Part II, Chapter 5

Main ideas and definitions

The goal of the three chapters composing Part II is to introduce the main concepts behind finite elements and to present various examples. This chapter introduces key notions such as degrees of freedom, shape functions, and interpolation operator. These notions are illustrated on Lagrange finite elements and modal finite elements, for which the degrees of freedom are values at specific nodes and moments against specific test functions, respectively.

5.1 Introductory example

This section introduces the notion of finite element in dimension one. Let $K := [-1, 1]$ and consider a continuous function $v \in C^0(K)$. Our objective is to devise an interpolation operator that approximates v in a finite-dimensional functional space, say P. For simplicity, we assume that $P = \mathbb{P}_k$ for some integer $k \geq 0$, where $\mathbb{P}_k$ is the real vector space composed of univariate polynomial functions of degree at most k, i.e., $p \in \mathbb{P}_k$ if $p(t) = \sum_{i\in\{0:k\}} \alpha_i t^i$ for all $t \in \mathbb{R}$, with $\alpha_i \in \mathbb{R}$ for every integer $i \in \{0:k\}$.

Let us consider $(k+1)$ distinct points $\{a_i\}_{i\in\{0:k\}}$ in K, which we call *nodes*. We want to construct an operator $\mathcal{I}_K : C^0(K) \to \mathbb{P}_k$ s.t. $\mathcal{I}_K(v)$ verifies

$$\mathcal{I}_K(v) \in \mathbb{P}_k, \qquad \mathcal{I}_K(v)(a_i) := v(a_i),\ \forall i \in \{0:k\}, \tag{5.1}$$

for every function $v \in C^0(K)$. These conditions uniquely determine $\mathcal{I}_K(v)$ since a polynomial in $\mathbb{P}_k$ is uniquely determined by the value it takes at $(k+1)$ distinct points in $\mathbb{R}$. For the same reason $\mathbb{P}_k$ is pointwise invariant under $\mathcal{I}_K$, i.e., $\mathcal{I}_K(p) = p$ for all $p \in \mathbb{P}_k$. To obtain an explicit representation of $\mathcal{I}_K(v)$, we introduce the Lagrange interpolation polynomials defined as follows:

$$\mathcal{L}_i^{[a]}(t) := \frac{\prod_{j\in\{0:k\}\setminus\{i\}}(t - a_j)}{\prod_{j\in\{0:k\}\setminus\{i\}}(a_i - a_j)}, \qquad \forall t \in \mathbb{R}, \qquad \forall i \in \{0:k\}. \tag{5.2}$$

We set $\mathcal{L}_0^{[a]} := 1$ if $k = 0$. By construction, the Lagrange interpolation polynomials satisfy $\mathcal{L}_i^{[a]}(a_i) = 1$ and $\mathcal{L}_i^{[a]}(a_j) = 0$ for all $j \neq i$, which we write concisely as

$$\mathcal{L}_i^{[a]}(a_j) = \delta_{ij}, \qquad \forall i, j \in \{0{:}k\}, \tag{5.3}$$

where δ_{ij} is the Kronecker symbol, i.e., $\delta_{ij} := 1$ if $i = j$ and $\delta_{ij} := 0$ otherwise. The Lagrange interpolation polynomials of degree $k \in \{1, 2, 3\}$ using equidistant nodes in K (including both endpoints) are shown in Figure 5.1. Let us show that the family $\{\mathcal{L}_i^{[a]}\}_{i\in\{0:k\}}$ forms a basis of $\mathbb{P}_k$. Since $\dim(\mathbb{P}_k) = k+1$, we only need to show linear independence. Assume that $\sum_{i\in\{0:k\}} \alpha_i \mathcal{L}_i^{[a]} = 0$. Evaluating this linear combination at the nodes $\{a_i\}_{i\in\{0:k\}}$ yields $\alpha_i = 0$ for all $i \in \{1{:}d\}$, which proves the assertion. In conclusion, the polynomial function $\mathcal{I}_K(v)$ defined in (5.1) is $\mathcal{I}_K(v)(t) := \sum_{i\in\{0:k\}} v(a_i)\mathcal{L}_i^{[a]}(t)$.

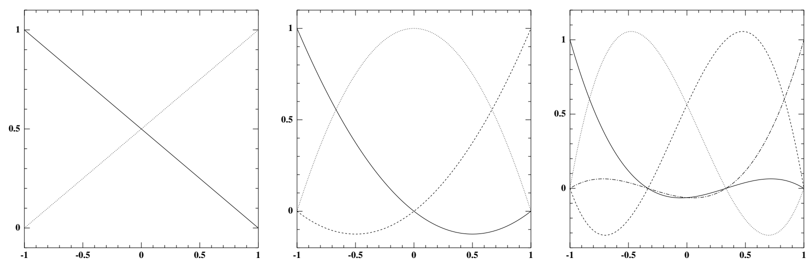

Fig. 5.1 Lagrange interpolation polynomials with equidistant nodes in the interval $K := [-1, 1]$ of degree $k = 1$ (left), 2 (center), and 3 (right).

