Finite Element Method

The finite element method is an alternative to the finite difference method from the previous lectures. It is usually better suited for domains with complex geometries.

Reminders:

1) Integration by parts

For $u, v : [0,1] \to \mathbb{R}$, we have

$$\int_0^1 u(x)\, v'(x)\, dx = [u\, v]_0^1 - \int_0^1 u'(x)\, v(x)\, dx .$$

2) Inner product for space of functions:

For $u, v : [0,1] \to \mathbb{C}$, we have an inner product given by

$$\langle u, v \rangle = \int_0^1 \bar{u}(x)\, v(x)\, dx$$

and the induced norm

$$\|u\|_2^2 = \langle u, u \rangle = \int_0^1 \underbrace{\bar{u}(x)\, u(x)}_{|u(x)|^2} dx$$

3) Logarithmic norm for a matrix A:

$$\mu[A] = \max_{x \neq 0} \frac{\langle Ax, x \rangle}{\langle x, x \rangle} \quad \text{with } \langle x, y \rangle = x^T y$$

and then $\langle Ax, x \rangle \leq \mu[A]\, \|x\|^2$

Note: $\mu[A] = \max_{x \neq 0} \frac{\mathrm{Re}\langle x, Ax \rangle}{\langle x, x \rangle} = \max_{x \neq 0} \frac{\langle Ax, x \rangle}{\langle x, x \rangle}$

as we only consider real-valued problems in the following.

Integration by parts

Next, we want to generalize the logarithmic norm to more general objects than matrices:
We want to find the logarithmic norm of the second derivative $\frac{d^2}{dx^2}$, i.e. we want to find the constant $\mu\left[\frac{d^2}{dx^2}\right]$ such that

$$\langle u, u'' \rangle \leq \mu\left[\frac{d^2}{dx^2}\right] \|u\|_2^2$$

for $u \in C_0^2[0,1] = \{u \in C^2(0,1) : u(0) = u(1) = 0\}$.
More precisely, we have:

$$\begin{aligned}\langle u, u'' \rangle &= \int_0^1 u\, u''\, dx \\ &= -\int_0^1 u'\, u'\, dx + \underbrace{[u\, u']_0^1}_{=0} \\ &= -\langle u', u' \rangle = -\|u'\|_2^2.\end{aligned}$$

Tool:

Sobolev's Lemma (Poincare's inequality)

For all functions u with $u(0) = u(1) = 0$, it holds that

$$\|u'\|_2 \geq \pi \|u\|_2.$$

Proof (idea)
Using Parseval's theorem (Fourier expansion), we can write

$$u(x) = \sqrt{2} \sum_{k=1}^{\infty} c_k \sin(k\pi x)$$

$$u'(x) = \pi\sqrt{2} \sum_{k=1}^{\infty} k\, c_k \cos(k\pi x)$$

which then shows that

$$\|u'\|_2 \geq \pi \|u\|_2 .$$

Note that for $u(x) = \sin(\pi x)$, we have an equality.

Note: Parseval's identity for a 2π periodic function f □

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos(kx) + b_k \sin(kx) \right).$$

Back to previous problem:

We have

$$\langle u, u'' \rangle = - \|u'\|_2^2 \leq -\pi^2 \|u\|_2^2 .$$

Theorem

The logarithmic norm of d^2/dx^2 on $C_0^2[0,1]$ is

$$\mu_2 \left[d^2/dx^2 \right] = - \pi^2 .$$

Corollary

The 2p BVP $u''(x) = f(x)$ with $u(0) = u(1) = 0$ has a unique solution with $\|u\|_2 \leq \frac{\|f\|_2}{\pi^2}$.

Proof

Since $\mu_2 \left[\frac{d^2}{dx^2} \right] = - \pi^2 < 0$, we can apply the uniform monotonicity theorem. Note that we have proved it for matrices but the same result indeed holds for operators too and find that $\left(\frac{d^2}{dx^2} \right)^{-1}$ is nonsingular and fulfills

$$\left\| \left(d^2/dx^2 \right)^{-1} \right\|_2 \leq - \frac{1}{\mu_2 \left[d^2/dx^2 \right]} = \frac{1}{\pi^2} .$$

Linear operators and adjoint operators

Definition

For an operator A, the adjoint operator A^* is given by $\langle v, Au \rangle = \langle A^* v, u \rangle$.

