Appendix B
Finite Element Methods

B.1 The Ritz Method and the Galerkin Method

Remark B.1. Contents. This section studies abstract problems in Hilbert spaces. The existence and uniqueness of solutions will be discussed. Approximating this solution with finite dimensional spaces is called Ritz method or Galerkin method. Some basic properties of this method will be proved.

In this section, a Hilbert space V will be considered with inner product $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ and norm $\|v\|_V = a(v,v)^{1/2}$. □

Theorem B.2. Representation theorem of Riesz. *Let $f \in V'$ be a continuous and linear functional, then there is a uniquely determined $u \in V$ with*

$$a(u,v) = f(v) \quad \forall\, v \in V. \tag{B.1}$$

In addition, u is the unique solution of the variational problem

$$F(v) = \frac{1}{2}a(v,v) - f(v) \to \min \ \forall\, v \in V. \tag{B.2}$$

Proof. First, the existence of a solution u of the variational problem will be proved. Since f is continuous, it holds

$$|f(v)| \le c\,\|v\|_V \quad \forall\, v \in V,$$

from what follows that

$$F(v) \ge \frac{1}{2}\,\|v\|_V^2 - c\,\|v\|_V \ge -\frac{1}{2}c^2,$$

where in the last estimate the necessary criterion for a local minimum of the expression of the first estimate is used. Hence, the function $F(\cdot)$ is bounded from below and

$$d = \inf_{v \in V} F(v)$$

exists.

Let $\{v_k\}_{k\in\mathbb{N}}$ be a sequence with $F(v_k) \to d$ for $k \to \infty$. A straightforward calculation (parallelogram identity in Hilbert spaces) gives

$$\|v_k - v_l\|_V^2 + \|v_k + v_l\|_V^2 = 2\,\|v_k\|_V^2 + 2\,\|v_l\|_V^2\,.$$

Using the linearity of $f(\cdot)$ and $d \leq F(v)$ for all $v \in V$, one obtains

$$\begin{aligned}
&\|v_k - v_l\|_V^2 \\
&= 2\|v_k\|_V^2 + 2\|v_l\|_V^2 - 4\left\|\frac{v_k + v_l}{2}\right\|_V^2 - 4f(v_k) - 4f(v_l) + 8f\left(\frac{v_k + v_l}{2}\right) \\
&= 4F(v_k) + 4F(v_l) - 8F\left(\frac{v_k + v_l}{2}\right) \\
&\leq 4F(v_k) + 4F(v_l) - 8d \to 0
\end{aligned}$$

for $k, l \to \infty$. Hence $\{v_k\}_{k\in\mathbb{N}}$ is a Cauchy sequence. Because V is a complete space, there exists a limit u of this sequence with $u \in V$. Because $F(\cdot)$ is continuous, it is $F(u) = d$ and u is a solution of the variational problem.

In the next step, it will be shown that each solution of the variational problem (B.2) is also a solution of (B.1). It is

$$\begin{aligned}
\Phi(\varepsilon) &= F(u + \varepsilon v) = \frac{1}{2}a(u + \varepsilon v, u + \varepsilon v) - f(u + \varepsilon v) \\
&= \frac{1}{2}a(u, u) + \varepsilon a(u, v) + \frac{\varepsilon^2}{2}a(v, v) - f(u) - \varepsilon f(v).
\end{aligned}$$

If u is a minimum of the variational problem, then the function $\Phi(\varepsilon)$ has a local minimum at $\varepsilon = 0$. The necessary condition for a local minimum leads to

$$0 = \Phi'(0) = a(u, v) - f(v) \quad \text{for all } v \in V.$$

Finally, the uniqueness of the solution will be proved. It is sufficient to prove the uniqueness of the solution of the equation (B.1). If the solution of (B.1) is unique, then the existence of two solutions of the variational problem (B.2) would be a contradiction to the fact proved in the previous step. Let u_1 and u_2 be two solutions of the equation (B.1). Computing the difference of both equations gives

$$a(u_1 - u_2, v) = 0 \quad \text{for all } v \in V.$$

This equation holds, in particular, for $v = u_1 - u_2$. Hence, $\|u_1 - u_2\|_V = 0$, such that $u_1 = u_2$. ■

Theorem B.3. Theorem of Lax–Milgram. *Let $b(\cdot, \cdot) : V \times V \to \mathbb{R}$ be a bounded and coercive bilinear form on the Hilbert space V. Then, for each bounded linear functional $f \in V'$ there is exactly one $u \in V$ with*

$$b(u, v) = f(v) \quad \forall\, v \in V. \tag{B.3}$$

Proof. One defines linear operators $T, T' : V \to V$ by

$$a(Tu, v) = b(u, v)\ \forall\, v \in V, \quad a(T'u, v) = b(v, u)\ \forall\, v \in V. \tag{B.4}$$

Since $b(u, \cdot)$ and $b(\cdot, u)$ are continuous linear functionals on V, it follows from Theorem B.2 that the elements Tu and $T'u$ exist and they are defined uniquely. Because the operators satisfy the relation

$$a(Tu, v) = b(u, v) = a(T'v, u) = a(u, T'v), \tag{B.5}$$

T' is called adjoint operator of T. Setting $v = Tu$ in (B.4) and using the boundedness of $b(\cdot, \cdot)$ yields

$$\|Tu\|_V^2 = a(Tu, Tu) = b(u, Tu) \leq M\|u\|_V \|Tu\|_V \implies \|Tu\|_V \leq M\|u\|_V$$

for all $u \in V$. Hence, T is bounded. Since T is linear, it follows that T is continuous. Using the same argument, one shows that T' is also bounded and continuous.

Define the bilinear form

$$d(u,v) := a(TT'u,v) = a(T'u,T'v) \quad \forall\, u,v \in V,$$

where (B.5) was used. Hence, this bilinear form is symmetric. Using the coercivity of $b(\cdot,\cdot)$ and the Cauchy–Schwarz inequality gives

$$m^2 \|v\|_V^4 \le b(v,v)^2 = a(T'v,v)^2 \le \|v\|_V^2 \left\|T'v\right\|_V^2 = \|v\|_V^2\, a(T'v,T'v) = \|v\|_V^2\, d(v,v).$$

Applying now the boundedness of $a(\cdot,\cdot)$ and of T' yields

$$m^2 \|v\|_V^2 \le d(v,v) = a(T'v,T'v) = \left\|T'v\right\|_V^2 \le M \|v\|_V^2 . \tag{B.6}$$

Hence, $d(\cdot,\cdot)$ is also coercive and, since it is symmetric, it defines an inner product on V. From (B.6) one has that the norm induced by $d(v,v)^{1/2}$ is equivalent to the norm $\|v\|_V$. From Theorem B.2 it follows that there is a exactly one $w \in V$ with

$$d(w,v) = f(v) \quad \forall\, v \in V.$$

Inserting $u = T'w$ into (B.3) gives with (B.4)

$$b(T'w,v) = a(TT'w,v) = d(w,v) = f(v) \quad \forall\, v \in V,$$

hence $u = T'w$ is a solution of (B.3).

The uniqueness of the solution is proved analogously as in the symmetric case. ■

Remark B.4. Basic idea of the Ritz method. For approximating the solution of (B.2) or (B.1) with a numerical method, it will be assumed that V has a countable orthonormal basis (Schauder basis), i.e., V is a separable Hilbert space. Then, using Parseval's equality, one finds that there are finite dimensional subspaces $V_1, V_2, \ldots \subset V$ with $\dim V_k = k$, which has the following property: for each $u \in V$ and each $\varepsilon > 0$ there is a $K \in \mathbb{N}$ and a $u_k \in V_k$ with

$$\|u - u_k\|_V \le \varepsilon \quad \forall\, k \ge K. \tag{B.7}$$

Note that it is not required that there holds an inclusion of the form $V_k \subset V_{k+1}$.

The Ritz approximation of (B.2) and (B.1) is defined by: Find $u_k \in V_k$ with

$$a(u_k,v_k) = f(v_k) \quad \forall\, v_k \in V_k. \tag{B.8}$$

□

Lemma B.5. Existence and uniqueness of a solution of (B.8)**.** *There exists exactly one solution of* (B.8)*.*

Proof. Finite dimensional subspaces of Hilbert spaces are Hilbert spaces as well. For this reason, one can apply the representation theorem of Riesz, Theorem B.2, to (B.8) which gives the statement of the lemma. In addition, the solution of (B.8) solves a minimization problem on V_k. ■

Lemma B.6. Best approximation property. *The solution of* (B.8) *is the best approximation of u in V_k, i.e., it is*

$$\|u - u_k\|_V = \inf_{v_k \in V_k} \|u - v_k\|_V . \tag{B.9}$$

Proof. Since $V_k \subset V$, one can use the test functions from V_k in the weak equation (B.1). Then, the difference of (B.1) and (B.8) gives the orthogonality, the so-called Galerkin orthogonality,

$$a(u - u_k, v_k) = 0 \quad \forall\, v_k \in V_k. \tag{B.10}$$

Hence, the error $u - u_k$ is orthogonal to the space V_k: $u - u_k \perp V_k$. That means, u_k is the orthogonal projection of u onto V_k with respect of the inner product of V.

Let now $w_k \in V_k$ be an arbitrary element, then it follows with the Galerkin orthogonality (B.10) and the Cauchy–Schwarz inequality that

$$\begin{aligned} \|u - u_k\|_V^2 &= a(u - u_k, u - u_k) = a(u - u_k, u - \underbrace{(u_k - w_k)}_{v_k}) = a(u - u_k, u - v_k) \\ &\leq \|u - u_k\|_V \, \|u - v_k\|_V . \end{aligned}$$

Since $w_k \in V_k$ was arbitrary, also $v_k \in V_k$ is arbitrary. If $\|u - u_k\|_V > 0$, division by $\|u - u_k\|_V$ gives the statement of the lemma. If $\|u - u_k\|_V = 0$, the statement of the lemma is trivially true. ∎

Theorem B.7. Convergence of the Ritz approximation. *The Ritz approximation converges*

$$\lim_{k \to \infty} \|u - u_k\|_V = 0.$$

Proof. The best approximation property (B.9) and property (B.7) give

$$\|u - u_k\|_V = \inf_{v_k \in V_k} \|u - v_k\|_V \leq \varepsilon$$

for each $\varepsilon > 0$ and $k \geq K(\varepsilon)$. Hence, the convergence is proved. ∎

Remark B.8. Formulation of the Ritz method as linear system of equations. One can use an arbitrary basis $\{\phi_i\}_{i=1}^k$ of V_k for the computation of u_k. First of all, the equation for the Ritz approximation (B.8) is satisfied for all $v_k \in V_k$ if and only if it is satisfied for each basis function ϕ_i. This statement follows from the linearity of both sides of the equation with respect to the test function and from the fact that each function $v_k \in V_k$ can be represented as linear combination of the basis functions. Let $v_k = \sum_{i=i}^k \alpha_i \phi_i$, then from (B.8) it follows that

$$a(u_k, v_k) = \sum_{k=1}^{k} \alpha_i a(u_k, \phi_i) = \sum_{k=1}^{k} \alpha_i f(\phi_i) = f(v_k).$$

This equation is satisfied if $a(u_k, \phi_i) = f(\phi_i)$, $i = 1, \ldots, k$. On the other hand, if (B.8) holds then it holds in particular for each basis function ϕ_i.

Then, one uses as ansatz for the solution also a linear combination of the basis functions

$$u_k = \sum_{j=1}^{k} u^j \phi_j$$

with unknown coefficients $u^j \in \mathbb{R}$. Using as test functions now the basis functions yields

$$\sum_{j=1}^{k} a(u^j \phi_j, \phi_i) = \sum_{j=1}^{k} a(\phi_j, \phi_i) u^j = f(\phi_i), \quad i = 1, \ldots, k.$$

This equation is equivalent to the linear system of equations $A\boldsymbol{u} = \boldsymbol{f}$, where

$$A = (a_{ij})_{i,j=1}^{k} = a(\phi_j, \phi_i)_{i,j=1}^{k}$$

is called stiffness matrix. Note that the order of the indices is different for the entries of the matrix and the arguments of the inner product. The right hand side is a vector of length k with the entries $f_i = f(\phi_i)$, $i = 1, \ldots, k$.

Using the one-to-one mapping between the coefficient vector $(v^1, \ldots, v^k)^T$ and the element $v_k = \sum_{i=1}^{k} v^i \phi_i$, one can show that the matrix A is symmetric and positive definite

$$\begin{aligned} A = A^T \quad &\Longleftrightarrow \quad a(v, w) = a(w, v) \quad \forall\, v, w \in V_k, \\ \boldsymbol{x}^T A \boldsymbol{x} > 0 \text{ for } \boldsymbol{x} \neq \boldsymbol{0} \quad &\Longleftrightarrow \quad a(v, v) > 0 \quad \forall\, v \in V_k, v \neq 0. \end{aligned}$$

□

Remark B.9. The case of a bounded and coercive bilinear form. If $b(\cdot, \cdot)$ is bounded and coercive, but not symmetric, it is possible to approximate the solution of (B.3) with the same idea as for the Ritz method. In this case, it is called Galerkin method. The discrete problem consists in finding $u_k \in V_k$ such that

$$b(u_k, v_k) = f(v_k) \quad \forall\, v_k \in V_k. \tag{B.11}$$

□

Lemma B.10. Existence and uniqueness of a solution of (B.11)**.** *There is exactly one solution of* (B.11)*.*

Proof. The statement of the lemma follows directly from the Theorem of Lax–Milgram, Theorem B.3. ■

Lemma B.11. Lemma of Cea, error estimate. *Let $b : V \times V \to \mathbb{R}$ be a bounded and coercive bilinear form on the Hilbert space V and let $f \in V'$ be a bounded linear functional. Let u be the solution of* (B.3) *and u_k be the solution of* (B.11)*, then the following error estimate holds*

$$\|u - u_k\|_V \leq \frac{M}{m} \inf_{v_k \in V_k} \|u - v_k\|_V , \tag{B.12}$$

where the constants M and m are given in (A.24) *and* (A.25)*.*

Proof. Considering the difference of the continuous equation (B.3) and the discrete equation (B.11), one obtains the error equation

$$b(u - u_k, v_k) = 0 \quad \forall\, v_k \in V_k,$$

which is also called Galerkin orthogonality. With (A.25), the Galerkin orthogonality, and (A.24) it follows that

$$\begin{aligned} \|u - u_k\|_V^2 &\leq \frac{1}{m} b(u - u_k, u - u_k) = \frac{1}{m} b(u - u_k, u - v_k) \\ &\leq \frac{M}{m} \|u - u_k\|_V \|u - v_k\|_V , \quad \forall\, v_k \in V_k, \end{aligned}$$

from what the statement of the lemma follows immediately. ■

Remark B.12. On the best approximation error. It follows from estimate (B.12) that the error is bounded by a multiple of the best approximation error, where the factor depends on properties of the bilinear form $b(\cdot, \cdot)$. Thus, concerning error estimates for concrete finite-dimensional spaces, the study of the best approximation error will be of importance. □

Remark B.13. The corresponding linear system of equations. The corresponding linear system of equations is derived analogously to the symmetric case. The system matrix is still positive definite but not symmetric. □

Remark B.14. Choice of the basis. The most important issue of the Ritz and Galerkin method is the choice of the spaces V_k, or more concretely, the choice of an appropriate basis $\{\phi_i\}_{i=1}^k$ that spans the space V_k. From the point of view of numerics, there are the following requirements:

- it should be possible to compute the entries a_{ij} of the stiffness matrix efficiently,
- the matrix A should be sparse.

