Chapter 3

A Priori Analysis

3.1 P1-Finite Element Method in 2D

A set $T \subset \mathbb{R}^2$ is called a **non-degenerate triangle** provided that there are nodes $x_T, y_T, z_T \in \mathbb{R}^2$ with $T = \operatorname{conv}\{x_T, y_T, z_T\}$ and provided that $|T| > 0$, i.e., T has positive measure. We note that T is in particular bounded and closed, whence compact. We denote by

$$\mathcal{K}_T := \{x_T, y_T, z_T\} \tag{3.1}$$

the **set of nodes** of T and by

$$\mathcal{E}_T := \big\{ \operatorname{conv}\{x_T, y_T\}, \operatorname{conv}\{y_T, z_T\}, \operatorname{conv}\{z_T, x_T\}\big\} \tag{3.2}$$

the **set of edges** of T. The **diameter** of T is denoted by

$$h_T := \operatorname{diam}(T) := \max \big\{|x - y| \,\big|\, x, y \in T\big\}. \tag{3.3}$$

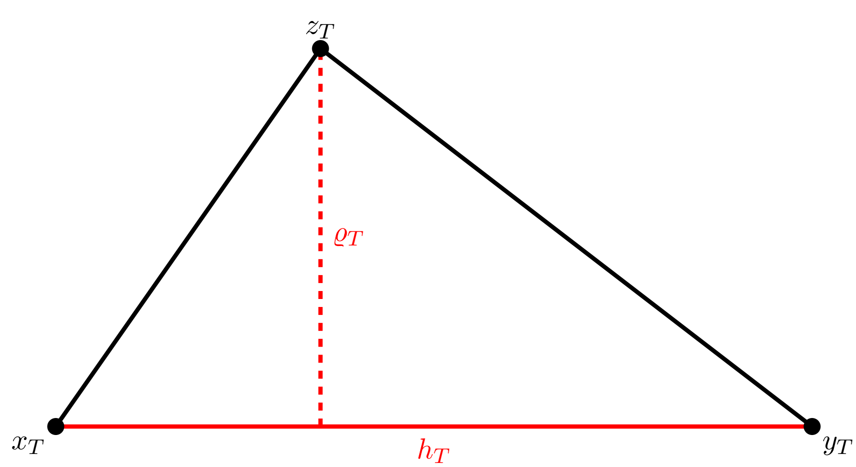

FIGURE 3.1. *The diameter h_T of the triangle T is the length of the longest edge (possibly non unique). The quantity ϱ_T denotes the corresponding height.*

Moreover, we define the **edge length**

$$h_E := \operatorname{diam}(E) := \max\{|x-y| \mid x, y \in E\} \tag{3.4}$$

for all edges $E \in \mathcal{E}_T$. Clearly, the diameter h_T of a triangle is the length of the longest edge (possibly non unique), i.e., there is some $E \in \mathcal{E}_T$ with $h_T = h_E$. The **height** over the longest edge E of T is denoted by ϱ_T, cf. Figure 3.1. Recall that the measure of the triangle reads

$$|T| = \frac{h_T \rho_T}{2}. \tag{3.5}$$

The most important example is the **reference triangle**

$$T_{\text{ref}} := \operatorname{conv}\{(0,0), (1,0), (0,1)\} \tag{3.6}$$

which has measure $|T_{\text{ref}}| = 1/2$.

Exercise 11. Give a formal proof that the diameter of a triangle T is the length of one longest edge, i.e., $h_T = \max_{E \in \mathcal{E}_T} h_E$. *Hint:* Use that the convex hull $\operatorname{conv}(M) := \bigcap\{\widehat{M} \subseteq \mathbb{R}^d \mid \widehat{M} \text{ is convex with } M \subseteq \widehat{M}\}$ of a set $M \subseteq \mathbb{R}^d$ is also characterized by $\operatorname{conv}(M) = \{\sum_{j=1}^N \lambda_j x_j \mid N \in \mathbb{N}, x_j \in M, \lambda_j \geq 0 \text{ with } \sum_{j=1}^N \lambda_j = 1\}$. The proof then directly applies to general simplices in $\mathbb{R}^d$, i.e., $T = \operatorname{conv}\{x_0, \ldots, x_d\} \subset \mathbb{R}^d$. □

Definition. A set $\mathcal{T}$ is a **triangulation** of Ω (consisting of triangles) if and only if

- $\mathcal{T}$ is a finite set of non-degenerate triangles,
- the closure of Ω is covered by $\mathcal{T}$, i.e., $\overline{\Omega} = \bigcup \mathcal{T}$,
- for all $T, T' \in \mathcal{T}$ with $T \neq T'$, it holds that $|T \cap T'| = 0$, i.e., the overlap is a set of measure zero.

By $\mathcal{K} := \bigcup\{x \in \mathcal{K}_T \mid T \in \mathcal{T}\}$, we then denote the **set of nodes** of the triangulation $\mathcal{T}$ and by $\mathcal{E} := \bigcup\{E \in \mathcal{E}_T \mid T \in \mathcal{T}\}$ the **set of edges** of the triangulation $\mathcal{T}$. A triangulation of Ω is called **conforming** or **regular (in the sense of Ciarlet)** provided that the intersection of two elements $T, T' \in \mathcal{T}$ with $T \neq T'$ is

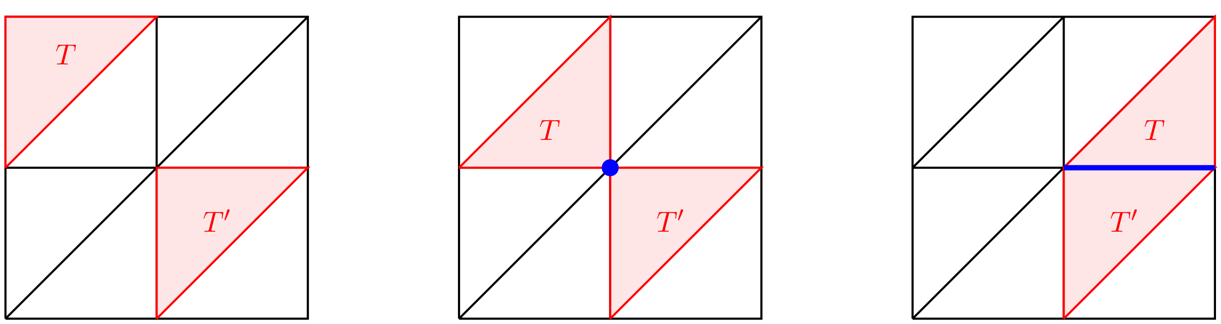

FIGURE 3.2. *For a regular triangulation $\mathcal{T}$, the intersection of two elements $T \neq T'$ is either empty, a joint node, or a joint edge.*

- either empty,
- or a joint node, i.e., $T \cap T' = \{z\} = \mathcal{K}_T \cap \mathcal{K}_{T'}$,
- or a joint edge, i.e., $E := T \cap T' \in \mathcal{E}_T \cap \mathcal{E}_{T'}$,

cf. Figure 3.2. According to this regularity assumption, an edge $E \in \mathcal{E}$ with surface measure $|E \cap \Gamma| > 0$ automatically satisfies $E \subseteq \Gamma$, i.e., an edge E is either a boundary edge or an interior edge. Additionally, we always assume that a regular triangulation resolves the boundary conditions: If $\Gamma = \partial\Omega$ is partitioned into Dirichlet and Neumann boundary Γ_D and Γ_N, respectively, each boundary edge $E \in \mathcal{E}$ with $E \subseteq \Gamma$ satisfies

- either $E \subseteq \overline{\Gamma}_D$
- or $E \subseteq \overline{\Gamma}_N$.

