FEM for parabolic problem

Praveen. C
praveen@math.tifrbng.res.in

Tata Institute of Fundamental Research
Center for Applicable Mathematics
Bangalore 560065
http://math.tifrbng.res.in

April 10, 2013

Heat equation

$$
\begin{aligned}
u_t - \Delta u &= f \qquad \text{in} \quad \Omega \times \mathbb{R}_+ \\
u &= 0 \qquad \text{on} \quad \Gamma \times \mathbb{R}_+ \\
u(\cdot, 0) &= u_0 \qquad \text{in} \quad \Omega
\end{aligned}
$$

Multiply by $v \in H_0^1(\Omega)$

$$\int_\Omega u_t v - \int_\Omega (\Delta u) v = \int_\Omega f v$$

and do integration by parts on the Laplacian term

$$\int_\Omega u_t v + \int_\Omega \nabla u \cdot \nabla v = \int_\Omega f v$$

Weak formulation: Find $u(\cdot, t) \in H_0^1(\Omega)$ such that

$$(\dot{u}, v) + a(u, v) = (f, v) \qquad \forall \; v \in H_0^1(\Omega) \tag{1}$$

where

$$a(u, v) = \int_\Omega \nabla u \cdot \nabla v$$

Regularity of solutions

In the case $f \equiv 0$, we can use the eigenfunctions of $-\Delta$ and construct the solution of the heat equation. Using this one can show that

Theorem

For any $u_0 \in L^2$, the solution u is a classical solution for $t > 0$, vanishes on Γ for $t > 0$, and satisfies the initial condition in the sense that $\lim_{t\to 0} \|u(t) - u_0\| = 0$. Moreover we have

$$\|\partial_t^m u(t)\|_s \le t^{-m-\frac{1}{2}s} \|u_0\|, \qquad t > 0$$

If the initial condition is more regular, then the solution is regular upto the initial time.

Theorem

If $u_0 \in H_0^1(\Omega)$, then the solution u satisfies

$$|u(t)|_1 \le |u_0|_1, \qquad t \ge 0$$

Remark: For a concise discussion, see Larsson and Thomee, Chapter 8.

Galerkin approximation

Let $\mathcal{T}_h$ be an admissible, shape-regular mesh and consider the piecewise polynomials spaces

$$V_h = X_h^k = \{v \in \mathcal{C}^0(\bar{\Omega}) : v|_K \in \mathbb{P}_k \ \forall \ K \in \mathcal{T}_h\} \cap H_0^1(\Omega)$$

Let $\{\varphi_j, j = 1, \ldots, N_h\}$ be a basis functions for V_h. We can write the approximate solution to the heat equation as

$$u_h(x,t) = \sum_{j=1}^{N_h} u_j(t)\varphi_j(x)$$

The Galerkin approximation is given by: find u_h such that

$$(\dot{u}_h, v_h) + a(u_h, v_h) = (f, v_h) \qquad \forall \ v_h \in V_h \tag{2}$$

We can write this in terms of the basis functions as:

$$(\dot{u}_h, \varphi_i) + a(u_h, \varphi_i) = (f, \varphi_i) \qquad i = 1, 2, \ldots, N_h$$

Galerkin approximation

or

$$\sum_j (\varphi_j, \varphi_i)\dot{u}_j + \sum_j a(\varphi_j, \varphi_i) u_j = \int_\Omega f(x,t)\varphi_i(x)$$

Define the mass matrix M, the stiffness matrix A and the right hand side vector b

$$M_{ij} = (\varphi_i, \varphi_j) \qquad A_{ij} = a(\varphi_j, \varphi_i) \qquad b_i(t) = \int_\Omega f(x,t)\varphi_i(x)$$