□

B.2 Finite Element Spaces

Remark B.15. Mesh cells, faces, edges, vertices. A mesh cell K is a compact polyhedron in $\mathbb{R}^d$, $d \in \{2, 3\}$, whose interior is not empty. The boundary ∂K of K consists of m-dimensional linear manifolds (points, pieces of straight lines, pieces of planes), $0 \leq m \leq d - 1$, which are called m-faces. The 0-faces are the vertices of the mesh cell, the 1-faces are the edges, and the $(d-1)$-faces are just called faces. □

Remark B.16. Finite dimensional spaces defined on K. Let $s \in \mathbb{N}$. Finite element methods use finite dimensional spaces $P(K) \subset C^s(K)$ which are defined on K. In general, $P(K)$ consists of polynomials. The dimension of $P(K)$ will be denoted by $\dim P(K) = N_K$. □

*Example B.17. The space $P(K) = P_1(K)$.*The space consisting of linear polynomials on a mesh cell K is denoted by $P_1(K)$:

$$P_1(K) = \left\{ a_0 + \sum_{i=1}^{d} a_i x_i \ : \ \boldsymbol{x} = (x_1, \ldots, x_d)^T \in K \right\}.$$

There are $d+1$ unknown coefficients a_i, $i = 0, \ldots, d$, such that $\dim P_1(K) = N_K = d + 1$. □

Remark B.18. Linear functionals defined on $P(K)$. For the definition of finite elements, linear functional which are defined on $P(K)$ are of importance.

Consider linear and continuous functionals $\Phi_{K,1}, \ldots, \Phi_{K,N_K} : C^s(K) \to \mathbb{R}$ which are linearly independent. There are different types of functionals which can be utilized in finite element methods:

- point values: $\Phi(v) = v(\boldsymbol{x})$, $\boldsymbol{x} \in K$,
- point values of a first partial derivative: $\Phi(v) = \partial_i v(\boldsymbol{x})$, $\boldsymbol{x} \in K$,
- point values of the normal derivative on a face E of K: $\Phi(v) = \nabla v(\boldsymbol{x}) \cdot \boldsymbol{n}_E$, $\boldsymbol{n}_E$ is the outward pointing unit normal vector on E,
- integral mean values on K: $\Phi(v) = \frac{1}{|K|} \int_K v(\boldsymbol{x}) \, d\boldsymbol{x}$,
- integral mean values on faces E: $\Phi(v) = \frac{1}{|E|} \int_E v(\boldsymbol{s}) \, d\boldsymbol{s}$.

The smoothness parameter s has to be chosen in such a way that the functionals $\Phi_{K,1}, \ldots, \Phi_{K,N_K}$ are continuous. If, e.g., a functional requires the evaluation of a partial derivative or a normal derivative, then one has to choose at least $s = 1$. For the other functionals given above, $s = 0$ is sufficient. □

Definition B.19. Unisolvence of $P(K)$ with respect to the functionals $\Phi_{K,1}, \ldots, \Phi_{K,N_K}$. The space $P(K)$ is called unisolvent with respect to the functionals $\Phi_{K,1}, \ldots, \Phi_{K,N_K}$ if there is for each $\mathbf{a} \in \mathbb{R}^{N_K}$, $\mathbf{a} = (a_1, \ldots, a_{N_K})^T$, exactly one $p \in P(K)$ with

$$\Phi_{K,i}(p) = a_i, \quad 1 \leq i \leq N_K.$$

□

Remark B.20. Local basis. Unisolvence means that for each vector $\mathbf{a} \in \mathbb{R}^{N_K}$, $\mathbf{a} = (a_1, \ldots, a_{N_K})^T$, there is exactly one element in $P(K)$ such that a_i is the image of the i-th functional, $i = 1, \ldots, N_K$.

Choosing in particular the Cartesian unit vectors for $\mathbf{a}$, then it follows from the unisolvence that a set $\{\phi_{K,i}\}_{i=1}^{N_K}$ exists with $\phi_{K,i} \in P(K)$ and

$$\Phi_{K,i}(\phi_{K,j}) = \delta_{ij}, \quad i, j = 1, \ldots, N_K.$$

Consequently, the set $\{\phi_{K,i}\}_{i=1}^{N_K}$ forms a basis of $P(K)$. This basis is called local basis. □

Remark B.21. Transform of an arbitrary basis to the local basis. If an arbitrary basis $\{p_i\}_{i=1}^{N_K}$ of $P(K)$ is known, then the local basis can be computed

by solving a linear system of equations. To this end, represent the local basis in terms of the known basis

$$\phi_{K,j} = \sum_{k=1}^{N_K} c_{jk} p_k, \quad c_{jk} \in \mathbb{R}, \ j = 1, \ldots, N_K,$$

with unknown coefficients c_{jk}. Applying the definition of the local basis leads to the linear system of equations

$$\Phi_{K,i}(\phi_{K,j}) = \sum_{k=1}^{N_K} c_{jk} a_{ik} = \delta_{ij}, \quad i, j = 1, \ldots, N_K, \quad a_{ik} = \Phi_{K,i}(p_k).$$

Because of the unisolvence, the matrix $A = (a_{ij})$ is non-singular and the coefficients c_{jk} are determined uniquely. □

Example B.22. Local basis for the space of linear functions on the reference triangle. Consider the reference triangle $\hat{K}$ with the vertices $(0,0)$, $(1,0)$, and $(0,1)$. A linear space on $\hat{K}$ is spanned by the functions $1, \hat{x}, \hat{y}$. Let the functionals be defined by the values of the functions in the vertices of the reference triangle. Then, the given basis is not a local basis because the function 1 does not vanish at the vertices.

Consider first the vertex $(0,0)$. A linear basis function $a\hat{x} + b\hat{y} + c$ which has the value 1 in $(0,0)$ and which vanishes in the other vertices has to satisfy the following set of equations

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

The solution is $a = -1, b = -1, c = 1$. The two other basis functions of the local basis are $\hat{x}$ and $\hat{y}$, such that the local basis has the form $\{1 - \hat{x} - \hat{y}, \hat{x}, \hat{y}\}$. □

Remark B.23. Triangulation, grid, mesh, grid cell. For the definition of global finite element spaces, a decomposition of the domain Ω into polyhedrons K is needed. This decomposition is called triangulation $\mathcal{T}^h$ and the polyhedrons K are called mesh cells. The union of the polyhedrons is called grid or mesh.

A triangulation is called admissible, see the definition in Ciarlet (1978), if: check Def. in Cia92

- It holds $\overline{\Omega} = \cup_{K \in \mathcal{T}^h} K$.
- Each mesh cell $K \in \mathcal{T}^h$ is closed and the interior $\mathring{K}$ is non-empty.
- For distinct mesh cells K_1 and K_2 there holds $\mathring{K}_1 \cap \mathring{K}_2 = \emptyset$.
- For each $K \in \mathcal{T}^h$, the boundary ∂K is Lipschitz-continuous.
- The intersection of two mesh cells is either empty or a common m-face, $m \in \{0, \ldots, d-1\}$.

□

Remark B.24. Global and local functionals. Let $\Phi_1, \ldots, \Phi_N : C^s(\overline{\Omega}) \to \mathbb{R}$ continuous linear functionals of the same types as given in Remark B.18. The restriction of the functionals to $C^s(K)$ defines local functionals $\Phi_{K,1}, \ldots, \Phi_{K,N_K}$, where it is assumed that the local functionals are unisolvent on $P(K)$. The union of all mesh cells K_j, for which there is a $p \in P(K_j)$ with $\Phi_i(p) \neq 0$, will be denoted by ω_i. □

Example B.25. On subdomains ω_i. Consider the two-dimensional case and let Φ_i be defined as nodal value of a function in $\boldsymbol{x} \in K$. If $\boldsymbol{x} \in \mathring{K}$, then $\omega_i = K$. In the case that $\boldsymbol{x}$ is on a face of K but not in a vertex, then ω_i is the union of K and the other mesh cell whose boundary contains this face. Last, if $\boldsymbol{x}$ is a vertex of K, then ω_i is the union of all mesh cells which possess this vertex, see Figure B.1. □

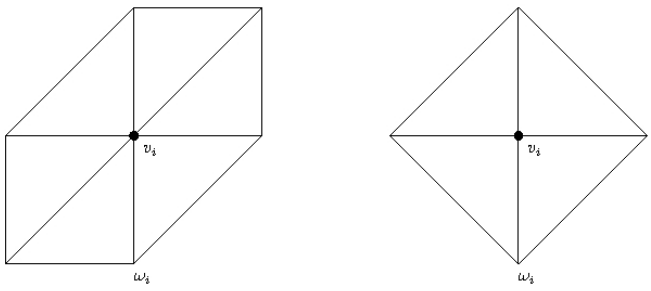

Fig. B.1 Subdomains ω_i.

Definition B.26. Finite element space, global basis. A function $v(\boldsymbol{x})$ defined on Ω with $v|_K \in P(K)$ for all $K \in \mathcal{T}^h$ is called continuous with respect to the functional $\Phi_i : \Omega \to \mathbb{R}$ if

$$\Phi_i(v|_{K_1}) = \Phi_i(v|_{K_2}), \quad \forall\, K_1, K_2 \in \omega_i.$$

The space

$$S = \Big\{ v \in L^\infty(\Omega) \ : \ v|_K \in P(K) \text{ and } v \text{ is continuous with respect to } \Phi_i, i = 1, \ldots, N \Big\}$$

is called finite element space.

The global basis $\{\phi_j\}_{j=1}^N$ of S is defined by the condition

$$\phi_j \in S, \quad \Phi_i(\phi_j) = \delta_{ij}, \quad i, j = 1, \ldots, N.$$

□

Example B.27. Piecewise linear global basis function. Figure B.2 shows a piecewise linear global basis function in two dimensions. Because of its form, such a function is called hat function. □

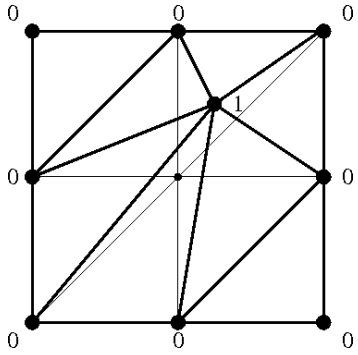

Fig. B.2 Piecewise linear global basis function (boldface lines), hat function.

Remark B.28. On global basis functions. A global basis function coincides on each mesh cell with a local basis function. This property implies the uniqueness of the global basis functions.

For many finite element spaces it follows from the continuity with respect to $\{\Phi_i\}_{i=1}^N$, the continuity of the finite element functions themselves. Only in this case, one can speak of values of finite element functions on m-faces with $m < d$. □

Definition B.29. Parametric finite elements. Let $\hat{K}$ be a reference mesh cell with the local space $P(\hat{K})$, the local functionals $\hat{\Phi}_1, \ldots, \hat{\Phi}_{\hat{N}}$, and a class of bijective mappings $\{F_K : \hat{K} \to K\}$. A finite element space is called a parametric finite element space if:

- The images $\{K\}$ of $\{F_K\}$ form the set of mesh cells.
- The local spaces are given by

$$P(K) = \left\{ p \ : \ p = \hat{p} \circ F_K^{-1}, \hat{p} \in \hat{P}(\hat{K}) \right\}. \tag{B.13}$$

- The local functionals are defined by

$$\Phi_{K,i}(v(\boldsymbol{x})) = \hat{\Phi}_i \left(v(F_K(\hat{\boldsymbol{x}})) \right), \tag{B.14}$$

  where $\hat{\boldsymbol{x}} = (\hat{x}_1, \ldots, \hat{x}_d)^T$ are the coordinates of the reference mesh cell and it holds $\boldsymbol{x} = F_K(\hat{\boldsymbol{x}})$.

□

Remark B.30. Motivations for using parametric finite elements. Definition B.26 of finite elements spaces is very general. For instance, different types of mesh cells are allowed. However, as well the finite element theory as the implementation of finite element methods become much simpler if only parametric finite elements are considered. □

B.3 Finite Elements on Simplices

Definition B.31. d-simplex. A d-simplex $K \subset \mathbb{R}^d$ is the convex hull of $(d+1)$ points $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{d+1} \in \mathbb{R}^d$ which form the vertices of K. □

Remark B.32. On d-simplices. It will be always assumed that the simplex is not degenerated, i.e., its d-dimensional measure is positive. This property is equivalent to the non-singularity of the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1,d+1} \\ a_{21} & a_{22} & \ldots & a_{2,d+1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{d1} & a_{d2} & \ldots & a_{d,d+1} \\ 1 & 1 & \ldots & 1 \end{pmatrix},$$

where $\boldsymbol{a}_i = (a_{1i}, a_{2i}, \ldots, a_{di})^T$, $i = 1, \ldots, d+1$.

For $d = 2$, the simplices are the triangles and for $d = 3$ they are the tetrahedrons. □

Definition B.33. Barycentric coordinates. Since K is the convex hull of the points $\{\boldsymbol{a}_i\}_{i=1}^{d+1}$, the parametrization of K with a convex combination of the vertices reads as follows

$$K = \left\{ \boldsymbol{x} \in \mathbb{R}^d \ : \ \boldsymbol{x} = \sum_{i=1}^{d+1} \lambda_i \boldsymbol{a}_i, \ 0 \le \lambda_i \le 1, \ \sum_{i=1}^{d+1} \lambda_i = 1 \right\}.$$

The coefficients $\lambda_1, \ldots, \lambda_{d+1}$ are called barycentric coordinates of $\boldsymbol{x} \in K$. □

Remark B.34. On barycentric coordinates. From the definition it follows that the barycentric coordinates are the solution of the linear system of equations

$$\sum_{i=1}^{d+1} a_{ji} \lambda_i = x_j, \quad 1 \le j \le d, \quad \sum_{i=1}^{d+1} \lambda_i = 1.$$

Since the system matrix is non-singular, see Remark B.32, the barycentric coordinates are determined uniquely.

The barycentric coordinates of the vertex $\boldsymbol{a}_i$, $i = 1, \ldots, d+1$, of the simplex is $\lambda_i = 1$ and $\lambda_j = 0$ if $i \neq j$. Since $\lambda_i(\boldsymbol{a}_j) = \delta_{ij}$, the barycentric coordinate

λ_i can be identified with the linear function which has the value 1 in the vertex $\boldsymbol{a}_i$ and which vanishes in all other vertices $\boldsymbol{a}_j$ with $j \neq i$.

The barycenter of the simplex is given by

$$S_K = \frac{1}{d+1} \sum_{i=1}^{d+1} \boldsymbol{a}_i = \sum_{i=1}^{d+1} \frac{1}{d+1} \boldsymbol{a}_i.$$

Hence, its barycentric coordinates are $\lambda_i = 1/(d+1)$, $i = 1, \ldots, d+1$. □

Remark B.35. Simplicial reference mesh cells. A commonly used reference mesh cell for triangles and tetrahedrons is the unit simplex

$$\hat{K} = \left\{ \hat{\boldsymbol{x}} \in \mathbb{R}^d \ : \ \sum_{i=1}^{d} \hat{x}_i \leq 1, \ \hat{x}_i \geq 0, \ i = 1, \ldots, d \right\},$$

see Figure B.3. The class $\{F_K\}$ of admissible mappings are the bijective affine mappings

$$F_K \hat{\boldsymbol{x}} = B\hat{\boldsymbol{x}} + \boldsymbol{b}, \qquad B \in \mathbb{R}^{d \times d}, \ \det(B) \neq 0, \ \boldsymbol{b} \in \mathbb{R}^d. \tag{B.15}$$

The images of these mappings generate the set of the non-degenerated simplices $\{K\} \subset \mathbb{R}^d$. □

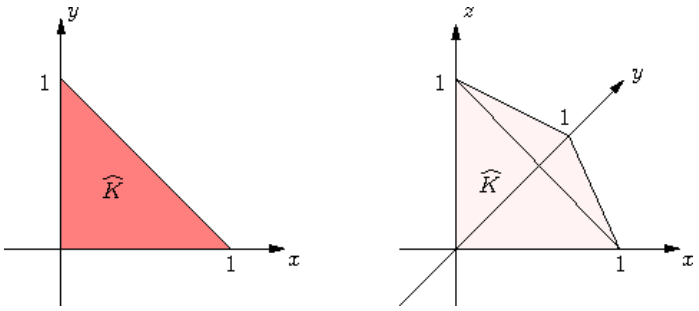

Fig. B.3 The unit simplices in two and three dimensions.

