NTNU
Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for

TMA4220 Numerical Solution of Partial Differential Equations Using Element Methods—SOLUTION

Academic contact during examination: Trond Kvamsdal

Phone: 93058702

Examination date: 13th of December 2021

Examination time (from–to): 09:00–13:00

Permitted examination support material: C:

- A. Quarteroni: *Numerical Models for Differential Problems, Springer 2014*
- S. Brenner and L. R. Scott: *The Mathematical Theory of Finite Element Methods, Springer 2008*
- *TMA4220 Lecture Notes Fall 2019* (Front page + 229 pages)
- *TMA4220-2019H-AFEM* (25 pages)
- Rottmann: *Matematisk formelsamling*
- Approved calculator

Other information:
All answers should be justified and include enough details to make it clear which methods and/or results have been used. All the (sub-)problems are worth 5 points each. The total value is 65 points.

Language: English

Number of pages: 11

Number of pages enclosed: 0

Checked by:

Date Signature

Problem 1 Consider the two-dimensional steady heat equation

$$\begin{aligned} -\nabla(\kappa\nabla u) &= f \quad \text{in} \quad \Omega \\ u &= 0 \quad \text{on} \quad \partial\Omega_D \\ H(u) &= \bar{t} \text{ on } \partial\Omega_N. \end{aligned}$$

a) (See, e.g., A. Quarteroni pag. 48) We now establish the weak formulation of the problem integrating the equation against a test function $v \in X$, for some test space X we determine later:

$$\int_\Omega -\nabla \cdot (k\nabla u)v dx = \int_\Omega f v dx.$$

By Green's theorem and the imposition of the boundary conditions, this turns out to be

$$a(u,v) = \int_\Omega k\nabla u \cdot \nabla v dx = \int_\Omega f v dx + \int_{\partial\Omega_N} v(k\nabla u) \cdot n d\Gamma.$$

We now introduce the Neumann operator

$$H(u) = k\nabla u \cdot n = k\frac{\partial u}{\partial n} = \bar{t} \text{ on } \partial\Omega_N$$

and conclude

$$l(v) = \int_\Omega f v dx + \int_{\partial\Omega_N} v\bar{t} d\Gamma.$$

Since the weak formulation involves up to first order weak derivatives, we require $u, v \in H^1(\Omega)$. Moreover, having Dirichlet boundary conditions on $\partial\Omega_D$, we set $X = \{w \in H^1(\Omega) : \ w = 0 \text{ in the sense of traces on } \partial\Omega_D\}$. This is our test and solution space, since we have homogeneous Dirichlet boundary conditions.

b) (See, e.g., A. Quarteroni pag. 116,117 ex. 7-8) First of all we notice that if $k(x,y) \equiv \bar{k} \in \mathbb{R}$ we need to assume $\partial\Omega_D \neq \emptyset$, otherwise even when the solution exists, it will be just defined up to a constant and hence not unique. Furthermore, we assume $f \in L^2(\Omega)$, $k \in L^\infty(\Omega)$ (for example it can be continuous on the compact set Ω), $\bar{t} \in L^2(\partial\Omega_N)$. We derive the existence and uniqueness of the solution via Lax-Milgram theorem. Indeed, a is coercive since

$$a(u,u) = \int_\Omega k\|\nabla u\|^2 dx \geq k_{min}|u|^2_{H^1(\Omega)} \geq C\|u\|^2_{H^1}$$

where the last step comes from Poincaré inequality. It is continuous since

$$a(u,v) \leq \|k\|_{L^\infty(\Omega}|v|_{H^1}|u|_{H^1} \leq \|k\|_{L^\infty(\Omega)}\|v\|_{H^1}\|u\|_{H^1}.$$

The functional l is linear. Moreover, it is bounded because of the assumptions made on $\bar{t}$ and f. We hence conclude existence and uniqueness by Lax-Milgram theorem.

