Part III, Chapter 13

Local interpolation on non-affine meshes

In this chapter we extend the results of Chapter 11 to non-affine meshes. For simplicity, the transformation ψ_K is the pullback by the geometric mapping, but this mapping is now nonaffine. The first difficulty consists of proving a counterpart of Lemma 11.7 to compare Sobolev norms. This is not a trivial task since the chain rule involves higher-order derivatives of the geometric mapping. The second difficulty is to define a notion of shape regularity for mesh sequences built using nonaffine geometric mappings. We show how to do this using a perturbation theory, and we present various examples.

13.1 Introductory example on curved simplices

If one wants to solve a problem in a domain D with a curved boundary ∂D, it may be useful to use non-affine cells, since they approximate ∂D better than affine cells. A relatively straightforward way to achieve this goal is as follows: (i) Construct a mesh composed of affine cells with all the vertices lying on the curved boundary ∂D. (ii) For each affine cell $\widetilde{K}$ having a non-empty intersection with ∂D, design a polynomial mapping (of degree larger than 1) that approximates the boundary more accurately than the first-order interpolation. Then replace $\widetilde{K}$ in the mesh by the resulting cell

Example 13.1 (Simple construction). An example relying on $\mathbb{P}_{2,S}$ or $\mathbb{Q}_{2,2}$ Lagrange elements in $\mathbb{R}^2$ (see Figure 13.1) is as follows:

(i) Let $\widetilde{K}$ be a triangle or a quadrangle having an edge whose vertices lie on ∂D. Let $\{\widetilde{\boldsymbol{a}}_1, \ldots, \widetilde{\boldsymbol{a}}_{n_{\rm geo}}\}$ be the geometric nodes of $\widetilde{K}$ ($n_{\rm geo} = 6$ for a triangle and $n_{\rm geo} = 9$ for a quadrangle).

(ii) For each $\widetilde{\boldsymbol{a}}_i$, $i \in \{1{:}n_{\rm geo}\}$, construct a new node $\boldsymbol{a}_i$ as follows: If $\widetilde{\boldsymbol{a}}_i$ is located at the middle of an edge whose vertices lie on ∂D, $\boldsymbol{a}_i$ is defined as the intersection with ∂D of the line normal to the corresponding edge and passing through the node $\widetilde{\boldsymbol{a}}_i$. Otherwise, set $\boldsymbol{a}_i := \widetilde{\boldsymbol{a}}_i$.

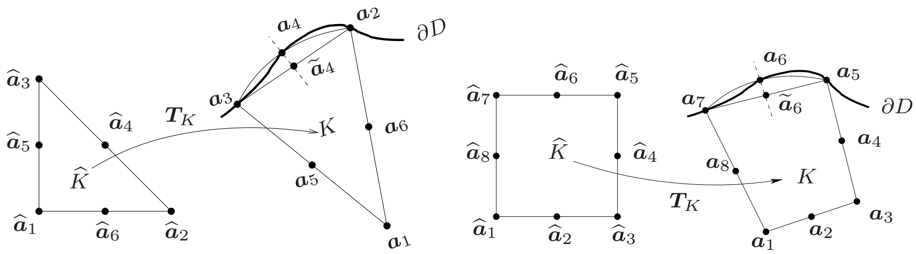

Fig. 13.1 Construction of a curved triangle (left) and a curved quadrangle (right).

(iii) Replace $\widetilde{K}$ by the curved triangle K by defining the mapping $\boldsymbol{T}_K : \widehat{K} \to K$ such that $\boldsymbol{T}_K(\widehat{\boldsymbol{x}}) := \sum_{i\in\{1:n_{\text{geo}}\}} \widehat{\psi}_i(\widehat{\boldsymbol{x}})\boldsymbol{a}_i$ for all $\widehat{\boldsymbol{x}} \in \widehat{K}$, where $\{\widehat{\psi}_i\}_{i\in\{1:n_{\text{geo}}\}}$ are the reference $\mathbb{P}_2$ or $\mathbb{Q}_2$ Lagrange shape functions. $\square$

13.2 A perturbation theory

This section presents a perturbation theory introduced by Ciarlet and Raviart [73] to analyze the finite element interpolation error on non-affine cells.

13.2.1 Motivation and notation

Let $(\widehat{K}, \widehat{P}_{n_{\text{geo}}}, \widehat{\Sigma}_{n_{\text{geo}}})$ be a reference geometric Lagrange finite element with nodes $\{\widehat{\boldsymbol{a}}_1, \ldots, \widehat{\boldsymbol{a}}_{n_{\text{geo}}}\}$ and shape functions $\{\widehat{\psi}_1, \ldots, \widehat{\psi}_{n_{\text{geo}}}\}$. Let us now consider two sets of points in $\mathbb{R}^d$,

$$\{\widetilde{\boldsymbol{a}}_1, \ldots, \widetilde{\boldsymbol{a}}_{n_{\text{geo}}}\}, \qquad \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n_{\text{geo}}}\}. \tag{13.1}$$

Let $\widetilde{\boldsymbol{T}} : \widehat{K} \to \mathbb{R}^d$ and $\boldsymbol{T} : \widehat{K} \to \mathbb{R}^d$ be the mappings defined as follows:

$$\widetilde{\boldsymbol{T}}(\widehat{\boldsymbol{x}}) := \sum_{i=1}^{n_{\text{geo}}} \widehat{\psi}_i(\widehat{\boldsymbol{x}})\widetilde{\boldsymbol{a}}_i, \qquad \boldsymbol{T}(\widehat{\boldsymbol{x}}) := \widetilde{\boldsymbol{T}}(\widehat{\boldsymbol{x}}) + \sum_{i=1}^{n_{\text{geo}}} \widehat{\psi}_i(\widehat{\boldsymbol{x}})(\boldsymbol{a}_i - \widetilde{\boldsymbol{a}}_i). \tag{13.2}$$

Let us set $\widetilde{K} := \widetilde{\boldsymbol{T}}(\widehat{K})$ and $K := \boldsymbol{T}(\widehat{K})$. The subscripts K and $\widetilde{K}$ are omitted in the rest of this section to simplify the notation. Assuming that $\widetilde{\boldsymbol{T}} : \widehat{K} \to \widetilde{K}$ is a reasonable diffeomorphism, we want to ascertain that $\boldsymbol{T} : \widehat{K} \to K$ is also a diffeomorphism with reasonable smoothness properties by making sure that K is close to $\widetilde{K}$.

The (Fréchet) derivatives of $\boldsymbol{T}$ and $\widetilde{\boldsymbol{T}}$ of order m at a point $\widehat{\boldsymbol{x}} \in \widehat{K}$ are denoted $D^m\boldsymbol{T}(\widehat{\boldsymbol{x}})$ and $D^m\widetilde{\boldsymbol{T}}(\widehat{\boldsymbol{x}})$, respectively (the superscript is omitted for $m = 1$). Recall from Appendix B that $D^m\boldsymbol{T}$ and $D^m\widetilde{\boldsymbol{T}}$ are members of $\mathcal{M}_m(\mathbb{R}^d, \ldots, \mathbb{R}^d; \mathbb{R}^d)$, i.e., they are multilinear maps from $\mathbb{R}^d \times \ldots \times \mathbb{R}^d$ to $\mathbb{R}^d$

(i.e., $D\boldsymbol{T}$ and $D\widetilde{\boldsymbol{T}}$ are linear maps in $\mathcal{L}(\mathbb{R}^d;\mathbb{R}^d)$, $D^2\boldsymbol{T}$ and $D^2\widetilde{\boldsymbol{T}}$ are bilinear maps in $\mathcal{M}_2(\mathbb{R}^d,\mathbb{R}^d;\mathbb{R}^d)$, etc.). We adopt the following notation: For every map $A \in C^m(\widehat{K};\mathbb{R}^q)$, every integer $q \geq 1$, and all $\widehat{\boldsymbol{x}} \in \widehat{K}$,

$$|D^m A(\widehat{\boldsymbol{x}})|_{\mathbb{P}} := \max_{\boldsymbol{h}_1,\ldots,\boldsymbol{h}_m \in \mathbb{R}^d} \frac{\|D^m A(\widehat{\boldsymbol{x}})(\boldsymbol{h}_1,\ldots,\boldsymbol{h}_m)\|_{\ell^2(\mathbb{R}^q)}}{\|\boldsymbol{h}_1\|_{\ell^2(\mathbb{R}^d)} \ldots \|\boldsymbol{h}_m\|_{\ell^2(\mathbb{R}^d)}}. \tag{13.3}$$

The notation for the subscript is motivated by the fact that for all $k \geq 0$, $|D^{k+1}A(\widehat{\boldsymbol{x}})|_{\mathbb{P}} = 0$ if and only if A is $[\mathbb{P}_k]^q$-valued. Note that the right-hand side of (13.3) is the canonical norm in $\mathcal{M}_m(\mathbb{R}^d,\ldots,\mathbb{R}^d;\mathbb{R}^q)$.