Definition B.36. Affine family of simplicial finite elements. Given a simplicial reference mesh cell $\hat{K}$, affine mappings $\{F_K\}$, and an unisolvent set of functionals on $\hat{K}$. Using (B.13) and (B.14), one obtains a local finite element space on each non-degenerated simplex. The set of these local spaces is called affine family of simplicial finite elements. □

Definition B.37. Polynomial space P_k. Let $\boldsymbol{x} = (x_1, \ldots, x_d)^T$, $k \in \mathbb{N} \cup \{0\}$, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^T$. Then, the polynomial space P_k is given by

$$P_k = \operatorname{span}\left\{ \prod_{i=1}^{d} x_i^{\alpha_i} = \boldsymbol{x}^{\boldsymbol{\alpha}} \ : \ \alpha_i \in \mathbb{N} \cup \{0\} \ \text{ for } \ i = 1, \ldots, d, \ \sum_{i=1}^{d} \alpha_i \leq k \right\}.$$

□

Remark B.38. Lagrangian finite elements. In all examples given below, the linear functionals on the reference mesh cell $\hat{K}$ are the values of the polynomials with the same barycentric coordinates as on the general mesh cell K. Finite elements whose linear functionals are values of the polynomials on certain points in K are called Lagrangian finite elements. □

Example B.39. P_0 : piecewise constant finite element. The piecewise constant finite element space consists of discontinuous functions. The linear functional is the value of the polynomial in the barycenter of the mesh cell, see Figure B.4. It is $\dim P_0(K) = 1$. □

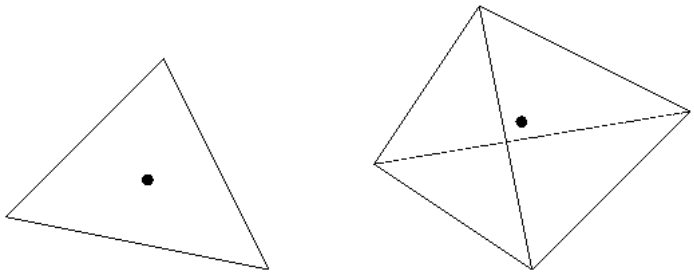

Fig. B.4 The finite element $P_0(K)$.

Example B.40. P_1 : conforming piecewise linear finite element. This finite element space is a subspace of $C(\overline{\Omega})$. The linear functionals are the values of the function in the vertices of the mesh cells, see Figure B.5. It follows that $\dim P_1(K) = d + 1$.

The local basis for the functionals $\{\Phi_i(v) = v(\boldsymbol{a}_i),\ i = 1, \ldots, d+1\}$, is $\{\lambda_i\}_{i=1}^{d+1}$ since $\Phi_i(\lambda_j) = \delta_{ij}$, see Remark B.34. Since a local basis exists, the functionals are unisolvent with respect to the polynomial space $P_1(K)$.

Now, it will be shown that the corresponding finite element space consists of continuous functions. Let K_1, K_2 be two mesh cells with the common face E and let $v \in P_1(= S)$. The restriction of v_{K_1} on E is a linear function on E as well as the restriction of v_{K_2} on E. It has to be shown that both linear functions are identical. A linear function on the $(d-1)$-dimensional face E is uniquely determined with d linearly independent functionals which are defined on E. These functionals can be chosen to be the values of the function in the d vertices of E. The functionals in S are continuous, by the definition of S. Thus, it must hold that both restrictions on E have the same

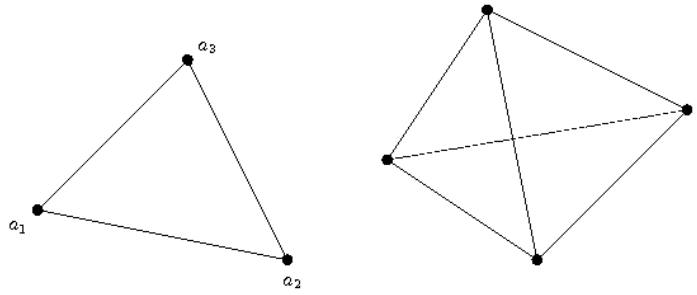

Fig. B.5 The finite element $P_1(K)$.

values in the vertices of E. Hence, it is $v_{K_1}|_E = v_{K_2}|_E$ and the functions from P_1 are continuous. □

Example B.41. P_2 : conforming piecewise quadratic finite element. This finite element space is also a subspace of $C(\overline{\Omega})$. It consists of piecewise quadratic functions. The functionals are the values of the functions in the $d+1$ vertices of the mesh cell and the values of the functions in the centers of the edges, see Figure B.6. Since each vertex is connected to each other vertex, there are $\sum_{i=1}^{d} i = d(d+1)/2$ edges. Hence, it follows that $\dim P_2(K) = (d+1)(d+2)/2$.

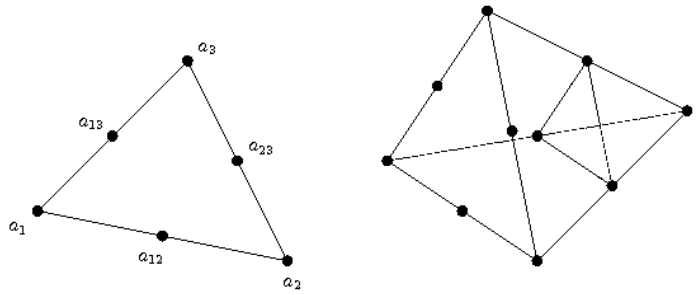

Fig. B.6 The finite element $P_2(K)$.

The part of the local basis which belongs to the functionals $\{\Phi_i(v) = v(\boldsymbol{a}_i), i = 1, \ldots, d+1\}$, is given by

$$\{\phi_i(\lambda) = \lambda_i(2\lambda_i - 1), \quad i = 1, \ldots, d+1\}.$$

Denote the center of the edges between the vertices $\boldsymbol{a}_i$ and $\boldsymbol{a}_j$ by $\boldsymbol{a}_{ij}$. The corresponding part of the local basis is given by

$$\{\phi_{ij} = 4\lambda_i\lambda_j, \quad i, j = 1, \ldots, d+1, \ i < j\}.$$

The unisolvence follows from the fact that there exists a local basis. The continuity of the corresponding finite element space is shown in the same way as for the P_1 finite element. The restriction of a quadratic function in a mesh cell to a face E is a quadratic function on that face. Hence, the function on E is determined uniquely with $d(d+1)/2$ linearly independent functionals on E.

The functions ϕ_{ij} are called in two dimensions edge bubble functions. □

Example B.42. P_3 : conforming piecewise cubic finite element. This finite element space consists of continuous piecewise cubic functions. It is a subspace of $C(\overline{\Omega})$. The functionals in a mesh cell K are defined to be the values in the vertices ($(d+1)$ values), two values on each edge (dividing the edge in three parts of equal length) ($2\sum_{i=1}^{d} i = d(d+1)$ values), and the values in the barycenter of the 2-faces of K, see Figure B.7. Each 2-face of K is defined by three vertices. If one considers for each vertex all possible pairs with other vertices, then each 2-face is counted three times. Hence, there are $(d+1)(d-1)d/6$ 2-faces. The dimension of $P_3(K)$ is given by

$$\dim P_3(K) = (d+1) + d(d+1) + \frac{(d-1)d(d+1)}{6} = \frac{(d+1)(d+2)(d+3)}{6}.$$

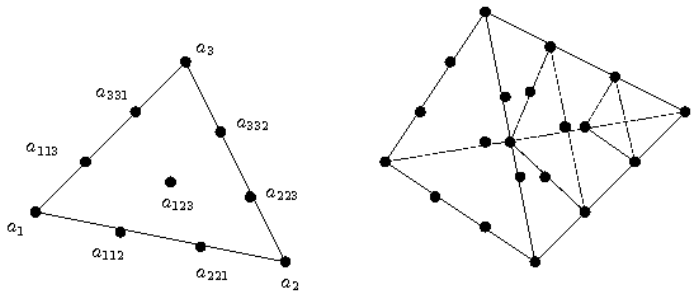

Fig. B.7 The finite element $P_3(K)$.

For the functionals

$$\begin{aligned}
&\Big\{\Phi_i(v) = v(\boldsymbol{a}_i),\ i = 1,\dots,d+1, && \text{(vertex)},\\
&\Phi_{iij}(v) = v(\boldsymbol{a}_{iij}),\ i,j = 1,\dots,d+1, i \neq j, && \text{(point on edge)},\\
&\Phi_{ijk}(v) = v(\boldsymbol{a}_{ijk}),\ i = 1,\dots,d+1, i < j < k && \text{(point on 2-face)}\Big\},
\end{aligned}$$

the local basis is given by

$$\left\{\begin{aligned}\phi_i(\lambda) &= \frac{1}{2}\lambda_i(3\lambda_i-1)(3\lambda_i-2),\\ \phi_{iij}(\lambda) &= \frac{9}{2}\lambda_i\lambda_j(3\lambda_i-1),\\ \phi_{ijk}(\lambda) &= 27\lambda_i\lambda_j\lambda_k\end{aligned}\right\}.$$

In two dimensions, the function $\phi_{ijk}(\lambda)$ is called cell bubble function. □

Example B.43. P_1^{bubble}. The P_1^{bubble} finite element is just the P_1 finite element enriched with mesh cell bubbles. In two dimensions, the functionals are given by the point values of a function $v(\boldsymbol{x})$ in the vertices $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$, and by

$$\frac{3\big[v(\boldsymbol{a}_1)+v(\boldsymbol{a}_2)+v(\boldsymbol{a}_3)\big]+8\big[v\left(\boldsymbol{a}_{12}\right)+v\left(\boldsymbol{a}_{13}\right)+v\left(\boldsymbol{a}_{23}\right)\big]+27v\left(\boldsymbol{a}_{123}\right)}{27},$$

see Figures B.6 and B.7 for the notations. The corresponding local basis is

$$\big\{\lambda_1-20\lambda_1\lambda_2\lambda_3,\ \lambda_1-20\lambda_1\lambda_2\lambda_3,\ \lambda_1-20\lambda_1\lambda_2\lambda_3,\ 27\lambda_1\lambda_2\lambda_3\big\}.$$

□

Example B.44. P_2^{bubble}. In this space, the P_2 finite element is enriched with bubble functions.

In two dimensions, one can take as nodal functionals the same functionals as for the P_2 element and as seventh functional

$$\frac{3\big[v(\boldsymbol{a}_1)+v(\boldsymbol{a}_2)+v(\boldsymbol{a}_3)\big]+8\big[v\left(\boldsymbol{a}_{12}\right)+v\left(\boldsymbol{a}_{13}\right)+v\left(\boldsymbol{a}_{23}\right)\big]+27v\left(\boldsymbol{a}_{123}\right)}{20},$$

see Figures B.6 and B.7 for the notations. Then, the local basis is given by

$$\begin{aligned}&\big\{4\lambda_1\lambda_2-20\lambda_1\lambda_2\lambda_3,\ 4\lambda_1\lambda_3-20\lambda_1\lambda_2\lambda_3,\ 4\lambda_2\lambda_3-20\lambda_1\lambda_2\lambda_3,\\ &2\lambda_1(\lambda_1-0.5),\ 2\lambda_2(\lambda_2-0.5),\ 2\lambda_1(\lambda_2-0.5),\ 20\lambda_1\lambda_2\lambda_3\big\}.\end{aligned}$$

The three-dimensional case, the enrichment is performed with the mesh cell bubble function and with the four bubble functions on the faces. The functionals are the four values in the vertices, the six values on the mid points of the edges, the four values in the barycenters of the faces, and the value in the barycenter of the mesh cell. Altogether, there are 15 functionals. The local basis is given by

$$\begin{aligned}&\big\{\lambda_1(2\lambda_1-1)+3\lambda_1(\lambda_2\lambda_3+\lambda_2\lambda_4+\lambda_3\lambda_4)-4\lambda_1\lambda_2\lambda_3\lambda_4,\dots,\\ &\lambda_1\lambda_2(4-12\lambda_4-12\lambda_3+32\lambda_3\lambda_4),\dots,\\ &27\lambda_1\lambda_2\lambda_3(1-4\lambda_4),\ 27\lambda_1\lambda_2(1-4\lambda_3)\lambda_4,\ 27\lambda_1(1-4\lambda_2)\lambda_3\lambda_4,\ 27(1-4\lambda_1)\lambda_2\lambda_3\lambda_4,\\ &256\lambda_1\lambda_2\lambda_3\lambda_4\big\},\end{aligned}$$

where the remaining basis functions are given by appropriate permutations of the indices. □

Example B.45. P_1^{nc} *: nonconforming linear finite element, Crouzeix–Raviart finite element Crouzeix and Raviart (1973).* This finite element consists of piecewise linear but discontinuous functions. The functionals are given by the values of the functions in the barycenters of the faces such that $\dim P_1^{\mathrm{nc}}(K) = (d+1)$. It follows from the definition of the finite element space, Definition B.26, that the functions from P_1^{nc} are continuous in the barycenter of the faces

$$P_1^{\mathrm{nc}} = \big\{ v \in L^2(\Omega) \ : \ v|_K \in P_1(K),\ v(\boldsymbol{x}) \text{ is continuous at the barycenter of all faces} \big\}. \tag{B.16}$$

Equivalently, the functionals can be defined to be the integral mean values on the faces and then the global space is defined to be

$$P_1^{\mathrm{nc}} = \left\{ v \in L^2(\Omega) \ : \ v|_K \in P_1(K), \right. \\ \left. \int_E v|_K \ d\boldsymbol{s} = \int_E v|_{K'} \ d\boldsymbol{s} \ \forall\, E \in \mathcal{E}(K) \cap \mathcal{E}(K') \right\}, \tag{B.17}$$

where $\mathcal{E}(K)$ is the set of all $(d-1)$ dimensional faces of K.

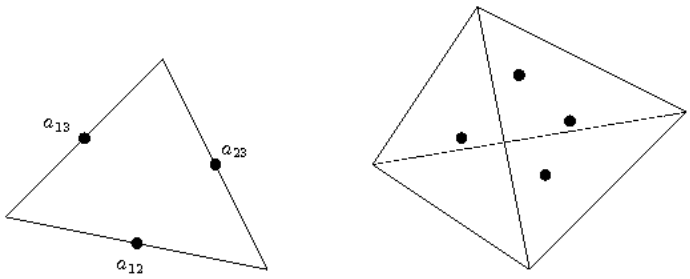

Fig. B.8 The finite element P_1^{nc}.

For the description of this finite element, one defines the functionals by

$$\Phi_i(v) = v(\boldsymbol{a}_{i-1,i+1}) \text{ for } d=2, \quad \Phi_i(v) = v(\boldsymbol{a}_{i-2,i-1,i+1}) \text{ for } d=3,$$

where the points are the barycenters of the faces with the vertices that correspond to the indices. This system is unisolvent with the local basis

$$\phi_i(\lambda) = 1 - d\lambda_i, \quad i = 1, \ldots, d+1.$$

□

Example B.46. P_1^{disc}. This space consists of piecewise linear but discontinuous functions.