With this assumption, we define the (disjoint) sets of boundary edges

$$\mathcal{E}_D := \big\{E \in \mathcal{E} \,\big|\, E \subseteq \overline{\Gamma}_D\big\} \quad \text{and} \quad \mathcal{E}_N := \big\{E \in \mathcal{E} \,\big|\, E \subseteq \overline{\Gamma}_N\big\} \tag{3.7}$$

as well as the set of all interior edges

$$\mathcal{E}_\Omega := \mathcal{E} \backslash (\mathcal{E}_D \cup \mathcal{E}_N). \tag{3.8}$$

We finally note that, for each $E \in \mathcal{E}_\Omega$, there are two elements $T, T' \in \mathcal{T}$ with $E = T \cap T'$.

Exercise 12. Let $\mathcal{T}$ be a regular triangulation of Ω and $v : \Omega \to \mathbb{R}$ such that $v|_T \in C^1(T)$ for all $T \in \mathcal{T}$. Prove that $v \in H^1(\Omega)$ if and only if $v \in C(\Omega)$. □

The following proposition essentially follows from the regularity of the triangulation $\mathcal{T}$.

Proposition 3.1. *For a regular triangulation $\mathcal{T}$ of Ω, we define the discrete space*

$$\mathcal{S}^1(\mathcal{T}) := \big\{v_h \in C(\Omega) \,\big|\, \forall T \in \mathcal{T} \quad v_h|_T \text{ affine}\big\} \tag{3.9}$$

of all $\mathcal{T}$-piecewise affine and globally continuous functions. Then, there holds the following:
(i) *$\mathcal{S}^1(\mathcal{T})$ is an N-dimensional subspace of $H^1(\Omega)$ with $N = \#\mathcal{K}$ the number of nodes.*
(ii) *For each node $z \in \mathcal{K}$, there is a unique **hat function***

$$\zeta_z \in \mathcal{S}^1(\mathcal{T}) \quad \text{with} \quad \zeta_z(z') = \delta_{zz'} \quad \text{for all } z' \in \mathcal{K}. \tag{3.10}$$

(iii) *The set $\mathcal{B} := \big\{\zeta_z \,\big|\, z \in \mathcal{K}\big\}$ is a basis of $\mathcal{S}^1(\mathcal{T})$, the so-called **nodal basis**.*

Proof. **1. step.** According to the regularity of $\mathcal{T}$, hat functions ζ_z are automatically continuous on Ω: For each element $T \in \mathcal{T}$, an affine function $v_h : T \to \mathbb{R}$ is uniquely determined by the nodal values $v_h(z)$ for $z \in \mathcal{K}_T$. Therefore, the $\mathcal{T}$-piecewise affine hat function ζ_z defined by $\zeta_z(z') = \delta_{zz'}$ is uniquely defined. We now show that $\zeta_z \in C(\Omega)$: If $T, T' \in \mathcal{T}$ are elements with $T \cap T' \neq \emptyset$, regularity of $\mathcal{T}$ implies that either $T = T'$ or $\{z'\} = T \cap T'$ is a joint point or $E = T \cap T'$ is a joint edge. In the latter case, note that the trace on E of the affine function $\zeta_z|_T$ as well as of $\zeta_z|_{T'}$ is

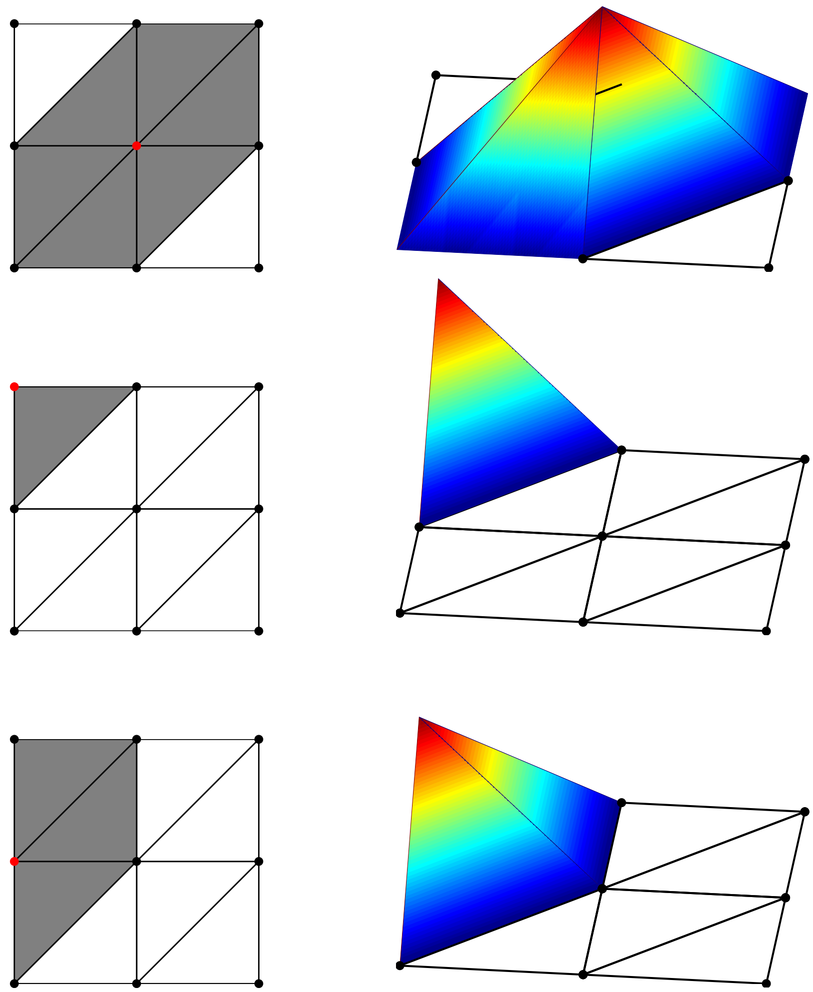

FIGURE 3.3. *Examples of P1 hat functions ζ_z: The left figures show the mesh as well as the support* $\operatorname{supp}(\zeta_z)$ *in grey, where the corresponding node $z \in \mathcal{K}$ is indicated in red. The right figures show the plots of the hat functions. Triangles $T \in \mathcal{T}$ with $\zeta_z|_T = 0$ are filled with white.*

uniquely defined on the edge E by the nodal values $\zeta_z(x_E)$ and $\zeta_z(y_E)$, where $E = \operatorname{conv}\{x_E, y_E\}$. Therefore the traces of $\zeta_z|_T$ and $\zeta_z|_{T'}$ on E coincide, i.e., ζ_z is continuous on each interior edge.

2. step. The nodal basis $\mathcal{B}$ is a basis of $\mathcal{S}^1(\mathcal{T})$ and $\dim \mathcal{S}^1(\mathcal{T}) = \#\mathcal{K}$: Clearly, the hat functions are linearly independent, $\mathcal{B} \subseteq \mathcal{S}^1(\mathcal{T})$, and $\#\mathcal{B} = \#\mathcal{K}$. Moreover, each function $v_h \in \mathcal{S}^1(\mathcal{T})$ is uniquely defined by the nodal values $v_h(z)$ for $z \in \mathcal{K}$ and can thus be written as the linear combination of the hat functions, i.e., $\mathcal{S}^1(\mathcal{T}) \subseteq \operatorname{span}(\mathcal{B})$.

