Chapter 6
Complements: Stabilisation Methods, Eigenvalue Problems

In this chapter, we shall consider two special topics related to the approximation of saddle-point problems. The first one is about stabilised methods, which are more and more widely used in many applications where it is difficult to build approximations satisfying both the ellipticity in the kernel and the *inf-sup* properties. The second section will be devoted to an abstract presentation of eigenvalue problems for mixed problems, where an emphasis will be put on both necessary and sufficient conditions.

6.1 Augmented Formulations

6.1.1 An Abstract Framework for Stabilised Methods

Stabilisation techniques have become quite popular and new methods have been introduced along many avenues. Taking into account the enormous variety of possible applications, stabilisation techniques would require a book of their own. On the other hand, we might conceive stabilisation techniques as an arsenal of tricks to manipulate the problem and transform it into one for which the general stability theories (as the ones described in this book), can be applied. Hence, we just give the flavour of some of these tricks, and refer to the specialised literature for applications on the various particular problems. We will start with some general considerations regarding *augmented formulations* (that are the basis of the so-called "stabilisations à la Hughes-Franca"). Then, following [123], we shall describe a general framework for the study of stability issues, in which one tries to reduce the stabilising modifications at the strictly necessary minimum (whence the name "minimal stabilisations").

We have seen previously, in Sect. 1.5, that some augmented formulations cannot be written as Euler's equations of a Lagrangian but rather through an antisymmetric bilinear form. To include these formulations, among others, in our framework, we

D. Boffi et al., *Mixed Finite Element Methods and Applications*, Springer Series in Computational Mathematics 44, DOI 10.1007/978-3-642-36519-5_6,

start by introducing an abstract framework, that contains the mixed methods studied in the previous chapters as a special case.

Let therefore $\mathcal{W}$ be a Hilbert space, let $\mathcal{A}$ be in $\mathcal{L}(\mathcal{W}, \mathcal{W}')$ (the space of linear continuous operators from $\mathcal{W}$ to $\mathcal{W}'$ as defined in Sect. 4.1.4), and let F be in $\mathcal{W}'$. We consider the problem: *find* $X \in \mathcal{W}$ *such that*,

$$\mathcal{A}X = F, \tag{6.1.1}$$

which in variational formulation can be written as

$$\langle \mathcal{A}X, Y\rangle_{\mathcal{W}'\times\mathcal{W}} = \langle F, Y\rangle_{\mathcal{W}'\times\mathcal{W}} \qquad \forall\, Y \in \mathcal{W}. \tag{6.1.2}$$

From now on, we shall always assume that the bilinear form associated to $\mathcal{A}$ is positive semi-definite, that is

$$\langle \mathcal{A}Y, Y\rangle \geq 0 \qquad \forall\, Y \in \mathcal{W}. \tag{6.1.3}$$

Remark 6.1.1. As we are mostly interested in **mixed problems**, it is worth showing that this abstract formalism contains the usual theory for these problems. Indeed, let $\mathcal{W} := V \times Q$, with $X := (u, p)$, and $Y := (v, q)$, and define

$$\begin{cases} \langle \mathcal{A}X, Y\rangle := a(u, v) + b(v, p) - b(u, q), \\ \langle F, Y\rangle := \langle f, v\rangle_{V'\times V} - \langle g, q\rangle_{Q'\times Q}. \end{cases} \tag{6.1.4}$$

In this context, it is clear that (6.1.2) is just another way of writing

$$\begin{cases} a(u, v) + b(v, p) = (f, v)_V & \forall\, v \in V, \\ b(u, q) = (g, q) & \forall\, q \in Q. \end{cases} \tag{6.1.5}$$

It must however be noted that we are implicitly using the non symmetric form

$$\begin{pmatrix} A & B^t \\ -B & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ -g \end{pmatrix} \tag{6.1.6}$$

rather than the symmetric one

$$\begin{pmatrix} A & B^t \\ B & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}. \tag{6.1.7}$$

As a consequence of this choice, assuming

$$a(v, v) \geq 0 \qquad \forall\, v \in V, \tag{6.1.8}$$

we clearly have that (6.1.3) holds. □

6.1.2 Stabilising Terms

We shall consider here a very wide class of stabilisations, the so-called *augmented formulations*. Philosophically, we could think of them as based on the following observation. Suppose that we are given a general problem of the type: *find* $X \in \mathcal{W}$ *such that*

$$\langle \mathcal{A}X, Y\rangle_{\mathcal{W}'\times\mathcal{W}} = \langle F, Y\rangle_{\mathcal{W}'\times\mathcal{W}} \qquad \forall\, Y \in \mathcal{W}, \tag{6.1.9}$$

and assume that $\mathcal{A}$ is an isomorphism from $\mathcal{W}$ to $\mathcal{W}'$ (so that our problem has a unique solution). In general, as we observed already in Remark 1.5.1, given a subspace $\mathcal{W}_h \subset \mathcal{W}$, we cannot be sure that the discretised problem: *find* $X_h \in \mathcal{W}_h$ *such that*

$$\langle \mathcal{A}X_h, Y\rangle_{\mathcal{W}'\times\mathcal{W}} = \langle F, Y\rangle_{\mathcal{W}'\times\mathcal{W}} \qquad \forall\, Y \in \mathcal{W}_h \tag{6.1.10}$$

has a unique solution as well. On the other hand, still following Remark 1.5.1, if we assume that we have an ellipticity condition of the form

$$\exists\, \alpha > 0 \ \text{ such that } \ \langle \mathcal{A}Y, Y\rangle_{\mathcal{W}'\times\mathcal{W}} \geq \alpha \|Y\|^2_{\mathcal{W}} \qquad \forall\, Y \in \mathcal{W}, \tag{6.1.11}$$

(that clearly implies stability, with constant $1/\alpha$, and unique solvability of (6.1.9)) then *for every subspace* $\mathcal{W}_h \subset \mathcal{W}$, we will immediately have

$$\exists\, \alpha > 0 \ \text{ such that } \ \langle \mathcal{A}Y_h, Y_h\rangle_{\mathcal{W}'\times\mathcal{W}} \geq \alpha \|Y_h\|^2_{\mathcal{W}} \qquad \forall\, Y_h \in \mathcal{W}_h, \tag{6.1.12}$$

and we have unique solvability of (6.1.10) (with the same stability constant $1/\alpha$) without any need to be smart. Hence, although simple-minded, the general idea is: given the problem (6.1.9), try to present its solution as the solution of *another problem* for which an ellipticity condition of the type (6.1.11) holds true. In these precise terms, this is very easy. Indeed, the solution X of (6.1.9) will also be a solution of the problem: *find* $X \in \mathcal{W}$ *such that*:

$$(\mathcal{A}X, \mathcal{A}Y)_{\mathcal{W}'} = (F, \mathcal{A}Y)_{\mathcal{W}'} \qquad \forall Y \in \mathcal{W}. \tag{6.1.13}$$

Note that if $\mathcal{A}$ is an isomorphism between $\mathcal{W}$ and its dual $\mathcal{W}'$, then for every $Y \in \mathcal{W}$, we obviously have $Y = \mathcal{A}^{-1}(\mathcal{A}Y)$ and problem (6.1.13) will satisfy the ellipticity condition

$$(\mathcal{A}Y, \mathcal{A}Y)_{\mathcal{W}'} = \|\mathcal{A}Y\|^2_{\mathcal{W}'} \geq \frac{\|Y\|^2_{\mathcal{W}}}{\|\mathcal{A}^{-1}\|^2} \quad \forall\, Y \in \mathcal{W}. \tag{6.1.14}$$

At this level of generality, it is difficult to explain why, in several applications, we are not happy with this “solution”, and we still want to look for some different trick.

Just to make an example, if $\mathcal{A}$ is a differential operator (say, the Laplace operator), then problem (6.1.14) will correspond to a differential operator in which the order is doubled (in our example: the biharmonic operator) and for which discretisation would produce a matrix that is more ill-conditioned than the original one discretising $\mathcal{A}$. In this subsection, we will see some of these possible alternative techniques. We develop our discussion in the general setting of [46] but we shall mostly restrict our examples to the specific case of mixed methods.

We come back to the operator $\mathcal{A}$. At a general level, the operator $\mathcal{A}$ has a *symmetric part* $\mathcal{A}_s$, defined as

$$\mathcal{A}_s = (\mathcal{A} + \mathcal{A}^t)/2 \tag{6.1.15}$$

and an *antisymmetric part* $\mathcal{A}_a$, defined as

$$\mathcal{A}_a = (\mathcal{A} - \mathcal{A}^t)/2. \tag{6.1.16}$$

It is immediate to see that, for every $Y \in \mathcal{W}$, we have

$$\langle \mathcal{A}_s Y, Y \rangle = \langle \mathcal{A} Y, Y \rangle \quad \text{and} \quad \langle \mathcal{A}_a Y, Y \rangle = 0. \tag{6.1.17}$$

We point out that, keeping the assumption (6.1.3), we now have that $\mathcal{A}_s$ is symmetric and non-negative. Hence, we can use Lemma 4.2.1 and then (6.1.17) to obtain

$$\|\mathcal{A}_s Y\|^2_{\mathcal{W}'} \leq \langle \mathcal{A}_s Y, Y \rangle \, \|\mathcal{A}_s\| = \langle \mathcal{A} Y, Y \rangle \, \|\mathcal{A}_s\| \qquad \forall\, Y \in \mathcal{W}. \tag{6.1.18}$$

We then define, for $t \in \mathbb{R}$,

$$\mathcal{A}_t = \mathcal{A}_a + t\mathcal{A}_s \tag{6.1.19}$$

and we consider for $\mu > 0$ the following augmented problem: *find* $X \in \mathcal{W}$ *such that*

$${}_{\mathcal{W}'}\langle \mathcal{A}X - F, Y \rangle_{\mathcal{W}} + \mu(\mathcal{A}X - F, \mathcal{A}_t Y)_{\mathcal{W}'} = 0 \quad \forall\, Y \in \mathcal{W}. \tag{6.1.20}$$

Remark 6.1.2. We call the attention of the reader on the difference between $\mathcal{A}^t$ (the transposed operator of $\mathcal{A}$) and $\mathcal{A}_t$, defined by (6.1.19). We apologise for the similarity of these two symbols that have, however, a totally different meaning. □

It is clear that every solution X of the original problem (6.1.9) will also be a solution of the augmented problem (6.1.20). A possible advantage of the formulation (6.1.20) over (6.1.13) is that we can hope to be allowed to take μ small enough, so that the condition number of the resulting matrix will not be much worse than the condition number of the matrix coming from the discretisation of $\mathcal{A}$.

Example 6.1.1. As we already stated, we shall restrict our examples here to the case of mixed methods, that we write again in the form (6.1.6):

$$\mathcal{A} = \begin{pmatrix} A & B^t \\ -B & 0 \end{pmatrix}. \tag{6.1.21}$$

We shall assume here that the bilinear form $a(\cdot,\cdot)$ defining A is symmetric and non-negative, as in (5.5.1) (or in (4.2.28)). As already pointed out, the non-negativity of $a(\cdot,\cdot)$ will imply, in particular, that (6.1.3) is satisfied. The symmetry of $a(\cdot,\cdot)$, on the other hand, will imply that the symmetric part of the operator $\mathcal{A}$ is given by

$$\mathcal{A}_s = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \tag{6.1.22}$$

and the antisymmetric part is given by

$$\mathcal{A}_a = \begin{pmatrix} 0 & B^t \\ -B & 0 \end{pmatrix}. \tag{6.1.23}$$

From the symmetry and non-negativity of a, using (4.2.30) we have

$$\|Au\|^2_{V'} \le \|a\| a(u,u) = \|A\| \, \langle Au, u\rangle_{V'\times V} \tag{6.1.24}$$

that represents (6.1.18) in our particular case. It is not difficult to check that the stabilising term $(\mathcal{A}X, \mathcal{A}_t Y)_{\mathcal{W}'}$, for $X = (u, p)$ and $Y = (v, q)$ now becomes

$$\begin{aligned}(\mathcal{A}X, \mathcal{A}_t Y)_{\mathcal{W}'} &= (Au + B^t p, tAv + B^t q)_{V'} + (Bu, Bv)_{Q'} \\ &= t(Au + B^t p, Av)_{V'} + (Bu, Bv)_{Q'} + (Au + B^t p, B^t q)_{V'}. \end{aligned} \tag{6.1.25}$$

□

Example 6.1.2. The treatment of advection dominated equations is surely outside the scope of this book. However, it might be interesting to see how the general setting above can deal with a problem of the type: *find* $u \in H^1_0(\Omega)$ *such that*

$$-\varepsilon \Delta u + \mathbf{c} \cdot \underline{\mathrm{grad}}\, u = f \qquad \text{in } \Omega \tag{6.1.26}$$

where ε is a given positive and "small" number, $\mathbf{c}$ is a given smooth vector field (that, for simplicity, we assume to be *divergence-free*), and f is a given forcing term, say, in $L^2(\Omega)$. In this case, the stabilising term would be

$$(\mathcal{A}u, \mathcal{A}_t v)_{\mathcal{W}'} = (-\varepsilon\Delta u + \mathbf{c}\cdot\underline{\mathrm{grad}}\, u, -t\varepsilon\Delta v + \mathbf{c}\cdot\underline{\mathrm{grad}}\, v)_{H^{-1}(\Omega)}. \tag{6.1.27}$$

□

Remark 6.1.3. The structure of an augmented problem can be described as follows. First, we observe that the equation $\mathcal{A}X - F = 0$ takes place in the dual space $\mathcal{W}'$. Indeed, in its variational formulation (6.1.2), the equation is tested on a generic element $Y \in \mathcal{W}$, giving $\langle \mathcal{A}X - F, Y \rangle = 0$ for all Y. In the augmented problem, we keep this term, and we sum to it (with a suitable multiplier μ) a term containing the same equation, but this time tested on a term of the type $\mathcal{A}_t Y$ (always for Y generic in $\mathcal{W}$). Since the term $\mathcal{A}_t Y$ is itself in $\mathcal{W}'$ (as the difference $\mathcal{A}X - F$), this new term cannot be written as a duality between $\mathcal{W}'$ and $\mathcal{W}$, and we must take the scalar product of the two terms in $\mathcal{W}'$. The idea of *adding a term made by the scalar product of the equation times a suitable operator acting on the test function* Y is, somehow, the essence of the original idea of Hughes and Brooks, that has been extended and exploited in a more general setting by Hughes, Franca, and various co-authors, and became popular under the name of *stabilisation à la Hughes-Franca*. However, as we shall see, to take the inner product in $\mathcal{W}'$ is, in general, not so easy and the stabilising terms that are found in the literature (starting from the earliest ones by Hughes and his group) do not have exactly this form. Indeed, a big variety of different stabilising terms have been introduced, studied, and used in the literature of the last two or three decades (see, for instance, [10], [193], [214], [250] and [339]), all (or almost all) based on L^2 inner products (possibly multiplied by some suitable power of the mesh-size h) rather than on the $\mathcal{W}'$ inner product. However, as pointed out in [46], we could think at most of these variants as being different attempts to mimic, in one way or another, the effect of $\mu(\mathcal{A}X - F, \mathcal{A}_t Y)_{\mathcal{W}'}$. □

6.1.3 Stability Conditions for Augmented Formulations

Now, we want to study the behaviour of augmented problems of the type of (6.1.20). To start with, we look for sufficient conditions on t and μ ensuring that the augmented problem (6.1.20) has a unique solution.

Theorem 6.1.1. *Let $\mathcal{W}$ be a Hilbert space, and $\mathcal{A} \in \mathcal{L}(\mathcal{W}, \mathcal{W}')$ be an isomorphism which verifies* (6.1.3)*. If $t \in \mathbb{R}$ and $\mu > 0$ verify*

$$\mu(1-t)^2 < 4\|\mathcal{A}_s\|^{-1}, \tag{6.1.28}$$

then there exists $\alpha_{stab} > 0$ such that

$$\langle \mathcal{A}Y, Y \rangle_{\mathcal{W}' \times \mathcal{W}} + \mu(\mathcal{A}Y, \mathcal{A}_t Y)_{\mathcal{W}'} \geq \alpha_{stab} \|\mathcal{A}Y\|^2_{\mathcal{W}'} \qquad \forall\, Y \in \mathcal{W}, \tag{6.1.29}$$

where $\mathcal{A}_t$ is defined in (6.1.19)*.*

Proof. We apply (6.1.18) and (6.1.19), and then Cauchy-Schwarz to obtain

$$\begin{aligned}
&\langle \mathcal{A}Y, Y\rangle_{\mathcal{W}'\times\mathcal{W}} + \mu(\mathcal{A}Y, \mathcal{A}_t Y)_{\mathcal{W}'} \\
&\geq \frac{1}{\|\mathcal{A}_s\|}\|\mathcal{A}_s Y\|^2_{\mathcal{W}'} + \mu\Big(\|\mathcal{A}_a Y\|^2_{\mathcal{W}'} + t\,\|\mathcal{A}_s Y\|^2_{\mathcal{W}'} + (1+t)\,(\mathcal{A}_a Y, \mathcal{A}_s Y)_{\mathcal{W}'}\Big) \\
&\geq \Big(\frac{1}{\|\mathcal{A}_s\|} + \mu t\Big)\|\mathcal{A}_s Y\|^2_{\mathcal{W}'} + \mu\|\mathcal{A}_a Y\|^2_{\mathcal{W}'} - \mu|1+t|\,\|\mathcal{A}_a Y\|_{\mathcal{W}'}\,\|\mathcal{A}_s Y\|_{\mathcal{W}'}.
\end{aligned} \tag{6.1.30}$$

The last line of (6.1.30) is a quadratic form in $\|\mathcal{A}_s Y\|$ and $\|\mathcal{A}_a Y\|$. Hence, the desired (6.1.29) will be satisfied for some $\alpha_{stab} > 0$ if

$$4\mu\Big(\frac{1}{\|\mathcal{A}_s\|} + \mu t\Big) > (\mu(1+t))^2. \tag{6.1.31}$$

This can be written as

$$\frac{4}{\|\mathcal{A}_s\|} > \mu(1+t)^2 - 4\mu t = \mu(t-1)^2, \tag{6.1.32}$$

and the result follows. □

Remark 6.1.4. We note that condition (6.1.28) implies in particular that the coefficient of $\|\mathcal{A}_s Y\|$ in the last line of (6.1.30) is positive. This is clear if we note that (6.1.31) is actually *equivalent* to (6.1.28). □

Remark 6.1.5. It is immediate to see that, for $t = 1$, we have that (6.1.28) is satisfied for every value of $\mu > 0$. This is not so unreasonable since, for $t = 1$, we have $\mathcal{A}_t = \mathcal{A}$. One could then argue that $t = 1$ is the best choice and that other values for t have no interest. However, as we shall see, in several applications, including mixed formulations and advection dominated elliptic equations, both the choices $t = 0$ and $t = -1$ have been abundantly used. □

Essentially with the same proof, one has the following result, which is slightly more general.

