Appendix B
Finite Element Methods

Remark B.1. Contents. This appendix provides a short introduction into finite element methods. In particular, notations are introduced that are used throughout this monograph and finite element spaces are described that are of importance for the discretization of incompressible flow problems. □

B.1 The Ritz Method and the Galerkin Method

Remark B.2. 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.3. 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 second estimate the necessary criterion for a local minimum of the expression of the first bound is used. Hence, the function $F(\cdot)$ is bounded from below and

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

exists.

Let $\{v_k\}_{k\in\mathbb{N}}$ be a sequence with $F(v_k) \to \kappa$ 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 $\kappa \le F(v)$ for all $v \in V$, one obtains

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

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 $u \in V$, see Definition A.5. Since $F(\cdot)$ is continuous, it is $F(u) = \kappa$ 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 for arbitrary $v \in V$

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

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 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 (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.4. 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.3 that the elements Tu and $T'u$ exist and they are defined uniquely. 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

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

for all $u \in V$. Hence, T is bounded. Since T is linear, it follows that T is continuous, see Lemma A.62. 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 (A.10) 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

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

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.3, it follows that there is a exactly one $w \in V$ with

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

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

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

and consequently, $u = T'w$ is a solution of (B.3).

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

Remark B.5. 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 have 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.6. 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.3, 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.7. 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 (A.10) that

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

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, since the error cannot be smaller than the best approximation error. If $\|u - u_k\|_V = 0$, the statement of the lemma is trivially true. ■

Theorem B.8. 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

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

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

Remark B.9. 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) hold, then it holds in particular for each basis function ϕ_i.

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 $A\underline{u} = \underline{f}$, 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

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

□

Remark B.10. 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.11. 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.4. ■

Lemma B.12. 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 let* 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.47) *and* (A.48)*.*

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.48), the Galerkin orthogonality, and (A.47), 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.13. 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.14. 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. □

Lemma B.15. Inf-sup criterion for finite-dimensional spaces. *Let* V^h *be a finite-dimensional space with inner product* $(\cdot, \cdot)_{V^h}$ *and induced norm* $\|\cdot\|_{V^h} = (\cdot, \cdot)_{V^h}^{1/2}$*. Consider a bilinear form* $a : V^h \times V^h \to \mathbb{R}$ *and a linear functional* $f : V^h \to \mathbb{R}$*. Then, the problem to find* $u^h \in V^h$ *such that*

$$a\left(u^h, v^h\right) = f\left(v^h\right) \tag{B.13}$$

has a unique solution for all $f(\cdot)$ *if and only if*

$$\inf_{w^h \in V^h \setminus \{0\}} \sup_{v^h \in V^h \setminus \{0\}} \frac{a\left(v^h, w^h\right)}{\|v^h\|_{V^h} \|w^h\|_{V^h}} \geq \beta_{\text{is}}^h > 0. \tag{B.14}$$

Proof. Denote by $n \in \mathbb{N}$ the dimension of V^h and let $\left\{\varphi_i^h\right\}_{i=1}^n$ be a basis of V^h. Then there are representations

$$v^h = \sum_{i=1}^n v_i \varphi_i^h, \quad w^h = \sum_{i=1}^n w_i \varphi_i^h,$$

with $\underline{v} = (v_1, \ldots, v_n)^T$, $\underline{w} = (w_1, \ldots, w_n)^T$ and there are matrices $A, M \in \mathbb{R}^{n\times n}$ such that

$$a\left(v^h, w^h\right) = \underline{v}^T A^T \underline{w}, \quad \left\|v^h\right\|_{V^h} = \left(\underline{v}^T M^T \underline{v}\right)^{1/2}, \quad \left\|w^h\right\|_{V^h} = \left(\underline{w}^T M^T \underline{w}\right)^{1/2}.$$

The matrix M is symmetric and positive definite. The inf-sup condition (B.14) can be written in the form

$$\inf_{\underline{w}\in\mathbb{R}^n\setminus\{\underline{0}\}} \sup_{\underline{v}\in\mathbb{R}^n\setminus\{\underline{0}\}} \frac{\underline{v}^T A^T \underline{w}}{\left(\underline{v}^T M \underline{v}\right)^{1/2} \left(\underline{w}^T M \underline{w}\right)^{1/2}} \ge \beta_{\text{is}}^h > 0. \tag{B.15}$$

Problem (B.13) has a unique solution for all f if and only if the matrix A is non-singular.

(B.15) *holds* $\Longrightarrow$ *A is non-singular.* Assume that (B.15) holds and A is a singular matrix. Then there is a vector $\underline{v} \neq \underline{0}$ such that $A\underline{v} = \underline{0}$ or equivalently $\underline{v}^T A^T = \underline{0}^T$. Hence, the supremum in (B.15) is zero such that (B.15) cannot hold, which is a contradiction to the assumption.