On the reference mesh cell $\hat{K}$ in two dimensions, one can use the functionals applied to a function $v(\hat{\boldsymbol{x}})$ given by

$$\int_{\hat{K}} 2v(\hat{\boldsymbol{x}})\ d\hat{\boldsymbol{x}}, \quad \int_{\hat{K}} (24\hat{x} - 8)\, v(\hat{\boldsymbol{x}})\ d\hat{\boldsymbol{x}}, \quad \int_{\hat{K}} (24\hat{y} - 8)\, v(\hat{\boldsymbol{x}})\ d\hat{\boldsymbol{x}}$$

and the corresponding local basis is

$$\left\{1,\ \hat{\lambda}_2 - \hat{\lambda}_1,\ \hat{\lambda}_3 - \hat{\lambda}_1\right\} = \{1,\ 2\hat{x} + \hat{y} - 1,\ \hat{x} + 2\hat{y} - 1\}\,.$$

In three dimensions, let $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \boldsymbol{a}_4$ be the vertices of the tetrahedron and S_K its barycenter. Then, the following functionals can be used

$$\frac{v(\boldsymbol{a}_1) + v(\boldsymbol{a}_2) + v(\boldsymbol{a}_3) + v(\boldsymbol{a}_4) + 16v(S_K)}{120}, \quad \frac{-v(\boldsymbol{a}_1) + 3v(\boldsymbol{a}_2) - v(\boldsymbol{a}_3) - v(\boldsymbol{a}_4)}{4},$$
$$\frac{-v(\boldsymbol{a}_1) - v(\boldsymbol{a}_2) + 3v(\boldsymbol{a}_3) - v(\boldsymbol{a}_4)}{4}, \quad \frac{-v(\boldsymbol{a}_1) - v(\boldsymbol{a}_2) - v(\boldsymbol{a}_3) + 3v(\boldsymbol{a}_4)}{4}.$$

The corresponding local basis is given by

$$\{6,\ \lambda_2 - \lambda_1,\ \lambda_3 - \lambda_1,\ \lambda_4 - \lambda_1\}.$$

□

B.4 Finite Elements on Parallelepipeds

Remark B.47. Reference mesh cells, reference map. On can find in the literature two reference cells: the unit cube $[0,1]^d$ and the large unit cube $[-1,1]^d$. It does not matter which reference cell is chosen. Here, the large unit cube will be used: $\hat{K} = [-1,1]^d$. The class of admissible reference maps $\{F_K\}$ consists of bijective affine mappings of the form

$$F_K\hat{\boldsymbol{x}} = B\hat{\boldsymbol{x}} + \boldsymbol{b}, \quad B \in \mathbb{R}^{d\times d},\ \boldsymbol{b} \in \mathbb{R}^d.$$

If B is a diagonal matrix, then $\hat{K}$ is mapped to d-rectangles.

The class of mesh cells which are obtained in this way is not sufficient to triangulate general domains. If one wants to use more general mesh cells than parallelepipeds, then the class of admissible reference maps has to be enlarged, see Section B.5. □

Definition B.48. Polynomial space Q_k. Let $\boldsymbol{x} = (x_1, \ldots, x_d)^T$ and denote by $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^T$ a multi-index. Then, the polynomial space Q_k is given

by

$$Q_k = \operatorname{span}\left\{ \prod_{i=1}^{d} x_i^{\alpha_i} = \boldsymbol{x}^{\boldsymbol{\alpha}} \ : \ 0 \le \alpha_i \le k \ \text{ for } \ i = 1, \ldots, d \right\}.$$

□

Example B.49. Q_1 vs. P_1. The space Q_1 consists of all polynomials which are d-linear. Let $d = 2$, then it is

$$Q_1 = \operatorname{span}\{1, x, y, xy\},$$

whereas

$$P_1 = \operatorname{span}\{1, x, y\}.$$

□

Remark B.50. Finite elements on d-rectangles. For simplicity of presentation, the examples below consider d-rectangles. In this case, the finite elements are just tensor products of one-dimensional finite elements. In particular, the basis functions can be written as products of one-dimensional basis functions. □

Example B.51. Q_0 : piecewise constant finite element. Similarly to the P_0 space, the space Q_0 consists of piecewise constant, discontinuous functions. The functional is the value of the function in the barycenter of the mesh cell K and it holds $\dim Q_0(K) = 1$. □

Example B.52. Q_1 : conforming piecewise d-linear finite element. This finite element space is a subspace of $C(\overline{\Omega})$. The functionals are the values of the function in the vertices of the mesh cell, see Figure B.9. Hence, it is $\dim Q_1(K) = 2^d$.

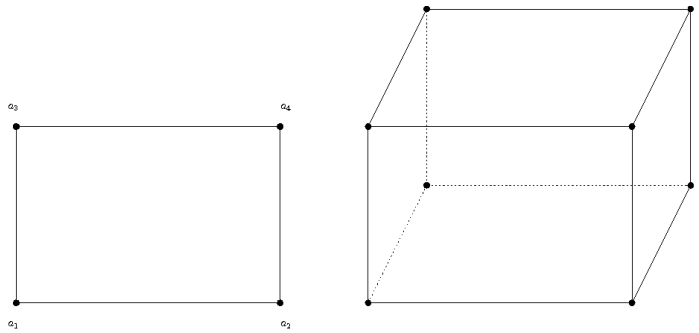

Fig. B.9 The finite element Q_1.

The one-dimensional local basis functions, which will be used for the tensor product, are given by

$$\hat{\phi}_1(\hat{x}) = \frac{1}{2}(1-\hat{x}), \quad \hat{\phi}_2(\hat{x}) = \frac{1}{2}(1+\hat{x}).$$

With these functions, e.g., the basis functions in two dimensions are computed by

$$\hat{\phi}_1(\hat{x})\hat{\phi}_1(\hat{y}),\ \hat{\phi}_1(\hat{x})\hat{\phi}_2(\hat{y}),\ \hat{\phi}_2(\hat{x})\hat{\phi}_1(\hat{y}),\ \hat{\phi}_2(\hat{x})\hat{\phi}_2(\hat{y}).$$

The continuity of the functions of the finite element space Q_1 is proved in the same way as for simplicial finite elements. It is used that the restriction of a function from $Q_k(K)$ to a face E is a function from the space $Q_k(E)$, $k \geq 1$. □

Example B.53. Q_2 : conforming piecewise d-quadratic finite element. It holds that $Q_2 \subset C(\overline{\Omega})$. The functionals in one dimension are the values of the function at both ends of the interval and in the center of the interval, see Figure B.10. In d dimensions, they are the corresponding values of the tensor product of the intervals. It follows that $\dim Q_2(K) = 3^d$.

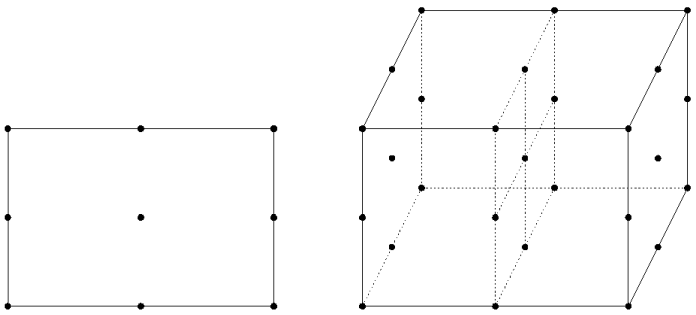

Fig. B.10 The finite element Q_2.

The one-dimensional basis function on the reference interval are defined by

$$\hat{\phi}_1(\hat{x}) = -\frac{1}{2}\hat{x}(1-\hat{x}), \quad \hat{\phi}_2(\hat{x}) = (1-\hat{x})(1+\hat{x}), \quad \hat{\phi}_3(\hat{x}) = \frac{1}{2}(1+\hat{x})\hat{x}.$$

The basis function $\prod_{i=1}^{d} \hat{\phi}_2(\hat{x}_i)$ is called cell bubble function. □

Example B.54. Q_3 : conforming piecewise d-quadratic finite element. This finite element space is a subspace of $C(\overline{\Omega})$. The functionals on the reference interval are given by the values at the end of the interval and the values at the points $\hat{x} = -1/3$, $\hat{x} = 1/3$. In multiple dimensions, it is the corresponding tensor product, see Figure B.11. The dimension of the local space is $\dim Q_3(K) = 4^d$.

The one-dimensional basis functions in the reference interval are given by

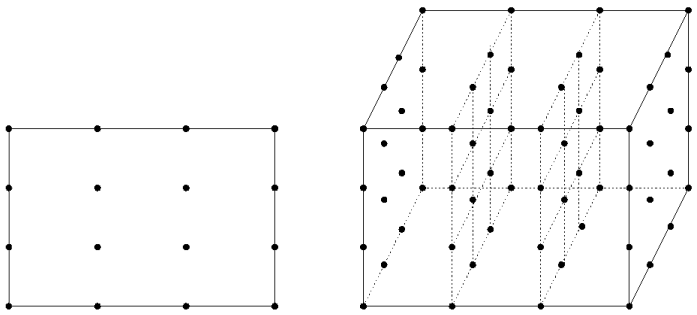

Fig. B.11 The finite element Q_3.

$$\hat{\phi}_1(\hat{x}) = -\frac{1}{16}(3\hat{x}+1)(3\hat{x}-1)(\hat{x}-1),$$
$$\hat{\phi}_2(\hat{x}) = \frac{9}{16}(\hat{x}+1)(3\hat{x}-1)(\hat{x}-1),$$
$$\hat{\phi}_3(\hat{x}) = -\frac{9}{16}(\hat{x}+1)(3\hat{x}+1)(\hat{x}-1),$$
$$\hat{\phi}_4(\hat{x}) = \frac{1}{16}(3\hat{x}+1)(3\hat{x}-1)(\hat{x}+1).$$

□

Example B.55. Q_1^{rot} *: rotated nonconforming element of lowest order, Rannacher–Turek element Rannacher and Turek (1992):* This finite element space is a generalization of the P_1^{nc} finite element to quadrilateral and hexahedral mesh cells. It consists of discontinuous functions which are continuous at the barycenter of the faces. The dimension of the local finite element space is $\dim Q_1^{\mathrm{rot}}(K) = 2d$. The space on the reference mesh cell is defined by

$$Q_1^{\mathrm{rot}}\left(\hat{K}\right) = \left\{\hat{p} \ : \ \hat{p} \in \mathrm{span}\{1, \hat{x}, \hat{y}, \hat{x}^2 - \hat{y}^2\}\right\} \qquad \text{for } d = 2,$$
$$Q_1^{\mathrm{rot}}\left(\hat{K}\right) = \left\{\hat{p} \ : \ \hat{p} \in \mathrm{span}\{1, \hat{x}, \hat{y}, \hat{z}, \hat{x}^2 - \hat{y}^2, \hat{y}^2 - \hat{z}^2\}\right\} \text{for } d = 3.$$

Note that the transformed space

$$Q_1^{\mathrm{rot}}(K) = \{p = \hat{p} \circ F_K^{-1}, \hat{p} \in Q_1^{\mathrm{rot}}(\hat{K})\}$$

contains polynomials of the form $ax^2 - by^2$, where a, b depend on F_K.

For $d = 2$, the local basis on the reference cell is given by

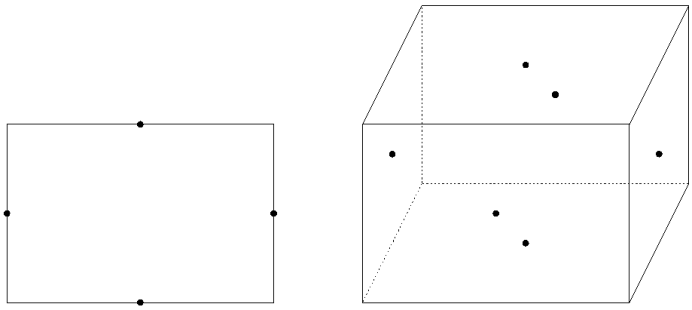

Fig. B.12 The finite element Q_1^{rot}.

$$\begin{aligned}
\phi_1(\hat{x},\hat{y}) &= -\frac{3}{8}(\hat{x}^2-\hat{y}^2) - \frac{1}{2}\hat{y} + \frac{1}{4},\\
\phi_2(\hat{x},\hat{y}) &= \frac{3}{8}(\hat{x}^2-\hat{y}^2) + \frac{1}{2}\hat{x} + \frac{1}{4},\\
\phi_3(\hat{x},\hat{y}) &= -\frac{3}{8}(\hat{x}^2-\hat{y}^2) + \frac{1}{2}\hat{y} + \frac{1}{4},\\
\phi_4(\hat{x},\hat{y}) &= \frac{3}{8}(\hat{x}^2-\hat{y}^2) - \frac{1}{2}\hat{x} + \frac{1}{4}.
\end{aligned}$$

Analogously to the Crouzeix–Raviart finite element, the functionals can be defined as point values of the functions in the barycenters of the faces, see Figure B.12, or as integral mean values of the functions at the faces. Consequently, the finite element spaces are defined in the same way as (B.16) or (B.17), with $P_1^{\mathrm{nc}}(K)$ replaced by $Q_1^{\mathrm{rot}}(K)$.

In the code MooNMD John and Matthies (2004), the mean value oriented Q_1^{rot} finite element space is implemented fro two dimensions and the point value oriented Q_1^{rot} finite element space for three dimensions. For $d = 3$, the integrals on the faces of mesh cells, whose equality is required in the mean value oriented Q_1^{rot} finite element space, involve a weighting function which depends on the particular mesh cell K. The computation of these weighting functions for all mesh cells is an additional computational overhead. For this reason, Schieweck (Schieweck, 1997, p. 21) suggested to use for $d = 3$ the simpler point value oriented form of the Q_1^{rot} finite element. □

Example B.56. P_1^{disc}.

$$P_1^{\mathrm{disc}} = \left\{ v \in L^2(\Omega) \ : \ v|_K \in P_1(K) \right\}.$$

functionals

$$\frac{1}{2^d}\int_{\hat{K}} \phi(\hat{\boldsymbol{x}})\, d\boldsymbol{x}, \quad \frac{2^{d-1}}{2^d}\int_{\hat{K}} \xi_i \phi(\hat{\boldsymbol{x}})\, d\boldsymbol{x}, \quad i = 1,\ldots,d$$

basis 2D: $1, 3\hat{x}, 3\hat{y}$.
basis 3D: $1, 3\hat{x}, 3\hat{y}, \hat{z}$. □

Example B.57. P_1^{disc}. This space consists of piecewise linear but discontinuous functions.

On the reference mesh cell $\hat{K}$, one can use the functionals

$$\frac{1}{2^d}\int_{\hat{K}} v(\hat{\boldsymbol{x}})\; d\hat{\boldsymbol{x}},$$
$$\frac{3}{2^d}\int_{\hat{K}} \hat{x}_i v(\hat{\boldsymbol{x}})\; d\hat{\boldsymbol{x}}, \quad i = 1, \ldots, d.$$

The corresponding local basis is given by

$$\{1, \hat{x}_1, \ldots, \hat{x}_d\}.$$

Im Programm ist es anders, probiere diese Wahl. □

Example B.58. P_2^{disc}.

$$\frac{1}{2^d}\int_{\hat{K}} \phi(\hat{\boldsymbol{x}})\; d\boldsymbol{x}, \quad \frac{3}{2^d}\int_{\hat{K}} \xi_i\phi(\hat{\boldsymbol{x}})\; d\boldsymbol{x}, \quad i = 1, \ldots, d$$

$$\frac{1}{2^d}\int_{\hat{K}} \xi_i\xi_j\phi(\hat{\boldsymbol{x}})\, d\boldsymbol{x}, \quad i = 1, \ldots, d, i < j, \quad \frac{1}{2^d}\int_{\hat{K}} \xi_i^2\phi(\hat{\boldsymbol{x}})\, d\boldsymbol{x}, \quad i = 1, \ldots, d, i < j$$

basis 2D: $1, 3\hat{x}, 3\hat{y}, 5(3\hat{x}^2 - 1)/4, 9\hat{x}\hat{y}, 5(3\hat{y}^2 - 1)/4$.
basis 3D: $1, 3\hat{x}, 3\hat{y}, \hat{z}, 3\hat{x}^2 - 1, \hat{x}\hat{y}, 3\hat{y}^2 - 1, \hat{x}\hat{z}, 3\hat{z}^2 - 1, \hat{y}\hat{z}$. □

Example B.59. P_2^{disc}. This space consists of discontinuous, piecewise quadratic functions.