3. step. The inclusion $\mathcal{S}^1(\mathcal{T}) \subset H^1(\Omega)$ follows from Exercise 12. ■

Remark. Examples for hat functions ζ_z are shown in Figure 3.3. Note that the support $\operatorname{supp}(\zeta_z)$ is always local. This leads to a sparse Galerkin matrix A, i.e., most of the entries of A are zero. □

For a given Dirichlet boundary $\Gamma_D \subseteq \Gamma$, we use the discrete space $\mathcal{S}^1_D(\mathcal{T})$ to discretize the weak form of the mixed boundary value problem. In case of $\Gamma_D = \Gamma$, we consider the space $\mathcal{S}^1_0(\mathcal{T})$.

Corollary 3.2. *Let $\mathcal{T}$ be a regular triangulation of Ω. Then, the space*

$$\mathcal{S}^1_D(\mathcal{T}) := \big\{ v_h \in \mathcal{S}^1(\mathcal{T}) \,\big|\, \forall z \in \mathcal{K} \cap \overline{\Gamma}_D \quad v_h(z) = 0 \big\} \tag{3.11}$$

is a finite dimensional subspace of $H^1_D(\Omega)$ of dimension $\#\big\{z \in \mathcal{K} \,\big|\, z \notin \overline{\Gamma}_D\big\}$. The space

$$\mathcal{S}^1_0(\mathcal{T}) := \big\{ v_h \in \mathcal{S}^1(\mathcal{T}) \,\big|\, \forall z \in \mathcal{K} \cap \Gamma \quad v_h(z) = 0 \big\} \tag{3.12}$$

is a finite dimensional subspace of $H^1_0(\Omega)$ of dimension $\#\big\{z \in \mathcal{K} \,\big|\, z \notin \Gamma\big\}$.

Proof. We only need to show that $v_h|_{\Gamma_D} = 0$ for $v_h \in \mathcal{S}^1_D(\mathcal{T})$. Let $x \in \Gamma_D$. According to the regularity of $\mathcal{T}$, there is an edge $E \in \mathcal{E}_D$ such that $x \in E$. Since the trace $v_h|_E$ is affine, it is uniquely determined by the nodal values $v_h(x_T) = 0 = v_h(y_T)$, where $E = \operatorname{conv}\{x_T, y_T\}$. Consequently, $v_h|_E = 0$ for all $E \in \mathcal{E}_D$ and hence $v_h \in H^1_D(\Omega)$. In particular, we obtain the claim for $\mathcal{S}^1_0(\mathcal{T})$ in case of $\Gamma_D = \Gamma$. ∎

For the discretization of the Neumann problem, we are dealing with $\mathcal{S}^1_*(\mathcal{T})$.

Corollary 3.3. *For a regular triangulation $\mathcal{T}$ of Ω, the space*

$$\mathcal{S}^1_*(\mathcal{T}) := \big\{ v_h \in \mathcal{S}^1(\mathcal{T}) \,\big|\, \textstyle\int_\Omega v_h \, dx = 0 \big\} \tag{3.13}$$

is a finite dimensional subspace of $H^1_*(\Omega)$ of dimension $\#\mathcal{K} - 1$.

Proof. Clearly, it holds that $\mathcal{S}^1_*(\mathcal{T}) \subseteq H^1_*(\Omega)$. Note that $I(v_h) := \int_\Omega v_h \, dx$ is a linear functional on $\mathcal{S}^1(\mathcal{T})$ with kernel $\mathcal{S}^1_*(\mathcal{T}) = \ker(I)$. Since $\operatorname{rank}(I) = 1$, Linear Algebra yields that $\dim \mathcal{S}^1_*(\mathcal{T}) = \dim \mathcal{S}^1(\mathcal{T}) - 1$. ∎

The **P1 Finite Element Method** now consists of using the Galerkin method with the discrete spaces $\mathcal{S}^1_0(\mathcal{T})$, $\mathcal{S}^1_D(\mathcal{T})$, and $\mathcal{S}^1_*(\mathcal{T})$ to approximate the weak solution of the Dirichlet problem, the mixed boundary value problem, and the Neumann problem, respectively. From now on, we shall assume that $\mathcal{T}$ is a regular triangulation of Ω. We start with the **Dirichlet problem**

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \Gamma, \end{aligned}$$

for given data $f \in L^2(\Omega)$. The P1-FEM then reads: Find $u_h \in \mathcal{S}^1_0(\mathcal{T})$ such that

$$(\nabla u_h \,;\, \nabla v_h)_{L^2(\Omega)} = (f \,;\, v_h)_{L^2(\Omega)} \quad \text{for all } v_h \in \mathcal{S}^1_0(\mathcal{T}). \tag{3.14}$$

Second, the **mixed boundary value problem** reads

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \Gamma_D, \\ \partial u/\partial n &= \phi \quad \text{on } \Gamma_N, \end{aligned}$$

with $\Gamma = \overline{\Gamma}_D \cup \overline{\Gamma}_N$, $\Gamma_D \cap \Gamma_N = \emptyset$, and $|\Gamma_D| > 0$. The data satisfy $f \in L^2(\Omega)$ and $\phi \in L^2(\Gamma_N)$. The P1-FEM for the mixed BVP reads: Find $u_h \in \mathcal{S}^1_D(\mathcal{T})$ such that

$$(\nabla u_h \,;\, \nabla v_h)_{L^2(\Omega)} = (f \,;\, v_h)_{L^2(\Omega)} + (\phi \,;\, v_h)_{L^2(\Gamma_N)} \quad \text{for all } v_h \in \mathcal{S}^1_D(\mathcal{T}). \tag{3.15}$$

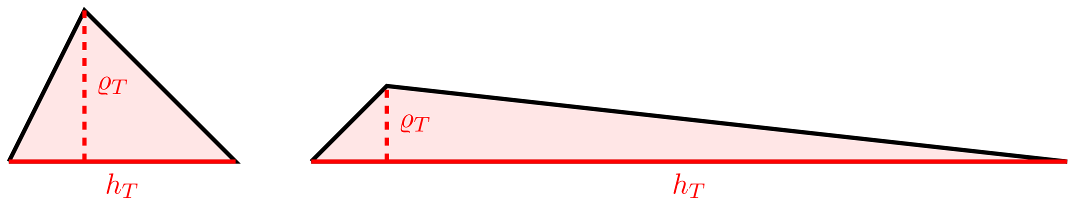

FIGURE 3.4. *two triangles with small (left) and large (right) shape regularity constant $\sigma(T) := h_T/\varrho_T$*

Finally, we consider the **Neumann problem**

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega, \\ \partial u/\partial n &= \phi \quad \text{on } \Gamma, \end{aligned}$$

where the data $f \in L^2(\Omega)$ and $\phi \in L^2(\Gamma)$ are assumed to satisfy $\int_\Omega f\,dx + \int_\Gamma \phi\,ds = 0$. The P1-FEM for the Neumann problem reads: Find $u_h \in \mathcal{S}^1_*(\mathcal{T})$ such that

$$(\nabla u_h\,;\,\nabla v_h)_{L^2(\Omega)} = (f\,;\,v_h)_{L^2(\Omega)} + (\phi\,;\,v_h)_{L^2(\Gamma)} \quad \text{for all } v_h \in \mathcal{S}^1_*(\mathcal{T}). \tag{3.16}$$

