Chapter 6
Saddle-Point Problems

6.1 Discrete Saddle-Point Problems

6.1.1 Limitations of the Lax–Milgram Framework

Suppose that a fluid is flowing through the fixed domain \( \Omega \subseteq \mathbb{R}^d, d = 2, 3 \). At each point \( x \in \Omega \), we let \( u(x) \in \mathbb{R}^d \) be the velocity of the fluid and \( p(x) \in \mathbb{R} \) its pressure, cf. Fig. 6.1.

If the process is stationary in the sense that the velocity is the same at all times, and if the fluid is viscous and incompressible, then a simplification of the Navier–Stokes equations leads to the Stokes system, which specifies \( u \) and \( p \) as the solution of the partial differential equations

\[
- \Delta u + \nabla p = f \quad \text{in } \Omega, \\
\text{div } u = 0 \quad \text{in } \Omega,
\]

subject to \( u|_{\Gamma_D} = 0 \), where we assume for simplicity that \( \Gamma_D = \partial \Omega \). Here \( \Delta u = [\Delta u_1, \ldots, \Delta u_d]^T \). To obtain a variational formulation, we multiply the first equation by \( v \in H^1_D(\Omega; \mathbb{R}^d) \), the second one by \( q \in L^2(\Omega) \), integrate over \( \Omega \), and use integration-by-parts to see that \( u \) and \( p \) satisfy

\[
\int_{\Omega} \nabla u : \nabla v \, dx - \int_{\Omega} p \, \text{div } v \, dx = \int_{\Omega} f \cdot v \, dx, \\
\int_{\Omega} q \, \text{div } u \, dx = 0,
\]

for all \( v \in H^1_D(\Omega; \mathbb{R}^d) \) and \( q \in L^2(\Omega) \). Note that \( \nabla u \) and \( \nabla v \) are square matrices, and \( \nabla u : \nabla v \) is the scalar product defined by the sum \( \sum_{j=1}^d \nabla u_j \cdot \nabla v_j \). Abbreviating
Fig. 6.1 Stokes flow through a domain $\Omega \subset \mathbb{R}^d$ described by a velocity field $u : \Omega \to \mathbb{R}^d$

$V = H^1_D(\Omega; \mathbb{R}^d)$ and $Q = L^2(\Omega)$, and introducing the bilinear forms

$$a(u, v) = \int_\Omega \nabla u : \nabla v \, dx, \quad b(v, p) = -\int_\Omega p \, \text{div} \, v \, dx,$$

and the linear functional

$$\ell_1(v) = \int_\Omega f \cdot v \, dx,$$

the weak formulation is equivalent to finding $(u, p) \in V \times Q$ such that

$$(S) \quad \begin{cases} a(u, v) + b(v, p) = \ell_1(v), \\ b(u, q) = 0 \end{cases}$$

for all $(v, q) \in V \times Q$. To relate this formulation to the Lax–Milgram lemma, we consider the product space $X = V \times Q$ and define for $x = (u, p)$ and $y = (v, q)$ the bilinear form

$$\Gamma(x, y) = a(u, v) + b(v, p) + b(u, q)$$

and the linear form

$$\ell(y) = \ell_1(v).$$

The formulation $(S)$ is then equivalent to finding $x \in X$ such that

$$\Gamma(x, y) = \ell(y)$$

for all $y \in X$. Unfortunately, the bilinear form is not coercive on $X$. To see this, we choose $x = (0, p)$ and verify that

$$\Gamma(x, x) = 0.$$

Therefore, we cannot apply the Lax–Milgram theorem to establish existence and uniqueness of a solution $(u, p)$ for $(S)$.

Remark 6.1 By restricting to the subspace $V_0 = \{v \in H^1_D(\Omega; \mathbb{R}^d) : \text{div} \, v = 0\}$, one can prove the existence of a unique velocity field $u \in H^1_D(\Omega; \mathbb{R}^d)$ with the Lax–Milgram lemma via the formulation

$$a(u, v) = \ell_1(v)$$
for all \( v \in V_0 \). The construction of discrete subspaces of \( V_0 \), i.e., of divergence-free finite element functions, is however a nontrivial task. Moreover, the pressure is a relevant quantity in many applications.

### 6.1.2 Variational Condition Number Estimate

Throughout what follows we adopt arguments from [7] and recall that the condition number of a regular matrix \( M \in \mathbb{R}^{n \times n} \) is with norms \( \| \cdot \|_\ell \) and \( \| \cdot \|_r \) defined as the smallest constant \( \text{cond},_r(M) \geq 0 \) such that

\[
\frac{\| x - \tilde{x} \|_\ell}{\| x \|_\ell} \leq \text{cond},_r(M) \frac{\| b - \tilde{b} \|_r}{\| b \|_r}
\]

for all \( x, \tilde{x} \in \mathbb{R}^n \) and \( b, \tilde{b} \in \mathbb{R}^n \) such that \( x, b \neq 0 \) and

\[ Mx = b, \quad M\tilde{x} = \tilde{b}. \]

The condition number measures the amplification of the relative errors for perturbations of the right-hand sides in solving the linear system \( Mx = b \). Defining the induced operator norms

\[
\| M \|_{\ell_r} = \sup_{y \in \mathbb{R}^n \setminus \{0\}} \frac{\|My\|_r}{\|y\|_\ell}, \quad \| M^{-1} \|_{\ell \ell} = \sup_{z \in \mathbb{R}^n \setminus \{0\}} \frac{\|M^{-1}z\|_\ell}{\|z\|_r},
\]

we have

\[
\text{cond},_r(M) = \| M \|_{\ell_r} \| M^{-1} \|_{\ell \ell} \geq 1.
\]

If \( \| \cdot \|_\ell \) and \( \| \cdot \|_r \) coincide with the Euclidean norm on \( \mathbb{R}^n \), and if \( M \) is symmetric and positive definite, then we have \( \text{cond},_r(M) = \lambda_{\text{max}}/\lambda_{\text{min}} \) with the maximal and minimal eigenvalues \( \lambda_{\text{max}} \) and \( \lambda_{\text{min}} \) of \( M \). It is often straightforward to construct a norm \( \| \cdot \|_\ell \) such that there exists \( k \geq 0 \) with

\[
y^T Mx \leq k \| x \|_\ell \| y \|_\ell
\]

for all \( x, y \in \mathbb{R}^n \) independently of \( n \in \mathbb{N} \). This induces a natural choice for \( \| \cdot \|_r \) that leads to a uniform bound for \( \| M \|_{\ell_r} \). The dual norm of \( \| \cdot \|_\ell \) is defined by

\[
\| z \|_{\ell'} = \sup_{y \in \mathbb{R}^n \setminus \{0\}} \frac{y^T z}{\|y\|_\ell}.
\]

This norm leads to a useful characterization of the operator norm.
Lemma 6.1 (Operator Norm $\|M\|_{\ell r}$) If $\| \cdot \|_{r} = \| \cdot \|_{\ell r}$, we have

$$\|M\|_{\ell r} = \sup_{x,y \in \mathbb{R}^n} \frac{y^T M x}{\|y\| \|x\|}. $$

Proof Due to the assumption and the definition of $\| \cdot \|_{\ell r}$, we have

$$\|M\|_{\ell r} = \sup_{x \in \mathbb{R}^n} \frac{\|Mx\|_r}{\|x\|_\ell}$$

$$= \sup_{x \in \mathbb{R}^n} \frac{\|Mx\|_r}{\|x\|_\ell}$$

$$= \sup_{x \in \mathbb{R}^n} \sup_{y \in \mathbb{R}^n} \frac{y^T M x}{\|y\|_\ell \|x\|_\ell}$$

$$= \sup_{x,y \in \mathbb{R}^n} \frac{y^T M x}{\|y\|_\ell \|x\|_\ell},$$

which proves the identity.

To specify a uniform upper bound for the condition numbers, it remains to bound $\|M^{-1}\|_{\ell r}$. It is desirable to avoid the explicit use of $M^{-1}$ for this.

Lemma 6.2 (Operator Norm $\|M^{-1}\|_{\ell r}$) If $\| \cdot \|_{r} = \| \cdot \|_{\ell r}$, then

$$\left(\|M^{-1}\|_{\ell r}\right)^{-1} = \inf_{x \in \mathbb{R}^n} \sup_{y \in \mathbb{R}^n} \frac{y^T M x}{\|y\|_\ell \|x\|_\ell}. $$

Proof With the transformation $z = Mx$, we find that

$$\left(\|M^{-1}\|_{\ell r}\right)^{-1} = \left(\sup_{z \in \mathbb{R}^n} \frac{\|M^{-1}z\|_r}{\|z\|_\ell}\right)^{-1}$$

$$= \inf_{z \in \mathbb{R}^n} \frac{\|z\|_r}{\|M^{-1}z\|_\ell}$$

$$= \inf_{x \in \mathbb{R}^n} \frac{\|Mx\|_r}{\|x\|_\ell}$$

$$= \inf_{x \in \mathbb{R}^n} \frac{\|Mx\|_\ell r}{\|x\|_\ell}$$

$$= \inf_{x \in \mathbb{R}^n} \sup_{y \in \mathbb{R}^n} \frac{y^T M x}{\|y\|_\ell \|x\|_\ell},$$

which proves the identity.

We summarize the observations in the following proposition.
6.1 Discrete Saddle-Point Problems

Proposition 6.1 (Condition Number Bound) Let $M \in \mathbb{R}^{n \times n}$ be regular, $\|\cdot\|_\ell$ a norm on $\mathbb{R}^n$, and $\|\cdot\|_r = \|\cdot\|_\ell$. If $k \geq 0$ and $\gamma > 0$ are such that

$$
\|M\|_{\ell r} = \sup_{x,y \in \mathbb{R}^n \setminus \{0\}} \frac{y^T M x}{\|y\|_\ell \|x\|_\ell} \leq k,
$$

$$
\|M^{-1}\|^{-1}_{r\ell} = \inf_{x \in \mathbb{R}^n \setminus \{0\}} \sup_{y \in \mathbb{R}^n \setminus \{0\}} \frac{y^T M x}{\|y\|_\ell \|x\|_\ell} \geq \gamma,
$$

then we have

$$\text{cond}_{\ell r}(M) \leq k/\gamma.$$ 

The characterization of the operator norms is particularly useful when the matrix $M$ results from the discretization of the weak formulation of a partial differential equation, and when uniform bounds for operator norms are needed.

