Part VI, Chapter 26

Basic error analysis

In Part VI, composed of Chapters 26 to 30, we introduce the Galerkin approximation technique and derive fundamental stability results and error estimates. We also investigate implementation aspects of the method (quadratures, linear algebra, assembling, storage). In this chapter, we consider the following problem, introduced in Chapter 25, and study its approximation by the Galerkin method:

$$\begin{cases} \text{Find } u \in V \text{ such that} \\ a(u,w) = \ell(w), \quad \forall w \in W. \end{cases} \tag{26.1}$$

Here, V and W are Banach spaces, a is a bounded sesquilinear form on $V \times W$, and ℓ is a bounded antilinear form on W. We focus on the well-posedness of the approximate problem, and we derive a bound on the approximation error in a simple setting. This bound is known in the literature as Céa's lemma. We also characterize the well-posedness of the discrete problem by using the notion of Fortin operator.

To stay general, we consider complex vector spaces. The case of real vector spaces is recovered by replacing the field $\mathbb{C}$ by $\mathbb{R}$, by removing the real part symbol $\Re(\cdot)$ and the complex conjugate symbol $\bar{\cdot}$, and by interpreting the symbol $|\cdot|$ as the absolute value instead of the modulus. Moreover, sesquilinear forms become bilinear forms, and antilinear forms are just linear forms. We denote by α and $\|a\|_{V\times W}$ the inf-sup and the boundedness constants of the sesquilinear form a on $V \times W$, i.e.,

$$\alpha := \inf_{v\in V} \sup_{w\in W} \frac{|a(v,w)|}{\|v\|_V \|w\|_W} \le \sup_{v\in V} \sup_{w\in W} \frac{|a(v,w)|}{\|v\|_V \|w\|_W} =: \|a\|_{V\times W}. \tag{26.2}$$

We assume that (26.1) is well-posed, i.e., $0 < \alpha$ and $\|a\|_{V\times W} < \infty$. Whenever the context is unambiguous, we write $\|a\|$ instead of $\|a\|_{V\times W}$.

26.1 The Galerkin method

The central idea in the Galerkin method is to replace in (26.1) the infinite-dimensional spaces V and W by *finite-dimensional* spaces V_h and W_h (we always assume that $V_h \neq \{0\}$ and $W_h \neq \{0\}$). The subscript $h \in \mathcal{H}$ refers to the fact that these spaces are constructed as explained in Volume I using finite elements and a mesh $\mathcal{T}_h$ belonging to some sequence of meshes $(\mathcal{T}_h)_{h\in\mathcal{H}}$. The discrete problem takes the following form:

$$\begin{cases} \text{Find } u_h \in V_h \text{ such that} \\ a_h(u_h, w_h) = \ell_h(w_h), \quad \forall w_h \in W_h, \end{cases} \tag{26.3}$$

where a_h is a bounded sesquilinear form on $V_h \times W_h$ and ℓ_h is a bounded antilinear form on W_h. Notice that a_h and ℓ_h possibly differ from a and ℓ, respectively. Since the spaces V_h and W_h are finite-dimensional, (26.3) is called *discrete problem*. The space V_h is called *discrete trial space* (or *discrete solution space*), and W_h *discrete test space*.

Definition 26.1 (Standard Galerkin, Petrov–Galerkin). *The discrete problem* (26.3) *is called* standard Galerkin approximation *when* $W_h = V_h$ *and* Petrov–Galerkin approximation *otherwise.*

Definition 26.2 (Conforming setting). *The approximation is said to be* conforming *if* $V_h \subset V$ *and* $W_h \subset W$.

There are circumstances when considering nonconforming approximations is useful. Two important examples are discontinuous Galerkin methods where discrete functions are discontinuous across the mesh interfaces (see Chapters 38 and 60) and boundary penalty methods where boundary conditions are enforced weakly (see Chapters 37 for elliptic PDEs and Chapters 57–59 for Friedrichs' systems). Very often, nonconforming approximations make it necessary to work with discrete forms that differ from their continuous counterparts. For instance, the bilinear form $\int_D \nabla v{\cdot}\nabla w \,\mathrm{d}x$ does not make sense if the functions v and w are discontinuous. Another important example leading to a modification of the forms at the discrete level is the use of quadratures (see Chapter 30).

26.2 Discrete well-posedness

Our goal in this section is to study the well-posedness of the discrete problem (26.3). We equip V_h and W_h with norms denoted by $\|\cdot\|_{V_h}$ and $\|\cdot\|_{W_h}$, respectively. These norms can differ from those of V and W. One reason can be that the approximation is nonconforming and the norm $\|\cdot\|_V$ is meaningless on V_h. This is the case for instance if the norm $\|\cdot\|_V$ includes the H^1-norm and the discrete functions are allowed to jump across the mesh interfaces.

26.2.1 Discrete Lax–Milgram

Lemma 26.3 (Discrete Lax–Milgram). *Let V_h be a finite-dimensional space. Assume that $W_h = V_h$ in (26.3). Let a_h be a bounded sesquilinear form on $V_h \times V_h$ and let $\ell_h \in V_h'$. Assume that a_h is* coercive *on V_h, i.e., there is a real number $\alpha_h > 0$ and a complex number ξ with $|\xi| = 1$ such that*

$$\Re\left(\xi a_h(v_h, v_h)\right) \geq \alpha_h \|v_h\|_{V_h}^2, \qquad \forall v_h \in V_h. \tag{26.4}$$

Then (26.3) is well-posed with the a priori estimate $\|u_h\|_{V_h} \leq \frac{1}{\alpha_h}\|\ell_h\|_{V_h'}$.