3.2 Approximation Theorem and Bramble-Hilbert Lemma

3.2.1 Uniform Mesh-Refinement and Shape Regularity

Let $h \in L^\infty(\Omega)$ and $\varrho \in L^\infty(\Omega)$ denote the **local mesh-width** functions which are defined by

$$h|_T := h_T = \operatorname{diam}(T) \quad \text{and} \quad \varrho|_T := \varrho_T \quad \text{for all } T \in \mathcal{T}. \tag{3.17}$$

Moreover, the quantities

$$\sigma(T) := \frac{h_T}{\varrho_T} \quad \text{and} \quad \sigma(\mathcal{T}) := \|h/\varrho\|_{L^\infty(\Omega)} = \max_{T\in\mathcal{T}} \frac{h_T}{\varrho_T} \geq 1 \tag{3.18}$$

denote the **shape regularity constant** of an element $T \in \mathcal{T}$ resp. the triangulation $\mathcal{T}$, see Fig. 3.4. Note that $|T| = h_T\varrho_T/2$ so that $2\,h_T/\varrho_T = h_T^2/|T|$. The shape regularity constant will affect all error estimates, so that mesh-refinement has to avoid a blow-up of $\sigma(\mathcal{T})$. We say that a regular mesh $\mathcal{T}$ is **γ-shape regular**, if $\sigma(\mathcal{T}) \leq \gamma < \infty$.

For this section, we stick with the so-called **uniform mesh-refinement**: Given a regular triangulation $\mathcal{T}^{(\mathrm{old})}$, we obtain a new triangulation $\mathcal{T}^{(\mathrm{new})}$ as follows: Each element $T \in \mathcal{T}^{(\mathrm{old})}$ is split into 4 similar triangles $T_1, \ldots, T_4 \in \mathcal{T}^{(\mathrm{new})}$, cf. Figure 3.5. Therefore, each node $z \in \mathcal{K}^{(\mathrm{new})}$ either belongs to $\mathcal{K}^{(\mathrm{old})}$ or is the midpoint of an edge $E \in \mathcal{E}^{(\mathrm{old})}$. We stress some simple observations:

- The new triangulation $\mathcal{T}^{(\mathrm{new})}$ is also regular.
- The local mesh-width functions satisfy $h^{(\mathrm{new})} = h^{(\mathrm{old})}/2$ and $\varrho^{(\mathrm{new})} = \varrho^{(\mathrm{old})}/2$.
- In particular, the shape regularity constant satisfies that $\sigma(\mathcal{T}^{(\mathrm{old})}) = \sigma(\mathcal{T}^{(\mathrm{new})})$.

Further mesh-refinement strategies are discussed in the following section.

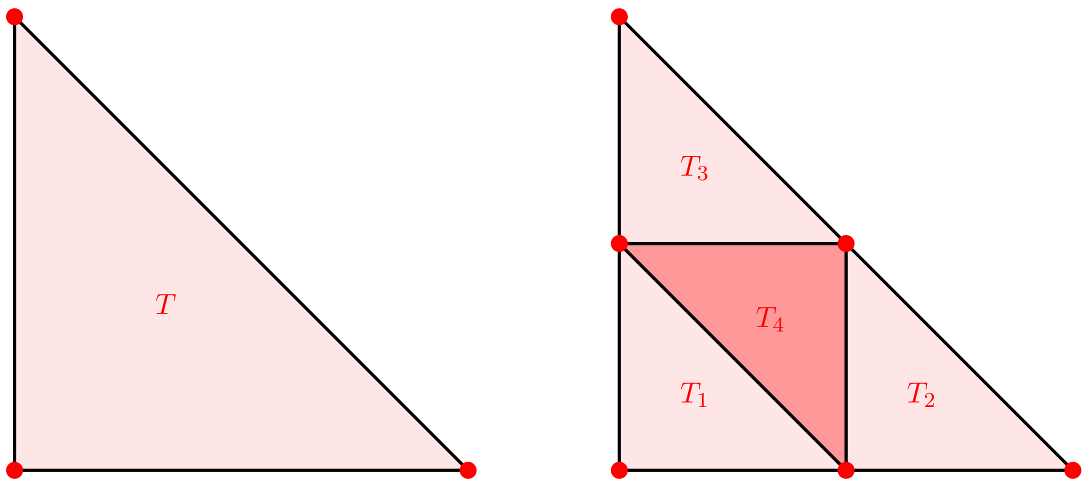

FIGURE 3.5. *Red-refinement refines the element $T \in \mathcal{T}^{(\mathrm{old})}$ into 4 similar elements $T_1, \dots, T_4 \in \mathcal{T}^{(\mathrm{new})}$. The new nodes $\mathcal{K}^{(\mathrm{new})} \backslash \mathcal{K}^{(\mathrm{old})}$ are just the edge midpoints for all edges $E \in \mathcal{E}^{(\mathrm{old})}$. In particular, regularity of $\mathcal{T}^{(\mathrm{old})}$ implies regularity of $\mathcal{T}^{(\mathrm{new})}$.*

Exercise 13. Let $T = \mathrm{conv}\{z_1, z_2, z_3\}$ be a non-degenerate triangle in $\mathbb{R}^2$. Prove that the shape regularity constant h_T/ϱ_T tends to infinity if and only if the smallest angle in T tends to zero. $\square$

Exercise 14. Often, the shape regularity constant is defined as the maximal quotient h_T/r_T, where $r_T > 0$ denotes the maximal radius of a ball $B(x, r_T) := \{y \in \mathbb{R}^2 \,\big|\, |x - y| \le r_T\}$ inscribed in T, i.e., $B(x, r_T) \subseteq T$. Let $T = \mathrm{conv}\{z_1, z_2, z_3\}$ be a non-degenerate triangle in $\mathbb{R}^2$. What is the relation between ϱ_T and r_T? $\square$

3.2.2 Statement and Interpretation of Approximation Theorem

To state our first main result in this section, we need to know that certain Sobolev functions are at least continuous.

Theorem 3.4 (Sobolev). *Let Ω be a Lipschitz domain in $\mathbb{R}^d$ and $m > d/2$. Then, there holds the continuous inclusion $H^m(\Omega) \subseteq C(\overline{\Omega})$.* $\blacksquare$

In particular, for $d = 2, 3$, each Sobolev function $u \in H^2(\Omega)$ is continuous so that evaluation of u at the nodes $z \in \mathcal{K}$ is well-defined. Throughout the remaining section, we assume that $\mathcal{T}$ is a regular triangulation of a bounded Lipschitz domain $\Omega \subset \mathbb{R}^2$. We stress, however, that the same results — even with the same proofs — hold for $d = 3$ as well. As in the previous section, the nodal basis function corresponding to a node $z \in \mathcal{K}$ is denoted by $\zeta_z \in \mathcal{S}^1(\mathcal{T})$.