Remarks 6.2

(i) Note that the second condition holds if and only if for all $x \in \mathbb{R}^n$, we have

$$
\sup_{y \in \mathbb{R}^n \setminus \{0\}} \frac{y^T M x}{\|y\|_\ell \|x\|_\ell} \geq \gamma
$$

which is equivalent to $\|Mx\|_{\ell r} \geq \gamma \|x\|_\ell$ for all $x \in \mathbb{R}^n$ and to $\|M^{-1}z\|_{r\ell} \leq \gamma^{-1} \|z\|_\ell$ for all $z \in \mathbb{R}^n$, i.e., $\gamma^{-1}$ is an upper bound for the operator norm of $M^{-1}$.

(ii) The matrix $M$ may result from the discretization of a bilinear form $a : V \times V \to \mathbb{R}$ so that

$$
y^T M x = a(u_h, v_h)
$$

with respect to an appropriate basis. The bound for the condition number can then be determined in terms of $a$.

We illustrate the application of the condition number bound for the discretization of an elliptic partial differential equation.

Example 6.1 Assume that $a : V \times V \to \mathbb{R}$ is a bounded and coercive bilinear form with constants $\beta, \alpha > 0$, $V_h = \text{span}\{v_1, v_2, \ldots, v_n\}$ is a finite-dimensional subspace, and $M_{jk} = a(v_j, v_k)$ for $j, k = 1, 2, \ldots, n$. Defining $\|x\|_\ell = \|u_h\|_V$ if $x \in \mathbb{R}^n$ is the coefficient vector of $u_h \in V_h$ with respect to the basis $(v_1, v_2, \ldots, v_n)$, we have

$$
\frac{y^T M x}{\|x\|_\ell \|y\|_\ell} = \frac{a(u_h, v_h)}{\|u_h\|_V \|v_h\|_V}
$$
and the boundedness and coercivity of $a$ lead to $\text{cond}_{\ell_r}(M) \leq \beta/\alpha$. This upper bound is also the constant that arises in Céa’s lemma. For the Poisson problem we thus obtain a uniform upper bound of the condition number with respect to the norm of $H^1(\Omega)$. For the Euclidean norm on $\mathbb{R}^n$, the condition number is of order $O(h^{-2})$ for typical finite element methods.

6.1.3 Well-Posedness

Motivated by the weak formulation ($S$) and following [7], we analyze linear systems of equations $Mx = b$, for which $M \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$ admit the block structures

$$M = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}, \quad b = \begin{bmatrix} f \\ g \end{bmatrix}.$$  

Here $A \in \mathbb{R}^{n_A \times n_A}$, $B \in \mathbb{R}^{n_B \times n_A}$, and $f \in \mathbb{R}^{n_A}$, $g \in \mathbb{R}^{n_B}$ with $n_A, n_B \in \mathbb{N}$ such that $n = n_A + n_B$. Accordingly, we partition the solution vector $x \in \mathbb{R}^n$ as

$$x = \begin{bmatrix} y \\ z \end{bmatrix}$$

with $y \in \mathbb{R}^{n_A}$ and $z \in \mathbb{R}^{n_B}$. The linear system of equations $Mx = b$ is thus equivalent to the vectorial equations

$$Ay + B^Tz = f, \quad By = g.$$  

Provided that $B$ is a surjection, the second equation determines $y$ up to an element from the kernel of $B$, i.e., in

$$\ker B = \{v \in \mathbb{R}^{n_A} : Bv = 0\}.$$  

We note that according to elementary results from linear algebra, we have

$$\mathbb{R}^{n_A} = \ker B \oplus \text{Im} B^T.$$  

In particular, every $v \in \mathbb{R}^{n_A}$ can be written as $v = v_K + v_I$ with $v_K \in \ker B$ and $v_I \in \text{Im} B^T$ such that $v_K \cdot v_I = 0$. By an appropriate choice of basis, we may assume
that with \( r = \dim \ker B \), we have

\[
v_K = \begin{bmatrix} v_1 \\ \vdots \\ v_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad v_I = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ v_{r+1} \\ \vdots \\ v_{n_A} \end{bmatrix},
\]

and we can identify \( v_K = [v_1, \ldots, v_r]^T \) and \( v_I = [v_{r+1}, \ldots, v_{n_A}]^T \). We then have

\[
Av = \begin{bmatrix} A_{KK} & A_{KI} \\ A_{IK} & A_{II} \end{bmatrix} \begin{bmatrix} v_K \\ v_I \end{bmatrix} = \begin{bmatrix} A_{KK}v_K + A_{KI}v_I \\ A_{IK}v_K + A_{II}v_I \end{bmatrix}
\]

and analogously

\[
Bv = \begin{bmatrix} B_K & B_I \end{bmatrix} \begin{bmatrix} v_K \\ v_I \end{bmatrix} = B_Kv_K + B_Iv_I = B_Iv_I.
\]

In particular, we have \( B_Iv_I = 0 \) if and only if \( v_I = 0 \). Moreover, up to appropriate identification of vectors, we have

\[
B^\top s = B_I^\top s \in \text{Im} B^\top
\]

for all \( s \in \mathbb{R}^{n_B} \), which follows from noting \((B^\top s) \cdot v = s \cdot (Bv) = s \cdot (B_Iv_I) = (B_I^\top s) \cdot v_I \). With the decomposition of vectors in \( \mathbb{R}^n \), the system \( Mx = b \) reads as

\[
\begin{align*}
A_{KK}y_K + A_{KI}y_I &= f_K, \\
A_{IK}y_K + A_{II}y_I + B_I^\top z &= f_I, \\
B_Iy_I &= g.
\end{align*}
\]

The conditions for unique solvability are:

- equation \( A_{KK}y_K = f_K \) is uniquely solvable for every \( f_K \in \ker B \);
- equation \( B_Iy_I = g \) is uniquely solvable for every \( g \in \mathbb{R}^{n_B} \);
- equation \( B_I^\top z = f_I \) is uniquely solvable for every \( f_I \in \text{Im} B^\top \).

To ensure the first condition, we need that \( B_I : \text{Im} B^\top \mapsto \mathbb{R}^{n_B} \) be invertible. Since \( B_I \) is by construction injective, this means that \( B \) has to be surjective, i.e., \( \dim \text{Im} B^\top = n_B \) and \( B_I \) is a regular square matrix. This automatically yields that also \( B_I^\top \) is regular, so that the third condition is satisfied. The second condition holds if \( A_{KK} : \ker B \mapsto \ker B \) is invertible.
Proposition 6.2 (Solvability) The linear system of equations

\[
\begin{bmatrix}
A & B^T \\
B & 0
\end{bmatrix}
\begin{bmatrix}
y \\
z
\end{bmatrix}
= \begin{bmatrix}
f \\
g
\end{bmatrix}
\]

is uniquely solvable for every right-hand side \( b = [f, g]^T \), if and only if \( B : \mathbb{R}^n \to \mathbb{R}^m \) is surjective, and the restriction of \( A \) to \( \ker B \) defines a bijection.

Example 6.2 The matrix

\[
\begin{bmatrix}
1 & 0 & 1 \\
0 & 0 & -1 \\
1 & -1 & 0
\end{bmatrix}
\]

is invertible. This follows from the proposition by letting \( A \) be the upper left \( 2 \times 2 \) submatrix and \( B = [1, -1] \).

We next investigate conditioning of the linear system. For this, we assume that there exist norms \( \| \cdot \|_V \) and \( \| \cdot \|_Q \) on \( \mathbb{R}^n \) and \( \mathbb{R}^m \), respectively, such that

\[
v^T A y \leq k_A \| v \|_V \| y \|_V, \quad v^T B^T z \leq k_B \| v \|_V \| z \|_Q
\]

for all \( v, y \in \mathbb{R}^n \) and all \( z \in \mathbb{R}^m \) with uniformly bounded constants \( k_A, k_B \geq 0 \). On the product space \( \mathbb{R}^n \times \mathbb{R}^m \), we employ the norm

\[
\|(y, z)\|_\ell = \|y\|_V + \|z\|_Q,
\]

note that the associated dual norm is given by

\[
\|(f, g)\|_{\ell'} = \max \{ \|f\|_{V'}, \|g\|_{Q'} \},
\]

and we use \( \| \cdot \|_r = \| \cdot \|_{\ell'} \). We then have

\[
\|M\|_{lr} \leq k_A + 2k_B.
\]

To provide an upper bound for the condition number, it remains to estimate \( \|M^{-1}\|_{rl} \), i.e., to specify a constant \( c \geq 0 \) such that \( \|M^{-1}b\|_\ell \leq c \|b\|_r \). In constructing \( x = [y, z]^T = M^{-1}[f, g]^T \), we have that

\[
\|y_I\|_V \leq \|B_I^{-1}\|_{Q'V} \|g\|_{Q'},
\]

\[
\|y_K\|_V \leq \|A_{KK}^{-1}\|_{V'V}(\|f_K\|_V + \|A_{KL}y_I\|_V)
\leq \|A_{KK}^{-1}\|_{V'V}(\|f_K\|_V + k_A \|y_I\|_V),
\]

\[
\|z\|_Q \leq \|(B_I^{-1})^{-1}\|_{V'Q}(\|f_I\|_{V'} + \|A_{II}y_I\|_{V'} + \|A_{IK}y_K\|_{V'})
\leq \|(B_I^{-1})^{-1}\|_{V'Q}(\|f_I\|_{V'} + k_A \|y_I\|_V + k_A \|y_K\|_V).
\]
Assuming that \( \|B_i^{-1}\|_{Q^*}, \|(B_i^T)^{-1}\|_{Q^*} \leq \beta^{-1} \) and \( \|A_{KK}^{-1}\|_{V^*} \leq \alpha^{-1} \), we obtain
\[
\| (y, z) \| \leq c(\alpha^{-1}, \beta^{-1}, k_A) \| (f, g) \|_r,
\]
and the constant \( c(\alpha^{-1}, \beta^{-1}, k_A) \) provides an upper bound for \( \|M^{-1}\|_{r, \ell} \). Note that \( (B_i^T)^{-1} = (B_i^{-1})^T \), and due to the choice of norms
\[
\|B_i^{-1}\|_{Q^*} = \|(B_i^T)^{-1}\|_{V^*}.
\]
It therefore suffices to bound \( \|(B_i^T)^{-1}\|_{V^*} \) and \( \|A_{KK}^{-1}\|_{V^*} \). With the transformation \( f_i = B_i^T q = B^T q \), we have
\[
\|(B_i^T)^{-1}\|_{V^*}^{-1} = \inf_{f_i \in \mathbb{R}^A \setminus \{0\}} \frac{\|f_i\|_{V^*}}{\|(B_i^T)^{-1} f_i\|_Q} = \inf_{q \in \mathbb{R}^B \setminus \{0\}} \frac{\|B_i^T q\|_{V^*}}{\|q\|_Q} = \inf_{q \in \mathbb{R}^B \setminus \{0\}} \frac{\|B^T q\|_{V^*}}{\|q\|_Q} = \inf_{q \in \mathbb{R}^B \setminus \{0\}} \sup_{v \in \mathbb{R}^A \setminus \{0\}} \frac{v^T B^T q}{\|q\|_Q \|v\|_V} = \inf_{q \in \mathbb{R}^B \setminus \{0\}} \sup_{v \in \mathbb{R}^A \setminus \{0\}} \frac{q^T B v}{\|q\|_Q \|v\|_V}.
\]
Similarly, with the transformation \( v = A_{KK} u \), we verify that
\[
\|A_{KK}^{-1}\|_{V^*}^{-1} = \inf_{u \in \mathbb{K} \setminus \{0\}} \frac{\|u\|_{V^*}}{\|A_{KK}^{-1} u\|_V} = \inf_{u \in \mathbb{K} \setminus \{0\}} \frac{\|A_{KK} u\|_{V^*}}{\|u\|_V} = \inf_{u \in \mathbb{K} \setminus \{0\}} \sup_{w \in \mathbb{K} \setminus \{0\}} \frac{w^T A_{KK} u}{\|u\|_V \|w\|_V} = \inf_{u \in \mathbb{K} \setminus \{0\}} \sup_{w \in \mathbb{K} \setminus \{0\}} \frac{w^T A u}{\|u\|_V \|w\|_V}.
\]
With these estimates we obtain a bound on the condition number.
Theorem 6.1 (Conditioning) Assume that we are given the linear systems of equations \( Mx = b \) with