Proof. A simple proof just consists of invoking the Lax–Milgram lemma (see Lemma 25.2). We now propose an elementary proof that relies on V_h being finite-dimensional. Let $A_h : V_h \to V_h'$ be the linear operator such that $\langle A_h(v_h), w_h\rangle_{V_h', V_h} := a_h(v_h, w_h)$ for all $v_h, w_h \in V_h$. Problem (26.3) amounts to seeking $u_h \in V_h$ such that $A_h(u_h) = \ell_h$ in V_h'. Hence, (26.3) is well-posed iff A_h is an isomorphism. Since $\dim(V_h) = \dim(V_h') < \infty$ this is equivalent to require that A_h be injective, i.e., $\ker(A_h) = \{0\}$. Let $v_h \in \ker(A_h)$ so that $0 = \xi\langle A_h(v_h), v_h\rangle_{V_h', V_h} = \xi a_h(v_h, v_h)$. From coercivity, we deduce that $0 \geq \alpha_h \|v_h\|_{V_h}^2$, which proves that $v_h = 0$. Hence, $\ker(A_h) = \{0\}$, thereby proving that A_h is bijective. □

Example 26.4 (Sufficient condition). (26.4) holds true if $V_h \subset V$ (conformity), $a_h := a_{|V_h \times V_h}$, and a is coercive on $V \times V$. □

Remark 26.5 (Variational formulation). As in the continuous setting (see Proposition 25.8), if V_h is a real Hilbert space and if a_h is symmetric and coercive (with $\xi := 1$ and $W_h = V_h$), then u_h solves (26.3) iff u_h minimizes the functional $\mathfrak{E}_h(v_h) := \frac{1}{2}a_h(v_h, v_h) - \ell_h(v_h)$ over V_h. If $V_h \subset V$, $a_h := a_{|V_h \times V_h}$, and $\ell_h := \ell_{|V_h}$, then $\mathfrak{E}_h = \mathfrak{E}_{|V_h}$ ($\mathfrak{E}$ is the exact energy functional), and $\mathfrak{E}(u_h) \geq \mathfrak{E}(u)$ since u minimizes $\mathfrak{E}$ over the larger space V. □

26.2.2 Discrete BNB

Theorem 26.6 (Discrete BNB). *Let V_h, W_h be finite-dimensional spaces. Let a_h be a bounded sesquilinear form on $V_h \times W_h$ and let $\ell_h \in W_h'$. Then the problem (26.3) is well-posed iff*

$$\inf_{v_h \in V_h} \sup_{w_h \in W_h} \frac{|a_h(v_h, w_h)|}{\|v_h\|_{V_h}\|w_h\|_{W_h}} =: \alpha_h > 0, \tag{26.5a}$$

$$\dim(V_h) = \dim(W_h). \tag{26.5b}$$

(Recall that arguments in the above infimum and supremum are understood to be nonzero.) Moreover, we have the a priori estimate $\|u_h\|_{V_h} \leq \frac{1}{\alpha_h}\|\ell_h\|_{W_h'}$.

Proof. Let $A_h : V_h \to W_h'$ be the linear operator such that

$$\langle A_h(v_h), w_h\rangle_{W_h', W_h} := a_h(v_h, w_h), \qquad \forall (v_h, w_h) \in V_h \times W_h. \tag{26.6}$$

The well-posedness of (26.3) is equivalent to A_h being an isomorphism, which owing to the finite-dimensional setting and the rank nullity theorem, is equivalent to (i) $\ker(A_h) = \{0\}$ (i.e., A_h is injective) and (ii) $\dim(V_h) = \dim(W_h')$. Since $\dim(W_h) = \dim(W_h')$, (26.5b) is equivalent to (ii). Let us prove that (i) is equivalent to the inf-sup condition (26.5a). By definition, we have

$$\sup_{w_h\in W_h} \frac{|a_h(v_h,w_h)|}{\|w_h\|_{W_h}} = \sup_{w_h\in W_h} \frac{|\langle A_h(v_h),w_h\rangle_{W_h',W_h}|}{\|w_h\|_{W_h}} =: \|A_h(v_h)\|_{W_h'}.$$

Assume first that (26.5a) holds true and let $v_h \in V_h$ be s.t. $A_h(v_h) = 0$. Then we have $\alpha_h \|v_h\|_{V_h} \leq \|A_h(v_h)\|_{W_h'} = 0$, which shows that $v_h = 0$. Hence, (26.5a) implies the injectivity of A_h. Conversely, assume $\ker(A_h) = \{0\}$ and let us prove (26.5a). An equivalent statement of (26.5a) is that there is $n_0 \in \mathbb{N}^*$ such that for all $v_h \in V_h$ with $\|v_h\|_{V_h} = 1$, one has $\|A_h(v_h)\|_{W_h'} > \frac{1}{n_0}$. Reasoning by contradiction, consider a sequence $(v_{hn})_{n\in\mathbb{N}^*}$ in V_h with $\|v_{hn}\|_{V_h} = 1$ and $\|A_h(v_{hn})\|_{W_h'} \leq \frac{1}{n}$. Since V_h is finite-dimensional, its unit sphere is compact. Hence, there is $v_h \in V_h$ such that, up to a subsequence, $v_{hn} \to v_h$. The limit v_h satisfies $\|v_h\|_{V_h} = 1$ and $A_h(v_h) = 0$, i.e., $v_h \in \ker(A_h) = \{0\}$, which contradicts $\|v_h\|_{V_h} = 1$. Hence, the injectivity of A_h implies (26.5a). In conclusion, $\ker(A_h) = \{0\}$ iff (26.5a) holds true. Finally, the a priori estimate follows from $\alpha_h\|u_h\|_{V_h} \leq \|A_h(u_h)\|_{W_h'} = \|\ell_h\|_{W_h'}$. □

Remark 26.7 (Link with BNB theorem). Condition (26.5a) is identical to (BNB1) from Theorem 25.9 applied to (26.3), and it is equivalent to the following *inf-sup condition*:

$$\exists\alpha_h > 0, \qquad \alpha_h\|v_h\|_{V_h} \leq \sup_{w_h\in W_h} \frac{|a_h(v_h,w_h)|}{\|w_h\|_{W_h}}, \qquad \forall v_h \in V_h. \tag{26.7}$$

Condition (26.5b) seemingly differs from (BNB2) applied to (26.3), which reads

$$\forall w_h \in W_h, \quad [\, a_h(v_h,w_h) = 0,\ \forall v_h \in V_h \,] \implies [\, w_h = 0 \,]. \tag{26.8}$$

To see that (26.5b) is equivalent to (26.8) provided (26.5a) holds true, let us introduce the adjoint operator $A_h^* : W_h \to V_h'$ (note that the space W_h is reflexive since it is finite-dimensional) such that

$$\overline{\langle A_h^*(w_h), v_h\rangle_{V_h',V_h}} = a_h(v_h,w_h), \qquad \forall (v_h,w_h) \in V_h{\times}W_h. \tag{26.9}$$