Theorem 3.5 (Approximation Theorem). *For $u \in H^2(\Omega)$, the* ***nodal interpolant*** *reads*

$$I_h u := \sum_{z \in \mathcal{K}} u(z)\zeta_z \in \mathcal{S}^1(\mathcal{T}). \tag{3.19}$$

For all $T \in \mathcal{T}$, there hold the elementwise error estimates

$$\|u - I_h u\|_{L^2(T)} \leq C\, \|h^2 D^2 u\|_{L^2(T)} \tag{3.20}$$

and

$$\|\nabla(u - I_h u)\|_{L^2(T)} \leq C\, \sigma(T)\, \|h D^2 u\|_{L^2(T)}, \tag{3.21}$$

where the generic constant $C > 0$ is independent of u, $\mathcal{T}$, and Ω, but depends only on the reference triangle. In particular, this proves for all $\alpha \in \mathbb{R}$ the global error estimates

$$\|h^\alpha(u - I_h u)\|_{L^2(\Omega)} \leq C\, \|h^{2+\alpha} D^2 u\|_{L^2(\Omega)} \tag{3.22}$$

and

$$\|h^\alpha \nabla(u - I_h u)\|_{L^2(\Omega)} \leq C\, \sigma(\mathcal{T})\, \|h^{1+\alpha} D^2 u\|_{L^2(\Omega)}. \tag{3.23}$$

This theorem will be shown later. First, we discuss the following immediate consequence:

Corollary 3.6. *For $u \in H^2(\Omega) \cap H^1_D(\Omega)$, it holds that $I_h u \in \mathcal{S}^1_D(\mathcal{T})$ and thus*

$$\min_{v_h \in \mathcal{S}^1_D(\mathcal{T})} \|u - v_h\|_{H^1(\Omega)} \leq \|u - I_h u\|_{H^1(\Omega)} \leq C\, \sigma(\mathcal{T})\, \|h D^2 u\|_{L^2(\Omega)}. \tag{3.24}$$

For $u \in H^2(\Omega) \cap H^1_*(\Omega)$, it holds that

$$\begin{aligned} \min_{v_h \in \mathcal{S}^1_*(\mathcal{T})} \|u - v_h\|_{H^1(\Omega)} = \min_{v_h \in \mathcal{S}^1(\mathcal{T})} \|u - v_h\|_{H^1(\Omega)} &\leq \|u - I_h u\|_{H^1(\Omega)} \\ &\leq C\, \sigma(\mathcal{T})\, \|h D^2 u\|_{L^2(\Omega)}. \end{aligned} \tag{3.25}$$

In either case, the constant $C > 0$ depends only on $\mathrm{diam}(\Omega)$.

Proof. Let $C_{\mathrm{apx}} > 0$ denote the constant from the approximation theorem. Then,

$$\|u - I_h u\|^2_{H^1(\Omega)} = \|u - I_h u\|^2_{L^2(\Omega)} + \|\nabla(u - I_h u)\|^2_{L^2(\Omega)} \leq C^2_{\mathrm{apx}}(\mathrm{diam}(\Omega)^2 + \sigma(\mathcal{T})^2)\|h D^2 u\|^2_{L^2(\Omega)}.$$

Since $\sigma(\mathcal{T}) \geq 1$, we obtain that

$$\|u - I_h u\|_{H^1(\Omega)} \leq C_{\mathrm{apx}} \sigma(\mathcal{T})(\mathrm{diam}(\Omega)^2 + 1)^{1/2}\, \|h D^2 u\|_{L^2(\Omega)}.$$

For $u \in H^2(\Omega) \cap H^1_D(\Omega)$, it holds that $u(z) = 0$ for all $z \in \overline{\Gamma}_D$. This implies that $I_h u \in \mathcal{S}^1_D(\mathcal{T})$ and hence (3.24). Before we prove (3.25), note that $I_h u \in \mathcal{S}^1(\mathcal{T})$ does not belong to $\mathcal{S}^1_*(\mathcal{T})$ in general. However, let $\mathbb{P}_h : H^1(\Omega) \to \mathcal{S}^1(\mathcal{T})$ denote the H^1-orthogonal projection onto $\mathcal{S}^1(\mathcal{T})$. Since $1 \in \mathcal{S}^1(\mathcal{T})$, it holds that

$$0 = \int_\Omega u\, dx = (u\,;1)_{H^1(\Omega)} = (\mathbb{P}_h u\,;1)_{H^1(\Omega)} = \int_\Omega \mathbb{P}_h u\, dx \quad \text{for all } u \in H^1_*(\Omega).$$

Therefore, $\mathbb{P}_h u \in \mathcal{S}^1_*(\mathcal{T})$, and the best approximation property of the orthogonal projection $\mathbb{P}_h$ thus implies that

$$\|u - \mathbb{P}_h u\|_{H^1(\Omega)} = \min_{v_h \in \mathcal{S}^1(\mathcal{T})} \|u - v_h\|_{H^1(\Omega)} \le \min_{v_h \in \mathcal{S}^1_*(\mathcal{T})} \|u - v_h\|_{H^1(\Omega)} \le \|u - \mathbb{P}_h u\|_{H^1(\Omega)}$$

and hence equality. As before, this proves (3.25). ■

Remark. Corollary 3.6 has two important consequences: First, according to Céa's lemma, the Galerkin error is up to a constant the best approximation error. For a smooth exact solution $u \in H^2(\Omega)$, the P1-FEM thus leads (at least and in fact even) to a convergence order $\mathcal{O}(h)$. Second, $C^\infty_D(\overline{\Omega})$ is dense in $H^1_D(\Omega)$ and $C^\infty_*(\overline{\Omega}) := \big\{v \in C^\infty(\overline{\Omega}) \,\big|\, \int_\Omega v\,dx = 0\big\}$ is dense in $H^1_*(\Omega)$. Corollary 3.6 therefore implies convergence of the Galerkin scheme on a dense subspace. The abstract framework provides convergence of the P1-FEM even without any regularity assumptions on u, cf. Proposition 1.7. □

Exercise 15. Use the Poincaré inequality and the Meyers-Serrin theorem to prove that $C^\infty_*(\overline{\Omega})$ is dense in $H^1_*(\Omega)$. □

3.2.3 Bramble-Hilbert Lemma

It now remains to prove the Approximation Theorem 3.5. The proof of which needs three lemmata. The first two lemmata provide the basis for general scaling arguments. We therefore state the results even in a slightly generalized setting.

Definition. For a multiindex $\alpha \in \mathbb{N}_0^d$ and $x \in \mathbb{R}^d$, we define the **monomial** $x^\alpha := \prod_{j=1}^d x_j^{\alpha_j}$, where $|\alpha| := \sum_{j=1}^d \alpha_j$ is the **(total) degree** of α. For a Lipschitz domain $T \subseteq \mathbb{R}^d$, we define

$$\mathcal{P}^m(T) := \big\{v : T \to \mathbb{R} \,\big|\, v \text{ is linear combination of monomials of degree } \le m\big\} \tag{3.26}$$

the space that consists of all **polynomials** of degree less than or equal to $m \in \mathbb{N}$.

Lemma 3.7 (Bramble-Hilbert). *For a Lipschitz domain $T \subset \mathbb{R}^d$ and a normed space X, let $A \in L(H^{m+1}(T); X)$ be a linear and continuous operator with $\mathcal{P}^m(T) \subseteq \ker(A)$. Besides the classical continuity estimate*

$$\|Av\|_X \le \|A\|\,\|v\|_{H^{m+1}(T)} \quad \text{for all } v \in H^{m+1}(T), \tag{3.27}$$

it holds that

$$\|Av\|_X \le C\,\|A\|\,\|D^{m+1}v\|_{L^2(T)} \quad \text{for all } v \in H^{m+1}(T), \tag{3.28}$$

where the constant $C > 0$ depends only on m and T.