\[
M = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}, \quad b = \begin{bmatrix} f \\ g \end{bmatrix},
\]

and assume that there exist constants \( k_A, k_B \geq 0 \) such that

\[
v^T Ay \leq k_A \|v\|_V \|y\|_V, \quad v^T Bz \leq k_B \|v\|_V \|z\|_Q.
\]

for all \( y, v \in \mathbb{R}^{nA} \) and \( z \in \mathbb{R}^{nB} \). Assume that \( \alpha, \beta > 0 \) are such that with \( K = \ker B \), we have the inf-sup conditions

\[
\inf_{v \in K \setminus \{0\}} \sup_{w \in K \setminus \{0\}} \frac{v^T Aw}{\|v\|_V \|w\|_V} \geq \alpha,
\]

\[
\inf_{q \in \mathbb{R}^{nB} \setminus \{0\}} \sup_{v \in \mathbb{R}^{nA} \setminus \{0\}} \frac{q^T Bv}{\|q\|_Q \|v\|_V} \geq \beta.
\]

Then the condition number of \( M \) with respect to \( \|\cdot\|_\ell r = \|\cdot\|_V + \|\cdot\|_Q \) and the associated dual norm \( \|\cdot\|_r \) satisfy

\[
\text{cond}_{\ell r}(M) \leq c(k_A, k_B, \alpha^{-1}, \beta^{-1}).
\]

Remarks 6.3

(i) The second inf-sup condition is equivalent to

\[
\|B^T q\|_{V'} = \sup_{v \in \mathbb{R}^{nA} \setminus \{0\}} \frac{q^T Bv}{\|v\|_V} \geq \beta \|q\|_Q
\]

for all \( q \in \mathbb{R}^{nB} \). It implies that \( B^T \) is injective and \( B \) is surjective. In particular, \( B^T \) and \( B \) have left and right inverses \((B^T)^{-1} : \text{Im} B^T = (\ker B)^\perp \to \mathbb{R}^{nB}\) and \( B^{-r} : \mathbb{R}^{nB} \to \mathbb{R}^{nA}\), respectively, whose operator norms are bounded by \( \beta^{-1} \).

(ii) The inf-sup conditions imply the solvability of the linear system of equations.

6.1.4 Constrained Quadratic Minimization

Quadratic minimization problems with a linear constraint consist in finding a solution for the problem

\[
\min_{y \in \mathbb{R}^{nA}} \frac{1}{2} y^T Ay - f^T y \quad \text{subject to } By = g.
\]
Here $A \in \mathbb{R}^{n_A \times n_A}$ is a symmetric and positive semidefinite and $B \in \mathbb{R}^{n_B \times n_A}$ is a surjective matrix. Noting that

$$By = g \iff \max_{z \in \mathbb{R}^{n_B}} z^T (By - g) < \infty,$$

we may equivalently consider solving the *saddle-point* or *min-max problem*

$$\min_{y \in \mathbb{R}^{n_A}} \max_{z \in \mathbb{R}^{n_B}} \frac{1}{2} y^T Ay - f^T y + z^T (By - g).$$

Letting $L(y, z) = (1/2)y^T Ay - f^T y + z^T (By - g)$ be the *Lagrange functional* associated with the constrained minimization problem, a solution $(y, z)$ of the saddle-point problem is characterized via the inequalities

$$L(y, s) \leq L(y, z) \leq L(r, z)$$

for all $(r, s) \in \mathbb{R}^{n_A} \times \mathbb{R}^{n_B}$. The optimality conditions for this problem require an optimal pair $(y, z)$ to satisfy

$$\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}.$$

The variable $z$ is also called the *Lagrange multiplier* subject to the constraint $By = g$.

**Example 6.3** The Poisson problem $-\Delta u = f$ with Neumann boundary conditions $\partial_n u = g$ on the entire boundary $\Gamma_N = \partial \Omega$ admits a unique solution, if

$$\int_{\Omega} f \, dx + \int_{\Gamma_N} g \, ds = 0$$

and if one imposes the constraint $\int_{\Omega} u \, dx = 0$. The solution can then be characterized as the unique minimizer of the functional

$$I(u) = \frac{1}{2} \int_{\Omega} |\nabla u|^2 \, dx - \int_{\Omega} fu \, dx - \int_{\Gamma_N} gu \, ds$$

subject to the constraint. Its discretization leads to a constrained quadratic minimization problem.
6.1.5 Iterative Solution

We consider a generalized saddle-point problem of the form

\[
\begin{bmatrix}
A & B^T \\
B & -C
\end{bmatrix}
\begin{bmatrix}
y \\
z
\end{bmatrix}
= \begin{bmatrix}
f \\
g
\end{bmatrix}
\]

and assume that
- \( A \in \mathbb{R}^{n_A \times n_A} \) is symmetric and positive definite,
- \( B \in \mathbb{R}^{n_B \times n_A} \) satisfies \( \text{rank}(B) = m \leq n \),
- \( C \in \mathbb{R}^{n_B \times n_B} \) is symmetric and positive semidefinite.

The system is equivalent to the decoupled system

\[
(BA^{-1}B^T + C)z = BA^{-1}f - g,
\]
\[
y = A^{-1}(f - B^Tz).
\]

The first equation is called the Schur complement (equation), with the symmetric and positive definite matrix \( BA^{-1}B^T + C \), and allows for the computation of \( z \). Solving the second equation then determines \( y \). The preconditioned Uzawa algorithm uses a preconditioner \( D \) for the Schur complement and iterates the following steps to approximate \( z \) and \( y \).

**Algorithm 6.1 (Uzawa Algorithm)** Choose \( z^0 \in \mathbb{R}^{n_B}, \tau, \epsilon_{\text{stop}} > 0 \), and set \( k = 0 \).

1. Solve \( Ay^{k+1} = f - B^Tz^k \).
2. Solve \( D(z^{k+1} - z^k) = \tau(By^{k+1} - Cz^k - g) \).
3. Stop if \( \|z^{k+1} - z^k\|_Q \leq \epsilon_{\text{stop}} \); otherwise set \( k \rightarrow k + 1 \) and continue with (1).

Since we have

\[
\frac{1}{\tau}D(z^{k+1} - z^k) + (BA^{-1}B^T + C)z^k = BA^{-1}f - g,
\]

we see that the algorithm is simply a Richardson iteration for the Schur complement. Convergence holds for \( \tau \) sufficiently small. The preconditioned Arrow–Hurwicz algorithm employs preconditioners \( G \) and \( D \) for \( A \) and \( BA^{-1}B^T + C \), respectively.

**Algorithm 6.2 (Arrow–Hurwicz Algorithm)** Choose \( y^0 \in \mathbb{R}^n, z^0 \in \mathbb{R}^m \) and parameters \( \omega, \tau, \epsilon_{\text{stop}} > 0 \), and set \( k = 0 \).

1. Solve \( G(y^{k+1} - y^k) = \omega(f - Ay^k - B^Tz^k) \).
2. Solve \( -D(z^{k+1} - z^k) = \tau(g - By^{k+1} + Cz^k) \).
3. Stop if \( \|z^{k+1} - z^k\|_Q \leq \epsilon_{\text{stop}} \); otherwise set \( k \rightarrow k + 1 \) and continue with (1).
Remark 6.4 Augmented Lagrangian methods introduce a quadratic term \( r|By - g|^2 \) with a parameter \( r > 0 \) in the saddle-point problem. The additional term vanishes for the solution of the saddle point problem, but often improves the performance of the iterative schemes since the matrix \( A + rB^T B \) is positive definite, provided that \( A \) is positive definite on the kernel of \( B \).

6.2 Continuous Saddle-Point Problems

6.2.1 Closed Range Theorem

Let \( X \) and \( Z \) be Banach spaces whose duals are denoted by \( X' \) and \( Z' \), respectively. Given any \( \varphi \in X' \) and \( \psi \in Z' \), we write the application of \( \varphi \) to \( x \in X \) and of \( \psi \) to \( z \in Z \) as

\[
\langle \varphi, x \rangle = \varphi(x), \quad \langle \psi, z \rangle = \psi(z).
\]

Let \( L : X \rightarrow Z \) be a bounded linear operator. For every fixed \( \psi \in Z' \), the mapping

\[
x \mapsto \langle \psi, Lx \rangle,
\]

specifies an element in \( X' \). This operation defines the adjoint operator

\[
L' : Z' \rightarrow X', \quad \langle L' \psi, x \rangle = \langle \psi, Lx \rangle
\]

for all \( \psi \in Z' \) and all \( x \in X \); it generalizes the transposition of matrices. Throughout this section we follow [6].

Definition 6.1 For a subset \( W \subset Z' \), we define its polar set \( W^\circ \subset Z \) by

\[
W^\circ = \{ z \in Z : \langle \psi, z \rangle = 0 \text{ for all } \psi \in W \}.
\]

If \( Z \) is a Hilbert space that is identified with its dual, then \( W^\circ \) coincides with the orthogonal complement \( W^\perp \) of \( W \). The closed range theorem generalizes the identity

\[
\text{Im} L = (\ker L^\top)^\perp,
\]

from finite-dimensional to infinite-dimensional situations. Its proof uses the Hahn–Banach theorem, i.e., the existence of a separating hyperplane for two disjoint convex sets.