A is non-singular $\Longrightarrow$ (B.15) *holds.* With A also A^T is non-singular. Then, for each $\underline{v} \in \mathbb{R}^n \setminus \{\underline{0}\}$ there is a unique $\underline{w} \in \mathbb{R}^n \setminus \{\underline{0}\}$ such that $\underline{v} = M^{-1}A^T\underline{w}$, since M and A^T are non-singular matrices. Inserting this vector in (B.15) and using the symmetry and positive definiteness of M gives

$$\begin{aligned}
&\inf_{\underline{w}\in\mathbb{R}^n\setminus\{\underline{0}\}} \frac{\underline{w}^T A M^{-1} A^T \underline{w}}{\left(\underline{w}^T A M^{-1} M M^{-1} A^T \underline{w}\right)^{1/2} \left(\underline{w}^T M \underline{w}\right)^{1/2}} \\
&= \inf_{\underline{w}\in\mathbb{R}^n\setminus\{\underline{0}\}} \frac{\underline{w}^T A M^{-1} A^T \underline{w}}{\left(\underline{w}^T A M^{-1} A^T \underline{w}\right)^{1/2} \left(\underline{w}^T M \underline{w}\right)^{1/2}} \\
&= \inf_{\underline{w}\in\mathbb{R}^n\setminus\{\underline{0}\}} \left(\frac{\underline{w}^T A M^{-1} A^T \underline{w}}{\underline{w}^T M \underline{w}}\right)^{1/2} \\
&= \inf_{\underline{w}\in\mathbb{R}^n\setminus\{\underline{0}\}} \left(\frac{\left(\underline{w}M^{1/2}\right)^T \left(M^{-1/2} A M^{-1/2}\right) \left(M^{-T/2} A^T M^{-T/2}\right) \left(M^{1/2}\underline{w}\right)}{\left(\underline{w}M^{1/2}\right)^T \left(M^{1/2}\underline{w}\right)}\right)^{1/2}.
\end{aligned}$$

Hence, one obtains a Rayleigh quotient. From Lemma A.19, it is known that the infimum of a Rayleigh quotient is attained and it is the smallest eigenvalue of the eigenvalue problem

$$\left(M^{-1/2} A M^{-1/2}\right) \left(M^{-T/2} A^T M^{-T/2}\right) \left(M^{1/2}\underline{w}\right) = \lambda \left(M^{1/2}\underline{w}\right).$$

This problem is an eigenvalue problem for a symmetric matrix, hence all eigenvalues are real. Since

$$\begin{aligned}
&\underline{w}^T \left(M^{-1/2} A M^{-1/2}\right) \left(M^{-T/2} A^T M^{-T/2}\right) \underline{w} \\
&= \left(\left(M^{-T/2} A^T M^{-T/2}\right) \underline{w}\right)^T \left(\left(M^{-T/2} A^T M^{-T/2}\right) \underline{w}\right) \\
&= \left\|\left(M^{-T/2} A^T M^{-T/2}\right) \underline{w}\right\|_2^2 \ge 0,
\end{aligned}$$

the matrix is positive semi-definite, such that all eigenvalues are non-negative. Finally, since A and M are non-singular matrices, the matrix of this eigenvalue problem is non-singular and all eigenvalues are positive. Hence, there is a positive constant β_{is}^h such that

$$\inf_{\underline{w}\in\mathbb{R}^n\setminus\{\underline{0}\}} \frac{\underline{v}^T A^T \underline{w}}{\left(\underline{v}^T M \underline{v}\right)^{1/2} \left(\underline{w}^T M \underline{w}\right)^{1/2}} \ge \beta_{\text{is}}^h$$

with $\underline{v} = M^{-1}A^T\underline{w}$. Taking now the supremum with respect to $\underline{v}$ might only increase the left-hand side and (B.15) follows. ■

B.2 Finite Element Spaces

Remark B.16. 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), $0 \le m \le d-1$, 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.17. Finite-dimensional spaces defined on K. Let $s \in \mathbb{N}$. Finite element methods use finite-dimensional spaces $P(K) \subset C^s(K)$ that are defined on K. In general, $P(K)$ consists of polynomials. The dimension of $P(K)$ will be denoted by $\dim P(K) = N_K$. □

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

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 that 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 $\underline{a} \in \mathbb{R}^{N_K}$, $\underline{a} = (a_1, \ldots, a_{N_K})^T$, exactly one $p \in P(K)$ with

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

□

Remark B.20. Local basis. Unisolvence means that for each vector $\underline{a} \in \mathbb{R}^{N_K}$, $\underline{a} = (a_1, \ldots, a_{N_K})^T$, 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 $\underline{a}$, 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. □

Remark B.22. Triangulation, grid, mesh, grid cell. For the definition of global finite element spaces, a decomposition of the domain Ω into polyhedra K is needed. This decomposition is called triangulation $\mathcal{T}^h$ and the polyhedra K are called mesh cells. The union of the polyhedra is called grid or mesh.

