the all-Neumann problem, and enforcing the Dirichlet conditions directly into the resulting system of equations.

Example : For the model problem (1.48)-(1.50), first consider the corresponding all-Neumann problem

$$-u''(x) = f, \text{ for } x \in (0,1) \tag{1.54}$$
$$u'(0) = u'(1) = 0 \tag{1.55}$$

with the solution space

$$V_h = \{v \in C^0([0,1]) : v|_K \in \mathbb{P}_1(K)\ \forall K \in T_h\}. \tag{1.56}$$

In our example with four elements of size $h = 1/4$, this gives 5 degrees of freedom and the resulting linear system (with $f = 1$):

$$\begin{bmatrix} 4 & -4 & 0 & 0 & 0 \\ -4 & 8 & -4 & 0 & 0 \\ 0 & -4 & 8 & -4 & 0 \\ 0 & 0 & -4 & 8 & -4 \\ 0 & 0 & 0 & -4 & 4 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} = \begin{bmatrix} 1/8 \\ 1/4 \\ 1/4 \\ 1/4 \\ 1/8 \end{bmatrix} \tag{1.57}$$

This system is singular (since the solution is undetermined up to a constant). We can now impose the Dirichlet conditions $u(0) = 0$ and $u(1) = 1$ directly by replacing the corresponding equations:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -4 & 8 & -4 & 0 & 0 \\ 0 & -4 & 8 & -4 & 0 \\ 0 & 0 & -4 & 8 & -4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 1/4 \\ 1/4 \\ 1/4 \\ 1 \end{bmatrix} \tag{1.58}$$

We can keep the symmetry of the matrix by eliminating the entries below/above the diagonal of the Dirichlet variables:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 8 & -4 & 0 & 0 \\ 0 & -4 & 8 & -4 & 0 \\ 0 & 0 & -4 & 8 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 1/4 - (-4) \cdot 0 \\ 1/4 \\ 1/4 - (-4) \cdot 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1/4 \\ 1/4 \\ 17/4 \\ 1 \end{bmatrix} \tag{1.59}$$

with solution $u = (u_1, \ldots, u_5)^T = (0, 11/32, 5/8, 27/32, 1)^T$. If necessary, the Dirichlet degress of freedom can be removed from the system, to obtain the smaller system of equations

$$\begin{bmatrix} 8 & -4 & 0 \\ -4 & 8 & -4 \\ 0 & -4 & 8 \end{bmatrix} \begin{bmatrix} u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} 1/4 \\ 1/4 \\ 17/4 \end{bmatrix} \tag{1.60}$$

1.6 The stamping method ("Assembly")

Consider a single element e_k, and its *local basis functions* $\mathcal{H}_i^k(x)$, $j = 1, 2$, given by the restriction of the *global basis functions* $\varphi_j(x)$ to the element. The connection between the local indices i and the global indices j are given by a *mesh representation*. For simplex elements, we use the form

- p: $N \times D$ node coordinates
- t: $T \times (D + 1)$ element indices

Here, row k of t is the *local-to-global* mapping for element k. The local basis functions are:

- Defined only inside the element e_k
- Polynomials of degree 1

We can then define an *elemental matrix* A^k, or a *local stiffness matrix*, by again considering only the contribution to the global stiffmatrix from element e_k:

$$A_{ij}^k = \int_{e_k} (\mathcal{H}_i^k)' \cdot (\mathcal{H}_j^k)' \, dx, \qquad i = 1, 2, \ j = 1, 2 \tag{1.61}$$

and, similarily, an *elemental load vector* $\boldsymbol{b}^k$ by

$$b_i^k = \int_{e_k} f(x) \cdot (\mathcal{H}_i^k) \, dx, \qquad i = 1, 2 \tag{1.62}$$

In the so-called *stamping method*, or the *assembly* process, each element matrix and load vector are added at the corresponding global position in the global stiffness matrix and right-hand side vector:

$A = 0$, $\boldsymbol{b} = 0$
for all elements k
    Calculate A^k, $\boldsymbol{b}^k$
    $A(t(k,:), t(k,:))$ += A^k
    $\boldsymbol{b}(t(k,:))$ += $\boldsymbol{b}^k$
end for

Example : Equidistant 1D mesh, element width h, piece-wise linears, $f = 1$. The elemental matrix and load vector are

$$A^k = \frac{1}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \qquad \boldsymbol{b}^k = \frac{h}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{1.63}$$

The mesh is represented by the arrays

$$p = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}, \qquad t = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ \vdots & \\ n & n+1 \end{bmatrix} \tag{1.64}$$

The stamping method gives the global matrices:

$$\frac{1}{h}\begin{bmatrix} 1 & -1 & & & \\ -1 & 1 & & & \\ & & & & \\ & & & & \\ & & & & \end{bmatrix} \to \frac{1}{h}\begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 1 & & \\ & & & & \\ & & & & \end{bmatrix} \to \frac{1}{h}\begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 1 & \\ & & & & \end{bmatrix}$$

$$\to \frac{1}{h}\begin{bmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & -1 & 2 & -1 & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{bmatrix} = A \tag{1.65}$$

and, similarily, the right-hand side vector

$$\boldsymbol{b} = \frac{h}{2}\begin{bmatrix} 1 & 2 & \cdots & 2 & 1 \end{bmatrix}^T \tag{1.66}$$

1.7 Higher order

Introduce the space of continuous piece-wise quadratics:

$$V_h = \{v \in C^0(\Omega) : v|_K \in \mathbb{P}_2(K)\ \forall K \in T_h\}. \tag{1.67}$$

Parameterize by adding degrees of freedom at element midpoints. Each element then has three local nodes: x_1^k, x_2^k, x_3^k, and three local basis functions $\mathcal{H}_1^k(x), \mathcal{H}_2^k(x), \mathcal{H}_3^k(x)$. These are determined by solving for the polynomial coefficients in

$$\mathcal{H}_i^k = a_i + b_i x + c_i x^2, \qquad i = 1, 2, 3 \tag{1.68}$$

and requiring that $\mathcal{H}_i^k(x_j) = \delta_{ij}$, for each $i, j = 1, 2, 3$. The leads to the linear system of equations $VC = I$:

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{bmatrix}\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.69}$$

which gives the coefficients $C = V^{-1}$. For example, with $x_1 = 0$, $x_2 = h/2$, $x_3 = h$, we get

$$\mathcal{H}_1(x) = \frac{2}{h^2}(x - \frac{h}{2})(x - h) \tag{1.70}$$

$$\mathcal{H}_2(x) = -\frac{4}{h^2}x(x - h) \tag{1.71}$$

$$\mathcal{H}_3(x) = \frac{2}{h^2}x(x - \frac{h}{2}) \tag{1.72}$$

The corresponding element matrix and the element load can then be calculated as before:

$$A^k_{ij} = \int_0^h \mathcal{H}^k_i(x)'\mathcal{H}^k_j(x)'\,dx, \qquad b^k_i = \int_0^h f(x)\mathcal{H}^k_i(x)\,dx \tag{1.73}$$

for $i, j = 1, 2, 3$, which gives

$$A^k = \frac{1}{3h}\begin{bmatrix} 7 & -8 & 1 \\ -8 & 16 & -8 \\ 1 & -8 & 7 \end{bmatrix}, \qquad \boldsymbol{b}^k = \frac{h}{6}\begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} \tag{1.74}$$

For a global element with two elements of width h, the stamping method gives the stiff matrix and right-hand side vector

$$A = \frac{1}{3h}\begin{bmatrix} 7 & -8 & 1 & & \\ -8 & 16 & -8 & & \\ 1 & -8 & 7+7 & -8 & 1 \\ & & -8 & 16 & -8 \\ & & 1 & -8 & 7 \end{bmatrix}, \qquad \boldsymbol{b} = \frac{h}{6}\begin{bmatrix} 1 \\ 4 \\ 1+1 \\ 4 \\ 1 \end{bmatrix} \tag{1.75}$$