Theorem 6.2 (Closed Range Theorem) Let \( L : X \rightarrow Z \) be a bounded and linear operator. Then \( \text{Im} \, L \) is closed in \( Z \) if and only if \( \text{Im} \, L = (\ker L')^\circ \).
Proof By the definition of \((\ker L')^o\), we have
\[
(\ker L')^o = \{ z \in Z : \langle \psi, z \rangle = 0 \text{ for all } \psi \in \ker L' \} = \bigcap_{\psi \in \ker L'} \ker \psi.
\]
Since for every \(\psi \in \ker L'\) and \(x \in X\) we have
\[
\langle \psi, Lx \rangle = \langle L'\psi, x \rangle = 0,
\]
we see that \(\text{Im } L \subseteq (\ker L')^o\).

(i) Assume that there exists \(z_0 \in (\ker L')^o \setminus \text{Im } L\). Since \(\text{Im } L\) is closed, there exists \(\varepsilon > 0\) such that \(B_\varepsilon(z_0) \cap \text{Im } L = \emptyset\) Applying the separation theorem, cf. Remarks 2.5, we deduce that there exist \(\psi \in Z'\) and \(m \in \mathbb{R}\) such that
\[
\langle \psi, z_0 \rangle \geq m \geq \langle \psi, Lx \rangle
\]
for all \(x \in X\). The second inequality can only be true if \(m \geq 0\) and \(\langle \psi, Lx \rangle = 0\) for all \(x \in X\). But this implies that \(L'\psi = 0\), i.e., \(\psi \in \ker L'\). Since \(z_0 \in (\ker L')^o\), we obtain the contradiction \(\langle \psi, z_0 \rangle = 0\).

(ii) The kernel of every \(\psi \in Z'\) is a closed set, and as the intersection of closed sets, \((\ker L')^o\) is closed. Hence, also \(\text{Im } L\) is closed. \(\square\)

Remark 6.5 Continuity of \(L\) alone is not sufficient to guarantee that \(\text{Im } L\) is closed.

6.2.2 Inf-Sup Condition

The inf-sup condition characterizes the boundedness of left inverse operators as in the discrete situation. We let \(X, Y, Z\) be Banach spaces in what follows.

Definition 6.2 A bijective linear operator \(L : X \to W\) for \(W \subset Z\) is an isomorphism if it is bounded and its inverse \(L^{-1} : W \to X\) is bounded.

The existence and boundedness of a left inverse of a bounded operator can be expressed by an inf-sup condition.

Lemma 6.3 (Inf-Sup Condition) Let \(L : X \to Y'\) be a bounded and linear operator. Then \(L : X \to \text{Im } L\) is an isomorphism if and only if there exists \(\beta > 0\) such that
\[
\inf_{x \in X \setminus \{0\}} \sup_{y \in Y \setminus \{0\}} \frac{\langle Lx, y \rangle}{\|x\|_X \|y\|_Y} \geq \beta.
\]
Proof

(i) If \( L : X \to \text{Im} L \) is an isomorphism, then for every \( x \in X \), we have the estimate \( \|x\|_X \leq c_L^{-1}\|Lx\|_{Y'} \) which is equivalent to

\[
c_L\|x\|_X \leq \sup_{y \in Y \setminus \{0\}} \frac{\langle Lx, y \rangle}{\|y\|_{Y'}},
\]

i.e., to the asserted estimate with \( \beta = c_L \).

(ii) Conversely, the inf-sup condition is equivalent to \( \|Lx\|_{Y'} \geq \beta\|x\|_X \) for all \( x \in X \), which implies that \( L \) is injective. Therefore, \( L : X \to \text{Im} L \) is a bijection. In particular, for every \( \psi \in \text{Im} L \), there exists a unique \( x \in X \) with \( Lx = \psi \) and

\[
\|x\|_X \leq \frac{1}{\beta}\|Lx\|_{Y'} = \frac{1}{\beta}\|\psi\|_{Y'},
\]

i.e., the left inverse \( L^{-\ell} : \text{Im} L \to X \) is bounded. \( \Box \)

Remarks 6.6

(i) The inf-sup condition is equivalent to \( \|Lx\|_{Y'} = \sup_{y \in Y \setminus \{0\}} \frac{\langle Lx, y \rangle}{\|y\|_{Y'}} \geq \beta\|x\|_X \) for some \( \beta > 0 \) and all \( x \in X \). Injectivity and closedness of the image of \( L \) are direct consequences of this bound.

(ii) Equivalent to the statements of the lemma is that \( L : X \to Y' \) is injective with bounded left inverse \( L^{-\ell} \) with \( \|L^{-\ell}\| \leq \beta^{-1} \).

If \( L \) is an isomorphism, then its image is closed. The converse implication is known as the inverse operator theorem.

Lemma 6.4 (Closedness) If \( L : X \to W \) for \( W \subset Z \) is an isomorphism, then \( W = \text{Im} L \) is closed.

Proof We have to show that for every Cauchy sequence \( (v_j)_{j \in \mathbb{N}} \subset \text{Im} L \) with \( v_j = Lx_j \) for a sequence \( (x_j)_{j \in \mathbb{N}} \subset X \), its limit \( v \) also belongs to \( \text{Im} L \). Since the left inverse is bounded, we have for all \( j, k \in \mathbb{N} \) that

\[
\|x_j - x_k\|_X \leq c_L^{-1}\|v_j - v_k\|_Z,
\]

i.e., \( (x_j)_{j \in \mathbb{N}} \) is a Cauchy sequence with limit \( x \in X \). By the continuity of \( L \), we have \( Lx = \lim_{j \to \infty} Lx_j = v \) which proves \( v \in \text{Im} L \). \( \Box \)

Remark 6.7 As a consequence of the previous lemma, we have \( \text{Im} L = (\ker L')^\circ \) whenever a bounded linear operator \( L \) satisfies the inf-sup condition.
6.2.3 **Generalized Lax–Milgram Lemma**

For Banach spaces \( X \) and \( Y \), and a continuous bilinear form

\[
\Gamma : X \times Y \to \mathbb{R},
\]

we define the bounded and linear operator \( L : X \to Y' \) via

\[
Lx = \Gamma(x, \cdot)
\]

for all \( x \in X \). The problem of finding \( x \in X \) such that

\[
\Gamma(x, y) = \ell(y)
\]

for all \( y \in Y \) for a given functional \( \ell \in Y' \) is thus equivalent to the operator equation

\[
Lx = \ell.
\]

**Theorem 6.3 (Generalized Lax–Milgram Lemma)** Assume that \( X \) and \( Y \) are reflexive Banach spaces. The linear operator \( L : X \to Y' \) is an isomorphism if and only if the associated bilinear form

\[
\Gamma(x, y) = \langle Lx, y \rangle
\]

is bounded, satisfies an inf-sup condition, and is nondegenerate, i.e.,

(a) there exists \( c_\Gamma \geq 0 \) such that

\[
|\Gamma(x, y)| \leq c_\Gamma \|x\|_X \|y\|_Y;
\]

(b) there exists \( \gamma > 0 \) such that

\[
\inf_{x \in X \setminus \{0\}} \sup_{y \in Y \setminus \{0\}} \frac{\Gamma(x, y)}{\|x\|_X \|y\|_Y} \geq \gamma;
\]

(c) for all \( y \in Y \setminus \{0\} \) there exists \( x \in X \) with \( \Gamma(x, y) \neq 0 \).

**Proof** We note that boundedness of \( L \) is equivalent to boundedness of \( \Gamma \).

(i) Assume that the conditions on \( \Gamma \) are satisfied. Lemma 6.3 then implies that \( L : X \to \text{Im} L \) is an isomorphism, and Lemma 6.4 shows that \( \text{Im} L \) is closed. The closed range theorem yields that

\[
\text{Im} L = (\ker L')^\circ.
\]
Note that $L' : Y'' \rightarrow X'$ is by identifying $Y'' \simeq Y$ regarded as an operator on $Y$, i.e., $L' : Y \rightarrow X'$, and with $X'' \simeq X$ we have $(L')' \simeq L$. The nondegeneracy condition implies that

$$\ker L' = \{0\},$$

since $L'y = 0$ means that $\Gamma(x, y) = \langle Lx, y \rangle = \langle x, L'y \rangle = 0$ for all $x \in X$. In particular, $(\ker L')^o = Y'$ and $\text{Im} L = Y'$. Hence $L : X \rightarrow Y'$ is an isomorphism.

(ii) Conversely, if $L$ is an isomorphism, then Lemma 6.3 implies the inf-sup condition. Since $\text{Im} L = Y'$ is closed, we have $Y' = \text{Im} L = (\ker L')^o$, which can only be the case if $\ker L' = \{0\}$. In particular, there is no $y \in Y \setminus \{0\}$ with $L'y = 0$, and hence no $y \in Y$, such that for all $x \in X$ we have

$$0 = \langle x, L'y \rangle = \langle Lx, y \rangle = \Gamma(x, y).$$

Therefore the nondegeneracy condition is also satisfied.

Remarks 6.8

(i) If we omit the nondegeneracy condition, then the mapping $L : X \rightarrow \text{Im} L = (\ker L')^o \subseteq Y'$ is an isomorphism.

(ii) If $X = Y$ and $\Gamma$ is bounded and coercive, then the conditions of the theorem are satisfied.

(iii) The inf-sup condition implies that we have $\|L^{-1}\| \leq \gamma^{-1}$.

(iv) If $\Gamma$ is symmetric, i.e., $\Gamma(x, y) = \Gamma(y, x)$ for all $x, y \in X$, then the inf-sup condition implies nondegeneracy.

6.2.4 Saddle-Point Problems

We return to the continuous saddle-point formulation which consists in finding $(u, p) \in V \times Q$ such that

$$\begin{cases}
a(u, v) + b(v, p) = \ell_1(v), \\
b(u, q) = \ell_2(q),
\end{cases}$$

for all $(v, q) \in V \times Q$. To formulate necessary and sufficient conditions for the unique solvability of $(S)$, we analyze properties of the bilinear form $b$.