On the reference mesh cell $\hat{K}$, one can apply the functionals

$$\frac{1}{2^d}\int_{\hat{K}} v(\hat{\boldsymbol{x}})\; d\hat{\boldsymbol{x}},$$
$$\frac{3}{2^d}\int_{\hat{K}} \hat{x}_i v(\hat{\boldsymbol{x}})\; d\hat{\boldsymbol{x}}, \quad i = 1, \ldots, d,$$
$$\frac{9}{2^d}\int_{\hat{K}} \hat{x}_i\hat{x}_j v(\hat{\boldsymbol{x}})\; d\hat{\boldsymbol{x}}, \quad i = 1, \ldots, d,\ j = i+1, \ldots, d,$$
$$\frac{5}{92^d}\int_{\hat{K}} \left(3\hat{x}_i^2 - 1\right) v(\hat{\boldsymbol{x}})\; d\hat{\boldsymbol{x}}, \quad i = 1, \ldots, d.$$

Then, the corresponding local basis is given by

$$\{1, \hat{x}_1, \ldots, \hat{x}_d, \hat{x}_1\hat{x}_2, \ldots, \hat{x}_1, \hat{x}_d, \ldots, \hat{x}_{d-1}\hat{x}_d, 3\hat{x}_1^2 - 1, \ldots, 3\hat{x}_d^2 - 1\}.$$

Im Programm ist es anders, probiere diese Wahl. □

B.5 Parametric Finite Elements on General *d*-Dimensional Quadrilaterals

Remark B.60. Parametric mappings. The image of an affine mapping of the reference mesh cell $\hat{K} = [-1,1]^d$, $d \in \{2,3\}$, is a parallelepiped. If one wants to consider finite elements on general q-quadrilaterals, then the class of admissible reference maps has to be enlarged.

The simplest parametric finite element on quadrilaterals in two dimensions uses bilinear mappings. Let $\hat{K} = [-1,1]^2$ and let

$$F_K(\hat{\boldsymbol{x}}) = \begin{pmatrix} F_K^1(\hat{\boldsymbol{x}}) \\ F_K^2(\hat{\boldsymbol{x}}) \end{pmatrix} = \begin{pmatrix} a_{11} + a_{12}\hat{x} + a_{13}\hat{y} + a_{14}\hat{x}\hat{y} \\ a_{21} + a_{22}\hat{x} + a_{23}\hat{y} + a_{24}\hat{x}\hat{y} \end{pmatrix}, F_K^i \in Q_1,\ i = 1,2,$$

be a bilinear mapping from $\hat{K}$ on the class of admissible quadrilaterals. A quadrilateral K is called admissible if

- the length of all edges of K is larger than zero,
- the interior angles of K are smaller than π, i.e. K is convex.

This class contains, e.g., trapezoids and rhombi. □

Remark B.61. Parametric finite element functions. The functions of the local space $P(K)$ on the mesh cell K are defined by $p = \hat{p} \circ F_K^{-1}$. These functions are in general rational functions. However, using d-linear mappings, then the restriction of F_K on an edge of $\hat{K}$ is an affine map. For instance, in the case of the Q_1 finite element, the functions on K are linear functions on each edge of K for this reason. It follows that the functions of the corresponding finite element space are continuous, see Example B.40. □

B.6 Transform of Integrals

Remark B.62. Motivation. The transformation of integrals from the reference mesh cell to mesh cells of the grid and vice versa is used as well for analysis as for the implementation of finite element methods. This section provides an overview of the most important formulae for transformations.

Let $\hat{K} \subset \mathbb{R}^d$ be the reference mesh cell, K be an arbitrary mesh cell, and $F_K : \hat{K} \to K$ with $\boldsymbol{x} = F_K(\boldsymbol{\xi})$ be the reference map. It is assumed that the reference map is a continuous differentiable one-to-one map. The inverse map is denoted by $F_K^{-1} : K \to \hat{K}$. For the integral transforms, the derivatives (Jacobians) of F_K and F_K^{-1} are needed

$$DF_K(\boldsymbol{\xi})_{ij} = \frac{\partial x_i}{\partial \xi_j}, \quad DF_K^{-1}(\boldsymbol{x})_{ij} = \frac{\partial \xi_i}{\partial x_j}, \quad i,j = 1,\ldots,d.$$

□

Remark B.63. Integral with a function without derivatives. This integral transforms with the standard rule of integral transforms

$$\int_K v(\boldsymbol{x})\ d\boldsymbol{x} = \int_{\hat{K}} \hat{v}(\boldsymbol{\xi})\,|\det DF_K(\boldsymbol{\xi})|\ d\boldsymbol{\xi}, \tag{B.18}$$

where $\hat{v}(\boldsymbol{\xi}) = v(F_K(\boldsymbol{\xi}))$. □

Remark B.64. Transform of derivatives. Using the chain rule, one obtains

$$\begin{aligned}\frac{\partial v}{\partial x_i}(\boldsymbol{x}) &= \sum_{j=1}^{d} \frac{\partial \hat{v}}{\partial \xi_j}(\boldsymbol{\xi})\frac{\partial \xi_j}{\partial x_i} = \nabla_{\boldsymbol{\xi}}\hat{v}(\boldsymbol{\xi}) \cdot \left(\left(DF_K^{-1}(\boldsymbol{x})\right)^T\right)_i \\ &= \nabla_{\boldsymbol{\xi}}\hat{v}(\boldsymbol{\xi}) \cdot \left(\left(DF_K^{-1}(F_K(\boldsymbol{\xi}))\right)^T\right)_i, \qquad \text{(B.19)} \\ \frac{\partial \hat{v}}{\partial \xi}(\boldsymbol{\xi}) &= \sum_{j=1}^{d} \frac{\partial v}{\partial x_j}(\boldsymbol{x})\frac{\partial x_j}{\partial \xi_i} = \nabla v(\boldsymbol{x}) \cdot \left(\left(DF_K(\boldsymbol{\xi})\right)^T\right)_i \\ &= \nabla v(\boldsymbol{x}) \cdot \left(\left(DF_K(F_K^{-1}(\boldsymbol{x}))\right)^T\right)_i. \qquad \text{(B.20)}\end{aligned}$$

The index i denotes the i-th row of a matrix. Derivatives on the reference mesh cell are marked with a symbol on the operator. □

Remark B.65. Integrals with a gradients. Using the rule for transforming integrals and (B.19) gives

$$\begin{aligned}&\int_K \boldsymbol{b}(\boldsymbol{x}) \cdot \nabla v(\boldsymbol{x})\ d\boldsymbol{x} \\ &= \int_{\hat{K}} \boldsymbol{b}\left(F_K(\boldsymbol{\xi})\right) \cdot \left[\left(DF_K^{-1}\right)^T\left(F_K(\boldsymbol{\xi})\right)\right] \nabla_{\boldsymbol{\xi}}\hat{v}(\boldsymbol{\xi})\,|\det DF_K(\boldsymbol{\xi})|\ d\boldsymbol{\xi}. \qquad \text{(B.21)}\end{aligned}$$

Similarly, one obtains

$$\begin{aligned}&\int_K \nabla v(\boldsymbol{x}) \cdot \nabla w(\boldsymbol{x})\ d\boldsymbol{x} \\ &= \int_{\hat{K}} \left[\left(DF_K^{-1}\right)^T\left(F_K(\boldsymbol{\xi})\right)\right] \nabla_{\boldsymbol{\xi}}\hat{v}(\boldsymbol{\xi}) \cdot \left[\left(DF_K^{-1}\right)^T\left(F_K(\boldsymbol{\xi})\right)\right] \nabla_{\boldsymbol{\xi}}\hat{w}(\boldsymbol{\xi}) \\ &\quad \times |\det DF_K(\boldsymbol{\xi})|\ d\boldsymbol{\xi}. \qquad \text{(B.22)}\end{aligned}$$

□

Remark B.66. Integral with the divergence. Integrals of the following type are important for the Navier–Stokes equations

$$\int_K \nabla \cdot v(\boldsymbol{x}) q(\boldsymbol{x})\ d\boldsymbol{x} = \int_K \sum_{i=1}^{d} \frac{\partial v_i}{\partial x_i}(\boldsymbol{x}) q(\boldsymbol{x})\ d\boldsymbol{x}$$
$$= \int_{\hat{K}} \sum_{i=1}^{d} \left[\left(\left(DF_K^{-1}(F_K(\boldsymbol{\xi})) \right)^T \right)_i \cdot \nabla_{\boldsymbol{\xi}} \hat{v}_i(\boldsymbol{\xi}) \right] \hat{q}(\boldsymbol{\xi}) \left| \det DF_K(\boldsymbol{\xi}) \right|\ d\boldsymbol{\xi}$$
$$= \int_{\hat{K}} \left[\left(DF_K^{-1}(F_K(\boldsymbol{\xi})) \right)^T : D_{\boldsymbol{\xi}} \boldsymbol{v}(\boldsymbol{\xi}) \right] \hat{q}(\boldsymbol{\xi}) \left| \det DF_K(\boldsymbol{\xi}) \right|\ d\boldsymbol{\xi}. \tag{B.23}$$

In the derivation, (B.19) was used. □

Example B.67. Affine transform. The most important class of reference maps are affine transforms

$$\boldsymbol{x} = B\boldsymbol{\xi} + \boldsymbol{b}, \quad B \in \mathbb{R}^{d \times d}, \boldsymbol{b} \in \mathbb{R}^d,$$

where the invertible matrix B and the vector $\boldsymbol{b}$ are constants. It follows that

$$\boldsymbol{\xi} = B^{-1}(\boldsymbol{x} - \boldsymbol{b}) = B^{-1}\boldsymbol{x} - B^{-1}\boldsymbol{b}.$$

In this case, there are

$$DF_K = B, \quad DF_K^{-1} = B^{-1}, \quad \det DF_K = \det(B).$$

One obtains for the integral transforms from (B.18), (B.21), (B.22), and (B.23)

$$\int_K v(\boldsymbol{x})\ d\boldsymbol{x} = |\det(B)| \int_{\hat{K}} \hat{v}(\boldsymbol{\xi})\ d\boldsymbol{\xi}, \tag{B.24}$$
$$\int_K \boldsymbol{b}(\boldsymbol{x}) \cdot \nabla v(\boldsymbol{x})\ d\boldsymbol{x} = |\det(B)| \int_{\hat{K}} \boldsymbol{b}(F_K(\boldsymbol{\xi})) \cdot B^{-T} \nabla_{\boldsymbol{\xi}} \hat{v}(\boldsymbol{\xi})\ d\boldsymbol{\xi}, \tag{B.25}$$
$$\int_K \nabla v(\boldsymbol{x}) \cdot \nabla w(\boldsymbol{x})\ d\boldsymbol{x} = |\det(B)| \int_{\hat{K}} B^{-T} \nabla_{\boldsymbol{\xi}} \hat{v}(\boldsymbol{\xi}) \cdot B^{-T} \nabla_{\boldsymbol{\xi}} \hat{w}(\boldsymbol{\xi})\ d\boldsymbol{\xi}, \tag{B.26}$$
$$\int_K \nabla \cdot v(\boldsymbol{x}) q(\boldsymbol{x})\ d\boldsymbol{x} = |\det(B)| \int_{\hat{K}} \left[B^{-T} : D_{\boldsymbol{\xi}} \boldsymbol{v}(\boldsymbol{\xi}) \right] \hat{q}(\boldsymbol{\xi})\ d\boldsymbol{\xi}. \tag{B.27}$$

□

Appendix C
Interpolation

Remark C.1. Motivation. Variational forms of partial differential equations use functions in Sobolev spaces. The solution of these equations shall be approximated with the Ritz method in finite dimensional spaces, the finite element spaces. The best possible approximation of an arbitrary function from the Sobolev space by a finite element function is a factor in the upper bound for the finite element error, e.g., see the Lemma of Cea, estimate (B.12).

This section studies the approximation quality of finite element spaces. Estimates are proved for interpolants of functions. Interpolation estimates are of course upper bounds for the best approximation error and they can serve as factors in finite element error estimates. □

C.1 Interpolation in Sobolev Spaces by Polynomials

Lemma C.2. Unique determination of a polynomial with integral conditions. *Let Ω be a bounded domain in $\mathbb{R}^d$ with Lipschitz boundary. Let $m \in \mathbb{N} \cup \{0\}$ be given and let for all derivatives with multi-index $\boldsymbol{\alpha}$, $|\boldsymbol{\alpha}| \leq m$, a value $a_{\boldsymbol{\alpha}} \in \mathbb{R}$ be given. Then, there is a uniquely determined polynomial $p \in P_m(\Omega)$ such that*

$$\int_\Omega \partial_{\boldsymbol{\alpha}} p(\boldsymbol{x}) \, d\boldsymbol{x} = a_{\boldsymbol{\alpha}}, \quad |\boldsymbol{\alpha}| \leq m. \tag{C.1}$$

Proof. Let $p \in P_m(\Omega)$ be an arbitrary polynomial. It has the form

$$p(\boldsymbol{x}) = \sum_{|\boldsymbol{\beta}| \leq m} b_{\boldsymbol{\beta}} \boldsymbol{x}^{\boldsymbol{\beta}}.$$

Inserting this representation into (C.1) leads to a linear system of equations $M\mathbf{b} = \mathbf{a}$ with

$$M = (M_{\boldsymbol{\alpha}\boldsymbol{\beta}}),\ M_{\boldsymbol{\alpha}\boldsymbol{\beta}} = \int_\Omega \partial_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\beta}}\ d\boldsymbol{x},\ \mathbf{b} = (b_{\boldsymbol{\beta}}),\ \mathbf{a} = (a_{\boldsymbol{\alpha}}),$$

for $|\boldsymbol{\alpha}|, |\boldsymbol{\beta}| \leq m$. Since M is a squared matrix, the linear system of equations possesses a unique solution if and only if M is non-singular.

The proof is performed by contradiction. Assume that M is singular. Then there exists a non-trivial solution of the homogeneous system. That means, there is a polynomial $q \in P_m(\Omega) \setminus \{0\}$ with

$$\int_\Omega \partial_{\boldsymbol{\alpha}} q(\boldsymbol{x})\ d\boldsymbol{x} = 0 \text{ for all } |\boldsymbol{\alpha}| \leq m.$$

The polynomial $q(\boldsymbol{x})$ has the representation $q(\boldsymbol{x}) = \sum_{|\boldsymbol{\beta}| \leq m} c_{\boldsymbol{\beta}} \boldsymbol{x}^{\boldsymbol{\beta}}$. Now, one can choose a $c_{\boldsymbol{\beta}} \neq 0$ with maximal value $|\boldsymbol{\beta}|$. Then, it is $\partial_{\boldsymbol{\beta}} q(\boldsymbol{x}) = C c_{\boldsymbol{\beta}} = const \neq 0$, where $C > 0$ comes from the differentiation rule for polynomials, which is a contradiction to the vanishing of the integral for $\partial_{\boldsymbol{\beta}} q(\boldsymbol{x})$. ■

Remark C.3. To Lemma C.2. Lemma C.2 states that a polynomial is uniquely determined if a condition on the integral on Ω is prescribed for each derivative. □

Lemma C.4. Poincaré-type inequality. *Denote by $D^k v(\boldsymbol{x})$, $k \in \mathbb{N} \cup \{0\}$, the total derivative of order k of a function $v(\boldsymbol{x})$, e.g., for $k = 1$ the gradient of $v(\boldsymbol{x})$. Let Ω be convex and be included into a ball of radius R. Let $k, l \in \mathbb{N} \cup \{0\}$ with $k \leq l$ and let $p \in \mathbb{R}$ with $p \in [1, \infty)$. Assume that $v \in W^{l,p}(\Omega)$ satisfies*

$$\int_\Omega \partial_{\boldsymbol{\alpha}} v(\boldsymbol{x})\ d\boldsymbol{x} = 0 \text{ for all } |\boldsymbol{\alpha}| \leq l - 1,$$

then it holds the estimate

$$\left\| D^k v \right\|_{L^p(\Omega)} \leq C R^{l-k} \left\| D^l v \right\|_{L^p(\Omega)},$$

where the constant C does not depend on Ω and on $v(\boldsymbol{x})$.