Then (26.8) says that A_h^* is injective, and this statement is equivalent to (26.5b) if $\ker(A_h) = \{0\}$; see Exercise 26.1. In summary, when the setting is finite-dimensional, the key property guaranteeing well-posedness is (26.5a), whereas the other condition (26.5b) is very simple to verify. □

Remark 26.8 (A_h^*). A_h is an isomorphism iff A_h^* is an isomorphism; see Exercise 26.2. Moreover, owing to Lemma C.53 (note that the space V_h is reflexive since it is finite-dimensional), A_h and A_h^* satisfy the inf-sup condition (26.5a) with *the same constant* α_h, i.e.,

$$\inf_{v_h\in V_h}\sup_{w_h\in W_h}\frac{|\langle A_h(v_h),w_h\rangle_{W_h',W_h}|}{\|v_h\|_{V_h}\|w_h\|_{W_h}}=\inf_{w_h\in W_h}\sup_{v_h\in V_h}\frac{|\langle A_h(v_h),w_h\rangle_{W_h',W_h}|}{\|v_h\|_{V_h}\|w_h\|_{W_h}}. \tag{26.10}$$

Note that $\langle A_h(v_h),w_h\rangle_{W_h',W_h}=\overline{\langle A_h^*(w_h),v_h\rangle_{V_h',V_h}}$. As shown in Remark C.54, the identity (26.10) may fail if A_h is not an isomorphism. □

26.2.3 Fortin's lemma

We focus on a conforming approximation, i.e., $V_h\subset V$ and $W_h\subset W$, we equip the spaces V_h and W_h with the norms of V and W, respectively, and we assume that $a_h:=a_{|V_h\times W_h}$. Our goal is to devise a criterion to ascertain that a_h satisfies the inf-sup condition (26.5a). To this purpose, we would like to use the inf-sup condition (26.2) satisfied by a on $V\times W$. Unfortunately, this condition does not imply its discrete counterpart on $V_h\times W_h$. Since $V_h\subset V$, (26.2) implies that $\alpha\|v_h\|_V\le\sup_{w\in W}\frac{|a(v_h,w)|}{\|w\|_W}$ for all $v_h\in V_h$, but it is not clear that the bound still holds when restricting the supremum to the subspace W_h. The Fortin operator provides the missing ingredient.

Lemma 26.9 (Fortin). *Let V,W be Hilbert spaces and let a be a bounded sesquilinear form on $V\times W$. Let α and $\|a\|$ be the inf-sup and boundedness constants of a defined in* (26.2). *Let $V_h\subset V$ and let $W_h\subset W$ be equipped with the norms of V and W, respectively. Consider the following two statements:*

(i) *There exists a map $\Pi_h:W\to W_h$, called* Fortin operator *such that:* (i.a) $a(v_h,\Pi_h(w)-w)=0$ *for all* $(v_h,w)\in V_h\times W$; (i.b) *There is* $\gamma_{\Pi_h}>0$ *such that* $\gamma_{\Pi_h}\|\Pi_h(w)\|_W\le\|w\|_W$ *for all* $w\in W$.
(ii) *The discrete inf-sup condition* (26.5a) *holds true.*

Then (i) $\Longrightarrow$ (ii) *with* $\alpha_h\ge\gamma_{\Pi_h}\alpha$. *Conversely,* (ii) $\Longrightarrow$ (i) *with* $\gamma_{\Pi_h}\ge\frac{\alpha_h}{\|a\|}$ *and Π_h can be constructed to be linear and idempotent ($\Pi_h\circ\Pi_h=\Pi_h$).*

Proof. (1) Let us assume (i). Let $\epsilon>0$. We have for all $v_h\in V_h$,

$$\begin{aligned}\sup_{w_h\in W_h}\frac{|a(v_h,w_h)|}{\|w_h\|_W}&\ge\sup_{w\in W}\frac{|a(v_h,\Pi_h(w))|}{\|\Pi_h(w)\|_W+\epsilon\|w\|_W}=\sup_{w\in W}\frac{|a(v_h,w)|}{\|\Pi_h(w)\|_W+\epsilon\|w\|_W}\\&\ge\gamma_{\Pi_h}\sup_{w\in W}\frac{|a(v_h,w)|}{\|w\|_W(1+\epsilon\gamma_{\Pi_h})}\ge\frac{\gamma_{\Pi_h}}{1+\epsilon\gamma_{\Pi_h}}\alpha\,\|v_h\|_V,\end{aligned}$$

since a satisfies (BNB1) and $V_h\subset V$. This proves (26.5a) with $\alpha_h\ge\gamma_{\Pi_h}\alpha$ since ϵ can be taken arbitrarily small. (Since Π_h cannot be injective, we introduced $\epsilon>0$ to avoid dividing by zero whenever $w\in\ker(\Pi_h)$.)
(2) Conversely, let us assume that a satisfies (26.5a). Let $A_h:V_h\to W_h'$ be defined in (26.6). Condition (26.5a) means that $\|A_h(v_h)\|_{W_h'}\ge\alpha_h\|v_h\|_V$ for all $v_h\in V_h$ ($\|\cdot\|_{W_h'}$ should not be confused with $\|\cdot\|_{W'}$). Hence, the operator

$B := A_h$ satisfies the assumptions of Lemma C.44 with $Y := V_h$, $Z := W_h'$, and $\beta := \alpha_h$. We infer that $A_h^* : W_h \to V_h'$ has a (linear) right inverse $A_h^{*\dagger} : V_h' \to W_h$ such that $\|A_h^{*\dagger}\|_{\mathcal{L}(V_h',W_h)} \leq \alpha_h^{-1}$. Let us now consider the operator $B : W \to V_h'$ s.t. $\langle B(w), v_h \rangle_{V_h',V_h} := \overline{a(v_h,w)}$ for all $(v_h,w) \in V_h \times W$, and let us set $\Pi_h := A_h^{*\dagger} \circ B : W \to W_h$. We have

$$a(v_h, \Pi_h(w)) = \langle A_h(v_h), A_h^{*\dagger}(B(w)) \rangle_{W_h',W_h} = \overline{\langle B(w), v_h \rangle_{V_h',V_h}} = a(v_h, w),$$

so that $a(v_h, \Pi_h(w) - w) = 0$. Moreover, we have $\|\Pi_h(w)\|_W \leq \frac{\|a\|}{\alpha_h} \|w\|_W$ since $\|A_h^{*\dagger}\|_{\mathcal{L}(V_h';W_h)} = \alpha_h^{-1}$ and $\|B\|_{\mathcal{L}(W;V_h')} \leq \|a\|$. Finally, since $B_{|W_h} = A_h^*$, we have $\Pi_h \circ \Pi_h = (A_h^{*\dagger} \circ B) \circ (A_h^{*\dagger} \circ B) = A_h^{*\dagger} \circ (A_h^* \circ A_h^{*\dagger}) \circ B = \Pi_h$, which proves that Π_h is idempotent. □