Lemma 6.5 (Properties of $b$) Let $V$ and $Q$ be Hilbert spaces and let $b : V \times Q \rightarrow \mathbb{R}$ be bounded and bilinear. Let

$$B : V \rightarrow Q', \ v \mapsto b(v, \cdot).$$
Then the following statements are equivalent:

(i) The operator \( B' : Q \to V' \) satisfies an inf-sup condition, i.e., there exists \( \beta > 0 \) such that

\[
\inf_{q \in Q \setminus \{0\}} \sup_{v \in V \setminus \{0\}} \frac{b(v, q)}{\|v\|_V \|q\|_Q} \geq \beta
\]

(ii) The operator \( B : (\ker B)\perp \to Q' \) is an isomorphism and there exists \( \beta > 0 \), such that for all \( v \in (\ker B)\perp \) we have

\[
\|Bv\|_{Q'} \geq \beta \|v\|_V.
\]

(iii) The operator \( B' : Q \to (\ker B)\perp \subseteq V' \) is an isomorphism and there exists \( \beta > 0 \), such that for all \( q \in Q \) we have

\[
\|B'q\|_{V'} \geq \beta \|q\|_Q.
\]

Proof (i) \( \iff \) (iii). The equivalence is the statement of the generalized Lax–Milgram lemma with \( X = Q, Y = V, L = B', L' = B, \Gamma'(q, v) = b(v, q) \), when the nondegeneracy condition is omitted, cf. Remark 6.8.

(iii) \( \implies \) (ii). We show that \( B : (\ker B)^\perp \to Q' \) satisfies the conditions of the generalized Lax–Milgram lemma with \( L = B, X = (\ker B)^\perp, \) and \( Y = Q \). To prove the inf-sup condition, let \( v \in (\ker B)^\perp \subset V \). With \( v \) we associate the functional \( g_v \in (\ker B)^\perp \subset V' \) by defining \( \langle g_v, w \rangle = (v, w)_V \) for all \( w \in V \). Then, \( \|g_v\|_{V'} = \|v\|_V \). Due to the statement (iii) there exists \( p \in Q \), such that \( B'p = g_v \) and \( \|p\|_Q \leq \beta^{-1}\|g_v\|_{V'} \). Using \( b(v, p) = \langle B'p, v \rangle = \langle g_v, v \rangle = (v, v)_V \) implies that

\[
\sup_{q \in Q \setminus \{0\}} \frac{b(v, q)}{\|q\|_Q} \geq \frac{b(v, p)}{\|p\|_Q} \geq \frac{b(v, p)}{\beta^{-1}\|g_v\|_{V'}} = \beta \|v\|_V,
\]

which is the inf-sup condition for \( B \) restricted to \( (\ker B)^\perp \). To verify the nondegeneracy condition, assume that there exists \( q \in Q \setminus \{0\} \), such that \( b(v, q) = 0 \) for all \( v \in (\ker B)^\perp \), which then holds for all \( v \in V \). This implies \( B'q = 0 \) which contradicts \( \|B'q\|_{V'} > 0 \). Hence, the conditions of the generalized Lax–Milgram lemma are satisfied and \( B \) is an isomorphism. The estimate is equivalent to the inf-sup condition.
(ii) \implies (i) For every $q \in Q$, using the bijectivity of $B$ and the estimate $\|Bv\|_Q' \geq \beta \|v\|_V$ for all $v \in (\ker B)^\perp$, we have that

$$\|q\|_Q = \sup_{g \in Q \setminus \{0\}} \frac{\langle g, q \rangle}{\|g\|_{Q'}} = \sup_{v \in (\ker B)^\perp \setminus \{0\}} \frac{\langle Bv, q \rangle}{\|Bv\|_{Q'}} = \sup_{v \in (\ker B)^\perp \setminus \{0\}} \frac{b(v, q)}{\|Bv\|_{Q'}} \leq \sup_{v \in (\ker B)^\perp \setminus \{0\}} \frac{b(v, q)}{\|Bv\|_{Q'}} \leq \beta^{-1} \sup_{v \in V \setminus \{0\}} \frac{b(v, q)}{\|v\|_V},$$

which is (i).

\[\square\]

**Remark 6.9** The second statement of the lemma is equivalent to the inf-sup condition

$$\inf_{v \in (\ker B)^\perp \setminus \{0\}} \sup_{q \in Q \setminus \{0\}} \frac{b(v, q)}{\|v\|_V \|q\|_Q} \geq \beta > 0,$$

in which the roles of $v$ and $q$ are exchanged and $V$ is replaced by $(\ker B)^\perp$. This inf-sup condition corresponds to the operator $Bv = b(v, \cdot)$.

With the help of the lemma we can specify conditions for the unique solvability of the saddle-point problem $(S)$.

**Theorem 6.4 (Brezzi’s Splitting Theorem)** Assume that $V$ and $Q$ are Hilbert spaces, $a : V \times V \to \mathbb{R}$ is a symmetric, bounded, and positive semidefinite bilinear form, and $b : V \times Q \to \mathbb{R}$ is a bounded and bilinear form. The operator

$$L : V \times Q \to V' \times Q', \quad (u, p) \mapsto (a(u, \cdot) + b(\cdot, p), b(u, \cdot))$$

is an isomorphism, i.e., $(S)$ is uniquely solvable, if and only if $a$ is coercive on $\ker B$, i.e., there exists $\alpha > 0$ such that

$$a(v, v) \geq \alpha \|v\|^2_V$$

for all $v \in \ker B$, and $b$ satisfies an inf-sup condition, i.e., there exists $\beta > 0$ such that

$$\inf_{q \in Q \setminus \{0\}} \sup_{v \in V \setminus \{0\}} \frac{b(v, q)}{\|v\|_V \|q\|_Q} \geq \beta.$$
Proof

(i) Assume that the conditions on $a$ and $b$ are satisfied, in particular, the equivalent statements of Lemma 6.5 are valid. Given $(\ell_1, \ell_2) \in V' \times Q'$, we first let $u_0 \in (\ker B)\perp$ be such that $Bu_0 = \ell_2$ and

$$\|u_0\|_V \leq \beta^{-1} \|\ell_2\|_{Q'}.$$  

We thus have that

$$b(u_0, q) = \ell_2(q)$$

for all $q \in Q$. We next try to find $u_1 \in \ker B$, such that for some $p \in Q$ we have

$$a(u_0 + u_1, v) + b(v, p) = \ell_1(v)$$

for all $v \in V$. By restricting to $v \in \ker B$, we have $b(v, p) = 0$, and the identity simplifies to the requirement that $u_1 \in \ker B$ be such that

$$a(u_1, v) = \ell_1(v) - a(u_0, v)$$

for all $v \in \ker B$. By the assumed coercivity of $a$ on $\ker B$, the Lax–Milgram lemma implies the existence of a unique solution $u_1 \in \ker B$ with

$$\alpha \|u_1\|^2_V \leq a(u_1, u_1) \leq \|\ell_1\|_{V'} \|u_1\|_V + k_a \|u_0\|_V \|u_1\|_V.$$  

It remains to determine $p \in Q$ such that

$$b(v, p) = \ell_1(v) - a(u_0 + u_1, v)$$

for all $v \in V$. The right-hand side defines a functional $\varphi \in V'$ and by the construction of $u_1$, we have $\varphi(v) = 0$ for all $v \in \ker B$. Hence $\varphi \in (\ker B)^\circ$. Since $B' : Q \to (\ker B)^\circ$ is an isomorphism, there exists a unique $p \in Q$ with $B'p = \varphi$, which is equivalent to the above equation. Moreover, we have

$$\|p\|_Q \leq \beta^{-1} \|\varphi\|_{V'}.$$  

The pair $(u, p)$ with $u = u_0 + u_1$ satisfies $(S)$, and by construction, we have

$$\|(u, p)\|_{V \times Q} \leq c(k_a, \alpha^{-1}, \beta^{-1}) \|(\ell_1, \ell_2)\|_{V' \times Q'}.$$  

To prove uniqueness of $(u, p)$, we note that for $(\ell_1, \ell_2) = 0$, the estimate implies $(u, p) = 0$. Together with the linearity of $L$, we conclude that $L$ is an isomorphism.
(ii) Assume that $L$ is an isomorphism and let $u \in \ker B$. We define $\ell_1 \in V'$ by setting

$$\langle \ell_1, v \rangle = a(u, v)$$

for all $v \in V$. Then, there exists $p \in Q$ such that $(u, p) = L^{-1}(\ell_1, 0)$, i.e., we have

$$a(u, v) + b(v, p) = \langle \ell_1, v \rangle,$$

$$b(u, q) = 0,$$

for all $(v, q) \in V \times Q$. Using the Cauchy–Schwarz inequality for the symmetric and positive semidefinite bilinear form $a$, we find that

$$\|\ell_1\|_{V'} = \sup_{v \in V \setminus \{0\}} \frac{a(u, v)}{\|v\|_V} \leq \sup_{v \in V \setminus \{0\}} \frac{a(u, u)^{1/2}a(v, v)^{1/2}}{\|v\|_V}.$$

Since $L^{-1}$ is bounded, we have

$$\|u\|_V \leq \|(u, p)\|_{V \times Q} \leq \|L^{-1}\| \| (\ell_1, 0) \|_{V' \times Q'} = \|L^{-1}\| \|\ell_1\|_{V'}.$$

The combination of the last two inequalities and the boundedness of $a$ show that

$$\|u\|_V \leq \|L^{-1}\| \sup_{v \in \ker B \setminus \{0\}} \frac{a(u, u)^{1/2}a(v, v)^{1/2}}{\|v\|_V} \leq \|L^{-1}\| k_a^{1/2} a(u, u)^{1/2}.$$

Since this holds for all $u \in \ker B$, we find that $a$ is coercive on $\ker B$. To prove the inf-sup condition for $b$, let $\ell_2 \in Q'$ and $(u, p) = L^{-1}(0, \ell_2)$. Then $\|u\|_V \leq \|L^{-1}\| \|\ell_2\|_{Q'}$, and for the orthogonal projection $u^\perp$ of $u$ onto $(\ker B)^\perp$, we have

$$\|u^\perp\|_V \leq \|u\|_V.$$

Therefore the mapping $\ell_2 \mapsto u \mapsto u^\perp$ is bounded such that $Bu^\perp = \ell_2$, i.e., $B : (\ker B)^\perp \to Q'$ is an isomorphism and Lemma 6.5 implies the inf-sup condition.