13.2.2 Bounds on the first-order derivatives

Our goal is to bound the first-order derivatives of the geometric mapping. We first identify a condition ensuring that $\boldsymbol{T}$ is a C^1-diffeomorphism with reasonable bounds on $D\boldsymbol{T}$, $D(\boldsymbol{T}^{-1})$, and $\det(D\boldsymbol{T})$. For the (Fréchet) derivatives, we use the norm $\|D^m A\| := \|\,|D^m A(\widehat{\boldsymbol{x}})|_{\mathbb{P}}\|_{L^\infty(\widehat{K})}$.

Lemma 13.2 (Bound on $D\boldsymbol{T}$, $D(\boldsymbol{T}^{-1})$). *Let $\widetilde{\boldsymbol{T}},\boldsymbol{T}$ be defined in* (13.2). *Assume that $\widetilde{\boldsymbol{T}}$ is a C^1-diffeomorphism, $\widehat{P}_{n_{\text{geo}}} \subset C^1(\widehat{K},\mathbb{R})$, and there is $c_1 \in \mathbb{R}$, $c_1 < 1$, s.t.*

$$\|(D\widetilde{\boldsymbol{T}})^{-1}\| \sum_{i=1}^{n_{\text{geo}}} \|D\widehat{\psi}_i\| \|\boldsymbol{a}_i - \widetilde{\boldsymbol{a}}_i\|_{\ell^2(\mathbb{R}^d)} \leq c_1. \tag{13.4}$$

Then $\boldsymbol{T}$ is a C^1-diffeomorphism and

$$\|D\boldsymbol{T}\| \leq (1+c_1)\|D\widetilde{\boldsymbol{T}}\|, \tag{13.5}$$

$$\|D(\boldsymbol{T}^{-1})\| \leq (1-c_1)^{-1}\|(D\widetilde{\boldsymbol{T}})^{-1}\| \tag{13.6}$$

$$(1-c_1)^d|\det(D\widetilde{\boldsymbol{T}}(\widehat{\boldsymbol{x}}))| \leq |\det(D\boldsymbol{T}(\widehat{\boldsymbol{x}}))| \leq (1+c_1)^d|\det(D\widetilde{\boldsymbol{T}}(\widehat{\boldsymbol{x}}))|, \ \forall \widehat{\boldsymbol{x}} \in \widehat{K}. \tag{13.7}$$

Proof. This is Theorem 3 in [73]. The definition of $\boldsymbol{T}$ in (13.2) implies that $D\boldsymbol{T}(\widehat{\boldsymbol{x}}) = D\widetilde{\boldsymbol{T}}(\widehat{\boldsymbol{x}}) + \boldsymbol{E}(\widehat{\boldsymbol{x}}) = D\widetilde{\boldsymbol{T}}(\widehat{\boldsymbol{x}})(\boldsymbol{I} + (D\widetilde{\boldsymbol{T}})^{-1}(\widehat{\boldsymbol{x}})\boldsymbol{E}(\widehat{\boldsymbol{x}}))$ where $\boldsymbol{E}(\widehat{\boldsymbol{x}})(\boldsymbol{\xi}) = \sum_{i=1}^{n_{\text{geo}}} D\widehat{\psi}_i(\widehat{\boldsymbol{x}})(\boldsymbol{\xi})(\boldsymbol{a}_i - \widetilde{\boldsymbol{a}}_i)$. The assumption (13.4) says that $\|(D\widetilde{\boldsymbol{T}})^{-1}\boldsymbol{E}\| \leq c_1 < 1$. This immediately implies that $\boldsymbol{I} + (D\widetilde{\boldsymbol{T}})^{-1}(\widehat{\boldsymbol{x}})\boldsymbol{E}(\widehat{\boldsymbol{x}})$ is invertible, i.e., $D\boldsymbol{T}(\widehat{\boldsymbol{x}})$ is invertible and

$$\|D\boldsymbol{T}\| = \|D\widetilde{\boldsymbol{T}}(\boldsymbol{I} + (D\widetilde{\boldsymbol{T}})^{-1}\boldsymbol{E})\| \leq (1+c_1)\|D\widetilde{\boldsymbol{T}}\|$$
$$\|(D\boldsymbol{T})^{-1}\| = \|(\boldsymbol{I} + (D\widetilde{\boldsymbol{T}})^{-1}\boldsymbol{E})^{-1}(D\widetilde{\boldsymbol{T}})^{-1}\| \leq (1-c_1)^{-1}\|(D\widetilde{\boldsymbol{T}})^{-1}\|.$$

The proof of the inequalities on the determinants is left as an exercise. □

Remark 13.3 (Regularity of $\widehat{P}_{n_{\rm geo}}$). In practice, the assumption on the smoothness of $\widehat{P}_{n_{\rm geo}}$ is satisfied since this space is composed of smooth (polynomial) functions; see (8.1). $\square$

13.2.3 Bounds on the higher-order derivatives

Lemma 13.4 (Higher-order derivatives). *Assume* (13.4). *Assume that there is an integer $k \geq 1$ s.t. $\widetilde{\boldsymbol{T}}$ is C^{k+1}-diffeomorphism, $\widehat{P}_{n_{\rm geo}} \in C^{k+1}(\widehat{K}, \mathbb{R})$, and there are real numbers $c_2, \ldots, c_{k+1}$ s.t.*

$$\|D^m \boldsymbol{T}\| \leq c_m \|D\widetilde{\boldsymbol{T}}\|, \qquad \forall m \in \{2{:}k+1\}. \tag{13.8}$$

Then $\boldsymbol{T}$ is a C^{k+1}-diffeomorphism, and letting $\kappa := \|D\widetilde{\boldsymbol{T}}\|\|D(\widetilde{\boldsymbol{T}}^{-1})\|$ and c_1 be defined in (13.4)*, for every integer $m \in \{2{:}k+1\}$, there is c_{-m} depending on $c_1, \ldots, c_m$ s.t.*

$$\|D^m(\boldsymbol{T}^{-1})\| \leq c_{-m}\kappa^{m-1}\|D(\widetilde{\boldsymbol{T}}^{-1})\|^m. \tag{13.9}$$

Proof. This is Theorem 4 in [73]. The assumption $\widehat{P}_{n_{\rm geo}} \in C^{k+1}(\widehat{K}, \mathbb{R})$ implies that $\boldsymbol{T}$ is of class C^{k+1}, and it has already been established in Lemma 13.2 that $\boldsymbol{T}$ is a diffeomorphism. Let us prove (13.9) for $m = 2$. Using the chain rule (see Lemma B.4) and the identity $\boldsymbol{T}^{-1}(\boldsymbol{T}(\widehat{\boldsymbol{x}})) = \widehat{\boldsymbol{x}}$, we infer that

$$D^2(\boldsymbol{T}^{-1})(\boldsymbol{h}_1, \boldsymbol{h}_2) = -D(\boldsymbol{T}^{-1})(D^2\boldsymbol{T}((D\boldsymbol{T})^{-1}\boldsymbol{h}_1, (D\boldsymbol{T})^{-1}\boldsymbol{h}_2)),$$

for all $\boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathbb{R}^d$. Using that $\|D(\boldsymbol{T}^{-1})\| = \|(D\boldsymbol{T})^{-1}\|$, this implies that

$$\|D^2(\boldsymbol{T}^{-1})\| \leq \|D(\boldsymbol{T}^{-1})\|\|D^2\boldsymbol{T}\|\|(D\boldsymbol{T})^{-1}\|^2 = \|D^2\boldsymbol{T}\|\|D(\boldsymbol{T}^{-1})\|^3.$$

Owing to (13.6) and (13.8), we infer that $\|D^2(\boldsymbol{T}^{-1})\| \leq c_{-2}\kappa\|D(\widetilde{\boldsymbol{T}}^{-1})\|^2$ with $c_{-2} = c_2(1-c_1)^{-3}$. The rest of the proof is left as an exercise. $\square$