1.8 Numerical Quadrature

For more general cases, the integrals cannot be computed analytically. We then use Gaussian quadrature rules of the following form (not necessarily the same f as above):

$$\int_{-1}^{1} f(x)\,dx \approx \sum_{i=1}^{n} w_i f(x_i) \tag{1.76}$$

when x_i, w_i, $i = 1, \ldots, n$ are specified points and weights. By choosing x_i as the zeros of the nth Legendre polynomial, and the weights such that the rule exactly integrates polynomials up to degree $n-1$, the rule (1.76) gives exact integration for polynomials of degree $\leq 2n-1$.

For example, the following rule has $n = 2$ and degree of precision $2 \cdot 2 - 1 = 3$:

$$\int_{-1}^{1} f(x)\,dx \approx f(-\frac{1}{\sqrt{3}}) + f(\frac{1}{\sqrt{3}}) \tag{1.77}$$

In our quadratic element example, the integrals for the element matrix were products of derivatives of quadratics, that is, quadratics. Therefore they can be exactly evaluated using this rule:

$$A^k_{11} = \left[\text{Set } f(x) = \left[\frac{2}{h^2}\left(2x - \frac{3h}{2}\right)\right]^2\right] = \int_0^h f(x)\,dx \tag{1.78}$$

$$= \frac{h}{2}\left[f(\frac{h}{2} - \frac{h}{2\sqrt{3}}) + f(\frac{h}{2} + \frac{h}{2\sqrt{3}})\right] = \cdots = \frac{7}{3h} \tag{1.79}$$

2 The Poisson Problem in 2-D

- Consider the problem

$$-\nabla^2 u = f \text{ in } \Omega \tag{2.1}$$
$$n \cdot \nabla u = g \text{ on } \Gamma \tag{2.2}$$

  for a domain Ω with boundary Γ

- Introduce the space of piecewise linear continuous functions on a mesh T_h:

$$V_h = \{v \in C^0(\Omega) : v|_K \in \mathbb{P}_1(K) \ \forall K \in T_h\}. \tag{2.3}$$

- Seek solution $u_h \in V_h$, multiply by a test function $v \in V_h$, and integrate:

$$\int_\Omega -\nabla^2 u_h v \, dx = \int_\Omega f v \, dx \tag{2.4}$$

- Apply the divergence theorem and use the Neumann condition, to get the Galerkin form

$$\int_\Omega \nabla u_h \cdot \nabla v \, dx = \int_\Omega f v \, dx + \oint_\Gamma g v \, ds \tag{2.5}$$

2.1 Finite Element Formulation

- Expand in basis $u_h = \sum_i u_{h,i}\varphi_i(x)$, insert into the Galerkin form, and set $v = \varphi_i$, $i = 1, \ldots, n$:

$$\int_\Omega \left[\sum_{j=1}^{n} u_{h,j} \nabla \varphi_j \right] \cdot \nabla \varphi_i \, dx = \int_\Omega f \varphi_i \, dx + \oint_\Gamma g \varphi_i \, ds \tag{2.6}$$

  Switch order of integration and summation to get the finite element formulation:

$$\sum_{j=1}^{n} A_{ij} u_{h,j} = b_i, \quad \text{or} \quad A\boldsymbol{u} = \boldsymbol{b} \tag{2.7}$$

  where

$$A_{ij} = \int_\Omega \nabla \varphi_i \cdot \nabla \varphi_j \, dx, \quad b_i = \int_\Omega f \varphi_i \, dx + \oint_\Gamma g \varphi_i \, ds \tag{2.8}$$

2.2 Discretization

- Find a tringulation of the domain Ω into triangular elements T^k, $k = 1, \ldots, K$ and nodes $\boldsymbol{x}_i$, $i = 1, \ldots, n$

- Consider the space V_h of continuous functions that are linear within each element