An operator A is self-adjoint (symmetric) if $A = A^*$.

An operator A is anti-self-adjoint (anti-symmetric) if $A^* = -A$.

Examples

1) For a matrix A, the adjoint is given by the transposed matrix A^T.

$$\langle v, Au \rangle = v^T Au = (A^T v)^T u = \langle A^T v, u \rangle$$

A matrix is self-adjoint if $A = A^T$.

2) $\mathcal{L} = \frac{d^2}{dx^2}$ is symmetric on $C_0^2[0,1]$:

We use integration by parts twice to find that

$$\langle v, \mathcal{L}u \rangle = \langle v, u'' \rangle = -\langle v', u' \rangle$$
$$= \langle v'', u \rangle = \langle \mathcal{L}v, u \rangle = \langle \mathcal{L}^* v, u \rangle$$

3) $\mathcal{L} = \frac{d}{dx}\left(p(x) \frac{d}{dx} \right) + q(x)$ is symmetric on $C_0^2[0,1]$

$$\langle v, \mathcal{L}u \rangle = \langle v, (pu')' + qu \rangle$$
$$= \langle v, (pu')' \rangle + \langle v, qu \rangle$$
$$= -\langle v', pu' \rangle + \langle qv, u \rangle$$

$\langle v, qu \rangle = \int v(qu)\,dx = \int (qv)\,u\,dx = \langle qv, u \rangle$

$$= - \langle pv', u' \rangle + \langle qv, u \rangle$$
$$= \langle (pv')', u \rangle + \langle qv, u \rangle$$
$$= \langle \mathcal{L}v, u \rangle = \langle \mathcal{L}^* v, u \rangle$$

4) $\mathcal{L} = \frac{d}{dx}$ is anti-symmetric on $C_0^2[0,1]$

$$\langle v, \mathcal{L}u \rangle = \langle v, u' \rangle = - \langle v', u \rangle$$
$$= - \langle \mathcal{L}v, u \rangle = \langle \mathcal{L}^* v, u \rangle, \quad \mathcal{L}^* = -\mathcal{L}.$$

5) $\mathcal{L} = \frac{d}{dx}\left(p(x) \frac{d}{dx} \right) + \frac{d}{dx} + q(x)$

is problematic: it is neither symmetric nor anti-sym.

Properties

- The eigenvalues of self-adjoint operators are real:
  Let $Au = \lambda u$ be fulfilled. Then it follows that
  $$\lambda \|u\|_2^2 = \langle u, \lambda u \rangle = \langle u, Au \rangle = \langle Au, u \rangle$$
  $$= \langle \lambda u, u \rangle = \bar{\lambda} \|u\|_2^2$$
- The eigenvalues of anti-symmetric operators are purely imaginary
- The eigenvectors to different eigenvalues of a self-adjoint operator are orthogonal to each other (in this context that means: u, v are orthogonal if $\langle u, v \rangle = 0$)

For $Au = \lambda u$, $Av = \mu v$, we find

$$\lambda \langle u, v \rangle = \langle Au, v \rangle = \langle u, Av \rangle$$
$$= \mu \langle u, v \rangle.$$

Thus $\lambda \neq \mu$ implies that $\langle u, v \rangle = 0$.

Elliptic operators

Definition

An operator is called elliptic if for all $u \neq 0$, it holds that

$$\langle u, Au \rangle > 0.$$

Examples

- $-\frac{d^2}{dx^2}$ is elliptic on $C_0^2[0,1]$.

$$\langle u, -u'' \rangle = \langle u', u' \rangle \geq \pi^2 \langle u, u \rangle$$

More generally, $-\frac{d}{dx}\left(p(x)\frac{d}{dx}\right) + q(x)$ is elliptic if if $p(x) > 0$ and $q(x) \geq 0$.

- $-\Delta = -\frac{d^2}{dx^2} - \frac{d^2}{dy^2}$ is an elliptic operator

Definition

An operator is positive definite if it is symmetric and elliptic.

Goal for FEM: When we discretize a differential operator, we want to preserve the symmetry, ellipticity using (possibly) higher order methods and adaptive grids.

