Chapter 6

Mixed Problems

6.1 Abstract Analysis of Petrov-Galerkin Schemes

Recall that for a continuous linear operator $T \in L(X,Y)$, the **adjoint operator** $T^* : Y^* \to X^*$ is formally defined by

$$T^*y^* \in X^* \quad \text{with} \quad (T^*y^*)(x) := y^*(Tx) \quad \text{for all } y^* \in Y^* \text{ and } x \in X. \tag{6.1}$$

It is an easy application of the Hahn-Banach extension theorem that $T^* \in L(Y^*, X^*)$ even with the same operator norm $\|T\| = \|T^*\|$. We start this section with some easy, but later on important, observations.

Lemma 6.1. *Let X and Y be normed spaces and $T \in L(X,Y)$. Then, T is an isomorphism between X and* range(T) *if and only if*

$$\tau := \inf_{x \in X\backslash\{0\}} \frac{\|Tx\|_Y}{\|x\|_X} > 0. \tag{6.2}$$

In this case, there holds $\|T^{-1} : \mathrm{range}(T) \to X\| = 1/\tau$. Moreover, the range$(T)$ *is closed provided that X is a Banach space.*

Proof. Clearly, $T^{-1} : \mathrm{range}(T) \to X$ is well-defined (and hence an isomorphism in the sense of Linear Algebra) if and only if T is injective. If T is not injective, there exists some $x \neq 0$ with $Tx = 0$, and hence it follows $\tau = 0$. In particular, $\tau > 0$ implies that T is injective. By elementary calculations, we see

$$\tau = \inf_{x \in X\backslash\{0\}} \frac{\|Tx\|_Y}{\|x\|_X} = \inf_{y \in \mathrm{range}(T)\backslash\{0\}} \frac{\|y\|_Y}{\|T^{-1}y\|_X} = \frac{1}{\sup\limits_{y \in \mathrm{range}(T)\backslash\{0\}} \frac{\|T^{-1}y\|_X}{\|y\|_Y}}$$
$$= \frac{1}{\|T^{-1} : \mathrm{range}(T) \to X\|}$$

Hence, $\tau > 0$ implies $\|T^{-1} : \mathrm{range}(T) \to X\| = 1/\tau < \infty$, i.e., T^{-1} is even continuous. The same calculation proves that well-posedness and continuity of T^{-1} imply $\tau > 0$. Finally, suppose that X

is a Banach space and $\tau > 0$. Then, $\mathrm{range}(T)$ is a Banach space as well and hence, in particular, a closed subspace of Y. ■

According to the Hahn-Banach extension theorem, the **Hahn-Banach embedding**

$$I_X : X \to X^{**}, \quad (I_X x)(x^*) := x^*(x) \quad \text{for } x \in X \text{ and } x^* \in X^* \tag{6.3}$$

is an isometric linear operator, whence injective and continuous. A normed space X is **reflexive** provided that I_X is also surjective and thus an isometric isomorphism between X and X^{**}. We stress that

- reflexive spaces are, in particular, complete and thus Banach spaces,
- finite dimensional spaces are reflexive,
- all Hilbert spaces are reflexive,
- closed subspaces of reflexive spaces are also reflexive.

All of these facts are simple exercises left to the reader.

Theorem 6.2. *Let X and Y be reflexive Banach spaces over $\mathbb{R}$, and $T \in L(X, Y^*)$. Then, T is an isomorphism if and only if the following two conditions hold:*

- ***inf-sup condition*** $\tau := \inf\limits_{x \in X \setminus \{0\}} \sup\limits_{y \in Y \setminus \{0\}} \dfrac{(Tx)(y)}{\|x\|_X \|y\|_Y} > 0,$
- ***non-degeneracy condition*** $\forall y \in Y \setminus \{0\} \exists x \in X \quad (Tx)(y) \neq 0.$

*In this case, there holds $\|T^{-1}\| = 1/\tau$ for the operator norm of the inverse. The combination of inf-sup condition and non-degeneracy condition is called **LBB condition** in the literature, named after Ladyshenskaja, Babuška, and Brezzi.*

Proof. According to Lemma 6.1, $\tau > 0$ is equivalent to $T : X \to \mathrm{range}(T)$ being an isomorphism with closed range. It thus remains to show that $\mathrm{range}(T) = Y^*$ is equivalent to the non-degeneracy condition (ND). Assume there exists $y^* \in Y^* \setminus \mathrm{range}(T)$. The Hahn-Banach separation theorem implies the existence of a functional $\psi \in Y^{**}$ such that $\psi(y^*) = 1$ and $\psi|_{\mathrm{range}(T)} = 0$. With the identification of Y^{**} and Y, we obtain some $y \in Y \setminus \{0\}$ with $I_Y y = \psi$ and

$$0 = \psi(Tx) = (I_Y y)(Tx) = (Tx)(y) \quad \forall x \in X.$$

This contradicts the non-degeneracy condition (ND). We showed that ND implies $\mathrm{range}(T) = Y^*$. For the converse direction assume $\mathrm{range}(T) = Y^*$ and $y \in Y \setminus \{0\}$. There exists $y^* \in Y^*$ with $y^*(y) \neq 0$. Hence, we find $x \in X$ with $Tx = y^*$ and hence $(Tx)(y) \neq 0$. This concludes (ND) and hence the proof. ■

The following simple exercise proves that the assumptions on X in Theorem 6.2 are sharp.

Exercise 41. Let X be a normed space and Y be a reflexive Banach space over $\mathbb{R}$. Let $T \in L(X, Y^*)$ be an isomorphism. Prove that X is also a reflexive Banach space. **Hint:** It is known that a Banach space Z is reflexive, if and only if Z^* is reflexive. Moreover, Z is reflexive, if and only if each bounded sequence has a weakly convergent subsequence (i.e., the unit ball of Z is weakly compact). □

We now turn to continuous bilinear forms $a : X \times Y \to \mathbb{R}$ on normed spaces X and Y. So far, we only considered weak formulations of the type: Find $x \in X$ such that

$$a(x, \cdot) = x^* \in X^*, \tag{6.4}$$

where $a(\cdot, \cdot)$ is a continuous bilinear form on $X = Y$. For the classical Galerkin scheme, we assumed that $a(\cdot, \cdot)$ is even elliptic. Note that the last theorem provides a mathematical framework for weak formulations of the following type: Find $x \in X$ such that

$$a(x, \cdot) = y^* \in Y^*, \tag{6.5}$$

where $a(\cdot, \cdot)$ now is a continuous bilinear form $a : X \times Y \to \mathbb{R}$. In the literature, this approach is named after Petrov-Galerkin.

Corollary 6.3. *Let X and Y be real Banach spaces, where Y is reflexive. Let $a : X \times Y \to \mathbb{R}$ be bilinear and continuous. Then, the following statements* (i)–(ii) *are equivalent:*

(i) *For each $y^* \in Y^*$, exists a unique $x \in X$ with $a(x, \cdot) = y^*$.*

(ii) *The bilinear form satisfies the* **LBB condition***:*

- ***inf-sup condition*** $\alpha := \inf_{x \in X \backslash \{0\}} \sup_{y \in Y \backslash \{0\}} \dfrac{a(x, y)}{\|x\|_X \|y\|_Y} > 0,$
- ***non-degeneracy condition*** $\forall y \in Y \backslash \{0\} \exists x \in X \quad a(x, y) \neq 0.$

In this case, it holds

$$\alpha \, \|x\|_X \leq \|y^*\|_{Y^*} \leq \|a\| \, \|x\|_X, \tag{6.6}$$

where $\|a\| := \sup_{\substack{x \in X \backslash \{0\} \\ y \in Y \backslash \{0\}}} \dfrac{a(x, y)}{\|x\|_X \|y\|_Y}$ *denotes the continuity bound of $a(\cdot, \cdot)$.*

Proof. We associate with $a(\cdot, \cdot)$ the operator $T \in L(X, Y^*)$ given by $Tx = a(x, \cdot)$. Note that (i) is equivalent to the fact that T is an isomorphism (according to the open mapping theorem). According to Theorem 6.2, the latter is characterized by the LBB condition for T which, in fact, coincides with that for $a(\cdot, \cdot)$. For given $y^* \in Y^*$ and $x \in X$ with $a(x, \cdot) = y^* \in Y^*$, it holds $Tx = y^*$. With $\|T : X \to Y^*\| = \|a\|$, we see $\|y^*\|_{Y^*} \leq \|a\| \, \|x\|_X$. With $x = T^{-1} y^*$ and $\|T^{-1} : Y^* \to X\| \leq 1/\alpha$, we derive $\|x\|_X \leq \|y^*\|_{Y^*}/\alpha$. This concludes the proof. ■

One important difference to the elliptic framework now is, that we may not simply replace X and Y by discrete spaces X_h and Y_h, respectively. Instead, Corollary 6.3 states that we need to satisfy the inf-sup condition and the non-degeneracy condition not only for the pairing (X, Y) of continuous spaces, but also for any pairing (X_h, Y_h) of discrete spaces. To underline this, note that

$$T = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

is an isomorphism on $Y = X = \mathbb{R}^3$. For $X_h = Y_h = \mathbb{R}^2$ and the canonical embedding, i.e., $x \in \mathbb{R}^2$ is identified with $(x, 0) \in \mathbb{R}^3$, the restricted matrix is

$$T_h = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

which is clearly singular. We finally note that in the discrete setting the inf-sup condition and the non-degeneracy condition are equivalent.

Proposition 6.4. *Let X and Y be real Banach spaces with $\dim X < \infty$ and $\dim Y < \infty$. Let $a : X \times Y \to \mathbb{R}$ be bilinear. Then, there holds the following:*

(i) *The inf-sup condition $\alpha := \inf_{x\in X\backslash\{0\}} \sup_{y\in Y\backslash\{0\}} \frac{a(x,y)}{\|x\|_X \|y\|_Y} > 0$ implies $\dim X \le \dim Y$.*

(ii) *The non-degeneracy condition $\big(\forall y \in Y\backslash\{0\} \exists x \in X \quad a(x, y) \neq 0\big)$ implies $\dim Y \le \dim X$.*

(iii) *For $\dim X = \dim Y$, the inf-sup condition is satisfied if and only if the non-degeneracy condition is satisfied.*

Proof. We define the operators $A_1 \in L(X, Y^*)$ and $A_2 \in L(Y, X^*)$ by $A_1 x := a(x, \cdot)$ and $A_2 y := a(\cdot, y)$. According to Linear Algebra, finite dimension implies

$$\dim X = \dim\ker(A_1) + \dim\operatorname{range}(A_1) \le \dim\ker(A_1) + \dim Y^* = \dim\ker(A_1) + \dim Y,$$
$$\dim Y = \dim\ker(A_2) + \dim\operatorname{range}(A_2) \le \dim\ker(A_2) + \dim X^* = \dim\ker(A_2) + \dim X.$$

1. step. If $\dim X > \dim Y$, we obtain $\dim\ker(A_1) > 0$. Hence, there exists $x \in X\backslash\{0\}$ with $A_1 x = 0$. This implies $a(x, y) = 0$ for all $y \in Y$ and hence $\alpha = 0$ for the inf-sup constant. By contraposition, this shows that the inf-sup condition implies $\dim\ker(A_1) = 0$ and hence $\dim X \le \dim Y$. This proves (i).