Remark 26.10 (Dimension, equivalence). We did not assume that V_h and W_h have the same dimension. This level of generality is useful to apply Lemma 26.9 to mixed finite element approximations; see Chapter 50. The implication (i) $\Longrightarrow$ (ii) in Lemma 26.9 is known in the literature as Fortin's lemma [201], and is useful to analyze mixed finite element approximations (see, e.g., Chapter 54 on the Stokes equations). The converse implication can be found in Girault and Raviart [217, p. 117]. This statement is useful in the analysis of Petrov–Galerkin methods; see Carstensen et al. [111], Muga and van der Zee [308], and also Exercise 50.7. Note that the gap in the stability constant γ_{Π_h} between the direct and the converse statements is equal to the condition number $\kappa(a) := \frac{\|a\|}{\alpha}$ of the sesquilinear form a (see Remark 25.12). Finally, we observe that the Fortin operator is not uniquely defined. □

Remark 26.11 (Banach spaces). Lemma 26.9 can be extended to Banach spaces. Such a construction is done in [187], where Lemma C.42 is invoked to build a (bounded) right inverse of A_h^*, and where the proposed map Π_h is nonlinear. Whether one can always construct a Fortin operator Π_h that is linear in Banach spaces seems to be an open question. □

26.3 Basic error estimates

In this section, we assume that the exact problem (26.1) and the discrete problem (26.3) are well-posed. Our goal is to bound the approximation error $(u - u_h)$ in the simple setting where the approximation is conforming ($V_h \subset V$, $W_h \subset W$, $a_h := a_{|V_h \times W_h}$, and $\ell_h := \ell_{|W_h}$).

26.3.1 Strong consistency: Galerkin orthogonality

The starting point of the error analysis is to make sure that the discrete problem (26.3) is *consistent* with the original problem (26.1). Loosely speaking one way of checking consistency is to insert the exact solution into the discrete problem and to verify that the discrepancy is small. We say that there is *strong consistency* whenever this operation is possible and the discrepancy is actually zero. A more general definition of consistency is given in the next chapter. The following result, known as the Galerkin orthogonality property, expresses the fact that strong consistency holds true in the present setting.

Lemma 26.12 (Galerkin orthogonality). *Assume that $V_h \subset V$, $W_h \subset W$, $a_h := a_{|V_h\times W_h}$, and $\ell_h := \ell_{|W_h}$. The following holds true:*

$$a(u, w_h) = \ell(w_h) = a(u_h, w_h), \qquad \forall w_h \in W_h. \tag{26.11}$$

In particular, we have $a(u - u_h, w_h) = 0$ for all $w_h \in W_h$.

Proof. The first equality follows from $W_h \subset W$ and the second one from $a_h := a_{|V_h\times W_h}$ and $\ell_h := \ell_{|W_h}$. □

26.3.2 Céa's and Babuška's lemmas

Lemma 26.13 (Céa). *Assume that $W_h = V_h \subset V = W$, $a_h := a_{|V_h\times V_h}$, and $\ell_h := \ell_{|V_h}$. Assume that the sesquilinear form a is V-coercive with constant $\alpha > 0$ and let $\|a\|$ be its boundedness constant defined in* (26.2) *(with $W = V$). Then the following error estimate holds true:*

$$\|u - u_h\|_V \le \frac{\|a\|}{\alpha} \inf_{v_h\in V_h} \|u - v_h\|_V. \tag{26.12}$$

Moreover, if the sesquilinear form a is Hermitian, the error estimate (26.12) *can be sharpened as follows:*

$$\|u - u_h\|_V \le \left(\frac{\|a\|}{\alpha}\right)^{\frac{1}{2}} \inf_{v_h\in V_h} \|u - v_h\|_V. \tag{26.13}$$

Proof. Invoking the coercivity of a (stability), followed by the Galerkin orthogonality property (strong consistency) and the boundedness of a, gives

$$\begin{aligned}\alpha \|u - u_h\|_V^2 &\le \Re(\xi a(u - u_h, u - u_h)) = \Re(\xi a(u - u_h, u - v_h)) \\ &\le \|a\| \, \|u - u_h\|_V \|u - v_h\|_V,\end{aligned}$$

for all v_h in V_h. This proves the error estimate (26.12). Assume now that the sesquilinear form a is Hermitian. Let v_h be arbitrary in V_h. Let us set $e := u - u_h$ and $\eta_h := u_h - v_h$. The Galerkin orthogonality property and the Hermitian symmetry of a imply that $a(e, \eta_h) = a(\eta_h, e) = 0$. Hence, we have

$$a(u - v_h, u - v_h) = a(e + \eta_h, e + \eta_h) = a(e, e) + a(\eta_h, \eta_h),$$

and the coercivity of a implies that $\Re(\xi a(e, e)) \le \Re(\xi a(u - v_h, u - v_h))$. Combining this bound with the stability and boundedness properties of a

yields

$$\begin{aligned}\alpha \|u - u_h\|_V^2 &\le \Re(\xi a(u-u_h, u-u_h)) = \Re(\xi a(e,e)) \\ &\le \Re(\xi a(u-v_h, u-v_h)) \le \|a\| \, \|u-v_h\|_V^2.\end{aligned}$$

Taking the infimum over $v_h \in V_h$ proves the error estimate (26.13). □

We now extend Céa's lemma to the more general case where stability relies on a discrete inf-sup condition rather than a coercivity argument. Thus, the discrete spaces V_h and W_h can differ.