Remark 5.1 (Key concepts). To sum up, we used three important ingredients to build the interpolation operator $\mathcal{I}_K$: the interval $K := [-1, 1]$, the finite-dimensional space $P := \mathbb{P}_k$, and a set of degrees of freedom, i.e., linear maps $\{\sigma_i\}_{i\in\{0:k\}}$ acting on continuous functions, which consist of evaluations at the nodes $\{a_i\}_{i\in\{0:k\}}$, i.e., $\sigma_i(v) := v(a_i)$. A key observation concerning the degrees of freedom is that they uniquely determine functions in P. □

5.2 Finite element as a triple

A *polyhedron* (also called *polytope*) in $\mathbb{R}^d$ is a compact interval if $d = 1$ and if $d \geq 2$, it is a compact, connected subset of $\mathbb{R}^d$ with nonempty interior such that its boundary is a finite union of images by affine mappings of polyhedra in $\mathbb{R}^{d-1}$. In $\mathbb{R}^2$, a polyhedron is also called polygon. Simple examples are presented in Figure 5.2 in dimensions two and three. A polyhedron in $\mathbb{R}^2$ (resp., $\mathbb{R}^3$) can always be described as a finite union of triangles (resp., tetrahedra).

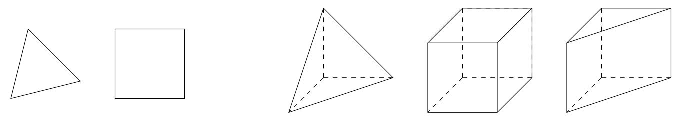

Fig. 5.2 Examples of polyhedra in $\mathbb{R}^2$ and $\mathbb{R}^3$. The hidden edges are shown with dashed lines in $\mathbb{R}^3$. From left to right: triangle, square, tetrahedron, hexahedron, prism.

The following definition of a finite element is due to Ciarlet [76, p. 93].

Definition 5.2 (Finite element). *Let $d \geq 1$, an integer $n_{\mathrm{sh}} \geq 1$, and the set $\mathcal{N} := \{1{:}n_{\mathrm{sh}}\}$. A* finite element *consists of a triple (K, P, Σ) where:*

(i) *K is a* polyhedron *in $\mathbb{R}^d$ or the image of a polyhedron in $\mathbb{R}^d$ by some smooth diffeomorphism. More generally, K could be the closure of a Lipschitz domain in $\mathbb{R}^d$ (see §3.1). K is nontrivial, i.e.,* $\mathrm{int}(K) \neq \emptyset$.

(ii) *P is a* finite-dimensional vector space *of functions $p : K \to \mathbb{R}^q$ for some integer $q \geq 1$ (typically $q \in \{1, d\}$). P is nontrivial, i.e., $P \neq \{0\}$. The members of P are polynomial functions, possibly composed with some smooth diffeomorphism.*

(iii) *Σ is a* set of n_{sh} linear forms *from P to $\mathbb{R}$, say $\Sigma := \{\sigma_i\}_{i\in\mathcal{N}}$, such that the map $\Phi_\Sigma : P \to \mathbb{R}^{n_{\mathrm{sh}}}$ defined by $\Phi_\Sigma(p) := \big(\sigma_i(p)\big)_{i\in\mathcal{N}}$ is an isomorphism. The linear forms σ_i are called* degrees of freedom *(in short* dofs*), and the bijectivity of the map Φ_Σ is referred to as* unisolvence.

Remark 5.3 (Proving unisolvence). To prove unisolvence, it suffices to show that $\dim P \geq n_{\mathrm{sh}} = \mathrm{card}\,\Sigma$ and that Φ_Σ is injective, i.e.,

$$[\,\sigma_i(p) = 0,\ \forall i \in \mathcal{N}\,] \implies [\,p = 0\,], \qquad \forall p \in P. \tag{5.4}$$

Owing to the rank nullity theorem, Φ_Σ is then bijective and $\dim P = n_{\mathrm{sh}}$. □

Remark 5.4 ($\mathcal{L}(P;\mathbb{R})$). Σ is a basis of the space of the linear forms over P, i.e., $\mathcal{L}(P;\mathbb{R})$. Indeed, $\dim(\mathcal{L}(P;\mathbb{R})) = \dim(P) = n_{\mathrm{sh}}$. Moreover, if the vector $X = (X_i)_{i\in\mathcal{N}} \in \mathbb{R}^{n_{\mathrm{sh}}}$ is s.t. $\sum_{i\in\mathcal{N}} X_i\sigma_i(p) = 0$ for all $p \in P$, taking $p := \Phi_\Sigma^{-1}(X)$ yields $\sum_{i\in\mathcal{N}} X_i^2 = 0$. Hence, $X_i = 0$ for all $i \in \mathcal{N}$. □

Proposition 5.5 (Shape functions). (i) *There is a basis $\{\theta_i\}_{i\in\mathcal{N}}$ of P s.t.*

$$\sigma_i(\theta_j) := \delta_{ij}, \qquad \forall i, j \in \mathcal{N}. \tag{5.5}$$

The functions θ_i are called shape functions. (ii) *Let $\{\phi_i\}_{i\in\mathcal{N}}$ be a basis of P. Then defining the* generalized Vandermonde matrix *$\mathcal{V} \in \mathbb{R}^{n_{\mathrm{sh}}\times n_{\mathrm{sh}}}$ with entries $\mathcal{V}_{ij} := \sigma_j(\phi_i)$ for all $i, j \in \mathcal{N}$, the shape functions are given by*

$$\theta_i = \sum_{j\in\mathcal{N}} (\mathcal{V}^{-1})_{ij}\phi_j, \qquad \forall i \in \mathcal{N}. \tag{5.6}$$