13.2.4 Interpolation error analysis

We now establish the approximation properties of the finite element defined in Proposition 9.2 using $\psi_K(v) = v \circ \boldsymbol{T}$, with $\boldsymbol{T}$ defined in (13.2). We first state a result similar to that in Lemma 11.7 on the comparison of Sobolev norms. We distinguish two cases for higher-order derivatives depending on whether mixed derivatives are involved or not. In particular, for every map $A \in C^m(\widehat{K}; \mathbb{R}^q)$ and every integer $q \geq 1$, we define the following seminorms:

$$|D^m A(\widehat{\boldsymbol{x}})|_{\mathbb{Q}} := \max_{\boldsymbol{e}\in\{\boldsymbol{e}_1,\ldots,\boldsymbol{e}_d\}} \|D^m A(\widehat{\boldsymbol{x}})(\boldsymbol{e}, \ldots, \boldsymbol{e})\|_{\ell^2(\mathbb{R}^q)}, \tag{13.10}$$

where $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$ is the canonical Cartesian basis of $\mathbb{R}^d$. The notation for the subscript is motivated by the fact that for all $k \geq 0$, $|D^{k+1}A(\widehat{\boldsymbol{x}})|_{\mathbb{Q}} = 0$ if and only if A is $[\mathbb{Q}_{k,d}]^q$-valued (see Exercise 13.3). Observe that (13.10) only defines a seminorm and that $|D^m A(\widehat{\boldsymbol{x}})|_{\mathbb{Q}} \le |D^m A(\widehat{\boldsymbol{x}})|_{\mathbb{P}}$. We define the seminorms $[\![D^m A]\!] := \| \, |D^m A(\widehat{\boldsymbol{x}})|_{\mathbb{Q}} \|_{L^\infty(\widehat{K})}$ and

$$[\![A]\!]_{W^{l,p}(\widehat{K})} := \| \, |D^l A(\widehat{\boldsymbol{x}})|_{\mathbb{Q}} \|_{L^p(\widehat{K})}. \tag{13.11}$$

This seminorm involves only the pure partial derivatives of A of order m, i.e., the mixed derivatives are not involved. Recall that the usual seminorm in $W^{m,p}(\widehat{K})$ involves all the derivatives, $|A|_{W^{m,p}(\widehat{K})} = \left(\sum_{|\alpha|=m} \|\partial^\alpha A\|^p_{L^p(\widehat{K})}\right)^{\frac{1}{p}}$, and that this seminorm is equivalent to $\| \, |D^m A(\widehat{\boldsymbol{x}})|_{\mathbb{P}} \|_{L^p(\widehat{K})}$.

Lemma 13.5 (Norm scaling by pullback). *Let $\widetilde{\boldsymbol{T}}, \boldsymbol{T}$ be defined in* (13.2). *Let the integer $k \ge 1$ satisfy the assumptions of Lemma* 13.4. *Assume that there are constants $c'_2, \ldots, c'_{k+1}$ s.t.*

$$\text{either} \quad \|D^m \boldsymbol{T}\| \le c'_m \, \|D\widetilde{\boldsymbol{T}}\|^m, \quad \forall m \in \{2{:}\,k+1\}, \tag{13.12}$$

$$\text{or} \quad [\![D^m \boldsymbol{T}]\!] \le c'_m \, \|D\widetilde{\boldsymbol{T}}\|^m, \quad \forall m \in \{2{:}\,k+1\}. \tag{13.13}$$

Then for all $l \ge 0$ and all $p \in [1,\infty]$, there is c, depending only on κ, $c_1, \ldots, c_{k+1}$, $c'_2, \ldots, c'_{k+1}$, p, and $\widehat{K}$, s.t. the following holds true for all $w \in W^{l,p}(K)$ with $K := \boldsymbol{T}(\widehat{K})$:

$$\text{either} \quad |w \circ \boldsymbol{T}|_{W^{l,p}(\widehat{K})} \le c \, \|\det(D\widetilde{\boldsymbol{T}})^{-1}\|^{\frac{1}{p}}_{L^\infty(\widehat{K})} \|D\widetilde{\boldsymbol{T}}\|^l \|w\|_{W^{l,p}(K)}, \tag{13.14}$$

$$\text{or} \quad [\![w \circ \boldsymbol{T}]\!]_{W^{l,p}(\widehat{K})} \le c \, \|\det(D\widetilde{\boldsymbol{T}})^{-1}\|^{\frac{1}{p}}_{L^\infty(\widehat{K})} \|D\widetilde{\boldsymbol{T}}\|^l \|w\|_{W^{l,p}(K)}, \tag{13.15}$$

and

$$|w|_{W^{l,p}(K)} \le c \, \|\det(D\widetilde{\boldsymbol{T}})\|^{\frac{1}{p}}_{L^\infty(\widehat{K})} \|D(\widetilde{\boldsymbol{T}}^{-1})\|^l \|w \circ \boldsymbol{T}\|_{W^{l,p}(\widehat{K})}. \tag{13.16}$$

Proof. Proof of (13.14). Assume first that $l \ge 2$. Using the chain rule (see Lemma B.4) together with the assumption (13.12), we infer that

$$\begin{aligned} |D^l(w \circ \boldsymbol{T})(\widehat{\boldsymbol{x}})|_{\mathbb{P}} &\le c \sum_{m=1}^{l} |(D^m w)(\boldsymbol{T}(\widehat{\boldsymbol{x}}))|_{\mathbb{P}} \sum_{|r|=l} |D^{r_1}\boldsymbol{T}(\widehat{\boldsymbol{x}})|_{\mathbb{P}} \ldots |D^{r_m}\boldsymbol{T}(\widehat{\boldsymbol{x}})|_{\mathbb{P}} \\ &\le c \, \|D\widetilde{\boldsymbol{T}}\|^l \sum_{m=1}^{l} |(D^m w)(\boldsymbol{T}(\widehat{\boldsymbol{x}}))|_{\mathbb{P}}, \end{aligned}$$

for all $\widehat{\boldsymbol{x}} \in \widehat{K}$, with $|r| = r_1 + \ldots + r_m$ and a generic constant c having the same dependencies as in the assertion. Then we have

$$\int_{\widehat{K}} |D^l(w \circ \boldsymbol{T})(\widehat{\boldsymbol{x}})|^p_{\mathbb{P}} \, \mathrm{d}\widehat{x} \le c \|D\widetilde{\boldsymbol{T}}\|^{pl} \int_K |D^m w(\boldsymbol{x})|^p_{\mathbb{P}} |\det(D\boldsymbol{T}^{-1}(\boldsymbol{x}))| \, \mathrm{d}x,$$

and we conclude using the estimate (13.7) on the determinant. The proof for $l = 0$ is evident. The proof for $l = 1$ can be done as above by using (13.5) instead of (13.12). The proof of the second estimate is similar once one realizes that the chain rule preserves pure derivatives of $\boldsymbol{T}$, i.e.,

$$|D^l(w \circ \boldsymbol{T})(\widehat{\boldsymbol{x}})|_{\mathbb{Q}} \leq c \sum_{m=0}^{l} |(D^m w)(\boldsymbol{T}(\widehat{\boldsymbol{x}}))|_{\mathbb{P}} \sum_{|r|=l} |D^{r_1}\boldsymbol{T}(\widehat{\boldsymbol{x}})|_{\mathbb{Q}} \ldots |D^{r_m}\boldsymbol{T}(\widehat{\boldsymbol{x}})|_{\mathbb{Q}}.$$

The third estimate is derived similarly by using the bound (13.9). □

Remark 13.6 (Assumption (13.8)). It is necessary to include the assumption (13.8) from Lemma 13.4 in the assumptions of Lemma 13.5 only when invoking assumption (13.13) (pure-derivatives case). Actually, assumption (13.12) implies (13.8) (recall that since $\|D\widetilde{\boldsymbol{T}}\|$ is proportional to the diameter of the cell generated by the nodes $\{\widetilde{\boldsymbol{a}}_1, \ldots, \widetilde{\boldsymbol{a}}_{n_{\text{geo}}}\}$, it is reasonable to assume that $\|D\widetilde{\boldsymbol{T}}\| \leq 1$ if the diameter of the cell is small enough). □

We now formulate the Bramble–Hilbert lemma for polynomials in $\mathbb{Q}_{k,d}$, (this is the counterpart of Lemma 11.9 stated for polynomials in $\mathbb{P}_{k,d}$). It is at this stage that the seminorm based on pure derivatives enters the analysis.

Lemma 13.7 ($\mathbb{Q}_k$-Bramble–Hilbert). *Let S be a Lipschitz domain in $\mathbb{R}^d$. Let p be a real number such that $p \in [1, \infty]$. Let $k \geq 0$ be an integer. There is c s.t.*

$$\inf_{q \in \mathbb{Q}_{k,d}} \|v + q\|_{W^{k+1,p}(S)} \leq c\, [\![v]\!]_{W^{k+1,p}(S)}. \tag{13.17}$$

Corollary 13.8 ($\mathbb{Q}_k$-Bramble–Hilbert for linear functionals). *Under the hypotheses of Lemma 13.7, there is c s.t. the following holds true for all $g \in W^{k+1,p}(S)'$ vanishing on $\mathbb{Q}_{k,d}$:*

$$|g(v)| \leq c\, \|g\|_{(W^{k+1,p}(S))'} [\![v]\!]_{W^{k+1,p}(S)}, \qquad \forall v \in W^{k+1,p}(S). \tag{13.18}$$

Proof. The estimate (13.17) is proved in Bramble and Hilbert [39, Thm. 1] The estimate (13.18) is proved in [39, Thm. 2]. □

We are now in the position to present the main result of this section.