Let

$$U = [u_1, \ldots, u_{N_h}]^\top$$

be the vector of dofs. Then the Galerkin method leads to the set of coupled ODE

$$M\dot{U} + AU = b$$

Galerkin approximation

We have to also approximate the initial condition u_0 in the space V_h. This can be done either by an interpolation if u_0 is continuous, or by a projection; in any case let $u_{0,h}$ be the approximation of u_0

$$u_{0,h} = \sum_j u_j(0)\varphi_j$$

which gives the initial condition $U(0)$. Now to solve the ODE, we need to invert the mass matrix

$$\dot{U} = -M^{-1}AU + M^{-1}b$$

Mass matrix is non-singular: Let $z_h = \sum_j z_j \varphi_j \in V_h$. Then

$$\sum_j \sum_k z_j z_k M_{j,k} = \sum_j \sum_k z_j z_k (\varphi_j, \varphi_k) = \left\| \sum_j z_j \varphi_j \right\|^2 \geq 0$$

Stability of semi-discrete scheme

$$(\dot{u}_h, v_h) + a(u_h, v_h) = (f, v_h) \qquad \forall\ v_h \in V_h$$

Take $v_h = u_h$. Since $a(u_h, u_h) \geq 0$, we get

$$(\dot{u}_h, u_h) \leq (f, u_h) \leq \|f\| \, \|u_h\|$$

But

$$(\dot{u}_h, u_h) = \int_\Omega \frac{\mathrm{d}u_h}{\mathrm{d}t} u_h = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega u_h^2 = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u_h\|^2 = \|u_h\| \frac{\mathrm{d}}{\mathrm{d}t}\|u_h\|$$

Hence we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\|u_h\| \leq \|f\|$$

which after integration in time shows the stability estimate

$$\|u_h(t)\| \leq \|u_{0,h}\| + \int_0^t \|f(s)\| \, \mathrm{d}s \tag{3}$$

Ritz projection

For any $v \in H_0^1(\Omega)$ define the Ritz projecton (elliptic projection) $R_h : H_0^1(\Omega) \to V_h$ by

$$a(R_h v, w_h) = a(v, w_h) \qquad \forall\ w_h \in V_h$$

The underlying PDE is

$$-\Delta u = f := -\Delta v$$

and its Galerkin approximation $u_h = R_h v$.

Error estimate for Ritz projection

Let Ω be convex. Then we have for $s = 1, 2$

$$\|R_h v - v\| \le C h^s \|v\|_s\,, \qquad |R_h v - v|_1 \le C h^{s-1} \|v\|_s \qquad \forall\ v \in H^s \cap H_0^1$$

Proof: For $s = 2$ we can use usual interpolation error estimates to conclude that

$$\|R_h v - v\| \le C h^2 \|v\|_2\,, \qquad |R_h v - v|_1 \le C h \|v\|_2 \qquad \forall\ v \in H^2 \cap H_0^1$$

Ritz projection

For the case $s = 1$, we first note that since R_h is the orthogonal projection wrt $a(\cdot,\cdot)$ we have

$$\|R_h v\|_a \le \|v\|_a \quad \Longrightarrow \quad |R_h v|_1 \le |v|_1 \quad \Longrightarrow \quad |R_h v - v|_1 \le 2|v|_1 \le 2\,\|v\|_1$$

For the L^2 error estimate, we use the adjoint technique. Let $\varphi \in H_0^1(\Omega)$ satisfy

$$-\Delta\varphi = R_h v - v$$

so that $\varphi \in H^2$ and $\|\varphi\|_2 \le C\,\|R_h v - v\|$. Then

$$\begin{aligned}
\|R_h v - v\|^2 &= a(\varphi, R_h v - v) \\
&= a(\varphi - I_h\varphi, R_h v - v) \\
&\le |\varphi - I_h\varphi|_1 |R_h v - v|_1 \\
&\le Ch\,\|\varphi\|_2\,\|v\|_1 \\
&\le Ch\,\|R_h v - v\|\,\|v\|_1
\end{aligned}$$

L^2 error of semi-discrete scheme

Let u and u_h be the solutions of (1) and (2). Then

$$\|u_h(t) - u(t)\| \le \|u_{0,h} - u_0\| + Ch^2 \left(\|u_0\|_2 + \int_0^t \|u_t\|_2 \, \mathrm{d}s \right)$$