c) We need to fix a finite dimensional space

$$X_h = \text{span}\{\varphi_i : \Omega \to \mathbb{R} : \ i = 1, ..., n\} \subset X$$

where the solution will be approximated. Thus, we look for a function

$$u_h = \sum_{i=1}^{n} u_h^i \varphi_i(x) \in X_h$$

such that, for any $v_h \in X_h$, it satisfies

$$a(u_h, v_h) = l(v_h).$$

d) To handle inhomogeneous Dirichlet boundary conditions, still working with symmetric solution and test spaces, we lift the solution via a function $R_{\bar{u}} \in H^1(\Omega)$ which satisfies $R_{\bar{u}}|_{\partial\Omega_D} = \bar{u}$. Then, we define the solution $u = \overset{\circ}{u} + R_{\bar{u}}$ where $\overset{\circ}{u} \in X$, i.e. it is valued 0 on the Dirichlet boundary. In this way, looking for the solution $\overset{\circ}{u}$ we get the variational formulation: Find $\overset{\circ}{u} \in X$ such that, for any $v \in X$:

$$a(\overset{\circ}{u}, v) = l(v) - a(R_{\bar{u}}, v).$$

In the FEM code, this can be translated into working with $R_{\bar{u}} = \sum_{i \in \mathcal{B}_\mathcal{D}} \bar{u}^i \varphi_i(x)$, where $\mathcal{B}_D$ are the indices of the boundary nodes of the domain. Hence the i-th row of the FEM linear system has an additional contribution on the right hand side which is due to $-\sum_{j \in \mathcal{B}_D} a(\varphi_j, \varphi_i)\bar{u}^j$. The solution, will then be $u = \overset{\circ}{u} + R_{\bar{u}}$.

e) Suppose to number the nodes as in Figure 1, and hence let z_9 be the inner node in both the triangulations. Since the two triangulations have all nodes but one on the Dirichlet boundary, we can already say that the only unknown of the linear systems $Au = b$ will be u_9. Thus, we can decide to impose the boundary conditions removing the rows and columns of the stiffness matrix associated to boundary nodes and what survives is a scalar equation of the

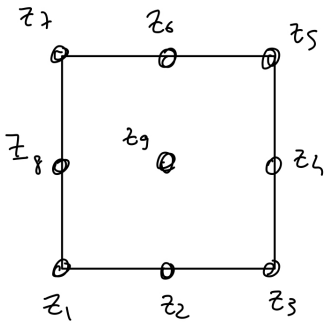

Figure 1: Labelling of the triangulation nodes

form $a_{99}u_9 = b_9$. We thus find b_9 and a_{99} for both the triangulations and then plot the diagonal profile of the solution. For the right hand side, we have

$$b_9 = f\int_\Omega \phi_9(x,y)dxdy$$

where $f \in \mathbb{R}$ is the constant scalar forcing term. This integral corresponds to the volume defined by the piecewise linear function ϕ_9. In the left triangulation, ϕ_9 defines a pyramid of basis given by the full square and hence $b_9 = f/3$, while for the right triangulation the basis has smaller area (given by the boundary nodes belonging to an element where z_9 is one of the vertices) and we get $b_9 = f/4$. For the left hand side, we have

$$a_{99} = k\int_\Omega \|\nabla\phi_9\|^2 dx = k\sum_{i=1}^{N}\int_{T_k}\|\nabla\phi_9\|^2 dx.$$

Recall that on a triangle of vertices $(x_1,y_1),(x_2,y_2),(x_3,y_3)$ the basis functions ϕ_1,ϕ_2,ϕ_3 that satisfy $\phi_i(z_j)=\delta_{ij}$ are defined as $\phi_i(x,y) = a_ix + b_iy + c_i$ where

$$\begin{bmatrix} x_1 & y_1 & 1\\ x_2 & y_2 & 1\\ x_3 & y_3 & 1\end{bmatrix}\begin{bmatrix} a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{bmatrix} = \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}.$$

Thus, for example on the triangle T_1 of vertices $(1/2,1/2),(1,1/2),(1,1)$, ϕ_9 takes the form $\phi_9 = -2x+2$. By symmetry, for the left triangulation, all the integrals take the same values and hence

$$a_{99} = 8k\int_{T_1} 4dx = 32k|T_1| = 4k.$$

This implies that for the left triangulation $u_9 = \frac{f}{12k}$. For the right triangulation, not all the integrals for the a_{99} terms coincide, in particular the triangles

sharing z_9 and living on the bottom right and top left squares have different integrals. Let us call T_2 the one with vertices $(1/2, 1/2), (1, 1/2), (1/2, 0)$. Then we have

$$a_{99} = k\left(4\int_{T_1} \|\nabla\phi_9\|^2 dx + 2\int_{T_2} \|\nabla\phi_9\|^2 dx\right).$$