Proof. There is nothing to prove if $k = l$. In addition, it suffices to prove the lemma for $k = 0$ and $l = 1$, since the general case follows by applying the result to $\partial_{\boldsymbol{\alpha}} v(\boldsymbol{x})$.

Since Ω is assumed to be convex, the integral mean value theorem can be written in the form

$$v(\boldsymbol{x}) - v(\boldsymbol{y}) = \int_0^1 \nabla v(t\boldsymbol{x} + (1-t)\boldsymbol{y}) \cdot (\boldsymbol{x} - \boldsymbol{y})\ dt, \quad \boldsymbol{x}, \boldsymbol{y} \in \Omega.$$

Integration with respect to $\boldsymbol{y}$ yields

$$v(\boldsymbol{x}) \int_\Omega d\boldsymbol{y} - \int_\Omega v(\boldsymbol{y})\ d\boldsymbol{y} = \int_\Omega \int_0^1 \nabla v(t\boldsymbol{x} + (1-t)\boldsymbol{y}) \cdot (\boldsymbol{x} - \boldsymbol{y})\ dt\ d\boldsymbol{y}.$$

It follows from the assumption that the second integral on the left hand side vanishes. Hence, one gets

$$v(\boldsymbol{x}) = \frac{1}{|\Omega|} \int_\Omega \int_0^1 \nabla v(t\boldsymbol{x} + (1-t)\boldsymbol{y}) \cdot (\boldsymbol{x} - \boldsymbol{y})\ dt\ d\boldsymbol{y}.$$

Now, taking the absolute value on both sides, using that the absolute value of an integral is estimated from above by the integral of the absolute value, applying the Cauchy–Schwarz inequality for vectors and the estimate $\|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq 2R$ yields

$$
\begin{aligned}
|v(\boldsymbol{x})| &= \frac{1}{|\Omega|}\left|\int_\Omega \int_0^1 \nabla v(t\boldsymbol{x}+(1-t)\boldsymbol{y})\cdot(\boldsymbol{x}-\boldsymbol{y})\ dt\ d\boldsymbol{y}\right| \\
&\leq \frac{1}{|\Omega|}\int_\Omega \int_0^1 |\nabla v(t\boldsymbol{x}+(1-t)\boldsymbol{y})\cdot(\boldsymbol{x}-\boldsymbol{y})|\ dt\ d\boldsymbol{y} \\
&\leq \frac{2R}{|\Omega|}\int_\Omega \int_0^1 \|\nabla v(t\boldsymbol{x}+(1-t)\boldsymbol{y})\|_2\ dt\ d\boldsymbol{y}. \qquad \text{(C.2)}
\end{aligned}
$$

Then (C.2) is raised to the power p and then integrated with respect to $\boldsymbol{x}$. One obtains with Hölder's inequality (A.7), with $p^{-1}+q^{-1}=1 \implies p/q-p=p(1/q-1)=-1$, that

$$
\begin{aligned}
\int_\Omega |v(\boldsymbol{x})|^p\ d\boldsymbol{x} &\leq \frac{CR^p}{|\Omega|^p}\int_\Omega \left(\int_\Omega \int_0^1 \|\nabla v(t\boldsymbol{x}+(1-t)\boldsymbol{y})\|_2\ dt\ d\boldsymbol{y}\right)^p d\boldsymbol{x} \\
&\leq \frac{CR^p}{|\Omega|^p}\int_\Omega \Bigg[\underbrace{\left(\int_\Omega \int_0^1 1^q\ dt\ d\boldsymbol{y}\right)^{p/q}}_{|\Omega|^{p/q}} \\
&\qquad \times \left(\int_\Omega \int_0^1 \|\nabla v(t\boldsymbol{x}+(1-t)\boldsymbol{y})\|_2^p\ dt\ d\boldsymbol{y}\right)\Bigg] d\boldsymbol{x} \\
&= \frac{CR^p}{|\Omega|}\int_\Omega \left(\int_\Omega \int_0^1 \|\nabla v(t\boldsymbol{x}+(1-t)\boldsymbol{y})\|_2^p\ dt\ d\boldsymbol{y}\right) d\boldsymbol{x}.
\end{aligned}
$$

Applying the theorem of Fubini allows the commutation of the integration

$$
\int_\Omega |v(\boldsymbol{x})|^p\ d\boldsymbol{x} \leq \frac{CR^p}{|\Omega|}\int_0^1 \int_\Omega \left(\int_\Omega \|\nabla v(t\boldsymbol{x}+(1-t)\boldsymbol{y})\|_2^p\ d\boldsymbol{y}\right) d\boldsymbol{x}\ dt.
$$

Using the integral mean value theorem in one dimension gives that there is a $t_0 \in [0,1]$, such that

$$
\int_\Omega |v(\boldsymbol{x})|^p\ d\boldsymbol{x} \leq \frac{CR^p}{|\Omega|}\int_\Omega \left(\int_\Omega \|\nabla v(t_0\boldsymbol{x}+(1-t_0)\boldsymbol{y})\|_2^p\ d\boldsymbol{y}\right) d\boldsymbol{x}.
$$

The function $\|\nabla v(\boldsymbol{x})\|_2^p$ will be extended to $\mathbb{R}^d$ by zero and the extension will be also denoted by $\|\nabla v(\boldsymbol{x})\|_2^p$. Then, it is

$$
\int_\Omega |v(\boldsymbol{x})|^p\ d\boldsymbol{x} \leq \frac{CR^p}{|\Omega|}\int_\Omega \left(\int_{\mathbb{R}^d} \|\nabla v(t_0\boldsymbol{x}+(1-t_0)\boldsymbol{y})\|_2^p\ d\boldsymbol{y}\right) d\boldsymbol{x}. \qquad \text{(C.3)}
$$

Let $t_0 \in [0,1/2]$. Since the domain of integration is $\mathbb{R}^d$, a substitution of variables $t_0\boldsymbol{x}+(1-t_0)\boldsymbol{y}=\boldsymbol{z}$ can be applied and leads to

$$
\int_{\mathbb{R}^d} \|\nabla v(t_0\boldsymbol{x}+(1-t_0)\boldsymbol{y})\|_2^p\ d\boldsymbol{y} = \frac{1}{1-t_0}\int_{\mathbb{R}^d} \|\nabla v(\boldsymbol{z})\|_2^p\ d\boldsymbol{z} \leq 2\,\|\nabla v\|_{L^p(\Omega)}^p,
$$

since $1/(1-t_0)\leq 2$. Inserting this expression into (C.3) gives

$$
\int_\Omega |v(\boldsymbol{x})|^p\ d\boldsymbol{x} \leq 2CR^p\,\|\nabla v\|_{L^p(\Omega)}^p.
$$

If $t_0 > 1/2$ then one changes the roles of $\boldsymbol{x}$ and $\boldsymbol{y}$, applies the theorem of Fubini to change the sequence of integration, and uses the same arguments. ■

Remark C.5. On Lemma C.4. The Lemma C.4 proves an inequality of Poincaré-type. It says that it is possible to estimate the $L^p(\Omega)$ norm of a lower derivative of a function $v(\boldsymbol{x})$ by the same norm of a higher derivative if the integral mean values of some lower derivatives vanish.

An important application of Lemma C.4 is in the proof of the Bramble–Hilbert lemma. The Bramble–Hilbert lemma considers a continuous linear functional which is defined on a Sobolev space and which vanishes for all polynomials of degree less or equal than m. It states that the value of the functional can be estimated by the Lebesgue norm of the $(m+1)$th total derivative of the functions from this Sobolev space. □

Theorem C.6. ***Bramble–Hilbert lemma.*** *Let $m \in \mathbb{N} \cup \{0\}$, $m \geq 0$, $p \in [1, \infty]$, and $F : W^{m+1,p}(\Omega) \to \mathbb{R}$ be a continuous linear functional, and let the conditions of Lemma C.2 and C.4 be satisfied. Let*

$$F(p) = 0 \quad \forall \ p \in P_m(\Omega),$$

then there is a constant $C(\Omega)$, which is independent of $v(\boldsymbol{x})$ and F, such that

$$|F(v)| \leq C(\Omega) \left\| D^{m+1} v \right\|_{L^p(\Omega)} \quad \forall \ v \in W^{m+1,p}(\Omega).$$

Proof. Let $v \in W^{m+1,p}(\Omega)$. It follows from Lemma C.2 that there is a polynomial from $P_m(\Omega)$ with

$$\int_\Omega \partial_{\boldsymbol{\alpha}}(v+p)(\boldsymbol{x}) \, d\boldsymbol{x} = 0 \text{ for } |\boldsymbol{\alpha}| \leq m.$$

Lemma C.4 gives, with $l = m+1$ and considering each term in $\|\cdot\|_{W^{m+1,p}(\Omega)}$ individually, the estimate

$$\|v+p\|_{W^{m+1,p}(\Omega)} \leq C(\Omega) \left\| D^{m+1}(v+p) \right\|_{L^p(\Omega)} = C(\Omega) \left\| D^{m+1} v \right\|_{L^p(\Omega)}.$$

From the vanishing of F for $p \in P_m(\Omega)$ and the continuity of F it follows that

$$|F(v)| = |F(v+p)| \leq c \|v+p\|_{W^{m+1,p}(\Omega)} \leq C(\Omega) \left\| D^{m+1} v \right\|_{L^p(\Omega)}.$$

■

Remark C.7. Strategy for estimating the interpolation error. The Bramble–Hilbert lemma will be used for estimating the interpolation error for an affine family of finite elements. The strategy is as follows:

- Show first the estimate on the reference mesh cell $\hat{K}$.
- Transform the estimate on an arbitrary mesh cell K to the reference mesh cell $\hat{K}$.
- Apply the estimate on $\hat{K}$.
- Transform back to K.

One has to study what happens if the transforms are applied to the estimate. □

Remark C.8. Assumptions, definition of the interpolant. Let $\hat{K} \subset \mathbb{R}^d, d \in \{2, 3\}$, be a reference mesh cell (compact polyhedron), $\hat{P}(\hat{K})$ a polynomial

space of dimension N, and $\hat{\Phi}_1, \ldots, \hat{\Phi}_N \ : \ C^s(\hat{K}) \to \mathbb{R}$ continuous linear functionals. It will be assumed that the space $\hat{P}(\hat{K})$ is unisolvent with respect to these functionals. Then, there is a local basis $\hat{\phi}_1, \ldots, \hat{\phi}_N \in \hat{P}(\hat{K})$.

Consider $\hat{v} \in C^s(\hat{K})$, then the interpolant $I_{\hat{K}}\hat{v} \in \hat{P}(\hat{K})$ is defined by

$$I_{\hat{K}}\hat{v}(\hat{\boldsymbol{x}}) = \sum_{i=1}^{N} \hat{\Phi}_i(\hat{v})\hat{\phi}_i(\hat{\boldsymbol{x}}).$$

The operator $I_{\hat{K}}$ is a continuous and linear operator from $C^s(\hat{K})$ to $\hat{P}(\hat{K})$. From the linearity it follows that $I_{\hat{K}}$ is the identity on $\hat{P}(\hat{K})$

$$I_{\hat{K}}\hat{p} = \hat{p} \quad \forall\, \hat{p} \in \hat{P}(\hat{K}).$$

□

Example C.9. Interpolation operators.

- Let $\hat{K} \subset \mathbb{R}^d$ be an arbitrary reference cell, $\hat{P}(\hat{K}) = P_0(\hat{K})$, and

$$\hat{\Phi}(\hat{v}) = \frac{1}{\left|\hat{K}\right|} \int_{\hat{K}} \hat{v}(\hat{\boldsymbol{x}}) \; d\hat{\boldsymbol{x}}.$$

  The functional $\hat{\Phi}$ is continuous on $C^0(\hat{K})$ since

$$\left|\hat{\Phi}(\hat{v})\right| \leq \frac{1}{\left|\hat{K}\right|} \int_{\hat{K}} |\hat{v}(\hat{\boldsymbol{x}})| \; d\hat{\boldsymbol{x}} \leq \frac{\left|\hat{K}\right|}{\left|\hat{K}\right|} \max_{\hat{\boldsymbol{x}} \in \hat{K}} |\hat{v}(\hat{\boldsymbol{x}})| = \|\hat{v}\|_{C^0(\hat{K})}.$$

  For the constant function $1 \in P_0(\hat{K})$ it is $\hat{\Phi}(1) = 1 \neq 0$. Hence, $\{\hat{\phi}\} = \{1\}$ is the local basis and the space is unisolvent with respect to $\hat{\Phi}$. The operator

$$I_{\hat{K}}\hat{v}(\hat{\boldsymbol{x}}) = \hat{\Phi}(\hat{v})\hat{\phi}(\hat{\boldsymbol{x}}) = \frac{1}{\left|\hat{K}\right|} \int_{\hat{K}} \hat{v}(\hat{\boldsymbol{x}}) \; d\hat{\boldsymbol{x}}$$

  is an integral mean value operator, i.e., each continuous function on $\hat{K}$ will be approximated by a constant function whose value equals the integral mean value, see Figure C.1

- It is possible to define $\hat{\Phi}(\hat{v}) = \hat{v}(\hat{\boldsymbol{x}}_0)$ for an arbitrary point $\hat{\boldsymbol{x}}_0 \in \hat{K}$. This functional is also linear and continuous in $C^0(\hat{K})$. The interpolation operator $I_{\hat{K}}$ defined in this way interpolates each continuous function by a constant function whose value is equal to the value of the function at $\hat{\boldsymbol{x}}_0$, see also Figure C.1.

  Interpolation operators which are defined by using values of functions, are called Lagrangian interpolation operators.

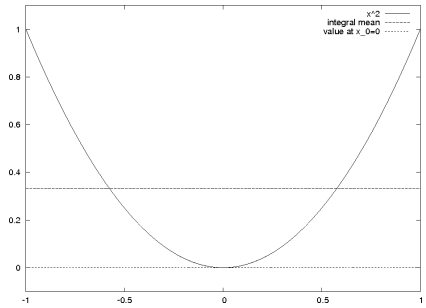

Fig. C.1 Interpolation of x^2 in $[-1,1]$ by a P_0 function with the integral mean value and with the value of the function at $x_0 = 0$.