Proof: We write

$$u_h - u = (u_h - R_h u) + (R_h u - u) = \theta + \rho$$

The second term is easily bounded using results on Ritz projection error

$$\|\rho(t)\| \le Ch^2 \|u(t)\|_2 = Ch^2 \left\| u_0 + \int_0^t u_t \mathrm{d}s \right\|_2 \le Ch^2 \left(\|u_0\|_2 + \int_0^t \|u_t\|_2 \, \mathrm{d}s \right)$$

In order to bound θ, we note that it satisfies

$$\begin{aligned}
(\dot\theta, v_h) + a(\theta, v_h) &= (\dot u_h, v_h) + a(u_h, v_h) - (R_h \dot u, v_h) - a(R_h u, v_h) \\
&= (f, v_h) - (R_h \dot u, v_h) - a(u, v_h) \\
&= (\dot u, v_h) + a(u, v_h) - (R_h \dot u, v_h) - a(u, v_h) \\
&= (\dot u - R_h \dot u, v_h)
\end{aligned}$$

i.e.,

$$(\theta_t, v_h) + a(\theta, v_h) = -(\rho_t, v_h) \tag{4}$$

Applying stability estimate (3) we obtain

$$\|\theta(t)\| \leq \|\theta(0)\| + \int_0^t \|\rho_t\| \,\mathrm{d}s$$

where

$$\|\theta(0)\| = \|u_{0,h} - R_h u_0\| \leq \|u_{0,h} - u_0\| + \|R_h u_0 - u_0\| \leq \|u_{0,h} - u_0\| + Ch^2 \|u_0\|_2$$

and

$$\|\rho_t\| = \|R_h u_t - u_t\| \leq Ch^2 \|u_t\|_2$$

which proves the desired result.

Gradient error of semi-discrete scheme

Under the assumptions of previous theorem, we have

$$|u_h(t) - u(t)|_1 \leq |u_{0,h} - u_0|_1 + Ch\left[\|u_0\|_2 + \|u(t)\|_2 + \left(\int_0^t \|u_t\|_1^2 \,\mathrm{d}s\right)^{\frac{1}{2}}\right]$$

Proof: We split the error in terms of θ and ρ. Then

$$|\rho(t)|_1 = |R_h u(t) - u(t)|_1 \le Ch \left\| u(t) \right\|_2 \tag{5}$$

To estimate $\nabla\theta$ we use equation (4) with $v_h = \theta_t$

$$\left\|\theta_t\right\|^2 + \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|\theta|_1^2 = -(\rho_t, \theta_t) \le \frac{1}{2}(\left\|\rho_t\right\|^2 + \left\|\theta_t\right\|^2)$$

so that

$$\frac{\mathrm{d}}{\mathrm{d}t}|\theta|_1^2 \le \left\|\rho_t\right\|^2 - \left\|\theta_t\right\|^2 \le \left\|\rho_t\right\|^2$$

and hence

$$\begin{aligned} |\theta(t)|_1^2 &\le |\theta(0)|_1^2 + \int_0^t \left\|\rho_t\right\|^2 \mathrm{d}s \\ &\le (|u_{0,h} - u_0|_1 + |R_h u_0 - u_0|_1)^2 + \int_0^1 \left\|\rho_t\right\|^2 \mathrm{d}s \end{aligned}$$

Using the inequality $a^2 + b^2 \le (|a| + |b|)^2$ we get

$$|\theta(t)|_1 \le |u_{0,h} - u_0|_1 + |R_h u_0 - u_0|_1 + \left(\int_0^t \left\|\rho_t\right\|^2 \mathrm{d}s\right)^{\frac{1}{2}}$$

Now we use the error of Ritz projection to say that

$$|R_h u_0 - u_0|_1 \leq Ch \|u_0\|_2 \quad \text{and} \quad \|\rho_t\| = \|R_h u_t - u_t\| \leq Ch \|u_t\|_1$$

which leads to

$$|\theta(t)|_1 \leq |u_{0,h} - u_0|_1 + Ch \left[\|u_0\|_2 + \left(\int_0^t \|u_t\|_1^2 \, \mathrm{d}s \right)^{\frac{1}{2}} \right] \tag{6}$$

The desired result is obtained by combining equation (5) and (6).