From FDM to FEM

For a differential equation

$$Au = f \quad + \text{boundary conditions}$$

(Ex.: $A = -\frac{d^2}{dx^2}$)

Main idea for FDM:

- replace function u and f by vectors
- replace differential operator A by a matrix
- Obtain a linear system of equations

Example: $\frac{d^2}{dx^2} u(x) = f(x) \longrightarrow T_{\Delta x} u_{\Delta x} = f_{\Delta x}$

Main idea for FEM:

- Approximate function u by piecewise polynomials v that satisfy boundary conditions
- Keep differential operator A as it is
- Insert v into original equation
- Choose v such that it minimizes the residual $\|Av - f\|_{L^2}$.

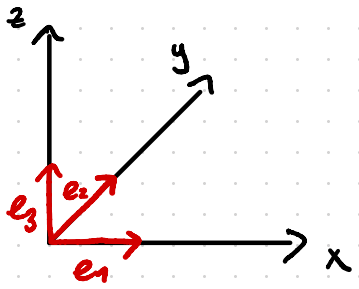
More precisely, we choose a basis $\{\varphi_i\}$ of piecewise polynomial functions. We can represent v as a linear combination of the basis functions:

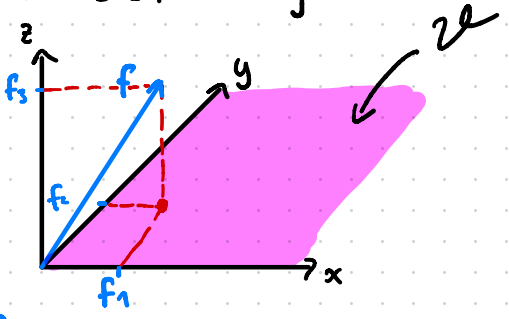
$$v(x) = \sum_{i=1}^{N} c_i \varphi_i(x).$$

We then minimize the residual

$$\min_{v \in V} \|Av - f\|_{L^2}^2$$

Example $A = I$, $\mathcal{V} = \{v = v_1 e_1 + v_2 e_2,\ v_1, v_2 \in \mathbb{R}\}$





$$f = f_1 e_1 + f_2 e_2 + f_3 e_3$$

$$\begin{aligned} v - f &= v_1 e_1 + v_2 e_2 - f_1 e_1 - f_2 e_2 - f_3 e_3 \\ &= (v_1 - f_1) e_1 + (v_2 - f_2) e_2 - f_3 e_3 \end{aligned}$$

Best approximation of f in $\mathcal{V}$:

$$v = f_1 e_1 + f_2 e_2$$

We then obtain

$$v - f = -f_3 e_3 .$$

Thus, we can see that

$$\langle \underbrace{v - f}_{-f_3 e_3}, e_1 \rangle = -f_3 \underbrace{\langle e_3, e_1 \rangle}_{=0} = 0$$

$$\langle \underbrace{v - f}_{-f_3 e_3}, e_2 \rangle = -f_3 \underbrace{\langle e_3, e_2 \rangle}_{=0} = 0$$

$$\langle \underbrace{v - f}_{-f_3 e_3}, e_3 \rangle = -f_3 \underbrace{\langle e_3, e_3 \rangle}_{=1} = -f_3$$

The best approximation v fulfills

$$\langle v - f, e_i \rangle = 0 \quad \text{for } i = 1, 2 \text{ where}$$

$\{e_1, e_2\}$ is a basis of $\mathcal{V}$.

Back to original problem: $\min_{v \in \mathcal{V}} \|Av - f\|_{L^2}^2$

Analogously, we can state a criteria for optimality:

For a basis $\{e_1, \dots, e_N\}$ of the piecewise polynomial space we then get the criteria

$$\langle Av - f, e_i \rangle = 0 \quad \forall i \in \{1, \dots, N\}$$

This is least-squares approximation.