By previous computation we have $\phi_9|_{T_2} = -2x + 2y + 1$ and hence $\|\nabla\phi_9\|^2 = 8$. Thus

$$a_{99} = 4k$$

again. We conclude that for the right triangulation $u_9 = \frac{f}{16k}$. To conclude, since for the right triangulation $u = 0$ both on z_6 and z_8, and the same happens for z_2 and z_4 the solution being piecewise linear, will be 0 even on the edges connecting them. This fact and the different support of ϕ_9 brings to the two following different profiles of the solution:

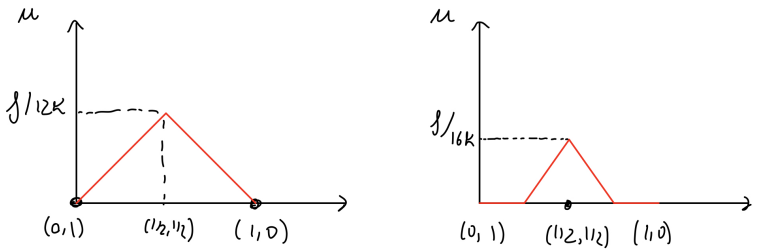

Figure 2: Diagonal profiles of the two solutions

Problem 2

a) (cfr. Brenner and Scott) Let

- $K \subseteq \mathbb{R}^n$ be a bounded closed set with nonempty interior and piecewise smooth boundary (the element domain),
- $\mathcal{P}$ be a finite-dimensional space of functions on K (the space of shape functions) and
- $\mathcal{N} = \{N_1, N_2, \ldots, N_k\}$ be a basis for $\mathcal{P}'$ (the set of nodal variables).

Then $(K, \mathcal{P}, \mathcal{N})$ is called a finite element.

A finite element is said to be compatible, or conforming, if the chosen finite dimensional space X_h is a true subset of the variational space X, i.e. $X_h \subset X$. This happens when there is continuity between the elements.

b) Let $z_1, ..., z_5$ be the nodes. We start considering just the 4 boundary nodes z_1, z_2, z_3, z_4, and we work with the reference element $[0,1]^2$. The basis functions

$$\Phi_1(x,y) = (1-x)(1-y), \quad \Phi_2(x,y) = x(1-y),$$
$$\Phi_3(x,y) = xy, \quad \Phi_4(x,y) = (1-x)y$$

fully determine the space $\mathcal{Q}_1 = \{\sum_j c_j p_j(x) q_j(y) : p_j, q_j \in \mathcal{P}_1\}$ where $\mathcal{P}_1$ is the space of one-variable polynomials of degree at most 1. Then, we define the bubble function $\Phi_5(x,y) = 16xy(1-x)(1-y)$ that is valued 1 on the node z_5 and 0 on all the edges. We then introduce 4 scalar coefficients $\alpha_1, ..., \alpha_4 \in \mathbb{R}$ that have the aim of constructing a compatible finite element involving all the 5 nodes. Let $\hat{\Phi}_i(x,y) = \Phi_i(x,y) - \alpha_i \Phi_5(x,y)$ for $i = 1, ..., 4$. Moreover, the basis functions have to satisfy the partition of unity and hence we have to have

$$\sum_{i=1}^{4} \hat{\Phi}_i + \Phi_5 = \sum_{i=1}^{4} \Phi_i + (1 - \sum_{i=1}^{4} \alpha_i)\Phi_5 = 1 + (1 - \sum_{i=1}^{4} \alpha_i)\Phi_5 = 1.$$

This implies that $\sum_{i=1}^{4} \alpha_i = 1$. Moreover, we already have that $\hat{\Phi}_i(z_i) = 1$, but it is not true in general that $\hat{\Phi}_i(z_5) = 0$ for $i = 1, ..., 4$. We hence impose this condition to recover the finite element space. By symmetry, we get that

$$\hat{\Phi}_i(1/2, 1/2) = 1/4 - \alpha_i = 0 \implies \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \frac{1}{4}.$$

We notice that these coefficients satisfy the partition of unity property and this concludes the derivation. Thus, the nodes $z_1, ..., z_5$ and the functions $\hat{\Phi}_1, ..., \hat{\Phi}_4, \Phi_5$ define a compatible finite element.