This example demonstrates that the interpolation operator $I_{\hat{K}}$ depends on $\hat{P}(\hat{K})$ and on the functionals $\hat{\Phi}_i$. □

Theorem C.10. Interpolation error estimate on a reference mesh cell. *Let $P_m(\hat{K}) \subset \hat{P}(\hat{K})$ and $p \in [1,\infty)$ with $(m+1-s)p > d$. Then there is a constant C that is independent of $\hat{v}(\hat{\boldsymbol{x}})$ such that*

$$\left\|\hat{v} - I_{\hat{K}}\hat{v}\right\|_{W^{m+1,p}(\hat{K})} \leq C \left\|D^{m+1}\hat{v}\right\|_{L^p(\hat{K})} \quad \forall\ \hat{v} \in W^{m+1,p}(\hat{K}). \tag{C.4}$$

Proof. Since $\hat{K}$ is a bounded and convex domain, one has the Sobolev imbedding (A.15)

$$W^{m+1,p}(\hat{K}) = W^{(m+1-s)+s,p}(\hat{K}) \to C^s(\hat{K})$$

if $(m+1-s)p > d$. That means, the interpolation operator is well defined in $W^{m+1,p}(\hat{K})$. From the identity of the interpolation operator in $P_m(\hat{K})$, the triangle inequality, the boundedness of the interpolation operator (it is a linear and continuous operator mapping $C^s(\hat{K}) \to \hat{P}(\hat{K}) \subset W^{m+1,p}(\hat{K})$), and the Sobolev imbedding, one obtains for $\hat{q} \in P_m(\hat{K})$

$$\begin{aligned}
\left\|\hat{v} - I_{\hat{K}}\hat{v}\right\|_{W^{m+1,p}(\hat{K})} &= \left\|\hat{v} + \hat{q} - I_{\hat{K}}(\hat{v}+\hat{q})\right\|_{W^{m+1,p}(\hat{K})} \\
&\leq \|\hat{v}+\hat{q}\|_{W^{m+1,p}(\hat{K})} + \left\|I_{\hat{K}}(\hat{v}+\hat{q})\right\|_{W^{m+1,p}(\hat{K})} \\
&\leq \|\hat{v}+\hat{q}\|_{W^{m+1,p}(\hat{K})} + c\,\|\hat{v}+\hat{q}\|_{C^s(\hat{K})} \\
&\leq c\,\|\hat{v}+\hat{q}\|_{W^{m+1,p}(\hat{K})}\,.
\end{aligned}$$

Choosing $\hat{q}(\hat{\boldsymbol{x}})$ in Lemma C.2 such that

$$\int_{\hat{K}} \partial_{\boldsymbol{\alpha}}(\hat{v}+\hat{q})\ d\hat{\boldsymbol{x}} = 0 \quad \forall\ |\boldsymbol{\alpha}| \leq m,$$

the assumptions of Lemma C.4 are satisfied. It follows that

$$\|\hat{v}+\hat{q}\|_{W^{m+1,p}(\hat{K})} \leq c\left\|D^{m+1}(\hat{v}+\hat{q})\right\|_{L^p(\hat{K})} = c\left\|D^{m+1}\hat{v}\right\|_{L^p(\hat{K})}.$$

■

Remark C.11. On Theorem C.10.

- One can construct examples which show that the Sobolev imbedding is not valid if $(m+1-s)p > d$ is not satisfied. In the case $(m+1-s)p \leq d$, the statement of Theorem C.10 is not true.
  Consider the interpolation of continuous functions ($s = 0$) with piecewise linear elements ($m = 1$) in Sobolev spaces that are also Hilbert spaces ($p = 2$). Then $(m + 1 - s)p = 4$ and it follows that the theorem can be applied only for $d \in \{2, 3\}$.
  For piecewise constant finite elements, the statement of the theorem is true only for $d = 1$.
- The theorem requires only that $P_m(\hat{K}) \subset \hat{P}(\hat{K})$. This requirement does not exclude that $\hat{P}(\hat{K})$ contains polynomials of higher degree, too. However, this property is not utilized and also not needed if the other assumptions of the theorem are satisfied.

□

Definition C.12. Quasi-uniform family of triangulations, regular family of triangulations, (Brenner and Scott, 2008, Def. 4.4.13). Let $\{\mathcal{T}^h\}$ with $0 < h \leq 1$, be a family of triangulations such that

$$\max_{K \in \mathcal{T}^h} h_K \leq h \operatorname{diam}(\Omega),$$

where h_K be the diameter of $K = F_K(\hat{K})$, i.e., the largest distance of two points that are contained in K. The family is called to be quasi-uniform if there exists a $c > 0$ such that

$$\min_{K \in \mathcal{T}^h} \rho_K \geq ch\operatorname{diam}(\Omega) \tag{C.5}$$

for all $h \in (0, 1]$, where ρ_K is the diameter of the largest ball contained in K.

The family is called to be regular if there is exists a $c > 0$ such that for all $K \in \mathcal{T}^h$ and for all $h \in (0, 1]$

$$\rho_K \geq ch_K.$$

□

Remark C.13. Assumptions on the reference mapping and the triangulation. For deriving the interpolation error estimate for arbitrary mesh cells K, and finally for the finite element space, one has to study the properties of the affine mapping from K to $\hat{K}$ and of the back mapping. Here, the case of an affine family of finite elements whose mesh cells are generated by affine mappings

$$F_K\hat{\boldsymbol{x}} = B\hat{\boldsymbol{x}} + \boldsymbol{b},$$

will be considered, where B is a non-singular $d \times d$ matrix and $\boldsymbol{b}$ is a d vector.

For the global estimate, a quasi-uniform family of triangulations will be considered. □

Lemma C.14. Estimates of matrix norms. *For each matrix norm $\|\cdot\|$ one has the estimates*

$$\|B\| \le ch_K, \quad \left\|B^{-1}\right\| \le ch_K^{-1}, \tag{C.6}$$

where the constants depend on the matrix norm and on C_R.

Proof. Since $\hat{K}$ is a Lipschitz domain with polyhedral boundary, it contains a ball $B(\hat{\boldsymbol{x}}_0, r)$ with $\hat{\boldsymbol{x}}_0 \in \hat{K}$ and some $r > 0$. Hence, $\hat{\boldsymbol{x}}_0 + \hat{\boldsymbol{y}} \in \hat{K}$ for all $\|\hat{\boldsymbol{y}}\|_2 = r$. It follows that the images

$$\boldsymbol{x}_0 = B\hat{\boldsymbol{x}}_0 + \boldsymbol{b}, \quad \boldsymbol{x} = B(\hat{\boldsymbol{x}}_0 + \hat{\boldsymbol{y}}) + \boldsymbol{b} = \boldsymbol{x}_0 + B\hat{\boldsymbol{y}}$$

are contained in K. Since the triangulation is assumed to be quasi-uniform, one obtains for all $\hat{\boldsymbol{y}}$

$$\|B\hat{\boldsymbol{y}}\|_2 = \|\boldsymbol{x} - \boldsymbol{x}_0\|_2 \le C_R h_K.$$

Now, it holds for the spectral norm that

$$\|B\|_2 = \sup_{\hat{\boldsymbol{z}} \neq \boldsymbol{0}} \frac{\|B\hat{\boldsymbol{z}}\|_2}{\|\hat{\boldsymbol{z}}\|_2} = \frac{1}{r} \sup_{\|\hat{\boldsymbol{z}}\|_2 = r} \|B\hat{\boldsymbol{z}}\|_2 \le \frac{C_R}{r} h_K.$$

An estimate of this form, with a possible different constant, holds also for all other matrix norms since all matrix norms are equivalent.

The estimate for $\left\|B^{-1}\right\|$ proceeds in the same way with interchanging the roles of K and $\hat{K}$. ∎

Theorem C.15. Local interpolation estimate. *Let an affine family of finite elements be given by its reference cell $\hat{K}$, the functionals $\{\hat{\Phi}_i\}$, and a space of polynomials $\hat{P}(\hat{K})$. Let all assumptions of Theorem C.10 be satisfied. Then, for all $v \in W^{m+1,p}(K)$, $p \in [1, \infty)$, there is a constant C, which is independent of $v(\boldsymbol{x})$ such that*

$$\left\|D^k(v - I_K v)\right\|_{L^p(K)} \le Ch_K^{m+1-k} \left\|D^{m+1} v\right\|_{L^p(K)}, \quad k \le m+1. \tag{C.7}$$

Proof. The idea of the proof consists in transforming left hand side of (C.7) to the reference cell, using the interpolation estimate on the reference cell and transforming back.

i). Denote the elements of the matrices B and B^{-1} by b_{ij} and $b_{ij}^{(-1)}$, respectively. Since $\|B\|_\infty = \max_{i,j} |b_{ij}|$ is also a matrix norm, it holds that

$$|b_{ij}| \le Ch_K, \quad \left|b_{ij}^{(-1)}\right| \le Ch_K^{-1}. \tag{C.8}$$

Using element-wise estimates for the matrix B (Leibniz formula for determinants), one obtains

$$|\det B| \le Ch_K^d, \quad \left|\det B^{-1}\right| \le Ch_K^{-d}. \tag{C.9}$$

ii). The next step consists in proving that the transformed interpolation operator is equal to the natural interpolation operator on K. The latter one is given by

$$I_K v = \sum_{i=1}^{N} \Phi_{K,i}(v) \phi_{K,i}, \tag{C.10}$$

where $\{\phi_{K,i}\}$ is the basis of the space

$$P(K) = \{p \ : \ K \to \mathbb{R} \ : \ p = \hat{p} \circ F_K^{-1}, \hat{p} \in \hat{P}(\hat{K})\},$$

which satisfies $\Phi_{K,i}(\phi_{K,j}) = \delta_{ij}$. The functionals are defined by

$$\Phi_{K,i}(v) = \hat{\Phi}_i(v \circ F_K)$$

Hence, it follows with $v = \hat{\phi}_j \circ F_K^{-1}$ from the condition on the local basis on $\hat{K}$ that

$$\Phi_{K,i}(\hat{\phi}_j \circ F_K^{-1}) = \hat{\Phi}_i(\hat{\phi}_j) = \delta_{ij},$$

i.e., the local basis on K is given by $\phi_{K,j} = \hat{\phi}_j \circ F_K^{-1}$. Using (C.10), one gets

$$\begin{aligned} I_{\hat{K}}\hat{v} &= \sum_{i=1}^{N} \hat{\Phi}_i(\hat{v})\hat{\phi}_i = \sum_{i=1}^{N} \Phi_{K,i}(\underbrace{\hat{v} \circ F_K^{-1}}_{=v})\, \phi_{K,i} \circ F_K = \left(\sum_{i=1}^{N} \Phi_{K,i}(v)\phi_{K,i} \right) \circ F_K \\ &= I_K v \circ F_K. \end{aligned}$$

Hence, $I_{\hat{K}}\hat{v}$ is transformed correctly.

iii). One obtains with the chain rule

$$\frac{\partial v(\boldsymbol{x})}{\partial \boldsymbol{x}_i} = \sum_{j=1}^{d} \frac{\partial \hat{v}(\hat{\boldsymbol{x}})}{\partial \hat{\boldsymbol{x}}_j} b_{ji}^{(-1)}, \quad \frac{\partial \hat{v}(\hat{\boldsymbol{x}})}{\partial \hat{\boldsymbol{x}}_i} = \sum_{j=1}^{d} \frac{\partial v(\boldsymbol{x})}{\partial \boldsymbol{x}_j} b_{ji}.$$

It follows with (C.8) that (with each derivative one obtains an additional factor of B or B^{-1}, respectively)

$$\left\| D_{\boldsymbol{x}}^k v(\boldsymbol{x}) \right\|_2 \leq C h_K^{-k} \left\| D_{\hat{\boldsymbol{x}}}^k \hat{v}(\hat{\boldsymbol{x}}) \right\|_2, \quad \left\| D_{\hat{\boldsymbol{x}}}^k \hat{v}(\hat{\boldsymbol{x}}) \right\|_2 \leq C h_K^{k} \left\| D_{\boldsymbol{x}}^k v(\boldsymbol{x}) \right\|_2.$$

One gets with (C.9)

$$\int_K \left\| D_{\boldsymbol{x}}^k v(\boldsymbol{x}) \right\|_2^p \, d\boldsymbol{x} \leq C h_K^{-kp} \left| \det B \right| \int_{\hat{K}} \left\| D_{\hat{\boldsymbol{x}}}^k \hat{v}(\hat{\boldsymbol{x}}) \right\|_2^p \, d\hat{\boldsymbol{x}} \leq C h_K^{-kp+d} \int_{\hat{K}} \left\| D_{\hat{\boldsymbol{x}}}^k \hat{v}(\hat{\boldsymbol{x}}) \right\|_2^p \, d\hat{\boldsymbol{x}}$$

and

$$\int_{\hat{K}} \left\| D_{\hat{\boldsymbol{x}}}^k \hat{v}(\hat{\boldsymbol{x}}) \right\|_2^p \, d\hat{\boldsymbol{x}} \leq C h_K^{kp} \left| \det B^{-1} \right| \int_K \left\| D_{\boldsymbol{x}}^k v(\boldsymbol{x}) \right\|_2^p \, d\boldsymbol{x} \leq C h_K^{kp-d} \int_K \left\| D_{\boldsymbol{x}}^k v(\boldsymbol{x}) \right\|_2^p \, d\boldsymbol{x}.$$

Using now the interpolation estimate on the reference cell (C.4) yields

$$\left\| D_{\hat{\boldsymbol{x}}}^k (\hat{v} - I_{\hat{K}}\hat{v}) \right\|_{L^p(\hat{K})}^p \leq C \left\| D_{\hat{\boldsymbol{x}}}^{m+1} \hat{v} \right\|_{L^p(\hat{K})}^p, \quad 0 \leq k \leq m+1.$$

It follows that

$$\begin{aligned} \left\| D_{\boldsymbol{x}}^k (v - I_K v) \right\|_{L^p(K)}^p &\leq C h_K^{-kp+d} \left\| D_{\hat{\boldsymbol{x}}}^k (\hat{v} - I_{\hat{K}}\hat{v}) \right\|_{L^p(\hat{K})}^p \\ &\leq C h_K^{-kp+d} \left\| D_{\hat{\boldsymbol{x}}}^{m+1} \hat{v} \right\|_{L^p(\hat{K})}^p \\ &\leq C h_K^{(m+1-k)p} \left\| D_{\boldsymbol{x}}^{m+1} v \right\|_{L^p(K)}^p. \end{aligned}$$

Taking the p-th root proves the statement of the theorem. ■

Remark C.16. On estimate (C.7).

- Note that the power of h_K does not depend on p and d.
- Consider a quasi-uniform triangulation and define

$$h = \max_{K \in \mathcal{T}^h} \{h_K\}.$$

Then, one obtains by summing over all mesh cells an interpolation estimate for the global finite element space

$$\begin{aligned}\left\|D^k(v-I_hv)\right\|_{L^p(\Omega)} &= \left(\sum_{K\in\mathcal{T}^h}\left\|D^k(v-I_Kv)\right\|^p_{L^p(K)}\right)^{1/p}\\ &\le \left(\sum_{K\in\mathcal{T}^h} ch_K^{(m+1-k)p}\left\|D^{m+1}v\right\|^p_{L^p(K)}\right)^{1/p}\\ &\le ch^{(m+1-k)}\left\|D^{m+1}v\right\|_{L^p(\Omega)}. \qquad \text{(C.11)}\end{aligned}$$

For linear finite elements P_1 ($m=1$) it is, in particular,

$$\|v-I_hv\|_{L^p(\Omega)} \le ch^2\left\|D^2v\right\|_{L^p(\Omega)}, \quad \|\nabla(v-I_hv)\|_{L^p(\Omega)} \le ch\left\|D^2v\right\|_{L^p(\Omega)},$$

if $v\in W^{2,p}(\Omega)$.

