Norwegian University of Science
and Technology
Department of Mathematical
Sciences

TMA4220 Numerical Solution of Partial Differential Equations Using Element Methods Fall 2023

Solutions to exercise set 3

1 **a)** We first introduce a generic functional space V and consider a test function $v : \Omega \times \mathbb{R} \to \mathbb{R}$ with $v(\cdot, t) \in V$ for every $t \in \mathbb{R}$. Integrating the PDE against v we get

$$(\partial_t u, v) + (\nabla u, \nabla v) = (f, v),$$

where for compactness we denoted $(u, v) = \int_\Omega uv\mathrm{d}x$. We see that V can be naturally defined as $H_0^1(\Omega)$.

b) By linearity we can get

$$\sum_{i=1}^{N}(\Phi_i, \Phi_j)\dot{u}_i(t) + \sum_{i=1}^{N}(\nabla\Phi_i, \nabla\Phi_j)u_i(t) = (f, \Phi_j),\ \ j = 1, ..., N$$

where $u(x, t) = \sum_{i=1}^{N} u_i(t)\Phi_i(x)$. By introducing the stiffness and mass matrices A and M, we get

$$M\dot{u}(t) + Au(t) = b,$$

where $M_{ij} = (\Phi_i, \Phi_j)$, $A_{ij} = (\nabla\Phi_i, \nabla\Phi_j)$ and $b_j = (f, \Phi_j)$. In order to prove the stiffness matrix A is symmetric positive semidefinite, we first notice symmetry is trivially verified, and the positive semidefiniteness can be checked by considering $v \in \mathbb{R}^N$ and computing

$$v^T Av = \sum_{i,j}(v_i\nabla\Phi_i, v_j\nabla\Phi_j) = (\sum_i v_i\nabla\Phi_i, \sum_j v_j\nabla\Phi_j) = (\sum_i v_i\nabla\Phi_i, \sum_i v_i\nabla\Phi_i)$$
$$= \|\sum_i v_i\nabla\Phi_i\|^2_{L^2(\Omega)} \geq 0.$$

This allows to conclude the desired result.

c) To prove the solution exists unique, we first can observe that M is symmetric and positive definite. Indeed, by a similar argument

$$v^T Mv = \|\sum_{i=1}^{N} v_i\Phi_i\|^2_{L^2(\Omega)} \geq 0$$

and we also observe that by linear independence of the basis functions $\Phi_1, ..., \Phi_N$, $\sum_{i=1}^{N} v_i\Phi_i = 0$ if and only if $v = 0$. This allows to get M is symmetric and positive definite, and hence in particular invertible. For this reason, we can rewrite the obtained system of linear differential equations as

$$\dot{u}(t) = -M^{-1}Au(t) + M^{-1}b =: X(u(t)).$$

The right hand side X is Lipschitz continuous, and hence the solution exists unique.

d) We start by considering the expansion

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

which leads to the relation $\nabla U^n - \nabla u(t_n) = \nabla\theta^n + \nabla\rho^n$, where $\rho^n = \rho(t_n)$. The norm $\|\nabla\rho(t_n)\|$ can be bounded by $Ch^{r-1}\|u(t_n)\|_r$ as a consequence of Lemma 1.1 in [1]. We then focus on the other term, and aim to provide a similar bound for $\|\nabla\theta^n\|$. If we plug θ^n in the weak formulation obtained by discretizing in time with the Backward-Euler method, we get

$$(\bar{\partial}\theta^n, \chi) + (\nabla\theta^n, \nabla\chi) = -(\omega^n, \chi),\ \forall\chi \in \mathbb{V}_h,\ n \geq 1,$$

where

$$\omega^n = R_h\bar{\partial}u\,(t_n) - u_t\,(t_n) = (R_h - I)\,\bar{\partial}u\,(t_n) + \left(\bar{\partial}u\,(t_n) - u_t\,(t_n)\right) = \omega_1^n + \omega_2^n.$$