Theorem 13.9 (Local interpolation). *Let $\widetilde{\boldsymbol{T}}, \boldsymbol{T}$ be defined in* (13.2) *and let $K := \boldsymbol{T}(\widehat{K})$. Let $p \in [1, \infty]$. Let the integer $k \geq 1$ satisfy the assumptions of Lemma* 13.4 *and assume* (13.12)-(13.13) *(i.e., the hypotheses of Lemma* 13.5 *are met). Assume that*

$$\text{either} \quad (13.12) \text{ and } \mathbb{P}_{k,d} \subset \widehat{P} \subset W^{k+1,p}(\widehat{K}) \hookrightarrow V(\widehat{K}), \tag{13.19}$$

$$\text{or} \quad (13.13) \text{ and } \mathbb{Q}_{k,d} \subset \widehat{P} \subset W^{k+1,p}(\widehat{K}) \hookrightarrow V(\widehat{K}). \tag{13.20}$$

Let $\mathcal{I}_K$ be the interpolation operator defined in (9.6)*. Let $l \in \{1{:}k+1\}$ be an integer s.t. $W^{l,p}(\widehat{K}) \hookrightarrow V(\widehat{K})$. Let $\lambda := \|\det(D\widetilde{\boldsymbol{T}})\|_{L^\infty(\widehat{K})}\|\det(D\widetilde{\boldsymbol{T}})^{-1}\|_{L^\infty(\widehat{K})}$ and recall the notation $\kappa := \|D\widetilde{\boldsymbol{T}}\|\|D(\widetilde{\boldsymbol{T}}^{-1})\|$. There is c, only depending on κ, $c_1, \ldots, c_{k+1}$, $c'_2, \ldots, c'_{k+1}$, p, and $\widehat{K}$, s.t. for all $v \in W^{l,p}(K)$ and all $m \in \{0{:}l\}$,*

$$|v - \mathcal{I}_K v|_{W^{m,p}(K)} \leq c\,\lambda^{\frac{1}{p}}\kappa^m \|D\widetilde{\boldsymbol{T}}\|^{l-m}\,\|v\|_{W^{l,p}(K)}. \tag{13.21}$$

Proof. (1) Let us prove the statement assuming (13.19). Using (13.16) from Lemma 13.5 and the commutation property $(\mathcal{I}_K w) \circ \boldsymbol{T} = \mathcal{I}_{\widehat{K}}(w \circ \boldsymbol{T})$ (see Proposition 9.3), we infer that

$$|w - \mathcal{I}_K w|_{W^{m,p}(K)} \leq c\|\det(D\widetilde{\boldsymbol{T}})\|^{\frac{1}{p}}_{L^\infty(\widehat{K})}\|D(\widetilde{\boldsymbol{T}}^{-1})\|^m\|\widehat{w} - \mathcal{I}_{\widehat{K}}\widehat{w}\|_{W^{m,p}(\widehat{K})},$$

with $\widehat{w} = w \circ \boldsymbol{T}$. Just like in the proof of Theorem 11.13, the assumptions (13.19) imply that there is c s.t. $\|\widehat{w} - \mathcal{I}_{\widehat{K}}\widehat{w}\|_{W^{m,p}(\widehat{K})} \leq c\,|\widehat{w}|_{W^{l,p}(\widehat{K})}$. This, together with (13.14), proves the claim since

$$\begin{aligned}
|w - \mathcal{I}_K w|_{W^{m,p}(\widehat{K})} &\leq c\,\|\det(D\widetilde{\boldsymbol{T}})\|^{\frac{1}{p}}_{L^\infty(\widehat{K})}\|D(\widetilde{\boldsymbol{T}}^{-1})\|^m|\widehat{w}|_{W^{l,p}(\widehat{K})} \\
&\leq c\,\|\det(D\widetilde{\boldsymbol{T}})\|^{\frac{1}{p}}_{L^\infty(\widehat{K})}\|\det(D\widetilde{\boldsymbol{T}})^{-1}\|^{\frac{1}{p}}_{L^\infty(\widehat{K})}\|D(\widetilde{\boldsymbol{T}}^{-1})\|^m\|D\widetilde{\boldsymbol{T}}\|^l\|w\|_{W^{l,p}(K)}.
\end{aligned}$$

(2) The only change in the above argument when proving (13.21) assuming (13.20) is that $\|\widehat{w} - \mathcal{I}_{\widehat{K}}\widehat{w}\|_{W^{m,p}(\widehat{K})} \leq c\,[\![\widehat{w}]\!]_{W^{l,p}(\widehat{K})}$, owing to the (13.17) from the Bramble–Hilbert lemma (Lemma 13.7). We conclude using (13.15). □

Remark 13.10 (Key assumptions). In conclusion the key assumptions to be verified for Theorem 13.9 to hold are (13.4) and (13.12) for $\mathbb{P}_k$-based finite elements or (13.4), (13.8), and (13.13) for $\mathbb{Q}_k$-based finite elements. Of course, the above theory makes sense only for meshes for which the numbers κ, $c_1, \ldots, c_{k+1}$, $c'_2, \ldots, c'_{k+1}$ are uniformly bounded with respect to K. □

Remark 13.11 (Extensions). Generalizations of the above ideas can be found in Bernardi [20], Brenner and Scott [45, §4.7], Ciarlet [71, §4.3-4.4], Ciarlet [71], Lenoir [127], and Zlámal [194, 195]. □

13.3 Curved simplices

Let us now describe how the above technique can be applied with curved $\mathbb{P}_2$-simplices, i.e., we set $k = 2$. Let us assume for the time being that we have at hand a mesh $\widetilde{\mathcal{T}}_h$ composed of affine simplices, say $\widetilde{K} \in \mathcal{T}_h$. From Lemma 11.1, we know that

$$\frac{\rho_{\widetilde{K}}}{h_{\widehat{K}}} \le \|D\widetilde{\boldsymbol{T}}\| \le \frac{h_{\widetilde{K}}}{\rho_{\widehat{K}}}, \qquad \frac{\rho_{\widehat{K}}}{h_{\widetilde{K}}} \le \|D(\widetilde{\boldsymbol{T}})^{-1}\| \le \frac{h_{\widehat{K}}}{\rho_{\widetilde{K}}}. \tag{13.22}$$

Let us define $\rho_K := \rho_{\widetilde{K}}$ and $h_K := h_{\widetilde{K}}$. Assume that the sequence of meshes $\{\widetilde{\mathcal{T}}_h\}_{h\in\mathcal{H}}$ composed of the affine simplices $\widetilde{K}$ is shape regular (see Definition 11.2), i.e., $\sigma_{\widetilde{K}} = \frac{h_{\widetilde{K}}}{\rho_{\widetilde{K}}} \le \sigma_\sharp$ for all $\widetilde{K}$.

Let us consider one element $\widetilde{K} \in \widetilde{\mathcal{T}}_h$, and let $\{\widetilde{\boldsymbol{a}}_1, \ldots, \widetilde{\boldsymbol{a}}_{n_{\text{geo}}}\}$ be its geometric nodes. Assume now that by means of some algorithm we construct the points $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n_{\text{geo}}}\}$ from the set $\{\widetilde{\boldsymbol{a}}_1, \ldots, \widetilde{\boldsymbol{a}}_{n_{\text{geo}}}\}$ and define the corresponding cell K using (13.2); see Figure 13.1 in dimension $d = 2$ where $n_{\text{geo}} = 6$. Assume that there is a constant c_a, uniform with respect to the cell $\widetilde{K} \in \widetilde{\mathcal{T}}_h$ and $h \in \mathcal{H}$, such that the following holds true:

$$\max_{i=1\ldots n_{\text{geo}}} \|\boldsymbol{a}_i - \widetilde{\boldsymbol{a}}_i\|_{\ell^2(\mathbb{R}^d)} \le c_a h_K^2. \tag{13.23}$$

This assumption is reasonable if the midpoint on each edge is constructed as explained in Example 13.1.