- Use a nodal basis $V_h = \text{span}\{\varphi_1, \ldots, \varphi_n\}$ defined by

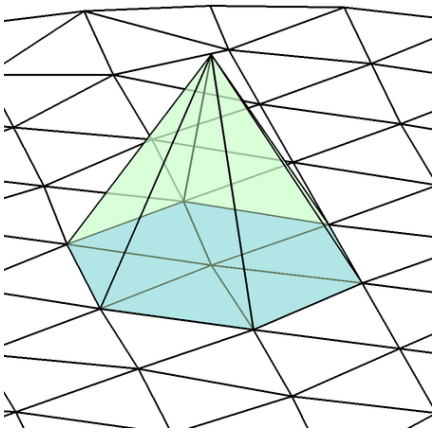
$$\varphi_i \in V_h, \quad \varphi_i(\boldsymbol{x}_j) = \delta_{ij}, \quad 1 \le i, j \le n \tag{2.9}$$

- A function $v \in V_h$ can then be written

$$v = \sum_{i=1}^{n} v_i \varphi_i(\boldsymbol{x}) \tag{2.10}$$

with the nodal interpretation

$$v(\boldsymbol{x}_j) = \sum_{i=1}^{n} v_i \varphi_i(\boldsymbol{x}_j) = \sum_{i=1}^{n} v_i \delta_{ij} = v_j \tag{2.11}$$



2.3 Local Basis Functions

- Consider a triangular element T^k with local nodes $\boldsymbol{x}_1^k, \boldsymbol{x}_2^k, \boldsymbol{x}_3^k$
- The local basis functions $\mathcal{H}_1^k, \mathcal{H}_2^k, \mathcal{H}_3^k$ are linear functions:

$$\mathcal{H}_\alpha^k = c_\alpha^k + c_{x,\alpha}^k x + c_{y,\alpha}^k y, \quad \alpha = 1,2,3 \tag{2.12}$$

with the property that $\mathcal{H}_\alpha^k(x_\beta) = \delta_{\alpha\beta}$, $\beta = 1,2,3$

- This leads to linear systems of equations for the coefficients:

$$\begin{pmatrix} 1 & x_1^k & y_1^k \\ 1 & x_2^k & y_2^k \\ 1 & x_3^k & y_3^k \end{pmatrix} \begin{pmatrix} c_\alpha^k \\ c_{x,\alpha}^k \\ x_{y,\alpha}^k \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \text{ or } \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \text{ or } \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{2.13}$$

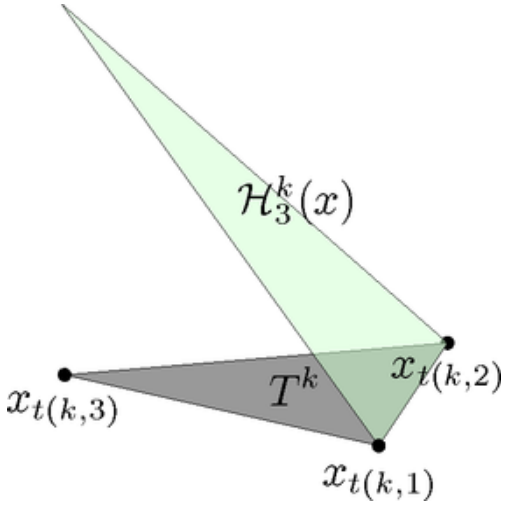
or $C = V^{-1}$ with coefficient matrix C and Vandermonde matrix V

2.4 Elementary Matrices and Loads

- The elementary matrix for an element T^k becomes

$$A_{\alpha\beta}^k = \int_{T^k} \frac{\partial \mathcal{H}_\alpha^k}{\partial x}\frac{\partial \mathcal{H}_\beta^k}{\partial x} + \frac{\partial \mathcal{H}_\alpha^k}{\partial y}\frac{\partial \mathcal{H}_\beta^k}{\partial y}\, dx \tag{2.14}$$

$$= \text{Area}^k (c_{x,\alpha}^k c_{x,\beta}^k + c_{y,\alpha}^k c_{y,\beta}^k), \quad \alpha,\beta = 1,2,3 \tag{2.15}$$