Proof. **1. step.** Construct an equivalent norm on $H^{m+1}(T)$: Note that $\mathcal{P}^m(T)$ is a finite dimensional space. Let $\Pi : L^2(T) \to \mathcal{P}^m(T)$ denote the L^2-orthogonal projection onto $\mathcal{P}^m(T)$. We define

$$|||v||| := \|D^{m+1}v\|_{L^2(T)} + \|\Pi v\|_{L^2(T)} \quad \text{for } v \in H^{m+1}(T).$$

From $\|\Pi v\|_{L^2(T)} \le \|v\|_{L^2(T)}$, we infer that

$$|||v||| \le \|D^{m+1}v\|_{L^2(T)} + \|v\|_{L^2(T)} \le \sqrt{2}\,\|v\|_{H^{m+1}(T)}.$$

Next, we prove the converse inequality, i.e., there exists a constant $C > 0$ such that

$$\|v\|_{H^{m+1}(T)} \le C\,|||v||| \quad \text{for all } v \in H^{m+1}(T).$$

As above, we use the Rellich theorem and argue by contradiction: If the claim is wrong, we find $v_n \in H^{m+1}(T)$ such that $\|v_n\|_{H^{m+1}(T)} > n|||v_n|||$. We define $w_n := v_n/\|v_n\|_{H^{m+1}(T)}$. Note that

$$\|w_n\|_{H^{m+1}(T)} = 1 \quad \text{as well as} \quad |||w_n||| \le \frac{1}{n}.$$

According to reflexivity, we may thus assume that $w_n \rightharpoonup w \in H^{m+1}(T)$. According to Lemma 2.6, convexity and continuity of $|||\cdot|||$ imply that $|||w||| = 0$. Therefore, it holds that $D^{m+1}w = 0$ as well as $\Pi w = 0$. With the help of Exercise 16, we deduce that $w \in \mathcal{P}^m(T)$ and consequently $\|w\|_{L^2(T)} = \|\Pi w\|_{L^2(T)} = 0$. According to Rellich's theorem, we have $w_n \to w = 0 \in H^m(T)$. Since $D^{m+1}w_n \to 0 \in L^2(T)$, we even conclude that $w_n \to 0 = w \in H^{m+1}(T)$. This however, contradicts $\|w_n\|_{H^{m+1}(T)} = 1$. Altogether, we have shown that $|||\cdot|||$ is an equivalent norm on $H^{m+1}(T)$.

2. step. With the norm equivalence constant $C > 0$ of step 1, it holds that

$$\|Av\|_X = \|A(v - \Pi v)\|_X \le \|A\|\,\|v - \Pi v\|_{H^{m+1}(T)} \le C\,\|A\|\,|||v - \Pi v||| = C\,\|A\|\,\|D^{m+1}v\|_{L^2(T)}$$

for all $v \in H^{m+1}(T)$. ■

Exercise 16. Prove that a function $v \in H^{m+1}(T)$ on a bounded Lipschitz domain $T \subset \mathbb{R}^d$ satisfies $D^{m+1}v = 0$ if and only if $v \in \mathcal{P}^m(T)$. **Hint:** You should use the case $m = 0$ without a proof, cf. Theorem 2.3. □

3.2.4 Scaling Argument and Proof of Approximation Theorem

Lemma 3.8 (Transformation Formula). *Let $T, \widehat{T} \subset \mathbb{R}^d$ be Lipschitz domains. Let $\Phi(x) := Bx + y$ with regular matrix $B \in \mathbb{R}^{d\times d}$ and vector $y \in \mathbb{R}^d$ be an affine diffeomorphism with $\Phi(\widehat{T}) = T$. For $u \in H^m(T)$, it holds that $u \circ \Phi \in H^m(\widehat{T})$ with*

$$\|D^m(u\circ\Phi)\|_{L^2(\widehat{T})} \le |\det B|^{-1/2}\|B\|_F^m\|D^m u\|_{L^2(T)}, \tag{3.29}$$

where $\|B\|_F$ denotes the Frobenius norm of B. Moreover, for $m = 0$, there even holds equality.

Proof. **1. step.** The case $m = 0$: According to the transformation theorem and $D\Phi(x) = B$, it holds that

$$\|u\|_{L^2(T)}^2 = \int_T u^2\,dy = \int_{\widehat{T}} (u\circ\Phi)^2\,|\det D\Phi|\,dx = |\det B|\,\|u\circ\Phi\|_{L^2(\widehat{T})}^2.$$

2. step. To treat the higher-order case for smooth functions $u \in C^\infty(\overline{T})$, we first prove by induction on m that for all $j_\ell \in \{1, \dots, d\}$, it holds that

$$\partial_{j_1} \cdots \partial_{j_m}(u \circ \Phi)(x) = \sum_{k_1=1}^{d} \cdots \sum_{k_m=1}^{d} \partial_{k_1} \cdots \partial_{k_m} u(\Phi(x)) \prod_{\ell=1}^{m} B_{k_\ell j_\ell}, \tag{3.30}$$

which is the special case of the Faà di Bruno formula (chain rule for partial derivatives): The case $m = 1$ follows from the chain rule $D(u \circ \Phi)(x) = Du(\Phi(x))D\Phi(x) = Du(\Phi(x))B$, where, e.g., $Du(y) = (\partial_1 u, \dots, \partial_d u)(y)$. Therefore,

$$\partial_j(u \circ \Phi)(x) = \sum_{k=1}^{d} \partial_k u(\Phi(x))B_{kj}.$$

Assuming that (3.30) holds up to $m \in \mathbb{N}$, we now prove the equality for $m + 1$:

$$\begin{aligned}
\partial_{j_1} \cdots \partial_{j_{m+1}}(u \circ \Phi)(x) &\overset{!}{=} \partial_{j_1}\Big(\sum_{k_2=1}^{d} \cdots \sum_{k_{m+1}=1}^{d} \partial_{k_2} \cdots \partial_{k_{m+1}} u(\Phi(x)) \prod_{\ell=2}^{m+1} B_{k_\ell j_\ell}\Big) \\
&= \sum_{k_2=1}^{d} \cdots \sum_{k_{m+1}=1}^{d} \partial_{j_1}\big(\partial_{k_2} \cdots \partial_{k_{m+1}} u(\Phi(x))\big) \prod_{\ell=2}^{m+1} B_{k_\ell j_\ell} \\
&\overset{!}{=} \sum_{k_2=1}^{d} \cdots \sum_{k_{m+1}=1}^{d} \sum_{k_1=1}^{d} \partial_{k_1}\partial_{k_2} \cdots \partial_{k_{m+1}} u(\Phi(x))B_{k_1 j_1} \prod_{\ell=2}^{m+1} B_{k_\ell j_\ell} \\
&= \sum_{k_1=1}^{d} \cdots \sum_{k_{m+1}=1}^{d} \partial_{k_1}\partial_{k_2} \cdots \partial_{k_{m+1}} u(\Phi(x)) \prod_{\ell=1}^{m+1} B_{k_\ell j_\ell},
\end{aligned}$$

where we have used the induction hypothesis for m and the initial step $m = 1$. This verifies (3.30).