The key assumptions to be verified for Theorem 13.9 to hold are (13.4) and (13.12). We first observe that the left-hand side of (13.4) can be bounded by $(c_a h_{\widehat{K}} \sum_{i\in\{1:n_{\text{geo}}\}} \|D\widehat{\psi}_i\|)\kappa_0 h_K$ which is less than 1 for h_K small enough, i.e., (13.4) holds true for h_K small enough. Using that $D^2\widetilde{\boldsymbol{T}} = 0$, we infer that

$$\begin{aligned} \|D^2\boldsymbol{T}\| &\le \Bigg(\sum_{i\in\{1:n_{\text{geo}}\}} \|D^2\widehat{\psi}_i\|\Bigg) \max_{i\in\{1:n_{\text{geo}}\}} \|\boldsymbol{a}_i - \widetilde{\boldsymbol{a}}_i\|_{\ell^2} \\ &\le \Bigg(c_a h_{\widehat{K}}^2 \sigma_\sharp^2 \sum_{i\in\{1:n_{\text{geo}}\}} \|D^2\widehat{\psi}_i\|\Bigg) \|D\widetilde{\boldsymbol{T}}\|^2. \end{aligned}$$

Notice also that $\|D^3\boldsymbol{T}\| = 0$. Hence (13.12) holds true for all $k \ge 1$. Moreover, since $\widetilde{\boldsymbol{T}}$ is affine, $\lambda = 1$ and $\kappa \le \sigma_\sharp \frac{h_{\widehat{K}}}{\rho_{\widehat{K}}}$. In conclusion, Theorem 13.9 implies that there is c s.t. for all K, all $h \in \mathcal{H}$ (small enough), all $v \in W^{l,p}(K)$, all $p \in [1,\infty]$, every integer $l \in \{0{:}k{+}1\}$ s.t. $W^{l,p}(K) \subset V(K)$, and every integer $m \in \{0{:}l\}$,

$$|v - \mathcal{I}_K v|_{W^{m,p}(K)} \le c\, \sigma_\sharp^m h_K^{l-m} \|v\|_{W^{l,p}(K)}. \tag{13.24}$$

Here $W^{l,p}(K)$ is in the domain of $\mathcal{I}_K$.

Remark 13.12 (Extensions). An algorithm that constructs $\mathbb{P}_3$-simplices in dimension two is described in Ciarlet and Raviart [73, p. 240], Ciarlet [72, §4.3], Ciarlet [71, p. 247]. An algorithm that constructs curved simplices of any order in any dimension and that satisfies the assumptions of the perturbation theory in §13.2 is described in Lenoir [127]. It is a recursive technique based on the following principle: the construction of curved $\mathbb{P}_{m+1}$ simplices that approximate the boundary with $\mathcal{O}(h^{m+2})$ accuracy relies on the existence of a construction technique of curved $\mathbb{P}_m$ simplices that approximate the boundary with $\mathcal{O}(h^{m+1})$ accuracy, $m \geq 1$. □

13.4 $\mathbb{Q}_1$-quadrangles

Let us now consider a mesh where all the cells are non-degenerate, convex quadrangles in $\mathbb{R}^2$. All the cells can be generated from the unit square $\widehat{K} = [0,1]^2$ using geometric mappings $\boldsymbol{T}_K \in [\mathbb{Q}_1(\widehat{K})]^2$; see Figure 13.2. We omit the subscript K in the rest of this section to simplify the notation. $\boldsymbol{T}$ maps the edges of $\widehat{K}$ to the edges of K, but unless K is a parallelogram, $\boldsymbol{T}$ is not affine. This means that the approximation Theorem 11.13 which is valid for affine cells only, cannot be applied here. We are going to apply instead the theory from §13.2 with $\widetilde{K} = K$, $\widetilde{\boldsymbol{T}} = \boldsymbol{T}$ and $\boldsymbol{a}_i = \widetilde{\boldsymbol{a}}_i$, $1 \leq i \leq 4$.

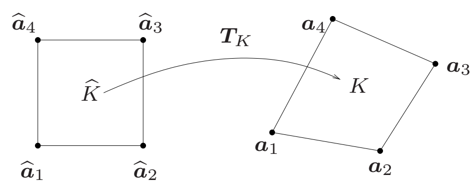

Fig. 13.2 Non-affine mapping from the unit square to a quadrangle.

Upon identifying the points $\boldsymbol{a}_{i\in\{1:4\}}$ with column vectors and $D\boldsymbol{T}$ with the Jacobian matrix, a simple computation shows that

$$\begin{aligned}
&D\boldsymbol{T}(\widehat{\boldsymbol{x}}) \\
&= (\boldsymbol{a}_2 - \boldsymbol{a}_1 + \widehat{x}_2(\boldsymbol{a}_3 - \boldsymbol{a}_4 + \boldsymbol{a}_1 - \boldsymbol{a}_2), \boldsymbol{a}_4 - \boldsymbol{a}_1 + \widehat{x}_1(\boldsymbol{a}_3 - \boldsymbol{a}_4 + \boldsymbol{a}_1 - \boldsymbol{a}_2)) \\
&= ((1-\widehat{x}_2)(\boldsymbol{a}_2 - \boldsymbol{a}_1) + \widehat{x}_2(\boldsymbol{a}_3 - \boldsymbol{a}_4), (1-\widehat{x}_1)(\boldsymbol{a}_4 - \boldsymbol{a}_1) + \widehat{x}_1(\boldsymbol{a}_3 - \boldsymbol{a}_2)),
\end{aligned}$$

where $\widehat{\boldsymbol{x}} := (\widehat{x}_1, \widehat{x}_2) \in \widehat{K}$. It follows from the first equality that $\det(D\boldsymbol{T}(\widehat{x}))$ is in $\mathbb{P}_1$, implying that $\max_{\widehat{\boldsymbol{x}}\in\widehat{K}} |\det(D\boldsymbol{T}(\widehat{x}))| = \max_{1\leq i\leq 4} |\det(D\boldsymbol{T}(\widehat{\boldsymbol{a}}_i))|$. Let P_i be the parallelogram formed by $\boldsymbol{a}_{i-1}$, $\boldsymbol{a}_i$, $\boldsymbol{a}_{i+1}$ (with the convention $\boldsymbol{a}_0 = \boldsymbol{a}_4$ and $\boldsymbol{a}_5 = \boldsymbol{a}_1$). It can be verified that $\det(D\boldsymbol{T}(\widehat{\boldsymbol{a}}_i)) = |P_i|$. As a result, letting $S_{\min} = \min_{1\leq i\leq 4} |P_i|$, $S_{\max} = \max_{1\leq i\leq 4} |P_i|$, we infer that

$$\|\det(D\boldsymbol{T})\|_{L^\infty(\widehat{K})} \leq S_{\max}, \qquad \|\det(D(\boldsymbol{T}^{-1}))\|_{L^\infty(\widehat{K})} \leq \frac{1}{S_{\min}}. \tag{13.25}$$

Let $\boldsymbol{c}_1(\widehat{\boldsymbol{x}})$ and $\boldsymbol{c}_2(\widehat{\boldsymbol{x}})$ be the columns of $D\boldsymbol{T}(\widehat{\boldsymbol{x}})$ and $\theta(\widehat{\boldsymbol{x}})$ be the angle formed by these two vectors. The vector $\boldsymbol{c}_1(\widehat{\boldsymbol{x}})$ is a convex combination of the sides $(\boldsymbol{a}_2 - \boldsymbol{a}_1)$ and $(\boldsymbol{a}_3 - \boldsymbol{a}_4)$, whereas the vector $\boldsymbol{c}_2(\widehat{\boldsymbol{x}})$ is a convex combination of the sides $(\boldsymbol{a}_4 - \boldsymbol{a}_1)$ and $(\boldsymbol{a}_3 - \boldsymbol{a}_2)$. The angle $\theta(\widehat{\boldsymbol{x}})$ takes its extreme values at the vertices of K, say $\theta_1, \ldots, \theta_4$. Let $\boldsymbol{h} = (h_1, h_2) \in \mathbb{R}^2$ with $\|\boldsymbol{h}\|_{\ell^2} = 1$. Then,

$$\begin{aligned}
\|D\boldsymbol{T}(\widehat{\boldsymbol{x}})(\boldsymbol{h})\|_{\ell^2}^2 &= h_1^2\|\boldsymbol{c}_1\|_{\ell^2}^2 + h_2^2\|\boldsymbol{c}_2\|_{\ell^2}^2 + 2h_1h_2\boldsymbol{c}_1\cdot\boldsymbol{c}_2 \\
&\geq h_1^2\|\boldsymbol{c}_1\|_{\ell^2}^2 + h_2^2\|\boldsymbol{c}_2\|_{\ell^2}^2 - 2|h_1||h_2|\|\boldsymbol{c}_1\|_{\ell^2}\|\boldsymbol{c}_2\|_{\ell^2}|\cos(\theta)| \\
&\geq h_1^2\|\boldsymbol{c}_1\|_{\ell^2}^2(1-|\cos(\theta)|) + (1-h_1^2)\|\boldsymbol{c}_2\|_{\ell^2}^2(1-|\cos(\theta)|) \\
&\geq \min(\|\boldsymbol{c}_1\|_{\ell^2}^2, \|\boldsymbol{c}_2\|_{\ell^2})^2)(1-|\cos(\theta)|),
\end{aligned}$$