Proof. (i) The shape functions are given by $\theta_i = \Phi_\Sigma^{-1}(e_i)$ for all $i \in \mathcal{N}$, where $(e_i)_{i\in\mathcal{N}}$ is the canonical basis of $\mathbb{R}^{n_{\rm sh}}$. (ii) To show that the matrix $\mathcal{V}$ is invertible, we consider $X \in \mathbb{R}^{n_{\rm sh}}$ s.t. $X^{\mathsf{T}}\mathcal{V} = 0$ and set $p := \sum_{i\in\mathcal{N}} X_i\phi_i$. Then $X^{\mathsf{T}}\mathcal{V} = 0$ implies that $\sigma_j(p) = 0$ for all $j \in \mathcal{N}$, and (5.4) in turn implies that $p = 0$. Hence, $X = 0$ since $\{\phi_i\}_{i\in\mathcal{N}}$ is a basis of P. Finally, $\sigma_k(\sum_{j\in\mathcal{N}}(\mathcal{V}^{-1})_{ij}\phi_j) = \sum_{j\in\mathcal{N}}(\mathcal{V}^{-1})_{ij}\sigma_k(\phi_j) = \sum_{j\in\mathcal{N}}(\mathcal{V}^{-1})_{ij}\mathcal{V}_{jk} = \delta_{ik}$ for all $k \in \mathcal{N}$. This proves that $\theta_i = \sum_{j\in\mathcal{N}}(\mathcal{V}^{-1})_{ij}\phi_j$. □

Proposition 5.5 gives a practical recipe to build the shape functions. One first chooses a basis of P and evaluates the associated Vandermonde matrix $\mathcal{V}$ and its inverse. The components of the shape function θ_i in the chosen basis are then $((\mathcal{V}^{-1})_{ij})_{j\in\mathcal{N}}$ for all $i \in \mathcal{N}$. One must be careful in choosing the basis $\{\phi_i\}_{i\in\mathcal{N}}$ when working with high-order polynomials, since the matrix $\mathcal{V}$ may become ill-conditioned if the basis is not chosen properly. The computation of the shape functions can be affected by roundoff errors if $\mathcal{V}$ is ill-conditioned.

Remark 5.6 (Vandermonde matrix). For $d = 1$, if one uses the monomial basis $\{x^i\}_{i\in\mathcal{N}}$ with the dofs $\sigma_i(p) := p(a_i)$, then $\mathcal{V}$ is a classical Vandermonde matrix with entries $\mathcal{V}_{ij} = a_j^i$ for all $i, j \in \mathcal{N}$. □

5.3 Interpolation: finite element as a quadruple

The notion of interpolation operator is central to the finite element theory. The term "interpolation" is used here in a broad sense, since the degrees of freedom (dofs) are not necessarily point evaluations. For the interpolation operator to be useful, one needs to extend the domain of the linear forms in Σ so that they can act on functions in a space larger than P, which we denote by $V(K)$. The space $V(K)$ is the fourth ingredient defining a finite element.

Definition 5.7 (Interpolation operator). *Let (K, P, Σ) be a finite element. Assume that there exists a Banach space $V(K) \subset L^1(K;\mathbb{R}^q)$ s.t.:*

(i) *$P \subset V(K)$.*
(ii) *The linear forms $\{\sigma_i\}_{i\in\mathcal{N}}$ can be extended to $\mathcal{L}(V(K);\mathbb{R})$, i.e., there exist $\{\tilde\sigma_i\}_{i\in\mathcal{N}}$ and c_Σ such that $\tilde\sigma_i(p) = \sigma_i(p)$ for all $p \in P$, and $|\tilde\sigma_i(v)| \le c_\Sigma\|v\|_{V(K)}$ for all $v \in V(K)$ and all $i \in \mathcal{N}$. We henceforth abuse the notation and use the symbol σ_i instead of $\tilde\sigma_i$.*

We define the interpolation operator $\mathcal{I}_K : V(K) \to P$ *by setting*

$$\mathcal{I}_K(v)(\boldsymbol{x}) := \sum_{i\in\mathcal{N}} \sigma_i(v)\theta_i(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in K, \tag{5.7}$$

for all $v \in V(K)$. $V(K)$ is the domain of $\mathcal{I}_K$, and P is its codomain.

Proposition 5.8 (Boundedness). *$\mathcal{I}_K$ belongs to $\mathcal{L}(V(K);P)$.*

Proof. Let $\|\cdot\|_P$ be a norm in P (all the norms are equivalent in the finite-dimensional space P). The triangle inequality and Definition 5.7(ii) imply that $\|\mathcal{I}_K(v)\|_P \le (c_\Sigma \sum_{i\in\mathcal{N}} \|\theta_i\|_P)\|v\|_{V(K)}$ for all $v \in V(K)$. □

Proposition 5.9 (P-invariance). *P is pointwise invariant under $\mathcal{I}_K$, i.e., $\mathcal{I}_K(p) = p$ for all $p \in P$. As a result, $\mathcal{I}_K$ is a projection, i.e., $\mathcal{I}_K \circ \mathcal{I}_K = \mathcal{I}_K$.*

Proof. Letting $p = \sum_{j\in\mathcal{N}} \alpha_j\theta_j$ yields $\mathcal{I}_K(p) = \sum_{i,j\in\mathcal{N}} \alpha_j\sigma_i(\theta_j)\theta_i = p$ owing to (5.5). This shows that P is pointwise invariant under $\mathcal{I}_K$, and it immediately follows that $\mathcal{I}_K$ is a projection. □