Theorem 6.1.2. *Under the same assumptions as in Theorem 6.1.1, let $\mathcal{M}$ be a continuous, bilinear form on $\mathcal{W}' \times \mathcal{W}'$ and let M and μ_0 be positive constants such that*

$$\mathcal{M}(X', Y') \leq M\|X'\|_{\mathcal{W}'}\,\|Y'\|_{\mathcal{W}'} \qquad \forall\, X',\, Y' \in \mathcal{W}' \tag{6.1.33}$$

and

$$\mu_0\|Y'\|^2_{\mathcal{W}'} \leq \mathcal{M}(Y', Y') \qquad \forall\, Y' \in \mathcal{W}'. \tag{6.1.34}$$

If $t \in \mathbb{R}$ *and* $\mu > 0$ *verify*

$$\mu\Big(M^2(1+t)^2 - 4\mu_0^2 t\Big) < \frac{4\mu_0}{\|\mathcal{A}_s\|}, \tag{6.1.35}$$

then there exists $\alpha_{stab} > 0$ *such that*

$$\langle \mathcal{A}Y, Y\rangle_{\mathcal{W}'\times\mathcal{W}} + \mu\,\mathcal{M}(\mathcal{A}Y, \mathcal{A}_t Y) \geq \alpha_{stab}\|\mathcal{A}Y\|^2_{\mathcal{W}'} \qquad \forall\, Y \in \mathcal{W}, \tag{6.1.36}$$

where $\mathcal{A}_t$ *is always defined in* (6.1.19)*.*

Proof. We use (6.1.18), (6.1.19), (6.1.34), and then (6.1.33) to obtain

$$\begin{aligned}
&\langle \mathcal{A}Y, Y\rangle_{\mathcal{W}'\times\mathcal{W}} + \mu\mathcal{M}(\mathcal{A}Y, \mathcal{A}_t Y)\\
&\geq \frac{1}{\|\mathcal{A}_s\|}\|\mathcal{A}_s Y\|^2_{\mathcal{W}'} + \mu\Big(\mu_0\|\mathcal{A}_a Y\|^2_{\mathcal{W}'} + t\mu_0\,\|\mathcal{A}_s Y\|^2_{\mathcal{W}'} + (1+t)\,\mathcal{M}(\mathcal{A}_a Y, \mathcal{A}_s Y)_{\mathcal{W}'}\Big)\\
&\geq \Big(\frac{1}{\|\mathcal{A}_s\|} + \mu\mu_0 t\Big)\|\mathcal{A}_s Y\|^2_{\mathcal{W}'} + \mu\mu_0\|\mathcal{A}_a Y\|^2_{\mathcal{W}'} - \mu|1+t|\,M\,\|\mathcal{A}_a Y\|_{\mathcal{W}'}\,\|\mathcal{A}_s Y\|_{\mathcal{W}'}.
\end{aligned} \tag{6.1.37}$$

The last line of (6.1.30) is a quadratic form in $\|\mathcal{A}_s Y\|$ and $\|\mathcal{A}_a Y\|$. Hence, the desired (6.1.36) will be satisfied for some $\alpha_{stab} > 0$ if

$$4\mu\mu_0\Big(\frac{1}{\|\mathcal{A}_s\|} + \mu\mu_0 t\Big) > \mu^2 M^2(1+t)^2, \tag{6.1.38}$$

and the result follows. □

Remark 6.1.6. We note that condition (6.1.35) implies in particular that the coefficient of $\|\mathcal{A}_s Y\|$ in the last line of (6.1.37) is positive. This is again clear if we note that (6.1.38) is actually *equivalent* to (6.1.35). □

Remark 6.1.7. Looking at the proof of Theorems 6.1.1 and 6.1.2, we see that we could also write more specialised estimates, of the type

$$\alpha\Big(\frac{\|\mathcal{A}_s Y\|^2_{\mathcal{W}'}}{\|\mathcal{A}_s\|} + \mu\|\mathcal{A}_a Y\|^2_{\mathcal{W}'}\Big) \leq \langle \mathcal{A}Y, Y\rangle_{\mathcal{W}'\times\mathcal{W}} + \mu\mathcal{M}(\mathcal{A}Y, \mathcal{A}_t Y). \tag{6.1.39}$$

This would prove relevant in cases like advection dominated problems (6.1.26), where $\|\mathcal{A}_s\| \simeq \varepsilon$ and $\|\mathcal{A}_s v\|^2_{\mathcal{W}'} \simeq \|\varepsilon\Delta v\|^2_{\mathcal{W}'} \simeq \|\varepsilon v\|^2_{H^1}$. □

Remark 6.1.8. Theorem 6.1.2 reproduces Theorem 6.1.1 when we use $\mathcal{M}$ to define the scalar product in $\mathcal{W}'$ (so that $M = \mu_0 = 1$). □

Remark 6.1.9. It is also obvious that the exact solution X of (6.1.2) will also satisfy the *augmented formulation* of the problem: *find* $X \in \mathcal{W}$ *such that*

$$\langle \mathcal{A}X - F, Y\rangle_{\mathcal{W}'\times\mathcal{W}} + \mu\,\mathcal{M}(\mathcal{A}X - F, \mathcal{A}_t Y) = 0 \qquad \forall Y \in \mathcal{W}. \tag{6.1.40}$$

□

Remark 6.1.10. In the same assumptions as in Theorem 6.1.2, and essentially with the same proof, one could show that there exists an ω_0, depending only on $\|\mathcal{A}_s\|$, t, M, and μ_0, and an $\alpha_0 > 0$ such that, *for every* ω with $0 < \omega \le \omega_0$ and for every $Y \in \mathcal{W}$, one has

$$\langle \mathcal{A}Y, Y\rangle_{\mathcal{W}'\times\mathcal{W}} + \omega^2\,\mathcal{M}(\mathcal{A}Y, \mathcal{A}_t Y) \ge \alpha_0\Big(\|\mathcal{A}_s Y\|^2_{\mathcal{W}'} + \omega^2\|\mathcal{A}_a Y\|^2_{\mathcal{W}'}\Big). \tag{6.1.41}$$

The interest of this variant, as we shall see, is that we would be allowed to take an $\omega = \omega(h)$ that goes to zero with h. □

We now consider our main example, that is, mixed formulations (6.1.5) inserted in the present framework through (6.1.4). We have seen that the general stabilising term takes the form (6.1.25). We point out that, in particular,

$$(\mathcal{A}Y, \mathcal{A}_t Y)_{\mathcal{W}'} = t\|Av\|^2_{V'} + (1+t)(Av, B^t q)_{V'} + \|B^t q\|^2_{V'} + \|Bv\|^2_{Q'}, \tag{6.1.42}$$

while

$$\|\mathcal{A}_s Y\|_{\mathcal{W}'} = \|Av\|_{V'} \quad \text{and} \quad \|\mathcal{A}_a Y\|^2_{\mathcal{W}'} = \|B^t q\|^2_{V'} + \|Bv\|^2_{Q'}. \tag{6.1.43}$$

It is clear that the general philosophy, requiring that the stabilising term vanishes when X is the exact solution, would still be respected by taking a more general term, instead of $(\mathcal{A}X - F, \mathcal{A}_t Y)$. Hence, in some sense, we could specialise the result of Theorem 6.1.2 and adapt it to the case (here most interesting) of mixed methods. For instance, for general positive constants μ_1 and μ_2, we could consider a stabilising term of the form

$$(Au + B^t p - f, tAv + \mu_1 B^t q)_{V'} + (Bu - g, \mu_2 Bv)_{Q'}. \tag{6.1.44}$$

It is clear that if (u, p) is a solution of (6.1.5), then it is also a solution of

$$\begin{aligned} &{}_{V'}\langle Au + B^t p - f, v\rangle_V - {}_{Q'}\langle Bu - g, q\rangle_{Q'} \\ &\quad + \mu\Big((Au + B^t p - f, tAv + \mu_1 B^t q)_{V'} + (Bu - g, \mu_2 Bv)_{Q'}\Big) = 0 \end{aligned} \tag{6.1.45}$$

for all $v \in V$ and for all $q \in Q$.

Concerning the stability (and hence, in particular, the uniqueness of the solution of (6.1.45)), we note that

$$\begin{aligned} &(Av + B^t q, tAv + \mu_1 B^t q)_{V'} + (Bv, \mu_2 Bv)_{Q'} \\ &\quad = t\|Av\|^2_{V'} + (t + \mu_1)(Av, B^t q)_{V'} + \mu_1\|B^t q\|^2_{V'} + \mu_2\|Bv\|^2_{Q'}. \end{aligned} \tag{6.1.46}$$

Hence, mimicking the proof of Theorem 6.1.1, we easily have that

$$
\begin{aligned}
&{}_{V'}\langle Av+B^tq, v\rangle_V - {}_{Q'}\langle Bv, q\rangle_Q + \mu(Av+B^tq, tAv+\mu_1 B^tq)_{V'} + \mu(Bv, \mu_2 Bv)_{Q'} \\
&\geq \frac{1}{\|\mathcal{A}_s\|}\|Av\|^2_{V'} + \mu\Big(t\|Av\|^2_{V'} + (t+\mu_1)(Av, B^tq)_{V'} + \mu_1\|B^tq\|^2_{V'} + \mu_2\|Bv\|^2_{Q'}\Big) \\
&\geq \Big(\frac{1}{\|\mathcal{A}_s\|} + \mu t\Big)\|Av\|^2_{V'} + \mu\Big(\mu_1\|B^tq\|^2_{V'} + \mu_2\|Bv\|^2_{Q'} - |t+\mu_1|\,\|Av\|_{V'}\|\,\|B^tq\|_{V'}\Big) \\
&\geq C\Big(\|Av\|^2_{V'} + \mu_1\|B^tq\|^2_{V'} + \mu_2\|Bv\|^2_{Q'}\Big)
\end{aligned}
\tag{6.1.47}
$$

whenever μ is small enough, and precisely,

$$
\mu(t-\mu_1)^2 < \frac{4\mu_1}{\|\mathcal{A}_s\|}. \tag{6.1.48}
$$

We can make this result more explicit in the following theorem.

Theorem 6.1.3. *Let V and Q be Hilbert spaces, and a and b bilinear forms on $V \times V$ and $V \times Q$, respectively, as in Assumption $\mathcal{AB}$ of Chap. 4 (Sect. 4.2.1). Assume that a is symmetric and positive semi-definite as in (6.1.8), and assume that the continuous problem (6.1.5) is well posed (that is, a is elliptic on the kernel of B, and b satisfies the inf-sup condition). Let $t \in \mathbb{R}$ and let μ, μ_1, and μ_2 be positive real numbers. If (6.1.48) is satisfied, then there exists an $\alpha_M > 0$ such that, for every $(v, q) \in V \times Q$, we have*

$$
\begin{aligned}
&\alpha_M\Big(\|Av\|^2_{V'} + \mu_1\|B^tq\|^2_{V'} + \mu_2\|Bv\|^2_{Q'}\Big) \\
&\quad\leq {}_{V'}\langle Av + B^tq, v\rangle_V - {}_{Q'}\langle Bv, q\rangle_Q \\
&\quad+ \mu(Av + B^tq, tAv + \mu_1 B^tq)_{V'} + \mu(Bv, \mu_2 Bv)_{Q'}.
\end{aligned}
\tag{6.1.49}
$$

6.1.4 Discretisations of Augmented Formulations

The augmented formulations (6.1.40) or (6.1.45) can then be transported into the discretised problem.

Starting from the more general case of (6.1.40), we consider therefore the discrete stabilised problem: *find $X_h \in \mathcal{W}_h$ such that*

$$
\begin{aligned}
&\langle \mathcal{A}X_h, Y_h\rangle_{\mathcal{W}'\times\mathcal{W}} + \mu\mathcal{M}(\mathcal{A}X_h, \mathcal{A}_t Y_h) \\
&\qquad = \langle F, Y_h\rangle_{\mathcal{W}'\times\mathcal{W}} + \mu\mathcal{M}(F, \mathcal{A}_t Y_h) \qquad \forall\, Y_h \in \mathcal{W}_h.
\end{aligned}
\tag{6.1.50}
$$

It is clear that, whenever (6.1.35) holds true, the ellipticity property (6.1.36) will be inherited by the discrete problem, that will therefore be stable. Hence, we immediately have the following result.

Theorem 6.1.4. *Let $\mathcal{W}$ be a Hilbert space, and $\mathcal{A} \in \mathcal{L}(\mathcal{W}, \mathcal{W}')$ be an isomorphism which verifies* (6.1.3)*. Let moreover $\mathcal{M}$ be a continuous, bilinear form on $\mathcal{W}' \times \mathcal{W}'$ and let M and μ_0 be positive constants such that* (6.1.33) *and* (6.1.34) *are satisfied. Finally, let $t \in \mathbb{R}$ and $\mu > 0$ verify* (6.1.35)*, and let $\mathcal{A}_t$ be defined as in* (6.1.19)*. Then, for every $F \in \mathcal{W}'$ and for every finite dimensional subspace $\mathcal{W}_h$, denoting by X and X_h the solutions of the continuous problem* (6.1.40) *and of the stabilised-discretised one* (6.1.50)*, respectively, we have*

$$\|X - X_h\|_{\mathcal{W}} \leq C \inf_{Y_h \in \mathcal{W}_h} \|X - Y_h\|_{\mathcal{W}}, \tag{6.1.51}$$

where C is a constant depending on $\|\mathcal{A}^{-1}\|$, $\|\mathcal{A}\|$, μ, M, t and on the constant α_{stab} appearing in (6.1.36)*, bounded on bounded subsets, but independent of the choice of $\mathcal{W}_h$.*

Proof (Hint). As usual, for every $X_I \in \mathcal{W}_h$, we apply the stability estimate (6.1.36) to the difference $\delta X := X_h - X_I$. Then, in the right-hand side, we substitute X *in lieu* of X_h, using the fact that they are the solutions of (6.1.2) and (6.1.50), respectively. Finally, we use the continuity of $\mathcal{A}$, of $\mathcal{A}^{-1}$ and $\mathcal{M}$ to have an estimate of $\|X_h - X_I\|_{\mathcal{W}}$ in terms of $\|X - X_I\|_{\mathcal{W}}$. Then, we add and subtract X and use the triangle inequality. Finally, since X_I is generic in $\mathcal{W}_h$, we replace $\|X - X_I\|$ with the infimum of $\|X - Y_h\|$ for Y_h varying in $\mathcal{W}_h$. □

Remark 6.1.11. In the simplified case where $\mathcal{M}$ is the scalar product in $\mathcal{W}'$, the above problem (6.1.50) could *formally* be obtained by writing

$$\langle \mathcal{A}X_h - F, Y_h + \mu \mathcal{A}_t Y_h \rangle = 0 \quad \forall\ Y_h \in \mathcal{W}_h \tag{6.1.52}$$

and we could call this a "Petrov-Galerkin" method as the test functions are not in the same space as the solution. However, unless $\mathcal{W}$ can be identified to $\mathcal{W}'$, (6.1.52) has no sense. One must make a certain number of additional manipulations in order to reach a viable formulation. □

Shifting now to the particular case of mixed formulations, and considering (6.1.45), we assume that V_h and Q_h are finite dimensional subspaces of V and Q, respectively. It might be convenient to recall some definitions from the previous chapters. We do it quickly:

$$B_h := \pi_{Q_h'} B E_V \quad B_h^t := \pi_{V_h'} B E_Q \quad A_h := \pi_{V_h'} A E_V \tag{6.1.53}$$

$$\begin{aligned} K &:= \operatorname{Ker} B = \{v \in V \text{ s.t. } Bv = 0\}, \\ K_h &:= \operatorname{Ker} B_h = \{v_h \in V_h \text{ s.t. } B_h v_h = 0\}. \end{aligned} \tag{6.1.54}$$

We now consider the discretised problem: *find* $(u_h, p_h) \in V_h \times Q_h$ *such that*

$$
\begin{aligned}
&{}_{V'}\langle Au_h + B^t p_h - f, v_h\rangle_V + {}_{Q'}\langle Bu_h - g, q_h\rangle_{Q'} \\
&+ \mu(Au_h + B^t p_f - f, tAv_h + \mu_1 B^t q_h)_{V'} \\
&+ \mu(Bu_h - g, \mu_2 Bv_h)_{Q'} = 0 \qquad \forall\, v_h \in V_h\, \forall q_h \in Q_h.
\end{aligned}
\tag{6.1.55}
$$

It is clear that the stability result of Theorem 6.1.3 can now be used to get the following error estimate.

Theorem 6.1.5. *In the same assumptions as in Theorem 6.1.3, assume further that the continuous problem (6.1.5) is stable (that is, a is elliptic in the kernel, and b satisfies the inf-sup condition). Let $V_h \subset V$ and $Q_h \subset Q$ be finite dimensional subspaces, and for $f \in V'$ and $g \in Q'$, let (u, p) and (u_h, p_h) be the solutions of the continuous problem (6.1.45) and of (6.1.55), respectively. Then, we have*

$$
\|u - u_h\|_V^2 + \|p - p_h\|_Q^2 \le C\Big(\inf_{v_h \in V_h} \|u - v_h\|_V + \inf_{v_h \in V_h} \|u - v_h\|_V\Big) \tag{6.1.56}
$$

where C is a constant depending on $\|A^{-1}\|$, $\|A\|$, μ, M, t and on the constant α_M appearing in (6.1.49), bounded on bounded subsets, but independent of the choices of V_h and Q_h.

Proof. The proof follows exactly the same lines as the proof of Theorem 6.1.4, and the classical form of all the “stability+consistency” error bound. □

We shall now make explicit problem (6.1.45) in a few special cases. It is not difficult to see that (6.1.45) corresponds to have a linear “augmented operator” of the type

$$
\begin{aligned}
\mathbb{M}_{stab} &= \begin{pmatrix} A & B^t \\ -B & 0 \end{pmatrix} \\
&+ \mu\Big(t\begin{pmatrix} A^tA & A^tB^t \\ 0 & 0 \end{pmatrix} + \mu_1\begin{pmatrix} 0 & 0 \\ BA & BB^t \end{pmatrix} + \mu_2\begin{pmatrix} B^tB & 0 \\ 0 & 0 \end{pmatrix}\Big).
\end{aligned}
\tag{6.1.57}
$$

Let us see, for $\mu = 1$, three typical values of t, namely $t = 1$, $t = 0$ and $t = -\mu_1$.