where dependencies of $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, and θ on $\widehat{\boldsymbol{x}}$ have been omitted. Denoting by $h_{\min}$ the length of the smallest side of K and $\gamma = \max_{1\leq i\leq 4}|\cos(\theta_i)|$, we infer that $\|D\boldsymbol{T}(\widehat{\boldsymbol{x}})(\boldsymbol{y})\|_{\ell^2} \geq h_{\min}(1-\gamma)$ for all $\boldsymbol{y}$ with $\|\boldsymbol{y}\|_{\ell^2} = 1$ and all $\widehat{\boldsymbol{x}} \in \widehat{K}$, implying that $\|(D\boldsymbol{T})^{-1}\| = \|D(\boldsymbol{T}^{-1})\| \leq (h_{\min}(1-\gamma))^{-1}$. By proceeding similarly we also obtain that $\|D\boldsymbol{T}\| \leq 2h_{\max}$ and $[\![D^l\boldsymbol{T}]\!] = 0$ for $l \geq 2$, where $h_{\max}$ is the length of the largest side of K. In conclusion we have

$$\|D(\boldsymbol{T}^{-1})\| \leq \frac{1}{h_{\min}(1-\gamma)}, \quad \|D\boldsymbol{T}\| \leq 2h_{\max}, \quad [\![D^l\boldsymbol{T}]\!] = 0,\ \forall l \geq 2. \tag{13.26}$$

The key assumptions to be verified for Theorem 13.9 to hold are (13.4), (13.8), and (13.13). Assumption (13.4) is trivial since $\boldsymbol{T} = \widetilde{\boldsymbol{T}}$, and so is (13.8). Assumption (13.13) is evident from (13.26) since $[\![D^l\boldsymbol{T}]\!] = 0 \leq c'_l\|D\boldsymbol{T}\|^l$ for all $l \geq 2$. Furthermore, owing to (13.25), $\lambda \leq \frac{S_{\max}}{S_{\min}}$ and owing to (13.26), $\kappa \leq \frac{h_{\max}}{h_{\min}(1-\gamma)}$. In conclusion the bound (13.21) in Theorem 13.9 gives

$$|v - \mathcal{I}_K v|_{W^{m,p}(K)} \leq c\left(\frac{S_{\max}}{S_{\min}}\right)^{\frac{1}{p}}\left(\frac{h_{\max}}{h_{\min}(1-\gamma)}\right)^m h_K^{l-m}\|v\|_{W^{l,p}(K)}, \tag{13.27}$$

for all K, all $v \in W^{l,p}(K)$, all $p \in [1,\infty]$, and every integers m, l with $0 \leq m \leq l \leq k+1$ and $W^{l,p}(K) \subset V(K)$ (i.e., $W^{l,p}(K)$ is in the domain of $\mathcal{I}_K$).

Remark 13.13 (Shape regularity). Shape regularity can be defined for mesh sequences $(\mathcal{T}_h)_{h\in\mathcal{H}}$ composed of $\mathbb{Q}_1$-quadrangles by requiring that the ratios $\sigma_K := \max(\frac{S_{\max}}{S_{\min}}, \frac{h_{\max}}{h_{\min}(1-\gamma)})$ be bounded uniformly w.r.t. $h \in \mathcal{H}$. □

Remark 13.14 (Pure derivatives). The proof of the critical assumption (13.13) hinges on the property that $[\![D^2\boldsymbol{T}]\!] = 0$. This assumption would not have been true if we had used the full seminorm (involving the mixed derivative), since a simple computation shows that $\|D^2\boldsymbol{T}\| = \|(\boldsymbol{a}_3 - \boldsymbol{a}_4) + (\boldsymbol{a}_1 - \boldsymbol{a}_2)\|_{\ell^2}$, yielding $\sqrt{2(1+\cos(\theta_1+\theta_4))h_{\min}^2 + (h_{\max}-h_{\min})^2} \leq \|D^2\boldsymbol{T}\| \leq 2h_{\max}$, thereby showing that $\|D^2\boldsymbol{T}\| \sim \|D\boldsymbol{T}\|$ (unless K is a parallelogram) which would transform the term $h_{\max}^l$ in (13.27) with $m = 1$ into $h_{\max}^{l-1}$. The reader is referred to Ciarlet and Raviart [74, pp. 245-247] and Girault and Raviart [103, p. 104] for more details. □

13.5 $\mathbb{Q}_2$-curved quadrangles

We now describe how to construct $\mathbb{Q}_2$-curved quadrangles. Assume that we have at hand a sequence of meshes $(\widetilde{\mathcal{T}}_h)_{h\in\mathcal{H}}$, each composed of $\mathbb{Q}_1$-quadrangles, and that $\sigma_{\widetilde{K}} := \max(\frac{S_{\max}(\widetilde{K})}{S_{\min}(\widetilde{K})}, \frac{h_{\max}(\widetilde{K})}{h_{\min}(\widetilde{K})(1-\gamma(\widetilde{K}))})$ is uniformly bounded from above by a constant $\sigma_\sharp$ for all $\widetilde{K} \in \widetilde{\mathcal{T}}_h$ and all $h \in \mathcal{H}$, i.e., the sequence is shape regular in the sense defined in §13.4. We now omit the dependency with respect to $\widetilde{K}$ to simplify the notation, i.e., we use $S_{\min} = S_{\min}(\widetilde{K})$, $h_{\min} = h_{\min}(\widetilde{K})$, and so on. Let $\{\widetilde{\boldsymbol{a}}_1, \ldots, \widetilde{\boldsymbol{a}}_4\}$ be the vertices of the $\mathbb{Q}_1$-quadrangle $\widetilde{K}$ and let $\{\widetilde{\boldsymbol{a}}_5, \ldots, \widetilde{\boldsymbol{a}}_8\}$ be the midpoints of the edges; see Figure 13.1. Assume now that the new curved element is constructed by means of the technique explained in Example 13.1, i.e., we assume that the new points $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_8\}$ are positioned so that the following criterion is satisfied uniformly:

$$\max_{1\le i\le 8} \|\widetilde{\boldsymbol{a}}_i - \boldsymbol{a}_i\|_{\ell^2(\mathbb{R}^2)} \le c\, h_{\max}^2. \tag{13.28}$$

Let $\mathcal{T}_h$ be the mesh thus constructed. We now verify that the assumptions (13.4), (13.8), and (13.13) hold uniformly, and that the quantities $\kappa := \|D\widetilde{\boldsymbol{T}}\|\|D(\widetilde{\boldsymbol{T}}^{-1})\|$ and $\lambda := \|\det(D\widetilde{\boldsymbol{T}})\|_{L^\infty(\widehat{K})}\|\det(D\widetilde{\boldsymbol{T}})^{-1}\|_{L^\infty(\widehat{K})}$ are bounded uniformly. Starting with (13.4), we observe that

$$\|D(\widetilde{\boldsymbol{T}}^{-1})\| \sum_{i=1}^{8} \|D\widehat{\psi}_i\|\|\widetilde{\boldsymbol{a}}_i - \boldsymbol{a}_i\|_{\ell^2(\mathbb{R}^2)} \le c\sigma_\sharp h_{\max}, \tag{13.29}$$

which is less than 1 provided the cells are small enough. Moreover, using the estimates $\|D^2\widetilde{\boldsymbol{T}}\| \le 2h_{\max}$ (see Remark 13.14) and $\|D\widetilde{\boldsymbol{T}}\| \ge h_{\min}(1-\gamma)$, we infer that $\|D^2\boldsymbol{T}\| \le \|D^2\widetilde{\boldsymbol{T}}\| + ch_{\max}^2 \le c' h_{\max} \le c'\sigma_\sharp\|D\widetilde{\boldsymbol{T}}\|$, which proves (13.8) for all $l = 2$. Moreover $\|D^l\boldsymbol{T}\| \le ch_{\max}^2 \le c'h_{\max} \le c'\sigma_\sharp\|D\widetilde{\boldsymbol{T}}\|$, since $D^l\widetilde{\boldsymbol{T}} = 0$ for $l \ge 3$, which proves (13.8) for all $l \ge 3$. Furthermore, $[\![D^2\boldsymbol{T}]\!] \le ch_{\max}^2 \le c\sigma_\sharp^2\|D\widetilde{\boldsymbol{T}}\|^2$ since $[\![D^2\widetilde{\boldsymbol{T}}]\!] = 0$, and $[\![D^l\boldsymbol{T}]\!] = 0$ for all $l \ge 3$, which proves (13.13). Finally we have already seen that $\lambda \le \frac{S_{\max}}{S_{\min}} \le \sigma_\sharp$, and we have $\kappa \le 2\sigma_\sharp$ since $\|D\widetilde{\boldsymbol{T}}\| \le 2h_{\max}$. In conclusion Theorem 13.9 gives

$$|v - \mathcal{I}_K v|_{W^{m,p}(K)} \le c\,\sigma_\sharp^m h_K^{l-m}\|v\|_{W^{l,p}(K)}, \tag{13.30}$$

for all K, all $v \in W^{l,p}(K)$, all $p \in [1,\infty]$, every integers m, l with $0 \le m \le l \le k+1$ and $W^{l,p}(K) \subset V(K)$ (i.e., $W^{l,p}(K)$ is in the domain of $\mathcal{I}_K$).