- The elementary load becomes

$$b_\alpha^k = \int_{T^k} f\, \mathcal{H}_\alpha^k\, dx \tag{2.16}$$

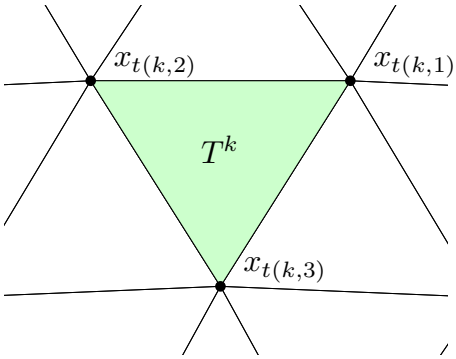
$$= (\text{if } f \text{ constant}) \tag{2.17}$$

$$= \frac{\text{Area}^k}{3} f, \quad \alpha = 1,2,3 \tag{2.18}$$

2.5 Assembly, The Stamping Method

- Assume a local-to-global mapping $t(k, \alpha)$, giving the global node number for local node number α in element k
- The global linear system is then obtained from the elementary matrices and loads by the stamping method:

$A = 0,\ b = 0$
for $k = 1, \ldots, K$
$\quad A(t(k,:), t(k,:)) = A(t(k,:), t(k,:)) + A^k$
$\quad b(t(k,:)) = b(t(k,:)) + b^k$



2.6 Dirichlet Conditions

- Suppose Dirichlet conditions $u = u_D$ are imposed on part of the boundary Γ_D
- Enforce $u_{h,i} = u_D$ for all nodes i on Γ_D directly in the linear system of equations:

$$\begin{array}{c} \\ i \end{array} \overset{\begin{array}{c} \\ i \end{array}}{\begin{pmatrix} & & & & & & \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ & & & & & & \end{pmatrix}} u_h = \begin{pmatrix} \\ u_D \\ \\ \end{pmatrix} \tag{2.19}$$

- Eliminate below/above diagonal of the Dirichlet nodes to retain symmetry

3 FEM Theory

3.1 Variational and minimization formulations

Consider the Dirichlet problem (D)

$$-\nabla^2 u = f \quad \text{in } \Omega \tag{3.1}$$
$$u = 0 \quad \text{on } \Gamma \tag{3.2}$$

Introduce the notation

$$a(u,v) = \int_\Omega \nabla u \cdot \nabla v \, dx \qquad \text{(bilinear form)} \tag{3.3}$$

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

Note that

$$a(u_1 + u_2, v) = a(u_1, v) + a(u_2, v) \qquad \text{(bilinearity)} \tag{3.5}$$

$$\ell(v_1 + v_2) = \ell(v_1) + \ell(v_2) \qquad \text{(linearity)} \tag{3.6}$$

$$a(u,v) = a(v,u) \qquad \text{(symmetry)} \tag{3.7}$$

Also,

$$a(u,u) = \int_\Omega \|\nabla u\|_2^2 \, dx \geq 0, \tag{3.8}$$

and if $u = 0$ on Γ, then equality only for $u = 0$. This can then be used to define the *energy norm*

$$|||u||| \equiv \sqrt{a(u,u)} \tag{3.9}$$

Now, define the spaces

$$L_2(\Omega) = \{v \ : \ v \text{ is defined on } \Omega \text{ and } \int_\Omega v^2 \, dx < \infty\} \tag{3.10}$$

$$H^1(\Omega) = \{v \in L_2(\Omega) \ : \ \frac{\partial v}{\partial x_i} \in L_2(\Omega) \text{ for } u = 1, \ldots, d\} \tag{3.11}$$

$$H_0^1(\Omega) = \{v \in H^1(\Omega) \ : \ v = 0 \text{ on } \Gamma\} \tag{3.12}$$

where d is the number of space dimensions. Let $V = H_0^1$. The solution u to (D) then satisfies a corresponding *variational problem* (V):

$$\text{Find } u \in V \text{ s.t. } a(u,v) = \ell(v) \quad \forall v \in V \tag{3.13}$$