A triangulation is called admissible, see the definition in (Ciarlet, 1978, p. 38, p. 51), if:

- 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.23. Global and local functionals. Let $\Phi_1,\ldots,\Phi_N \ : \ C^s(\overline{\Omega}) \to \mathbb{R}$ be continuous linear functionals of the same types as given in Remark B.18. The restriction of the functionals to $C^s(K)$ defines a set of 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.24. 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 that possess this vertex. □

Definition B.25. 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

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

□

Remark B.26. 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.

Whether the continuity with respect to $\{\Phi_i\}_{i=1}^N$ implies the continuity of the finite element functions depends on the functionals that define the finite element space. □

Definition B.27. Parametric finite elements. Let $\hat{K}$ be a reference mesh cell with the local space $\hat{P}(\hat{K})$, 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

$$P(K) = \Big\{ p \ : \ p = \hat{p} \circ F_K^{-1}, \hat{p} \in \hat{P}(\hat{K}) \Big\}. \tag{B.16}$$

- The local functionals are defined by

$$\Phi_{K,i}(v(\boldsymbol{x})) = \hat{\Phi}_i \left(v(F_K(\hat{\boldsymbol{x}})) \right), \tag{B.17}$$

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

□

Remark B.28. Motivations for using parametric finite elements. Definition B.25 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.29. *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.30. 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 tetrahedra. □

Definition B.31. 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.32. On barycentric coordinates. From the definition, it follows that the barycentric coordinates are the solution of the linear system of equations

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

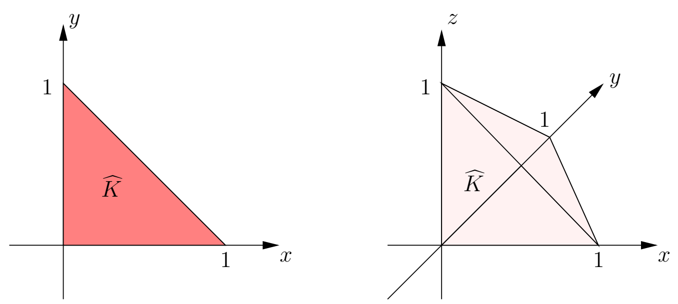

Fig. B.1 The unit simplices in two and three dimensions.

Since the system matrix is non-singular, see Remark B.30, the barycentric coordinates are determined uniquely.

The barycentric coordinates of the vertex $\boldsymbol{a}_i$, $i = 1, \ldots, d{+}1$, of the simplex are $\lambda_i = 1$ 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 that has the value 1 in the vertex $\boldsymbol{a}_i$ and that 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.33. Simplicial reference mesh cells. A commonly used reference mesh cell for triangles and tetrahedra 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.1. The class $\{F_K\}$ of admissible mappings are the bijective affine mappings

$$F_K \hat{\boldsymbol{x}} = B_K \hat{\boldsymbol{x}} + \boldsymbol{b}, \qquad B_K \in \mathbb{R}^{d \times d}, \ \det(B_K) \neq 0, \ \boldsymbol{b} \in \mathbb{R}^d. \tag{B.18}$$

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

Definition B.34. 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.16) and (B.17), 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. □

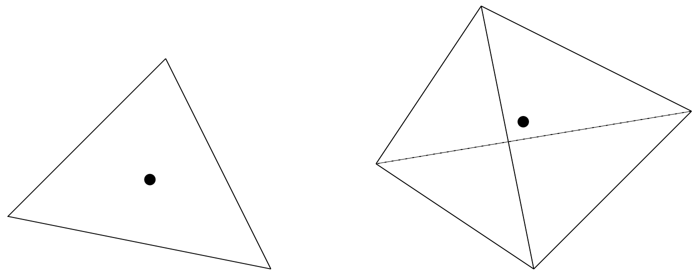

Fig. B.2 The finite element $P_0(K)$.

Definition B.35. 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

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

□

Remark B.36. Lagrangian finite elements. In many 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.37. 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.2. It is $\dim P_0(K) = 1$. □

Example B.38. 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.3. 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 $\Phi_i(\lambda_j) = \delta_{ij}$, compare Remark B.32. 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 that are defined on E. These functionals can be chosen to be the values of the function in

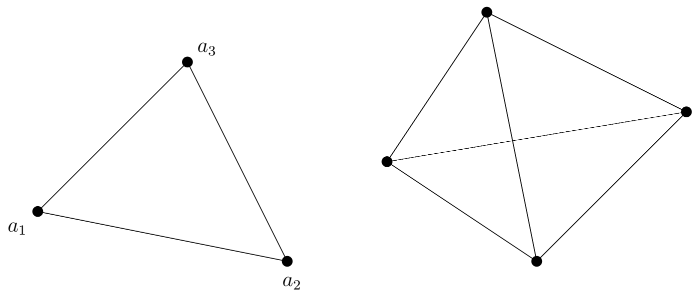

Fig. B.3 The finite element $P_1(K)$.