□

Corollary C.17. Finite element error estimate. *Let $u(\boldsymbol{x})$ be the solution of the model problem (??) with $u\in H^{m+1}(\Omega)$ and let $u^h(\boldsymbol{x})$ be the solution of the corresponding finite element problem. Consider a family of quasi-uniform triangulations and let the finite element spaces V^h contain polynomials of degree m. Then, the following finite element error estimate holds*

$$\left\|\nabla(u-u^h)\right\|_{L^2(\Omega)} \le ch^m\left\|D^{m+1}u\right\|_{L^2(\Omega)} = ch^m\left|u\right|_{H^{m+1}(\Omega)}. \qquad \text{(C.12)}$$

Proof. The statement follows by combining Lemma ?? (for $V=H^1_0(\Omega)$) and (C.11)

$$\left\|\nabla(u-u^h)\right\|_{L^2(\Omega)} \le \inf_{v^h\in V^h}\left\|\nabla(u-v^h)\right\|_{L^2(\Omega)} \le \|\nabla(u-I_hu)\|_{L^2(\Omega)} \le ch^m\left|u\right|_{H^{m+1}(\Omega)}.$$

■

Remark C.18. To (C.12). Note that Lemma ?? provides only information about the error in the norm on the left-hand side of (C.12), but not in other norms. □

C.2 Inverse Estimate

Remark C.19. On inverse estimates. The approach for proving interpolation error estimates can be uses also to prove so-called inverse estimates. In contrast to interpolation error estimates, a norm of a higher order derivative of a finite element function will be estimated by a norm of a lower order derivative of this function. One obtains as penalty a factor with negative powers of the diameter of the mesh cell. □

Theorem C.20. Inverse estimate. *Let $0 \leq k \leq l$ be natural numbers and let $p, q \in [1, \infty]$. Then there is a constant C_{inv}, which depends only on $k, l, p, q, \hat{K}, \hat{P}(\hat{K})$ such that*

$$\left\|D^l v^h\right\|_{L^q(K)} \leq C_{\text{inv}} h_K^{(k-l)-d(p^{-1}-q^{-1})} \left\|D^k v^h\right\|_{L^p(K)} \quad \forall\, v^h \in P(K). \tag{C.13}$$

Proof. In the first step, (C.13) is shown for $h_{\hat{K}} = 1$ and $k = 0$ on the reference mesh cell. Since all norms are equivalent in finite dimensional spaces, one obtains

$$\left\|D^l \hat{v}^h\right\|_{L^q(\hat{K})} \leq \left\|\hat{v}^h\right\|_{W^{l,q}(\hat{K})} \leq C \left\|\hat{v}^h\right\|_{L^p(\hat{K})} \quad \forall\, \hat{v}^h \in \hat{P}(\hat{K}).$$

If $k > 0$, then one sets

$$\tilde{P}(\hat{K}) = \left\{\partial_{\boldsymbol{\alpha}} \hat{v}^h \;:\; \hat{v}^h \in \hat{P}(\hat{K}), |\boldsymbol{\alpha}| = k\right\},$$

which is also a space consisting of polynomials. The application of the first estimate of the proof to $\tilde{P}(\hat{K})$ gives

$$\begin{aligned}\left\|D^l \hat{v}^h\right\|_{L^q(\hat{K})} &= \sum_{|\boldsymbol{\alpha}|=k} \left\|D^{l-k}\left(\partial_{\boldsymbol{\alpha}} \hat{v}^h\right)\right\|_{L^q(\hat{K})} \leq C \sum_{|\boldsymbol{\alpha}|=k} \left\|\partial_{\boldsymbol{\alpha}} \hat{v}^h\right\|_{L^p(\hat{K})} \\ &= C \left\|D^k \hat{v}^h\right\|_{L^p(\hat{K})}.\end{aligned}$$

This estimate is transformed to an arbitrary mesh cell K analogously as for the interpolation error estimates. From the estimates for the transformations, one obtains

$$\begin{aligned}\left\|D^l v^h\right\|_{L^q(K)} &\leq C h_K^{-l+d/q} \left\|D^l \hat{v}^h\right\|_{L^q(\hat{K})} \leq C h_K^{-l+d/q} \left\|D^k \hat{v}^h\right\|_{L^p(\hat{K})} \\ &\leq C_{\text{inv}} h_K^{k-l+d/q-d/p} \left\|D^k v^h\right\|_{L^p(K)}.\end{aligned}$$

■

Remark C.21. On the proof. The crucial point in the proof was the equivalence of all norms in finite dimensional spaces. Such a property does not exist in infinite dimensional spaces. □

Corollary C.22. Global inverse estimate. *Let $p = q$ and let $\mathcal{T}^h$ be a regular triangulation of Ω, then*

$$\left\|D^l v^h\right\|_{L^{p,h}(\Omega)} \leq C_{\text{inv}} h^{k-l} \left\|D^k v^h\right\|_{L^{p,h}(\Omega)},$$

where

$$\|\cdot\|_{L^{p,h}(\Omega)} = \left(\sum_{K \in \mathcal{T}^h} \|\cdot\|_{L^p(K)}^p\right)^{1/p}.$$

Remark C.23. On $\|\cdot\|_{L^{p,h}(\Omega)}$. The cell wise definition of the norm is important for $l \geq 2$ since in this case finite element functions generally do not possess the regularity for the global norm to be well defined. It is also important for $l \geq 1$ and non-conforming finite element functions. □

C.3 Interpolation of Non-Smooth Functions

Remark C.24. Motivation. The interpolation theory of Section C.1 requires that the interpolation operator is continuous on the Sobolev space to which the function belongs that should be interpolated. But if one, e.g., wants to interpolate discontinuous functions with continuous, piecewise linear elements, then Section C.1 does not provide estimates.

A simple remedy seems to be first to apply some smoothing operator to the function to be interpolated and then to interpolate the smoothed function. However, this approach leads to difficulties at the boundary of Ω and it will not be considered further.

There are two often used interpolation operators for non-smooth functions. The interpolation operator of Clément (1975) is defined for functions from $L^1(\Omega)$ and it can be generalized to more or less all finite elements. The interpolation operator of Scott and Zhang (1990) is more special. It has the advantage that it preserves homogeneous Dirichlet boundary conditions in a natural way. For the Clément interpolation operator, one needs a modification for the preservation of homogeneous Dirichlet boundary conditions, which cannot be generalized easily to the non-homogeneous case. Here, only the interpolation operator of Clément, for linear finite elements, will be considered.

Let $\mathcal{T}^h$ be a regular triangulation of the polyhedral domain $\Omega \subset \mathbb{R}^d, d \in \{2, 3\}$, with simplicies K. Denote by P_1 the space of continuous, piecewise linear finite elements on $\mathcal{T}^h$. □

Remark C.25. Construction of the interpolation Operator of Clément. For each vertex V_i of the triangulation, the union of all grid cells which possess V_i as vertex will be denoted by ω_i, see Figure B.1.

The interpolation operator of Clément is defined with the help of local $L^2(\omega_i)$ projections. Let $v \in L^1(\Omega)$ and let $P_1(\omega_i)$ be the space of continuous piecewise linear finite elements on ω_i. Then, the local $L^2(\omega_i)$ projection of $v \in L^1(\omega_i)$ is the solution $p_i \in P_1(\omega_i)$ of

$$\int_{\omega_i} (v - p_i)(\boldsymbol{x}) q(\boldsymbol{x}) \, d\boldsymbol{x} = 0 \quad \forall\, q \in P_1(\omega_i) \tag{C.14}$$

or equivalently of

$$(v - p_i, q)_{L^2(\omega_i)} = 0 \quad \forall\, q \in P_1(\omega_i).$$

Then, the Clément interpolation operator is defined by

$$P^h_{\mathrm{Cle}} v(\boldsymbol{x}) = \sum_{i=1}^{N} p_i(V_i) \phi^h_i(\boldsymbol{x}), \tag{C.15}$$

where $\{\phi_i^h\}_{i=1}^N$ is the standard basis of the global finite element space P_1. Since $P^h_{\mathrm{Cle}} v(\boldsymbol{x})$ is a linear combination of basis functions of P_1, it defines a map $P^h_{\mathrm{Cle}} : L^1(\Omega) \to P_1$. □

Theorem C.26. Interpolation estimate. *Let $k, l \in \mathbb{N} \cup \{0\}$ and $q \in \mathbb{R}$ with $k \le l \le 2$, $1 \le q \le \infty$, and let ω_K be the union of all subdomains ω_i that contain the mesh cell K, see Figure C.2. Then it holds for all $v \in W^{l,q}(\omega_K)$ the estimate*

$$\left\| D^k (v - P^h_{\mathrm{Cle}} v) \right\|_{L^q(K)} \le C h^{l-k} \left\| D^l v \right\|_{L^q(\omega_K)}, \tag{C.16}$$

with $h = diam(\omega_K)$, where the constant C is independent of $v(\boldsymbol{x})$ and h.

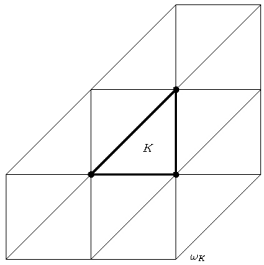

Fig. C.2 A subdomain ω_K.

Proof. The statement of the lemma is obvious in the case $k = l = 2$ since it is $D^2 P^h_{\mathrm{Cle}} v(\boldsymbol{x})|_K = 0$.

Let $k \in \{0, 1\}$. Because the $L^2(\Omega)$ projection gives an element with best approximation, one gets with (C.14)

$$P^h_{\mathrm{Cle}} p = p \quad \text{in } K \quad \forall\, p \in P_1(\omega_K). \tag{C.17}$$

One says that P^h_{Cle} is a consistent operator.

The next step consists in proving the stability of P^h_{Cle}. One obtains with the inverse inequality (C.13)

$$\|p\|_{L^\infty(\omega_i)} \le c h^{-d/2} \|p\|_{L^2(\omega_i)} \qquad \text{for all } p \in P_1(\omega_i).$$

The inverse inequality and definition (C.14) of the local L^2 projection with the test function $q = p_i$ gives

$$\|p_i\|^2_{L^\infty(\omega_i)} \le c h^{-d} \|p_i\|^2_{L^2(\omega_i)} \le c h^{-d} \|v\|_{L^1(\omega_i)} \|p_i\|_{L^\infty(\omega_i)}.$$

Dividing by $\|p_i\|_{L^\infty(\omega_i)}$ and applying Hölder's inequality, one obtains for $p^{-1} = 1 - q^{-1}$ (*exercise*)

$$\begin{aligned} |p_i(V_i)| &\le \|p_i\|_{L^\infty(\omega_i)} \le c h^{-d} \|v\|_{L^1(\omega_i)} = c h^{-d} \|1v\|_{L^1(\omega_i)} \\ &\le c h^{-d} \|v\|_{L^q(\omega_i)} \underbrace{\|1\|_{L^p(\omega_i)}}_{= c h^{d/p}} = c h^{d(1/p-1)} \|v\|_{L^q(\omega_i)} = c h^{-d/q} \|v\|_{L^q(\omega_i)} \end{aligned} \tag{C.18}$$

for all $V_i \in K$. From the regularity of the triangulation, it follows for the basis functions that (inverse estimate)

$$\left\|D^k \phi_i\right\|_{L^\infty(K)} \leq c h^{-k}, \quad k = 0, 1. \tag{C.19}$$

Using the triangle inequality, combining (C.18) and (C.19) yields the stability of P^h_{Cle}

$$\begin{aligned}
\left\|D^k P^h_{\mathrm{Cle}} v\right\|_{L^q(K)} &\leq \sum_{V_i \in K} |p_i(V_i)| \left\|D^k \phi_i\right\|_{L^q(K)} \\
&\leq c \sum_{V_i \in K} h^{-d/q} \|v\|_{L^q(\omega_i)} \left\|D^k \phi_i\right\|_{L^\infty(K)} \|1\|_{L^q(K)} \\
&\leq c \sum_{V_i \in K} h^{-d/q} \|v\|_{L^q(\omega_i)} h^{-k} h^{d/q} \\
&= c h^{-k} \|v\|_{L^q(\omega_K)} .
\end{aligned} \tag{C.20}$$

The remainder of the proof follows the proof of the interpolation error estimate for the polynomial interpolation, Theorem C.10, apart from the fact that a reference cell is not used for the Clément interpolation operator. Using Lemma C.2 and C.4, one can find a polynomial $p \in P_1(\omega_K)$ with

$$\left\|D^j (v - p)\right\|_{L^q(\omega_K)} \leq c h^{l-j} \left\|D^l v\right\|_{L^q(\omega_K)}, \quad 0 \leq j \leq l \leq 2. \tag{C.21}$$

With (C.17), the triangle inequality, $\|\cdot\|_{L^q(K)} \leq \|\cdot\|_{L^q(\omega_K)}$, (C.20), and (C.21), one obtains

$$\begin{aligned}
\left\|D^k \left(v - P^h_{\mathrm{Cle}} v\right)\right\|_{L^q(K)} &= \left\|D^k \left(v - p + P^h_{\mathrm{Cle}} p - P^h_{\mathrm{Cle}} v\right)\right\|_{L^q(K)} \\
&\leq \left\|D^k (v - p)\right\|_{L^q(K)} + \left\|D^k P^h_{\mathrm{Cle}} (v - p)\right\|_{L^q(K)} \\
&\leq \left\|D^k (v - p)\right\|_{L^q(\omega_K)} + c h^{-k} \|v - p\|_{L^q(\omega_K)} \\
&\leq c h^{l-k} \left\|D^l v\right\|_{L^q(\omega_K)} + c h^{-k} h^l \left\|D^l v\right\|_{L^q(\omega_K)} \\
&= c h^{l-k} \left\|D^l v\right\|_{L^q(\omega_K)} .
\end{aligned}$$

■

Remark C.27. Uniform meshes.

- If all mesh cells in ω_K are of the same size, then one can replace h by h_K in the interpolation error estimate (C.16). This property is given in many cases.
- If one assumes that the number of mesh cells in ω_K is bounded uniformly for all considered triangulations, the global interpolation estimate

$$\left\|D^k (v - P^h_{\mathrm{Cle}} v)\right\|_{L^q(\Omega)} \leq C h^{l-k} \left\|D^l v\right\|_{L^q(\Omega)}, \quad 0 \leq k \leq l \leq 2,$$

  follows directly from (C.16).

□

Remark C.28. Other finite element spaces. The idea of the Clément interpolation can be extended to other finite element spaces, see Clément (1975). In this paper, it is just assumed that the global functionals are values or derivatives of the function in the nodes. Optimal interpolation estimates are given in Clément (1975). □

Remark C.29. Preservation of homogeneous Dirichlet boundary conditions. For global finite element spaces $V^h \subset H_0^1(\Omega)$, it is shown in Clément (1975) that homogeneous Dirichlet boundary conditions can be preserved under some (weak) assumptions on the finite element space. First, the analysis of Clément (1975) is restricted to finite element spaces with certain global functionals as mentioned in Remark C.28. In addition, it is assumed that for the nodes on the boundary the functionals are only values of the function (and no derivatives). For the definition of the global Clément interpolation operator, these values are left unchanged, i.e., equal to zero, and the interpolation is computed for all other degrees of freedom. For this construction, optimal interpolation estimates were proved in Clément (1975).

As a consequence, for finite element spaces $V^h = P_k \cap H_0^1(\Omega)$ or $V^h = Q_k \cap H_0^1(\Omega)$, the Clément interpolant of $v \in H_0^1(\Omega)$ into V^h is well defined and, in particular, the homogeneous Dirichlet boundary values are preserved. □

Appendix D
Examples for Numerical Simulations

Remark D.1. General considerations. The definition of good test examples is of importance for the assessment of numerical schemes. There are different classes of test problems:

- *Academic test examples with prescribed solution.* In these examples, the velocity field $\boldsymbol{u}$ and the pressure p are prescribed by analytical functions. The right hand side, boundary conditions, and the initial condition are chosen such that the strong form of the considered equation is fulfilled. These examples serve for supporting convergence estimates. A connection to real life problems is generally not given.
  In the definition of these examples, one has to take care that the velocity field is divergence-free. Depending on the type of boundary condition, see Section 1.4, the integral mean value of the pressure has to vanish or the integral of the Dirichlet boundary data has to fulfill the compatibility condition (1.25).
- *Academic test examples with features of real life flows.* These examples contain on the one hand some important features of real life flow problems, but on the other hand, a number of simplifications are used to facilitate their implementation and the assessment of the results. Generally, an analytical solution is not known. Reference values for quantities of interest are obtained by performing simulations on very fine grids in space and time.
- *Real life examples.* A derived method should work well for this type of examples. However, often the data are incomplete in real life examples, e.g. a temporal and spatial resolved boundary condition is generally not known. In addition, reference values to compare with are coming often from measurements. These values are generally mean values in space or in time. A certain measurement error has always to be expected.
  In conclusion, special care and special techniques are necessary to assess numerical methods at real world problems. Generally, none of the methods will produce results which agree completely with the real world flow,