2. step. If $\dim Y > \dim X$, we obtain $\dim\ker(A_2) > 0$. Hence, there exists $y \in Y\backslash\{0\}$ with $A_2 y = 0$. This implies $a(x, y) = 0$ for all $x \in X$, and hence the non-degeneracy condition fails. By contraposition, this shows that the non-degeneracy condition implies $\dim\ker(A_2) = 0$ and hence $\dim Y \le \dim X$. This proves (ii).

3. step.

In Step (ii), we have shown that (ND) implies injectivity of A_2. Since $\dim X = \dim Y = \dim Y^*$, this proves that A_2 is bijective. The converse implication is obvious, i.e., A_2 is bijective if and only if (ND) holds. In Step (i), we showed that the inf-sup condition implies injectivity of A_1. Since $\dim X = \dim Y = \dim Y^*$, this proves that A_1 is bijective. Again, the converse implication is easy, i.e., A_1 is bijective if and only if the inf-sup condition holds. To conlcude (iii), we only have to show that bijectivity of A_1 and A_2 are equivalent. To that end, let $\{x_1, \ldots, x_n\} \subset X$ and $\{y_1, \ldots, y_n\} \subset Y$ be bases. We define the matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{A}_{jk} := a(x_k, y_j)$ and note that $(A_1 x_k)(y_j) = a(x_k, y_j) = \boldsymbol{A}_{jk}$ as well as $(A_2 y_j)(x_k) = a(x_k, y_j) = \boldsymbol{A}_{jk}$. Therefore, $\boldsymbol{A}$ is the Petrov-Galerkin matrix corresponding to A_1 and its transpose $\boldsymbol{A}^T$ is the Petrov-Galerkin matrix corresponding to A_2. Therefore, Linear Algebra proves the equivalence

$$A_1 \text{ is bijective} \quad \Longleftrightarrow \quad \boldsymbol{A} \text{ is regular} \quad \Longleftrightarrow \quad \boldsymbol{A}^T \text{ is regular} \quad \Longleftrightarrow \quad A_2 \text{ is bijective}$$

This concludes the proof. ■

Exercise 42. Prove that a bilinear form $a : X \times Y \to \mathbb{R}$ on normed spaces X and Y is continuous if and only if $\|a\| := \sup_{\substack{x \in X \setminus \{0\} \\ y \in Y \setminus \{0\}}} \frac{a(x,y)}{\|x\|_X \|y\|_Y} < \infty$. □

The following exercise states the quasi optimality of Petrov-Galerkin schemes. We stress, however, that the quasi-optimality constant depends on the discrete inf-sup condition.

Exercise 43 (Céa's Lemma for Petrov-Galerkin Schemes). We consider the weak form (6.5) with a continuous bilinear form $a : X \times Y \to \mathbb{R}$ on Banach spaces X and Y. Let $y^* \in Y^*$. Let X_h and Y_h be finite dimensional subspaces of X resp. Y with $\dim X_h = \dim Y_h$. We assume the

- **discrete inf-sup condition** $\alpha_h := \inf_{x_h \in X_h \setminus \{0\}} \sup_{y_h \in Y_h \setminus \{0\}} \frac{a(x_h, y_h)}{\|x_h\|_X \|y_h\|_Y} > 0,$

Then, there is a unique $x_h \in X_h$ with

$$a(x_h, \cdot) = y^* \in Y_h^*. \tag{6.7}$$

If $x \in X$ solves the weak form (6.5), we have quasi optimality

$$\|x - x_h\|_X \leq \big(1 + \|a\|/\alpha_h\big) \min_{v_h \in X_h} \|x - v_h\|_X, \tag{6.8}$$

where $\|a\| := \sup_{\substack{x \in X \setminus \{0\} \\ y \in Y \setminus \{0\}}} \frac{a(x,y)}{\|x\|_X \|y\|_Y}$ denotes the continuity bound of $a(\cdot, \cdot)$. □

A simple observation is that the LBB theory allows a generalization of the Lax-Milgram lemma to the case of reflexive Banach spaces.

Exercise 44 (Lax-Milgram Lemma for Reflexive Spaces). Let $a : X \times X \to \mathbb{R}$ be a continuous and elliptic bilinear form on the reflexive Banach space X. Prove that $a(\cdot, \cdot)$ satisfies the inf-sup condition

$$\tau := \inf_{x \in X \setminus \{0\}} \sup_{y \in X \setminus \{0\}} \frac{a(x,y)}{\|x\|_X \|y\|_X} > 0$$

as well as the non-degeneracy condition

$$\forall y \in X\backslash\{0\}\exists x \in X \quad a(x, y) \neq 0.$$

For each given right-hand side $x^* \in X^*$, the weak form (6.4) thus has a unique solution $x \in X$. □

Another observation is that for reflexive spaces, it is immaterial whether the LBB condition is stated for the first or the second component.

Exercise 45. Let X, Y be reflexive Banach spaces and $a : X \times Y \to \mathbb{R}$ be a continuous bilinear form. Prove that the following statements (i)–(ii) are equivalent:

(i) The bilinear form satisfies the **LBB condition for the first argument**:

- $\alpha_1 := \inf\limits_{x\in X\backslash\{0\}} \sup\limits_{y\in Y\backslash\{0\}} \dfrac{a(x,y)}{\|x\|_X\|y\|_Y} > 0,$
- $\forall y \in Y\backslash\{0\}\exists x \in X \quad a(x, y) \neq 0.$

(ii) The bilinear form satisfies the **LBB condition for the second argument**:

- $\alpha_2 := \inf\limits_{y\in Y\backslash\{0\}} \sup\limits_{x\in X\backslash\{0\}} \dfrac{a(x,y)}{\|x\|_X\|y\|_Y} > 0,$
- $\forall x \in X\backslash\{0\}\exists y \in Y \quad a(x, y) \neq 0.$

Moreover, in this case there holds $\alpha_1 = \alpha_2$. □

6.2 Abstract Analysis of Mixed Formulations

Instead of the general mixed formulation (6.5), we consider linear problems with side constraints in the following. These arise, for instance, for the Stokes problem.

Before we focus on the abstract solution theory, we explain why these problems are called *saddle point problems*: Plotting a function $f : \mathbb{R}^2 \to \mathbb{R}$ over the two-dimensional plane, we call a point (x, y) saddle point of f if the real function $f(x + t, y)$ has a minimum at $t = 0$ and the function $f(x, y + t)$ has a maximum for $t = 0$. This is, what is stated in the following proposition for the so-called *Lagrange functional.*

Proposition 6.5.

Let $a : X \times X \to \mathbb{R}$ and $b : X \times Y \to \mathbb{R}$ be bilinear forms on normed spaces X and Y. Assume that $a(\cdot, \cdot)$ is positive semidefinite, i.e., $a(x, x) \geq 0$ and symmetric. Then, given $(x^*, y^*) \in X^* \times Y^*$, $(x, y) \in X \times Y$ is a solution of the saddle point problem

$$\begin{array}{lcll} a(x,\cdot) & + & b(\cdot,y) & = & x^* \in X^* \\ b(x,\cdot) & & & = & y^* \in Y^*. \end{array} \tag{6.9}$$

if and only if the Lagrange functional $\mathcal{L}(v,w) := \dfrac{1}{2}\,a(v,v) - x^*(v) + b(v,w) - y^*(w)$ *satisfies*

$$\mathcal{L}(x,w) \le \mathcal{L}(x,y) \le \mathcal{L}(v,y) \quad \textit{for all } (v,w) \in X \times Y, \tag{6.10}$$

i.e., (x,y) *is a saddle point of* $\mathcal{L}(\cdot,\cdot)$*. In this case, the first estimate in* (6.10) *holds with equality.*

Proof. First, assume that $(x,y) \in X \times Y$ is a solution of the saddle point problem (6.9). For $w \in Y$, the second equality in (6.9) implies

$$\mathcal{L}(x,y) - \mathcal{L}(x,w) = b(x,y-w) - y^*(y-w) = 0.$$

This proves the lower estimate of (6.10) even with equality. For $v \in X$, symmetry of $a(\cdot,\cdot)$ and the first equality in (6.9) prove

$$\mathcal{L}(v,y) - \mathcal{L}(x,y) = \frac{1}{2}a(x-v,x-v) + \underbrace{a(x,v-x) - x^*(v-x) + b(v-x,y)}_{=0} \ge 0,$$

and we obtain the upper estimate. Altogether, (x,y) is a saddle point of the Lagrange functional. The proof of the converse implication follows from a classical argument from the calculus of variations: Let $(x,y) \in X \times Y$ satisfy (6.10). For fixed $v \in X$, the real function $f(t) := \mathcal{L}(x+tv,y)$ has a global minimum at $t = 0$. There holds

$$f(t) = \frac{1}{2}\,a(x,x) - x^*(x) + b(x,y) - y^*(y) + \frac{t^2}{2}\,a(v,v) + t\{a(x,v) - x^*(v) + b(v,y)\}.$$

Hence $0 = f'(0) = a(x,v) - x^*(v) + b(v,y)$ for all $v \in X$. This proves the first equality in (6.9). To prove the second equality, consider, for fixed $w \in Y$, the real function $g(t) := \mathcal{L}(x,y+tw)$ which has a global maximum at $t = 0$. There holds

$$g(t) = \frac{1}{2}\,a(x,x) - x^*(x) + b(x,y) - y^*(y) + t\{b(x,w) - y^*(w)\}$$

and thus $0 = g'(0) = b(x,w) - y^*(w)$ for all $w \in Y$, i.e., $b(x,\cdot) = y^* \in Y^*$. ■

The following theorem of Brezzi provides existence and uniqueness of the solution of saddle point problems.

Theorem 6.6 (Brezzi). *Let X be a Hilbert space and Y be a reflexive Banach space. Let $a : X \times X \to \mathbb{R}$ and $b : X \times Y \to \mathbb{R}$ be continuous bilinear forms. We define* $X_0 := \big\{x \in X \,\big|\, b(x,\cdot) = 0 \in Y^*\big\}$ *and assume*

- $\alpha := \inf\limits_{v \in X_0\setminus\{0\}} \dfrac{a(v,v)}{\|v\|_X^2} > 0$, *i.e.,* $a(\cdot,\cdot)$ *is elliptic on* X_0,

- $\beta := \inf\limits_{y \in Y\setminus\{0\}} \sup\limits_{x \in X\setminus\{0\}} \dfrac{b(x,y)}{\|x\|_X\|y\|_Y} > 0.$

Then, for any $(x^*, y^*) \in X^* \times Y^*$, there is a unique solution $(x, y) \in X \times Y$ of

$$\begin{array}{lcl} a(x,\cdot) \;+\; b(\cdot,y) &=& x^* \in X^* \\ b(x,\cdot) &=& y^* \in Y^*. \end{array} \tag{6.11}$$

Moreover, we have the stability estimates

$$\|x\|_X \le \frac{1}{\alpha}\,\|x^*\|_{X^*} + \frac{1}{\beta}\Big(1 + \frac{\|a\|}{\alpha}\Big)\|y^*\|_{Y^*} \tag{6.12}$$

and

$$\|y\|_Y \le \frac{1}{\beta}\,\Big(1 + \frac{\|a\|}{\alpha}\Big)\Big(\|x^*\|_{X^*} + \frac{\|a\|}{\beta}\,\|y^*\|_{Y^*}\Big) \tag{6.13}$$

Remark. (i) Note that one can identify $X^* \times Y^* = (X \times Y)^*$ as follows: For $x^* \in X^*$ and $y^* \in Y^*$, the definition $z^*(x, y) := x^*(x) + y^*(y)$ yields $z^* \in (X \times Y)^*$. Conversely, $z^* \in (X \times Y)^*$ gives rise to $x^*(x) := z^*(x, 0)$ and $y^*(y) := z^*(0, y)$ with $(x^*, y^*) \in X^* \times Y^*$.