Example 5.10 ($V(K)$). If one builds $\mathcal{I}_K(v)$ by using values of the function v at some points in K, like we did in §5.1, then it is natural to set $V(K) := C^0(K;\mathbb{R}^q)$ (recall that K is a closed set in $\mathbb{R}^d$, so that functions in $C^0(K;\mathbb{R}^q)$ are continuous up to the boundary). Another possibility is to set $V(K) := W^{s,p}(K;\mathbb{R}^q)$ for some real numbers $s \ge 0$ and $p \in [1,\infty]$ such that $sp > d$ (or $s \ge d$ if $p = 1$); see Theorem 2.31. If $\mathcal{I}_K(v)$ involves integrals over the faces of K, then one can take $V(K) := W^{s,p}(K;\mathbb{R}^q)$ with $sp > 1$ (or $s \ge 1$ if $p = 1$). More generally, if $\mathcal{I}_K(v)$ involves integrals over manifolds of codimension d', then it is legitimate to set $V(K) := W^{s,p}(K;\mathbb{R}^q)$ with $sp > d'$ (or $s \ge d'$ if $p = 1$). We abuse the notation since we should write $W^{s,p}(\mathrm{int}(K);\mathbb{R}^q)$, where $\mathrm{int}(K)$ denotes the interior of the set K in $\mathbb{R}^d$. □

5.4 Basic examples

5.4.1 Lagrange (nodal) finite elements

The dofs of scalar-valued Lagrange (or nodal) finite elements are point values. The extension to vector-valued Lagrange elements is done by proceeding componentwise.

Definition 5.11 (Lagrange finite element). *Let (K, P, Σ) be a scalar-valued finite element ($q := 1$ in Definition 5.2). If there is a set of points $\{\boldsymbol{a}_i\}_{i\in\mathcal{N}}$ in K such that for all $i \in \mathcal{N}$,*

$$\sigma_i(p) := p(\boldsymbol{a}_i), \qquad \forall p \in P, \tag{5.8}$$

the triple (K, P, Σ) is called Lagrange finite element*. The points $\{\boldsymbol{a}_i\}_{i\in\mathcal{N}}$ are called* nodes*, and the shape functions $\{\theta_i\}_{i\in\mathcal{N}}$, which are s.t.*

$$\theta_i(\boldsymbol{a}_j) := \delta_{ij}, \qquad \forall i, j \in \mathcal{N}, \tag{5.9}$$

form the nodal basis *of P associated with the nodes $\{\boldsymbol{a}_i\}_{i\in\mathcal{N}}$.*

Examples are presented in Chapters 6 and 7. Following Definition 5.7, the *Lagrange interpolation operator* $\mathcal{I}_K^{\mathrm{L}}$ acts as follows:

$$\mathcal{I}_K^{\mathrm{L}}(v)(\boldsymbol{x}) := \sum_{i\in\mathcal{N}} v(\boldsymbol{a}_i)\theta_i(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in K. \tag{5.10}$$

By construction, $\mathcal{I}_K^{\mathrm{L}}(v)$ matches the values of v at all the Lagrange nodes, i.e., $\mathcal{I}_K^{\mathrm{L}}(v)(\boldsymbol{a}_j) = v(\boldsymbol{a}_j)$ for all $j \in \mathcal{N}$. The domain of $\mathcal{I}_K^{\mathrm{L}}$ can be $V(K) := C^0(K)$ or $V(K) := W^{s,p}(K)$ with $p \in [1,\infty]$ and $ps > d$ ($s \geq d$ if $p = 1$).

5.4.2 Modal finite elements

The dofs of modal finite elements are moments against test functions using some measure over K. For simplicity, we consider the uniform measure and work in $L^2(K;\mathbb{R}^q)$ with $q \geq 1$. We are going to use the notation $(v,w)_{L^2(K;\mathbb{R}^q)} := \int_K (v,w)_{\ell^2(\mathbb{R}^q)}\,\mathrm{d}x$.

Proposition 5.12 (Modal finite element). *Let K be as in Definition* 5.2. *Let P be a finite-dimensional subspace of $L^2(K;\mathbb{R}^q)$ and let $\{\zeta_i\}_{i\in\mathcal{N}}$ be a basis of P. Let $\Sigma := \{\sigma_i\}_{i\in\mathcal{N}}$ be composed of the following linear forms $\sigma_i : P \to \mathbb{R}$:*

$$\sigma_i(p) := |K|^{-1}(\zeta_i, p)_{L^2(K;\mathbb{R}^q)}, \qquad \forall p \in P,\ \forall i \in \mathcal{N}. \tag{5.11}$$

(The factor $|K|^{-1}$ is meant to make σ_i independent of the size of K.) Then the triple (K,P,Σ) is a finite element called modal finite element.

Proof. We use Remark 5.3. By definition, $\dim(P) = \mathrm{card}(\Sigma)$. Let $p \in P$ be such that $\sigma_i(p) = 0$ for all $i \in \mathcal{N}$. Writing $p = \sum_{i\in\mathcal{N}} \alpha_i\zeta_i$, we infer that $|K|^{-1}\|p\|^2_{L^2(K;\mathbb{R}^q)} = \sum_{j\in\mathcal{N}} \alpha_j\sigma_j(p) = 0$, so that $p = 0$. □

Examples of modal finite elements are presented in Chapter 6. Let us introduce the *mass matrix* $\mathcal{M}$ of order n_{sh} with entries

$$\mathcal{M}_{ij} := |K|^{-1}(\zeta_i,\zeta_j)_{L^2(K;\mathbb{R}^q)}, \qquad \forall i,j \in \mathcal{N}. \tag{5.12}$$