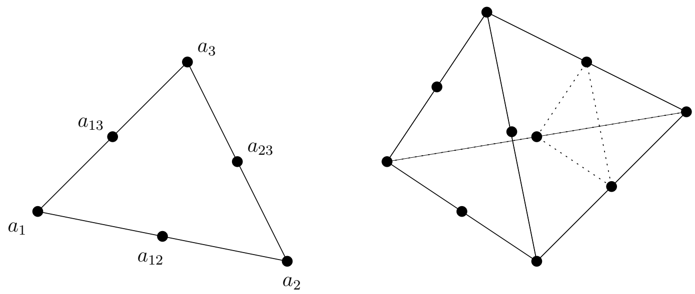

Fig. B.4 The finite element $P_2(K)$.

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 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.39. 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.4. 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 $\dim P_2(K) = (d+1)(d+2)/2$.

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

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

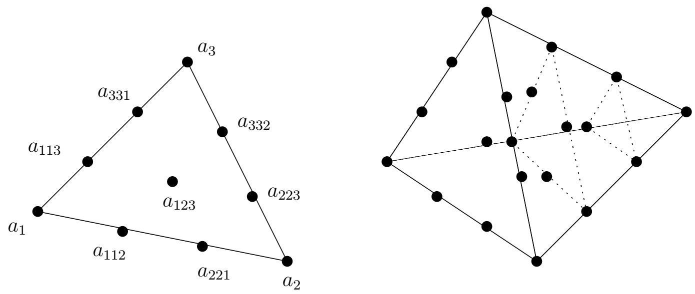

Fig. B.5 The finite element $P_3(K)$.

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 defined 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.40. 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.5. 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

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

For the functionals

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

the local basis is given by

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

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

Example B.41. 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.4 and B.5 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.42. 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

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

compare Figures B.4 and B.5 for the notations. Then, the local basis is given by

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

In 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

$$\big\{\lambda_1(2\lambda_1-1)+3\lambda_1(\lambda_2\lambda_3+\lambda_2\lambda_4+\lambda_3\lambda_4)-4\lambda_1\lambda_2\lambda_3\lambda_4,\ldots,$$
$$\lambda_1\lambda_2(4-12\lambda_4-12\lambda_3+32\lambda_3\lambda_4),\ldots,$$
$$27\lambda_1\lambda_2\lambda_3(1-4\lambda_4),\ 27\lambda_1\lambda_2(1-4\lambda_3)\lambda_4,\ 27\lambda_1(1-4\lambda_2)\lambda_3\lambda_4,$$
$$27(1-4\lambda_1)\lambda_2\lambda_3\lambda_4,\ 256\lambda_1\lambda_2\lambda_3\lambda_4\big\},$$

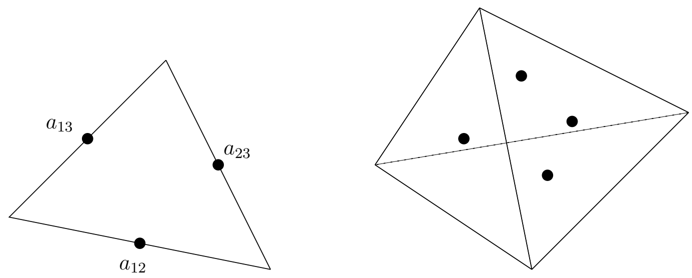

Fig. B.6 The finite element $P_1^{\mathrm{nc}}(K)$.

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

Example B.43. P_1^{nc} : non-conforming linear finite element, Crouzeix–Raviart finite element, Crouzeix & 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.25, that the functions from P_1^{nc} are continuous in the barycenter of the faces

$$P_1^{\mathrm{nc}} = \big\{ v \in L^2(\Omega) \ : \ v|_K \in P_1(K),\ v(\boldsymbol{x}) \text{ is continuous at the barycenter of all faces} \big\}. \tag{B.19}$$

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

$$P_1^{\mathrm{nc}} = \left\{ v \in L^2(\Omega) \ : \ v|_K \in P_1(K), \int_E v|_K \, d\boldsymbol{s} = \int_E v|_{K'} \, d\boldsymbol{s} \ \forall\, E \in \mathcal{E}(K) \cap \mathcal{E}(K') \right\}, \tag{B.20}$$

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

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, see Figure B.6. This system is unisolvent with the local basis

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

□

Example B.44. 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

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

The corresponding local basis is given by

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

□

Example B.45. Raviart–Thomas finite elements RT_k. Raviart–Thomas finite elements are a class of vector-valued finite elements that approximate the space $H(\operatorname{div}, \Omega)$, see (2.22). Details of their definition and important properties can be found, e.g., in (Boffi *et al.*, 2013, pp. 84).