(i) **Case t = 1**

The augmented system is:

$$
\begin{cases}
\langle Au_h + B^t p_h - f, v_h\rangle_{V'\times V} + (Au_h + B^t p_h - f, Av_h)_{V'} \\
\qquad + \mu_2(Bu_h - g, Bv_h)_{Q'} = 0 \quad \forall\, v_h \in V_h, \\
\langle -Bu_h + g, q\rangle_{Q'\times Q} \\
\qquad + \mu_1(Au_h + B^t p_h - f, B^t q_h)_{V'} = 0 \quad \forall\, q_h \in Q_h.
\end{cases}
\tag{6.1.58}
$$

(ii) **Case t = 0**
The augmented system is:

$$\begin{cases} \langle Au_h + B^t p_h - f, v_h \rangle_{V' \times V} \\ \qquad + \mu_2 (Bu_h - g, Bv_h)_{Q'} = 0 \quad \forall\, v_h \in V_h, \\ \langle -Bu_h + g, q \rangle_{Q' \times Q} \\ \qquad + \mu_1 (Au_h + B^t p_h - f, B^t q_h)_{V'} = 0 \quad \forall\, q_h \in Q_h. \end{cases} \tag{6.1.59}$$

(iii) **Case t** $= -\mu_1$
The augmented system is:

$$\begin{cases} \langle Au_h + B^t p_h - f, v_h \rangle_{V' \times V} - \mu_1 (Au_h + B^t p_h - f, Av_h)_{V'} \\ \qquad + \mu_2 (Bu_h - g, Bv_h)_{Q'} = 0 \quad \forall\, v_h \in V_h, \\ \langle -Bu_h + g, q \rangle_{Q' \times Q} \\ \qquad + \mu_1 (Au_h + B^t p_h - f, B^t q_h)_{V'} = 0 \quad \forall\, q_h \in Q_h. \end{cases} \tag{6.1.60}$$

Remark 6.1.12. From the point of view of "economy", the case $t = 0$ implies the smallest number of extra terms and would be our favourite. On the other hand, we have seen that, in several cases, the choice $t = 1$ guarantees stability *for every value of the stabilisation parameter* μ *in* (6.1.50), and this is also a nice feature. Finally, for the choice $t = -\mu_1$, we can see that the final expression of $\mathbb{M}_{stab}$ in (6.1.57) is

$$\mathbb{M}_{stab} = \begin{pmatrix} A - \mu_1 A^t A + \mu_2 B^t B & B^t - \mu_1 A^t B^t \\ -B + \mu_1 BA & \mu_1 BB^t \end{pmatrix} \tag{6.1.61}$$

which, changing the sign of the second equation, becomes *symmetric* since obviously we have $(B^t - \mu_1 A^t B^t)^t = B - \mu_1 BA$. In conclusion, all the three choices present some interesting aspects. □

Remark 6.1.13. It is not too difficult to spot the role of each of the extra terms in (6.1.58) and (6.1.57). Indeed, we can easily see that if A is coercive on the kernel of B (a property that, in general, will not be inherited by the discretised problem), then, according to Proposition 4.3.4,

$$\langle Au, u \rangle_{V' \times V} + \mu_2 (Bu, Bu)_{Q' \times Q'} \geq \tilde{\alpha} \|u\|_V^2 \quad \forall\, u \in V, \tag{6.1.62}$$

for a suitable constant $\tilde{\alpha}$, a property that will be inherited by the discretised problem. It is then clear that the extra term on the first equation (that is, the term containing μ_2) will allow to bypass problems related to the coercivity of the bilinear form a. On the other hand, the extra term in the second equation will help in controlling p as the (continuous) *inf-sup* condition implies

$$\mu_1 (B^t q, B^t q)_{V' \times V'} \equiv \mu_1 \|B^t q\|_{V'}^2 \geq \mu_1 \beta^2 \|q\|_Q^2. \tag{6.1.63}$$

The choice between the three possibilities above will obviously depend on the type of discretisation that we want to use, as well as on many other possible considerations. We will see some of them in the following chapters. □

Remark 6.1.14. We point out that, in most applications, things are usually not *totally bad*: in general, we will either have a lack of coercivity on the kernel but a good *inf-sup* condition or the reverse. Very roughly speaking, the lack of the *inf-sup* condition will occur when V_h is not big enough, compared with Q_h (so that, for instance, the image of B_h will not fill Q'_h). On the other hand, if V_h, compared with Q_h, is too big, then the kernel K_h of B_h will contain elements that are not in the kernel K of the operator B, and the ellipticity in the kernel, for the discrete problem, would fail. In these cases, we can limit ourselves to a *lighter* stabilisation. Two typical cases are

1. To retrieve coercivity of a: in (6.1.61), take $t = \mu_1 = 0$

$$\begin{cases} \langle Au_h + B^t p_h - f, v_h\rangle_{V'\times V} + \mu_2(Bu_h - g, Bv_h)_{Q'} = 0 \quad \forall\, v_h \in V_h, \\ \langle -Bu_h + g, q_h\rangle_{Q'\times Q} = 0 \quad \forall\, q_h \in Q_h. \end{cases} \tag{6.1.64}$$

2. To retrieve the *inf-sup* condition for b: in (6.1.61) take $t = \mu_2 = 0$

$$\begin{cases} \langle Au_h + B^t p_h - f, v_h\rangle_{V'\times V} = 0 \quad \forall\, v_h \in V_h, \\ \langle -Bu_h + g, q_h\rangle_{Q'\times Q} + \mu_1(Au_h + B^t p_h - f, B^t q_h)_{V'} = 0\; \forall\, q_h \in Q_h. \end{cases} \tag{6.1.65}$$

It is easily seen, following the path of Theorem 6.1.1, that the above problems are stable under suitable conditions. The simplest case would be that A is defined by a bilinear form which is coercive on V such as in the Stokes problem. This will be developed in Chap. 8. In that case, a stabilisation such as in (6.1.65) would be sufficient. The first case (6.1.64) is nothing but the discretised version of (1.5.10). We will come back to this line of thought in the next section. □

Remark 6.1.15 (Caveat emptor). We recall that we have used the **exact** norms in V' and Q'. In many cases, (e.g. when this would imply the use of the H^{-1} norm) this may well be impossible (or very difficult) to implement numerically, and we shall have to introduce an approximation of our stabilised problem. This will typically be done by applying, in the discretised problem, the differential operators *element-wise*, and then substituting the H^{-1} scalar product with h^2 times the L^2 scalar product. □

6.1.5 Stabilising with the "Element-Wise Equations"

To give an idea of the techniques mentioned in the above remark, we consider the following variant of Theorem 6.1.4. As we shall see, the variant follows the spirit of Remark 6.1.10, and is closely connected with the family of methods of the next subsection. For this, however, we have to introduce some new objects. We assume that we have a space $\mathcal{W}^+$ (made of smoother functions) and a Hilbert space $\mathcal{H}$ (that we identify with its dual space $\mathcal{H}'$) such that

$$\mathcal{W}^+ \subset \mathcal{W} \subseteq \mathcal{H} \equiv \mathcal{H}' \subseteq \mathcal{W}' \tag{6.1.66}$$

and

$$\mathcal{A}_s(\mathcal{W}^+) \subseteq \mathcal{H}, \quad \mathcal{A}_a(\mathcal{W}^+) \subseteq \mathcal{H}, \tag{6.1.67}$$

and for all h

$$\mathcal{A}_a(\mathcal{W}_h) \subseteq \mathcal{H}. \tag{6.1.68}$$

We also assume that we have, for all h, a linear operator

$$\mathcal{S}_h : \mathcal{H} + \mathcal{A}_s(\mathcal{W}_h) + \mathcal{A}_a(\mathcal{W}_h) \to \mathcal{H} \tag{6.1.69}$$

such that

$$\|\mathcal{S}_h Y\|_{\mathcal{H}} = \|Y\|_{\mathcal{H}} \qquad \forall Y \in \mathcal{H}, \tag{6.1.70}$$

and we note that, together with (6.1.68), this gives

$$\|\mathcal{S}_h(\mathcal{A}_a Y_h)\|_{\mathcal{H}} \geq \|\mathcal{A}_a Y_h\|_{\mathcal{H}} \qquad \forall Y_h \in \mathcal{W}_h. \tag{6.1.71}$$

We assume further that there exists a monotonically increasing function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\omega(h)\|\mathcal{S}_h(\mathcal{A}_r Y_h)\|_{\mathcal{H}} \leq \|\mathcal{A}_r Y_h\|_{\mathcal{W}'} \quad \text{where } r = s \text{ or } a, \ \forall\, Y_h \in \mathcal{W}_h. \tag{6.1.72}$$

Remark 6.1.16. In the applications that we have in mind, the space $\mathcal{H}$ will be either L^2 or a Cartesian product of several copies of L^2, and the operator $\mathcal{S}_h$ will be the one that allows to take the element-by-element derivatives of functions that are smooth (typically, polynomials) inside each element but might be discontinuous from one element to the next (or are continuous but not C^1, when you take second derivatives). In mathematical words, $\mathcal{S}_h(\chi)$ would take the restriction $\chi_{|T}$ to each individual *open* triangle T, and then consider the L^2 function that in each triangle T is equal to $\chi_{|T}$. In this way, possible Dirac masses concentrated on the inter-element boundaries would be dropped. Having this in mind, it should be clear that the assumption in (6.1.68) is a *very strong* one, and in all the applications that we considered, it requires either that the antisymmetric part of $\mathcal{A}$ is an operator of *lower order* (as it happens for advection dominated flows) or that the elements of $\mathcal{W}_h$ have, in a certain sense, *more continuity* than strictly necessary (as when using continuous pressures in the Stokes problem). □

Assuming further that

$$F \in \mathcal{H} \tag{6.1.73}$$

where F is the right-hand side of (6.1.1), we can consider the discretised problem: *find* $X_h \in \mathcal{W}_h$ *such that*

$$\langle \mathcal{A}X_h - F, Y_h\rangle_{\mathcal{W}'\times\mathcal{W}} + \mu\omega^2(h)(\mathcal{S}_h(\mathcal{A}X_h - F), \mathcal{S}_h(\mathcal{A}_t Y_h))_{\mathcal{H}} = 0 \tag{6.1.74}$$

for all $Y_h \in \mathcal{W}_h$. Proceeding as in Theorems 6.1.1 and 6.1.2, it is not difficult to see that if

$$(1-t)^2\mu\,\omega^2(h) < \frac{4}{\|\mathcal{A}_s\|}, \tag{6.1.75}$$

then there exists $\alpha_0 > 0$ such that

$$\begin{aligned}\langle \mathcal{A}Y_h, Y_h\rangle_{\mathcal{W}'\times\mathcal{W}} &+ \mu\,\omega^2(h)\,(\mathcal{S}_h(\mathcal{A}Y_h), \mathcal{S}_h(\mathcal{A}_t Y_h))_{\mathcal{H}}\\ &\geq \alpha_0\Big(\|\mathcal{A}_s Y_h\|^2_{\mathcal{W}'} + \mu\,\omega^2(h)\|\mathcal{S}_h(\mathcal{A}_a Y_h)\|^2_{\mathcal{H}}\Big) \quad \forall Y_h \in \mathcal{W}_h.\end{aligned} \tag{6.1.76}$$

We can now apply the above estimate to have a bound on the error.

Theorem 6.1.6. *Let* $\mathcal{W}$ *be a Hilbert space and* $\mathcal{A} \in \mathcal{L}(\mathcal{W}, \mathcal{W}')$ *be an isomorphism which verifies* (6.1.3)*. Assume that all the additional assumptions* (6.1.66)–(6.1.73) *are satisfied, and assume further that the solution* X *of problem* (6.1.2) *belongs to* $\mathcal{W}^+$*. For* $t \in \mathbb{R}$ *and for* $\mu > 0$*, let* X_h *be the solution of* (6.1.74)*. If* (6.1.75) *is satisfied, then there exists a constant* C*, depending only on* α_0*,* t*, and* μ*, such that*

$$\begin{aligned}\|\mathcal{A}_s(X - X_h)\|_{\mathcal{W}'} &+ \omega(h)\|\mathcal{S}_h(\mathcal{A}_a(X - X_h))\|_{\mathcal{H}}\\ &\leq C \inf_{Y_h\in\mathcal{W}_h}\Big((\|X - Y_h\|_{\mathcal{W}} + \omega^{-1}(h)\|X - Y_h\|_{\mathcal{H}}\\ &\quad + \omega(h)\|\mathcal{S}_h(\mathcal{A}_a(X - Y_h))\|_{\mathcal{H}}\Big).\end{aligned} \tag{6.1.77}$$

Proof. We first observe that the Galerkin orthogonality equation

$$\langle \mathcal{A}(X - X_h), Y_h\rangle_{\mathcal{W}'\times\mathcal{W}} + \mu\omega^2(h)(\mathcal{S}_h(\mathcal{A}(X - X_h)), \mathcal{S}_h(\mathcal{A}_t Y_h))_{\mathcal{H}} = 0 \tag{6.1.78}$$

holds for all $Y_h \in \mathcal{W}_h$. Then let X_I be a generic element of $\mathcal{W}_h$, and set as before $\delta X := X_h - X_I$ and $\delta_I := X - X_I$. We apply the estimate (6.1.76) to δX and then we add and subtract X and use (6.1.78) to obtain

$$
\begin{aligned}
&\alpha_0\Big(\|\mathcal{A}_s\delta X\|^2_{\mathcal{W}'} + \mu\,\omega^2(h)\|\mathcal{S}_h(\mathcal{A}_a\delta X)\|^2_{\mathcal{H}}\Big)\\
&\quad\le \langle\mathcal{A}\,\delta X,\delta X\rangle_{\mathcal{W}'\times\mathcal{W}} + \mu\,\omega^2(h)\,(\mathcal{S}_h(\mathcal{A}\delta X),\mathcal{S}_h(\mathcal{A}_t\delta X))_{\mathcal{H}}\\
&\quad= \langle\mathcal{A}\,\delta_I,\delta X\rangle_{\mathcal{W}'\times\mathcal{W}} + \mu\,\omega^2(h)\,(\mathcal{S}_h(\mathcal{A}\delta_I),\mathcal{S}_h(\mathcal{A}_t\delta X))_{\mathcal{H}}.
\end{aligned}
\tag{6.1.79}
$$

The first term in the last line of (6.1.79), using (6.1.19), (6.1.68), and then (6.1.71), can be estimated by

$$
\begin{aligned}
&\langle\mathcal{A}\,\delta_I,\delta X\rangle_{\mathcal{W}'\times\mathcal{W}} \le \|\delta_I\|_{\mathcal{W}}\cdot\|\mathcal{A}_s\delta X\|_{\mathcal{W}'} + \|\delta_I\|_{\mathcal{H}}\cdot\|\mathcal{A}_a\delta X\|_{\mathcal{H}},\\
&\le \|\delta_I\|_{\mathcal{W}}\cdot\|\mathcal{A}_s\delta X\|_{\mathcal{W}'} + \omega^{-1}(h)\|\delta_I\|_{\mathcal{H}}\cdot\omega(h)\|\mathcal{S}_h(\mathcal{A}_a\delta X)\|_{\mathcal{H}},\\
&\le (\|\delta_I\|_{\mathcal{W}} + \omega^{-1}(h)\|\delta_I\|_{\mathcal{H}})\cdot(\|\mathcal{A}_s\delta X\|_{\mathcal{W}'} + \omega(h)\|\mathcal{S}_h(\mathcal{A}_a\delta X)\|_{\mathcal{H}})
\end{aligned}
\tag{6.1.80}
$$

while the second term, using (6.1.19) and then (6.1.72), is easily estimated by

$$
\begin{aligned}
&\omega^2(h)\,(\mathcal{S}_h(\mathcal{A}\delta_I),\mathcal{S}_h(\mathcal{A}_t\delta X))_{\mathcal{H}}\\
&\quad\le \omega(h)\,\|\mathcal{S}_h(\mathcal{A}\delta_I)\|_{\mathcal{H}}\cdot\omega(h)\|\mathcal{S}_h(\mathcal{A}_t\delta X)\|_{\mathcal{H}}\\
&\quad\le \omega(h)\,\|\mathcal{S}_h(\mathcal{A}\delta_I)\|_{\mathcal{H}}\cdot\Big(|t|\|\mathcal{A}_s\delta X\|_{\mathcal{W}'} + \omega(h)\|\mathcal{S}_h(\mathcal{A}_a\delta X)\|_{\mathcal{H}}\Big).
\end{aligned}
\tag{6.1.81}
$$

The result (6.1.77) now follows easily by a repeated use of the arithmetic-geometric mean inequality and finally, the use of the triangle inequality to estimate $X - X_h$ in terms of δX and δ_I. Note that the last term in the right-hand side of (6.1.77) appears only in this final step (using the triangle inequality).

Remark 6.1.17. In most applications, the constant $\omega(h)$ corresponds to some *inverse inequality* applied to piecewise polynomial functions. The same constant (in terms of powers of h) will often appear if we compare the best approximation of a smooth function X taken in the norm of $\mathcal{H}$ rather than in the (stronger) norm of $\mathcal{W}$. As a result, the first two terms appearing in the right-hand side of (6.1.77) will, in general, be of *the same order*, and the third will be either of the same order or smaller. □

Remark 6.1.18. As we can see, the strong assumption (6.1.68) has been used only to estimate the term $\langle\mathcal{A}_a\delta_I,\delta X\rangle$ in (6.1.80). In a certain number of applications, one could take advantage of some particular feature of the problem at hand, and survive without it. To do so when dealing with the abstract problem would be, however, very complicated. Hence, we defer the analysis of the different applications of the above theory to the following chapters, mostly to Chap. 8 concerning the Stokes problem, and we just consider here below some example of the possible stabilisations of Laplace operator in mixed form. □

Example 6.1.3 (Stabilisation of the mixed Poisson problem). In Sect. 1.5.1 of Chap. 1, we have considered many augmented methods for the mixed formulation of the Dirichlet problem. Most of these methods can be written in the framework that we have just developed. For simplicity, we refer to the simplest formulations (1.5.2) and (1.5.9), that we briefly recall for the convenience of the reader:

$$(\underline{u},\underline{v})+(\operatorname{div}\underline{v},p)-(\operatorname{div}\underline{u}+f,q)=0,\quad \forall\,\underline{v}\in H(\operatorname{div};\Omega)\ \forall\, q\in L^2(\Omega)\tag{6.1.82}$$

and

$$(\underline{u}-\underline{\operatorname{grad}}\,p,\underline{v})+(\underline{u},\underline{\operatorname{grad}}\,q)-(f,q)=0\quad \forall\,\underline{v}\in L^2(\Omega)\ \forall\, q\in H^1_0(\Omega).\tag{6.1.83}$$

Other examples will be seen in the following chapters. We have therefore, for the formulation (6.1.82), $\mathcal{W}=H(\operatorname{div};\Omega)\times L^2(\Omega)$ and, for the formulation (6.1.83), $\mathcal{W}=(L^2(\Omega))^d\times H^1_0(\Omega)$. In all cases, we take $\mathcal{H}:=(L^2(\Omega))^d\times L^2(\Omega)$, and we use the symbol $(\cdot\,,\cdot)$ to denote the inner product in $L^2(\Omega)$ or in $(L^2(\Omega))^d$. We also assume, for simplicity, that the solution $(\underline{u},p)$ belongs to $(H^1(\Omega))^d\times H^2(\Omega)\cap H^1_0(\Omega)$. Finally, following the common usage, we denote by $\underline{\operatorname{grad}}_h\,q$ the element-wise gradient $\mathcal{S}_h(\underline{\operatorname{grad}}\,q)$ and by $\operatorname{div}_h\underline{v}$ the element-wise divergence $\mathcal{S}_h(\operatorname{div}\underline{v})$. In the first case (that is when using the formulation (6.1.82)), we have

$$\begin{aligned}(\underline{u},\underline{v})&+(\operatorname{div}\underline{v},p)-(\operatorname{div}\underline{u}+f,q)\\&+\mu h^2\Big(t((\underline{u},\underline{v})+(\operatorname{div}\underline{v},p))\\&-\mu_1(\underline{u}-\underline{\operatorname{grad}}_h\,p,\underline{\operatorname{grad}}_h\,q)+\mu_2(\operatorname{div}\underline{u}+f,\operatorname{div}\underline{v})\Big)=0,\end{aligned}\tag{6.1.84}$$

while in the second case (that is when using the formulation (6.1.83)), we have instead

$$\begin{aligned}(\underline{u}&-\underline{\operatorname{grad}}\,p,\underline{v})+(\underline{u},\underline{\operatorname{grad}}\,q)-(f,q)\\&+\mu h^2\Big(t(\underline{u}-\underline{\operatorname{grad}}\,p,\underline{v})\\&-\mu_1(\underline{u}-\underline{\operatorname{grad}}\,p,\underline{\operatorname{grad}}\,q)+\mu_2(\operatorname{div}_h\underline{u}+f,\operatorname{div}_h\underline{v})\Big)=0.\end{aligned}\tag{6.1.85}$$

Let us see some particular cases related to this last example. In all cases, we will take, for simplicity, $\mu=1$.