This is also called the *weak form* of (D), with a *weak solution* u. Note that a solution u of (D) is also a solution of (V), since for any $v \in V$:

$$\int_\Omega -\nabla^2 u v \, dx = \int_\Omega f v \, dx \tag{3.14}$$

$$\int_\Omega \nabla u \cdot \nabla v \, dx - \oint_\Gamma \boldsymbol{n} \cdot \nabla u \, v \, ds = \int_\Omega f v \, dx \tag{3.15}$$

$$\int_\Omega \nabla u \cdot \nabla v \, dx = \int_\Omega f v \, dx \tag{3.16}$$

$$a(u,v) = \ell(v) \tag{3.17}$$

where we used the fact that $v = 0$ on Γ. The reverse can also be shown, (V) $\Longrightarrow$ (D), if u is sufficiently regular.

Now consider the minimization problem (M):

$$\text{Find } u \in V \text{ s.t. } F(u) \leq F(v) \quad \forall v \in V, \quad \text{where } F(v) = \frac{1}{2} a(v,v) - \ell(v) \tag{3.18}$$

Suppose u solves (V). Then it also solves (M), since for any $v \in V$, set $w = v - u \in V$, and

$$F(v) = F(u+w) = \frac{1}{2}a(u+w, u+w) - \ell(u+w) \tag{3.19}$$

$$= \frac{1}{2}a(u,u) + \frac{1}{2}a(u,w) + \frac{1}{2}a(w,v) + \frac{1}{2}a(w,w) - \ell(u) - \ell(w) \tag{3.20}$$

$$= \underbrace{\frac{1}{2}a(u,u) - \ell(u)}_{F(u)} + \underbrace{a(u,w) - \ell(w)}_{=0 \ \forall w \in V} + \underbrace{\frac{1}{2}a(w,w)}_{\geq 0} \geq F(u) \tag{3.21}$$

The reverse can also easily be shown, that is, (M) $\Longrightarrow$ (V).

3.2 FEM formulation

Now define the FEM formulation

$$\text{Find } u_h \in V_h \text{ s.t. } a(u_h, v) = \ell(v) \quad \forall v \in V_h \tag{3.22}$$

for some finite-dimensional subspace V_h of V. Recall that the solution u to (V) satisfies $a(u,v) = \ell(v)$ for all $v \in V$, and in particular for all $v \in V_h$, we have

$$a(u - u_h, v) = 0 \quad \forall v \in V_h, \tag{3.23}$$

or $a(e,v) = 0$ with the *error* $e = u - u_h$. This means that the error also satisfies a corresponding minimization problem (M), with $\ell(v) = 0$, which leads to the property:

$$\frac{1}{2}a(v,v) - 0 = \frac{1}{2}|||v|||^2 \text{ is minimized by } v = u - u_h = e \tag{3.24}$$

In other words, for any $w_h \in V_h$, $|||u - w_h||| \geq |||u - u_h|||$, or the FEM formulation finds the best possible solution in the energy norm.

More generally, for any Hilbert space V with corresponding norm $\|\cdot\|_V$, we require that

(i) $a(\cdot,\cdot)$ is symmetric

(ii) $a(\cdot,\cdot)$ is *continuous*, i.e., $\exists \gamma > 0$ s.t. $|a(v,w)| \leq \gamma \|v\|_V \|w\|_V$, $\forall v, w \in V$

(iii) $a(\cdot,\cdot)$ is *V-elliptic*, i.e., $\exists \alpha > 0$ s.t. $a(v,v) \geq \alpha \|v\|_V^2$, $\forall v \in V$