How to use this in practice:

1) For a basis $\{e_1, \dots, e_N\}$, we want to find $v = \sum_{i=1}^{N} c_i e_i$ such that $\langle Av - f, e_j \rangle = 0$ for all $j \in \{1, \dots, N\}$.
2) That means we are foremost interested in finding $c = (c_1, \dots, c_N)$.
3) We can find c by solving the linear system

$$K c = b,$$

where $K_{i,j} = \langle A e_j, e_i \rangle, \quad b_j = \langle f, e_j \rangle$

K is referred to as stiffness matrix

Explanation:

$$\langle Av - f, \varphi_i \rangle = 0 \qquad \forall i$$

$$\Leftrightarrow \sum_{j=1}^{N} \underbrace{\langle A c_j \varphi_j, \varphi_i \rangle}_{\langle A\varphi_j, \varphi_i \rangle c_j} = \langle f, \varphi_i \rangle \qquad \forall i$$

$$\Leftrightarrow \quad (\langle A\varphi_j, \varphi_i \rangle)_{ij} \, c = (\langle f, \varphi_i \rangle)_i$$

$$\Leftrightarrow \quad K c = b$$

Important question for FEM:

How should we choose the basis $\{\varphi_1, \dots, \varphi_N\}$ and thus the space of functions $v = \text{span}\{\varphi_1, \dots, \varphi_N\}$.

We want the following:

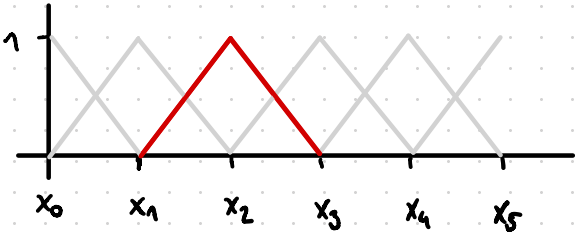
1) The functions in v fulfill the boundary cond. of the problem.
2) The function should be connected to a given grid.
3) The matrix $(\langle A\varphi_j, \varphi_i \rangle)_{ij}$ should have as many zero entries as possible.

Common choice of basis functions:

CG(1): piecewise linear basis functions

"hat functions"

$$\varphi_j(x_i) = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{else} \end{cases}$$

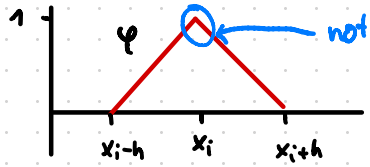


For a piecewise linear interpolant $v \approx u$, we have

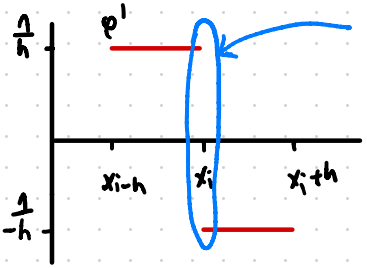
$$v(x) = \sum_{i=1}^{N} c_i \varphi_i(x) \quad \text{with} \quad v(x_i) = c_i \approx u(x_i).$$

Weak formulation of a BVP

Problem: Is this theory actually well-defined for $A = \frac{d^2}{dx^2}$? No...



not twice differentiable



not even continuous, so we cannot find a second derivative of φ here

Note: In a very generalized setting one could say $\varphi''(x) = \begin{cases} \infty & , x=x_i \\ 0 & , \text{else} \end{cases}$

Solution: Reformulate the problem.

Original problem:

$$-u'' = f \qquad \text{with } u(0) = u(1) = 0$$

Multiply with v (test function) that fulfills the boundary condition $v(0) = v(1) = 0$ and integrate:

$$-\langle u'', v\rangle = \int -u''v\,dx = \int fv\,dx = \langle f, v\rangle$$

Use integration by parts:

$$\langle u', v'\rangle = \langle f, v\rangle.$$

Weak formulation of the problem:

Find u such that

$$a(u,v): \langle u', v'\rangle = \langle f, v\rangle \qquad \forall\, v \in \mathcal{V}.$$

Advantage: only one derivative of u, v are needed.

Note: $a(u,v) = \langle u', v'\rangle$ is a bilinear form and $a(u,u) = \int |u'(x)|^2\,dx$ defines the energy norm.

Linear system for cG(1)

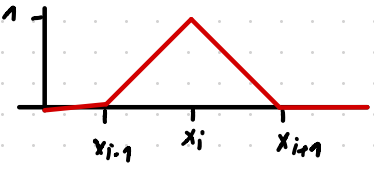
Problem we want to solve

$$-u'' = f \qquad \text{with} \qquad u(0) = u(1) = 0$$

We discretize the interval $[0,1]$ using the equidistant grid $0 < x_1 < \dots < x_N = 1$ with $x_i = i \cdot \Delta x$.