By construction, $\mathcal{M}$ is symmetric, and since

$$(\mathcal{M}X, X)_{\ell^2(\mathbb{R}^{n_{\mathrm{sh}}})} = \sum_{i,j\in\mathcal{N}} \mathcal{M}_{ij}X_iX_j = |K|^{-1}\|\xi\|^2_{L^2(K;\mathbb{R}^q)},$$

for all $X \in \mathbb{R}^{n_{\mathrm{sh}}}$ with $\xi = \sum_{j\in\mathcal{N}} X_j\zeta_j$, we infer that $(\mathcal{M}X,X)_{\ell^2(\mathbb{R}^{n_{\mathrm{sh}}})} \geq 0$. Moreover, $(\mathcal{M}X,X)_{\ell^2(\mathbb{R}^{n_{\mathrm{sh}}})} = 0$ implies $\xi = 0$, i.e., $X = 0$ since $\{\zeta_i\}_{i\in\mathcal{N}}$ is a basis of P. In conclusion, $\mathcal{M}$ is symmetric positive definite. Furthermore, one readily sees that $\mathcal{M} = \mathcal{V}$, where the Vandermonde matrix $\mathcal{V}$ is defined in Proposition 5.5. Hence, $\theta_i = \sum_{j\in\mathcal{N}} (\mathcal{M}^{-1})_{ij}\zeta_j$ for all $i \in \mathcal{N}$. Following Definition 5.7, the *modal interpolation operator* $\mathcal{I}_K^{\mathrm{m}}$ acts as follows:

$$\mathcal{I}_K^{\mathrm{m}}(v)(\boldsymbol{x}) := \sum_{i\in\mathcal{N}} \left(\frac{1}{|K|}(\zeta_i, v)_{L^2(K;\mathbb{R}^q)}\right)\theta_i(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in K. \tag{5.13}$$

The domain of $\mathcal{I}_K^{\mathrm{m}}$ can be defined to be $V(K) := L^2(K;\mathbb{R}^q)$, or even $V(K) := L^1(K;\mathbb{R}^q)$ if $P \subset L^\infty(K;\mathbb{R}^q)$. One can verify that $\mathcal{I}_K^{\mathrm{m}}$ is the L^2-orthogonal projection onto P; see Exercise 5.2. Finally, if the basis $\{\zeta_i\}_{i\in\mathcal{N}}$ is L^2-orthogonal, the mass matrix is diagonal, and in that case the shape functions are given by $\theta_i := (|K|/\|\zeta_i\|^2_{L^2(K;\mathbb{R}^q)})\zeta_i$ for all $i \in \mathcal{N}$.

5.5 The Lebesgue constant

Recall from Definition 5.7 that the interpolation operator $\mathcal{I}_K$ is in $\mathcal{L}(V(K);P)$. Since $P \subset V(K)$, we can equip P with the norm of $V(K)$ and view $\mathcal{I}_K$ a member of $\mathcal{L}(V(K))$. In this section, we study the quantity

$$\|\mathcal{I}_K\|_{\mathcal{L}(V(K))} := \sup_{v\in V(K)} \frac{\|\mathcal{I}_K(v)\|_{V(K)}}{\|v\|_{V(K)}}, \tag{5.14}$$

which is called the *Lebesgue constant* for $\mathcal{I}_K$. We abuse the notation by writing the supremum over $v \in V(K)$ instead of $v \in V(K)\setminus\{0\}$.

Lemma 5.13 (Lower bound). $\|\mathcal{I}_K\|_{\mathcal{L}(V(K))} \geq 1$.

Proof. Since P is nontrivial (i.e., $P \neq \{0\}$) and since $\mathcal{I}_K(p) = p$ for all $p \in P$ owing to Proposition 5.9, we infer that

$$\sup_{v\in V(K)} \frac{\|\mathcal{I}_K(v)\|_{V(K)}}{\|v\|_{V(K)}} \geq \sup_{p\in P} \frac{\|\mathcal{I}_K(p)\|_{V(K)}}{\|p\|_{V(K)}} = 1. \qquad \square$$

The Lebesgue constant arises naturally in the estimate of the interpolation error in terms of the *best-approximation error* of a function $v \in V(K)$ by a function in P, that is, $\inf_{p\in P}\|v-p\|_{V(K)}$. In particular, the next result shows that a large value of the Lebesgue constant is associated with poor approximation properties of $\mathcal{I}_K$.

Theorem 5.14 (Interpolation error). *For all $v \in V(K)$, we have*

$$\|v - \mathcal{I}_K(v)\|_{V(K)} \leq (1 + \|\mathcal{I}_K\|_{\mathcal{L}(V(K))}) \inf_{p\in P} \|v-p\|_{V(K)}, \tag{5.15}$$

and $\|v-\mathcal{I}_K(v)\|_{V(K)} \leq \|\mathcal{I}_K\|_{\mathcal{L}(V(K))} \inf_{p\in P} \|v-p\|_{V(K)}$ *if $V(K)$ is a Hilbert space.*

Proof. Since $\mathcal{I}_K(p) = p$ for all $p \in P$, we infer that $v - \mathcal{I}_K(v) = (I - \mathcal{I}_K)(v) = (I - \mathcal{I}_K)(v-p)$, where I is the identity operator in $V(K)$, so that

$$\|v-\mathcal{I}_K(v)\|_{V(K)} \leq \|(I-\mathcal{I}_K)(v-p)\|_{V(K)} \leq (1 + \|\mathcal{I}_K\|_{\mathcal{L}(V(K))})\|v-p\|_{V(K)},$$

where we used the triangle inequality. We obtain (5.15) by taking the infimum over $p \in P$. Assume now that $V(K)$ is a Hilbert space. We use the fact that in any Hilbert space H, any operator $T \in \mathcal{L}(H)$ such that $0 \neq T \circ T = T \neq I$ satisfies $\|T\|_{\mathcal{L}(H)} = \|I - T\|_{\mathcal{L}(H)}$; see Kato [124], Xu and Zikatanov [201, Lem. 5], Szyld [188]. We can apply this result with $H := V(K)$ and $T := \mathcal{I}_K$. Indeed, $\mathcal{I}_K \neq 0$ since P is nontrivial, $\mathcal{I}_K \neq I$ since P is a proper subset of $V(K)$, and $\mathcal{I}_K \circ \mathcal{I}_K = \mathcal{I}_K$ owing to Proposition 5.9. We infer that