A different solution can be to reconduce this setting to the known one in which the square is triangulated in 4 triangles, with shared internal vertex. In this case, linear polynomials are fully determined by the 3 degrees of freedom (i.e. the 3 vertices). Thus, we have that linear polynomials are fully determined by $\mathcal{N} = \{N_1, ..., N_5\}$, where $N_i(z_j) = \delta_{ij}$ and they are linear.

c) We define the space $\mathcal{Q}_3 = \{\sum_j c_j p_j(x) q_j(y) : \ p_j, q_j \in \mathcal{P}_3\}$ where $\mathcal{P}_3$ is the space of one-variable polynomials of degree at most 3. We now show that the grid presented in the Figure 3 and their basis functions determine the element space. Then, we choose to find the basis functions associated to the nodes, z_7, z_{13} and z_6. To show that the 16 basis functions defined for these nodes determine $\mathcal{Q}_3$, we check that if a polynomial $P \in \mathcal{Q}_3$ vanishes on all the nodes, then it is identically 0. Because when restricted on the lines L_i,

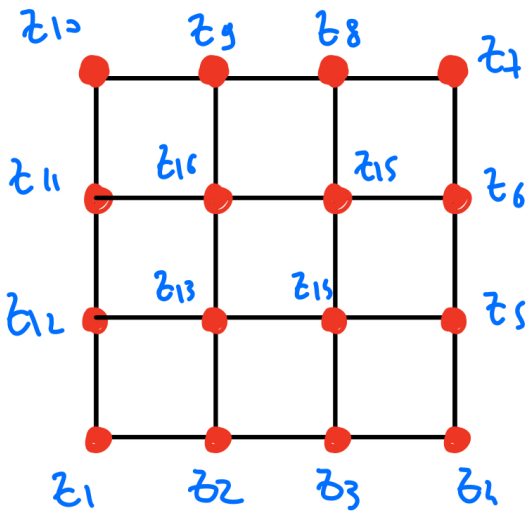

Figure 3: Nodes of the cubic basis functions

$i = 1, .., 6$, depicted in Figure 4 we have that P has degree 3 and it vanishes on the 4 nodes on L_i, we can write $P = cL_1L_2L_3L_4L_5L_6$. Moreover, since $P(z_7) = 0$, we conclude $c = 0$ and hence that $P \equiv 0$. Thus, these nodes (and the associated basis functions) determine the finite element space. To build the 3 required basis functions, we define the 4 cubic basis functions on the reference interval $[-1, 1]$:

$$\varphi_0(x) = -\frac{9}{16}(x + 1/3)(x - 1/3)(x - 1)$$

$$\varphi_1(x) = \frac{27}{16}(x - 1/3)(x + 1)(x - 1)$$

$$\varphi_2(x) = -\frac{27}{16}(x + 1/3)(x + 1)(x - 1)$$

$$\varphi_3(x) = \frac{9}{16}(x + 1)(x - 1/3)(x + 1/3)$$

Because of the structure of $\mathcal{Q}_3$, we define the basis functions for the square element as follows:

$$\phi_7(x, y) = \varphi_3(x)\varphi_3(y),$$
$$\phi_{13}(x, y) = \varphi_1(x)\varphi_1(y)$$
$$\phi_6(x, y) = \varphi_3(x)\varphi_2(y).$$

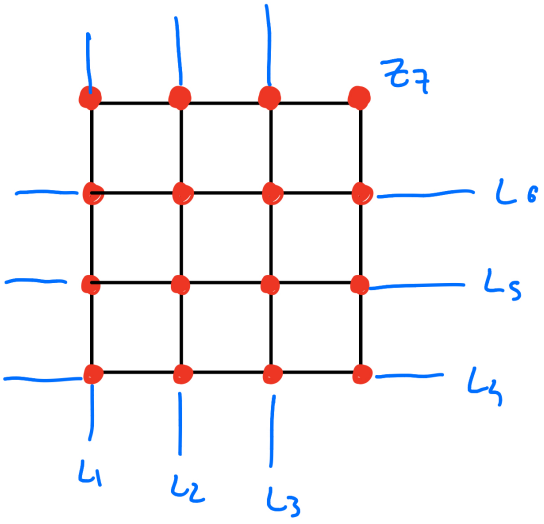

Figure 4: Check the nodes determine the finite element

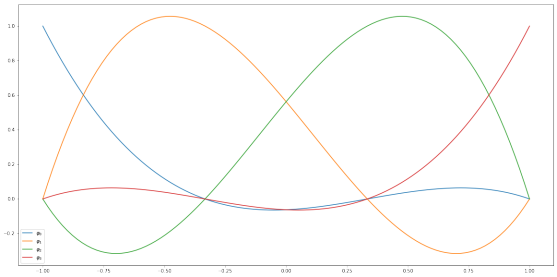

Figure 5: Cubic 1d basis functions

d) Consider a PDE with solution of the variational problem $u \in X$ and let u_h be the solution of the associated Galerkin variational formulation: find $u_h \in X_h$ such that $a(u_h, v_h) = l(v_h)$ for any $v_h \in X_h$. Then, Galerkin orthogonality is the property