Consider a simplicial triangulation with mesh cells $\{K\}$ and let $\boldsymbol{P}_k(K) = (P_k(K))^d$, $k \ge 0$. Then, the following local polynomial space is defined directly on K

$$\mathrm{RT}_k(K) = \{\boldsymbol{v} \in (\boldsymbol{P}_k(K) + \boldsymbol{x}P_k(K))\ :\ \boldsymbol{v}\cdot\boldsymbol{n}_{\partial K} \in R_k(\partial K)\}\,, \quad k \ge 0,$$

where

$$R_k(\partial K) = \left\{\varphi \in L^2(\partial K)\ :\ \varphi|_E \in P_k(E) \text{ for all faces } E \subset \partial K\right\}.$$

It is noted in (Boffi *et al.*, 2013, Rem. 2.3.1) that this definition of $\mathrm{RT}_k(K)$ is different than the original definition in Raviart & Thomas (1977).

In particular, the Raviart–Thomas element of lowest order is given by

$$\mathrm{RT}_0(K) = \{\boldsymbol{v} \in (\boldsymbol{P}_0(K) + \boldsymbol{x}P_0(K))\ :\ \boldsymbol{v}\cdot\boldsymbol{n}_{\partial K} \in R_0(\partial K)\}\,,$$

i.e., it is linear on K. Its local functionals are $\boldsymbol{v} \cdot \boldsymbol{n}_{E_i}$, $i = 1, \ldots, d+1$. A function from $\mathrm{RT}_0(K)$ can be written in the form

$$\boldsymbol{v}(\boldsymbol{x}) = \boldsymbol{a} + b\boldsymbol{x}, \quad \boldsymbol{a} \in \mathbb{R}^d, b \in \mathbb{R}, \boldsymbol{x} \in K. \tag{B.21}$$

A face $E \subset \partial K$ is a hyperplane that can be represented in the form

$$\boldsymbol{x} \cdot \boldsymbol{n}_E = c, \quad c \in \mathbb{R}, \ \forall\, \boldsymbol{x} \in E.$$

Inserting this representation in (B.21) yields

$$\boldsymbol{v}(\boldsymbol{x}) \cdot \boldsymbol{n}_E = \boldsymbol{a} \cdot \boldsymbol{n}_E + b\boldsymbol{x} \cdot \boldsymbol{n}_E = \boldsymbol{a} \cdot \boldsymbol{n}_E + bc = \mathrm{const} \quad \forall\, \boldsymbol{x} \in E.$$

Thus, the normal component of $\boldsymbol{v}$ on each face is a constant.

The global space RT_0 is defined as usual by defining global functionals on the basis of the local functionals and requiring the continuity of the global functionals, see Definition B.25. Consequently, the normal component of functions from RT_0 is continuous across faces of the mesh cells. Since the normal component on each face is a constant, it is sufficient for requiring its continuity to require the continuity in the barycenters $\{\boldsymbol{m}_E\}$ of $\{E\}$. From Lemma 2.56, it follows that $\mathrm{RT}_0 \subset H(\mathrm{div}, \Omega)$. □

B.4 Finite Elements on Parallelepipeds and Quadrilaterals

Remark B.46. Reference mesh cells, reference map to parallelepipeds. One 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 $\{F_K\}$ to parallelepipeds consists of bijective affine mappings of the form

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

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

The class of mesh cells that is 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 Remark B.55. □

Definition B.47. 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

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

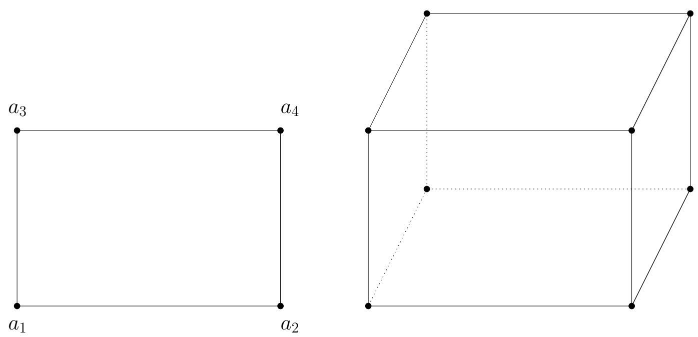

Fig. B.7 The finite element $Q_1(K)$.

□

Remark B.48. 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.49. 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.50. 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.7. Hence, it is $\dim Q_1(K) = 2^d$.