Remark: If we take $u_{0,h} = I_h u_0$ or $u_{0,h} = R_h u_0$ then

$$\|u_{0,h} - u_0\| \leq Ch^2 \|u_0\|_2 , \qquad |u_{0,h} - u_0|_1 \leq Ch \|u_0\|_2$$

which leads to the error estimates

$$\begin{aligned} \|u_h(t) - u(t)\| &\leq Ch^2 \left(\|u_0\|_2 + \int_0^t \|u_t\|_2 \, \mathrm{d}s \right) \\ |u_h(t) - u(t)|_1 &\leq Ch \left[\|u_0\|_2 + \|u(t)\|_2 + \left(\int_0^t \|u_t\|_1^2 \, \mathrm{d}s \right)^{\frac{1}{2}} \right] \end{aligned}$$

Weaker assumptions

The previous results require the data to be very regular as seen in the norms appearing in the error estimates. If the initial condition is only in L^2 we cannot expect the solution to be regular small times. But for larges times, the solutions of heat equation are very regular. The solution of the semi-discrete scheme also has the smoothing property. Using this property it is possible to show that

Theorem

Assume that $f \equiv 0$ and let u and u_h be the solutions of (1) and (2), respectively, with the initial data being chosen as $u_{0,h} = P_h u_0$. Then

$$\|u_h(t) - u(t)\| \leq Ch^2 t^{-1} \|u_0\|, \qquad t > 0$$

Remark: See Larsson and Thomee, problem 10.4

Fully discrete, backward Euler scheme

In the Galerkin approximation

$$(\dot{u}_h(t), v_h) + a(u_h(t), v_h) = (f(t), v_h) \qquad \forall\ v_h \in V_h$$

we approximate the time derivative term by a backward difference in time

$$\dot{u}_h(t_{n+1}) \approx \frac{u_h(t_{n+1}) - u_h(t_n)}{\Delta t} =: \Delta_t^- u_h(t_{n+1})$$

Then the fully discrete scheme is given by

$$\frac{1}{\Delta t}(u_h(t_{n+1}) - u_h(t_n), v_h) + a(u_h(t_{n+1}), v_h) = (f(t_{n+1}), v_h) \qquad \forall\ v_h \in V_h$$

or

$$(u_h^{n+1}, v_h) + \Delta t\ a(u_h^{n+1}, v_h) = (u_h^n + \Delta t\ f^{n+1}, v_h) \qquad \forall\ v_h \in V_h \tag{7}$$

The discrete problem can be written in terms of matrices as

$$(M + \Delta t\ A)U^{n+1} = MU^n + \Delta t\ b^{n+1}$$

Fully discrete, backward Euler scheme

Unconditional stability: Choose $v_h = u_h^{n+1}$, then

$$\left\|u_h^{n+1}\right\|^2 \leq (u_h^n, u_h^{n+1}) + \Delta t(f^{n+1}, u_h^{n+1}) \leq \|u_h^n\| \left\|u_h^{n+1}\right\| + \Delta t \left\|f^{n+1}\right\| \left\|u_h^{n+1}\right\|$$

which implies that

$$\left\|u_h^{n+1}\right\| \leq \|u_h^n\| + \Delta t \left\|f^{n+1}\right\|$$

Iterating over time, we obtain

$$\|u_h^n\| \leq \|u_{0,h}\| + \Delta t \sum_{j=1}^{n} \left\|f^j\right\| \tag{8}$$

Error estimate for backward Euler scheme

Let u_h and u be solutions of (7) and (1), and let $u_{0,h}$ be such that $\|u_{0,h} - u_0\| \leq Ch^2 \|u_0\|_2$. Then for $n \geq 0$

$$\|u_h^n - u(t_n)\| \leq Ch^2 \left(\|u_0\|_2 + \int_0^{t_n} \|u_t\|_2 \,\mathrm{d}s \right) + \Delta t \int_0^{t_n} \|u_{tt}\| \,\mathrm{d}s$$

Fully discrete, backward Euler scheme

Proof: As before we split the error as

$$u_h^n - u(t_n) = (u_h^n - R_h u(t_n)) + (R_h u(t_n) - u(t_n)) = \theta^n + \rho^n$$