$$a(u - u_h, v_h) = 0 \quad \forall v_h \in X_h.$$

This property coincides with the definition of strong consistency of the finite element method. Moreover, in the case $a(\cdot, \cdot)$ is symmetric and positive definite, i.e. it defines an inner product and a metric $\| \cdot \|_a$, then u_h is the orthogonal projection of the exact solution $u \in X$ onto the finite dimensional subspace X_h. This motivates the name of this property. Even without these assumptions on $a(\cdot, \cdot)$, the Galerkin orthogonality property implies the optimality of the solution u_h in the space X_h. Indeed,

$$\|u - u_h\|_a \leq C \inf_{w_h \in X_h} \|u - w_h\|_a$$

and hence if X_h tends to fill X as $h \to 0$, $\|u - u_h\|_a \to 0$ and hence we can say that Galerkin orthogonality implies convergence of the finite element method.

e) We want to find the 4 basis functions $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4$ for this transfer element. We start from the linear basis functions on the triangle z_1, z_2, z_3. In barycentric coordinates ξ, η, they are

$$\phi_1(\xi, \eta) = 1 - \xi - \eta, \ \phi_2(\xi, \eta) = \xi, \ \phi_3(\xi, \eta) = \eta.$$

We need to introduce a 4th function $\hat{\phi}_4$ that is the standard quadratic basis function for triangular elements, associated to the midpoint z_4. More explicitly, $\hat{\phi}_4 = 4\eta\xi$. We define $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ and set

$$\hat{\phi}_i = \phi_i - \alpha_i \phi_4.$$

Because the basis functions need to partition unity, i.e. $\sum_{i=1}^4 \hat{\phi}_i = 1$, and since $\sum_{i=1}^3 \phi_i = 1$, we get

$$\hat{\phi}_4 + \sum_{i=1}^3 \phi_i - \hat{\phi}_4 \sum_{i=1}^3 \alpha_i = \hat{\phi}_4 + 1 - \hat{\phi}_4 \sum_{i=1}^3 \alpha_i = 1$$

and hence $\alpha_1 + \alpha_2 + \alpha_3 = 1$. We then impose $\hat{\phi}_i(z_j) = \delta_{ij}$ to recover the expressions of the αs:

$$\hat{\phi}_1(1/2, 1/2) = -\alpha_1 = 0$$
$$\hat{\phi}_2(1, 0) = \hat{\phi}_3(0, 1) = 1$$

$$\hat{\phi}_2(1/2, 1/2) = 1/2 - \alpha_2 = 0$$
$$\hat{\phi}_3(1/2, 1/2) = 1/2 - \alpha_3 = 0$$

and hence $\alpha_1 = 0$, $\alpha_2 = 1/2$ and $\alpha_3 = 1/2$. This gives

$$\hat{\phi}_1 = 1 - \xi - \eta$$
$$\hat{\phi}_2 = \xi - 2\xi\eta$$
$$\hat{\phi}_3 = \eta - 2\xi\eta$$
$$\hat{\phi}_4 = 4\xi\eta.$$

Problem 3

a) Consider a PDE that defines a bilinear form $a : X \times X \to \mathbb{R}$ that is symmetric and positive definite. Then, the solution u_h of the Galerkin method, because of the Galerkin orthogonality property, is the orthogonal projection of the exact solution $u \in X$ onto the finite dimensional space $X_h \subset X$ with respect to the energy norm. By energy norm we mean $\|v\|_a = \sqrt{a(v, v)}$. In other words, $\|u - u_h\|_a \leq \|u - v_h\|_a$ for any other $v_h \in X_h$. In the PDE introduced in the first exercise, we have that

$$a(u, v) = \int_\Omega k(x, y) \nabla u \cdot \nabla v dx dy$$

and since $k(x, y) \geq k_{min} > 0$ by assumption, we have

$$a(u, v) = a(v, u) \text{ and } a(u, u) = \int_\Omega k(x, y) \|\nabla u\|^2 dx dy \geq k_{min} \|\nabla u\|^2_{L^2(\Omega)} \geq 0.$$