(iv) $\ell(\cdot)$ is *continuous*, i.e., $\exists \Lambda > 0$ s.t. $|L(v)| \leq \Lambda \|v\|_V$, $\forall v \in V$

The energy norm $|||\cdot|||$ is then equivalent to $\|\cdot\|_V$, that is, $\exists c, C > 0$ s.t.

$$c\|v\|_V \leq |||v||| \leq C\|v\|_V \quad \forall v \in V \tag{3.25}$$

for example with $c = \sqrt{\alpha}$ and $C = \sqrt{\gamma}$. This gives

$$\|u - u_j\|_V \leq \frac{\gamma}{\alpha}\|u - v\|_V \quad \forall v \in V_h \tag{3.26}$$

3.3 Interpolants

Consider a 1-D function $w \in V$ and its *linear interpolant* $w_I \in V_h$, where V_h is a space of piecewise linear functions, defined by $w_I(x_i) = w(x_i)$, or

$$w_I(x) = \sum_{i=1}^{n} w(x_i)\varphi_i(x) \tag{3.27}$$

To find a bound on the energy norm of the difference $w - w_I$, we first bound the derivative. Consider a single element, and note that the difference $w - w_I$ is zero at the endpoints. A point $x = z$ must then exists, such that

$$[w(z) - w_I(z)]' = 0 \tag{3.28}$$

For any other point x inside the element, we then have from the fundamental theorem of calculus,

$$[w(x) - w_I(x)]' = \int_z^x [w(y) - w_I(y)]'' \, dy \tag{3.29}$$

but $w_I(y)$ is linear, so $w_I''(y) = 0$ and

$$[w(x) - w_I(x)]' = \int_z^x w''(y) \, dy \tag{3.30}$$

If the element width is h, this gives the bound

$$\left|[w(x) - w_I(x)]'\right| \leq h \max |w''| \tag{3.31}$$

A bound for the energy norm of $w - w_I$ can then be derived as follows. Note that the number of elements T can be written as C_1/h, for some constant C_1.

$$|||w - w_I|||^2 = \sum_{k=1}^{T} \int_{e_k} [(w - w_I)']^2 \, dx \leq \frac{C_1}{h} \cdot h \cdot h^2 \cdot [\max |w''|]^2 \tag{3.32}$$

$$= C_1 h^2 [\max |w''|]^2 \tag{3.33}$$

or

$$|||w - w_I||| \leq Ch \max |w''| \tag{3.34}$$

3.4 FEM error bounds

For a finite element solution u_h, the optimality in the energy norm and the boundary on the interpolat leads to

$$|||u - u_h||| \leq |||u - u_I||| \leq Kh \max |u''| = \mathcal{O}(h) \tag{3.35}$$

This implies convergence in the energy norm, at a linear rate w.r.t. the element sizes h.

This result does not imply that the solution itself in, for example, the L_2-norm is quadratically convergent. However, using other techniques it can be shown that

$$\|u - u_h\|_{L_2} = \mathcal{O}(h^2) \tag{3.36}$$

under suitable assumptions.

3.5 2-D interpolants

For piecewise interpolant on a triangular mesh in 2-D, it can be shown that

$$|||w - w_I||| \leq Ch\|w\|_{H^2} \tag{3.37}$$

$$\|w - w_I\|_{L_2} \leq Ch^2\|w\|_{H^2} \tag{3.38}$$

which again leads to the energy norm bound for a finite element solution u_h

$$|||u - u_h||| \leq C_1 h\|u\|_{H^2} \tag{3.39}$$

However, two new problems appear in 2-D:

- "Bad elements", consider e.g. interpolation of the function $w(x, y) = x^2$ on a triangle with corners at $(x, y) = (-1, 0)$, $(1, 0)$, and $(0, \varepsilon)$. Since $w(-1, 0) = w(1, 0) = 1$, linear interpolation gives $w_I(0, 0) = 1$. But $w(0, \varepsilon) = 0$, so the derivative along the y-axis:

$$\frac{\partial w_I}{\partial y}(0, 0) = -\frac{1}{\varepsilon} \to -\infty \text{ as } \theta \to 180^\circ \tag{3.40}$$

  where θ is the top angle. This does not affect the linear convergence of the energy norm as $h \to 0$, since the worst angle θ remains fixed, but it can cause a large constant C_1 if the mesh contains bad elements.