Lemma 26.14 (Babuška). *Assume that $V_h \subset V$, $W_h \subset W$, $a_h := a_{|V_h \times W_h}$, $\ell_h := \ell_{|W_h}$, and $\dim(V_h) = \dim(W_h)$. Equip V_h and W_h with the norms of V and W, respectively. Assume the following discrete inf-sup condition:*

$$\inf_{v_h \in V_h} \sup_{w_h \in W_h} \frac{|a(v_h, w_h)|}{\|v_h\|_V \|w_h\|_W} =: \alpha_h > 0. \tag{26.14}$$

Let $\|a\|$ be the boundedness constant of a defined in (26.2)*. The following error estimate holds true:*

$$\|u - u_h\|_V \le \left(1 + \frac{\|a\|}{\alpha_h}\right) \inf_{v_h \in V_h} \|u - v_h\|_V. \tag{26.15}$$

Proof. Let $v_h \in V_h$. Using stability (i.e., (26.14)), strong consistency (i.e., the Galerkin orthogonality property), and the boundedness of a, we infer that

$$\begin{aligned}\alpha_h \|u_h - v_h\|_V &\le \sup_{w_h \in W_h} \frac{|a(u_h - v_h, w_h)|}{\|w_h\|_W} \\ &= \sup_{w_h \in W_h} \frac{|a(u - v_h, w_h)|}{\|w_h\|_W} \le \|a\| \, \|u - v_h\|_V,\end{aligned}$$

and (26.15) follows from the triangle inequality. □

The error estimates from Lemma 26.13 and from Lemma 26.14 are said to be *quasi-optimal* since $\inf_{v_h \in V_h} \|u - v_h\|_V$ is the best-approximation error of u by an element in V_h, and by definition $\|u - u_h\|_V$ cannot be smaller than the best-approximation error, i.e., the following two-sided error bound holds:

$$\inf_{v_h \in V_h} \|u - v_h\|_V \le \|u - u_h\|_V \le c \inf_{v_h \in V_h} \|u - v_h\|_V, \tag{26.16}$$

with $c := \frac{\|a\|}{\alpha}$ for Céa's lemma and $c := 1 + \frac{\|a\|}{\alpha_h}$ for Babuška's lemma. One noteworthy consequence of (26.16) is that $u_h = u$ whenever the exact solution turns out to be in V_h.

Corollary 26.15 (Convergence). *We have $\lim_{h \to 0} \|u - u_h\|_V = 0$ if the assumptions of Lemma* 26.14 *hold true together with the following properties:*

(i) Uniform stability: $\alpha_h \geq \alpha_0 > 0$ *for all* $h \in \mathcal{H}$.
(ii) Approximability: $\lim_{h\to 0}(\inf_{v_h\in V_h} \|v - v_h\|_V) = 0$ *for all* $v \in V$.

Proof. Direct consequence of the assumptions. □

Remark 26.16 (Céa). In the context of Céa's lemma, uniform stability follows from coercivity. Thus, approximability implies convergence. □

Remark 26.17 (Literature). Lemma 26.13 is derived in [114, Prop. 3.1] and is usually called Céa's lemma in the literature; see, e.g., Ciarlet [124, Thm. 2.4.1], Brenner and Scott [87, Thm. 2.8.1]. Lemma 26.14 is derived in Babuška [33, Thm. 2.2]. □

26.3.3 Approximability by finite elements

Let us present an important example where the approximability property identified in Corollary 26.15 holds true. Let $V := H^1(D)$ where D is a Lipschitz polyhedron in $\mathbb{R}^d$. Let $V_h := P_k^{\mathrm{g}}(\mathcal{T}_h) \subset H^1(D)$ be the H^1-conforming finite element space of degree $k \geq 1$ (see (20.1)), where $(\mathcal{T}_h)_{h\in\mathcal{H}}$ is a shape-regular sequence of affine meshes so that each mesh covers D exactly. One way to prove approximability is to consider the Lagrange interpolation operator or the canonical interpolation operator (see §19.3), i.e., let us set either $\mathcal{I}_h := \mathcal{I}_h^{\mathrm{L}}$ or $\mathcal{I}_h := \mathcal{I}_h^{\mathrm{g}}$ (we omit the subscript k for simplicity), so that $\mathcal{I}_h : V^{\mathrm{g}}(D) \to P_k^{\mathrm{g}}(\mathcal{T}_h)$ with domain $V^{\mathrm{g}}(D) := H^s(D)$, $s > \frac{d}{2}$ (see (19.19) with $p := 2$). Let l be the smallest integer s.t. $l > \frac{d}{2}$. Setting $r := \min(l-1, k)$, Corollary 19.8 with $m := 1$ (note that $r \geq 1$) implies that

$$\inf_{v_h\in V_h} \|v - v_h\|_{H^1(D)} \leq \|v - \mathcal{I}_h(v)\|_{H^1(D)} \leq c\, h^r \ell_D |v|_{H^{1+r}(D)},$$

for all $v \in H^{1+r}(D)$, where ℓ_D is a characteristic length of D, e.g., $\ell_D := \mathrm{diam}(D)$. Another possibility consists of using the quasi-interpolation operator $\mathcal{I}_h^{\mathrm{g,av}} : L^1(D) \to V_h$ from Chapter 22 since Theorem 22.6 implies that

$$\inf_{v_h\in V_h} \|v - v_h\|_{H^1(D)} \leq \|v - \mathcal{I}_h^{\mathrm{g,av}}(v)\|_{H^1(D)} \leq c\, h^r \ell_D |v|_{H^{1+r}(D)},$$

for all $v \in H^{1+r}(D)$ and all $r \in (0, k]$. We now establish approximability by invoking a density argument. Let $v \in V$ and let $\epsilon > 0$. Since $H^{1+r}(D)$ is dense in V for all $r > 0$, there is $v_\epsilon \in H^{1+r}(D)$ s.t. $\|v - v_\epsilon\|_{H^1(D)} \leq \epsilon$. Using the triangle inequality and the above interpolation estimates, we infer that

$$\begin{aligned}\inf_{v_h\in V_h} \|v - v_h\|_{H^1(D)} &\leq \|v - \mathcal{I}_h^{\mathrm{g,av}}(v_\epsilon)\|_{H^1(D)} \\ &\leq \|v - v_\epsilon\|_{H^1(D)} + \|v_\epsilon - \mathcal{I}_h^{\mathrm{g,av}}(v_\epsilon)\|_{H^1(D)} \\ &\leq \epsilon + c\, h^r \ell_D |v_\epsilon|_{H^{1+r}(D)}.\end{aligned}$$

Letting $h \to 0$ shows that $\limsup_{h\to 0}(\inf_{v_h\in V_h} \|v - v_h\|_{H^1(D)}) \leq \epsilon$, and since $\epsilon > 0$ is arbitrary, we conclude that approximability holds true, i.e., the best-approximation error in V_h of any function $v \in V$ tends to zero as $h \to 0$. The above arguments can be readily adapted when homogeneous Dirichlet conditions are strongly enforced.