(ii) If we define operators $A_1 \in L(X, X^*)$, $B_1 \in L(X, Y^*)$, and $B_2 \in L(Y, X^*)$ by

$$A_1 x := a(x,\cdot), \quad B_1 x := b(x,\cdot), \quad \text{and} \quad B_2 y := b(\cdot, y),$$

Equation (6.11) can be written in the form

$$\begin{pmatrix} A_1 & B_2 \\ B_1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x^* \\ y^* \end{pmatrix}. \tag{6.14}$$

In this form, the Brezzi theorem states that this operator matrix is an isomorphism from $X \times Y$ to $X^* \times Y^* = (X \times Y)^*$ and so fits into the abstract framework given above.

(iii) We stress that the original proof of Brezzi works for reflexive Banach spaces X and Y. Therein, it is proved directly that the operator matrix from (6.14) satisfies the inf-sup condition as well as the non-degeneracy condition. Our stronger assumption that X is not only a reflexive Banach space, but even a Hilbert space, reduces the technical difficulties and leads to a much simpler proof. □

Sketch of Proof of Theorem 6.6. Let $(x, y) \in X \times Y$. With the orthogonal decomposition $X = X_0 \oplus X_0^\perp$, we write $x = x_1 + x_2$ with $x_1 \in X_0$ and $x_2 \in X_0^\perp$. Note that (6.11) is equivalent to the following three identities:

- $b(x_2,\cdot) = y^* \in Y^*$,
- $a(x_1,\cdot) = x^* - a(x_2,\cdot) \in X_0^*$,
- $b(\cdot, y) = x^* - a(x_1 + x_2,\cdot) \in X^*$.

For the proof of Theorem 6.6 we are going to show that these three equations — proved in the stated order — admit unique solutions $x_2 \in X_0^\perp$, $x_1 \in X_0$, and $y \in Y^*$. This proves existence and uniqueness of the solution $(x, y) = (x_1 + x_2, y) \in X \times Y$ of (6.11). ■

The main ingredient of the proof of Theorem 6.6 is the closed range theorem:

Theorem 6.7 (Banach's Closed Range Theorem). *For an operator $T \in L(X,Y)$ between Banach spaces X and Y, the following is equivalent:*
(i) $\operatorname{range}(T) \subseteq Y$ *is closed,*
(ii) $\operatorname{range}(T) = (\ker T^*)_\circ = \big\{y \in Y \,\big|\, \forall y^* \in \ker T^* \quad y^*(y) = 0\big\}$,
(iii) $\operatorname{range}(T^*) \subseteq X^*$ *is closed,*
(iv) $\operatorname{range}(T^*) = (\ker T)^\circ = \big\{x^* \in X^* \,\big|\, \forall x \in \ker T \quad x^*(x) = 0\big\}$. ■

Proof of Theorem 6.6. The essential steps of the proof are based on operator arguments for the operators defined by $B_1 x := b(x,\cdot)$ and $B_2 y := b(\cdot, y)$. We are going to consider the four operators

$$B_1 \in L(X, Y^*), \qquad B_1^* \in L(Y^{**}, X^*),$$
$$B_2 \in L(Y, X^*), \qquad B_2^* \in L(X^{**}, Y^*).$$

More precisely, the first three steps state the essential observations about these operators, whereas the remaining proof follows the line of the sketch given before.

1. step. B_2 is injective with closed range and $\|B_2^{-1} : \operatorname{range}(B_2) \to Y\| = 1/\beta$, which follows from Lemma 6.1 and

$$\beta = \inf_{y \in Y \setminus \{0\}} \frac{\|B_2 y\|_{X^*}}{\|y\|_Y}.$$

2. step. There holds $B_2 = B_1^* I_Y$, which follows from

$$(B_2 y)(x) = b(x,y) = (B_1 x)(y) = (I_Y y)(B_1 x) = (B_1^* I_Y y)(x) \quad \text{for all } x \in X, y \in Y.$$

3. step. Since Y is reflexive, B_1^* is injective with closed $\operatorname{range}(B_1^*) = \operatorname{range}(B_2)$. Moreover, the closed range theorem even proves

$$\operatorname{range}(B_2) = \operatorname{range}(B_1^*) = (\ker B_1)^\circ = (X_0)^\circ \quad \text{as well as} \quad \operatorname{range}(B_1) = (\ker B_1^*)_\circ = Y^*.$$

4. step. There is a unique $x_2 \in X_0^\perp$ with $b(x_2,\cdot) = y^* \in Y^*$: According to step 3, there is at least one $x \in X$ with $b(x,\cdot) = B_1 x = y^*$. The decomposition $x = x_1 + x_2$ with $x_1 \in X_0$ and $x_2 \in X_0^\perp$ proves $b(x_2,\cdot) = b(x,\cdot) = y^* \in Y^*$, which concludes existence. To prove uniqueness, let $\widetilde{x}_2 \in X_0^\perp$ with $b(\widetilde{x}_2,\cdot) = y^* \in Y^*$. Then, $b(x_2 - \widetilde{x}_2,\cdot) = 0 \in Y^*$, whence $x_2 - \widetilde{x}_2 \in \ker B_1 = X_0$. From $x_2 - \widetilde{x}_2 \in X_0^\perp$, we thus obtain $x_2 = \widetilde{x}_2$.

5. step. There is a unique element $x_1 \in X_0$ with $a(x_1,\cdot) = x^* - a(x_2,\cdot) \in X_0^*$ which immediately follows from the Lax-Milgram lemma and the observation that X_0 is a closed subspace of a Hilbert space and hence a Hilbert space as well.

6. step. There is a unique element $y \in Y$ with $b(\cdot, y) = x^* - a(x,\cdot)$, where $x := x_1 + x_2 \in X$: By construction in step 5, there holds

$$x^* - a(x,\cdot) \in (X_0)^\circ = \big\{v^* \in X^* \,\big|\, \forall v \in X_0 \quad v^*(v) = 0\big\}.$$

According to step 1 and step 3, B_2 is injective with $\operatorname{range}(B_2) = (X_0)^\circ$. Thus, there is a unique $y \in Y$ with $b(\cdot, y) = B_2 y = x^* - a(x,\cdot)$.

7. step. There holds $\|x_2\|_X \le \|y^*\|_{Y^*}/\beta$: From $x_2 \in X_0^\perp$ follows $(x_2\,;\cdot)_X \in (X_0)^\circ = \mathrm{range}(B_2)$. Thus, we may choose $\widetilde{y} \in Y$ with $B_2\widetilde{y} = (x_2\,;\cdot)_X$. From $\|B_2^{-1} : (X_0)^\circ \to Y\| = 1/\beta$, we infer $\|\widetilde{y}\|_Y \le \|(x_2\,;\cdot)_X\|_{X^*}/\beta = \|x_2\|_X/\beta$. Together with $b(x_2,\cdot) = y^*$, we conclude

$$\|x_2\|_X^2 = (x_2\,;x_2)_X = (B_2\widetilde{y})(x_2) = b(x_2,\widetilde{y}) = y^*(\widetilde{y}) \le \|y^*\|_{Y^*}\|\widetilde{y}\|_Y \le \frac{\|y^*\|_{Y^*}}{\beta}\,\|x_2\|_X.$$

8. step. There holds $\|x_1\|_X \le \alpha^{-1}\,(\|x^*\|_{X^*} + \|a\|\|x_2\|_X)$: Note that $A_1 \in L(X_0, X_0^*)$ is an isomorphism with $\|A_1^{-1} : X_0^* \to X_0\| \le 1/\alpha$. From $A_1x_1 = a(x_1,\cdot) = x^* - a(x_2,\cdot)$, we thus infer

$$\|x_1\|_X \le \frac{1}{\alpha}\,\|x^* - a(x_2,\cdot)\|_{X_0^*} \le \frac{1}{\alpha}\,\big(\|x^*\|_{X^*} + \|a\|\|x_2\|_X\big).$$

9. step. The triangle inequality leads to

$$\|x\|_X \le \|x_1\|_X + \|x_2\|_X \le \frac{1}{\alpha}\,\|x^*\|_{X^*} + \Big(\frac{\|a\|}{\alpha} + 1\Big)\,\|x_2\|_X \le \frac{1}{\alpha}\,\|x^*\|_{X^*} + \frac{1}{\beta}\,\Big(\frac{\|a\|}{\alpha} + 1\Big)\,\|y^*\|_{Y^*}.$$

10. step. It finally remains to dominate $\|y\|_Y$, where $B_2y = b(\cdot,y) = x^* - a(x,\cdot) \in (X_0)^\circ$. We use $\|B_2^{-1} : (X_0)^\circ \to Y\| = 1/\beta$ to see

$$\begin{aligned}\|y\|_Y \le \frac{1}{\beta}\,\|x^* - a(x,\cdot)\|_{X^*} &\le \frac{1}{\beta}\,\|x^*\|_{X^*} + \frac{\|a\|}{\beta}\,\|x\|_X \\ &\le \frac{1}{\beta}\,\|x^*\|_{X^*} + \frac{\|a\|}{\beta}\,\frac{1}{\alpha}\,\|x^*\|_{X^*} + \frac{\|a\|}{\beta^2}\Big(1 + \frac{\|a\|}{\alpha}\Big)\,\|y^*\|_{Y^*} \\ &= \frac{1}{\beta}\,\Big(1 + \frac{\|a\|}{\alpha}\Big)\,\Big(\|x^*\|_{X^*} + \frac{\|a\|}{\beta}\,\|y^*\|_{Y^*}\Big).\end{aligned}$$

This concludes the proof. ■

Remark. (i) Let $B_1 \in L(X, Y^*)$ and $B_2 \in L(Y, X^*)$ be defined as in the proof of Theorem 6.6. In the proof, we have seen that $\beta > 0$ implies surjectivity of B_1. We note that even the converse implication holds, i.e.,

$$\beta := \inf_{y\in Y\setminus\{0\}}\ \sup_{x\in X\setminus\{0\}} \frac{b(x,y)}{\|x\|_X\|y\|_Y} > 0 \quad\Longleftrightarrow\quad B_1 \text{ is surjective.} \tag{6.15}$$

Suppose that B_1 is surjective. As in step 3 of the preceding proof, the closed range theorem proves that B_1^* is injective with closed range. Moreover, $B_2 = B_1^*I_Y$ proves that B_2 is injective with closed $\mathrm{range}(B_2) = \mathrm{range}(B_1^*) = (\ker B_1)^\circ = (X_0)^\circ$, i.e., $B_2 : Y \to \mathrm{range}(B_2)$ is continuous and bijective between the Banach spaces Y and $\mathrm{range}(B_2) \subseteq X^*$. According to the open mapping theorem, $B_2 : Y \to \mathrm{range}(B_2)$ even is an isomorphism, i.e., $\beta^{-1} = \|B_2 : Y \to \mathrm{range}(B_2)\| < \infty$, whence $\beta > 0$.