$$\|v - \mathcal{I}_K(v)\|_{V(K)} \leq \|I - \mathcal{I}_K\|_{\mathcal{L}(V(K))}\|v - p\|_{V(K)} = \|\mathcal{I}_K\|_{\mathcal{L}(V(K))}\|v - p\|_{V(K)},$$

and we conclude by taking the infimum over $p \in P$. □

Example 5.15 (Lagrange elements). The Lebesgue constant for the Lagrange interpolation operator $\mathcal{I}_K^{\mathrm{L}}$ with nodes $\{\boldsymbol{a}_i\}_{i\in\mathcal{N}}$ and space $V(K) := C^0(K)$ is denoted by $\Lambda^{\mathcal{N}} := \|\mathcal{I}_K^{\mathrm{L}}\|_{\mathcal{L}(C^0(K))}$. Owing to Theorem 5.14, we have $\|v - \mathcal{I}_K^{\mathrm{L}}(v)\|_{C^0(K)} \leq (1 + \Lambda^{\mathcal{N}}) \inf_{p\in P} \|v - p\|_{C^0(K)}$. One can verify (see Exercise 5.6) that $\Lambda^{\mathcal{N}} = \|\lambda^{\mathcal{N}}\|_{C^0(K)}$ with the Lebesgue function $\lambda^{\mathcal{N}}(\boldsymbol{x}) := \sum_{i\in\mathcal{N}} |\theta_i(\boldsymbol{x})|$ for all $\boldsymbol{x} \in K$. □

Example 5.16 (Modal elements). Consider a modal finite element with $V(K) := L^2(K;\mathbb{R}^q)$ (see Proposition 5.12). Since $\mathcal{I}_K^{\mathrm{m}}$ is the L^2-orthogonal projection from $L^2(K;\mathbb{R}^q)$ onto P, the Pythagorean identity $\|v\|^2_{L^2(K;\mathbb{R}^q)} = \|\mathcal{I}_K^{\mathrm{m}}(v)\|^2_{L^2(K;\mathbb{R}^q)} + \|v - \mathcal{I}_K^{\mathrm{m}}(v)\|^2_{L^2(K;\mathbb{R}^q)}$ implies that $\|\mathcal{I}_K^{\mathrm{m}}\|_{\mathcal{L}(L^2(K;\mathbb{R}^q))} \leq 1$, which in turn gives $\|\mathcal{I}_K^{\mathrm{m}}\|_{\mathcal{L}(L^2(K;\mathbb{R}^q))} = 1$ owing to Lemma 5.13. □

Let assume that $V(K)$ is a Hilbert space with inner product $(\cdot,\cdot)_{V(K)}$. Following ideas developed in Maday et al. [137], we now show that the Lebesgue constant can be related to the stability of an oblique projection. Owing to Theorem A.16 (or Exercise 5.9), we introduce the functions $q_i \in V(K)$ for all $i \in \mathcal{N}$ s.t. $(q_i, v)_{V(K)} = \sigma_i(v)$ for all $v \in V(K)$. Let us set $Q := \mathrm{span}\{q_i\}_{i\in\mathcal{N}}$, and let $Q^\perp$ be the orthogonal to Q in $V(K)$ for the inner product $(\cdot,\cdot)_{V(K)}$.

Lemma 5.17 (Oblique projection). *Let $\mathcal{I}_K$ be defined in (5.7). Then $\mathcal{I}_K$ is the oblique projection onto P along $Q^\perp$, and the Lebesgue constant is $\|\mathcal{I}_K\|_{\mathcal{L}(V(K))} = \alpha_{PQ}^{-1}$ with $\alpha_{PQ} := \inf_{p\in P} \sup_{q\in Q} \frac{(p,q)_{V(K)}}{\|p\|_{V(K)}\|q\|_{V(K)}}$.*

Proof. (1) Unisolvence implies that $P \cap Q^\perp = \{0\}$. Indeed, if $p \in P \cap Q^\perp$, then $p \in P$ and $\sigma_i(p) = 0$ for all $i \in \mathcal{N}$, so that $p = 0$. Let now $v \in V(K)$. We observe that $\mathcal{I}_K(v) \in P$ and

$$(q_i, \mathcal{I}_K(v) - v)_{V(K)} = \sigma_i(\mathcal{I}_K(v)) - \sigma_i(v) = 0, \qquad \forall i \in \mathcal{N}.$$

Hence, $\mathcal{I}_K(v) - v \in Q^\perp$. From the decomposition $v = \mathcal{I}_K(v) + (v - \mathcal{I}_K(v))$, we infer that $V(K) = P + Q^\perp$. Therefore, the sum is direct, and $\mathcal{I}_K(v)$ is the oblique projection of v onto P along $Q^\perp$.
(2) We have

$$\alpha_{PQ}\|\mathcal{I}_K(v)\|_{V(K)} \leq \sup_{q\in Q} \frac{(\mathcal{I}_K(v), q)_{V(K)}}{\|q\|_{V(K)}} = \sup_{q\in Q} \frac{(v, q)_{V(K)}}{\|q\|_{V(K)}} \leq \|v\|_{V(K)},$$

for all $v \in V(K)$, showing that $\|\mathcal{I}_K\|_{\mathcal{L}(V(K))} \leq \alpha_{PQ}^{-1}$. To prove the lower bound, let us first show that $\mathcal{I}_K(\Pi_Q(p)) = p$ for all $p \in P$, where Π_Q is the $V(K)$-orthogonal projection onto Q. We first observe that