26.3.4 Sharper error estimates

We now sharpen the constant appearing in the error estimate (26.15) from Lemma 26.14. Let $V_h \subset V$ and $W_h \subset W$ with $\dim(V_h) = \dim(W_h)$, and let a be a bounded sesquilinear form on $V \times W$ satisfying the discrete inf-sup condition (26.14) on $V_h \times W_h$. We define the *discrete solution map* $G_h : V \to V_h$ s.t. for all $v \in V$, $G_h(v)$ is the unique element in V_h satisfying

$$a(G_h(v) - v, w_h) = 0, \qquad \forall w_h \in W_h. \tag{26.17}$$

Note that $G_h(v)$ is well defined owing to the discrete inf-sup condition (26.14) and since $a(v, \cdot) : W_h \to \mathbb{C}$ is a bounded antilinear form on W_h. Moreover, G_h is linear and V_h is pointwise invariant under G_h.

Lemma 26.18 (Xu–Zikatanov). *Let* $\{0\} \subsetneq V_h \subsetneq V$ *and* $W_h \subset W$ *with* $\dim(V_h) = \dim(W_h)$ *where* V, W *are Hilbert spaces, and let* a *be a bounded sesquilinear form on* $V \times W$ *with constant* $\|a\|$ *defined in* (26.2) *satisfying the discrete inf-sup condition* (26.14) *on* $V_h \times W_h$ *with constant* α_h. *Then,*

$$\|u - u_h\|_V \le \frac{\|a\|}{\alpha_h} \inf_{v_h \in V_h} \|u - v_h\|_V. \tag{26.18}$$

Proof. Since G_h is linear and V_h is pointwise invariant under G_h, we have

$$u - u_h = u - G_h(u) = (u - v_h) - G_h(u - v_h),$$

for all $v_h \in V_h$. We infer that

$$\|u - u_h\|_V \le \|I - G_h\|_{\mathcal{L}(V)} \|u - v_h\|_V = \|G_h\|_{\mathcal{L}(V)} \|u - v_h\|_V,$$

where the last equality follows from the fact that in any Hilbert space H, any operator $T \in \mathcal{L}(H)$ such that $0 \neq T \circ T = T \neq I$ verifies $\|T\|_{\mathcal{L}(H)} = \|I - T\|_{\mathcal{L}(H)}$ (see the proof of Theorem 5.14). We can apply this result to the discrete solution map since $G_h \neq 0$ (since $V_h \neq \{0\}$), $G_h \circ G_h = G_h$ (since V_h is pointwise invariant under G_h), and $G_h \neq I$ (since $V_h \neq V$). To conclude the proof, we bound $\|G_h\|_{\mathcal{L}(V)}$ as follows: For all $v \in V$,

$$\alpha_h \|G_h(v)\|_V \le \sup_{w_h \in W_h} \frac{|a(G_h(v), w_h)|}{\|w_h\|_W} = \sup_{w_h \in W_h} \frac{|a(v, w_h)|}{\|w_h\|_W} \le \|a\| \, \|v\|_V,$$

which shows that $\|G_h\|_{\mathcal{L}(V)} \le \frac{\|a\|}{\alpha_h}$. □

Let Λ be the smallest c so that the inequality $\frac{\|u-u_h\|_V}{\inf_{v_h\in V_h}\|u-v_h\|_V} \le c$ holds for every $u \in V$. Then $\Lambda = \sup_{u\in V}\sup_{v_h\in V_h}\frac{\|u-G_h(u)\|_V}{\|u-v_h\|_V}$ since $u_h = G_h(u)$. But the proof of Lemma 26.18 shows that $\Lambda = \|I - G_h\|_{\mathcal{L}(V)} = \|G_h\|_{\mathcal{L}(V)}$. Hence, $\|G_h\|_{\mathcal{L}(V)}$ is the smallest constant such that the following quasi-optimal error estimate holds:

$$\|u - u_h\|_V \le \|G_h\|_{\mathcal{L}(V)} \inf_{v_h\in V_h} \|u - v_h\|_V.$$

Thus, sharp estimates on $\|G_h\|_{\mathcal{L}(V)}$ are important to determine whether the approximation error is close or not to the best-approximation error. The following result shows in particular that $\|G_h\|_{\mathcal{L}(V)}$ is, up to a factor in the interval $[\alpha, \|a\|]$, proportional to the inverse of the discrete inf-sup constant α_h.

Lemma 26.19 (Tantardini–Veeser). *Under the assumptions of Lemma* 26.18*, the following holds true:*

$$\|G_h\|_{\mathcal{L}(V)} = \sup_{w_h\in W_h} \frac{\left(\displaystyle\sup_{v\in V}\frac{|a(v,w_h)|}{\|v\|_V}\right)}{\left(\displaystyle\sup_{v_h\in V_h}\frac{|a(v_h,w_h)|}{\|v_h\|_V}\right)} \ge 1, \tag{26.19a}$$

$$\frac{\alpha}{\alpha_h} \le \|G_h\|_{\mathcal{L}(V)} \le \frac{\|a\|}{\alpha_h}. \tag{26.19b}$$

Proof. (1) Let $A \in \mathcal{L}(V;W')$ be the operator associated with the sesquilinear form a, i.e., $\langle A(v), w\rangle_{W',W} := a(v,w)$ for all $(v,w) \in V\times W$, and let $A^* \in \mathcal{L}(W;V')$ be its adjoint (where we used the reflexivity of W). We have

$$\alpha\,\|w\|_W \le \|A^*(w)\|_{V'} = \sup_{v\in V}\frac{|a(v,w)|}{\|v\|_V} \le \|a\|\,\|w\|_W, \tag{26.20}$$

for all $w \in W$. Indeed, the first bound follows from Lemma C.53 and the inf-sup stability of a, and the second one follows from the boundedness of a. This shows that the norms $\|\cdot\|_W$ and $\|A^*(\cdot)\|_{V'}$ are equivalent on W.
(2) Since $W_h \subset W$, we have $A^*(w_h) \in V'$ for all $w_h \in W_h$. Upon setting

$$\gamma_h := \inf_{w_h\in W_h}\sup_{v_h\in V_h}\frac{|a(v_h,w_h)|}{\|v_h\|_V\|A^*(w_h)\|_{V'}},$$