(i) **Case $\mathbf{t}=\mathbf{1}$**. In this case, the augmented formulation is

$$\begin{aligned}&(1+h^2)(\underline{u}-\underline{\operatorname{grad}}\,p,\underline{v})+\mu_2h^2(\operatorname{div}_h\underline{u}+f,\operatorname{div}_h\underline{v})=0\quad\forall\,\underline{v}\in(L^2(\Omega))^2,\\&(1-\mu_1h^2)(\underline{u},\underline{\operatorname{grad}}\,q)+\mu_1h^2(\underline{\operatorname{grad}}\,p,\underline{\operatorname{grad}}\,q)-(f,q)=0\quad\forall\,q\in H^1_0(\Omega).\end{aligned}\tag{6.1.86}$$

Note that stability holds for every choice of $\mu_2\geq 0$.

(ii) **Case $\underline{t = 0}$**. In this case, the augmented formulation is

$$(\underline{u} - \underline{\mathrm{grad}}\, p, \underline{v}) + \mu_2 h^2 (\mathrm{div}_h\, \underline{u} + f, \mathrm{div}_h\, \underline{v}) = 0 \quad \forall\, \underline{v} \in (L^2(\Omega))^2,$$
$$(1 - \mu_1 h^2)(\underline{u}, \underline{\mathrm{grad}}\, q) + \mu_1 h^2 (\underline{\mathrm{grad}}\, p, \underline{\mathrm{grad}}\, q) - (f, q) = 0 \quad \forall\, q \in H_0^1(\Omega). \tag{6.1.87}$$

This formulation, in particular with $\mu_1 = 0$, is particularly appealing for discretisations in which the *inf-sup* condition holds already but the ellipticity in the kernel is lacking.

(iii) **Case $\mathbf{t} = -\boldsymbol{\mu}_1$**. In this case, the augmented formulation is

$$(1 - \mu_1 h^2)(\underline{u} - \underline{\mathrm{grad}}\, p, \underline{v}) + \mu_2 h^2 (\mathrm{div}_h\, \underline{u} + f, \mathrm{div}_h\, \underline{v}) = 0\ \forall\, \underline{v} \in (L^2(\Omega))^2,$$
$$(1 - \mu_1 h^2)(\underline{u}, \underline{\mathrm{grad}}\, q) + \mu_1 h^2 (\underline{\mathrm{grad}}\, p, \underline{\mathrm{grad}}\, q) - (f, q) = 0 \quad \forall\, q \in H_0^1(\Omega). \tag{6.1.88}$$

Note that, changing the sign of the second equation, we reach a *symmetric* problem, as already pointed out in Remark 6.1.12. □

6.2 Other Stabilisations

In this subsection, we still want to deal with methods for transforming the problem in a stable one, but not necessarily reaching a formulation where ellipticity holds. In particular, here, we want to analyse methods to fix discretisations that have already some sort of stability, in a spirit similar to the one of Remark 6.1.14.

6.2.1 General Stability Conditions

We go back to our original abstract formulation (6.1.1) which we re-write for the convenience of the reader. We consider the problem: *find $X \in \mathcal{W}$ such that*

$$\mathcal{A}X = F, \tag{6.2.1}$$

together with its variational formulation

$$\langle \mathcal{A}X, Y \rangle_{\mathcal{W}' \times \mathcal{W}} = \langle F, Y \rangle_{\mathcal{W}' \times \mathcal{W}} \qquad \forall\, Y \in \mathcal{W}. \tag{6.2.2}$$

We also recall that we assumed the non-negativity condition (6.1.3) that we also repeat here

$$\langle \mathcal{A}Y, Y \rangle_{\mathcal{W}' \times \mathcal{W}} \geq 0, \qquad \forall Y \in \mathcal{W}. \tag{6.2.3}$$

The following result is an exercise of functional analysis, but, for the convenience of the readers, we sketch a proof.

Proposition 6.2.1. *If* (6.2.3) *holds, then the two following conditions are equivalent:*

$$(i) \quad \mathcal{A} \text{ is an isomorphism from } \mathcal{W} \text{ onto } \mathcal{W}' \tag{6.2.4}$$

(ii) $\exists\, \Phi \in \mathcal{L}(\mathcal{W}, \mathcal{W})$ *and a constant* $\alpha_\Phi > 0$ *such that*

$$\langle \mathcal{A}Y, \Phi(Y)\rangle_{\mathcal{W}'\times\mathcal{W}} \geq \alpha_\Phi \|Y\|^2_{\mathcal{W}} \qquad \forall\, Y \in \mathcal{W}. \tag{6.2.5}$$

Proof. Let $J = R_{\mathcal{W}'}$ be the Ritz operator from $\mathcal{W}'$ to $\mathcal{W}$ as defined in Theorem 4.1.2. The implication $(i) \implies (ii)$ follows by taking $\Phi = J\mathcal{A}$. To prove the converse implication, we denote by Id the identity operator in $\mathcal{W}$, and we remark that, if (6.1.3) holds, then for every positive real number s, we have, for all $Y \in \mathcal{W}$,

$$\langle (s\Phi + Id)^t \mathcal{A}Y, Y\rangle_{\mathcal{W}'\times\mathcal{W}} = \langle \mathcal{A}Y, (s\Phi + Id)Y\rangle_{\mathcal{W}'\times\mathcal{W}} \geq s\,\alpha_\Phi \|Y\|^2_{\mathcal{W}}.$$

This easily implies that $(s\Phi + Id)^t\mathcal{A}$ is an isomorphism from $\mathcal{W}$ onto $\mathcal{W}'$. On the other hand, we know that $s\Phi + Id$ is an isomorphism for s small enough (see for instance Theorem 4.1.3), so that $(s\Phi + Id)^t$ will also be an isomorphism, as well as its inverse $(s\Phi + Id)^{-t}$. Hence, $\mathcal{A} = [(s\Phi + Id)^{-t}][(s\Phi + Id)^t\mathcal{A}]$ (as product of two isomorphisms) is also an isomorphism, and (i) holds. □

Remark 6.2.1. If we further assume $\mathcal{A} = \mathcal{A}^t$ (that is, if we assume the bilinear form $\langle \mathcal{A}Y, Y\rangle$ to be *symmetric*), then, using Lemma 4.2.2, we see that in (6.2.5) we could always use $\Phi = Id$, and the equivalence would still hold. □

Remark 6.2.2. If (6.1.3) is not satisfied, we always have $(i) \implies (ii)$ but the converse is false. This can be seen by considering in $L^2(\,]0, +\infty[\,)$ the mapping:

$$\begin{cases} (\mathcal{A}u)(x) = u(x-1) \text{ for } x > 1 \\ (\mathcal{A}u)(x) = 0 \text{ for } 0 < x \leq 1 \end{cases}$$

(corresponding to shifting the graph of u to the right by 1, and inserting 0 in the interval $(0, 1)$). Clearly, (ii) is satisfied by taking $\Phi u := \mathcal{A}u$, but (i) is not, as $\mathcal{A}$ is injective but not surjective. For an operator that does not satisfy (6.1.3), we would need two conditions instead of (6.2.5), that is: $\exists\, \Phi_1, \Phi_2 \in \mathcal{L}(\mathcal{W}, \mathcal{W})$ such that, for all $Y \in \mathcal{W}$,

$$\begin{cases} \langle \mathcal{A}Y, \Phi_1(Y)\rangle_{\mathcal{W}'\times\mathcal{W}} \geq \alpha_1 \|Y\|^2_{\mathcal{W}}, \\ \langle \Phi_2(Y), \mathcal{A}^tY\rangle_{\mathcal{W}\times\mathcal{W}'} \geq \alpha_2 \|Y\|^2_{\mathcal{W}}, \end{cases} \tag{6.2.6}$$

implying that $\mathcal{A}$ is both injective and surjective. □

Remark 6.2.3. It must be noted that the **stability constant** of Problem (6.1.2), that is, the smallest constant C such that

$$\|X\| \le C\|\mathcal{A}X\| \qquad \forall\, X \in \mathcal{W}, \tag{6.2.7}$$

is not $1/\alpha_\Phi$ (see (6.2.5)) but rather

$$C = \|\Phi\|/\alpha_\Phi. \tag{6.2.8}$$

□

Remark 6.2.4. We now consider again the case of (6.1.4), in which the abstract problem (6.2.2) is just a different way of writing the mixed problem (6.1.5). For this case, we want to get an explicit construction of some Φ that satisfies (6.2.5) starting from the usual stability conditions developed previously in Chap. 4. In other words, we are going to see the equivalence of (6.2.5) with the *ellipticity in the kernel* and *inf-sup* conditions. We thus consider, for any given $X^* = (u^*, p^*)$ in $V \times Q$, two auxiliary problems, which have a unique solution if the mixed problem (6.1.5) is well posed:

– *Find* (u_1, p_1), *solution of*

$$\begin{cases} a(v,u_1) + b(v,p_1) = (u^*, v)_V & \forall\, v \in V, \\ b(u_1, q) = 0 & \forall\, q \in Q, \end{cases} \tag{6.2.9}$$

– *Find* (u_2, p_2), *solution of*

$$\begin{cases} a(v,u_2) + b(v,p_2) = 0 & \forall\, v \in V, \\ b(u_2, q) = (p^*, q)_Q & \forall\, q \in Q. \end{cases} \tag{6.2.10}$$

In other words, we take $(u_1, p_1) = \mathcal{A}^{-1}(R_V u^*, 0)$ and $(u_2, p_2) = \mathcal{A}^{-1}(0, R_Q p^*)$, where R_V and R_Q are the Ritz operators from V to V' and from Q to Q', respectively (see (4.1.37)). We now set $\Phi((u^*, p^*)) := (u_1 + u_2, -p_1 - p_2)$ and we have:

$$\begin{aligned} A(X^*, \Phi(X^*)) &= a(u^*, u_1 + u_2) + b(u_1 + u_2, p^*) + b(u^*, p_1 + p_2) \\ &= \|u^*\|_V^2 + \|p^*\|_Q^2 = \|X\|_{\mathcal{W}}^2. \end{aligned} \tag{6.2.11}$$

□

Remark 6.2.5. Problems (6.2.9) and (6.2.10) could, by linearity, be combined into one. We preferred to make more explicit the separate control of $\|u^*\|_V$ and $\|p^*\|_Q$. One should also note that (see Remark 6.2.3) the stability constant in (6.2.7) (which, using (6.2.8), is now equal to $\|\Phi\|$, since by (6.2.11) we have $\alpha_\Phi = 1$) depends through (6.2.9) and (6.2.10) on the usual constants defining, for example, the coercivity in the kernel and the *inf-sup* condition. No *free lunch*. □

6.2.2 Stability of Discretised Formulations

Let us now turn to the discretisation of problem (6.1.2). For a given sequence of subspaces $\mathcal{W}_h$ of $\mathcal{W}$ (usually of finite dimension), we consider, for each h, the discrete problem: *find* $X_h \in \mathcal{W}_h$ *such that*

$$\langle \mathcal{A}X_h, Y_h\rangle = \langle F, Y_h\rangle \qquad \forall\, Y_h \in \mathcal{W}_h. \tag{6.2.12}$$

In general, for an arbitrary choice of the sequence $\{\mathcal{W}_h\}$, (6.2.12) will not be stable, that is, we cannot ensure that there exists a sequence of linear operators $\Phi_h \in \mathcal{L}(\mathcal{W}_h, \mathcal{W}_h)$, uniformly bounded in h, such that for some $\alpha_\Phi > 0$ independent of h:

$$\langle \mathcal{A}Y_h, \Phi_h(Y_h)\rangle \geq \alpha_\Phi \|Y_h\|^2_{\mathcal{W}} \qquad \forall\, Y_h \in \mathcal{W}_h. \tag{6.2.13}$$

We suppose that we have, for each h, a stabilising term R with the structure

$$R(X_h, Y_h) := L(X_h, Y_h) + \langle N, Y_h\rangle \tag{6.2.14}$$

where N, which is possibly null, will depend on F, and where $L(X_h, Y_h)$ is a continuous bilinear form on $\mathcal{W}_h$ with a continuity constant c_L,

$$|L(X_h, Y_h)| \leq c_L \|X_h\|_{\mathcal{W}} \|Y_h\|_{\mathcal{W}}. \tag{6.2.15}$$

In practice, we shall build $R(X_h, Y_h)$ in such a way that it can be used as a *stabilising term* in a sense that will be defined in hypothesis **H.0** below. All the stabilisations of the previous section (see, for instance, (6.1.20) or (6.1.40)) had indeed the above structure. Here, however, we shall often use just the bilinear part $L(X_h, Y_h)$.

We shall now consider an abstract error estimate based on the following hypothesis.

H.0 We have:

(i) A continuous problem

$$\langle \mathcal{A}X, Y\rangle = \langle F, Y\rangle \qquad \forall\, Y \in \mathcal{W}, \tag{6.2.16}$$

which we assume to have a unique solution,

(ii) A sequence of stabilised discrete problems

$$\langle \mathcal{A}X_h, Y_h\rangle + rR(X_h, Y_h) = \langle F, Y_h\rangle \qquad \forall\, Y_h \in \mathcal{W}_h, \tag{6.2.17}$$

where $R(X_h, Y_h)$ is of the form (6.2.14) and $r > 0$ is a scalar,

(iii) Two constants $\tilde{c}_\Phi$ and $\tilde{\alpha}_\Phi$, and an operator $\tilde{\Phi}_h \in \mathcal{L}(\mathcal{W}_h, \mathcal{W}_h)$ such that

$$\|\tilde{\Phi}_h(Y_h)\| \le \tilde{c}_\Phi \|Y_h\| \quad \forall\, Y_h \in \mathcal{W}_h \tag{6.2.18}$$

and

$$\langle \mathcal{A}Y_h, \tilde{\Phi}_h(Y_h)\rangle_{\mathcal{W}'\times\mathcal{W}} + rL(Y_h, \tilde{\Phi}_h(Y_h)) \ge \tilde{\alpha}_\Phi \|Y_h\|^2_{\mathcal{W}}. \tag{6.2.19}$$

□

Under the assumption **H.0**, we have the following error bound.

Proposition 6.2.2. *Assume that* **H.0** *holds, and let X and X_h be the solutions of* (6.2.16) *and* (6.2.17) *respectively. For every $X_I \in \mathcal{W}_h$, let us set*

$$\mathcal{R}(X_I) := \sup_{Y_h \in \mathcal{W}_h} \frac{R(X_I, Y_h)}{\|Y_h\|}. \tag{6.2.20}$$

We then have

$$\frac{\tilde{\alpha}_\Phi}{\tilde{c}_\Phi} \|X_I - X_h\| \le \|\mathcal{A}\|\, \|X - X_I\| + r\mathcal{R}(X_I), \tag{6.2.21}$$

and consequently

$$\|X - X_h\| \le \frac{\tilde{c}_\Phi \|\mathcal{A}\| + \tilde{\alpha}_\Phi}{\tilde{\alpha}_\Phi} \|X - X_I\| + \frac{\tilde{c}_\Phi r\mathcal{R}(X_I)}{\tilde{\alpha}_\Phi}. \tag{6.2.22}$$

Proof. Set $\delta X := X_I - X_h$ and $\tilde{Y}_h := \tilde{\Phi}_h(\delta X)$. From (6.2.18), we immediately have

$$\|\tilde{Y}_h\| \le \tilde{c}_\Phi\, \|\delta X\|. \tag{6.2.23}$$

On the other hand, using (6.2.19), adding and subtracting X and using (6.2.14), then using (6.2.16) and (6.2.17), and finally (6.2.20), we obtain:

$$\begin{aligned}
\tilde{\alpha}_\Phi\, \|\delta X\|^2 &\le \langle \mathcal{A}\delta X, \tilde{Y}_h\rangle + r\, L(\delta X, \tilde{Y}_h) \\
&= \langle \mathcal{A}(X_I - X), \tilde{Y}_h\rangle + \langle \mathcal{A}X, \tilde{Y}_h\rangle - \langle \mathcal{A}X_h, \tilde{Y}_h\rangle - rR(X_h, \tilde{Y}_h) + rR(X_I, \tilde{Y}_h) \\
&= \langle \mathcal{A}(X_I - X), \tilde{Y}_h\rangle + rR(X_I, \tilde{Y}_h) \\
&\le \|\tilde{Y}_h\|\, (\|\mathcal{A}\|\, \|X_I - X\| + r\mathcal{R}(X_I))
\end{aligned}$$

and (6.2.21) follows immediately using (6.2.23). Finally, (6.2.22) follows from (6.2.21) using the triangle inequality. □

On many occasions, as we have seen in the previous section, the perturbation term R can be chosen in such a way that the strong consistency property (usually called *Galerkin orthogonality*) still holds. In these cases, the solution X of (6.2.16) would verify

$$R(X, Y_h) = 0 \qquad \forall\, Y_h \in \mathcal{W}_h, \tag{6.2.24}$$

implying that for the discrete stabilised problem we have

$$\langle \mathcal{A}X, Y_h\rangle + rR(X, Y_h) = \langle F, Y_h\rangle \qquad \forall\, Y_h \in \mathcal{W}_h. \tag{6.2.25}$$

In this case, we have, essentially by the same proof as in Proposition 6.2.2, the following corollary.

Corollary 6.2.1. *Assume that* **H.0** *holds, and let X and X_h be the solutions of* (6.2.16) *and* (6.2.17) *respectively. Assume moreover that X satisfies the strong consistency condition* (6.2.24)*. Then, we have*

$$\|X - X_h\| \le \frac{\tilde{c}_\Phi(\|\mathcal{A}\| + rc_L) + \tilde{\alpha}_\Phi}{\tilde{\alpha}_\Phi} \inf_{Y_h \in \mathcal{W}_h} \|X - Y_h\| \tag{6.2.26}$$

where c_L is defined in (6.2.15)*.*

Remark 6.2.6. It is clear that the above results, and in particular Corollary 6.2.1, could be applied to the methods of the previous section. □

The results of Proposition 6.2.2 and of Corollary 6.2.1 are of a *general* nature and, in order to obtain sharper results, we shall have to specialise somehow the construction of $R(X_h, Y_h)$ and its properties. This will be done in the following subsection.

6.3 Minimal Stabilisations

In several applications, we will have that there exists a subspace $\overline{\mathcal{W}}_h \subseteq \mathcal{W}$ and a positive constant $\overline{\alpha}$ such that

$$\langle \mathcal{A}Z, Z\rangle \ge \overline{\alpha}\|\pi_{\overline{\mathcal{W}}_h} Z\|^2 \qquad \forall Z \in \overline{\mathcal{W}}_h, \tag{6.3.1}$$

or, more generally,

$${}_{\mathcal{W}'}\langle \mathcal{A}Z, \overline{\Phi}_h(Z)\rangle_{\mathcal{W}} \ge \overline{\alpha}\|\pi_{\overline{\mathcal{W}}_h} Z\|_{\mathcal{W}}\, \|\overline{\Phi}_h(Z)\| \qquad \forall Z \in \overline{\mathcal{W}}_h, \tag{6.3.2}$$

for some linear mapping $\overline{\Phi}_h$ from $\mathcal{W}_h$ to $\mathcal{W}_h$. In these cases, we can consider that the part of the solution that belongs to $\overline{W}_h$ will somehow be "*under control*" and will not need to be stabilised.