(ii) Altogether, the two main assumptions on $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$ can equivalently be stated as follows:

- The bilinear form $a(\cdot,\cdot)$ is elliptic on $X_0 = \ker B_1$.
- The operator $B_1 \in L(X, Y^*)$ is surjective.

We hope that the reader may keep this (abstract) formulation in mind much easier. For the statement of Theorem 6.6, we used the definition of α and β instead, to provide the stability estimates (6.12)–(6.13) with explicit constants. □

Going through the proof of Theorem 6.6, one realizes that ellipticity of $a(\cdot,\cdot)$ on X_0 is only used to provide a unique $x_1 \in X_0$ with $a(x_1,\cdot) = x_0^* \in X_0^*$ in step 5. To prove unique existence of x_1, it is, however, sufficient to assume that the operator $A_1 : X_0 \to X_0^*$ defined by $A_1 x := a(x,\cdot)$ is an isomorphism. This is done in the following exercise.

Exercise 46. Let X, Y, $a(\cdot,\cdot)$, and $b(\cdot,\cdot)$ be as in Theorem 6.6. Then, the following statements are equivalent:

(i) For all $(x^*, y^*) \in X^* \times Y^*$, there exists a unique solution $(x, y) \in X \times Y$ of the saddle point problem (6.11).

(ii) The bilinear forms $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$ satisfy the following three assumptions:

- $\alpha := \inf\limits_{v\in X_0\backslash\{0\}} \sup\limits_{w\in X_0\backslash\{0\}} \dfrac{a(v,w)}{\|v\|_X\|w\|_X} > 0,$
- $\forall w \in X_0\backslash\{0\} \exists v \in X_0 \quad a(v,w) \neq 0,$
- $\beta := \inf\limits_{y\in Y\backslash\{0\}} \sup\limits_{x\in X\backslash\{0\}} \dfrac{b(x,y)}{\|x\|_X\|y\|_Y} > 0.$

The first two assumptions state that $A_1 : X_0 \to X_0^*$ is an isomorphism, cf. Theorem 6.2. The assumption on β is the same as in the above statement of the Brezzi theorem. □

The following corollary provides the relation between saddle point problems and the abstract Petrov-Galerkin scheme from Section 6.1.

Corollary 6.8. *Suppose that X is a Hilbert space, Y is a reflexive Banach space, and $a : X \times X \to \mathbb{R}$ and $b : X \times Y \to \mathbb{R}$ are continuous bilinear forms. Then, $Z := X \times Y$ is a reflexive Banach space, and $c((x,y),(\widetilde{x},\widetilde{y})) := a(x,\widetilde{x}) + b(\widetilde{x},y) + b(x,\widetilde{y})$ defines a continuous bilinear form $c : Z \times Z \to \mathbb{R}$. Moreover, for $(x,y) \in X \times Y$ and $(x^*,y^*) \in X^* \times Y^*$, the saddle point problem* (6.11) *is equivalent to*

$$c((x,y),(\widetilde{x},\widetilde{y})) = x^*(\widetilde{x}) + y^*(\widetilde{y}) \quad \textit{for all } (\widetilde{x},\widetilde{y}) \in X \times Y. \tag{6.16}$$

Finally, the following three statements are equivalent:

(i) *$a(\cdot,\cdot)$ and $b(\cdot,\cdot)$ satisfy the assumptions of the Brezzi theorem, i.e.,*

- $\alpha := \inf\limits_{v\in X_0\backslash\{0\}} \sup\limits_{w\in X_0\backslash\{0\}} \dfrac{a(v,w)}{\|v\|_X\|w\|_X} > 0$ *with* $X_0 := \{x \in X \,|\, b(x,\cdot) = 0 \in Y^*\},$
- $\forall w \in X_0\backslash\{0\} \exists v \in X_0 \quad a(v,w) \neq 0,$
- $\beta := \inf\limits_{y\in Y\backslash\{0\}} \sup\limits_{x\in X\backslash\{0\}} \dfrac{b(x,y)}{\|x\|_X\|y\|_Y} > 0.$

(ii) *$c(\cdot,\cdot)$ satisfies the LBB conditions*

- $\gamma := \inf\limits_{z\in Z\backslash\{0\}} \sup\limits_{w\in Z\backslash\{0\}} \dfrac{c(z,w)}{\|z\|_Z\|w\|_Z} > 0,$
- $\forall w \in Z\backslash\{0\}\exists z \in Z \quad c(z,w) \neq 0.$

(iii) *For all $(x^*,y^*) \in X^* \times Y$, the variational formulation* (6.16) *has a unique solution $(x,y) \in X \times Y$.*

In particular, it holds $\|c\| \leq \|a\| + 2\,\|b\|$ for the corresponding norms and there exists a constant $C > 0$ such that

$$\gamma \geq C\Big[\frac{1}{\alpha} + \frac{1}{\beta}\Big(1 + \frac{\|a\|}{\alpha}\Big)\Big(1 + \frac{\|a\|}{\beta}\Big)\Big]^{-1}. \tag{6.17}$$

Proof. **1. step.** Since X and Y are reflexive, their closed unit balls $B_X \subset X$ and $B_Y \subset Y$ are weakly compact. According to the Tychonov theorem, $B_X \times B_Y$ and hence B_Z are weakly compact as well. Consequently, Z is reflexive. Moreover, it is obvious that $c(\cdot,\cdot)$ is bilinear and continuous with $\|c\| \leq \|a\| + 2\,\|b\|$.

2. step. Summing the equations of (6.11), we obtain the variational form (6.16). Testing (6.16) with test functions of the type $(\widetilde{x}, 0)$ or $(0, \widetilde{y})$, we see that (6.11) and (6.16) are, in fact, equivalent.

3. step. The equivalence of (ii) and (iii) is stated in Corollary 6.3. The equivalence of (i) and (iii) follows from step 2 and Exercise 46.

4. step. It remains to prove (6.17): From (6.12)–(6.13), we obtain

$$\begin{aligned}
\|x\|_X + \|y\|_Y &\leq \frac{1}{\alpha}\,\|x^*\|_{X^*} + \frac{1}{\beta}\Big(1 + \frac{\|a\|}{\alpha}\Big)\|y^*\|_{Y^*} + \frac{1}{\beta}\Big(1 + \frac{\|a\|}{\alpha}\Big)\Big(\|x^*\|_{X^*} + \frac{\|a\|}{\beta}\,\|y^*\|_{Y^*}\Big) \\
&= \Big[\frac{1}{\alpha} + \frac{1}{\beta}\Big(1 + \frac{\|a\|}{\alpha}\Big)\Big]\,\|x^*\|_{X^*} + \frac{1}{\beta}\Big(1 + \frac{\|a\|}{\alpha}\Big)\Big(1 + \frac{\|a\|}{\beta}\Big)\,\|y^*\|_{Y^*} \\
&\leq \Big[\frac{1}{\alpha} + \frac{1}{\beta}\Big(1 + \frac{\|a\|}{\alpha}\Big)\Big(1 + \frac{\|a\|}{\beta}\Big)\Big]\,\big[\|x^*\|_{X^*} + \|y^*\|_{Y^*}\big].
\end{aligned}$$

With the operator $Tz := c(z,\cdot)$, this proves that the solution operator $T^{-1} : X^* \times Y^* \to X \times Y$ has operator norm $\|T^{-1}\| \leq C\,\Big[\frac{1}{\alpha} + \frac{1}{\beta}\Big(1 + \frac{\|a\|}{\alpha}\Big)\frac{1}{\beta}\Big(1 + \frac{\|a\|}{\beta}\Big)\Big]$, where $C > 0$ depends only on the norms chosen on $Z = X \times Y$ and $Z^* = X^* \times Y^*$. According to Theorem 6.2, it holds $\|T^{-1}\| = 1/\gamma$. This concludes the proof. ■

Exercise 47. Give a direct proof that $c(\cdot,\cdot)$ from Corollary 6.8 satisfies the LBB condition, i.e., prove directly that (i) implies (ii). **Hint.** For $(x,y) \neq 0$ use the orthogonal decomposition $x = x_1 + x_2 \in X_0 + X_0^\perp$ and estimate $\|x_1\|_X$, $\|x_2\|_X$, and $\|y\|_Y$ separately. □

Corollary 6.8 together with Exercise 43 provides a solvability theory and the Céa lemma for Galerkin discretizations of saddle point problems.

Corollary 6.9 (Céa Lemma for Saddle Point Problems, Version I). *Let $a : X \times X \to \mathbb{R}$ and $b : X \times Y \to \mathbb{R}$ be continuous bilinear forms on a Hilbert space X and a reflexive Banach space Y. Given $(x^*, y^*) \in X^* \times Y^*$, let $(x, y) \in X \times Y$ be a solution of the saddle point problem (6.11). Let $X_h \subset X$ and $Y_h \subset Y$ be finite dimensional subspaces and define $X_{0h} := \{x_h \in X_h \,|\, b(x_h, \cdot) = 0 \in Y_h^*\}$. Suppose that*

- $\alpha_h := \inf\limits_{v_h \in X_{0h}\setminus\{0\}} \sup\limits_{w_h \in X_{0h}\setminus\{0\}} \dfrac{a(v_h, w_h)}{\|v_h\|_X \|w_h\|_X} > 0,$
- $\beta_h := \inf\limits_{y_h \in Y_h\setminus\{0\}} \sup\limits_{x_h \in X_h\setminus\{0\}} \dfrac{b(x_h, y_h)}{\|x_h\|_X \|y_h\|_Y} > 0.$

Then, there is a unique solution $(x_h, y_h) \in X_h \times Y_h$ of the discrete saddle point problem

$$\begin{array}{lcccl} a(x_h,\cdot) & + & b(\cdot, y_h) & = & x^* \in X_h^*, \\ b(x_h,\cdot) & & & = & y^* \in Y_h^*, \end{array} \tag{6.18}$$

and there holds

$$\|x - x_h\|_X + \|y - y_h\|_Y \le C_h \Big(\min_{\widetilde{x}_h \in X_h} \|x - \widetilde{x}_h\|_X + \min_{\widetilde{y}_h \in Y_h} \|y - \widetilde{y}_h\|_Y \Big)$$

The constant $C_h > 0$ depends only on $(\|a\| + 2\|b\|)\gamma_h$ with $\gamma_h := \Big[\frac{1}{\alpha_h} + \frac{1}{\beta_h}\Big(1 + \frac{\|a\|}{\alpha_h}\Big)\frac{1}{\beta_h}\Big(1 + \frac{\|a\|}{\beta_h}\Big)\Big]$.