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.51. 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

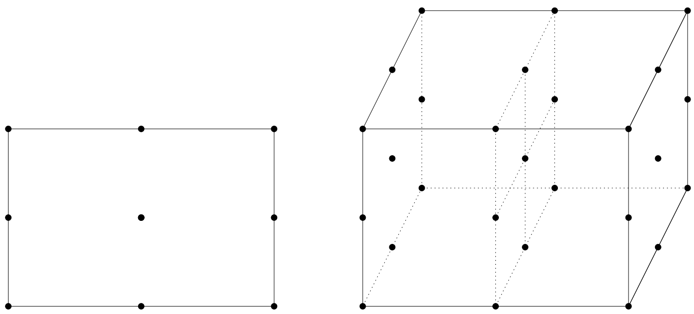

Fig. B.8 The finite element $Q_2(K)$.

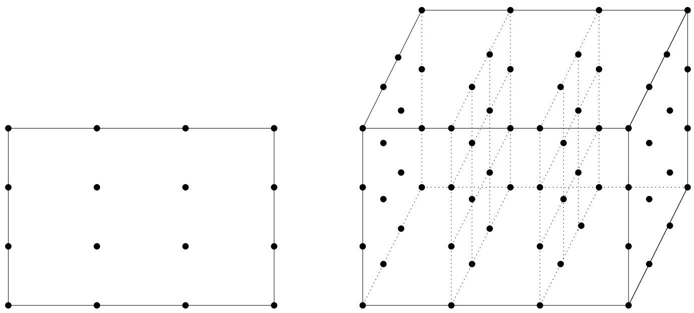

Fig. B.9 The finite element $Q_3(K)$.

function at both ends of the interval and in the center of the interval, see Figure B.8. In d dimensions, they are the corresponding values of the tensor product of the intervals. It follows that $\dim Q_2(K) = 3^d$.

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.52. Q_3 : conforming piecewise d-cubic 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.9. 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

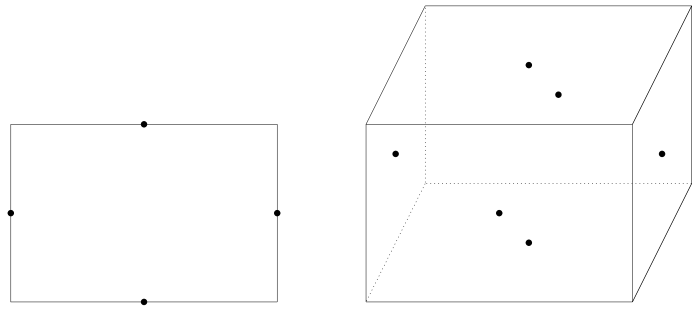

Fig. B.10 The finite element $Q_1^{\mathrm{rot}}(K)$.

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

□

Example B.53. Q_1^{rot} : rotated non-conforming element of lowest order, Rannacher–Turek element, Rannacher & 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 that 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

$$\begin{aligned} \hat{\phi}_1(\hat{x},\hat{y}) &= -\frac{3}{8}(\hat{x}^2-\hat{y}^2) - \frac{1}{2}\hat{y} + \frac{1}{4}, \quad \hat{\phi}_2(\hat{x},\hat{y}) = \frac{3}{8}(\hat{x}^2-\hat{y}^2) + \frac{1}{2}\hat{x} + \frac{1}{4}, \\ \hat{\phi}_3(\hat{x},\hat{y}) &= -\frac{3}{8}(\hat{x}^2-\hat{y}^2) + \frac{1}{2}\hat{y} + \frac{1}{4}, \quad \hat{\phi}_4(\hat{x},\hat{y}) = \frac{3}{8}(\hat{x}^2-\hat{y}^2) - \frac{1}{2}\hat{x} + \frac{1}{4}. \end{aligned} \tag{B.22}$$

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.10, or as integral mean values of the functions at the faces.

Consequently, the finite element spaces are defined in the same way as (B.19) or (B.20), with $P_1^{\mathrm{nc}}(K)$ replaced by $Q_1^{\mathrm{rot}}(K)$.

For a discussion of the practical use of this finite element, it is referred to Remark 2.145. □

Example B.54. P_k^{disc}, $k \geq 1$. The space P_k^{disc}, $k \geq 1$, is given by

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

The construction of a basis on the reference mesh cell is based on the Legendre polynomials in $[-1,1]$, which are given by

$$1, \hat{x}, \frac{1}{2}\left(3\hat{x}^2 - 1\right), \ldots. \tag{B.23}$$

Then, the basis of $P_k^{\mathrm{disc}}(\hat{K})$ in multiple dimensions is defined as a tensor product of polynomials of type (B.23) that gives a polynomial of degree smaller than or equal to k. Concretely, the basis of $P_1^{\mathrm{disc}}(\hat{K})$ is given by

$$\begin{cases} 2d: & \{1, \hat{x}, \hat{y}\}, \\ 3d: & \{1, \hat{x}, \hat{y}, \hat{z}\}, \end{cases} \tag{B.24}$$

and of $P_2^{\mathrm{disc}}(\hat{K})$

$$\begin{cases} 2d: & \left\{1, \hat{x}, \hat{y}, \hat{x}\hat{y}, \frac{1}{2}\left(3\hat{x}^2 - 1\right), \frac{1}{2}\left(3\hat{y}^2 - 1\right)\right\}, \\ 3d: & \left\{1, \hat{x}, \hat{y}, \hat{z}, \hat{x}\hat{y}, \hat{x}\hat{z}, \hat{y}\hat{z}, \frac{1}{2}\left(3\hat{x}^2 - 1\right), \frac{1}{2}\left(3\hat{y}^2 - 1\right), \frac{1}{2}\left(3\hat{z}^2 - 1\right)\right\}. \end{cases} \tag{B.25}$$