we have $\gamma_h \ge \frac{\alpha_h}{\|a\|} > 0$ owing to the inf-sup condition satisfied by a on $V_h\times W_h$, the norm equivalence (26.20), and Lemma C.53. Recalling that $\|A^*(w_h)\|_{V'} = \sup_{v\in V}\frac{|a(v,w_h)|}{\|v\|_V}$, the assertion (26.19a) amounts to $\|G_h\|_{\mathcal{L}(V)} = \gamma_h^{-1} \ge 1$.
(3) Let $w_h \in W_h$. Using the definition (26.17) of the discrete solution map and the definition of the dual norm $\|A^*(w_h)\|_{V'}$, we have

$$\|A^*(w_h)\|_{V'} = \sup_{v\in V} \frac{|a(G_h(v), w_h)|}{\|v\|_V} \le \sup_{v\in V} \frac{|a(G_h(v), w_h)|}{\|G_h(v)\|_V} \sup_{v\in V} \frac{\|G_h(v)\|_V}{\|v\|_V}$$
$$\le \sup_{v_h\in V_h} \frac{|a(v_h, w_h)|}{\|v_h\|_V} \|G_h\|_{\mathcal{L}(V)}.$$

Rearranging the terms and taking the infimum over $w_h \in W_h$ shows that $\gamma_h \ge \|G_h\|^{-1}_{\mathcal{L}(V)}$, i.e., $\|G_h\|_{\mathcal{L}(V)} \ge \gamma_h^{-1}$.
(4) Since $\gamma_h > 0$, Remark 26.8 implies that

$$\gamma_h = \inf_{v_h\in V_h} \sup_{w_h\in W_h} \frac{|a(v_h, w_h)|}{\|v_h\|_V \|A^*(w_h)\|_{V'}}. \tag{26.21}$$

Let $v \in V$. Applying the above identity to $G_h(v) \in V_h$, we infer that

$$\gamma_h \|G_h(v)\|_V \le \sup_{w_h\in W_h} \frac{|a(G_h(v), w_h)|}{\|A^*(w_h)\|_{V'}} = \sup_{w_h\in W_h} \frac{|a(v, w_h)|}{\|A^*(w_h)\|_{V'}} \le \|v\|_V,$$

since $|a(v, w_h)| = |\langle A^*(w_h), v\rangle_{V',V}| \le \|A^*(w_h)\|_{V'}\|v\|_V$. Taking the supremum over $v \in V$ shows that $\|G_h\|_{\mathcal{L}(V)} \le \gamma_h^{-1}$. Thus, we have proved that $\|G_h\|_{\mathcal{L}(V)} = \gamma_h^{-1}$, and the lower bound in (26.19a) is a direct consequence of $V_h \subset V$.
(5) It remains to prove (26.19b). Using the norm equivalence (26.20) in (26.21) to bound from below and from above $\|A^*(w_h)\|_{V'}$, we infer that

$$\frac{1}{\|a\|} \inf_{v_h\in V_h} \sup_{w_h\in W_h} \frac{|a(v_h, w_h)|}{\|v_h\|_V \|w_h\|_W} \le \gamma_h \le \frac{1}{\alpha} \inf_{v_h\in V_h} \sup_{w_h\in W_h} \frac{|a(v_h, w_h)|}{\|v_h\|_V \|w_h\|_W},$$

so that $\frac{\alpha_h}{\|a\|} \le \gamma_h \le \frac{\alpha_h}{\alpha}$, and (26.19b) follows from $\|G_h\|_{\mathcal{L}(V)} = \gamma_h^{-1}$. □

Remark 26.20 (Literature). Lemma 26.18 is proved in Xu and Zikatanov [397, Thm. 2], and Lemma 26.19 in Tantardini and Veeser [361, Thm. 2.1]. See also Arnold et al. [18] for the lower bound $\frac{\alpha}{\alpha_h} \le \|G_h\|_{\mathcal{L}(V)}$. □

Remark 26.21 (Discrete dual norm). For all $w_h \in W_h$, $A^*(w_h) \in V'$ can be viewed as a member of V_h' by restricting its action to the subspace $V_h \subset V$. We use the same notation and simply write $A^*(w_h) \in V_h'$. The statement (26.19a) in Lemma 26.19 can be rewritten as follows:

$$\|G_h\|_{\mathcal{L}(V)} = \sup_{w_h\in W_h} \frac{\|A^*(w_h)\|_{V'}}{\|A^*(w_h)\|_{V_h'}}, \tag{26.22}$$

where $\|A^*(w_h)\|_{V_h'} := \sup_{v_h\in V_h} \frac{|\langle A^*(w_h), v_h\rangle_{V',V}|}{\|v_h\|_V} = \sup_{v_h\in V_h} \frac{|a(v_h,w_h)|}{\|v_h\|_V}$. □

Example 26.22 (Orthogonal projection). Let $V \hookrightarrow L$ be two Hilbert spaces with continuous and dense embedding. Using the Riesz–Fréchet theorem (Theorem C.24), we identify L with its dual L' by means of the inner product $(\cdot,\cdot)_L$ in L. This allows us to define the continuous embedding

$E_{V'} : V \to V'$ s.t. $\langle E_{V'}(v), w\rangle_{V',V} := (v, w)_L$ for all $v, w \in V$. Note that $E_{V'}$ is self-adjoint. Consider a subspace $\{0\} \subsetneq V_h \subsetneq V$. Let $\mathcal{P}_h$ be the discrete solution map associated with the sesquilinear form $a(v, w) := \langle E_{V'}(v), w\rangle_{V',V}$ for all $v, w \in V$. Note that $\mathcal{P}_h$ is the L-orthogonal projection onto V_h since

$$(\mathcal{P}_h(v), w_h)_L = \langle E_{V'}(\mathcal{P}_h(v)), w_h\rangle_{V',V} := \langle E_{V'}(v), w_h\rangle_{V',V} = (v, w_h)_L,$$

for all $v \in V$ and all $w_h \in V_h$. Then Lemma 26.19 provides a precise estimate on the V-stability of $\mathcal{P}_h$ in the form

$$\|\mathcal{P}_h\|^{-1}_{\mathcal{L}(V)} = \inf_{w_h \in V_h} \frac{\|E_{V'}(w_h)\|_{V'_h}}{\|E_{V'}(w_h)\|_{V'}} = \inf_{w_h \in V_h} \sup_{v_h \in V_h} \frac{|(w_h, v_h)_L|}{\|E_{V'}(w_h)\|_{V'}\|v_h\|_V}. \tag{26.23}$$

See also Tantardini and Veeser [361, Prop. 2.5], Andreev [11, Lem. 6.2]. An important example is $V := H^1_0(D)$ and $L := L^2(D)$. The reader is referred to §22.5 for further discussion on the L^2-orthogonal projection onto conforming finite element spaces (see in particular Remark 22.23 for sufficient conditions on the underlying mesh to ensure H^1-stability). □

Exercises

Exercise 26.1 ((BNB2)). Prove that (26.8) is equivalent to (26.5b) provided (26.5a) holds true. (*Hint*: use that $\dim(W_h) = \mathrm{rank}(A_h) + \dim(\ker(A_h^*))$ (A_h^* is defined in (26.9)) together with the rank nullity theorem.)