The second term is the Ritz projection error which is bounded as before

$$\|\rho(t)\| \leq Ch^2 \|u(t)\|_2 = Ch^2 \left\| u_0 + \int_0^t u_t \mathrm{d}s \right\|_2 \leq Ch^2 \left(\|u_0\|_2 + \int_0^t \|u_t\|_2 \mathrm{d}s \right)$$

Then we can show that θ satisfies

$$(\Delta_t^- \theta^n, v_h) + a(\theta^n, v_h) = -(\omega^n, v_h) \tag{9}$$

where

$$\omega^n = R_h \Delta_t^- u(t_n) - u_t(t_n) = \underbrace{(R_h - I)\Delta_t^- u(t_n)}_{\omega_1^n} + \underbrace{(\Delta_t^- u(t_n) - u_t(t_n))}_{\omega_2^n}$$

Applying stability estimate (8) to (9) we obtain

$$\|\theta^n\| \leq \left\|\theta^0\right\| + \Delta t \sum_{j=1}^{n} (\|\omega_1^n\| + \|\omega_2^n\|)$$

Fully discrete, backward Euler scheme

Now

$$\left\|\theta^0\right\| = \|u_{0,h} - R_h u_0\| \leq \|u_{0,h} - u_0\| + \|u_0 - R_h u_0\| \leq Ch^2 \|u_0\|_2$$

Note that

$$\omega_1^j = (R_h - I)\frac{1}{\Delta t}(u(t_j) - u(t_{j-1})) = (R_h - I)\frac{1}{\Delta t}\int_{t_{j-1}}^{t_j} u_t \mathrm{d}s = \frac{1}{\Delta t}\int_{t_{j-1}}^{t_j} (R_h - I)u_t \mathrm{d}s$$

whence

$$\Delta t \sum_{j=1}^{n} \left\|\omega_1^j\right\| \leq \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} Ch^2 \|u_t\|_2 \,\mathrm{d}s = Ch^2 \int_0^{t_n} \|u_t\|_2 \,\mathrm{d}s$$

Further we can write ω_2 as

$$\omega_2^j = \frac{1}{\Delta t}(u(t_j) - u(t_{j-1})) - u_t(t_j) = -\frac{1}{\Delta t}\int_{t_{j-1}}^{t_j} (s - t_{j-1})u_{tt}(s)\mathrm{d}s$$

Fully discrete, backward Euler scheme

so that

$$\Delta t \sum_{j=1}^{n} \left\| \omega_2^j \right\| \leq \sum_{j=1}^{n} \left\| \int_{t_{j-1}}^{t_j} (s - t_{j-1}) u_{tt}(s) \mathrm{d}s \right\| \leq \Delta t \int_0^{t_n} \| u_{tt} \| \, \mathrm{d}s$$

Together these inequalities complete the proof of the theorem.

Remark: The backward Euler scheme is first order in time and second order accurate in space, as is evident from the previous error estimate. To improve the accuracy in time, we can use the Crank-Nicholson scheme, which uses a centered in time discretization. For an equation like

$$\frac{\mathrm{d}}{\mathrm{d}t} u(t) = f(u(t), t)$$

the CN scheme is

$$\frac{u^{n+1} - u^n}{\Delta t} = f\left(\frac{u^n + u^{n+1}}{2}, \frac{t_n + t_{n+1}}{2} \right)$$

Crank-Nicholson scheme

The scheme is given by

$$\frac{1}{\Delta t}(u_h^{n+1} - u_h^n, v_h) + a(\frac{u_h^n + u_h^{n+1}}{2}, v_h) = (f(t_{n+\frac{1}{2}}), v_h) \tag{10}$$

Unconditional stability: Choose $v_h = (u_h^n + u_h^{n+1})$ which leads to

$$\frac{1}{\Delta t}(\left\|u_h^{n+1}\right\|^2 - \|u_h^n\|^2) \leq \left\|f^{n+\frac{1}{2}}\right\| \left\|u_h^n + u_h^{n+1}\right\| \leq \left\|f^{n+\frac{1}{2}}\right\| (\|u_h^n\| + \left\|u_h^{n+1}\right\|)$$