Proof. The existence and uniqueness of (x_h, y_h) follows from the abstract Brezzi theorem; see Corollary 6.8. For Petrov-Galerkin schemes, the constant in the Céa lemma depends only on the quotient of the continuity bound and the discrete inf-sup constant; see Exercise 43. Both constants have been estimated in Corollary 6.8. ■

Remark. The Galerkin discretization of saddle point problems is structurally much more difficult than for problems of the Lax-Milgram lemma:

(i) Note that $X_{0h} \not\subseteq X_0 := \{v \in X \,|\, b(v, \cdot) = 0 \in Y^*\}$. There may be even no relation between X_0 and X_{0h} besides the trivial $X_0 \cap X_h \subseteq X_{0h}$. In particular, there is no relation between α and α_h even if $a(\cdot, \cdot)$ is elliptic on X_0.

(ii) However, if $a(\cdot, \cdot)$ is already elliptic on X, i.e., $\tau := \inf_{x \in X\setminus\{0\}} \frac{a(x,x)}{\|x\|_X^2} > 0$ this implies $\alpha \ge \tau$ and $\alpha_h \ge \tau$ for the continuous and discrete inf-sup constant of $a(\cdot, \cdot)$.

(iii) Moreover, $\beta > 0$ from the continuous formulation does not imply $\beta_h > 0$ for the discrete formulation. Below, we introduce Fortin's criterium which provides some help on this matter.

(iv) Finally, we recall that $\beta_h > 0$ implies necessarily $\dim Y_h \le \dim X_h$; see Proposition 6.4. □

Exercise 48. For a matrix $A \in \mathbb{R}^{m\times n}$ holds $\ker(A^T) = (\mathrm{range} A)^\perp$ as well as $\mathrm{range}(A^T) = (\ker A)^\perp$, where $(\cdot)^\perp$ denotes the orthogonal complement with respect to the usual Euclidean product in $\mathbb{R}^m$ resp. $\mathbb{R}^n$. □

The following two exercises consider the discretization of the mixed problem (6.11). We stress that a linear system similar to the one here, also appeared for the discretization of the Neumann problem, where we had to realize the linear side constraint $\int_\Omega u_h \, dx = 0$.

Exercise 49. Let $a : X \times X \to \mathbb{R}$ and $b : X \times Y \to \mathbb{R}$ be continuous bilinear forms on a Hilbert space X and a reflexive Banach space Y. We replace X and Y by finite dimensional subspaces X_h and Y_h, respectively. Show that the computation of a discrete solution $(x_h, y_h) \in X_h \times Y_h$ of

$$\begin{array}{lcccl} a(x_h,\cdot) & + & b(\cdot,y_h) & = & x^* \in X_h^*, \\ b(x_h,\cdot) & & & = & y^* \in Y_h^*, \end{array} \tag{6.19}$$

is equivalent to the solution of a linear system with a matrix of the type $M := \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}$. □

Exercise 50. Let $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{m\times n}$, and $M := \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}$. Assume that A is positive definite on the kernel of B. Prove that M is regular if and only if $\operatorname{range}(B) = \mathbb{R}^m$. □

We conclude this section with an improved Céa lemma for saddle point problems; cf. Corollary 6.9.

Theorem 6.10 (Céa Lemma for Saddle Point Problems, Version II). *Let $a : X \times X \to \mathbb{R}$ and $b : X \times Y \to \mathbb{R}$ be continuous bilinear forms on a Hilbert space X and a reflexive Banach space Y. Given $(x^*, y^*) \in X^* \times Y^*$, let $(x, y) \in X \times Y$ be a solution of the saddle point problem (6.11). Let $X_h \subset X$ and $Y_h \subset Y$ be finite dimensional subspaces and define $X_{0h} := \{x_h \in X_h \,|\, b(x_h,\cdot) = 0 \in Y_h^*\}$. Suppose that*

- $\alpha_h := \inf\limits_{v_h \in X_{0h}\setminus\{0\}} \dfrac{a(v_h,v_h)}{\|v_h\|_X^2} > 0,$
- $\beta_h := \inf\limits_{y_h \in Y_h\setminus\{0\}} \sup\limits_{x_h \in X_h\setminus\{0\}} \dfrac{b(x_h,y_h)}{\|x_h\|_X \|y_h\|_Y} > 0.$

Then, there is a unique solution $(x_h, y_h) \in X_h \times Y_h$ of the discrete saddle point problem

$$\begin{array}{lcccl} a(x_h,\cdot) & + & b(\cdot,y_h) & = & x^* \in X_h^*, \\ b(x_h,\cdot) & & & = & y^* \in Y_h^*, \end{array} \tag{6.20}$$

and there holds

$$\|x - x_h\|_X \le \Big(1 + \frac{\|a\|}{\alpha_h}\Big)\Big(1 + \frac{\|b\|}{\beta_h}\Big) \min_{\widetilde{x}_h \in X_h} \|x - \widetilde{x}_h\|_X + \frac{\|b\|}{\alpha_h} \min_{\widetilde{y}_h \in Y_h} \|y - \widetilde{y}_h\|_Y \tag{6.21}$$

and

$$\|y-y_h\|_Y \le \Big(1+\frac{\|b\|}{\beta_h}\Big) \min_{\widetilde{y}_h\in Y_h} \|y-\widetilde{y}_h\|_Y + \frac{\|a\|}{\beta_h}\,\|x-x_h\|_X. \tag{6.22}$$

Sketch of Proof of Theorem 6.10. The unique existence of a discrete solution $(x_h, y_h) \in X_h \times Y_h$ follows from the Brezzi Theorem 6.6 applied for $X_h \times Y_h$. The quasioptimality is proven in three steps:

- First, we prove estimate (6.22).
- Second, we prove quasioptimality of $\|x-x_h\|_X$ with respect to the affine space $Z_h := \big\{\widetilde{x}_h \in X_h \,\big|\, b(\widetilde{x}_h,\cdot) = y^* \in Y_h^*\big\}$.
- In a final step, we estimate the bestapproximation error with respect to Z_h by the bestapproximation error with respect to the entire discrete space X_h which then leads to (6.21).

This general concept even works for nonlinear problems with linear side constraint. ■

Proof. We first note the Galerkin orthogonality, which now reads

$$\begin{array}{lcll} a(x-x_h,\cdot) & + & b(\cdot, y-y_h) & = \; 0 \in X_h^*, \\ b(x-x_h,\cdot) & & & = \; 0 \in Y_h^*, \end{array} \tag{6.23}$$

1. step. There holds

$$\|y-y_h\|_Y \le \Big(1+\frac{\|b\|}{\beta_h}\Big)\,\|y-\widetilde{y}_h\|_Y + \frac{\|a\|}{\beta_h}\,\|x-x_h\|_X \quad \text{for all } \widetilde{y}_h \in Y_h :$$

According to the definition of β_h, there holds

$$\beta_h\,\|\widetilde{y}_h - y_h\|_Y \le \sup_{\widetilde{x}_h\in X_h\setminus\{0\}} \frac{b(\widetilde{x}_h, \widetilde{y}_h - y_h)}{\|\widetilde{x}_h\|_X}.$$

With the Galerkin orthogonality, the nominator may be written as

$$\begin{aligned} b(\widetilde{x}_h, \widetilde{y}_h - y_h) &= -\big(a(x-x_h, \widetilde{x}_h) + b(\widetilde{x}_h, y-y_h)\big) + b(\widetilde{x}_h, \widetilde{y}_h - y_h) \\ &= -a(x-x_h,\widetilde{x}_h) + b(\widetilde{x}_h, \widetilde{y}_h - y) \end{aligned}$$

Therefore, continuity of $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$ lead to

$$\beta_h\,\|\widetilde{y}_h - y_h\|_Y \le \|a\|\|x-x_h\|_X + \|b\|\|\widetilde{y}_h - y\|_Y.$$

Altogether, a triangle inequality $\|y-y_h\|_Y \le \|y-\widetilde{y}_h\|_Y + \|\widetilde{y}_h - y_h\|_Y$ yields step 1.

2. step. With the affine space $Z_h := \big\{\widetilde{x}_h \in X_h \,\big|\, b(\widetilde{x}_h,\cdot) = y^* \in Y_h^*\big\}$, there holds

$$\|x-x_h\|_X \le \Big(1+\frac{\|a\|}{\alpha_h}\Big)\,\|x-z_h\|_X + \frac{\|b\|}{\alpha_h}\,\|y-\widetilde{y}_h\|_Y \quad \text{for all } z_h \in Z_h \text{ and } \widetilde{y}_h \in Y_h :$$

Since $x_h, z_h \in Z_h$, there holds $x_h - z_h \in X_{0h}$. According to the definition of α_h, we see

$$\alpha_h \|x_h - z_h\|_X^2 \leq a(x_h - z_h, x_h - z_h) = a(x_h - x, x_h - z_h) + a(x - z_h, x_h - z_h).$$

For the first term, the Galerkin orthogonality implies

$$a(x_h - x, x_h - z_h) = b(x_h - z_h, y - y_h) = b(x_h - z_h, \widetilde{y}_h - y_h) + b(x_h - z_h, y - \widetilde{y}_h),$$

where the first summand $b(x_h - z_h, \widetilde{y}_h - y_h) = 0$ drops out by use of $x_h - z_h \in X_{0h}$. By continuity of $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$, we see

$$\alpha_h \|x_h - z_h\|_X \leq \|a\| \|x - z_h\|_X + \|b\| \|y - \widetilde{y}_h\|_Y.$$

Again, a triangle inequality $\|x - x_h\|_X \leq \|x - z_h\|_X + \|x_h - z_h\|_X$ yields step 2.