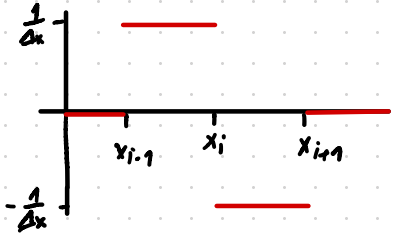
We use piecewise linear basis functions

$$\varphi_i(x) = \begin{cases} \frac{1}{\Delta x}(x - x_{i-1}), & x \in (x_{i-1}, x_i] \\ \frac{1}{\Delta x}(x_{i+1} - x), & x \in (x_i, x_{i+1}] \\ 0, & \text{else} \end{cases}$$



With the (weak) derivative

$$\varphi_i'(x) = \begin{cases} \frac{1}{\Delta x}, & x \in (x_{i-1}, x_i) \\ -\frac{1}{\Delta x}, & x \in (x_i, x_{i+1}) \\ 0, & \text{else} \end{cases}$$



The weak formulation

$$\langle u', v' \rangle = \langle f, v \rangle \quad \forall\, v \in \mathcal{V}$$

is equivalent to

$$\left\langle \sum_{i=1}^{N} c_i \varphi_i', \varphi_j' \right\rangle = \langle f, \varphi_j \rangle \quad \forall\, \varphi_j \in \{\varphi_1, \dots, \varphi_N\}$$

for $u = \sum_{i=1}^{N} c_i \varphi_i$. We then obtain the linear system

$$(Kc)_j = \sum_{i=1}^{N} \langle \varphi_i', \varphi_j' \rangle c_i = \langle f, \varphi_j \rangle = b_j$$

with $K = (\langle \varphi_i', \varphi_j' \rangle)_{i,j}$ and $b = (\langle f, \varphi_j \rangle)_j$ and unknown vector $c = (c_j)_j$. $(Kc = b)$

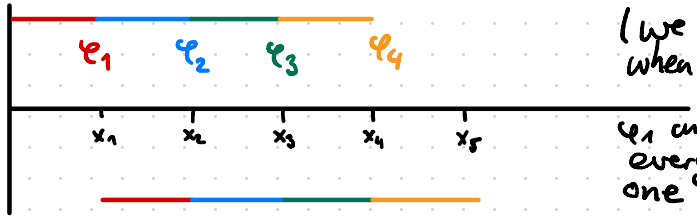
For our choice of basis functions this means

$$\langle \varphi_i', \varphi_i' \rangle = \int_{x_{i-1}}^{x_i} \frac{1}{(\Delta x)^2}\, dx + \int_{x_i}^{x_{i+1}} \frac{1}{(-\Delta x)^2}\, dx = \frac{2}{\Delta x}$$

$$\langle \varphi_i^1, \varphi_{i+1}^1 \rangle = \int_{x_i}^{x_{i+1}} -\frac{1}{(\Delta x)^2}\, dx = -\frac{1}{\Delta x}$$

$$\langle \varphi_{i-1}^1, \varphi_i^1 \rangle = -\frac{1}{\Delta x}$$

$$\langle \varphi_i^1, \varphi_j^1 \rangle = 0 \qquad \text{otherwise}$$



(we can see that when multiplying, e.g., φ_1 and φ_3 that at every point at least one function is zero)

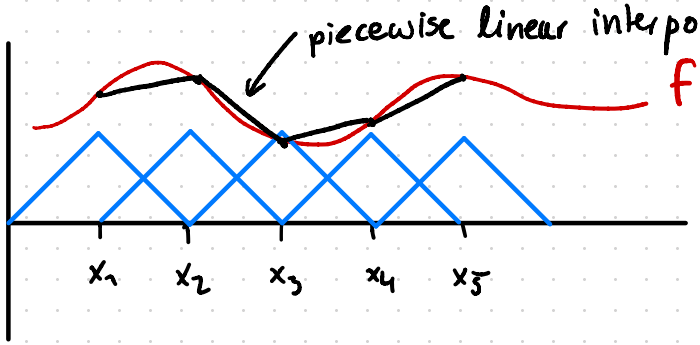
Thus, we obtain the stiffness matrix

$$K_{\Delta x} = \frac{1}{\Delta x}\,\text{tridiag}\,(-1, 2, -1)$$

Notes:

-) $K_{\Delta x} = -\Delta x\, T_{\Delta x}$ ← from FDM
-) $K_{\Delta x}$ is positive definite and therefore non-singular.
-) smallest eigenvalue $\lambda_1[K_{\Delta x}] \approx \pi^2 \Delta x$

It remains to find a representation for $b = (\langle f, \varphi_j \rangle)_j$



$$\langle f, \varphi_j \rangle \approx \left\langle \sum_{i=1}^{N} f(x_i)\,\varphi_i\,, \varphi_j \right\rangle = \sum_{i=1}^{N} f(x_i)\,\langle \varphi_i, \varphi_j \rangle,$$

where

$$\langle \varphi_i, \varphi_i \rangle = \int_{x_{i-1}}^{x_i} \frac{1}{(\Delta x)^2}(x - x_{i-1})^2\,dx + \int_{x_i}^{x_{i+1}} \frac{1}{(\Delta x)^2}(x_{i+1} - x)\,dx$$

$$= \frac{1}{(\Delta x)^2}\,\frac{1}{3}(x - x_{i-1})^3 \Big|_{x_{i-1}}^{x_i} - \frac{1}{(\Delta x)^2}\,\frac{1}{3}(x_{i+1} - x)^3 \Big|_{x_i}^{x_{i+1}}$$

$$= \frac{1}{3}\Delta x + \frac{1}{3}\Delta x = \frac{2}{3}\Delta x = \frac{\Delta x}{6} \cdot 4$$

$$\langle \varphi_i, \varphi_{i+1} \rangle = \int_{x_i}^{x_{i+1}} \frac{1}{(\Delta x)^2}(x_{i+1} - x)(x - x_i)\,dx$$

$$= \frac{1}{(\Delta x)^2} \int_0^{\Delta x} (x_{i+1} - s - x_i)(s + x_i - x_i)\,ds$$

$x = s + x_i$

$$= \frac{1}{(\Delta x)^2} \int_0^{\Delta x} (h - s)s\,ds$$

$$= \frac{1}{(\Delta x)^2}\,\frac{1}{2}hs^2 - \frac{1}{3}s^3 \Big|_0^{\Delta x}$$

$$= \frac{1}{(\Delta x)^2}\,\frac{1}{6}\Delta x^3 = \frac{\Delta x}{6}$$

Analogously $\langle \varphi_{i-1}, \varphi_i \rangle = \frac{\Delta x}{6}$. Moreover, it follows that $\langle \varphi_i, \varphi_j \rangle = 0$ in other cases, as the functions have no common support (compare picture for derivatives).

We then get the mass matrix $M_{\Delta x}$

$$M_{\Delta x} = \frac{\Delta x}{6}\,\text{tridiag}(1\ \ 4\ \ 1).$$

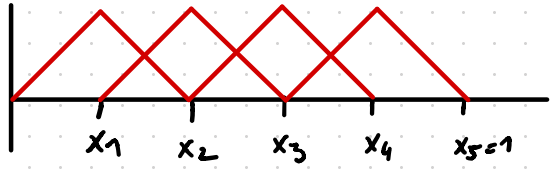
Altogether, we find

$$K_{\Delta x}\,c = M_{\Delta x}\,f.$$ ← $f = (f(x_1), f(x_2), \ldots, f(x_N))$

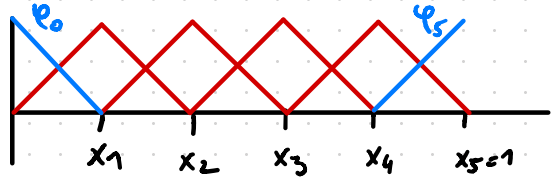
Remarks

- $u_{\Delta x} = \sum_{i=1}^{N} c_i \varphi_i$ produces a "continuous solution" instead of approximation at grid points.
- Boundary condition built into test functions

Dirichlet $u(0) = u(1) = 0$



Neumann $u'(0) = u'(1) = 0$



Add functions φ_0, φ_N.

- Compared to FDM, it is more flexible to consider complex domains or non uniform grids.
- Piecewise linear basis functions can be replaced by higher degree splines to obtain higher convergence orders.
- There exists a rich theoretical foundation.