We now choose $\chi = \bar{\partial}\theta^n$ and replace it in the equation we just obtained, so to get

$$\|\bar{\partial}\theta^n\|^2 + (\nabla\theta^n, \nabla(\bar{\partial}\theta^n)) = -(\omega_1^n, \bar{\partial}\theta^n) - (\omega_2^n, \bar{\partial}\theta^n).$$

We now apply Young's inequality to the left hand side to get

$$\|\bar{\partial}\theta^n\|^2 + (\nabla\theta^n, \nabla(\bar{\partial}\theta^n)) \leq \frac{1}{2}\|\omega^n\|^2 + \frac{1}{2}\|\bar{\partial}\theta^n\|^2$$

and hence

$$(\nabla\theta^n, \nabla(\bar{\partial}\theta^n)) \leq \frac{1}{2}\|\omega^n\|^2.$$

We now introduce the definition of $\bar{\partial}\theta^n$ so to get

$$\|\nabla\theta^n\|^2 \leq \frac{\delta t}{2}\|\omega^n\|^2 + (\nabla\theta^n, \theta^{n-1}).$$

We further apply Young's inequality and get

$$\|\nabla\theta^n\|^2 \leq \frac{\delta t}{2}\|\omega^n\|^2 + \frac{1}{2}\|\nabla\theta^n\|^2 + \frac{1}{2}\|\nabla\theta^{n-1}\|^2$$

and hence

$$\frac{1}{2\delta t}(\|\nabla\theta^n\|^2 - \|\nabla\theta^{n-1}\|^2) \leq \frac{1}{2}\|\omega^n\|^2.$$

This allows to conclude

$$\bar{\partial}\|\nabla\theta^n\|^2 \leq \|\omega^n\|^2.$$

By expanding we get

$$\begin{aligned}\|\nabla\theta^n\|^2 &\leq \|\nabla\theta^{n-1}\|^2 + \delta t\|\omega^n\|^2 \\ &\leq (\|\nabla\theta^{n-2}\|^2 + \delta t\|\omega^{n-1}\|^2) + \delta t\|\omega^n\|^2 \\ &\leq \|\nabla\theta^0\|^2 + \delta t\sum_{k=1}^{n}\|\omega^k\|^2.\end{aligned}$$

The term $\|\nabla\theta^0\| = \|\nabla U^0 - \nabla u(t_0)\|$ can be controlled as

$$\begin{aligned}\|\nabla\theta^0\| = \|\nabla U^0 - \nabla u(t_0)\| &= \|\nabla U^0 - \nabla R_h v + \nabla R_h v - \nabla u(t_0)\| \\ &\leq \|\nabla U^0 - \nabla R_h v\| + \|\nabla R_h v - \nabla v\| \\ &\leq C_1 h^{r-1}\|v\|_r.\end{aligned}$$

As derived in the proof of Theorem 1.5 in [1], one has

$$\delta t \sum_{k=1}^{n} \|\omega^k\| \leq C_2 h^r \int_0^{t_n} \|u_t\|_r \mathrm{d}s + \delta t \int_0^{t_n} \|u_{tt}\| \mathrm{d}s.$$

Now, since

$$\left(\sum_{k=1}^{n} \|\omega^k\| \right)^2 \geq \sum_{k=1}^{n} \|\omega^k\|^2,$$

we can conclude

$$\sum_{k=1}^{n} \|\omega^k\|^2 \leq C_3 h^{2r} \int_0^{t_n} \|u_t\|_r^2 \mathrm{d}s + C_4 \delta t^2 \int_0^{t_n} \|u_{tt}\|^2 \mathrm{d}s.$$

We remark that

$$\|\nabla U^n - \nabla u(t_n)\|^2 \leq 2(\|\nabla \rho_n\|^2 + \|\nabla \theta_n\|^2).$$

Combining the two estimates, we get

$$\begin{aligned} \|\nabla U^n - \nabla u(t_n)\|^2 \leq 2 \Bigg(& C_1^2 h^{2(r-1)} \|v\|_r^2 + C_3 h^r \int_0^{t_n} \|u_t\|_r^2 \mathrm{d}s \\ & + C_4 \delta t \int_0^{t_n} \|u_{tt}\|^2 \mathrm{d}s + C_5 h^{2(r-1)} \|u(t_n)\|_r^2 \Bigg), \end{aligned}$$

which allows to conclude the desired result.