Exercises

Exercise 13.1 (Bound on determinant). (i) Let $\mathbb{A}$ be a $d\times d$ real-valued matrix. Prove that $\det(\mathbb{A}) \le \|\mathbb{A}\|^d$, where $\|\mathbb{A}\|$ is the matrix norm subordinate to the Euclidean norm in $\mathbb{R}^d$. (*Hint*: $\det(\mathbb{A})^2 = \det(\mathbb{A}^\mathsf{T}\mathbb{A})$.) (ii) Prove (13.7). (*Hint*: $\det(D\boldsymbol{T}) = \det(D\widetilde{\boldsymbol{T}})\det(\boldsymbol{I} + (D\widetilde{\boldsymbol{T}})^{-1}\boldsymbol{E})$.)

Exercise 13.2 (Lemma 13.4). Complete the proof of Lemma 13.4. (*Hint*: use induction and the chain rule.)

Exercise 13.3 (Pure derivatives, $\mathbb{Q}_k$-polynomials). Let $\{\boldsymbol{e}_i\}_{i\in\{1:d\}}$ be the canonical Cartesian basis of $\mathbb{R}^d$. Let $k \ge 1$. Verify that $D^{k+1}q(\boldsymbol{e}_i,\ldots,\boldsymbol{e}_i) = 0$ for all $i \in \{1:d\}$ if and only if $q \in \mathbb{Q}_{k,d}$. (*Hint*: by induction on d.) What is instead the characterization of polynomials in $\mathbb{P}_k$ in terms of $D^{k+1}q$?

Exercise 13.4 (Chain rule). Let $f \in \mathcal{C}^3(U;W_1)$ and $g \in \mathcal{C}^3(W_1;W_2)$, $n \ge 1$, where V, W_1, W_2 are Banach spaces and U is an open set in V. (i) Evaluate the pure derivatives $D^2(g\circ f)(x)(h,h)$ and $D^3(g\circ f)(x)(h,h,h)$ for $x \in U$ and $h \in V$. (ii) Rewrite these expressions when f and g map from $\mathbb{R}$ to $\mathbb{R}$.

Exercise 13.5 (Tensor-product transformation). Assume the transformation $\boldsymbol{T}$ has the tensor-product form $\boldsymbol{T}(\widehat{\boldsymbol{x}}) = \sum_{j=1}^d t_j(\widehat{\boldsymbol{x}}_j)\boldsymbol{e}_j$ for some univariate function t_j, $j \in \{1:d\}$, where $\{\boldsymbol{e}_1,\ldots,\boldsymbol{e}_d\}$ is the canonical Cartesian basis of $\mathbb{R}^d$. (i) Show that (13.15) can be sharpened as $[\![w\circ\boldsymbol{T}]\!]_{W^{s,p}(\widehat{K})} \le c\|\det(D\widetilde{\boldsymbol{T}})^{-1}\|^{\frac{1}{p}}_{L^\infty(\widehat{K})}\|D\widetilde{\boldsymbol{T}}\|^s[\![w]\!]_{W^{s,p}(K)}$. (ii) What is the consequence of this new bound on the error estimate (13.21) under the assumption (13.20)?

Exercise 13.6 ($\mathbb{Q}_1$-quadrangles). Prove that $\det(D\boldsymbol{T}(\widehat{\boldsymbol{a}}_i)) = |P_i|$ where P_i is the parallelogram formed by $\boldsymbol{a}_{i-1}$, $\boldsymbol{a}_i$, $\boldsymbol{a}_{i+1}$ (with $\boldsymbol{a}_0 := \boldsymbol{a}_4$ and $\boldsymbol{a}_5 := \boldsymbol{a}_1$). (*Hint*: see §13.4.)

Exercise 13.7 (Butterfly subdivision algorithm). Consider a mesh composed of four triangles with the following connectivity array $\texttt{j\_geo}(1,1{:}3) = (3,4,5)$, $\texttt{j\_geo}(2,1{:}3) = (0,4,5)$, $\texttt{j\_geo}(3,1{:}3) = (1,3,5)$, and $\texttt{j\_geo}(4,1{:}3) = (2,3,4)$. Let $\boldsymbol{m}$ be the midpoint of the edge $(\boldsymbol{z}_3,\boldsymbol{z}_4)$. Let $\widehat{\boldsymbol{z}}_0 = (0,0)$, $\widehat{\boldsymbol{z}}_1 = (1,0)$, $\widehat{\boldsymbol{z}}_2 = (0,1)$, $\widehat{\boldsymbol{z}}_3 = (\frac12,\frac12)$, $\widehat{\boldsymbol{z}}_4 = (0,\frac12)$, $\widehat{\boldsymbol{z}}_5 = (\frac12,0)$. Consider now the curved triangle given by the $\mathbb{P}_2$ geometric transformation $\boldsymbol{T}$ that transforms $\widehat{\boldsymbol{z}}_i$ to $\boldsymbol{z}_i$ for all $i \in \{0{:}5\}$. Let $f_0,\ldots,f_7 \in \mathbb{R}$. Let $\widehat{p} \in \mathbb{P}_{2,2}$ be the polynomial defined by $\widehat{p}(\widehat{\boldsymbol{z}}_i) := f_i$ for all $i \in \{0{:}5\}$. (i) Compute $\widehat{p}(\boldsymbol{T}^{-1}(\boldsymbol{m}))$. (ii) Consider two additional points $\boldsymbol{z}_6$, $\boldsymbol{z}_7$ and two more triangles given by $\texttt{j\_geo}(5,1{:}3) = (2,3,6)$, $\texttt{j\_geo}(6,1{:}3) = (2,4,7)$. Let $\boldsymbol{T}'$ be the $\mathbb{P}_2$ geometric mapping that transforms $\widehat{\boldsymbol{z}}_i$ to $\boldsymbol{z}_i$, for all $i \in \{2{:}7\}$. Let $\widehat{p}' \in \mathbb{P}_{2,2}$ be defined by $\widehat{p}'(\widehat{\boldsymbol{z}}_i) := f_i$ for all $i \in \{2{:}7\}$. Compute $\frac12\big(\widehat{p}(\boldsymbol{T}^{-1}(\boldsymbol{m})) + \widehat{p}'((\boldsymbol{T}')^{-1}(\boldsymbol{m}))\big)$. (Note: The name of the algorithm comes from the shape of the generic configuration. The algorithm is used for 3D computer graphics applications. It allows the representation of smooth surfaces via the specification of coarser piecewise linear polygonal meshes. Given an initial polygonal mesh, a smooth surface is obtained by recursively applying the butterfly subdivision algorithm to the Cartesian coordinates of the vertices; see Dyn et al. [88].)

Solution to exercises

Exercise 13.1 (Bound on determinant). (i) The matrix $\mathbb{A}^{\mathsf{T}}\mathbb{A}$ is symmetric, so it can be diagonalized in $\mathbb{R}$ with eigenvalues λ_i, $i \in \{1{:}d\}$. Moreover $0 \le \lambda_i \le \|\mathbb{A}\|^2$ for all $i \in \{1{:}d\}$, where $\|\mathbb{A}\|$ is the matrix norm subordinate to the Euclidean norm in $\mathbb{R}^d$. This implies that

$$\det(\mathbb{A})^2 = \det(\mathbb{A}^{\mathsf{T}}\mathbb{A}) = \Pi_{i=1}^d \lambda_i \le \|\mathbb{A}\|^{2d}.$$

(ii) Since $\det(D\boldsymbol{T}) = \det(D\widetilde{\boldsymbol{T}})\det(\boldsymbol{I} + (D\widetilde{\boldsymbol{T}})^{-1}\boldsymbol{E})$ and $\|\boldsymbol{I} + (D\widetilde{\boldsymbol{T}})^{-1}\boldsymbol{E}\| \le 1 + c_1$, the upper bound on $\det(D\boldsymbol{T})$ results from Step (i) (using the same notation for endomorphisms in $\mathbb{R}^d$ and their matrix representation). The lower bound results from $\det(D\widetilde{\boldsymbol{T}}) = \det(D\boldsymbol{T})\det(\boldsymbol{I} + (D\widetilde{\boldsymbol{T}})^{-1}\boldsymbol{E})^{-1}$ and $\|(\boldsymbol{I} + (D\widetilde{\boldsymbol{T}})^{-1}\boldsymbol{E})^{-1}\| \le (1 - c_1)^{-1}$.