Moreover, since the space X where u lives does not admit $u \equiv c$ for $c \neq 0$, we have that $a(u, u) = 0$ if and only if $u \equiv 0$, when this is compatible with the Neumann BCs. Thus, in the case of interest the solution u_h is optimal with respect to the energy norm.

b) Following the a priori error estimate presented in Quarteroni, page 95, we can bound the H^1 norm of the error $u - u_h$ as follows. Assume $u \in H^{r+1}(\Omega)$, then

$$\|u - u_h\|_{H^1(\Omega)} \leq \frac{M}{\alpha} C \left(\sum_{T \in \mathcal{T}_h} h_T^{2r} |u|^2_{H^{r+1}(T)} \right)^{1/2}$$

and hence

$$\|u - u_h\|_{H^1(\Omega)} \leq \frac{M}{\alpha} C h^r |u|_{H^{r+1}(\Omega)}.$$

This even allows to recover that if $u \in H^{p+1}(\Omega)$ for some $p > 0$, then

$$|u - u_h|_{H^1(\Omega)} \leq \bar{C} h^s |u|_{H^{s+1}(\Omega)}, \quad s = \min\{r, p\}$$

where r is the degree of the finite element space used. We now explicitly recall the continuity constant M and the coercivity one, α, of the problem:

$$M = \|k\|_{L^\infty(\Omega)}$$

$$\alpha = \frac{k_{min}}{1 + C_\Omega^2}$$

where C_Ω is such that $\|v\|_{L^2(\Omega)} \leq C_\Omega |v|_{H^1(\Omega)}$ for any $v \in H_0^1(\Omega)$. C_Ω is the constant defining Poincaré's inequality. This a priori estimate comes from

$$\begin{aligned}\alpha |u - u_h|_{H^1}^2 &\leq a(u - u_h, u - v_h) = a(u - u_h, u - u_h) + a(u - u_h, u_h - v_h) \leq \\ &\leq M |u - u_h|_{H^1} |u - v_h|_{H^1}\end{aligned}$$

for any $v_h \in X_h$, by Galerkin orthogonality. This gives

$$|u - u_h|_{H^1} \leq \inf_{v_h \in X_h} |u - v_h|_{H^1} \leq |u - \Pi_h^1 u|_{H^1},$$

where Π_h^1 is the interpolation operator. This term on the right hand side can be then bounded by the estimates on the interpolation error, and this allows to conclude.

c) A priori estimates for the finite element approximation error suggest that refining the mesh allows to get better accuracies from the finite element approximation. However, it is not necessary to restrict the diameter of all the elements, since the error balance $|u - u_h|_{H^1(\Omega)}$ is controlled by the products between the elements diameters and the $r + 1$ seminorm on the respective elements of the exact solution (where r is the polynomial degree adopted in the FEM approximation). Thus, we would like to equi-distribute the error on each element of the triangulation. A larger contribution from $|u|_{H^{r+1}(K)}$, i.e. a larger variability of the exact solution on the element, should be balanced by a smaller local grid-size or higher polynomial degree. This gives rise to adaptive finite elements approaches. There a number of ways in which one can adaptively refine meshes. The basic structure of the overall algorithm usually follows the following steps:

- Solve the PDE on the current mesh
- Estimate the error on each element using some criterion indicative for the error

- Mark those with large errors for refinement, and those with small one for coarsening, and leave the rest intact,
- Refine and coarsen the marked cells obtaining a new mesh, or alternatively change the polynomial approximation degree on those (when this makes sense from the point of view of solution's regularity).
- Repeat the steps up to reaching the desired error.

This refinement procedure, allows to get an efficient grid, without decreasing the diameters of the elements where it is not needed. The labelling of elements to coarsen and refine is where various possibilities are available. For techniques based on a priori estimates, this decision is based on reconstruction techniques aiming to get similar inequalities to the a priori estimates but without using the exact (unknown) solution. Those based on a posteriori estimates, on the other hand, base this choice on computable quantities starting from the current mesh and the obtained numerical solution. (cfr. `https://www.dealii.org/current/doxygen/deal.II/step_6.html#Intro` and A. Quarteroni)