$$(\mathcal{I}_K(\Pi_Q(p)), q)_{V(K)} = (\Pi_Q(p), q)_{V(K)} = (p, q)_{V(K)},$$

for all $q \in Q$, where we used the fact that both $\mathcal{I}_K$ and Π_Q are projections along $Q^\perp$. The above identity implies that $\mathcal{I}_K(\Pi_Q(p)) - p \in P \cap Q^\perp = \{0\}$. Hence, $\mathcal{I}_K(\Pi_Q(p)) = p$. Since P is a finite-dimensional space, a compactness argument shows that there is $p^* \in P$ with $\|p^*\|_{V(K)} = 1$ such that $\alpha_{PQ} = \sup_{q\in Q} \frac{(p^*,q)_{V(K)}}{\|q\|_{V(K)}}$. Since $(p^*, q)_{V(K)} = (\Pi_Q(p^*), q)_{V(K)}$, we infer that $\alpha_{PQ} = \sup_{q\in Q} \frac{(\Pi_Q(p^*),q)_{V(K)}}{\|q\|_{V(K)}} = \|\Pi_Q(p^*)\|_{V(K)}$. We conclude that

$$\|\mathcal{I}_K\|_{\mathcal{L}(V(K))} \geq \frac{\|\mathcal{I}_K(\Pi_Q(p^*))\|_{V(K)}}{\|\Pi_Q(p^*)\|_{V(K)}} = \frac{\|p^*\|_{V(K)}}{\|\Pi_Q(p^*)\|_{V(K)}} = \frac{1}{\alpha_{PQ}}. \qquad \square$$

Further results on the Lebesgue constant for one-dimensional Lagrange elements can be found in §6.3.1.

Exercises

Exercise 5.1 (Linear combination). Let $\mathcal{S} \in \mathbb{R}^{n_{\rm sh}\times n_{\rm sh}}$ be an invertible matrix. Let (K, P, Σ) be a finite element. Let $\tilde{\Sigma} := \{\tilde{\sigma}_i\}_{i\in\mathcal{N}}$ with dofs $\tilde{\sigma}_i := \sum_{i'\in\mathcal{N}} \mathcal{S}_{ii'}\sigma_{i'}$ for all $i \in \mathcal{N}$. Prove that $(K, P, \tilde{\Sigma})$ is a finite element. Write the shape functions $\{\tilde{\theta}_j\}_{j\in\mathcal{N}}$ and verify that the interpolation operator does not depend on $\mathcal{S}$, i.e., $\tilde{\mathcal{I}}_K(v)(\boldsymbol{x}) = \mathcal{I}_K(v)(\boldsymbol{x})$ for all $v \in V(K)$ and all $\boldsymbol{x} \in K$.

Exercise 5.2 (Modal finite element). (i) Let (K, P, Σ) and $(K, P, \widetilde{\Sigma})$ be two modal finite elements. Let $\{\zeta_i\}_{i\in\mathcal{N}}$, $\{\tilde{\zeta}_i\}_{i\in\mathcal{N}}$, be the two bases of P s.t. the dofs in Σ and $\widetilde{\Sigma}$ are given by $\sigma_i(p) := |K|^{-1}(\zeta_i, p)_{L^2(K;\mathbb{R}^q)}$ and $\tilde{\sigma}_i(p) := |K|^{-1}(\tilde{\zeta}_i, p)_{L^2(K;\mathbb{R}^q)}$ for all $i \in \mathcal{N}$. Prove that the interpolation operators $\mathcal{I}_K^{\rm m}$ and $\widetilde{\mathcal{I}}_K^{\rm m}$ are identical. (ii) Prove that $(p, \mathcal{I}_K^{\rm m}(v) - v)_{L^2(K;\mathbb{R}^q)} = 0$ for all $p \in P$. (iii) Let $\mathcal{M}$ be defined by (5.12), and let $\mathcal{M}_{ij}^\theta := |K|^{-1}(\theta_i, \theta_j)_{L^2(K;\mathbb{R}^q)}$ for all $i, j \in \mathcal{N}$, where $\{\theta_i\}_{i\in\mathcal{N}}$ are the shape functions associated with (K, P, Σ). Prove that $\mathcal{M}^\theta = \mathcal{M}^{-1}$.

Exercise 5.3 (Variation on $\mathbb{P}_2$). Let $K := [0, 1]$, $P := \mathbb{P}_2$, and $\Sigma := \{\sigma_1, \sigma_2, \sigma_3\}$ be the linear forms on P s.t. $\sigma_1(p) := p(0)$, $\sigma_2(p) := 2p(\frac{1}{2}) - p(0) - p(1)$, $\sigma_3(p) := p(1)$ for all $p \in P$. Show that (K, P, Σ) is a finite element, compute the shape functions, and indicate possible choices for $V(K)$.

Exercise 5.4 (Hermite). Let $K := [0, 1]$, $P := \mathbb{P}_3$, and $\Sigma := \{\sigma_1, \sigma_2, \sigma_3, \sigma_4\}$ be the linear forms on P s.t. $\sigma_1(p) := p(0)$, $\sigma_2(p) := p'(0)$, $\sigma_3(p) := p(1)$,

$\sigma_4(p) := p'(1)$ for all $p \in P$. Show that (K, P, Σ) is a finite element, compute the shape functions, and indicate possible choices for $V(K)$.