We illustrate the application of the theorem with a saddle-point formulation of the Poisson problem.
Example 6.4 We consider the Poisson problem $-\Delta p = g$ in $\Omega \subset \mathbb{R}^d$ with Dirichlet boundary condition $p|_{\partial \Omega} = 0$. By introducing $u = \nabla p$, this is equivalent to finding $(u, p) \in L^2(\Omega; \mathbb{R}^d) \times H^1_0(\Omega)$ such that

$$\int_{\Omega} u \cdot v \, dx - \int_{\Omega} v \cdot \nabla p \, dx = 0,$$

$$- \int_{\Omega} u \cdot \nabla q \, dx = - \int_{\Omega} g q \, dx,$$

for all $(v, q) \in L^2(\Omega; \mathbb{R}^d) \times H^1_0(\Omega)$. This is called the primal mixed form of the Poisson problem. The bilinear form

$$a(u, v) = \int_{\Omega} u \cdot v \, dx$$

coincides with the inner product on $L^2(\Omega; \mathbb{R}^d)$ and is thus elliptic on $L^2(\Omega; \mathbb{R}^d)$. The bilinear form

$$b(v, q) = - \int_{\Omega} v \cdot \nabla q \, dx$$

satisfies an inf-sup condition, since by choosing $v = -\nabla q$, we have

$$\sup_{v \in L^2(\Omega; \mathbb{R}^d) \setminus \{0\}} \frac{b(v, q)}{\|v\|_{L^2(\Omega)}} \geq \frac{b(-\nabla q, q)}{\|\nabla q\|_{L^2(\Omega)}} = \|\nabla q\|_{L^2(\Omega)}$$

for all $q \in H^1_0(\Omega)$.

6.2.5 Perturbed Saddle-Point Problems

In some applications perturbed saddle-point problems occur. These consist in finding $(u, p) \in V \times Q$ such that

$$(S_t) \quad \begin{cases} a(u, v) + b(v, p) = \ell_1(v), \\ b(u, q) - t^2 c(p, q) = \ell_2(q), \end{cases}$$

for all $(v, q) \in V \times Q$. Here $c : Q \times Q \to \mathbb{R}$ is a positive semidefinite symmetric bilinear form and $t \geq 0$ is a small parameter. We follow [6].

Theorem 6.5 (Perturbed Formulation) Assume that the bilinear forms $a$ and $b$ satisfy the conditions of Brezzi’s splitting theorem, and assume that $c$ is bounded and positive semidefinite. Then $(S_t)$ has a unique solution for every $t \in [0, 1]$ and the solution operator $L^{-1} : (\ell_1, \ell_2) \to (u, p)$ is bounded $t$-independently.

The proof is based on the following lemma.
Lemma 6.6 (Inf-Sup Condition for $\Gamma_i$) Suppose that the assumptions of the theorem are satisfied, set $|q|_c = c(q, q)^{1/2}$, and assume that for some $\alpha > 0$ and all $u \in V \setminus \{0\}$, we have

$$\frac{a(u, u)}{\|u\|_V} + \sup_{q \in Q \setminus \{0\}} \frac{b(u, q)}{\|q\|_Q + t|q|_c} \geq \alpha \|u\|_V.$$ 

Then the bilinear form associated with $(S_i)$, i.e.,

$$\Gamma_i(u, p; v, q) = a(u, v) + b(v, p) + b(u, q) - t^2 c(p, q),$$

satisfies an inf-sup condition with a $t$-independent constant $\gamma' > 0$, and with the norm

$$\|(v, q)\| = \|v\|_V + \|q\|_Q + t|q|_c$$
on $V \times Q$.

Proof Let $(u, p) \in V \times Q$. We distinguish three cases with a parameter $0 < \delta \leq 1$ to be determined later but which will not depend on $(u, p)$ and $t$.

Case 1: Assume that $\|u\|_V + \|p\|_Q \leq \delta^{-1} t|p|_c$. We then have $\Gamma_i(u, p; u, -p) = a(u, u) + t^2 c(p, p) \geq (\delta^2 / 4) \|(u, p)\|^2$ and hence

$$\frac{\delta^2}{4} \|(u, p)\| \leq \frac{\Gamma_i(u, p; u, -p)}{\|(u, -p)\|} \leq \sup_{(v, q) \neq 0} \frac{\Gamma_i(u, p; v, q)}{\|(v, q)\|},$$

i.e., an inf-sup condition for $\Gamma_i$.

Case 2a: Assume that $\|u\|_V + \|p\|_Q > \delta^{-1} t|p|_c$ and $\|u\|_V \leq (\beta / (2k_a)) \|p\|_Q$. Since $b$ satisfies the inf-sup condition, we have

$$\beta \|p\|_Q \leq \sup_{v \neq 0} \frac{b(v, p)}{\|v\|_V} \leq \sup_{v \neq 0} \frac{\Gamma_i(u, p; v, 0)}{\|v\|_V} + \sup_{v \neq 0} \frac{a(u, v)}{\|v\|_V}$$

$$\leq \sup_{(v, q) \neq 0} \frac{\Gamma_i(u, p; v, q)}{\|(v, q)\|} + k_a \|u\|_V.$$ 

The assumption that $\|u\|_V \leq (\beta / (2k_a)) \|p\|_Q$ leads to

$$k_a \|u\|_V \leq \sup_{(v, q) \neq 0} \frac{\Gamma_i(u, p; v, q)}{\|(v, q)\|}.$$ 

Incorporating the assumed bound for $t|p|_c$, we deduce the inf-sup condition for $\Gamma_i$. 

Case 2b: Assume that \( \|u\|_V + \|p\|_Q > \delta^{-1}|p|_c \) and \( \|u\|_V > (\beta/(2k_a))\|p\|_Q. \) The bounds imply that \( \|u\|_V > (2 + 4k_a/\beta)\|u\|_V, \) and for \( \delta \leq 1/(2 + 4k_a/\beta) \) we have \( \delta \|u\|_V \leq \|u\|_V. \) The assumption of the lemma shows that

\[
\alpha \|u\|_V \leq \frac{a(u, u)}{\|u\|_V} + \frac{\|b(u, q)\|}{\|u\|_V} + \frac{\|\Gamma_1(u, p; 0, q)\|}{\|u\|_V} + \frac{\|\Gamma_2(c(p, p))\|}{\|u\|_V}.
\]

With \( c(p, q) \leq |p|_c|q|_c, \) we deduce that

\[
\alpha \|u\|_V \leq \frac{a(u, u) + r^2c(p, p)}{\delta \|u\|_V} + \frac{\Gamma_1(u, p; 0, q) + r^2c(p, p)}{\|u\|_V} + \frac{\|\Gamma_1(u, p; 0, q)\|}{\|u\|_V} + \frac{\|\Gamma_1(u, p; 0, q)\|}{\|u\|_V}.
\]

For \( \delta \) small enough we may assume that \( |p|_c \leq \delta\|u\|_V + \|p\|_Q \leq (\alpha/2)\|u\|_V. \) Incorporating the bound \( \delta \|u\|_V \leq \|u\|_V \) then leads to

\[
\frac{\alpha \delta}{2} \|u\|_V \leq \frac{a(u, u) + r^2c(p, p)}{\delta \|u\|_V} + \frac{\Gamma_1(u, p; 0, q) + r^2c(p, p)}{\|u\|_V} + \frac{\|\Gamma_1(u, p; 0, q)\|}{\|u\|_V} + \frac{\|\Gamma_1(u, p; 0, q)\|}{\|u\|_V}.
\]

This completes the proof of the inf-sup condition for \( \Gamma_1. \) 

We are now in position to prove Theorem 6.5.

**Proof (of Theorem 6.5)** The requirement that the unperturbed saddle-point problem with \( t = 0 \) satisfy the conditions of Brezzi’s splitting theorem implies that the bilinear form \( \Gamma_0 \) satisfies an inf-sup condition with a constant \( \gamma > 0. \) For the pair \( (u, 0) \in V \times Q, \) we therefore have that

\[
\gamma \|u\|_V \leq \frac{\Gamma_0(u, 0; v, q)}{\|v\|_V + \|q\|_Q} \leq \frac{a(u, v)}{\|v\|_V} + \frac{b(u, q)}{\|q\|_Q}.
\]

Using \( \|q\|_Q + t|q|_c \leq (1 + k_c^{1/2})\|q\|_Q \) and the Cauchy–Schwarz inequality for \( a, \) we deduce that

\[
\gamma \|u\|_V \leq k_a^{1/2}a(u, u)^{1/2} + (1 + k_c^{1/2})\sup_{q \neq 0} \frac{b(u, q)}{\|q\|_Q + t|q|_c}.
\]

We use that for nonnegative real numbers \( x, y, z \) with \( x > 0 \) and \( x \leq y + z \) it follows that \( x \leq y^2/x + 2z \) (which follows from a case distinction \( y \leq x \) and \( y > x \) with a
binomial formula) to estimate
\[ \gamma \|u\|_V \leq k_u a(u, u) + 2(1 + k_u^{1/2}) \sup_{q \neq 0} \frac{b(u, q)}{\|q\|_Q + r|q|}. \]

This is the condition of Lemma 6.6 which therefore implies an inf-sup condition for \( \Gamma \) with a positive constant that does not depend on \( t \). Noting that symmetry and the inf-sup condition imply nondegeneracy and verifying continuity, the result follows from the generalized Lax–Milgram lemma.

Remark 6.10  The theorem motivates treating the optimality conditions related to minimizing
\[ u \mapsto \frac{1}{2} \int_{\Omega} |\nabla u|^2 \, dx + \frac{t^{-2}}{2} \int_{\Omega} |\text{div} u|^2 \, dx - \int_{\Omega} f \cdot u \, dx \]
in the set of \( u \in H^1_0(\Omega; \mathbb{R}^d) \) for \( 0 < t \ll 1 \) as a perturbed saddle-point problem. By introducing the variable \( p = t^{-2} \text{div} u \), we find that every minimizer \( u \in H^1_0(\Omega; \mathbb{R}^d) \) satisfies
\[ \int_{\Omega} \nabla u : \nabla v \, dx + \int_{\Omega} p \text{div} v \, dx = \int_{\Omega} f \cdot v \, dx, \]
\[ \int_{\Omega} q \text{div} u \, dx - t^2 \int_{\Omega} pq \, dx = 0, \]
for all \((v, q) \in H^1_0(\Omega; \mathbb{R}^d) \times L^2(\Omega)\). This approach avoids a critical dependence of stability bounds on the parameter \( t \).