In these cases, the stabilising term R in (6.2.14) could be chosen of the form

$$R(X_h, Y_h) = (G_h X_h - N, G_h Y_h)_{\mathcal{H}} \tag{6.3.3}$$

for some suitable Hilbert space $\mathcal{H}$, a suitable N in $\mathcal{H}$ (either equal to 0 or depending on f), and a suitable linear operator G_h from $\mathcal{W}_h$ to $\mathcal{H}$. Conditions for $(G_h X_h, G_h Y_h)_{\mathcal{H}}$ to be stabilising will be given in the subsections below. Often, roughly speaking, one would take $N = 0$ and use a G_h with

$$\operatorname{Ker} G_h = \overline{\mathcal{W}}_h, \tag{6.3.4}$$

so that G_h will act only on the part of $\mathcal{W}_h$ that is *not* in $\overline{\mathcal{W}}_h$. In other cases, as we already did at the end of the previous section, we have to deal with several equations, and G_h will act differently on each of them. Moreover, in several cases, $\overline{\mathcal{W}}_h$ will contain all the *low frequencies* of $\mathcal{W}_h$, so that a smooth solution $X \in \mathcal{W}$ could be approximated fairly well by elements $\overline{X}_h \in \overline{W}_h$. Then, for every $X_I \in \mathcal{W}_h$ and for every $\overline{X}_h \in \overline{\mathcal{W}}_h$, we will have that the term $\mathcal{R}(X_I)$ in (6.2.21) can be estimated by

$$\frac{R(X_I, Y_h)}{\|Y_h\|} = \frac{R(X_I - \overline{X}_h, Y_h)}{\|Y_h\|} \le c_L \|X_I - \overline{X}_h\| \le c_L(\|X_I - X\| + \|X - \overline{X}_h\|),$$

so that, from (6.2.21), we have in this case

$$\frac{\tilde{\alpha}_\Phi}{\tilde{c}_\Phi} \|X_I - X_h\| \le (\|\mathcal{A}\| + r c_L) \|X - X_I\| + r\, c_L \|X - \overline{X}_h\| \tag{6.3.5}$$

and the error estimate will depend on the approximation properties of both $\mathcal{W}_h$ and $\overline{\mathcal{W}}_h$, on the value of c_L and on the choice of r. We shall now provide a precise and sharper analysis of some of these situations.

We still suppose that the discrete problem defined by (6.2.12) is not stable. We may however suppose that a partial stability holds for some semi-norm $[Y_h]_h$ on $\mathcal{W}_h$.

Remark 6.3.1. In general, the "biggest semi-norm" one could consider is clearly

$$[X]_h := \sup_{Y_h \in \mathcal{W}_h} \frac{{}_{\mathcal{W}'}\langle \mathcal{A}X, Y_h \rangle_{\mathcal{W}}}{\|Y_h\|_{\mathcal{W}}}. \tag{6.3.6}$$

However, in many applications, simpler (and more explicit) norms can be preferred. □

The following assumption expresses in a precise way the fact that a certain semi-norm $[X_h]_h$ is "under control":

H.1 *For every h, there exists*

(i) A semi-norm $[\cdot]_h$ on $\mathcal{W}$,
(ii) An operator $\Phi_h \in \mathcal{L}(\mathcal{W}_h, \mathcal{W}_h)$,
(iii) A constant c_Φ such that

$$\|\Phi_h(Y_h)\|_{\mathcal{W}} \le c_\Phi \|Y_h\|_{\mathcal{W}} \quad \forall\, Y_h \in \mathcal{W}_h, \tag{6.3.7}$$

(iv) A constant $\alpha_\Phi > 0$ such that

$$\langle A Y_h, \Phi_h(Y_h)\rangle \ge \alpha_\Phi [Y_h]_h^2 \quad \forall\, Y_h \in \mathcal{W}_h. \tag{6.3.8}$$

□

Assumption **H.1** might seem cumbersome or difficult to realise in practice. This is not the case. Indeed, before proceeding, we point out that Assumption **H.1** is indeed verified in a number of applications. In particular, we consider the following (rather typical) situations.

Minimal stabilisation of mixed formulations. Assume that $\langle \mathcal{A}X, Y\rangle$ is defined as in (6.1.4), and recall the definitions (6.1.53) and (6.1.54).

For every fixed $X_h \equiv (u_h, p_h) \in \mathcal{W}_h$, we consider, in the spirit of Remark 6.3.1,

$$S(X_h) := \sup_{(v_h,q_h)\in V_h\times Q_h} \frac{{}_{\mathcal{W}'}\langle \mathcal{A}(u_h, p_h), (v_h, q_h)\rangle_{\mathcal{W}}}{\|(v_h, q_h)\|_{\mathcal{W}}}. \tag{6.3.9}$$

Assuming that (6.1.8) holds, we will always have

$$\begin{aligned} S(X_h) &\ge \frac{{}_{\mathcal{W}'}\langle \mathcal{A}(u_h, p_p), (u_h, p_h)\rangle_{\mathcal{W}}}{\|(u_h, p_h)\|_{\mathcal{W}}} \\ &\ge \frac{a(u_h, u_h)}{\|(u_h, p_h)\|_{\mathcal{W}}} =: \frac{|u_h|_a^2}{\|(u_h, p_h)\|_{\mathcal{W}}}. \end{aligned} \tag{6.3.10}$$

Similarly, we have (always without any assumptions on V_h or Q_h)

$$S(X_h) \ge \sup_{(v_h,0)\in K_h\times\{0\}} \frac{a(u_h, v_h)}{\|v_h\|_V} = \|\pi_{K_h'} A\, u_h\|_{V'}, \tag{6.3.11}$$

$$S(X_h) \ge \sup_{(0,q_h)\in \{0\}\times Q_h} \frac{b(u_h, q_h)}{\|q_h\|_Q} = \|B_h\, u_h\|_{Q'}, \tag{6.3.12}$$

$$S(X_h) \ge \sup_{(v_h,0)\in K_h^\perp\times\{0\}} \frac{a(u_h, v_h) + b(v_h, p_h)}{\|v_h\|_V} = \|\pi_{K_h^0} A\, v_h + B_h^t\, q_h\|_{V'}, \tag{6.3.13}$$

but, in general, we would not be able to get estimates of the type

$$S(X_h) \geq C\|A_h u_h\|_V \quad \text{or} \quad S(X_h) \geq \|B_h^t\|_{Q'} \tag{6.3.14}$$

separately. In most particular cases, however, one might have some good property and exploit it. We have seen in the previous chapters that sufficient conditions that ensure stability and error estimates are the *ellipticity in the kernel* K_h (*elker*) of the bilinear form $a(\cdot\,,\cdot)$ and the *discrete inf-sup condition for the bilinear form* $b(\cdot\,,\cdot)$. We also pointed out that the two conditions play, in a certain sense, one against the other: taking a bigger V_h helps in ensuring the *inf-sup condition* but increases the kernel and makes *elker* more at risk and the other way round. It is therefore not unreasonable to assume that we already took care of *one of the two conditions* (just by increasing or decreasing one of the two spaces), and ask the help of some stabilising trick in order to take care of the other. More precisely, we assume, to start with, that we have a continuous problem that is well posed.

A.1 *We suppose that* $\langle \mathcal{A}X, Y\rangle$ *is defined as in* (6.1.4) *and that*

(*i*) *The bilinear form* $a(\cdot,\cdot)$ *is* K*-elliptic, that is,*

$$\exists\, \alpha_0 > 0 \; s.t. \quad a(v,v) \geq \alpha\|v\|_V^2 \qquad \forall\, v \in K = \mathrm{Ker}B; \tag{6.3.15}$$

(*ii*) *The bilinear form* $b(v,q)$ *satisfies the inf-sup condition in* $V \times Q$ *where*

$$\beta := \inf_{v\in V}\sup_{q\in Q}\frac{b(v,q)}{\|v\|_V\,\|q\|_Q}. \tag{6.3.16}$$

□

In what follows, we will discuss the cases in which one of the two conditions is not satisfied for the discretised problem.

Example 6.3.1 (Minimal stabilisation of the inf-sup condition). For simplicity, we further assume that the ellipticity condition holds on the whole V, that is,

$$\exists\, \alpha > 0 \; s.t. \quad a(v,v) \geq \alpha\|v\|_V^2 \qquad \forall\, v \in V. \tag{6.3.17}$$

We know that the full ellipticity in V (6.3.17) implies automatically the full ellipticity in V_h. On the other hand, this is not true, in general, for the *inf-sup* condition. Hence, as we consider methods in need to be stabilised, we suppose that the discrete *inf-sup* condition *does not hold* with a constant independent of h. In order to see that, however, Assumption **H.1** is satisfied; we consider the following semi-norm:

$$[Y_h]_h^2 = [(v_h, q_h)]^2 := \|v_h\|_V^2 + [\![q_h]\!]_h^2 \tag{6.3.18}$$

where

$$[\![q_h]\!]_h := \sup_{v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_V} \equiv \|B_h^t q_h\|_{V'} \qquad \forall\, q_h \in Q_h. \tag{6.3.19}$$

The following proposition states that, in this framework, Assumption **H.1** holds.

Proposition 6.3.1. *Let $\mathcal{A}$ be of the form* (6.1.4) *and assume that* **A.1** *holds, together with the full ellipticity* (6.3.17)*. Then,* **H.1** *also holds. In particular,* (6.3.7) *and* (6.3.8) *hold with*

$$\alpha_{\Phi} = \frac{\alpha}{2}\, min\left(1, \frac{1}{\|a\|^2}\right), \tag{6.3.20}$$

$$c_{\Phi} = 1 + \frac{\alpha}{\|a\|^2} \tag{6.3.21}$$

and with the semi-norm $[\cdot]_h$ *defined in* (6.3.18) *and* (6.3.19).

Proof. For a given $Y_h := (v_h, q_h)$, let $v_h^* \in V_h$ be such that

$$\frac{b(v_h^*, q_h)}{\|v_h^*\|_V} = \sup_{w_h \in V_h} \frac{b(w_h, q_h)}{\|w_h\|_V} =: [\![q_h]\!]_h \tag{6.3.22}$$

scaled in such a way that

$$\|v_h^*\|_V = [\![q_h]\!]_h. \tag{6.3.23}$$

We now choose

$$\Phi_h(Y_h) = (v_h + \delta v_h^*, q_h), \tag{6.3.24}$$

with δ a positive real number to be specified later on. We have from (6.1.4) and (6.3.24):

$$\begin{aligned}
\langle \mathcal{A} Y_h, \Phi_h(Y_h)\rangle = a(v_h, v_h) + \delta a(v_h, v_h^*) \\
+ b(v_h, q_h) + \delta b(v_h^*, q_h) - b(v_h, q_h) \\
\geq \alpha\|v_h\|_V^2 - \delta\|a\|\,\|v_h\|_V\|v_h^*\|_V + \delta[\![q_h]\!]_h\|v_h^*\|_V \\
= \alpha\|v_h\|_V^2 - \delta\|a\|\,\|v_h\|_V[\![q_h]\!]_h + \delta[\![q_h]\!]_h^2,
\end{aligned} \tag{6.3.25}$$

having used (6.3.17), (6.3.22), and, in the last step, (6.3.23). It is now clear that, choosing $\delta = \alpha/\|a\|^2$, (6.3.25) implies

$$\langle A Y_h, \Phi_h(Y_h)\rangle \geq \frac{\alpha}{2}\|v_h\|_V^2 + \frac{\delta}{2}[\![q_h]\!]_h^2 \tag{6.3.26}$$

having used $2ab \leq a^2 + b^2$. Hence, we have (6.3.8) with the constant α_Φ given by (6.3.20). On the other hand, (6.3.23) and the choice of δ imply (6.3.7) and (6.3.21) since

$$\|v_h - \delta v_h^*\| \leq \|v_h\| + \delta\|v_h^*\| = \|v_h\| + \delta[\![q_h]\!]_h. \qquad \square$$

Remark 6.3.2. Looking at the above proof, we can see that, actually, we proved, instead of (6.3.7), the stronger inequality

$$\|\Phi_h(Y_h)\| \leq c_\Phi [Y_h]_h \quad \forall\, Y_h \in \mathcal{W}_h. \tag{6.3.27}$$

$\square$

Example 6.3.2 (Minimal stabilisations of the ellipticity condition). Another possible case in which **H.1** is satisfied is the following one, in which we suppose, this time, that the discrete *inf-sup* condition *does hold* with a constant independent of h, but the ellipticity in the kernel does not. For instance, we might have that a is elliptic on the kernel K of B, but the kernel K_h of B_h is not a subset of K, and ellipticity does not hold for all $v_h \in K_h$.

In particular, we assume that **A1** holds, that the discrete *inf-sup* condition

$$\exists \beta^* > 0 \text{ such that } := \inf_{v\in V} \sup_{q\in Q} \frac{b(v,q)}{\|v\|_V\, \|q\|_Q} \geq \beta_* \tag{6.3.28}$$

holds with β^* independent of h, and that, moreover, as in (5.2.37) and (5.2.38), there exists a Hilbert space V^* with $V \hookrightarrow V^*$ such that

$$\exists\, \alpha^* > 0 \text{ such that } a(v,v) \geq \alpha^* \|v\|^2_{V^*} \qquad \forall\, v \in V, \tag{6.3.29}$$

together with

$$\exists\, M_a^* \text{ such that } a(u,v) \leq M_a^* \|u\|_{V^*}\, \|v\|_{V^*} \qquad \forall\, u,\, v \in V. \tag{6.3.30}$$

Then, we consider the following semi-norm:

$$[Y_h]_h^2 = [(v_h, q_h)]^2 := \|v_h\|^2_{V^*} + \|q_h\|^2_Q. \tag{6.3.31}$$

The following proposition states that in this framework, Assumption **H.1** holds.

Proposition 6.3.2. *Let $\mathcal{A}$ be of the form* (6.1.4) *and assume that* **A.1** *holds, together with assumptions* (6.3.28)–(6.3.30)*. Then,* **H.1** *also holds. In particular,* (6.3.7) *and* (6.3.8) *hold with*

$$\alpha_\Phi = \frac{\alpha}{2}\, min\left(1, \frac{\beta_*^2}{(M_a^*)^2}\right), \tag{6.3.32}$$

$$c_\Phi = 1 + \frac{\alpha^*\beta_*}{(M_a^*)^2} \tag{6.3.33}$$

and with the semi-norm $[\cdot]_h$ *defined in* (6.3.31)*.*

Proof. For a given $Y_h := (v_h, q_h)$, we use (6.3.28) to choose $v_h^* \in V_h$ such that

$$\frac{b(v_h^*, q_h)}{\|v_h^*\|_V} = \sup_{w_h \in V_h} \frac{b(w_h, q_h)}{\|w_h\|_V} \geq \beta_* \|q_h\|_Q^2, \tag{6.3.34}$$

scaled in such a way that

$$\|v_h^*\|_V = \|q_h\|_Q. \tag{6.3.35}$$

We now choose

$$\Phi_h(Y_h) = (v_h + \delta v_h^*, q_h), \tag{6.3.36}$$

with δ a positive real number to be specified later on. We have from (6.1.4) and (6.3.36):

$$\begin{aligned}\langle \mathcal{A}Y_h, \Phi_h(Y_h)\rangle &= a(v_h, v_h) + \delta a(v_h, v_h^*)\\ &\quad + b(v_h, q_h) + \delta b(v_h^*, q_h) - b(v_h, q_h)\\ &\geq \alpha^* \|v_h\|_{V^*}^2 - \delta M_a^* \|v_h\|_{V^*} \|v_h^*\|_{V^*} + \delta\beta_* \|q_h\|_Q \|v_h^*\|_V\\ &= \alpha^* \|v_h\|_V^2 - \delta M_a^* \|v_h\|_V \|q_h\|_Q + \delta\beta_* \|q_h\|_Q^2, \end{aligned} \tag{6.3.37}$$

having used (6.3.15), (6.3.34), and, in the last step, (6.3.35). It is now clear that, choosing $\delta = \alpha^*\beta_*/(M_a^*)^2$, (6.3.25) implies

$$\langle \mathcal{A}Y_h, \Phi_h(Y_h)\rangle \geq \frac{\alpha^*}{2}\|v_h\|_V^2 + \frac{\delta\beta_*}{2}[\![q_h]\!]_h^2, \tag{6.3.38}$$

having used $2ab \leq a^2 + b^2$. Hence, we have (6.3.8) with the constant α_Φ given by (6.3.32). On the other hand, (6.3.35) and the choice of δ imply (6.3.7) with (6.3.33), since

$$\|v_h - \delta v_h^*\| \leq \|v_h\| + \delta\|v_h^*\| = \|v_h\| + \delta\|q_h\|. \qquad \square$$

Remark 6.3.3. Looking at the above proof, we can see that, together with (6.3.7), we could also prove the inequality

$$[\Phi_h(Y_h)]_h \leq c_\Phi [Y_h]_h \quad \forall\, Y_h \in \mathcal{W}_h, \tag{6.3.39}$$

which in this case is stronger than (6.3.7). □

Remark 6.3.4. It is important to note that $\Phi_h(v_h, q_h)$, defined by (6.3.24) or (6.3.36), leaves the **second component** of (v_h, q_h) **unchanged**. This property can be useful in several circumstances. □

General results on minimal stabilisations. Having seen that Assumption **H.1** is indeed a reasonable one, we are now going to see how to use it in order to stabilise the problem. Roughly speaking, as we have already mentioned, we are going to add a bilinear form $L(X_h, Y_h)$ on $\mathcal{W}_h \times \mathcal{W}_h$, assuming that it could take care of "the remaining part of the $\mathcal{W}$ norm", that is "the part of the $\mathcal{W}$ norm which is not controlled by the semi-norm $[\,\cdot\,]_h$".

For technical reasons, we are going to make this assumption *in two steps*: we shall first assume in **H.2** that $L(Y_h, Y_h)$ controls a suitable intermediate term $\|G_h Y_h\|^2_{\mathcal{H}}$ (to be discussed later on), and then we shall assume, in **H.3**, that this intermediate term, together with the semi-norm $[\,\cdot\,]_h$, can control the whole $\mathcal{W}$ norm. Let us see this in a more precise way.

H.2 *There exist a Hilbert space $\mathcal{H}$, a bilinear form $L \in \mathcal{L}(\mathcal{H}, \mathcal{H})$, three positive constants c_G, c_L and α_G, and, for every h, an operator $G_h \in \mathcal{L}(\mathcal{W}_h, \mathcal{H})$, with $\|G_h\|_{\mathcal{L}(\mathcal{W}_h,\mathcal{H})} \leq c_G$, such that*

$$L(G_h Z_h, G_h Y_h) \leq c_L \|G_h Z_h\|_{\mathcal{H}}\, \|G_h Y_h\|_{\mathcal{H}} \qquad \forall\, Y_h,\ Z_h \in \mathcal{W}_h, \tag{6.3.40}$$

$$L(G_h Y_h, G_h Y_h) \geq \alpha_G \|G_h Y_h\|^2_{\mathcal{H}} \qquad \forall\, Y_h \in \mathcal{W}_h. \tag{6.3.41}$$

□

Remark 6.3.5. It is clear that hypothesis **H2** is tailored for using a stabilising term R of the form (6.3.3). □

We now consider, for some positive real number r, the stabilised operator $\tilde{\mathcal{A}}$ defined as

$$\langle \tilde{\mathcal{A}} X_h, Y_h \rangle := \langle \mathcal{A} X_h, Y_h \rangle + r L(G_h X_h, G_h Y_h) \qquad \forall\, X_h, Y_h \in \mathcal{W}_h \tag{6.3.42}$$

and the corresponding regularised problem

$$\langle \tilde{\mathcal{A}} X_h, Y_h \rangle = r L(N, G_h Y_h) + \langle F, Y_h \rangle \qquad \forall\, Y_h \in \mathcal{W}_h. \tag{6.3.43}$$

We have the following result.