3. step. There holds

$$\|x - z_h\|_X \leq \Big(1 + \frac{\|b\|}{\beta_h}\Big) \|x - \widetilde{x}_h\|_X \quad \text{for all } \widetilde{x}_h \in X_h \text{ and some } z_h \in Z_h \text{ depending on } \widetilde{x}_h :$$

We define $W_h := (X_{0h})^\perp \subseteq X_h$ and consider the operators $B_1 \in L(W_h, Y_h^*)$ and $B_2 \in L(Y_h, W_h^*)$ defined by $B_1 w_h := b(w_h, \cdot)$ and $B_2 y_h := b(\cdot, y_h)$. Note that

$$0 < \beta_h = \inf_{\widetilde{y}_h \in Y_h \setminus \{0\}} \sup_{\widetilde{x}_h \in X_h \setminus \{0\}} \frac{b(\widetilde{x}_h, \widetilde{y}_h)}{\|\widetilde{x}_h\|_X \|\widetilde{y}_h\|_Y} = \inf_{\widetilde{y}_h \in Y_h \setminus \{0\}} \sup_{w_h \in W_h \setminus \{0\}} \frac{b(w_h, \widetilde{y}_h)}{\|w_h\|_X \|\widetilde{y}_h\|_Y}.$$

According to Lemma 6.1, the operator B_2 is injective with closed range and $1/\beta_h = \|B_2^{-1} : \text{range}(B_2) \to Y_h\|$. From this, we derive that $B_1 = B_2^* \circ I_{Y_h}$ is surjective due to $\text{range}(B_1) = \text{range}(B_2^*) = (\ker B_2)^\circ = Y_h^*$. Note that by definition of $W_h := (X_{0h})^\perp \subseteq X_h$, the operator B_1 is injective and thus an isomorphism between W_h and Y_h^*. In particular, this yields bijectivity of B_2 as well as

$$\|B_1^{-1}\| = \|I_{Y_h}^{-1} (B_2^*)^{-1}\| = \|(B_2^{-1})^*\| = \|B_2^{-1}\| = 1/\beta_h.$$

In particular, there is a unique element $w_h \in W_h$ with $b(w_h, \cdot) = B_1 w_h = b(x - \widetilde{x}_h, \cdot) \in Y_h^*$ and there holds $\|w_h\|_X \leq \beta_h^{-1} \|b(x - \widetilde{x}_h, \cdot)\|_{X^*} \leq (\|b\|/\beta_h) \|x - \widetilde{x}_h\|_X$. The element $z_h := \widetilde{x}_h + w_h \in X_h$ satisfies $b(z_h, \cdot) = b(x, \cdot) = y^* \in Y_h^*$ and thus $z_h \in Z_h$. Now, we finally see

$$\|x - z_h\|_X \leq \|x - \widetilde{x}_h\|_X + \|w_h\|_X \leq \Big(1 + \frac{\|b\|}{\beta_h}\Big) \|x - \widetilde{x}_h\|_X.$$

This concludes step 3.

4. step. The proof of (6.22) follows by finite dimension: Note that step 1 implies

$$\|y - y_h\|_Y \leq \Big(1 + \frac{\|b\|}{\beta_h}\Big) \inf_{\widetilde{y}_h \in Y_h} \|y - \widetilde{y}_h\|_Y + \frac{\|a\|}{\beta_h} \|x - x_h\|_X,$$

and it only remains to see that the infimum is, in fact, attained: To that end, choose an infimizing sequence (y_k) in Y_h, i.e.

$$\lim_{k \to \infty} \|y - y_k\|_Y = \inf_{\widetilde{y}_h \in Y_h} \|y - \widetilde{y}_h\|_Y.$$

According to the triangle inequality, there holds $\|y_k\|_Y \leq \|y\|_Y + \|y - y_k\|_Y$, i.e. the sequence (y_k) is a bounded sequence in the finite dimensional space Y_h. Thus, the Bolzano-Weierstrass theorem yields the existence of a convergent subsequence (y_{k_ℓ}) with limit $y_0 \in Y_h$. By continuity, we conclude

$$\inf_{\widetilde{y}_h \in Y_h} \|y - \widetilde{y}_h\|_Y = \lim_{\ell \to \infty} \|y - y_{k_\ell}\|_Y = \|y - y_0\|_Y.$$

5. step. The proof of (6.21) now follows from a combination of step 2 and step 3: For arbitrary $\widetilde{x}_h \in X_h$, choose $z_h \in Z_h$ by use of step 3. Let $\widetilde{y}_h \in Y_h$ and be arbitrary. We then infer

$$\begin{aligned} \|x - x_h\|_X &\leq \Big(1 + \frac{\|a\|}{\alpha_h}\Big) \|x - z_h\|_X + \frac{\|b\|}{\alpha_h} \|y - \widetilde{y}_h\|_Y \\ &\leq \Big(1 + \frac{\|a\|}{\alpha_h}\Big)\Big(1 + \frac{\|b\|}{\beta_h}\Big) \|x - \widetilde{x}_h\|_X + \frac{\|b\|}{\alpha_h} \|y - \widetilde{y}_h\|_Y. \end{aligned}$$

Now, we take the infimum over $\widetilde{x}_h$ and $\widetilde{y}_h$ and note that, according to finite dimension, this infimum is attained by independent minima. ■

6.2.1 Discrete inf-sup conditions

Often, the continuous inf-sup condition is not that hard to prove, but the discrete one is the problem. The next two lemmata provide a tool to derive the discrete inf-sup condition from the continuous condition.

Lemma 6.11 (M. Fortin). *Let $b : X \times Y \to \mathbb{R}$ denote a continuous bilinearform which satisfies the continuous inf-sup condition*

$$\inf_{0 \neq \lambda \in Y} \sup_{0 \neq u \in X} \frac{b(u, \lambda)}{\|u\|_X \|\lambda\|_Y} \geq \gamma > 0. \tag{6.24}$$

Let $X_h \subset X$ and $Y_h \subset Y$ denote closed subspaces and let $\Pi : X \to X_h$ denote a linear mapping with

$$\begin{aligned} b(u - \Pi u, \lambda) &= 0 \qquad \forall \lambda \in Y_h && (6.25) \\ \|\Pi u\|_X &\leq C_\Pi \|u\|_X \qquad \forall u \in X. && (6.26) \end{aligned}$$

Then, there holds

$$\inf_{0 \neq \lambda \in Y_h} \sup_{0 \neq u \in X_h} \frac{b(u, \lambda)}{\|u\|_X \|\lambda\|_Y} \geq \gamma_N := \frac{\gamma}{C_\Pi} > 0.$$

Proof. Let $\lambda \in Y_h$ and note

$$\begin{aligned} \gamma \|\lambda\|_Y &\overset{(6.24)}{\leq} \sup_{0 \neq v \in X} \frac{b(v, \lambda)}{\|v\|_X} \overset{(6.25)}{=} \sup_{0 \neq v \in X} \frac{b(\Pi v, \lambda)}{\|v\|_X} \\ &\overset{(6.26)}{\leq} C_\Pi \sup_{0 \neq v \in X} \frac{b(\Pi v, \lambda)}{\|\Pi v\|_X} = C_\Pi \sup_{0 \neq v \in \mathrm{range}\Pi} \frac{b(v, \lambda)}{\|v\|_X} \leq C_\Pi \sup_{0 \neq v \in X_h} \frac{b(v, \lambda)}{\|v\|_X} \end{aligned}$$

■

Often, it is easier to generate the operator Π in two steps, as done in the following lemma.

Lemma 6.12. *Let $\Pi_i : X \to X_h$, $i = 1, 2$ denote linear mappings with*

$$\begin{aligned} \|\Pi_1 u\|_X &\leq C_1 \|u\|_X \qquad \forall u \in X \\ \|\Pi_2(\boldsymbol{I} - \Pi_1)u\|_X &\leq C_2 \|u\|_X \qquad \forall u \in X \\ b(u - \Pi_2 u, \lambda) &= 0 \qquad \forall \lambda \in Y_h. \end{aligned}$$

Then, (6.24) *implies the discrete inf-sup condition*

$$\inf_{0 \neq \lambda \in Y_h} \sup_{0 \neq u \in X_h} \frac{b(u,\lambda)}{\|u\|_X \|\lambda\|_Y} \geq \frac{\gamma}{C_1 + C_2}.$$

Proof. Let $\lambda \in Y_h$ and define $\Pi : X \to X_h$ via $\Pi u := \Pi_2(\boldsymbol{I} - \Pi_1)u + \Pi_1 u$. Then, we have

$$b(\Pi u, \lambda) = b(\Pi_2(u - \Pi_1 u), \lambda) + b(\Pi_1 u, \lambda) = b(u - \Pi_1 u, \lambda) + b(\Pi_1 u, \lambda) = b(u, \lambda).$$

Moreover, there holds

$$\|\Pi u\|_X \leq \|\Pi_2(\boldsymbol{I} - \Pi_1)u\|_X + \|\Pi_1 u\|_X \leq (C_1 + C_2)\|u\|_X.$$

This concludes the proof. ■

6.3 The Stokes problem

6.3.1 Setting

We apply the general theory of saddle point-problems from the previous section to the Stokes problem: Let $\Omega \subset \mathbb{R}^2$ be a Lipschitz domain. Find $u = (u_1, u_2) \in H_0^1(\Omega) \times H_0^1(\Omega)$ and $p \in L^2(\Omega)$ such that

$$\begin{aligned} -\Delta u + \nabla p &= f \qquad \text{in } \Omega \qquad (6.27a) \\ \nabla \cdot u &= 0 \qquad \text{in } \Omega \qquad (6.27b) \end{aligned}$$

for given $f = (f_1, f_2)^\top \in L^2(\Omega) \times L^2(\Omega)$. Here, the operator $-\Delta$ is understood component wise, i.e. $\Delta u = (\Delta u_1, \Delta u_2)^\top$.

Remark. From a physical perspective, u denotes the velocity and p the pressure of a fluid in a case where an equilibrium has been reached and the quantities do not depend on time anymore. The incompressibility condition $\nabla \cdot u = 0$ implies that the fluid can not be compressed (e.g. water). The equation $-\Delta u + \nabla p = f$ describes conservation of momentum. The stationary Stokes problem (6.27) stems from a severe simplification of the Navier-Stokes Equations and are physically meaningful only in slow flowing fluids with high viscosity, e.g., honey. □

A weak form can be formulated as

$$\begin{aligned} \int_\Omega \nabla u : \nabla v - \int_\Omega p \nabla \cdot v &= \int_\Omega f v \qquad \forall v \in (H^1_0(\Omega))^2 & (6.28a)\\ -\int_\Omega q \nabla \cdot u &= 0 \qquad \forall q \in L^2(\Omega). & (6.28b)\end{aligned}$$

Obviously, the pressure is unique only up to an additive constant and hence one usually chooses to satisfy $\int_\Omega p = 0$. This motivates the choice of space

$$L^2_\star(\Omega) := \{p \in L^2(\Omega) \mid \int_\Omega p = 0\}. \qquad (6.29)$$

With this side-constraint, (6.28) is equivalent to the problem: Find $(u,p) \in (H^1_0(\Omega))^2 \times L^2_\star(\Omega)$ such that

$$\begin{aligned} a(u,v) + b(v,p) &= x^\star(v) \qquad \forall v \in (H^1_0(\Omega))^2 & (6.30a)\\ b(u,q) &= 0 \qquad \forall q \in L^2_\star(\Omega), & (6.30b)\end{aligned}$$

where

$$\begin{aligned} a(u,v) &= \int_\Omega \nabla u : \nabla v & (6.31a)\\ b(v,p) &= -\int_\Omega p \nabla \cdot v & (6.31b)\end{aligned}$$

Existence of a unique solution for the Stokes problem results from Theorem 6.6 together with the following theorem. (Note that the bilinearform $a(u,v)$ satisfies $a(u,u) = \|\nabla u\|^2_{L^2(\Omega)}$ and is hence elliptic.)