6.3 Approximation of Saddle-Point Problems

6.3.1 Generalized Céa Lemma

To approximate a general variational formulation defined by the bilinear form \( \Gamma : X \times Y \to \mathbb{R} \) numerically, we choose finite-dimensional subspaces \( X_h \subset X \) and \( Y_h \subset Y \). The generalized Lax–Milgram lemma guarantees the existence of a unique solution \( x_h \in X_h \) such that
\[ \Gamma(x_h, y_h) = \ell(y_h) \]
for all \( y_h \in Y_h \), provided that \( \Gamma \) is bounded, and satisfies an inf-sup condition, and is nondegenerate with respect to the spaces \( X_h \) and \( Y_h \). In this case we have the following generalization of Céa’s lemma.
Theorem 6.6 (Generalized Céa Lemma) Assume that $\Gamma : X \times Y \to \mathbb{R}$ is a bounded bilinear form, such that for every $\ell \in Y'$, there exists a unique $x \in X$ with

$$\Gamma(x, y) = \ell(y)$$

for all $y \in Y$, and such that $\|x\|_X \leq \gamma^{-1}\|\ell\|_{Y'}$. Assume that $(X_h)_{h>0}$ and $(Y_h)_{h>0}$ are families of subspaces of $X$ and $Y$, respectively, with the property that for every $h > 0$, there exists $\gamma_h > 0$ with

$$\inf_{x_h \in X_h \setminus \{0\}} \sup_{y_h \in Y_h \setminus \{0\}} \frac{\Gamma(x_h, y_h)}{\|x_h\|_X \|y_h\|_Y} \geq \gamma_h,$$

and for every $y_h \in Y \setminus \{0\}$, there exists $x_h \in X_h$ so that $\Gamma(x_h, y_h) \neq 0$. Suppose that there exists $\gamma > 0$ such that $\gamma_h \geq \gamma$ for all $h > 0$. Then for every $\ell \in Y'$, there exists a unique $x_h \in X_h$ with

$$\Gamma(x_h, y_h) = \ell(y_h)$$

for all $y_h \in Y_h$, and such that

$$\|x - x_h\|_X \leq (1 + k\Gamma / \gamma) \inf_{w_h \in X_h} \|x - w_h\|_X.$$

Proof The generalized Lax–Milgram lemma implies the existence of a uniquely defined solution $x_h \in X_h$. By subtracting the variational formulations, we obtain the Galerkin orthogonality

$$\Gamma(x - x_h, y_h) = 0$$

for all $y_h \in Y_h$. Let $w_h \in X_h$ be arbitrary. The discrete inf-sup condition implies that

$$\gamma_h \|x_h - w_h\|_X \leq \sup_{y_h \in Y_h \setminus \{0\}} \frac{\Gamma(x_h - w_h, y_h)}{\|y_h\|_Y}.$$

Using Galerkin orthogonality and boundedness of $\Gamma$, we infer that

$$\gamma_h \|x_h - w_h\|_X \leq \sup_{y_h \in Y_h \setminus \{0\}} \frac{\Gamma(x_h - w_h, y_h)}{\|y_h\|_Y} \leq \sup_{y_h \in Y_h \setminus \{0\}} \frac{k\Gamma \|x - w_h\|_X \|y_h\|_Y}{\|y_h\|_Y} \leq k\Gamma \|x - w_h\|_X.$$
6.3 Approximation of Saddle-Point Problems

With the triangle inequality, we deduce that

\[ \| x - x_h \|_X \leq \| x - w_h \|_X + \| w_h - x_h \|_X \]

Using that \( w_h \in X_h \) is arbitrary and that \( \gamma_h \geq \gamma \) for all \( h > 0 \) proves the estimate. \( \square \)

Remarks 6.11

(i) The constant in the estimate of the theorem can be improved. For every \( v \in X \) there exists a unique \( v_h = P_h v \) with \( \Gamma(v, y_h) = \Gamma(v, y) \) for all \( y_h \in Y_h \), and we have \( \| P_h \| \leq k \gamma / \gamma_h \). The operator \( P_h \) defines a projection in the sense that \( P_h^2 = P_h \) which leads to the nontrivial identity \( \| P_h \| = \| I - P_h \| \). With this we deduce that

\[ \| x - x_h \|_X = \| (I - P_h)(x - x_h) \|_X = \| (I - P_h)(x - w_h) \|_X \leq \| P_h \| \| x - w_h \|_X. \]

(ii) In contrast to boundedness, inf-sup condition and nondegeneracy are in general not inherited by subspaces.

6.3.2 Saddle-Point Problems

As in the continuous setting, it is desirable to formulate conditions on the bilinear forms involved in a saddle-point problem to determine their solvability. We consider again the problem of finding \( (u, p) \in V \times Q \) with

\[
\begin{align*}
a(u, v) + b(v, p) &= \ell_1(v), \\
b(u, q) &= \ell_2(q),
\end{align*}
\]

for all \( (v, q) \in V \times Q \). Here, \( \ell_1 \in V' \) and \( \ell_2 \in Q' \) are given functionals.

Theorem 6.7 (Babuška–Brezzi Conditions) Assume that the conditions of Brezzi’s splitting theorem are satisfied, and \( (V_h)_{h>0} \) and \( (Q_h)_{h>0} \) are families of subspaces of the Hilbert spaces \( V \) and \( Q \), respectively, such that for all \( h > 0 \) we have:

1. the bilinear form \( a \) is elliptic on the discrete kernel of \( b \), i.e., there exists \( \alpha_h > 0 \) such that for all \( v_h \in K_h = \{ v_h \in V_h : b(v_h, q_h) = 0 \text{ for all } q_h \in Q_h \} \), we have

\[ a(v_h, v_h) \geq \alpha_h \| v_h \|_{V_h}^2. \]
(2) the bilinear form $b$ satisfies the inf-sup condition with respect to $V_h$ and $Q_h$, i.e., there exists $\beta_h > 0$ such that

$$\inf_{q_h \in Q_h} \sup_{v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_V \|q_h\|_Q} \geq \beta_h.$$ 

Suppose that there exist $\alpha, \beta > 0$ such that $\alpha_h \geq \alpha$ and $\beta_h \geq \beta$ for all $h > 0$. Then for every $h > 0$, there exists a pair $(u_h, p_h) \in V_h \times Q_h$ such that

$$a(u_h, v_h) + b(v_h, p_h) = \ell_1(v_h),$$

$$b(u_h, q_h) = \ell_2(q_h),$$

for all $(v_h, q_h) \in V_h \times Q_h$. Moreover, there exists $c > 0$ such that

$$\|u - u_h\|_V + \|p - p_h\|_Q \leq c \inf_{(v_h, q_h)} (\|u - v_h\|_V + \|p - q_h\|_Q).$$

Proof The conditions on the subspaces $V_h$ and $Q_h$ imply that the discrete problems for every $h > 0$ are uniquely solvable, i.e., the induced bilinear form $I'$ on $X_h = Y_h = V_h \times Q_h$ satisfies an inf-sup condition uniformly in $h$ and is nondegenerate for all $h > 0$. The application of the generalized Céa lemma implies the estimate. 

The nondegeneracy and the inf-sup condition are in general not inherited by the subspaces. Also the discrete coercivity has to be verified for every $h > 0$, since in general we have

$$K_h \not\subset \ker B = \{v \in V : b(v, q) = 0 \text{ for all } q \in Q\},$$

i.e., significantly more conditions are involved in the definition of $\ker B$ than in the definition of $K_h$. In fact, a sharper error estimate can be proved if $K_h \subset \ker B$ for all $h > 0$.

Remark 6.12 The spaces $V_h$ and $Q_h$ have to be carefully chosen so that the operator

$$B_h : K_h^\perp \to Q_h' \simeq Q_h, \quad v_h \mapsto b(v_h, \cdot)$$

is an isomorphism, i.e., $\dim V_h - \dim K_h = \dim Q_h$ is a necessary requirement. Moreover, $K_h$ has to be sufficiently small so that $a$ is coercive on $K_h$.

We verify the conditions of Theorem 6.7 in a simple model problem.

Example 6.5 Consider the primal mixed formulation of the Poisson problem, and for a sequence of regular triangulations $(\mathcal{T}_h)_{h > 0}$, the subspaces

$$V_h = \mathcal{L}^0(\mathcal{T}_h)^d = \{v_h \in L^\infty(\Omega; \mathbb{R}^d) : v_h|_T \in \mathcal{P}_0(T) \text{ for all } T \in \mathcal{T}_h\},$$

$$Q_h = \mathcal{L}^0(\mathcal{T}_h) = \{q_h \in C(\overline{\Omega}) : q_h|_T \in \mathcal{P}_1(T) \text{ for all } T \in \mathcal{T}_h\}. $$
The bilinear form

\[ a(u, v) = \int_{\Omega} u \cdot v \, dx \]

is coercive on \( V = L^2(\Omega; \mathbb{R}^d) \) and hence uniformly on the subspaces \( V_h \). For the bilinear form \( b : V \times Q \to \mathbb{R} \) with \( Q = H^1_0(\Omega) \), we find with the choice \( v_h = -\nabla q_h \) that

\[
\sup_{v_h \in V_h \setminus \{0\}} \frac{b(v_h, q_h)}{\|v_h\|_{L^2(\Omega)}} = \sup_{v_h \in V_h \setminus \{0\}} \frac{-\int_{\Omega} v_h \cdot \nabla q_h \, dx}{\|v_h\|_{L^2(\Omega)}} \geq \|\nabla q_h\|_{L^2(\Omega)},
\]

so that the discrete inf-sup condition for \( b \) holds uniformly in \( h \).

The error estimate resulting from Theorem 6.7 in fact coincides up to the involved constants with the estimate one obtains with Céa’s lemma for the direct treatment of the Poisson equation. In the case of the Stokes problem, the subspaces have to be chosen more carefully.

**Example 6.6** The Stokes problem leads to the bilinear form \( b : V \times Q \to \mathbb{R} \) defined by

\[ b(v, q) = \int_{\Omega} q \, \text{div} \, v \, dx \]

for \( v \in V = H^1_0(\Omega; \mathbb{R}^d) \) and \( q \in L^2(\Omega) = \{ q \in L^2(\Omega) : \int_{\Omega} q \, dx = 0 \} \). The choice of

\[ V_h = \mathcal{S}_0^1(\mathcal{T}_h)^d, \quad Q_h = \mathcal{L}^0(\mathcal{T}_h) \cap L^2(\Omega) \]

in general does not lead to the inf-sup condition for \( b \). For a triangulation \( \mathcal{T}_h \) of \( \Omega = (0, 1)^2 \) consisting of halved squares with two triangles \( T_1, T_2 \in \mathcal{T}_h \) that have all vertices on \( \partial \Omega \), cf. Fig. 6.2, we have \( \text{div} \, v_h|_{T_j} = 0 \) for \( j = 1, 2 \) and every \( v_h \in V_h \). Hence \( \text{div} : V_h \to Q_h \) is not surjective. More generally, for a triangulation \( \mathcal{T}_h \) of \( \Omega = (0, 1)^2 \) with nodes \( \mathcal{N}_h \) and consisting of halved squares, we have

\[ \dim Q_h = |\mathcal{T}_h| - 1, \quad \dim V_h = 2|\mathcal{N}_h \cap \Omega| \]

**Fig. 6.2** Triangles \( T_1, T_2 \) in a triangulation for which all vertices belong to \( \partial \Omega \) (left); triangulations consisting of halved squares with \( |\mathcal{T}_h| = 2|\mathcal{N}_h \cap \Omega| + |\mathcal{N}_h \cap \partial \Omega| - 2 \) (left and right)
6.3.3 Fortin Interpolation

In general, it is difficult to verify the Babuška–Brezzi conditions. The following lemma provides a useful equivalent characterization of the discrete inf-sup condition.