Lemma 6.3.1. *Assume that* **H.1** *and* **H.2** *hold, and assume moreover that, for the mapping Φ_h considered in* **H.1** *and the map G_h considered in* **H.2**, *we have*

$$\|G_h(\Phi_h(Y_h))\|_{\mathcal{H}} \leq c_{G\Phi} \|G_h(Y_h)\|_{\mathcal{H}} \quad \forall Y_h \in \mathcal{W}_h \tag{6.3.44}$$

for some constant $C_{G\Phi}$ independent of h. Then, there exist a linear mapping $\Phi_h^* \in \mathcal{L}(\mathcal{W}_h, \mathcal{W}_h)$ and two constants α_Φ^* and c_Φ^*, depending only on α_Φ, c_Φ α_R, and c_R such that, for every h, for every r and for every $Y_h \in \mathcal{W}_h$, we have

$$\|\Phi_h^*(Y_h)\|_{\mathcal{W}} \le c_\Phi^* \|Y_h\|_{\mathcal{W}}, \tag{6.3.45}$$

$$\|G_h(\Phi_h^*(Y_h))\|_{\mathcal{H}} \le c_\Phi^* \|G_h(Y_h)\|_{\mathcal{H}}, \tag{6.3.46}$$

and

$$\langle \tilde{\mathcal{A}} Y_h, \Phi_h^*(Y_h)\rangle \ge \alpha_\Phi^* \Big([Y_h]_h^2 + r\|G_h(Y_h)\|_{\mathcal{H}}^2\Big), \tag{6.3.47}$$

where $\tilde{\mathcal{A}}$ is given in (6.3.42). □

Proof. We set

$$\Phi_h^*(Y_h) := Y_h + \delta\Phi_h(Y_h) \tag{6.3.48}$$

with δ to be chosen later on. Then, using first (6.3.42) and (6.3.48), then (6.3.8), (6.3.41), and using (6.3.44) to bound the last term, we have:

$$\begin{aligned} &\langle \tilde{\mathcal{A}} Y_h, \Phi_h^*(Y_h)\rangle \\ &\ge \delta\alpha_\Phi [Y_h]_h^2 + r\alpha_G \|G_h(Y_h)\|_{\mathcal{H}}^2 - r c_L \|G_h(Y_h)\|_{\mathcal{H}} \delta c_{G\Phi} \|G_h(Y_h)\|_{\mathcal{H}}, \end{aligned} \tag{6.3.49}$$

and the result follows easily for $r\delta$ smaller than $4\alpha_G\alpha_\Phi/(c_L\, c_{G\Phi})^2$. □

Remark 6.3.6. It is easy to check that (6.3.44) holds easily whenever

$$L(X_h, \Phi_h(Y_h)) = L(X_h, Y_h), \tag{6.3.50}$$

implying that $L(X_h, Y_h)$ depends only on the part of Y_h which is left unchanged by Φ_h. □

We finally need a further assumption that connects the right-hand side of (6.3.47) with the norm in $\mathcal{W}_h$.

H.3 *With the notation of assumptions* **H.1** *and* **H.2***, we further assume that there exist two positive constants γ_2 and γ_3 such that*

$$[Y_h]_h^2 + \gamma_2\|G_h Y_h\|_{\mathcal{H}}^2 \ge \gamma_3 \|Y_h\|_{\mathcal{W}}^2 \qquad \forall\, Y_h \in \mathcal{W}_h. \tag{6.3.51}$$

□

It is clear that, if Assumption **H.3** is also verified, then (6.3.47) will give a stability result of type (6.2.13), where the explicit value of the constant α_Φ can be easily deduced from the values of the other constants. On the other hand, the estimate (6.3.47) will also be used in the sequel in cases when some constant

(r mostly, and sometimes γ_2) might depend on h, so that it is convenient to leave it in its present form.

We now consider the problem of error estimates. As we introduced sufficient conditions to ensure stability, the question will be to check consistency, and in particular the effect on consistency of the extra stabilising terms. As we said, in many applications, the constants r, γ_2, and γ_3 appearing in (6.3.51) might be allowed to depend on h. Hence, it is important that we keep track of them in our abstract estimates. Mimicking (6.2.20), we now set

$$\mathcal{R}(X_I) := \sup_{Y_h \in \mathcal{W}_h} \frac{L(G_h X_I - N, G_h Y_h)}{\|G_h Y_h\|_{\mathcal{H}}}. \tag{6.3.52}$$

Theorem 6.3.1. *Let X and X_h be the solutions of* (6.2.16) *and (6.3.43) respectively. Assume that* **H.1**, **H.2**, *and* **H.3** *hold. Then, for every $X_I \in \mathcal{W}_h$, we have*

$$\begin{aligned}
&[X_I - X_h]_h^2 + r\|G_h(X_I - X_h)\|_{\mathcal{H}}^2 \\
&\qquad \le C\left(\frac{r+\gamma_2}{r\gamma_3}\|\mathcal{A}\|^2\|X_I - X\|^2 + r(\mathcal{R}(X_I))^2\right),
\end{aligned} \tag{6.3.53}$$

where the constant C depends on α_G, α_Φ, c_L and c_Φ, but does not depend on the other parameters.

Proof. We set $\delta X := X_I - X_h$ and $Y_h := \Phi_h^*(\delta X)$, with Φ_h^* given in Lemma 6.3.1. Using Lemma 6.3.1, then the continuous equation (6.2.16) and the stabilised discrete one (6.3.43), and then (6.3.40), (6.3.45), and (6.3.46), we get

$$\begin{aligned}
\alpha_\Phi[\delta X]_h^2 + r\alpha_G\|G_h\delta X\|^2 &\le \langle \mathcal{A}(\delta X), \Phi_h^*(\delta X)\rangle + rL(G_h(\delta X), G_h(\Phi_h^*(\delta X))) \\
&= \langle \mathcal{A}(X_I - X), \Phi_h^*(\delta X)\rangle + rL(G_h X_I - N, G_h(\Phi_h^*(\delta X))) \\
&\le c_\Phi^*\|\mathcal{A}\|\,\|X_I - X\|\,\|\delta X\| + r\mathcal{R}(X_I)\|G_h\delta X\|.
\end{aligned} \tag{6.3.54}$$

We now use **H.3** to bound $\|\delta X\|$:

$$\|\delta X\| \le \left(\frac{[\delta X]_h^2 + \gamma_2\|G_h\delta X\|^2}{\gamma_3}\right)^{1/2} \le \frac{[\delta X]_h + \gamma_2^{1/2}\|G_h\delta X\|}{\gamma_3^{1/2}}. \tag{6.3.55}$$

At this point, we need, just for a while, a lighter notation. We denote one of the two terms on the right-hand side of (6.3.53) by $D_1 := \|\mathcal{A}\|\,\|X_I - X\|$ and the other by $D_2 := \|G_h(X_I)\|_{\mathcal{H}}$. We also denote the second term in the left-hand side by $g := \|G_h(\delta X)\|_{\mathcal{H}}$. With this notation, the inequality that we have to prove becomes

$$[\delta X]_h^2 + rg^2 \le C\Big(\frac{r+\gamma_2}{r\gamma_3} D_1^2 + rD_2^2\Big) \tag{6.3.56}$$

and what we have got, inserting (6.3.55) into (6.3.54) and using the new notation, can be written as

$$[\delta X]_h^2 + rg^2 \le C\Big(\frac{D_1}{\gamma_3{}^{1/2}}[\delta X]_h + D_1\Big(\frac{\gamma_2}{r\gamma_3}\Big)^{1/2} r^{1/2}g + r^{1/2}D_2 r^{1/2}g\Big). \tag{6.3.57}$$

Then, we apply the inequality $ab \le \frac{c}{2}a^2 + \frac{1}{2c}b^2$ with suitable choices of c, move three terms to the left and multiply the resulting equation by a suitable constant to get (6.3.53). □

As an immediate consequence of Theorem 6.3.1, we have the following error estimate.

Theorem 6.3.2. *Let X and X_h be the solutions of* (6.2.16) *and (6.3.43) respectively. Assume that* **H.1**, **H.2** *and* **H.3** *hold, and assume that the operator G_h could be extended to a space $\mathcal{W}(h) \subseteq \mathcal{W}$ containing both $\mathcal{W}_h$ and X. Then, there exists a constant $C = C(\alpha_G, \alpha_\Phi, c_L, c_\Phi)$ such that*

$$[X - X_h]_h^2 + r\|G_h(X - X_h)\|_{\mathcal{H}}^2 \le C \inf_{X_I \in \mathcal{W}_h} \Big(\frac{r+\gamma_2}{r\gamma_3}\|\mathcal{A}\|^2 \|X - X_I\|^2 + r(\mathcal{R}(X_I))^2\Big). \tag{6.3.58}$$

Moreover, we have the following important corollary.

Corollary 6.3.1. *Keep the same assumptions as in Theorem 6.3.2, and assume moreover that for every h we have a space $\mathcal{W}(h)$ containing both $\mathcal{W}_h$ and the exact solution X of* (6.2.16) *such that G_h could be extended to an operator in $\mathcal{L}(\mathcal{W}(h), \mathcal{H})$, with norm c_G uniformly bounded in h, and such that* (6.3.40) *and* (6.3.41) *still hold for Y_h and Z_h in $\mathcal{W}(h)$. Finally, assume that, for the exact solution X of* (6.2.16)*, we have*

$$L(G_h X - N, G_h Y_h) = 0 \quad \forall Y_h \in \mathcal{W}_h, \tag{6.3.59}$$

so that, with the notation of (6.3.52)*, we have $\mathcal{R}(X) = 0$. Then, there exists a constant $C = C(\alpha_G, \alpha_\Phi, c_L, c_\Phi)$ such that*

$$[X - X_h]_h^2 + r\|G_h(X - X_h)\|_{\mathcal{H}}^2 \le C \inf_{X_I \in \mathcal{W}_h} \Big(\frac{r+\gamma_2}{r\gamma_3}\|\mathcal{A}\|^2 \|X - X_I\|^2 + r\|G_h(X - X_I)\|_{\mathcal{H}}^2\Big). \tag{6.3.60}$$

Remark 6.3.7. It is not difficult to see that the approach described here, and in particular the result of Corollary 6.3.1, has much in common with the ones described and analysed in the previous section. Actually, many stabilising techniques available on the market can be set equally well in the framework of the previous section as in that of the present one. Still, in certain cases, one of the two would be easier to use, and in other cases, only one of these two approaches will be usable. □

As we mentioned already, in different applications, we might allow some of the constants (and mainly r and γ_2) to depend on h. Hereafter, we shall rapidly see some typical cases, taking as simplest example the one-dimensional version of mixed formulations for the Poisson problem, already seen in (6.1.82) and (6.1.83). Other examples will be seen in the following chapters, for different applications.

The first, and main difference, is whether the constant γ_2 can be assumed to tend to zero (and fast enough) when h tends to zero. We first consider the case in which it is more convenient to take a γ_2 that does not depend on h. In a certain number of cases, Theorem 6.3.2 or Corollary 6.3.1 can be applied with all the constants (r, γ_2, and γ_3) independent of h.

Example 6.3.3 (Both r and γ_2 are independent of h). It is clear that Corollary 6.3.1 is the natural candidate to be applied in these cases. Most augmented formulations and their variants can be analysed in this way. Just to see an example, consider the one-dimensional Poisson problem (6.1.82), and assume that we take $V_h := \mathcal{L}^2_1$ and $Q_h := \mathcal{L}^0_0$. We have already seen in the previous Chapter (in Sect. 5.2.4) that this choice leads to a total disaster, due to the failure of the *elker* condition. However, adding a term

$$R(X_h, Y_h) \equiv R((u_h, p_h), (v_h, q_h)) = (u'_h + f, v'_h) \tag{6.3.61}$$

will restore the full ellipticity and give a good solution. On the other hand, the bound (6.3.12) tells us, in this case, that the projection of $Bu_h \equiv u'_h$ onto Q_h is already under control, and a further analysis would show that indeed the term in (6.3.61) could be multiplied by h^2 and still provide a sufficient stabilisation (see [125]). □

Example 6.3.4 (Taking γ_2 fixed and r depending on h). In this case, we are allowed to use directly Theorem 6.3.1. The bound (6.3.53) will provide (for r "small") an estimate of the type

$$[\delta X]_h^2 + r\|G_h(\delta X)\|^2_{\mathcal{H}} \le C\left(\frac{1}{r}\|X_I - X\|^2 + r\|G_h(X_I)\|^2_{\mathcal{H}}\right), \tag{6.3.62}$$

which, when (6.3.4) holds, can become

$$\begin{aligned}[\delta X]_h^2 + r\|G_h\delta X\|^2 &\le C\left(\frac{1}{r}\|X_I - X\|^2 + r\|\overline{X}_h - X\|^2\right)\\ &\le C\left(\frac{1}{r}h^{s_1} + rh^{s_2}\right),\end{aligned} \tag{6.3.63}$$

by usual interpolation estimates with, in general, $s_1 \geq s_2 \geq 0$. Then, by taking $r = h^s$, we get

$$[\delta X]_h^2 + h^s \|G_h \delta X\|^2 \leq C \left(h^{s_1 - s} + h^{s_2 + s}\right), \tag{6.3.64}$$

with the optimal choice given by $s = (s_1 - s_2)/2$. We further develop such a case in the following example. □

Example 6.3.5 (Penalty methods for the inf-sup condition). Coming back to the case of Proposition 6.3.1, a large class of stabilisations to cure methods where the *inf-sup* condition fails can be built taking a subspace $\tilde{Q}_h$ of Q_h, denoting by $\tilde{P}$ the projection operator on $\tilde{Q}_h$, and setting

$$\begin{cases} G_h((v_h, q_h)) := \tilde{P}(q_h), \\ R((u_h, p_h), (v_h, q_h)) := (\tilde{P} p_h, \tilde{P} q_h)_Q. \end{cases} \tag{6.3.65}$$

The subspace $\tilde{Q}_h$ will be chosen so that **H.3** holds. This means that, using the notation (6.3.19), we should have

$$[\![q_h]\!]_h^2 + \gamma_2 \|P_{\tilde{Q}_h} q_h\|^2 \geq \gamma_3 \|q_h\|_Q^2, \tag{6.3.66}$$

for some positive constants γ_2 and γ_3. The stabilised problem then becomes

$$\begin{cases} a(u_h, v_h) + b(v_h, p_h) = (f, v_h) & \forall\, v_h \in V_h, \\ b(u_h, q_h) - r(\tilde{P} p_h, \tilde{P} q_h) = (g, q_h) & \forall\, q_h \in Q_h. \end{cases} \tag{6.3.67}$$

In this case, the estimate (6.3.53) would yield

$$\begin{aligned} \|u_I - u_h\|_V^2 &+ [\![p_I - p_h]\!]_h^2 + r\|\tilde{P}(p_I - p_h)\|_Q^2 \\ &\leq C \frac{r + \gamma_2}{r} (\|A\|^2 \|u_I - u\|_V^2 + \|p_I - p\|_Q^2) \\ &\qquad + r\|\tilde{P}(p_I - p)\|_Q^2 + r\|\tilde{P}(p)\|_Q^2. \end{aligned} \tag{6.3.68}$$

We now consider three cases:

(i) **Stable penalty**. The *inf-sup* condition is satisfied. In this case, we obviously have $[\![q_h]\!]_h \simeq \|q_h\|_Q$. We can then take $\gamma_2 = 0$ and $\tilde{Q}_h = Q_h$, which means that $\tilde{P} = I$. We can take r as small as we want and there is an $O(r)$ term in the right-hand side. This means that the penalty is used as a computational trick to obtain an otherwise good solution.

(ii) **Brute force penalty**. We have no (usable) *inf-sup* condition (roughly speaking: $[\![q_h]\!]_h = 0$ for any q_h). We take again $\tilde{Q}_h = Q_h$ and $\tilde{P} = I$ but we now need $\gamma_2 = O(1)$. The error estimate becomes,

$$\|u_I - u_h\|_V^2 + r\|p_I - p_h\|_Q^2$$
$$\leq C\frac{1}{r}(\|A\|^2\|u_I - u\|_V^2 + \|p_I - p\|_Q^2) + r\|p\|_Q^2. \quad (6.3.69)$$

In that case, we only get a bound on $\|u_I - u_h\|$. Suppose, to fix ideas, that we would expect an $O(h^k)$ from the usual error estimates. It is clear that the best that we can do is to take $r = O(h^k)$ and get a final bound in $O(h^{k/2})$ instead of $O(h^k)$.

(iii) **Clever-penalty**. Let us suppose that there is a subspace $\overline{Q}_h \subset Q_h$ such that $V_h \times \overline{Q}_h$ satisfies the *inf-sup* condition and let $\overline{P}$ be the projection onto $\overline{Q}_h$. We set $\tilde{P} = (I - \overline{P})$ and we now need $\gamma_2 = O(1)$. The bound would now be

$$\|u_I - u_h\|_V^2 + \|\bar{p}_I - \overline{p}_h\|^2 + r\|(I - \overline{P})(p_I - p_h)\|_Q^2$$
$$\leq C\frac{1}{r}(\|A\|^2\|u_I - u\|_V^2 + \|p_I - p\|_Q^2) + r\|(I - \overline{P})p\|_Q^2. \quad (6.3.70)$$

Two possibilities arise, depending on the approximation properties in $\overline{Q}_h$. If all the terms $\|u_I - u\|_V$, $\|p_I - p\|_Q$, and $\|(I - \overline{P})p\|_Q$ in the right-hand side have a similar order in h, then any positive r will provide an estimate of the best possible order. Such stabilising methods have been considered in [349]. On the other hand, suppose that the last term is of lower order than the other ones. One could use a small value of r to get a better accuracy but to the expense of loosing on the first terms. If, for instance, we expect

$$(\|u_I - u\|_V^2 + \|p_I - p\|_Q^2) = O(h^4) \quad \text{and} \quad \|(I - \overline{P})(p)\|_Q^2 = O(h^2),$$

then the choice $r = O(h)$ will yield an estimate of $O(h^{3/2})$ on $\|u - u_h\|_V$ and of $O(h)$ on $\|p - p_h\|_Q$. Such a procedure was introduced by Lovadina and Auricchio [284] for the Stokes problem. □

Example 6.3.6 (Using a γ_2 that depends on h). We now consider the cases in which it is possible, and convenient, to use a γ_2 that tends to zero with h. At first sight, one might think that this *never* (or almost never) occurs. However, this is not true. Assume, for instance, that

$$[q_h]_h \geq \|\overline{q}_h\|_{L^2}, \quad (6.3.71)$$

where $\overline{q}_h$ is the L^2 projection of q_h onto the space of piecewise constant functions. It is elementary, by the Poincaré inequality, to see that, for a q_h piecewise in H^1, we have

$$\|q_h\|_{L^2}^2 \leq \|\overline{q}_h\|_{L^2}^2 + C\,h^2\|\,\underline{\mathrm{grad}}_h\,q_h\|_{L^2}^2, \quad (6.3.72)$$

where, as before, $\underline{\mathrm{grad}}_h$ is the piecewise gradient and C is a constant that depends only on the minimum angle of the decomposition (and where, for simplicity, we assumed a quasi-uniform decomposition). Hence, in this case, we would have

$$[q_h]_h^2 + h^2 \| \underline{\mathrm{grad}}_h\, q_h \|_{L^2}^2 \geq \gamma_3 \|q_h\|_{L^2}^2 \tag{6.3.73}$$

with a constant γ_3 independent of h: a formula of the type of (6.3.51) with $\gamma_2 = h^2$. Looking at (6.3.53), it seems natural, when γ_2 is small, to take r as small as γ_2 (that is, going to zero with the same order). It is what we consider in the next example. □

Example 6.3.7 (Both r and γ_2 depend on h). We consider the case in which both r and γ_2 depend on h, and go to zero. For simplicity, we assume that we take, brutally, $r \geq \gamma_2$, but everything will work just taking, say, $r \geq \kappa\gamma_2$ with a constant κ independent of h. Then, we have

$$[\delta X]_h^2 + r\|G_h(\delta X)\|_{\mathcal{H}}^2 \geq [\delta X]_h^2 + \gamma_2\|G_h(\delta X)\|_{\mathcal{H}}^2 \geq \gamma_3\|\delta X\|^2 \tag{6.3.74}$$

so that applying (6.3.53) to the left-hand side of (6.3.74) gives

$$\gamma_3 \|\delta X\|^2 \leq C\left(\frac{2}{\gamma_3}\|\mathcal{A}\|^2\|X_I - X\|^2 + r\|G_h(X_I)\|_{\mathcal{H}}^2\right). \tag{6.3.75}$$

For instance, dealing with the one-dimensional version of (6.1.82) and starting from $V_h := \mathcal{L}_1^1$ and $Q_h := \mathcal{L}_2^1$, it is easy to see that $[(v_h, q_h)]_h^2 \geq \|v_h\|_1^2 + \|\overline{q}_h\|_0^2$, where again $\overline{q}_h$ is the projection of q_h on piecewise constants. In view of (6.3.72), we can then take

$$R((u,p),(v,q)) = (p', q') \tag{6.3.76}$$

and $r = \gamma_2 = h^2$. This will give linear convergence for both u and p. □

6.3.1 Another Form of Minimal Stabilisation

We now develop a more sophisticated variant of the previous case (where γ_2 depends on h) that is based, instead of (6.3.71), on a (possible) estimate of the type

$$[q_h]_h \geq \|q_h\|_{L^2} - C\, h^2 \| \underline{\mathrm{grad}}\, q_h \|_{L^2}^2. \tag{6.3.77}$$

Estimates of this type are met in situations like the one analysed in Sect. 5.4.5 (see, in particular Eq. (5.4.22)) and related to the technique known as *Verfürth's trick* [375] that we discussed in the previous chapter. We still suppose that Assumption **A.1** holds and we complete it by the following assumption.