2 **a)** We first introduce two generic functional spaces W and V, and consider the test functions $\boldsymbol{v} \in W$ and $q \in V$. The two equations can then be integrated by parts. For conciseness in notation we denote

$$(f, g) = \int_\Omega f(x) g(x) \mathrm{d}x.$$

The weak formulation is then

$$\begin{aligned} (-\Delta \boldsymbol{u}, \boldsymbol{v}) + (\nabla p, \boldsymbol{v}) &= (f, \boldsymbol{v}) \\ (\nabla \cdot \boldsymbol{u}, q) &= 0. \end{aligned}$$

Integrating by parts we can conclude

$$\begin{aligned} (\nabla \boldsymbol{u}, \nabla \boldsymbol{v}) - (p, \nabla \cdot \boldsymbol{v}) &= (f, \boldsymbol{v}) \\ -(\nabla \cdot \boldsymbol{u}, q) &= 0, \end{aligned}$$

where we denote

$$\nabla \boldsymbol{u}(x) = \begin{bmatrix} \nabla \boldsymbol{u}_1(x) \\ \vdots \\ \nabla \boldsymbol{u}_d(x) \end{bmatrix}$$

and

$$(\nabla \boldsymbol{u}, \nabla \boldsymbol{v}) = \sum_{i,j=1}^{d} (\nabla \boldsymbol{u}_i, \nabla \boldsymbol{v}_j).$$

We then set $a(\boldsymbol{u}, \boldsymbol{v}) := (\nabla \boldsymbol{u}, \nabla \boldsymbol{v})$ and $b(\boldsymbol{u}, q) = -(\nabla \cdot \boldsymbol{u}, q)$. In order for these two bilinear forms to be well defined, the minimal regularity we can ask is $W = (H_0^1(\Omega))^d$ and $V = L^2(\Omega)$. The linear functional is instead $F(\boldsymbol{v}) := (f, \boldsymbol{v})$.

b) To prove continuity it is enough to use Cauchy-Schwartz:

$$a(\boldsymbol{u},\boldsymbol{v}) = (\nabla\boldsymbol{u},\nabla\boldsymbol{v}) \leq \|\nabla\boldsymbol{u}\|_{L^2(\Omega)}\|\nabla\boldsymbol{v}\|_{L^2(\Omega)} = |\boldsymbol{u}|_{H^1(\Omega)}|\boldsymbol{v}|_{H^1(\Omega)},$$

$$|b(\boldsymbol{u},q)| = |(\nabla\cdot\boldsymbol{u},q)| \leq \|\nabla\cdot\boldsymbol{u}\|_{L^2(\Omega)}\|q\|_{L^2(\Omega))} \leq d\|\nabla\boldsymbol{u}\|_{L^2(\Omega)}\|q\|_{L^2(\Omega))},$$

where we used

$$\|\nabla\cdot\boldsymbol{u}\|_{L^2(\Omega)} \leq d\|\nabla\boldsymbol{u}\|_{L^2(\Omega)}$$

which is a consequence of triangular inequality.

c) To prove a is coercive in the $|\cdot|_{H^1(\Omega)}$ norm, i.e. the H^1-seminorm, one does not have to do anything since $a(\boldsymbol{u},\boldsymbol{u}) = |\boldsymbol{u}|^2_{H^1(\Omega)}$. If the interest is in studying also the L^2-norm, a further step is needed and it is based on Poincarè's inequality.

d) If we set $q \equiv 1$ we notice that

$$b(\boldsymbol{u},1) = -(\nabla\cdot\boldsymbol{u},1) = -\int_{\partial\Omega}\boldsymbol{u}\cdot n\mathrm{d}s = 0$$

since $\boldsymbol{u}$ vanishes on $\partial\Omega$. This implies that b can not satisfy the inf-sup condition if we work with $q \in L^2(\Omega)$. If instead we work with $V = L^2_0(\Omega)$ such a problem is removed. Indeed by the above calculation we understood that $\nabla\cdot : (H^1_0(\Omega))^d \to L^2_0(\Omega)$, and hence L^2_0 is a natural space also for the test function q.