**Lemma 6.7 (Fortin Criterion)** Suppose that $V, Q$ are Hilbert spaces and there exists $\beta > 0$ such that the bounded bilinear form $b : V \times Q \to \mathbb{R}$ satisfies

$$\sup_{v \in V \setminus \{0\}} \frac{b(v, q)}{\|v\|_V} \geq \beta \|q\|_Q$$

for all $q \in Q$. Let $V_h$ and $Q_h$ be subspaces of $V$ and $Q$, respectively, such that there exist $v \in V$ and $q_h \in Q_h$ with $b(v, q_h) \neq 0$. Then the discrete inf-sup condition

$$\sup_{v_h \in V_h \setminus \{0\}} \frac{b(v_h, q_h)}{\|v_h\|_V} \geq c'_F \beta \|q_h\|_Q$$

is satisfied with an $h$-independent constant $c'_F > 0$ if and only if there exists a uniformly bounded, linear, and nontrivial operator $I_F : V \to V_h$ called a Fortin interpolant, such that for all $v \in V$, we have

$$b(v - I_F v, q_h) = 0$$

for all $q_h \in Q_h$.

**Proof**

(i) Assume that a bounded linear operator $I_F : V \to V_h$ with the specified properties exists, and let $q_h \in Q_h$. The continuous inf-sup condition, the property $b(v - I_F v, q) = 0$, and the boundedness $\|I_F v\|_V \leq c_F \|v\|_V$ imply
that
\[
\beta \|q_h\|_Q \leq \sup_{v \in V \setminus \{0\}} \frac{b(v, q_h)}{\|v\|_V} = \sup_{v \in V \setminus \{0\}} \frac{b(I_F v, q_h)}{\|v\|_V} \leq \sup_{v \in V \setminus \{0\}} \frac{b(I_F v, q_h)}{c_F^{-1} \|I_F v\|_V} = c_F \sup_{v_h \in \text{Im} I_F \setminus \{0\}} \frac{b(v_h, q_h)}{\|v_h\|_V},
\]
where \(c_F\) denotes the operator norm of \(I_F\). This proves the discrete inf-sup condition.

(ii) Assume that the discrete inf-sup condition is satisfied, and let \(v \in V\). With the inner product \((\cdot, \cdot)_V\) of \(V\), we consider the discrete saddle point problem that consists in finding \((v_h, p_h) \in V_h \times Q_h\) with

\[
(v_h, w_h)_V + b(w_h, p_h) = 0, \\
b(v_h, q_h) = b(v, q_h),
\]
for all \((w_h, q_h) \in V_h \times Q_h\). Since the Babuška–Brezzi conditions are satisfied, there exists a unique solution \((v_h, p_h) \in V_h \times Q_h\) such that

\[
\|v_h\|_V \leq \|(v_h, p_h)\|_{V \times Q} \leq c_L \|b(v, \cdot)\|_{Q'} \leq c_L k_b \|v\|_V.
\]

Setting \(I_F v = v_h\) defines a Fortin interpolant. \(\Box\)

### 6.3.4 Locking and Softening

Saddle-point formulations are often used to avoid the locking effect of a standard numerical method. To illustrate this, we consider Hilbert spaces \(V\) and \(Q\) and the problem of finding \(u \in V\) such that

\[
a(u, v) + t^{-2} (Bu, Bv)_Q = \ell(v)
\]
for all \( v \in V \). Here, \( a : V \times V \to \mathbb{R} \) is a symmetric, continuous, and coercive bilinear form, \( B : V \to Q \) a bounded linear operator, \( \ell \in V' \) a functional, \( (\cdot , \cdot)_Q \) the inner product on \( Q \), and \( 0 < t \ll 1 \) a small parameter. We follow [6].

**Proposition 6.3 (Locking)** Assume that \((V_h)_{h>0}\) is a family of finite-dimensional subspaces of \( V \), and for every \( h > 0 \), let \( u_h \in V_h \) be the Galerkin approximation of \( u \), i.e.,

\[
a(u_h, v_h) + t^{-2}(Bu_h, Bv_h)_Q = \ell(v_h)
\]

for all \( v_h \in V_h \). Suppose that there exists \( \overline{u} \in V \) with \( B\overline{u} = 0 \), and \( \ell(\overline{u}) > 0 \). Assume that

\[
V_h \cap \ker B = \{0\},
\]

in particular, and assume that there exist \( \sigma > 0 \) and \( c_b > 0 \) such that \( \|Bu_h\|_Q \geq c_b h^\sigma \|v_h\|_V \) for all \( v_h \in V_h \). Then for all \( h > 0 \), we have

\[
\|u - u_h\|_V \geq c_1 - c_2 t^2 h^{-2\sigma}
\]

with constants \( c_1, c_2 > 0 \) that do not depend on \( t \) and \( h \).

**Proof** The Lax–Milgram lemma implies the existence of uniquely defined solutions \( u \in V \) and \( u_h \in V_h \) for every \( h > 0 \). By replacing \( \overline{u} \) by \( s\overline{u} \) with a sufficiently small number \( s > 0 \), we may assume that

\[
a(\overline{u}, \overline{u}) \leq \ell(\overline{u}).
\]

Since \( u \in V \) is the unique minimizer of the functional

\[
I_v(v) = \frac{1}{2} a(v, v) + \frac{t^{-2}}{2}(Bv, Bv)_Q - \ell(v),
\]

we have that

\[
-\ell(u) \leq I_v(u) \leq I_v(\overline{u}) = \frac{1}{2} a(\overline{u}, \overline{u}) - \ell(\overline{u}) \leq -\frac{1}{2} \ell(\overline{u}).
\]

This yields that

\[
\|\ell\|_{V'} \geq \frac{\ell(u)}{\|u\|_V} \geq \frac{\ell(\overline{u})}{2\|u\|_V},
\]

i.e., \( \|u\|_V \geq c_1 = \ell(\overline{u})/(2\|\ell\|_{V'}) \). On the other hand, we have for the discrete solution \( u_h \in V_h \) that

\[
t^{-2} c_b h^{2\sigma} \|u_h\|_V^2 \leq a(u_h, u_h) + t^{-2}(Bu_h, Bu_h)_Q = \ell(u_h) \leq \|\ell\|_{V'} \|u_h\|_V.
\]
6.3 Approximation of Saddle-Point Problems

i.e., \( \|u_h\|_V \leq c_2 t^2 h^{-2\sigma} \). The reverse triangle inequality

\[ \|u - u_h\|_V \geq \|u\|_V - \|u_h\|_V \]

implies the estimate. \( \square \)

Remark 6.13 The proposition states that unless \( h^\sigma \) is small compared to \( t \), the approximation error is large, called a locking effect of the numerical method, which occurs when the kernel of \( B \) is not sufficiently resolved.

Example 6.7 If \( T_h \) is a triangulation of \( \Omega = (0,1)^2 \) consisting of halved squares with diagonals parallel to the vector \((1,1)\), then for every \( v_h \in \mathcal{S}_0^1(T_h)^2 \) with \( \text{div} \ v_h = 0 \), we have \( v_h = 0 \). Because of the approximation properties of linear finite elements we expect \( \sigma = 1 \).

A way to avoid the locking effect is to introduce the additional variable \( p = t^{-2} Bu \) and to consider the perturbed saddle-point formulation

\[
\begin{align*}
    a(u,v) + (p,Bv)_Q &= \ell(v), \\
    (q,Bu)_Q - t^2(p,q)_Q &= 0.
\end{align*}
\]

If the Babuška–Brezzi conditions are satisfied, then this formulation can be approximated robustly. This is related to a softening effect of the saddle-point formulation.

Proposition 6.4 (Softening) Assume that the families of subspaces \((V_h)_{h>0}\) and \((Q_h)_{h>0}\) satisfy the Babuška–Brezzi conditions, and let \((u_h,p_h) \in V_h \times Q_h\) be for every \( h > 0 \) the solution of

\[
\begin{align*}
    a(u_h,v_h) + (p_h,Bv_h)_Q &= \ell(v_h), \\
    (q_h,Bu_h)_Q - t^2(p_h,q_h)_Q &= 0,
\end{align*}
\]

for all \((v_h,q_h) \in V_h \times Q_h\). Let \( \Pi_h : Q \rightarrow Q_h \) denote the orthogonal projection onto \( Q_h \), i.e., for every \( q \in Q \), the element \( \Pi_h q \in Q_h \) is defined by

\[
(\Pi_h q, r_h)_Q = (q,r_h)_Q
\]

for all \( r_h \in Q_h \). Then the function \( u_h \in V_h \) satisfies

\[
a(u_h,v_h) + t^{-2}(\Pi_h Bu_h, \Pi_h Bv_h)_Q = \ell(v_h)
\]

for all \( v_h \in V_h \).
Proof The second identity in the saddle-point formulation implies that
\[ p_h = r^{-2} \Pi_h B u_h. \]
Since for every \( v_h \in V_h \) we have
\[ (r_h, B v_h)_Q = (r_h, \Pi_h B v_h)_Q \]
for all \( r_h \in Q_h \), the choice \( r_h = r^{-2} \Pi_h B u_h \) implies the result. \( \square \)

Remark 6.14 The interpretation of the proposition is that the variational formulation involves the operator \( B_h = \Pi_h \circ B \) instead of \( B \), which increases the discrete kernel, and thereby softens the formulation and avoids a locking effect.

References

A version of the generalized Lax–Milgram lemma can be found in [12]. The formulation of the inf-sup condition and its relevance for the well-posedness of saddle-point problems is due to [3, 7]. Further important contributions to the understanding and approximation of saddle-point problems are the references [1, 2, 13]. A derivation of the inf-sup condition in a finite-dimensional setting is given in [8]. Chapters on saddle-point problems and their efficient numerical solution are contained in [4–6, 9–11].