A.2 There exists a Hilbert space H with $V \hookrightarrow H \equiv H' \hookrightarrow V'$ such that

$$B^t(Q_h) \subset H \tag{6.3.78}$$

(where $B^t : Q \longrightarrow V'$ is, as usual, the linear operator associated with the bilinear form $b(v, q)$), and there exist a monotone function $\omega : \mathbb{R}^+ \to \mathbb{R}^+$ and a positive constant σ, independent of h, such that

$$\omega(h)\|v_h\|_V \leq \|v_h\|_H \qquad \forall\, v_h \in V_h, \tag{6.3.79}$$

$$\omega(h)\|B^t q_h\|_H \leq \|q_h\|_Q \quad \forall q_h \in Q_h \tag{6.3.80}$$

and

$$\|v - \pi_{V_h} v\|_H \leq \sigma\omega(h)\|v\|_V \quad \forall\, v \in V. \tag{6.3.81}$$

□

We note that, setting

$$[\![q_h]\!]_h := \sup_{v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_V} = \sup_{v_h \in V_h} \frac{(v_h, B^t q_h)_H}{\|v_h\|_V}, \tag{6.3.82}$$

from (6.3.79) we have

$$[\![q_h]\!]_h = \sup_{v_h \in V_h} \frac{(v_h, B^t q_h)_H}{\|v_h\|_H} \frac{\|v_h\|_H}{\|v_h\|_V} \geq \omega(h)\,\|\pi_{V_h'} B^t q_h\|_H \equiv \omega(h)\|B_h^t q_h\|_H. \tag{6.3.83}$$

In agreement with the general procedure of this section, we can now take $\mathcal{H} = H$ with

$$G_h((v_h, q_h)) = B^t q_h - B_h^t q_h = (I - \pi_{V_h'})B^t q_h \tag{6.3.84}$$

and define:

$$R((u_h, p_h), (v_h, q_h)) = \left(B^t p_h - B_h^t p_h, B^t q_h - B_h^t q_h\right)_H. \tag{6.3.85}$$

It is clear that both (6.3.40) and (6.3.41) will hold with constants independent of h, so that **H.2** holds. We are left with **H.3** which will be proved in the next two propositions using essentially the so-called Verfürth's trick [375] that we already discussed in Sect. 5.4.5.

Lemma 6.3.2. *Assume that* **A.1** *and* **A.2** *hold. Then,*

$$[\![q_h]\!]_h := \sup_{v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_V} \geq \beta\|q\|_Q - \sigma\omega(h)\|B^t q_h\|_H \quad \forall\, q_h \in Q_h, \tag{6.3.86}$$

where β is the inf-sup constant appearing in (6.3.16)*, $\omega(h)$ is given in* (6.3.79)*–*(6.3.81)*, and σ is given in* (6.3.81)*.*

Proof. The proof is essentially the same as the one that was used to prove (5.4.22) in the last chapter. Let us see it briefly. We start from the *inf-sup* condition (6.3.16), we add and subtract the projection $\pi_{V_h} v$ of v over V_h and then use (6.3.82) and (6.3.81):

$$\begin{aligned}
\beta \|q_h\|_Q &\le \sup_{v\in V} \frac{b(v,q_h)}{\|v\|_V} = \sup_{v\in V}\left(\frac{b(\pi_{V_h} v, q_h)}{\|v\|_V} + \frac{b(v-\pi_{V_h}, q_h)}{\|v\|_V}\right)\\
&\le \sup_{v\in V} \frac{b(\pi_{V_h} v, q_h)}{\|\pi_{V_h} v\|_V} + \sup_{v\in V} \frac{(v-\pi_{V_h} v, B^t q_h)_H}{\|v\|_V}\\
&\le \sup_{v_h\in V_h} \frac{b(v_h, q_h)}{\|v_h\|_V} + \sup_{v\in V} \frac{\|v-\pi_{V_h} v\|_H \|B^t(q_h)\|_H}{\|v\|_V}\\
&\le [\![q_h]\!]_h + \sigma\omega(h)\|B^t q_h\|_H \quad \forall\, q_h \in Q_h.
\end{aligned} \tag{6.3.87}$$

□

We can now easily get the following result.

Lemma 6.3.3. *Under Assumptions* **A.1** *and* **A.2***, there exists a constant $\tilde\beta$, independent of h, such that*

$$[\![q_h]\!]_h^2 + \omega^2(h)\|B^t q_h - \pi_{V_h} B^t q_h\|_H^2 \ge \tilde\beta \|q_h\|_Q^2 \quad \forall\, q_h \in Q_h. \tag{6.3.88}$$

Proof. Indeed, by the triangle inequality, we have, for every $q_h \in Q_h$:

$$\|B^t q_h - \pi_{V_h'} B^t q_h\|_H + \|\pi_{V_h'} B^t q_h\|_H \ge \|B^t q_h\|_H. \tag{6.3.89}$$

On the other hand, summing (6.3.86) plus σ times (6.3.83), we have

$$(1+\sigma)[\![q_h]\!]_h \ge \beta\|q_h\|_Q + \sigma\omega(h)\Big(\|\pi_{V_h'} B^t q_h\|_H - \|B^t q_h\|_H\Big) \tag{6.3.90}$$

so that

$$(1+\sigma)[\![q_h]\!]_h + \sigma\omega(h)\|B^t q_h - \pi_{V_h'} B^t q_h\|_H \ge \beta\|q_h\|_Q \tag{6.3.91}$$

and the result follows easily. □

Remark 6.3.8. Actually, by Pythagora's theorem, we obviously have, for every $q_h \in Q_h$:

$$\|B^t q_h\|_H^2 = \|B^t q_h - \pi_{V_h'} B^t q_h\|_H^2 + \|\pi_{V_h'} B^t q_h\|_H^2. \tag{6.3.92}$$

However, as we have seen, the triangle inequality (6.3.89) is enough for our proof. □

Lemma 6.3.3 implies that **H.3** holds, with the above choices for $[\,\cdot\,]_h$ and G_h, with a constant γ_3 independent of h, and with $\gamma_2 = \omega^2(h)$.

Remark 6.3.9. In certain cases, it would be more convenient to introduce another finite element space $\tilde{V}_h \subseteq H$, and use a stabilising term like

$$R((u_h, p_h), (v_h, q_h)) = \left(B^t p_h - \pi_{\tilde{V}_h'} B^t p_h, B^t q_h - \pi_{\tilde{V}_h'} B^t q_h\right)_H \tag{6.3.93}$$

instead of (6.3.85). Then, Lemma 6.3.3 will still hold (and consequently **H.3** will also hold), provided that we have some additional result that guarantees, in the particular case under study, that

$$\|B^t q_h\|_H \geq C(\|B^t q_h - \pi_{\tilde{V}_h'} B^t q_h\|_H + \|\pi_{V_h'} B^t q_h\|_H) \quad \forall q_h \in Q_h \tag{6.3.94}$$

for some constant C independent of h. □

In view of the above remark, it might be convenient to treat the two cases (that is: using (6.3.85) or (6.3.93) when (6.3.94) also holds) together. For this, we introduce the following assumption.

A.3 *With the notation of Assumption* **A.2***, we consider a space* $\tilde{V}_h \subseteq H$ *and we assume that there exists a positive constant* $\tilde{\kappa}$*, independent of* h*, such that*

$$\|\pi_{V_h} B^t q_h\|_H + \|B^t q_h - \pi_{\tilde{V}_h} B^t q_h\|_H \geq \tilde{\kappa}\, \|B^t q_h\|_H \quad \forall\, q_h \in Q_h. \tag{6.3.95}$$

□

Assumption **A.3** obviously holds, for instance, if $\tilde{V}_h = \{0\}$, or more generally whenever $\tilde{V}_h \subseteq V_h$. The case of a $\tilde{V}_h$ larger than V_h, instead, will work only in some special case, and will require an *ad hoc* (and sometimes delicate) proof. We can collect the result of Lemma 6.3.3 and the above discussion in the following theorem.

Theorem 6.3.3. *Assume that Assumptions* **A.1***,* **A.2***, and* **A.3** *hold. Assume moreover that the full ellipticity condition* (6.3.17) *holds. Assume that we are given subspaces* $V_h \subset V$ *and* $Q_h \subset Q$*, and we take* $\mathcal{W}_h := V_h \times Q_h$ *with* (6.3.18) *and* (6.3.19)*. Set*

$$G_h((v_h, q_h)) := B^t q_h - \pi_{\tilde{V}_h} B^t q_h. \tag{6.3.96}$$

Then, **H.3** *holds with a constant* γ_3 *independent of* h*, and with* $\gamma_2 = \omega^2(h)$*.*

We are therefore in a situation very similar to that of Example 6.3.6. A very reasonable choice would then be to use an r that also behaves as $\omega(h)^2$ as in Example 6.3.7. Then using Theorem 6.3.1 as in (6.3.75), we have the following theorem.

Theorem 6.3.4. *Assume that* **A.1***,* **A.2***, and* **A.3** *hold, and let* (u, p) *be the solution of Problem* (6.1.5)*. Assume that, in (6.2.17),* R *is defined through (6.3.85), and that* r *is a positive number* $\geq \omega(h)^2$*. Then, Problem (6.2.17) has a unique solution*

(u_h, p_h) *and there exists a constant* C, *independent of* h *and* r, *such that, for every* $(u_I, p_I) \in V_h \times Q_h$, *we have:*

$$\|u_I - u_h\|_V^2 + \|p_I - p_h\|_Q^2 \\ \leq C \Big(\|u - u_I\|_V^2 + \|p - p_I\|_Q^2 + r\|G_h(0, p_I)\|^2 \Big). \tag{6.3.97}$$

Proof. The proof is an immediate consequence of (6.3.53) as in (6.3.74) and (6.3.75). □

Remark 6.3.10. The convenience in using $r > \omega^2(h)$ (including the case of an r fixed that does not depend on h) can occur when the term $\|G_h(0, p_I)\|_{\mathcal{H}}$ in (6.3.97) is already small. Indeed, with a choice of the type (6.3.93), we would have

$$\begin{aligned} \|G_h(0, p_I)\|_{\mathcal{H}} &= \|B^t p_I - \pi_{\tilde{V}_h} B^t p_I\|_{\mathcal{H}} \\ &\leq \|B^t(p_I - p) - \pi_{\tilde{V}_h} B^t(p_I - p)\|_{\mathcal{H}} + \|B^t p - \pi_{\tilde{V}_h} B^t p\| \\ &\leq C\|B^t(p - p_I)\|_{\mathcal{H}} + \|(I - \pi_{Vth})B^t p\|_{\mathcal{H}} \end{aligned} \tag{6.3.98}$$

that could be small whenever $B^t p$ is smooth and $\tilde{V}_h$ has good approximation properties. □

Remark 6.3.11. It is clear that the condition $r \geq \omega^2(h)$ could be replaced with $r \geq \kappa\omega^2(h)$ for some $\kappa > 0$ independent of h. This, indeed, will make the notation in the proof heavier, but the final result will end up in a different value of the constant C in (6.3.97). □

Remark 6.3.12. As usual, we can then take u_I and p_I as the best approximations of u and p, respectively (in the respective norms), and then deduce an estimate for $\|u - u_h\|_V + \|p - p_I\|_Q$ by the triangle inequality. □

We now consider the case of an r *smaller* that $\omega^2(h)$.

Theorem 6.3.5. *In the same assumptions as in Theorem 6.3.4, taking* $r \leq \omega^2(h)$, *we have:*

$$\|u_I - u_h\|_V^2 + [\![p_I - p_h]\!]^2 + r\|p_I - p_h\|_Q^2 \\ \leq C\Big(\frac{2\omega^2(h)}{r}(\|u - u_I\|_V^2 + \|p - p_I\|_Q^2) + r\|G_h(0, p_I)\|^2\Big). \tag{6.3.99}$$

Proof. The proof is again an easy consequence of (6.3.53) in Theorem 6.3.1.

Remark 6.3.13. In applications, the choice of the form of $r(h)$ will be done in order to get the best possible estimate. In particular, the choice $r = \kappa\omega(h)^2$ will be the best choice when $\tilde{V}_h = 0$ and first order approximations are employed.

This situation will be met for instance in the Brezzi-Pitkäranta stabilisation for the Stokes problem considered in Chap. 8, that however could be treated by the simpler estimates of Example 6.3.7. □

6.4 Enhanced Strain Methods

The so-called "enhanced strain" methods have become popular as a stabilising device. We shall try to give a feeling of how they work and the way they can be applied to mixed methods. We shall however first consider a more classical setting. Let us thus consider two Hilbert spaces V and Q. To simplify the notation, we assume, from the very beginning, that Q is identified with its dual space, that is $Q \equiv Q'$. We then assume that we have a continuous operator B from V to $Q = Q'$ and a continuous isomorphism $\mathfrak{C}$ from Q onto Q. We want to solve a variational problem of the form,

$$\inf_{v\in V} \frac{1}{2}(\mathfrak{C}Bv, Bv)_Q - \langle f, v\rangle_{V'\times V}. \tag{6.4.1}$$

As usual, we consider an analogous problem in subspaces V_h and Q_h where we make the assumption that $B(V_h) \subseteq Q_h$. We want to solve

$$\inf_{v_h\in V_h} \frac{1}{2}(\mathfrak{C}Bv_h, Bv_h)_Q - \langle f, v_h\rangle_{V'\times V}. \tag{6.4.2}$$

In some cases, for instance in an almost incompressible elasticity problem, the numerical solution may behave badly: a locking phenomenon can occur when problem (6.4.2) is too stiff. To build an enhanced method, we introduce a new space $E_h \subset Q$ and we change the problem into

$$\inf_{v_h\in V, \eta\in E_h} \frac{1}{2}(\mathfrak{C}(Bv_h + \eta), (Bv_h + \eta))_Q - \langle f, v_h\rangle_{V'\times V}, \tag{6.4.3}$$

where $\eta \in E_h$ is some "enhancement" of Bv_h. In terms of mathematical programming, this would be called a "slack variable". The optimality conditions of (6.4.3) are

$$\begin{aligned} &(\mathfrak{C}(Bu_h + \eta), Bv_h)_Q - \langle f, v_h\rangle_{V'\times V} = 0, \quad \forall\, v_h \in V_h,\\ &(\mathfrak{C}(Bu_h + \eta), \delta)_Q = 0, \quad \forall\, \delta \in E_h. \end{aligned} \tag{6.4.4}$$

Assuming, for simplicity, that $\mathfrak{C}(E_h) \subseteq E_h$ (as it is almost always the case in practice), and denoting by P_E the projection on E_h, the last equation of (6.4.4) can be read:

$$\mathfrak{C}\eta = -P_E\mathfrak{C}Bu_h \tag{6.4.5}$$

and taking this expression into the first equation, we obtain

$$(\mathfrak{C}(I - P_E)Bu_h, Bv_h)_Q - \langle f, v_h\rangle_{V'\times V} = 0, \quad \forall\, v_h \in V_h, \tag{6.4.6}$$

which is clearly a weaker formulation of the original problem. This idea has been used to obtain stable formulations for a variety of problems such as nearly incompressible elasticity or simulation of very thin structures.

We shall now see briefly, following [284], how this idea can be extended to the case of a mixed formulation (that we repeat once more for the sake of convenience)

$$\begin{cases} a(u, v) + b(v, p) = (f, v)_V & \forall\, v \in V, \\ b(u, q) = (g, q)_Q & \forall\, q \in Q, \end{cases} \tag{6.4.7}$$

assuming again that Q is identified with its dual space.
We shall not try to cover all cases: to avoid unnecessary technicalities, we shall concentrate on the *inf-sup* condition. We shall then suppose that the bilinear form $a(u, v)$ can be decomposed as

$$a(u, v) = a_D(u, v) + \chi(Bu, Bv), \tag{6.4.8}$$

where $a_D(u, v)$ is coercive on the kernel of B so that, according to Proposition 4.3.4, $a(u, v)$ is coercive on the whole space V if $\chi > 0$. We write explicitly:

$$\begin{cases} a_D(u_h, v_h) + \chi(Bu_h, Bv_h) + (Bv_h, p_h) = (f, v_h) & \forall\, v_h \in V_h, \\ (Bu_h, q_h) = (g, q_h) & \forall\, q_h \in Q_h. \end{cases} \tag{6.4.9}$$

Following the idea of enhanced methods, we introduce a subspace E_h of Q and we change the problem into

$$\begin{cases} a_D(u_h, v_h) + \chi(Bu_h + \eta, Bv_h) + (Bv_h, p_h) = (f, v_h) & \forall\, v_h \in V_h, \\ \chi(Bu_h + \eta, \delta) + (\delta, p_h) = 0 & \forall\, \delta \in E_h, \\ (Bu_h + \eta, q_h) = (g, q_h) & \forall\, q_h \in Q_h. \end{cases} \tag{6.4.10}$$

The second equation of (6.4.10) can be read as

$$\eta = -P_E Bu_h - (1/\chi) P_E p_h. \tag{6.4.11}$$

Bringing (6.4.11) into the first and the last equation of (6.4.10), we get:

$$\begin{cases} a_D(u_h, v_h) + \chi((I - P_E)Bu_h, Bv_h) + b(v_h, (I - P_E)p_h) = (f, v_h) & \forall\, v_h \in V_h, \\ b(u_h, (I - P_E)q_h) - (1/\chi)(P_E p_h, P_E q_h) = (g, q_h) & \forall\, q_h \in Q_h, \end{cases} \tag{6.4.12}$$

where in the second equation we used the fact that $((I - P_E)Bu_h, q_h)$ is equal to $b(u_h, (I - P_E)q_h)$.