For the definition of the local nodal functionals, the $L^2(\hat{K})$ orthogonality of the Legendre polynomials is used. Denoting the basis functions of (B.24) and (B.25) by $\hat{\phi}_j(\hat{\boldsymbol{x}})$, then these functionals are defined by

$$\Phi_{\hat{K},i}\left(\hat{\phi}_j\right) = \left(\int_{\hat{K}} \hat{\phi}_j^2 \ d\hat{\boldsymbol{x}}\right)^{-1} \int_{\hat{K}} \hat{\phi}_i \hat{\phi}_j \ d\hat{\boldsymbol{x}},$$

such that $\Phi_{\hat{K},i}\left(\hat{\phi}_j\right) = \delta_{ij}$. □

Remark B.55. 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 d-quadrilaterals, then the class of admissible reference maps has to be enlarged.

The simplest non-affine 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.56. Parametric finite element functions. The functions of the local space $P(K)$ on the mesh cell K are defined by $p = \hat{p} \circ F_K^{-1}$, where $\hat{p}$ is a polynomial. 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. It follows that the functions of the corresponding finite element space are continuous, compare Example B.38. □

B.5 Transform of Integrals

Remark B.57. Motivation. The transformation of integrals from the reference mesh cell to mesh cells of the grid and vice versa is used as well for the 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 $\boldsymbol{x} = F_K(\hat{\boldsymbol{x}})$ 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

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

□

Remark B.58. Integral with a function without derivatives. This integral transforms with the standard rule of integral transforms

$$\int_K v(\boldsymbol{x})\ d\boldsymbol{x} = \int_{\hat{K}} \hat{v}(\hat{\boldsymbol{x}})\,|\det DF_K(\hat{\boldsymbol{x}})|\ d\hat{\boldsymbol{x}}, \tag{B.26}$$

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

Remark B.59. Transform of derivatives. Using the chain rule, one obtains

$$\begin{aligned}
\frac{\partial v}{\partial x_i}(\boldsymbol{x}) &= \sum_{j=1}^{d} \frac{\partial \hat{v}}{\partial \xi_j}(\hat{\boldsymbol{x}}) \frac{\partial \xi_j}{\partial x_i} = \nabla_{\hat{\boldsymbol{x}}} \hat{v}(\hat{\boldsymbol{x}}) \cdot \left(\left(DF_K^{-1}(\boldsymbol{x}) \right)^T \right)_i \\
&= \nabla_{\hat{\boldsymbol{x}}} \hat{v}(\hat{\boldsymbol{x}}) \cdot \left(\left(DF_K^{-1}(F_K(\hat{\boldsymbol{x}})) \right)^T \right)_i, \qquad \text{(B.27)} \\
\frac{\partial \hat{v}}{\partial \xi_i}(\hat{\boldsymbol{x}}) &= \sum_{j=1}^{d} \frac{\partial v}{\partial x_j}(\boldsymbol{x}) \frac{\partial x_j}{\partial \xi_i} = \nabla v(\boldsymbol{x}) \cdot \left(\left(DF_K(\hat{\boldsymbol{x}}) \right)^T \right)_i \\
&= \nabla v(\boldsymbol{x}) \cdot \left(\left(DF_K(F_K^{-1}(\boldsymbol{x})) \right)^T \right)_i. \qquad \text{(B.28)}
\end{aligned}$$

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.60. Integrals with a gradients. Using the rule for transforming integrals and (B.27) gives

$$\begin{aligned}
&\int_K \boldsymbol{b}(\boldsymbol{x}) \cdot \nabla v(\boldsymbol{x}) \, d\boldsymbol{x} \\
&= \int_{\hat{K}} \boldsymbol{b}\left(F_K(\hat{\boldsymbol{x}})\right) \cdot \left[\left(DF_K^{-1}\right)^T \left(F_K(\hat{\boldsymbol{x}})\right) \right] \nabla_{\hat{\boldsymbol{x}}} \hat{v}(\hat{\boldsymbol{x}}) \left|\det DF_K(\hat{\boldsymbol{x}})\right| \, d\hat{\boldsymbol{x}}. \qquad \text{(B.29)}
\end{aligned}$$