3. step. We apply the Cauchy inequality to (3.30) to see that

$$\begin{aligned}
\big|\partial_{j_1} \cdots \partial_{j_m}(u \circ \Phi)(x)\big|^2 &\le \Big(\sum_{k_1=1}^{d} \cdots \sum_{k_m=1}^{d} \big|\partial_{k_1} \cdots \partial_{k_m} u(\Phi(x))\big|^2 \Big)\Big(\sum_{k_1=1}^{d} \cdots \sum_{k_m=1}^{d} \Big|\prod_{\ell=1}^{m} B_{k_\ell j_\ell}\Big|^2 \Big) \\
&= \Big(\sum_{k_1=1}^{d} \cdots \sum_{k_m=1}^{d} \big|\partial_{k_1} \cdots \partial_{k_m} u(\Phi(x))\big|^2 \Big)\Big(\sum_{k_1=1}^{d} \cdots \sum_{k_m=1}^{d} \prod_{\ell=1}^{m} B_{k_\ell j_\ell}^2 \Big) \\
&\overset{!}{=} \Big(\sum_{k_1=1}^{d} \cdots \sum_{k_m=1}^{d} \big|\partial_{k_1} \cdots \partial_{k_m} u(\Phi(x))\big|^2 \Big)\Big(\prod_{\ell=1}^{m} \sum_{k_\ell=1}^{d} B_{k_\ell j_\ell}^2 \Big),
\end{aligned}$$

where the last equality follows from another simple induction argument.

4. step. We prove the transformation formula (3.29) for $u \in C^\infty(\overline{T})$:

$$
\begin{aligned}
|\det B| \|D^m(u\circ\Phi)\|^2_{L^2(\widehat{T})} &= \int_{\widehat{T}} \sum_{j_1=1}^d \cdots \sum_{j_m=1}^d |\partial_{j_1}\cdots\partial_{j_m}(u\circ\Phi)(x)|^2 |\det D\Phi(x)|\,dx \\
&\le \underbrace{\Big(\sum_{j_1=1}^d \cdots \sum_{j_m=1}^d \prod_{\ell=1}^m \sum_{k_\ell=1}^d B^2_{k_\ell j_\ell}\Big)}_{=\prod_{\ell=1}^m \sum_{j_\ell=1}^d \sum_{k_\ell=1}^d B^2_{k_\ell j_\ell}} \underbrace{\Big(\int_{\widehat{T}} \sum_{k_1=1}^d \cdots \sum_{k_m=1}^d \big|\partial_{k_1}\cdots\partial_{k_m} u(\Phi(x))\big|^2 |\det D\Phi(x)|\,dx\Big)}_{=\|D^m u\|^2_{L^2(T)}} \\
&= \|B\|_F^{2m} \|D^m u\|^2_{L^2(T)}.
\end{aligned}
$$

5. step. We prove the transformation formula (3.29) for general $u \in H^m(T)$: According to the Meyers-Serrin theorem, $C^\infty(\overline{T})$ is a dense subspace of $H^m(T)$. Note that (3.29) implies for $u \in C^\infty(\overline{T})$ the estimate $\|u\circ\Phi\|_{H^m(\widehat{T})} \le C\,\|u\|_{H^m(T)}$, where $C > 0$ depends only on m and B. Hence, $\Psi u := u\circ\Phi$ extends uniquely to a linear and continuous mapping $\Psi : H^m(T) \to H^m(\widehat{T})$. For $u \in H^m(T)$, choose $(u_n) \subset C^\infty(\overline{T})$ with $u_n \to u \in H^m(T)$. By continuity of Ψ, it holds that $u_n\circ\Phi = \Psi u_n \to \Psi u$ in $H^m(\widehat{T})$. Moreover, according to step 1, it holds that $u_n\circ\Phi \to u\circ\Phi \in L^2(\widehat{T})$. This implies that $u\circ\Phi = \Psi u \in H^m(\widehat{T})$, i.e., the (unique) extension of Ψ from $C^\infty(\overline{T})$ to $H^m(T)$ is, in fact, the composition. Moreover, the left-hand side and the right-hand side of (3.29) depend continuously (with respect to $H^m(T)$) on u. This and (3.29) for $u_n \in C^\infty(\overline{T})$ prove that

$$
\begin{aligned}
\|D^m(u\circ\Phi)\|_{L^2(\widehat{T})} = \lim_{n\to\infty} \|D^m(u_n\circ\Phi)\|_{L^2(\widehat{T})} &\le \lim_{n\to\infty} |\det B|^{-1/2}\|B\|_F^m \|D^m u_n\|_{L^2(T)} \\
&= |\det B|^{-1/2}\|B\|_F^m \|D^m u\|_{L^2(T)}
\end{aligned}
$$

and conclude the proof. ■

Lemma 3.9. *For $\widehat{T} = T_{\mathrm{ref}}$ the reference element and $T = \mathrm{conv}\{z_1, z_2, z_3\} \subset \mathbb{R}^2$ being a non-degenerate triangle, we define*

$$
\Phi_T : T_{\mathrm{ref}} \to T, \quad \Phi_T(s,t) := z_1 + B\begin{pmatrix} s \\ t \end{pmatrix}, \quad \textit{where } B := \begin{pmatrix} z_2 - z_1 & z_3 - z_1 \end{pmatrix} \in \mathbb{R}^{2\times 2}. \tag{3.31}
$$

Then, it holds that $|\det B| = 2\,|T|$ and

$$
h_T/\sqrt{2} \le \|B\|_F \le \sqrt{2}\,h_T \quad \textit{as well as} \quad \varrho_T^{-1}/\sqrt{2} \le \|B^{-1}\|_F \le \sqrt{2}\,\varrho_T^{-1}. \tag{3.32}
$$

Proof. It holds that

$$
\|B\|_F^2 = |z_2 - z_1|^2 + |z_3 - z_1|^2 \le 2h_T^2.
$$

Moreover,

$$
|z_3 - z_2| \le |z_3 - z_1| + |z_2 - z_1| \le \sqrt{2}\,\big(|z_3 - z_1|^2 + |z_2 - z_1|^2\big)^{1/2} \le \sqrt{2}\,\|B\|_F.
$$

In particular, $h_T = \max\{|z_2 - z_1|, |z_3 - z_1|, |z_3 - z_2|\} \le \sqrt{2}\,\|B\|_F$. The transformation theorem gives

$$
\frac{1}{2}\,|\det B| = |T_{\mathrm{ref}}|\,|\det B| = \int_{T_{\mathrm{ref}}} |\det D\Phi_T|\,dx = \int_T dx = |T| > 0.
$$

Hence, $0 < |\det B| = 2\,|T| = h_T \varrho_T$. In particular, B^{-1} as well as ϱ_T^{-1} are well-defined. It holds that

$$B^{-1} = \frac{1}{\det B} \begin{pmatrix} b_{22} & -b_{12} \\ -b_{21} & b_{11} \end{pmatrix} \quad \text{for} \quad B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}.$$