- Lack of regularity of the solution, or geometry-induced singularities. For example on a domain with convex corners, it can be shown that $\|u\|_{H^2}$ is not bounded which reduces the convergence rate of the FEM solution.

4 Some extensions

4.1 Higher order elements in 2D

For the piecewise linear case on triangular meshes, we parameterized the space of continuous functions as:

- Representing a solution u_i at the global nodes $\boldsymbol{x}_i$
- Using a uniquely defined linear function $u(x, y) = a + bx + cy$ on each triangle, from the 3 corner nodes
- Continuity is enforced since the node values are single-valued, and along an edge between two elements there is a uniquely defined linear function.

This extends natually to piecewise quadratic elements:

- Introduce the edge midpoints as global nodes
- Each triangle is now associated with 6 nodes, with determines a unique quadratic function $u(x, y) = a + bx + cy + dx^2 + exy + fy^2$
- Continuity obtained since along an edge, there is a uniquely defined quadratic

This can be generalized to *Lagrange elements* of any degree p in any dimension D, using $\binom{p+D}{D}$ nodes in a regular pattern.

4.2 Other PDEs

Consider the more general linear parabolic PDE:

$$\rho\frac{\partial u}{\partial t} - \nabla\cdot(D\nabla u + \boldsymbol{\alpha}u) + au = f \quad \text{in } \Omega \tag{4.1}$$

with boundary conditions

$$\boldsymbol{n}\cdot(D\nabla u + \boldsymbol{\alpha}u) = g \qquad \text{on } \Gamma_N \tag{4.2}$$

$$u = u_D \qquad \text{on } \Gamma_D \tag{4.3}$$

Note that the fields rho, D, $\boldsymbol{\alpha}$, g, and u_D in general are time and space dependent.

A standard FEM formulation for (4.1) is: Find $u_h \in V_h$ s.t.

$$\int_\Omega \rho u_h v\,dx + \int_\Omega (D\nabla u_h + \boldsymbol{\alpha}u_h)\cdot\nabla v\,dx + \int_\Omega a u_h v\,dx = \int_\Omega fv\,dx + \int_{\Gamma_N} gv\,ds \tag{4.4}$$

for all $v \in V_{0,h}$, where V_h is an appropriate finite dimensional space satisfying the Dirichlet conditions. Considering the all-Neumann problem, discretization with basis functions $\varphi_i(\boldsymbol{x})$ leads to the linear system of ODEs

$$M_\rho\dot{\boldsymbol{u}} + K\boldsymbol{u} + C\boldsymbol{u} + M_a\boldsymbol{u} = \boldsymbol{f} + \boldsymbol{g} \tag{4.5}$$

where

$$M_{\rho,ij} = \int_\Omega \rho\varphi_i\varphi_j\,dx \tag{4.6}$$

$$M_{a,ij} = \int_\Omega a\varphi_i\varphi_j\,dx \tag{4.7}$$

$$K_{ij} = \int_\Omega (D\nabla\varphi_j)\cdot\nabla\varphi_i\,dx \tag{4.8}$$

$$C_{ij} = \int_\Omega (\boldsymbol{\alpha}\varphi_j)\cdot(\nabla\varphi_i)\,dx \tag{4.9}$$

$$f_i = \int_\Omega f\varphi_i\,dx \tag{4.10}$$

$$g_i = \oint_\Gamma g\varphi_i\,ds \tag{4.11}$$

This can be integrated in time using method of lines, with e.g. a BDF method or an implicit Runge-Kutta. Note that explicit methods can be used, but they require inversion of M_ρ and will put stability constrains on the timestep.