e) We solve the Poisson boundary value problem

$$\begin{cases} -\Delta\phi = q \in L^2_0(\Omega),\ \Omega \\ \partial_n\phi = 0,\ \partial\Omega, \end{cases}$$

and by regularity of Ω one has $\phi \in H^2(\Omega)$. Furthermore, even though on $H^2(\Omega)$ the solution is defined up to a constant, due to the zero-average condition, on $H^2(\Omega)\cap L^2_0(\Omega)$ ϕ is unique. Thus we pose the problem on this space, and hence have $\phi \in H^2(\Omega)\cap L^2_0(\Omega)$. We then set $\boldsymbol{u} = \nabla\phi$ so that for sure $\boldsymbol{u}\cdot n = 0$ on $\partial\Omega$. One can then also show, with a bit of technicalities, that it is also true that $\boldsymbol{u} = 0$ on $\partial\Omega$. We then proceed with the following derivation

$$b(\boldsymbol{u},q) = -(\nabla\cdot\boldsymbol{u},q) = -(\nabla\cdot\nabla\phi,q) = (-\Delta\phi,q) = \|q\|^2_{L^2(\Omega)},$$

and hence since

$$\|\nabla\boldsymbol{u}\|^2_{L^2(\Omega)} \leq \|\phi\|^2_{H^2(\Omega)} \leq \beta\|q\|^2_{L^2(\Omega)},$$

one can conclude

$$\frac{b(\boldsymbol{u},q)}{\|\nabla\boldsymbol{u}\|_{L^2(\Omega)}} = \frac{\|q\|^2_{L^2(\Omega)}}{\|\nabla\boldsymbol{u}\|_{L^2(\Omega)}} \geq \frac{1}{\sqrt{\beta}}\|q\|_{L^2(\Omega)}.$$

As a consequence

$$\frac{1}{\sqrt{\beta}}\|q\|_{L^2(\Omega)} \leq \sup_{\boldsymbol{v}\in W}\frac{b(\boldsymbol{v},q)}{\|\nabla\boldsymbol{v}\|_{L^2(\Omega)}}$$

for every $q \in L^2_0(\Omega)$, which allows to conclude the derivation.

f) The problem is well posed since we have just proved the continuity and coercivity of a, the continuity of b and also the inf-sup condition. What is missing is only the boundedness of the linear functional F, which follows from Cauchy-Schwartz inequality.

3 **a)** This point is an exercise in applying the discrete version of Cauchy-Schwartz inequality. The first term can be controlled as follows:

$$\begin{aligned}\left|\sum_{T\in\mathcal{T}_h}\int_T \nabla w\cdot\nabla v\mathrm{d}x\right| &\leq \sum_{T\in\mathcal{T}_h}\|\nabla w\|_{L^2(T)}\|\nabla v\|_{L^2(T)}\\ &\leq \left(\sum_{T\in\mathcal{T}_h}\|\nabla w\|^2_{L^2(T)}\right)^{1/2}\left(\sum_{T\in\mathcal{T}_h}\|\nabla v\|^2_{L^2(T)}\right)^{1/2}\\ &\leq \|v\|_{\mathrm{ip}}\|w\|_{\mathrm{ip}}.\end{aligned}$$

We now consider the less intuitive terms, i.e. the second and the third. For this, we first focus on the following integral

$$\begin{aligned}\left|\int_e\{\nabla w\}\cdot[v]\mathrm{d}s\right| &= \left|\int_e(|e|^{1/2}\{\nabla w\})(|e|^{-1/2}[v])\mathrm{d}s\right|\\ &\leq \left(\int_e|e|\{\nabla w\}^2\mathrm{d}s\right)^{1/2}\left(\int_e|e|^{-1}[v]^2\mathrm{d}s\right)^{1/2}\\ &= \left(\eta\eta^{-1}\int_e|e|\{\nabla w\}^2\mathrm{d}s\right)^{1/2}\left(2\eta(2\eta)^{-1}\int_e|e|^{-1}[v]^2\mathrm{d}s\right)^{1/2}\\ &= \eta^{1/2}\sqrt{\frac{1}{2}}\eta^{-1/2}\|v\|_{\mathrm{ip}}\|w\|_{\mathrm{ip}} = \sqrt{\frac{1}{2}}\|v\|_{\mathrm{ip}}\|w\|_{\mathrm{ip}}.\end{aligned}$$