In particular, this proves that

$$\|B^{-1}\|_F = \frac{\|B\|_F}{|\det B|} = \frac{\|B\|_F}{h_T \varrho_T},$$

and the second estimate in (3.32) follows from the first. ■

Proof of Approximation Theorem 3.5. **1. step.** Estimate on the reference element T_{ref}: Let $I_h^{\mathrm{ref}} : H^2(T_{\mathrm{ref}}) \to \mathcal{P}^1(T_{\mathrm{ref}})$ denote the nodal interpolation operator on the reference element. We consider the operator

$$A := 1 - I_h^{\mathrm{ref}} : H^2(T_{\mathrm{ref}}) \to H^k(T_{\mathrm{ref}}) \quad \text{for } k = 0, 1$$

and observe that $\mathcal{P}^1(T_{\mathrm{ref}}) \subseteq \ker(A)$. To see that A is continuous, we estimate

$$\|Av\|_{H^k(T_{\mathrm{ref}})} \le \|v\|_{H^2(T_{\mathrm{ref}})} + \|I_h^{\mathrm{ref}} v\|_{H^k(T_{\mathrm{ref}})}.$$

Let z_1, z_2, z_3 denote the nodes of the reference element. Since all norms on the finite dimensional space $\mathcal{P}^1(T_{\mathrm{ref}})$ are equivalent, we use the Sobolev inequality to see that

$$\|I_h^{\mathrm{ref}} v\|_{H^k(T_{\mathrm{ref}})} \le C_{\mathrm{norm}} \max_{j=1,\dots,3} |I_h^{\mathrm{ref}} v(z_j)| \le C_{\mathrm{norm}} \|v\|_{\infty, T_{\mathrm{ref}}} \le C_{\mathrm{norm}} C_{\mathrm{sobolev}} \|v\|_{H^2(T_{\mathrm{ref}})}.$$

Altogether, we obtain that $\|Av\|_{H^k(T_{\mathrm{ref}})} \le (1 + C_{\mathrm{norm}} C_{\mathrm{sobolev}}) \|v\|_{H^2(T_{\mathrm{ref}})}$, whence continuity of the operator A. Consequently, the Bramble-Hilbert lemma provides a constant $C_{\mathrm{ref}} > 0$ that depends only on T_{ref} with

$$\|v - I_h^{\mathrm{ref}} v\|_{H^k(T_{\mathrm{ref}})} \le C_{\mathrm{ref}} \|D^2 v\|_{L^2(T_{\mathrm{ref}})} \quad \text{for all } v \in H^2(T_{\mathrm{ref}}) \text{ and } k = 0, 1.$$

2. step. Scaling arguments provide the estimate on each element T: Let $\Phi = \Phi_T$ denote the affine diffeomorphism from Lemma 3.9. Note that $I_h^{\mathrm{ref}}(u \circ \Phi) = (I_h u) \circ \Phi$. Define $v := u \circ \Phi$ and observe that $(u - I_h u) \circ \Phi = (1 - I_h^{\mathrm{ref}}) v$. First, we apply the transformation formula to Φ^{-1},

$$\begin{aligned} \|D^k(u - I_h u)\|_{L^2(T)} &= \|D^k((v - I_h^{\mathrm{ref}} v) \circ \Phi^{-1})\|_{L^2(T)} \\ &\le |\det B^{-1}|^{-1/2} \|B^{-1}\|_F^k \|D^k(v - I_h^{\mathrm{ref}} v)\|_{L^2(T_{\mathrm{ref}})} \\ &\le C_{\mathrm{ref}} |\det B|^{1/2} \|B^{-1}\|_F^k \|D^2 v\|_{L^2(T_{\mathrm{ref}})}. \end{aligned}$$

Second, we plug-in $v = u \circ \Phi$ and apply the transformation formula to Φ,

$$\|D^2 v\|_{L^2(T_{\mathrm{ref}})} = \|D^2(u \circ \Phi)\|_{L^2(T_{\mathrm{ref}})} \le |\det B|^{-1/2} \|B\|_F^2 \|D^2 u\|_{L^2(T)}.$$

The combination of the last two estimates proves that

$$\|D^k(u - I_h u)\|_{L^2(T)} \le C_{\mathrm{ref}} \|B^{-1}\|_F^k \|B\|_F^2 \|D^2 u\|_{L^2(T)} \le C_{\mathrm{ref}}\, 2^{(k+2)/2} h_T^2 \varrho_T^{-k} \|D^2 u\|_{L^2(T)},$$

where we have used the geometric interpretation of $\|B\|_F$ and $\|B^{-1}\|_F$. This proves that

$$\|u - I_h u\|_{L^2(T)} \le 2C_{\mathrm{ref}}\|h^2 D^2 u\|_{L^2(T)} \quad\text{and}\quad \|\nabla(u - I_h u)\|_{L^2(T)} \le 2^{3/2} C_{\mathrm{ref}}\,\sigma(\mathcal{T})\|hD^2 u\|_{L^2(T)}.$$

and thus concludes the proof. ■

Remark. The proof of Theorem 3.5 shows that it is enough to assume $u \in C(\overline{\Omega}) \cap H^2(\mathcal{T})$, where $H^k(\mathcal{T}) := \big\{u \in L^2(\Omega) \,\big|\, \forall T \in \mathcal{T} \quad u \in H^k(T)\big\}$ for $k \ge 1$. According to the Sobolev inequality, it holds that $H^2(\Omega) \subseteq C(\overline{\Omega}) \cap H^2(\mathcal{T})$. For the *broken Sobolev spaces* $H^k(\mathcal{T})$, we write $D_h^k v$ for the $\mathcal{T}$-piecewise k-th derivative of v and, in particular, $\nabla_h v = D_h^1 v$ for the $\mathcal{T}$-piecewise gradient. □

Remark. We recall the procedure of a scaling argument for proving an estimate. To that end, let $\Phi_T : T_{\mathrm{ref}} \to T$ be the affine diffeomorphism with linear part B.

- First, transfer the left-hand side from T to T_{ref}:

$$\begin{aligned}\|D^k v\|_{L^2(T)} = \|D^k(v \circ \Phi_T \circ \Phi_T^{-1})\|_{L^2(T)} &\le |\det B^{-1}|^{-1/2}\,\|B^{-1}\|_F^k\,\|D^k(v \circ \Phi_T)\|_{L^2(T_{\mathrm{ref}})} \\ &\simeq |T|\,\varrho_T^{-k}\,\|D^k(v \circ \Phi_T)\|_{L^2(T_{\mathrm{ref}})},\end{aligned}$$

  i.e., derivative on the left-hand side give rise to negative powers of ϱ_T.

- Second, prove an appropriate estimate on the reference element T_{ref}.

- Third, transfer the right-hand side from T_{ref} to T:

$$\|D^\ell(w \circ \Phi_T)\|_{L^2(T_{\mathrm{ref}})} \le |\det B|^{-1/2}\,\|B\|_F^\ell\;\|D^\ell w\|_{L^2(T)} \simeq |T|^{-1/2}\,h_T^\ell\;\|D^\ell w\|_{L^2(T)},$$

  i.e., derivatives on the right-hand side give rise to positive powers of h_T.

Plugging everything together, proves the desired estimate. □

Note that the heart of the proof of the approximation theorem is the Rellich theorem and thus a compactness argument. The following exercise shows that approximation results are necessarily proved by use of compactness.

Exercise 17. Let X be a Banach space and Y be a normed space with continuous inclusion $Y \subseteq X$. For $h \to 0$, let X_h be finite dimensional subspaces of X and $I_h \in L(Y; X_h)$ be a continuous and linear operator with

$$\|u - I_h u\|_X \le C h^\alpha\,\|u\|_Y \quad \text{for all } u \in Y,$$

where the constants $C, \alpha > 0$ are independent of u and h. Then, the continuous inclusion $Y \subseteq X$ is already compact. □