Exercise 13.2 (Lemma 13.4). To be done

Exercise 13.3 (Pure derivatives, $\mathbb{Q}_k$-polynomials). A direct verification shows that any polynomial $q \in \mathbb{Q}_{k,d}$ verifies $D^{k+1}q(x)(\boldsymbol{e}_i, \ldots, \boldsymbol{e}_i) = 0$ for all $i \in \{1{:}d\}$. Conversely, assume that q is such that $D^{k+1}q(x)(\boldsymbol{e}_i, \ldots, \boldsymbol{e}_i) = 0$, for all $i \in \{1{:}d\}$. We proceed by induction on d. If $d = 1$, then $q \in \mathbb{Q}_{k,1}$. For $d \ge 2$, writing $\boldsymbol{x} = (x', x_d)$ and fixing x', we infer that the $(k+1)$-th derivative of the function $x_d \mapsto q(x', x_d)$ is zero, so that there are functions $q_0(x'), \ldots, q_k(x')$ s.t. $q(x) = \sum_{m=0}^k q_m(x')x_d^m$. Since for all $j < d$ we have

$$0 = D^{k+1}q(x)(\boldsymbol{e}_j, \ldots, \boldsymbol{e}_j) = \sum_{m=0}^{k} D^{k+1}q_m(x')(\boldsymbol{e}_j, \ldots, \boldsymbol{e}_j)x_d^m,$$

and the monomials $\{x_d^m\}$ are linearly independent, we infer that

$$D^{k+1}q_m(x')(\boldsymbol{e}_j, \ldots, \boldsymbol{e}_j) = 0, \qquad \forall j \in \{1{:}(d-1)\}.$$

By the induction hypothesis we have $q_m \in \mathbb{Q}_{k,d-1}$, so that $q \in \mathbb{Q}_{k,d}$. By proceeding as above we finally show that $q \in \mathbb{P}_k$ if and only if $D^{k+1}q = 0$, that is, $D^{k+1}q(\boldsymbol{h}_1, \ldots, \boldsymbol{h}_{k+1}) = 0$ for all $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_{k+1} \in \mathbb{R}^d$.

Exercise 13.4 (Chain rule). (i) We apply Lemma B.4. For $n = 2$, the summation in l has two terms and we obtain (we omit the point x in the (Fréchet) derivatives of f)

$$D^2(g \circ f)(x)(h, h) = Dg(f(x))(D^2 f(h, h)) + D^2 g(f(x))(Df(h), Df(h)).$$

For $n = 3$, the summation in l has three terms and we obtain

$$\begin{aligned} &D^3(g \circ f)(x)(h, h, h) = Dg(f(x))(D^3 f(h, h, h)) \\ &\quad + 3D^2 g(f(x))(Df(h), D^2 f(h, h)) + D^3 g(f(x))(Df(h), Df(h), Df(h)), \end{aligned}$$

where we used Lemma B.3 for the second term on the right-hand side.
(ii) When f and g map from $\mathbb{R}$ to $\mathbb{R}$, we obtain

$$(g \circ f)''(x) = g'(f(x))(f'(x))^2 + g''(f(x)),$$

and

$$(g \circ f)'''(x) = g'(f(x))(f'(x))^3 + 3g''(f(x))f'(x)f''(x) + g'''(f(x))(f'(x))^3.$$

Exercise 13.5 (Tensor-product transformation). (i) When $\boldsymbol{T}$ has a tensor-product form, we obtain $D^r\boldsymbol{T}(\widehat{\boldsymbol{x}})(\boldsymbol{e}_i,\ldots,\boldsymbol{e}_i) = t_i^{(r)}(x_i)\boldsymbol{e}_i$ for all $i \in \{1{:}d\}$, so that using the chain rule now leads to

$$\begin{aligned}|D^s(w \circ \boldsymbol{T})(\widehat{\boldsymbol{x}})|_{\mathbb{Q}} \le c \sum_{l=0}^{s} &|(D^l w)(\boldsymbol{T}(\widehat{\boldsymbol{x}}))|_{\mathbb{Q}} \\ &\times \sum_{1\le r_1+\ldots+r_l=s} |D^{r_1}\boldsymbol{T}(\widehat{\boldsymbol{x}})|_{\mathbb{Q}} \ldots |D^{r_l}\boldsymbol{T}(\widehat{\boldsymbol{x}})|_{\mathbb{Q}}.\end{aligned}$$

The expected estimate readily follows.
(ii) The error estimate (13.21) under the assumption (13.20) becomes

$$|v - \mathcal{I}_K v|_{W^{m,p}(K)} \le c\,\lambda^{\frac{1}{p}}\kappa^m \|D\widetilde{\boldsymbol{T}}\|^{l-m}\, [\![v]\!]_{W^{l,p}(K)}.$$

Note that such an error estimate cannot hold under the assumption (13.19) (think of $k = l = 1$, $d = 2$, and $v = x_1x_2$ for which $[\![v]\!]_{W^{1,p}(K)} = 0$).

Exercise 13.6 ($\mathbb{Q}_1$-quadrangles). Consider the (Fréchet) derivative $D\boldsymbol{T}$ at $\widehat{a}_1$ which corresponds to $\widehat{x}_1 = \widehat{x}_2 = 0$. Then, $D\boldsymbol{T}(\widehat{\boldsymbol{x}}) = (a_2 - a_1, a_4 - a_1)$, so that, taking into account the orientation of the enumeration of vertices leads to the expected result.

Exercise 13.7 (Butterfly subdivision algorithm). (i) Let us set $\widehat{\boldsymbol{m}} := \boldsymbol{T}^{-1}(\boldsymbol{m})$. Using the following expression of the $\mathbb{P}_2$ shape functions:

$$\begin{aligned}&\widehat{\theta}_0 = \widehat{\lambda}_0(2\widehat{\lambda}_0 - 1), && \widehat{\theta}_1 = \widehat{\lambda}_1(2\widehat{\lambda}_1 - 1), && \widehat{\theta}_2 = \widehat{\lambda}_2(2\widehat{\lambda}_2 - 1),\\ &\widehat{\theta}_3 = 4\widehat{\lambda}_1\widehat{\lambda}_2, && \widehat{\theta}_4 = 4\widehat{\lambda}_2\widehat{\lambda}_0, && \widehat{\theta}_5 = 4\widehat{\lambda}_0\widehat{\lambda}_1,\end{aligned}$$

together with $\widehat{\lambda}_0(\widehat{\boldsymbol{m}}) = \frac{1}{4}$, $\widehat{\lambda}_1(\widehat{\boldsymbol{m}}) = \frac{1}{4}$, $\widehat{\lambda}_2(\widehat{\boldsymbol{m}}) = \frac{1}{2}$, we obtain

$$\widehat{p}(\widehat{\boldsymbol{m}}) = \sum_{i=0}^{5} f_i\widehat{\theta}_i(\widehat{\boldsymbol{m}}) = -\frac{1}{8}f_0 - \frac{1}{8}f_1 + \frac{1}{2}f_3 + \frac{1}{2}f_4 + \frac{1}{4}f_5.$$

(ii) Similarly, we have

$$\widehat{p}'(\widehat{\boldsymbol{m}}) = -\frac{1}{8}f_6 - \frac{1}{8}f_7 + \frac{1}{2}f_3 + \frac{1}{2}f_4 + \frac{1}{4}f_2.$$

Hence,

$$\frac{1}{2}(\widehat{p}(\widehat{\boldsymbol{m}}) + \widehat{p}'(\widehat{\boldsymbol{m}})) = -\frac{1}{16}f_0 - \frac{1}{16}f_1 + \frac{1}{8}f_2 + \frac{1}{2}f_3 + \frac{1}{2}f_4 + \frac{1}{8}f_5 - \frac{1}{16}f_6 - \frac{1}{16}f_7.$$

The generic configuration is shown in the right panel of the figure. The mesh mapped to the reference space is shown in the left panel of the figure.