Exercise 26.2 (Bijectivity of A_h^*). Prove that A_h is an isomorphism if and only if A_h^* is an isomorphism. (*Hint*: use $\dim(V_h) = \mathrm{rank}(A_h^*) + \dim(\ker(A_h))$ and $\dim(W_h) = \mathrm{rank}(A_h) + \dim(\ker(A_h^*))$.)

Exercise 26.3 (Petrov–Galerkin). Let V, W be real Hilbert spaces, let $A \in \mathcal{L}(V; W')$ be an isomorphism, and let $\ell \in W'$. Consider a conforming Petrov–Galerkin approximation with a finite-dimensional subspace $V_h \subset V$ and $W_h := (J_W^{\mathrm{RF}})^{-1} A V_h \subset W$, where $J_W^{\mathrm{RF}} : W \to W'$ is the Riesz–Fréchet isomorphism. The discrete bilinear form is $a_h(v_h, w_h) := \langle A(v_h), w_h\rangle_{W',W}$, and the discrete linear form is $\ell_h(w_h) := \ell(w_h)$ for all $v_h \in V_h$ and all $w_h \in W_h$. (i) Prove that the discrete problem (26.3) is well-posed. (ii) Show that its unique solution minimizes the residual functional $\mathfrak{R}(v) := \|A(v) - \ell\|_{W'}$ over V_h.

Exercise 26.4 (Fortin's lemma). (i) Prove that Π_h in the converse statement of Lemma 26.9 is idempotent. (*Hint*: prove that $B \circ A_h^{*\dagger} = I_{V'_h}$.) (ii) Assume that there are two maps $\Pi_{1,h}, \Pi_{2,h} : W \to W_h$ and two uniform constants $c_1, c_2 > 0$ such that $\|\Pi_{1,h}(w)\|_W \le c_1\|w\|_W$, $\|\Pi_{2,h}((I - \Pi_{1,h})(w))\|_W \le c_2\|w\|_W$ and $a(v_h, \Pi_{2,h}(w) - w) = 0$ for all $v_h \in V_h$, $w \in W$. Prove that $\Pi_h := \Pi_{1,h} + \Pi_{2,h}(I - \Pi_{1,h})$ is a Fortin operator. (iii) Write a variant of the direct statement in Lemma 26.9 assuming V, W reflexive, $A \in \mathcal{L}(V;W')$ bijective, and using this time an operator $\Pi_h : V \to V_h$ such that $a(\Pi_h(v)-v, w_h) = 0$ for all $(v, w_h) \in V \times W_h$ and $\gamma_{\Pi_h}\|\Pi_h(v)\|_V \leq \|v\|_V$ for all $v \in V$ for some $\gamma_{\Pi_h} > 0$. (*Hint*: use (26.10) and Lemma C.53.)

Exercise 26.5 (Compact perturbation). Let V, W be Banach spaces with W reflexive. Let $A_0 \in \mathcal{L}(V;W')$ be bijective, let $T \in \mathcal{L}(V;W')$ be compact, and assume that $A := A_0 + T$ is injective. Let $a_0(v, w) := \langle A_0(v), w\rangle_{W',W}$ and $a(v, w) := \langle A(v), w\rangle_{W',W}$ for all $(v, w) \in V \times W$. Let $V_h \subset V$ and $W_h \subset W$ be s.t. $\dim(V_h) = \dim(W_h)$ for all $h \in \mathcal{H}$. Assume that approximability holds, and that the sesquilinear form a_0 satisfies the inf-sup condition

$$\inf_{v_h \in V_h} \sup_{w_h \in W_h} \frac{|a_0(v_h, w_h)|}{\|v_h\|_V \|w_h\|_W} =: \alpha_0 > 0, \qquad \forall h \in \mathcal{H}.$$

Following Wendland [392], the goal is to show that there is $h_0 > 0$ s.t.

$$\inf_{v_h \in V_h} \sup_{w_h \in W_h} \frac{|a(v_h, w_h)|}{\|v_h\|_V \|w_h\|_W} =: \alpha > 0, \qquad \forall h \in \mathcal{H} \cap (0, h_0].$$

(i) Prove that $A \in \mathcal{L}(V;W')$ is bijective. (*Hint*: recall that a compact operator is bijective iff it is injective; this follows from the Fredholm alternative, Theorem 46.13.) (ii) Consider $R_h \in \mathcal{L}(V;V_h)$ s.t. for all $v \in V$, $R_h(v) \in V_h$ satisfies $a_0(R_h(v)-v, w_h) = 0$ for all $w_h \in W_h$. Prove that $R_h \in \mathcal{L}(V;V_h)$ and that $R_h(v)$ converges to v as $h \downarrow 0$ for all $v \in V$. (*Hint*: proceed as in the proof of Céa's lemma.) (iii) Set $L := I_V + A_0^{-1}T$ and $L_h := I_V + R_h A_0^{-1} T$ where I_V is the identity operator in V (observe that both L and L_h are in $\mathcal{L}(V)$). Prove that L_h converges to L in $\mathcal{L}(V)$. (*Hint*: use Remark C.5.) (iv) Show that if $h \in \mathcal{H}$ is small enough, L_h is bijective and there is C, independent of $h \in \mathcal{H}$, such that $\|L_h^{-1}\|_{\mathcal{L}(V)} \leq C$. (*Hint*: observe that $L^{-1}L_h = I_V - L^{-1}(L - L_h)$ and consider the Neumann series.) (v) Conclude.