The left hand side can be factorised

$$\frac{1}{\Delta t}(\left\|u_h^{n+1}\right\| - \|u_h^n\|)(\left\|u_h^{n+1}\right\| + \|u_h^n\|) \leq \left\|f^{n+\frac{1}{2}}\right\| (\|u_h^n\| + \left\|u_h^{n+1}\right\|)$$

which implies

$$\left\|u_h^{n+1}\right\| \leq \|u_h^n\| + \Delta t \left\|f^{n+\frac{1}{2}}\right\|$$

Iterating over time yields

$$\|u_h^n\| \leq \|u_{0,h}\| + \Delta t \sum_{j=1}^{n} \left\|f^{j-\frac{1}{2}}\right\|$$

Crank-Nicholson scheme

Error estimate for Crank-Nicholson scheme

Let u_h and u be solutions of (10) and (1), and let $u_{0,h}$ be such that $\|u_{0,h} - u_0\| \leq Ch^2 \|u_0\|_2$. Then for $n \geq 0$

$$\|u_h^n - u(t_n)\| \leq Ch^2 \left(\|u_0\|_2 + \int_0^{t_n} \|u_t\|_2 \,\mathrm{d}s \right) + C\Delta t^2 \int_0^{t_n} (\|u_{ttt}\| + \|\Delta u_{tt}\|)\mathrm{d}s$$

<u>Proof</u>: Similar to previous theorem.

Condition number of stiffness matrix

Consider the bilinear form of the Laplace equation

$$a(u, v) = \int_\Omega \nabla u \cdot \nabla v$$

and let A be the stiffness matrix corresponding to continuous, piecewise affine basis functions, i.e.,

$$A_{ij} = a(\varphi_i, \varphi_j)$$

As usual, we assume the mesh belongs to a *shape-regular* family of triangulations, i.e., $\exists\ \kappa > 0$ such that

$$\frac{h_K}{\rho_K} \leq \kappa$$

Moreover, let us assume that the mesh is also *quasi-uniform*, i.e., $\exists\ \beta > 0$ such that

$$h_K \geq \beta h$$

where $h = \max_{K \in \mathcal{T}_h} h_K$ is the diameter of $\mathcal{T}_h$. This means that all elements $K \in \mathcal{T}_h$ are roughly the same size.

Condition number of stiffness matrix

Lemma (Inverse estimate)

There are constants c and C only depending on κ, β such that for all $v_h = \sum_{i=1}^{N_h} \eta_i \varphi_i \in V_h$ with $\eta = (\eta_i)_i \in \mathbb{R}^{N_h}$ and $|\eta| = \left(\sum_{i=1}^{N_h} |\eta_i|^2\right)^{\frac{1}{2}}$

$$ch^2|\eta|^2 \leq \|v_h\|^2 \leq Ch^2|\eta|^2 \tag{11}$$

$$a(v_h, v_h) = \int_\Omega |\nabla v_h|^2 \leq Ch^{-2} \|v_h\|^2 \tag{12}$$

Proof: It is enough to show that for each triangle $K \in \mathcal{T}_h$ with vertices $\{a^i\}$ and $v_h \in \mathbb{P}_1(K)$, we have

$$ch_K^2 \sum_{i=1}^{3} |v_h(a^i)|^2 \leq \|v_h\|_K^2 \leq Ch_K^2 \sum_{i=1}^{3} |v_h(a^i)|^2 \tag{13}$$

$$\int_K |\nabla v_h|^2 \leq Ch_K^{-2} \|v_h\|_K^2 \tag{14}$$

Condition number of stiffness matrix

with c, C independent of K and v_h. From these local estimates, the desired result follows by summation over all $K \in \mathcal{T}_h$ and using the quasi-uniform property.