Similarly, one obtains

$$\begin{aligned}
&\int_K \nabla v(\boldsymbol{x}) \cdot \nabla w(\boldsymbol{x}) \, d\boldsymbol{x} \\
&= \int_{\hat{K}} \left[\left(DF_K^{-1}\right)^T \left(F_K(\hat{\boldsymbol{x}})\right) \right] \nabla_{\hat{\boldsymbol{x}}} \hat{v}(\hat{\boldsymbol{x}}) \cdot \left[\left(DF_K^{-1}\right)^T \left(F_K(\hat{\boldsymbol{x}})\right) \right] \nabla_{\hat{\boldsymbol{x}}} \hat{w}(\hat{\boldsymbol{x}}) \\
&\quad \times \left|\det DF_K(\hat{\boldsymbol{x}})\right| \, d\hat{\boldsymbol{x}}. \qquad \text{(B.30)}
\end{aligned}$$

□

Remark B.61. Integral with the divergence. Integrals of the following type are important for the Navier–Stokes equations

$$\begin{aligned}
&\int_K \nabla \cdot v(\boldsymbol{x}) q(\boldsymbol{x}) \, d\boldsymbol{x} = \int_K \sum_{i=1}^{d} \frac{\partial v_i}{\partial x_i}(\boldsymbol{x}) q(\boldsymbol{x}) \, d\boldsymbol{x} \\
&= \int_{\hat{K}} \sum_{i=1}^{d} \left[\left(\left(DF_K^{-1}(F_K(\hat{\boldsymbol{x}})) \right)^T \right)_i \cdot \nabla_{\hat{\boldsymbol{x}}} \hat{v}_i(\hat{\boldsymbol{x}}) \right] \hat{q}(\hat{\boldsymbol{x}}) \left|\det DF_K(\hat{\boldsymbol{x}})\right| \, d\hat{\boldsymbol{x}} \\
&= \int_{\hat{K}} \left[\left(DF_K^{-1}(F_K(\hat{\boldsymbol{x}})) \right)^T : D_{\hat{\boldsymbol{x}}} \hat{\boldsymbol{v}}(\hat{\boldsymbol{x}}) \right] \hat{q}(\hat{\boldsymbol{x}}) \left|\det DF_K(\hat{\boldsymbol{x}})\right| \, d\hat{\boldsymbol{x}}. \qquad \text{(B.31)}
\end{aligned}$$

In the derivation, (B.27) was used. □

Example B.62. Affine transform. The most important class of reference maps are affine transforms (B.18), where the invertible matrix B_K and the vector

$\boldsymbol{b}$ are constants. It follows that

$$\hat{\boldsymbol{x}} = B_K^{-1}(\boldsymbol{x} - \boldsymbol{b}) = B_K^{-1}\boldsymbol{x} - B_K^{-1}\boldsymbol{b}.$$

In this case, there are

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

One obtains for the integral transforms from (B.26), (B.29), (B.30), and (B.31)

$$\int_K v(\boldsymbol{x})\ d\boldsymbol{x} = |\det(B_K)| \int_{\hat{K}} \hat{v}(\hat{\boldsymbol{x}})\ d\hat{\boldsymbol{x}}, \tag{B.32}$$

$$\int_K \boldsymbol{b}(\boldsymbol{x}) \cdot \nabla v(\boldsymbol{x})\ d\boldsymbol{x} = |\det(B_K)| \int_{\hat{K}} \boldsymbol{b}(F_K(\hat{\boldsymbol{x}})) \cdot B_K^{-T} \nabla_{\hat{\boldsymbol{x}}} \hat{v}(\hat{\boldsymbol{x}})\ d\hat{\boldsymbol{x}}, \tag{B.33}$$

$$\int_K \nabla v(\boldsymbol{x}) \cdot \nabla w(\boldsymbol{x})\ d\boldsymbol{x} = |\det(B_K)| \int_{\hat{K}} B_K^{-T} \nabla_{\hat{\boldsymbol{x}}} \hat{v}(\hat{\boldsymbol{x}}) \cdot B_K^{-T} \nabla_{\hat{\boldsymbol{x}}} \hat{w}(\hat{\boldsymbol{x}})\ d\hat{\boldsymbol{x}}, \tag{B.34}$$

$$\int_K \nabla \cdot v(\boldsymbol{x}) q(\boldsymbol{x})\ d\boldsymbol{x} = |\det(B_K)| \int_{\hat{K}} \left[B_K^{-T} : D_{\hat{\boldsymbol{x}}} \hat{\boldsymbol{v}}(\hat{\boldsymbol{x}})\right] \hat{q}(\hat{\boldsymbol{x}})\ d\hat{\boldsymbol{x}}. \tag{B.35}$$

Setting $v(\boldsymbol{x}) = 1$ in (B.32) yields

$$|\det(B_K)| = \frac{|K|}{\left|\hat{K}\right|}. \tag{B.36}$$

□