Applying this reasoning, together with the discrete Cauchy-Schwartz inequality, we can then conclude

$$\left|\sum_{e\in\mathcal{E}_h}\int_e\{\nabla w\}\cdot[v]\mathrm{d}s\right| + \left|\sum_{e\in\mathcal{E}_h}\int_e\{\nabla v\}\cdot[w]\mathrm{d}s\right| \leq \sqrt{2}\|v\|_{\mathrm{ip}}\|w\|_{\mathrm{ip}}.$$

A similar reasoning applies also for the last term. We conclude that the bilinear form is continuous.

b) This is immediate since the bilinear form has two terms which are anti-symmetric in v, w. This allows to say that $a_h^+(v,v) = \|v\|^2_{\mathrm{dG}}$. Indeed by direct replacement of $v = w$ in the bilinear form, one notices that the terms

$$\int_e\{\nabla v\}[v]\mathrm{d}s - \int_e\{\nabla v\}[v]\mathrm{d}s$$

cancel out and what survives is just the $\|\cdot\|_{\mathrm{dG}}$ norm.

c) The difference between the two norms is given by

$$\|v\|^2_{\mathrm{ip}} - \|v\|^2_{\mathrm{dG}} = \eta^{-1}\sum_{e\in\mathcal{E}_h}|e|\|\{\nabla v\}\|^2_{L^2(e)} + \eta\sum_{e\in\mathcal{E}_h}|e|^{-1}\|[v]\|^2_{L^2(e)}.$$

Let us call $\mathcal{P}_e$ the subset of $\{T\in\mathcal{T}_h\}$ having $e\in\partial T$. We first aim to bound the first term:

$$\begin{aligned}\eta^{-1}\sum_{e\in\mathcal{E}_h}|e|\|\{\nabla v\}\|^2_{L^2(e)} &\leq C\eta^{-1}\sum_{e\in\mathcal{E}_h}|\partial T|^2\sum_{T\in\mathcal{P}_e}\|\nabla v|_T\|^2_{L^2(T)}\\ &\leq C_1\eta^{-1}\sum_{e\in\mathcal{E}_h}\sum_{T\in\mathcal{P}_e}\|\nabla v|_T\|^2_{L^2(T)}\\ &\leq C_2\eta^{-1}\sum_{T\in\mathcal{T}_h}\|\nabla v|_T\|^2_{L^2(T)}.\end{aligned}$$

This allows to conclude the desired result. We will call $C_2 = C$ for simplicity in the next question.

d) To prove Céa's lemma we proceed as follows

$$\begin{aligned}
\|u - u_h^+\|_{\mathrm{ip}} &\leq \|u - v\|_{\mathrm{ip}} + \|u_h^+ - v\|_{\mathrm{ip}} \\
&\leq \|u - v\|_{\mathrm{ip}} + C(1+\eta^{-1})\|u_h^+ - v\|_{\mathrm{dG}} \\
&\leq \|u - v\|_{\mathrm{ip}} + C(1+\eta^{-1}) \sup_{w \in V_h \setminus \{0\}} \frac{|a_h^+(u_h^+ - v, w)|}{\|w\|_{\mathrm{dG}}} \\
&= \|u - v\|_{\mathrm{ip}} + C(1+\eta^{-1}) \sup_{w \in V_h \setminus \{0\}} \frac{|a_h^+(u - v, w)|}{\|w\|_{\mathrm{dG}}} \\
&\leq C(1+\eta^{-1})^2 \|u - v\|_{\mathrm{ip}}.
\end{aligned}$$

The result follows by arbitrarity of $v \in V_h$, and taking the infimum over $v \in V_h$ on both sides of the inequality.

References

[1] Vidar Thomée. *Galerkin finite element methods for parabolic problems*, volume 25. Springer Science & Business Media, 2007.