Proof of (14): Let us map K to the reference element $\hat{K}$ by an affine map. Then we will show that

$$\int_{\hat{K}} |\nabla_{\hat{x}} \hat{v}_h|^2 \leq C \int_{\hat{K}} |\hat{v}_h|^2$$

Note that on K

$$v_h = \sum_{i=1}^{3} v_h(a^i)\varphi_i = \sum_{i=1}^{3} \eta_i \varphi_i \quad \Longrightarrow \quad \hat{v}_h = \sum_{i=1}^{3} \eta_i \hat{\varphi}_i$$

and with $\eta = (\eta_1, \eta_2, \eta_3)$, let us define

$$f_1(\eta) = \int_{\hat{K}} |\nabla_{\hat{x}} \hat{v}_h|^2, \qquad f_2(\eta) = \int_{\hat{K}} |\hat{v}_h|^2, \qquad f_3(\eta) = \frac{f_1(\eta)}{f_2(\eta)}$$

Condition number of stiffness matrix

We see that $f_3(\eta)$ is a homogeneous function of degree zero since $f_3(\alpha\eta) = f_3(\eta)$. Hence it is enough to show that

$$f_3(\eta) \leq C \qquad \forall\ \eta \in B = \{x \in \mathbb{R}^3 : |x| = 1\}$$

Since $f_3(\eta)$ is a continuous function and B is closed and bounded subset of $\mathbb{R}^3$, it attains its maximum on B.

To show the inequality on K we make use of the theorem relating the norms on K and $\hat{K}$.

$$\begin{aligned}
\int_K |\nabla v_h|^2 &\leq C \left\|B_K^{-1}\right\|^2 |\det(B_K)| \int_{\hat{K}} |\nabla_{\hat{x}} \hat{v}_h|^2 \\
&\leq C \left\|B_K^{-1}\right\|^2 |\det(B_K)| \int_{\hat{K}} |\hat{v}_h|^2 \\
&\leq C \left\|B_K^{-1}\right\|^2 |\det(B_K)| |\det(B_K)|^{-1} \int_K |v_h|^2 \\
&\leq C \rho_K^{-2} \int_K |v_h|^2 \leq C h_K^{-2} \int_K |v_h|^2
\end{aligned}$$

Condition number of stiffness matrix

Inequality (13) can be shown along similar lines.

Remark: Inequality (12) is refered to as an *inverse estimate.*

Condition number of stiffness matrix: For any $v_h \in V_h$

$$v_h = \sum_{i=1}^{N_h} \eta_i \varphi_i, \qquad a(v_h, v_h) = \eta^\top A \eta$$

Hence, using (12) and right inequality of (11)

$$\frac{\eta^\top A\eta}{|\eta|^2} = \frac{a(v_h, v_h)}{|\eta|^2} \leq Ch^{-2}\frac{\|v_h\|^2}{|\eta|^2} \leq C \qquad \forall\ \eta \in \mathbb{R}^{N_h}$$

On the other hand, using coercivity of $a(\cdot,\cdot)$ and left inequality of (11)

$$\frac{\eta^\top A\eta}{|\eta|^2} = \frac{a(v_h, v_h)}{|\eta|^2} \geq \frac{\alpha\,\|v_h\|_1^2}{|\eta|^2} \geq \alpha\frac{\|v_h\|^2}{|\eta|^2} \geq \alpha c h^2 \qquad \forall\ \eta \in \mathbb{R}^{N_h}$$

Condition number of stiffness matrix

These inequalities show that there are constants c, C such that

$$\lambda_{\max}(A) \leq C, \qquad \lambda_{\min}(A) \geq ch^2$$

Then the condition number can be estimated as

$$\text{cond}(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)} \leq Ch^{-2}$$

Remark: While the stiffness matrix A is positive definite for any mesh size h, its condition number increases with decreasing mesh size, and the matrix becomes badly ill-conditioned. This affects the convergence rate of iterative methods for solving the matrix equation and one has to resort to some *pre-conditioning* technique to improve the condition number.

Condition number of mass matrix

Using similar techniques and results as above, we can estimate the eigenvalues of the mass matrix.

$$ch^2 \leq \frac{\eta^\top M \eta}{|\eta|^2} = \frac{\|v_h\|^2}{|\eta|^2} \leq Ch^2 \qquad \forall\ \eta \in \mathbb{R}^{N_h}$$

which shows that

$$\lambda_{\min}(M) \geq ch^2, \qquad \lambda_{\max}(M) \leq Ch^2, \qquad \mathrm{cond}(M) = \frac{\lambda_{\max}(M)}{\lambda_{\min}(M)} \leq \frac{C}{c}$$

The mass matrix is well conditioned as the condition number if independent of the mesh size.