Exercise 5.5 (Powell–Sabin). Consider $K := [0,1]$ and let P be composed of the functions that are piecewise quadratic over the intervals $[0, \frac{1}{2}] \cup [\frac{1}{2}, 1]$ and are of class C^1 over K, i.e., functions in P and their first derivatives are continuous. Let $\Sigma := \{\sigma_1, \ldots, \sigma_4\}$ be the linear forms on P s.t. $\sigma_1(p) := p(0)$, $\sigma_2(p) := p'(0)$, $\sigma_3(p) := p(1)$, $\sigma_4(p) := p'(1)$. Prove that the triple (K, P, Σ) is a finite element. Verify that the first two shape functions are

$$\theta_1(t) = \begin{cases} 1 - 2t^2 & \text{if } t \in [0, \frac{1}{2}], \\ 2(1-t)^2 & \text{if } t \in [\frac{1}{2}, 1], \end{cases} \qquad \theta_2(t) = \begin{cases} t(1 - \frac{3}{2}t) & \text{if } t \in [0, \frac{1}{2}], \\ \frac{1}{2}(1-t)^2 & \text{if } t \in [\frac{1}{2}, 1], \end{cases}$$

and compute the other two shape functions. *Note*: a two-dimensional version of this finite element on triangles has been developed in [161].

Exercise 5.6 (Lebesgue constant for Lagrange element). Prove that the Lebesgue constant $\Lambda^{\mathcal{N}}$ defined in Example 5.15 is equal to $\|\mathcal{I}_K^{\mathrm{L}}\|_{\mathcal{L}(C^0(K))}$. (*Hint*: to prove $\|\mathcal{I}_K^{\mathrm{L}}\|_{\mathcal{L}(C^0(K))} \geq \Lambda^{\mathcal{N}}$, consider functions $\{\psi_i\}_{i\in\mathcal{N}}$ taking values in $[0,1]$ s.t. $\sum_{i\in\mathcal{N}} \psi_i = 1$ in K and $\psi_i(\boldsymbol{a}_j) = \delta_{ij}$ for all $i, j \in \mathcal{N}$.)

Exercise 5.7 (Lagrange interpolation). Let $K := [a, b]$ and let $p \in [1, \infty)$. (i) Prove that $\|v\|_{L^\infty(K)} \leq (b-a)^{-\frac{1}{p}} \|v\|_{L^p(K)} + (b-a)^{1-\frac{1}{p}} \|v'\|_{L^p(K)}$ for all $v \in W^{1,p}(K)$ (*Hint*: use $v(x) - v(y) = \int_x^y v'(t)\,\mathrm{d}t$ for all $v \in C^1(K)$, where $|v(y)| := \min_{z\in K} |v(z)|$, then use the density of $C^1(K)$ in $W^{1,p}(K)$.) (ii) Prove that $W^{1,p}(K)$ embeds continuously in $C^0(K)$. (iii) Let $\mathcal{I}_K^{\mathrm{L}}$ be the interpolation operator based on the linear Lagrange finite element using the nodes a and b. Determine the two shape functions and prove that $\mathcal{I}_K^{\mathrm{L}}$ can be extended to $W^{1,p}(K)$. (iv) Assuming that $w \in W^{1,p}(K)$ is zero at some point in K, show that $\|w\|_{L^p(K)} \leq (b-a)\|w'\|_{L^p(K)}$. (v) Prove the following estimates: $\|(v - \mathcal{I}_K^{\mathrm{L}}(v))'\|_{L^p(K)} \leq (b-a)\|v''\|_{L^p(K)}$, $\|v - \mathcal{I}_K^{\mathrm{L}}(v)\|_{L^p(K)} \leq (b-a)\|(v - \mathcal{I}_K^{\mathrm{L}}(v))'\|_{L^p(K)}$, $\|(\mathcal{I}_K^{\mathrm{L}}(v))'\|_{L^p(K)} \leq \|v'\|_{L^p(K)}$, for all $p \in (1, \infty]$ and all $v \in W^{2,p}(K)$.

Exercise 5.8 (Cross approximation). Let X, Y be nonempty subsets of $\mathbb{R}$ and $f : X \times Y \to \mathbb{R}$ be a bivariate function. Let $\mathcal{N} := \{1{:}n_{\mathrm{sh}}\}$ with $n_{\mathrm{sh}} \geq 1$, and consider n_{sh} points $\{x_i\}_{i\in\mathcal{N}}$ in X and n_{sh} points $\{y_j\}_{j\in\mathcal{N}}$ in Y. Assume that the matrix $\mathcal{F} \in \mathbb{R}^{n_{\mathrm{sh}} \times n_{\mathrm{sh}}}$ with entries $\mathcal{F}_{ij} := f(x_i, y_j)$ is invertible. Let $\mathcal{I}^{\mathrm{CA}}(f) : X \times Y \to \mathbb{R}$ be s.t. $\mathcal{I}^{\mathrm{CA}}(f)(x, y) := \sum_{i,j\in\mathcal{N}} (\mathcal{F}^{-\mathsf{T}})_{ij} f(x, y_j) f(x_i, y)$. Prove that $\mathcal{I}^{\mathrm{CA}}(f)(x, y_k) = f(x, y_k)$ for all $x \in X$ and all $k \in \mathcal{N}$, and that $\mathcal{I}^{\mathrm{CA}}(f)(x_k, y) = f(x_k, y)$ for all $y \in Y$ and all $k \in \mathcal{N}$.

Exercise 5.9 (Riesz–Fréchet in finite dimension). Let V be a finite-dimensional complex Hilbert space. Show that for every antilinear form $A \in V'$, there is a unique $v \in V$ s.t. $(v, w)_V = \langle A, w \rangle_{V',V}$ for all $w \in V$, with $\|v\|_V = \|A\|_{V'}$.