Theorem 6.13 (deRham). *Let Ω be a Lipschitz domain and recall the bilinearform $b(\cdot,\cdot)$ from* (6.31)*. Then, there exists $\gamma > 0$ such that*

$$\inf_{0 \neq p \in L^2_\star(\Omega)} \sup_{0 \neq u \in (H^1_0(\Omega))^2} \frac{|b(v,u)|}{\|p\|_{L^2(\Omega)} \|v\|_{H^1(\Omega)}} \geq \gamma > 0.$$

6.3.2 FEM for Stokes

The finite element method corresponding to (6.30) reads: For $X_h \subset (H^1_0(\Omega))^2$ and $Y_h \subset L^2_\star(\Omega)$ find $(u_h, p_h) \in X_h \times Y_h$ such that

$$\begin{aligned} a(u_h,v) + b(v,p_h) &= x^\star(v) \qquad \forall v \in X_h & (6.32a)\\ b(u_h,q) &= 0 \qquad \forall q \in Y_h. & (6.32b)\end{aligned}$$

From the abstract theory of saddle-point problems (particularly Theorem 6.6) we know that the discrete spaces also need to satisfy an inf-sup condition, i.e.

$$\inf_{0 \neq p \in Y_h} \sup_{0 \neq v \in X_h} \frac{b(v,p)}{\|p\|_{L^2(\Omega)} \|v\|_{H^1(\Omega)}} \geq \gamma_h > 0. \qquad (6.33)$$

Particularly from the Céa Lemma for saddle-point problems (Theorem 6.10), we want to choose discrete spaces which lead to the same rate of convergence

$$\inf_{v\in X_h}\|u-v\|_{H^1(\Omega)}, \qquad \inf_{q\in Y_h}\|p-q\|_{L^2(\Omega)}.$$

This motivates the choice $X_h = (\mathcal{S}_0^1(\mathcal{T}))^2$ and $Y_h = P^0(\mathcal{T}) \cap L^2_\star(\Omega)$. However, this choice does not satisfy a discrete inf-sup condition as we will show using a version of Euler's formula for planar graphs.

Theorem 6.14. *Let $\mathcal{T}$ denote a regular triangulation of a simply connected domain Ω. Then, there holds*

$$\#\mathcal{T} = 2\#(\mathcal{K}\cap\Omega) + \#(\mathcal{K}\cap\partial\Omega) - 2.$$

With this, we see

$$\dim(Y_h) = \#\mathcal{T} - 1 = 2\#(\mathcal{K}\cap\Omega) + \#(\mathcal{K}\cap\partial\Omega) - 3 = \dim(X_h) + \#(\mathcal{K}\cap\partial\Omega) - 3.$$

Since $\#(\mathcal{K}\cap\partial\Omega) - 3 > 0$ for all meshes with more than one element, we see $\dim(Y_h) > \dim(X_h)$. This contradicts the inf-sup condition (see also Proposition 6.4) and hence this discretization does not lead to regular linear systems.

In the following, we discuss a couple of valid choices of discrete spaces.

Theorem 6.15 (Taylor-Hood-type element). *Let $\mathcal{T}$ denote a regular triangulation of Ω. Let*

$$X_h := \left(\mathcal{S}_0^2(\mathcal{T})\right)^2, \qquad Y_h := P^0(\mathcal{T}) \cap L^2_\star(\Omega).$$

Then, there exists a constant $\gamma > 0$, which depends only on the shape regularity of $\mathcal{T}$ such that

$$\inf_{0\neq p\in Y_h}\sup_{0\neq u\in X_M}\frac{b(u,p)}{\|p\|_{L^2(\Omega)}\|u\|_{H^1(\Omega)}} \geq \gamma > 0.$$

Proof. We apply Lemma 6.12. For that, we choose $\Pi_1 : (H_0^1(\Omega))^2 \to (\mathcal{S}_0^1(\mathcal{T}))^2 \subset X_h$ as the Scott-Zhang interpolation operator (or any Clément operator). Particularly, this shows

$$\begin{aligned}\|u-\Pi_1 u\|_{L^2(T)} &\leq Ch_T\|u\|_{H^1(\Omega_T)} \qquad \forall T\in\mathcal{T},\\ \|\Pi_1 u\|_{H^1(\Omega)} &\leq C\|u\|_{H^1(\Omega)}.\end{aligned}$$

The operator Π_2 is defined elementwise via

$$\Pi_2 u \in (\mathcal{S}_0^2(\mathcal{T}))^2 \tag{6.34a}$$

$$(\Pi_2 u)(V) = 0 \qquad \forall V\in\mathcal{N}(\mathcal{T}) \tag{6.34b}$$

$$\int_e \Pi_2 u - u = 0 \qquad \forall e\in\mathcal{E}(\mathcal{T}). \tag{6.34c}$$

Obviously, Π_2 is a well-defined linear operator. Moreover, a scaling argument shows for all $T \in \mathcal{T}$ that

$$\begin{aligned}\|\Pi_2 u\|^2_{L^2(T)} &\le Ch_T^2 \|(\Pi_2 u)\circ\Phi_T\|^2_{L^2(T_{\rm ref})} \le Ch_T^2\|u\circ\Phi_T\|^2_{L^2(\partial T_{\rm ref})} \le Ch_T^2\|u\circ\Phi_T\|^2_{H^1(T_{\rm ref})}\\ &= Ch_T^2\left[\|u\circ\Phi_T\|^2_{L^2(T_{\rm ref})} + \|\nabla u\circ\Phi_T\|^2_{L^2(T_{\rm ref})}\right]\\ &\le C\|u\|^2_{L^2(T)} + Ch_T^2\|\nabla u\|^2_{H^1(T)}, \text{ as well as}\end{aligned}$$

$$\|\nabla\Pi_2 u\|^2_{L^2(T)} \le C\|\nabla(\Pi_2 u)\circ\Phi_T\|^2_{L^2(T_{\rm ref})} \le C\|u\circ\Phi_T\|^2_{L^2(\partial T_{\rm ref})} \le \cdots \le C\left[h_T^{-2}\|u\|^2_{L^2(T)} + \|\nabla u\|^2_{H^1(T)}\right].$$

Altogether, this shows

$$\|\Pi_2 u\|_{H^1(T)} \le C\left[h_T^{-1}\|u\|_{L^2(T)} + |u|_{H^1(T)}\right] \qquad \forall u \in (H^1(T))^2.$$

This implies

$$\|\Pi_2(\boldsymbol{I}-\Pi_1)u\|_{H^1(T)} \le Ch_T^{-1}\|u-\Pi_1 u\|_{L^2(T)} + C\|u-\Pi_1 u\|_{H^1(T)} \le C\|u\|_{H^1(\widetilde{\Omega}_T)}.$$

Summing up over all $T \in \mathcal{T}$ shows $\|\Pi_2(\boldsymbol{I}-\Pi_1)u\|_{H^1(\Omega)} \le C\|u\|_{H^1(\Omega)}$.

For $p \in Y_h$ und $u \in (H^1_0(\Omega))^2$, we obtain

$$-b(u-\Pi_2 u, p) \;=\; \sum_{T\in\mathcal{T}}\int_T p\nabla\cdot(u-\Pi_2 u) = \sum_{T\in\mathcal{T}}\underbrace{\int_{\partial T} p(u-\Pi_2 u)\cdot n_T}_{=0\text{by construction of }\Pi_2} - \int_T \underbrace{\nabla p}_{=0}(u-\Pi_2 u) = 0.$$

■

Theorem 6.16 (MINI-Element). *Let $\mathcal{T}$ a regular triangulation of Ω. Let $B_3 := \{u \in H^1(\Omega)\,|\,u|_T\circ\Phi_T \in \operatorname{span}\{b_3\}\}$, where b_3 is the cubic element bubble function $b_3(x,y) := xy(1-x-y)$ on the reference triangle $T_{\rm ref}$. Let*

$$X_h := \left(\mathcal{S}^1_0(\mathcal{T}) + B_3\right)^2, \qquad\qquad Y_h := \mathcal{S}^1(\mathcal{T}) \cap L^2_\star(\Omega).$$

Then, there exists a constant $\gamma > 0$, which depends only on the shape regularity of $\mathcal{T}$ such that

$$\inf_{0\ne p\in Y_h}\sup_{0\ne u\in X_M}\frac{b(u,p)}{\|p\|_{L^2(\Omega)}\|u\|_{H^1(\Omega)}} \ge \gamma > 0.$$

Proof. Again, we use 6.12. Let Π_1 denote again the Scott-Zhang operator. The operator Π_2 is defined elementwise as follows: The bubble-functions $b_T := b_3\circ\Phi_T^{-1}$ satisfy $\operatorname{supp} b_T \subset T$ und $B_3 = \operatorname{span}\{b_T\,|\,T\in\mathcal{T}\}$. We define

$$\Pi_2 u|_T := \frac{1}{\int_T b_T} b_T \begin{pmatrix}\int_T u_1\\ \int_T u_2\end{pmatrix},$$

with $u = (u_1, u_2)$. Then, there holds

$$\begin{aligned}\Pi_2 : (H^1_0(\Omega))^2 &\to B_3^2 &&\text{is a linear operator}\\ \|\Pi_2 u\|_{L^2(T)} &\le C\|u\|_{L^2(T)} &&\forall T\in\mathcal{T}\\ \|\Pi_2 u\|_{H^1(T)} &\le Ch_T^{-1}\|u\|_{L^2(T)} &&\forall T\in\mathcal{T}.\end{aligned}$$

Analogously to the proof of Theorem 6.15, we obtain $\|\Pi_2(\boldsymbol{I}-\Pi_1)u\|_{H^1(\Omega)} \leq C\|u\|_{H^1(\Omega)}$. Moreover, for $p \in \mathcal{S}^1(\mathcal{T})$:

$$\begin{aligned} b(u-\Pi_2 u, p) &= \int_\Omega p\nabla\cdot(u-\Pi_2 u) = \underbrace{\int_{\partial\Omega} p(u-\Pi_2 u)}_{=0 \text{ due to boundary condition}} - \int_\Omega \nabla p\cdot(u-\Pi_2 u) \\ &= \sum_{T\in\mathcal{T}}\int_K \underbrace{\nabla p|_T}_{=\text{constant}} \cdot(u-\Pi_2 u) \overset{\text{by construction of } \Pi_2}{=} 0 \end{aligned}$$

■

The most widely used discretization for Stokes is the following Tayler-Hood element.