Forward Euler scheme

The scheme is given by

$$\frac{1}{\Delta t}(u_h^{n+1} - u_h^n, v_h) + a(u_h^n, v_h) = (f(t_n), v_h) \tag{15}$$

The discrete equations are given by

$$\frac{1}{\Delta t} M(U^{n+1} - U^n) + AU^n = b^n$$

or

$$MU^{n+1} = (M - \Delta t A)U^n + \Delta t b^n$$

The scheme is not fully explicit since we have a matrix on the left hand side. Solving we obtain

$$U^{n+1} = (I - \Delta t M^{-1} A)U^n + \Delta t M^{-1} b^n$$

Let us take $b \equiv 0$ and iterating over time

$$U^n = (I - \Delta t M^{-1} A)^n U^0$$

Forward Euler scheme

For stability we require the eigenvalues of the iteration matrix $I - \Delta t M^{-1} A$ to lie in $[-1, +1]$, or

$$-1 \leq 1 - \Delta t \lambda(M^{-1}A) \leq 1$$

Conditional stability: Consider the case $f \equiv 0$. Set $v_h = u_h^{n+1} - u_h^n$ and $v_h = u_h^{n+1} + u_h^n$ in (15) to obtain

$$\frac{1}{\Delta t} \left\| u_h^{n+1} - u_h^n \right\|^2 + a(u_h^n, u_h^{n+1} - u_h^n) = 0 \qquad (16)$$

$$\frac{1}{\Delta t} (\left\| u_h^{n+1} \right\|^2 - \| u_h^n \|^2) + a(u_h^n, u_h^{n+1} + u_h^n) = 0 \qquad (17)$$

Then (17)-(16) yields after using coercivity

$$\left\| u_h^{n+1} \right\|^2 - \| u_h^n \|^2 - \left\| u_h^{n+1} - u_h^n \right\|^2 = -2\Delta t a(u_h^n, u_h^n) \leq -2\alpha \Delta t \left\| u_h^n \right\|_1^2 \qquad (18)$$

Now use the continuity of the bilinear form in (16)

$$\left\| u_h^{n+1} - u_h^n \right\|^2 = -\Delta t a(u_h^n, u_h^{n+1} - u_h^n) \leq \gamma \Delta t \left\| u_h^n \right\|_1 \left\| u_h^{n+1} - u_h^n \right\|_1$$

Forward Euler scheme

The last factor can be estimated using inverse estimate

$$\left\|u_h^{n+1}-u_h^n\right\|_1^2 = \left\|u_h^{n+1}-u_h^n\right\|^2 + |u_h^{n+1}-u_h^n|_1^2 \le (1+Ch^{-2})\left\|u_h^{n+1}-u_h^n\right\|^2$$

so that

$$\left\|u_h^{n+1}-u_h^n\right\|_1 \le (1+Ch^{-2})^{\frac{1}{2}}\left\|u_h^{n+1}-u_h^n\right\|$$

which leads to

$$\left\|u_h^{n+1}-u_h^n\right\| \le \gamma\Delta t(1+Ch^{-2})^{\frac{1}{2}}\left\|u_h^n\right\|_1$$

Now (18) becomes

$$\left\|u_h^{n+1}\right\|^2 \le \left\|u_h^n\right\|^2 + \Delta t[\gamma^2\Delta t(1+Ch^{-2})-2\alpha]\left\|u_h^n\right\|_1^2 \le \left\|u_h^n\right\|^2$$

where the last inequality holds provided

$$\Delta t \le \frac{2\alpha}{(1+Ch^{-2})\gamma^2} \quad \Longrightarrow \quad \Delta t = \mathcal{O}\left(h^2\right)$$

The time step is thus restricted by the spatial mesh size h and the scheme is *conditionally stable*. If $h \ll 1$, then the problem is very stiff requiring very small time steps.