Theorem 6.17 (Taylor-Hood). *Let $\mathcal{T}$ denote a regular triangulation such that each element $T \in \mathcal{T}$ has at most one edge on $\partial\Omega$. Define*

$$X_h := \left(\mathcal{S}_0^2(\mathcal{T})\right)^2, \qquad Y_h := \mathcal{S}^1(\mathcal{T}) \cap L^2_\star(\Omega).$$

Then, there exists a constant $\gamma > 0$ which depends only on the shape regularity of $\mathcal{T}$ such that

$$\inf_{0\neq p\in Y_h}\sup_{0\neq u\in X_M}\frac{b(u,p)}{\|p\|_{L^2(\Omega)}\|u\|_{H^1(\Omega)}} \geq \gamma > 0.$$

6.4 Further remarks on mixed methods

Mixed methods can be useful if a direct discretization of a problem is difficult. We demonstrate this for the biharmonic equation:

$$\begin{aligned} \Delta^2 u &= f \qquad \text{in } \Omega, && (6.35a) \\ u &= 0 \qquad \text{on } \partial\Omega && (6.35b) \\ \partial_n u &= 0 \qquad \text{on } \partial\Omega && (6.35c) \end{aligned}$$

The classical weak form of the biharmonic equation is

$$\text{Find } u \in H_0^2(\Omega) \text{ such that} \qquad \int_\Omega \Delta u \Delta v = \int_\Omega fv \qquad \forall v \in H_0^2(\Omega). \tag{6.36}$$

To derive a FEM for the above problem, we need to choose discrete subspaces $X_h \subseteq H_0^2(\Omega)$. Note that the standard spaces $\mathcal{S}_0^p \not\subset H^2(\Omega)$ do not work. By ensuring C^1-regularity over element interfaces, it is possible to construct piecewise polynomial spaces which are subspaces of $H^2(\Omega)$ (for example the Argyris-element or the Hsieh-Clough-Tocher-element). However, such an implementation is complicated and not very popular among users. It is easier to change the weak form. To that end, we introduce a new variable $\sigma = -\Delta u$. This leads to the following problem: Find

$(u, \sigma) \in H_0^1(\Omega) \times H^1(\Omega)$ such that

$$\begin{aligned} \int_\Omega \nabla\sigma \cdot \nabla w &= \int_\Omega f w \qquad \forall w \in H_0^1(\Omega), && (6.37a) \\ \int_\Omega \nabla u \cdot \nabla v - \int_\Omega \sigma v &= 0 \qquad \forall v \in H^1(\Omega) && (6.37b) \end{aligned}$$

Without looking into the solution theory of the mixed FEM, we note that we want to find $(u, \sigma) \in H_0^1(\Omega) \times H^1(\Omega)$. Hence, we only need to choose finite dimensional subspaces of $H_0^1(\Omega) \times H^1(\Omega)$, which can be done by using classical polynomial spaces.

6.5 The Gårding inequality

Often, an elliptic problem is perturbed by a lower order term such that the resulting problem is no longer elliptic but satisfies a Gårding inequality.

Definition. Let X_0, X_1 denote Hilbert spaces with compact embedding $X_1 \subset X_0$. A bilinearform $a : X_1 \times X_1 \to \mathbb{R}$ satisfies a Gårding inequality if there exist constants C_0, $C_1 > 0$ with

$$a(u,u) \geq C_1 \|u\|_{X_1}^2 - C_0 \|u\|_{X_0}^2 \qquad \forall u \in X_1.$$

Problems which satisfy a Gårding inequality arise for example if one considers PDEs with lower order terms.

Exercise 51. Consider

$$\begin{aligned} -\Delta u - b(x) \cdot \nabla u + c(x) u &= f \quad \text{in } \Omega, \\ u &= 0 \qquad \text{on } \partial\Omega \end{aligned}$$

Show that the corresponding bilinear form satisfies a Gårding inequality with $X_1 = H_0^1(\Omega)$ and $X_0 = L^2(\Omega)$. □

We use the following result from functional analysis:

Exercise 52. Let X, Y denote Banach spaces and let $K : X \to Y$ be a compact operator. Let $(\Pi_N)_{N \in \mathbb{N}}$ a sequence of linear operators $\Pi_N : Y \to Y$ with $\|\Pi_N\| \leq 1$ and $\Pi_N \to \mathrm{Id}$ pointwise (i.e. $\lim_{N\to\infty} \Pi_N u = u$ for all $u \in Y$). Then, there holds

$$\lim_{N\to\infty} \|(\mathrm{Id} - \Pi_h) K\|_{Y \leftarrow X} = 0.$$

□

For well-posed problems (i.e. the continuous equation has a unique solution) with Gårding inequality, the following result shows that their discretization is asymptotically quasi-optimal.

Theorem 6.18. *Let X_1, X_0 be Hilbert spaces with compact embedding $X_1 \subset X_0$. Let $a(\cdot,\cdot)$ satisfy a Gårding inequality and let the induced operator $\mathbf{A} : X_1 \to X_1'$, $\mathbf{A}u := a(u, \cdot)$ be* *bijective. Let $(X_h)_{h>0} \subset X_1$ denote a sequence of closed subspaces such that*

$$\lim_{h\to 0} \inf_{v\in X_h} \|u - v\|_{X_1} = 0 \qquad \forall u \in X_h.$$

Then, there exists $h_0 > 0$ and $\gamma > 0$ such that for all $0 < h < h_0$

$$\inf_{u\in X_h} \sup_{v\in X_h} \frac{a(u,v)}{\|u\|_{X_1}\|v\|_{X_1}} \geq \gamma > 0.$$

In particular, there holds for the FEM error

$$\|u - u_h\|_{X_1} \leq \left(1 + \frac{\|a\|}{\gamma}\right) \inf_{v\in X_h} \|u - v\|_{X_1}$$

Proof. *Step 1:* We show that there exists $\widetilde{C} > 0$ such that for all $u \in X_1$ we find $v \in X_1$ of the form $v = u + z$ such that

$$\begin{aligned} a(u,v) &= a(u, u+z) \geq C_1\|u\|^2_{X_1} && (6.38)\\ \|v\|_{X_1} &\leq \widetilde{C}\|u\|_{X_1}. && (6.39)\end{aligned}$$

The choice of z is motivated by the Gårding inequality $a(u,u) \geq C_1\|u\|^2_{X_1} - C_0\|u\|^2_{X_0}$, i.e

$$a(u, u+z) = a(u,u) + a(u,z) \geq C_1\|u\|^2_{X_1} - C_0\|u\|^2_{X_0} + a(u,z).$$

Hence, we choose $z \in X_1$ as solution of the (adjoint) problem

$$\text{Find } z \in X_1 \text{ s.t.} \quad a(w,z) = C_0\langle w, u\rangle_{X_0} \qquad \forall w \in X_1$$

In operator notation, this reads as

$$\mathbf{A}^\top z = \mathbf{K}u,$$

where $\mathbf{A} : X_1 \to X_1'$ is induced by the bilinearform $a(\cdot,\cdot)$ and $\mathbf{K} : X_1 \to X_1'$ is defined via

$$\langle \cdot, \mathbf{K}u\rangle_{X_1\times X_1'} = \langle \cdot, C_0 u\rangle_{X_0}.$$

We note that

(i) Since $\mathbf{A}$ is bijective, also $\mathbf{A}^\top$ bijective and $\|\mathbf{A}^{-\top}\| = \|\mathbf{A}^{-1}\|$.

(ii) Since $X_1 \subset X_0$ is compact, also $\mathbf{K} : X_1 \to X_1'$ is a compact operator.

The operator $\mathbf{A}^{-\top}\mathbf{K} : X_1 \to X_1$, which maps u to z is compact. We obtain

$$\begin{aligned} a(u,v) &= a(u, u+z) \geq C_1\|u\|^2_{X_1}\\ \|v\|_{X_1} &\leq \|u\|_{X_1} + \|z\|_{X_1} \leq \left(1 + \|\mathbf{A}^{-\top}\mathbf{K}\|\right)\|u\|_{X_1}.\end{aligned}$$

Step 2: For given $u \in X_h$, we construct $v \in X_h$ such that

$$\begin{aligned} a(u,v) &\geq \frac{C_1}{2}\|u\|^2_{X_1}\\ \|v\|_{X_1} &\leq \widetilde{C}\|u\|_{X_1}\end{aligned}$$

Let $\Pi_h : X_1 \to X_h$ denote the orthogonal projection (in X_1). Following Step 1, we define $v = u + \Pi_h z \in X_h$ for given $u \in X_h$. This shows

$$\begin{aligned} a(u,v) &= a(u,u+z) + a(u,\Pi_h z - z) \geq C_1\|u\|_{X_1}^2 - \|a\|\|u\|_{X_1}\|(\mathrm{Id} - \Pi_h)z\|_{X_1} \\ &\geq C_1\|u\|_{X_1}^2 - \|a\|_{X_1}\|(\mathrm{Id} - \Pi_h)\mathbf{A}^{-\top}\mathbf{K}\|\|u\|_{X_1}^2 \end{aligned}$$

Since $\mathbf{K}$ is compact and $\mathbf{A}^{-\top}$ bounded, also their composition is compact. Since $\mathrm{Id} - \Pi_h$ converges to zero pointwise (according to the assumption), we obtain with Exercise 52 that $\lim_{h\to 0} \|(\mathrm{Id} - \Pi_h)\mathbf{A}^{-\top}\mathbf{K}\| = 0$. Hence, there exists $h_0 > 0$ (independently of u) such that all $0 < h < h_0$ satisfy

$$a(u, u + \Pi_h z) \geq \frac{C_1}{2}\|u\|_{X_1}^2.$$

Furthermore, $v = u + \Pi_h z \in X_h$ satisfies

$$\|v\|_{X_1} \leq \|u\|_{X_1} + \|\Pi_h z\|_{X_1} \leq \|u\|_{X_1} + \|z\|_{X_1} \leq (1 + \|\mathbf{A}^{-\top}\mathbf{K}\|)\|u\|_{X_1}.$$

Step 3: This shows the discrete inf-sup condition. Quasi-optimality follows from the Céa lemma. ■

A bilinear form which satisfies a Gårding inequality does not necessarily induce a bijective operator. A famous result from functional analysis states however, that injectivity implies already bijectivity.

Theorem 6.19 (Fredholmalternative). *Let a denote a bounded bilinearform on the Hilbertspace X_1, which satisfies a Gårding inequality. Let the induced operator $\mathbf{A} : X_1 \to X_1'$ be injective, i.e.*

$$a(u,v) = 0 \qquad \forall v \in X_1 \qquad \Longrightarrow u = 0.$$

Then, $\mathbf{A}$ is already bijective.

Proof. The Gårding inequality states

$$a(u,u) \geq C_1\|u\|_{X_1}^2 - C_0\|u\|_{X_0}^2$$

Consider $\widetilde{a} : X_1 \times X_1 \to \mathbb{R}$ defined by

$$\widetilde{a}(u,v) := a(u,v) + C_0\langle u,v\rangle_{X_0}$$

Due to the Lax-Milgram Lemma $\widetilde{a}(\cdot,\cdot)$ induces a bijective operator $\widetilde{\mathbf{A}} : X_1 \to X_1'$. The difference $\mathbf{K} := \widetilde{\mathbf{A}} - \mathbf{A} : X_1 \to X_1'$ is compact since

$$\langle \mathbf{K}u, v\rangle_{X_1' \times X_1} = C_0\langle u,v\rangle_{X_0}$$

and $X_1 \subset X_0$ is compact. Hence $\mathbf{A}$ reads as

$$\mathbf{A} = \widetilde{\mathbf{A}} - \mathbf{K} = \widetilde{\mathbf{A}}\left(\mathrm{Id} - \widetilde{\mathbf{A}}^{-1}\mathbf{K}\right).$$

The injectivity of $\mathbf{A}$ implies that 1 is not an Eigenvalue of the compact operator $\widetilde{\mathbf{A}}^{-1}\mathbf{K}$. The theory of compact operators shows that this implies that $\mathrm{Id} - \widetilde{\mathbf{A}}^{-1}\mathbf{K}$ is invertible and hence $\mathbf{A}$ is bijective. ■