We can now consider two interesting special cases. In the first one, we choose E_h so that $P_E B v_h = 0$ for any v_h. Equations (6.4.12) now simplify to

$$\begin{cases} a_D(u_h, v_h) + \chi(Bu_h, Bv_h) + b(v_h, p_h) = (f, v_h) \; \forall\, v_h \in V_h, \\ b(u_h, q_h) - (1/\chi)(P_E p_h, P_E q_h) = (g, q_h) \qquad \forall\, q_h \in Q_h. \end{cases} \tag{6.4.13}$$

This is clearly like using a penalty method to stabilise p_h, as we have seen already in Remark 4.3.7. We have already seen these methods in Example 6.3.5, and we shall discuss their applications at various occasions in the following chapters, and in particular in Chap. 8.

It is also interesting to give a look at the case where we choose as E_h a subspace of Q_h. Let us denote by $\overline{Q}_h$ the orthogonal complement of E_h and by $\overline{P} \equiv I - P_E$ the projection onto $\overline{Q}_h$. We can now write (6.4.12) as

$$\begin{cases} a_D(u_h, v_h) + \chi(\overline{P}Bu_h, \overline{P}Bv_h) + b(v_h, \overline{P}p_h) = (f, v_h) \; \forall\, v_h \in V_h, \\ b(u_h, \overline{P}q_h) - (1/\chi)(P_E p_h, P_E q_h) = (g, q_h) \qquad \forall\, q_h \in Q_h. \end{cases} \tag{6.4.14}$$

Writing the second equation for $q_h = \overline{q}_h \in \overline{Q}_h$ and then for $q_h = \delta \in E_h$, we have

$$b(u_h, \overline{q}_h) = (g, \overline{q}_h) \quad \forall\, \overline{q}_h \in \overline{Q}_h \tag{6.4.15}$$

plus

$$c P_E p_h = P_E g. \tag{6.4.16}$$

Thus, we have simply written a discrete problem in $V_h \times \overline{Q}_h$:

$$\begin{cases} a_D(u_h, v_h) + \chi(\overline{B}u_h, \overline{B}v_h) + b(v_h, \overline{p}_h) = (f, v_h) \; \forall\, v_h \in V_h, \\ b(u_h, \overline{q}_h) = (g, \overline{q}_h) \quad \forall\, \overline{q}_h \in \overline{Q}_h, \end{cases} \tag{6.4.17}$$

with $\overline{B} = \overline{P}(B)$, and then corrected $p_h = \overline{p}_h + (1/\chi)P_E g$.

Is that all? By no means! *There are more things in Heaven and Earth, Horatio, than are dreamt of in your philosophy.*
And among all those things, "stabilisation methods" hold a non negligible place.

6.5 Eigenvalue Problems

We shall consider in this section a general setting for the approximation of eigenvalue problems associated with the mixed problems introduced in Sect. 5.1. To make the presentation clearer, we recall some basic assumptions. We thus have two Hilbert spaces, V and Q. Moreover, $a(v, v)$ and $b(v, q)$ are continuous bilinear forms on $V \times V$ and $V \times Q$,

$$\begin{aligned} &\exists\, M_a > 0 \quad \forall u, v \in V \quad a(u, v) \leq M_a ||u||_V ||v||_V \\ &\exists\, M_b > 0 \quad \forall v \in V,\ \forall q \in Q \quad b(v, q) \leq M_b ||v||_V ||q||_Q. \end{aligned} \tag{6.5.1}$$

To simplify the presentation, we also assume that

$$a(\cdot, \cdot) \text{ is symmetric and positive semi-definite.} \tag{6.5.2}$$

Setting $||v||_a := (a(v, v))^{1/2}$ (which in general will only be a semi-norm on V), this immediately gives

$$\forall u, v \in V \quad a(u, v) \leq ||u||_a ||v||_a. \tag{6.5.3}$$

These properties will be assumed to hold throughout all the section. For any given pair (f, g) in $V' \times Q'$, the standard mixed problem is then to find (u, p) in $V \times Q$ such that

$$\begin{cases} a(u, v) + b(v, p) = \langle f, v \rangle & \forall v \in V \\ b(u, q) = \langle g, q \rangle & \forall q \in Q. \end{cases} \tag{6.5.4}$$

We now know that in order to have existence, uniqueness and continuous dependence from the data for problem (6.5.4), it is necessary and sufficient that the bilinear forms $a(\cdot, \cdot)$ and $b(\cdot, \cdot)$ satisfy conditions (5.1.6) and (5.1.1). We thus suppose to have on $b(\cdot, \cdot)$ the *inf-sup condition*.

There exists $\beta > 0$ such that

$$\inf_{q \in Q} \sup_{v \in V} \frac{b(v, q)}{||v||_V ||q||_Q} \geq \beta. \tag{6.5.5}$$

We shall, for simplicity, assume the *ellipticity on the kernel* (5.1.7) instead of (5.1.1).

There exists $\alpha > 0$ such that

$$a(v, v) \geq \alpha ||v||_V^2 \quad \forall v \in \mathrm{Ker}B \tag{6.5.6}$$

where the kernel $\mathrm{Ker}B$ is defined as:

$$\mathrm{Ker}B = \{v \in V \text{ such that } b(v, q) = 0 \quad \forall q \in Q\}. \tag{6.5.7}$$

In Chap. 1 (see Sect. 1.3.4), we have seen many examples of mixed formulations of boundary value problems related to various applications in fluid mechanics and in continuous mechanics and we have shown that there are eigenvalue problems associated with most of them. We shall be interested, here, in the approximation of these eigenvalue problems. We thus consider the discrete analogue of (6.5.4).

We assume that we are given two families of finite dimensional subspaces V_h and Q_h of V and Q, respectively, and we consider the discretised problem: *find* (u_h, p_h) *in* $V_h \times Q_h$ *such that*

$$\begin{cases} a(u_h, v_h) + b(v_h, p_h) = \langle f, v_h\rangle & \forall v_h \in V_h \\ b(u_h, q_h) = \langle g, q_h\rangle & \forall q_h \in Q_h. \end{cases} \tag{6.5.8}$$

We have seen in Chap. 5 that discrete analogues of (6.5.5) and (6.5.6) are sufficient to ensure solvability of the discrete problem together with optimal error bounds. More precisely, the spaces V_h and Q_h should satisfy two conditions:

- The *discrete ellipticity on the kernel*: there exists $\alpha > 0$, independent of h, such that

  $$a(v_h, v_h) \geq \alpha ||v_h||_V^2 \quad \forall v_h \in \mathrm{Ker}B_h, \tag{6.5.9}$$

  where the discrete kernel $\mathrm{Ker}B_h$ is defined as

  $$\mathrm{Ker}B_h = \{v_h \in V_h \text{ such that } b(v_h, v_h) = 0 \ \forall v_h \in V_h\},$$

- The *discrete inf-sup* condition: there exists $\beta > 0$, independent of h, such that

  $$\inf_{q_h \in Q_h} \sup_{v_h \in V_h} \frac{b(v_h, q_h)}{||v_h||_V ||q_h||_Q} \geq \beta. \tag{6.5.10}$$

  Then, we have unique solvability of (6.5.8) and the following error estimate

  $$||u - u_h||_V + ||p - p_h||_V \leq C \left(\inf_{v \in V_h} ||u - v||_V + \inf_{q \in Q_h} ||p - q||_Q \right). \tag{6.5.11}$$

We now turn to the eigenvalue problems. As we have seen, the eigenvalue problem which is naturally associated with the corresponding boundary value problem in strong form does not correspond to taking $(\lambda u, \lambda p)$ as right-hand side of (6.5.4). Instead, according to the different cases, the *natural* eigenvalue problem is obtained by taking $(\lambda u, 0)$ or $(0, -\lambda p)$ as right-hand side of (6.5.4). One expects, as for instance in [299], that (6.5.9) and (6.5.10), together with suitable compactness properties, are sufficient to ensure good convergence of the eigenvalues. However, when the problem is set in mixed variational form, compactness is more delicate to deal with. It was shown in [82] that, for the particular case of (1.3.85) for the mixed Poisson problem, even if the operator mapping g into u is clearly compact, assumptions (6.5.9) and (6.5.10) are not sufficient to avoid, for instance, the presence of spurious eigenvalues in the discrete spectrum. Here, we address a more general problem, in abstract form, and we look for sufficient (and, possibly, necessary) conditions in order to have good approximation properties for the eigenvalue problems having either $(\lambda\, u, 0)$ or $(0, -\lambda\, p)$ at the right-hand side. As we shall see, in each of the two cases, (6.5.9) and (6.5.10) might be neither necessary nor sufficient for that.

Our approach will be more similar to the one of [188] than to the one of [112] or [40]. Important references for the study of eigenvalue problems in mixed form are [43, 299, 316]. With respect to *sufficient* conditions, our development introduces minor differences. For instance, our bilinear form $a(\cdot, \cdot)$ is not supposed to be positive definite. Moreover, previous related papers deal mostly with cases in which the two components of the solution of the direct problem are both convergent, while we accept discretisations that can produce singular global matrices. On the other hand, having assumed symmetry of $a(\cdot, \cdot)$, we do not have to consider adjoint problems as in [188]. However, in practical cases, the actual gain is negligible. The major interest of the present setting consists in showing that our sufficient conditions are, mostly, also *necessary*, thus providing a severe test for assessing whether a given discretisation is suitable for computing eigenvalues or not. This justifies, in our opinion, the apparently excessive generality of our abstract approach. Indeed, as we shall see, convergence of discrete eigenvalues does not even imply, for mixed formulations, the non-singularity of the corresponding global matrices.

Finally, we point out that we do not look here for a priori estimates for eigenvalues and eigenvectors, but only deal with convergence. This is somehow in agreement with the fact that necessary conditions are a major issue here. However, in most cases, a priori error estimates can be readily deduced, checking the last step in the proofs of sufficient conditions and/or applying the general instruments of, say, [43, 109, 299] (see also [76] for a review).

6.5.1 Some Classical Results

Before considering the case of eigenvalue problems in mixed form, we need to recall some classical facts. Let H be a Hilbert space and $T : H \to H$ be a self-adjoint compact operator. To simplify the presentation, we assume that T is non-negative.

We are interested in the eigenvalues $\lambda \in \mathbb{R}$ defined by

$$\lambda\, Tu = u, \quad \text{with } u \in H \setminus \{0\}. \tag{6.5.12}$$

In the above assumptions, it is well-known that there exists a sequence $\{\lambda_i\}$ and an associated orthonormal basis $\{u_i\}$ such that

$$\begin{aligned} &\lambda_i T u_i = u_i, \\ &0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_i \le \cdots, \\ &\lim_{i\to\infty} \lambda_i = +\infty. \end{aligned} \tag{6.5.13}$$

We also set, for $i \in \mathbb{N}$, $E_i = \operatorname{span}(u_i)$.

The following mapping will be useful. Let $m : \mathbb{N} \to \mathbb{N}$ be the application which to every N associates the dimension of the space generated by the eigenspaces of the first N distinct eigenvalues; that is

$$m(1) = \dim \{\oplus_i E_i : \lambda_i = \lambda_1\}, \\ m(N+1) = m(N) + \dim \left\{\oplus_i E_i : \lambda_i = \lambda_{m(N)+1}\right\}. \tag{6.5.14}$$

Clearly, $\lambda_{m(1)}, \ldots, \lambda_{m(N)}$ ($N \in \mathbb{N}$) will now be the first N *distinct* eigenvalues of (6.5.12).

Assume that we are given, for every $h > 0$, a self-adjoint non-negative operator $T_h : H \to H$ with finite range. We denote by $\lambda_i^h \in \mathbb{R}$ the eigenvalues of the problem

$$\lambda\, T_h u = u, \quad \text{with } u \in H \setminus \{0\}. \tag{6.5.15}$$

Let H_h be the finite-dimensional range of T_h and $\dim H_h =: N(h)$; then, T_h admits $N(h)$ real eigenvalues denoted λ_i^h such that

$$0 \leq \lambda_1^h \leq \cdots \leq \lambda_i^h \leq \ldots \lambda_{N(h)}^h. \tag{6.5.16}$$

The associated discrete eigenfunctions u_i^h, $i = 1, \ldots, N(h)$, give rise to an orthonormal basis of H_h with respect to the scalar product of H. Let $E_i^h := \operatorname{span}(u_i^h)$.

We assume that

$$\lim_{h \to 0} ||T - T_h||_{\mathcal{L}(H)} = 0. \tag{6.5.17}$$

It is a classical result in spectrum perturbation theory that (6.5.17) implies the following convergence property for eigenvalues and eigenvectors:

$$\forall \epsilon > 0, \ \forall N \in \mathbb{N} \quad \exists h_0 > 0 \text{ such that } \forall h \leq h_0 \\ \max_{i=1,\ldots,m(N)} |\lambda_i - \lambda_i^h| \leq \epsilon, \\ \hat{\delta}(\oplus_{i=1}^{m(N)} E_i, \oplus_{i=1}^{m(N)} E_i^h) \leq \epsilon, \tag{6.5.18}$$

where $\hat{\delta}(E, F)$, for E and F linear subspaces of H, represents the gap between E and F and is defined by

$$\hat{\delta}(E, F) = \max[\delta(E, F), \delta(F, E)], \\ \delta(E, F) = \sup_{u \in E,\ ||u||_H = 1} \inf_{v \in F} ||u - v||_H. \tag{6.5.19}$$

Vice versa, it is not difficult to prove that (6.5.18) is a sufficient condition for (6.5.17).

6.5.2 Eigenvalue Problems in Mixed Form

Let us go back to the abstract framework introduced above. In particular, assume, for the moment, that (6.5.5) and (6.5.6) are satisfied and that (6.5.8) has a solution for every (f, g) in $V' \times Q'$. Problems (6.5.4) and (6.5.8) then define, in a natural way, two operators $S(f, g) = (u, p)$ (solution of (6.5.4)) and $S_h(f, g) = (u_h, p_h)$ (solution of (6.5.8)). To make things precise, we introduce, for every $h > 0$, the dual norms:

$$||f||_{V_h'} = \sup_{v_h \in V_h} \frac{\langle f, v_h \rangle}{||v_h||_V} \qquad ||g||_{Q_h'} = \sup_{q_h \in Q_h} \frac{\langle g, q_h \rangle}{||q_h||_Q}. \tag{6.5.20}$$

From Theorem 3.4.4 of Chap. 3, we know that (6.5.10) and (6.5.9) imply that the discrete operator S_h is bounded from $V_h' \times Q_h'$ to $V \times Q$, uniformly in h, and we have the bounds (3.4.103) and (3.4.104) (with $\mathbf{x} = u_h$ and $\mathbf{y} = p_h$). Moreover, Lemma 3.5.2 tells us that the converse holds true.

Lemma 6.5.1. *If there exists a constant $C > 0$ such that, for every $h > 0$ and for every quadruplet $(u_h, p_h, f, g) \in V_h \times Q_h \times V' \times Q'$ satisfying (6.5.8), one has*

$$||S_h(f, g)||_{V \times Q} \leq C(||f||_{V_h'} + ||g||_{Q_h'}), \tag{6.5.21}$$

then (6.5.10) and (6.5.9) are verified with $\beta = 1/C$ and $\alpha = 1/(C^2 M_a)$. Then, (6.5.8) has a solution for all $f \in V_h'$ and $g \in Q_H'$.

Proof. This is a mere rewriting of Lemma 3.5.2. □

We now consider the eigenvalue problem. For the sake of simplicity, let us assume for the moment that there exist two Hilbert spaces H_V and H_Q such that we can identify

$$\begin{aligned} H_V &\equiv H_V', \\ H_Q &\equiv H_Q' \end{aligned} \tag{6.5.22}$$

and such that

$$\begin{aligned} V \subseteq H_V \subseteq V' \\ Q \subseteq H_Q \subseteq Q' \end{aligned} \tag{6.5.23}$$

hold with dense and continuous embedding, in a compatible way.

The restrictions of S and S_h to $H_V \times H_Q$ now define two operators from $H_V \times H_Q$ into itself.

As a consequence of (6.5.11) and Lemma 6.5.1, it is immediate to prove the following proposition.

Proposition 6.5.1. *Assume that (6.5.10) and (6.5.9) hold. Then, S_h converges uniformly to S in $\mathcal{L}(H_V \times H_Q)$ if and only if S (from $H_V \times H_Q$ into itself) is compact.* □

This proposition concludes the convergence analysis for the eigenvalue problems associated to (6.5.4) and (6.5.8). However, in the applications, one usually finds eigenvalue problems associated to (6.5.4) and (6.5.8) when one of the two components of the datum is zero. Let us set these eigenvalue problems in their appropriate abstract framework introducing the following operators:

$$\begin{array}{ll} C_V : V' \to V' \times Q' & C_Q : Q' \to V' \times Q' \\ C_V(f) = (f, 0) & C_Q(g) = (0, g) \end{array} \tag{6.5.24}$$

and their adjoints

$$\begin{array}{ll} C_V^* : V \times Q \to V & C_Q^* : V \times Q \to Q \\ C_V^*(v, q) = v & C_Q^*(v, q) = q. \end{array} \tag{6.5.25}$$

We shall say that (6.5.4) is a *problem of the type* $(f,\ 0)$ if the right-hand side in (6.5.4) satisfies $g = 0$. Similarly, we shall say that (6.5.4) is a *problem of the type* $(0,\ g)$ if the right-hand side in (6.5.4) satisfies $f = 0$. Correspondingly, we shall study the approximation of the eigenvalues of the following operators:

$$\begin{array}{ll} T_V = C_V^* \circ S \circ C_V : V' \to V, & \text{for problems of the type } (f,\ 0), \\ T_Q = C_Q^* \circ S \circ C_Q : Q' \to Q, & \text{for problems of the type } (0,\ g). \end{array} \tag{6.5.26}$$

Whenever the associated discrete problems are solvable, we can introduce the discrete counterparts of T_V and T_Q as:

$$\begin{array}{ll} T_V^h = C_V^* \circ S_h \circ C_V : V' \to V, & \text{for problems of the type } (f,\ 0), \\ T_Q^h = C_Q^* \circ S_h \circ C_Q : Q' \to Q, & \text{for problems of the type } (0,\ g). \end{array} \tag{6.5.27}$$

6.5.3 Special Results for Problems of Type $(f, 0)$ *and* $(0, g)$

In the remaining part of this section, we recall the results obtained in Sect. 3.5.3 on the solvability and boundedness of the discrete operators with either the discrete *inf-sup* condition or the discrete ellipticity on the kernel for the special type of data associated with our eigenvalue problems.

Problems of the type $(f, 0)$: From Proposition 3.5.2, we have the following result.

Proposition 6.5.2. *If the discrete ellipticity on the kernel* (6.5.9) *holds and* $g = 0$, *then problem* (6.5.8) *has at least one solution* (u_h, p_h). *Moreover,* u_h *is uniquely determined by* f *and*

$$||u_h||_V \leq \frac{1}{\alpha} ||f||_{V_h'}, \tag{6.5.28}$$

where α *is the constant appearing in* (6.5.9). □

We also have the reciprocal from Proposition 3.5.3.

Proposition 6.5.3. *Assume that there exists a constant* $C > 0$ *such that for every* $h > 0$ *and for every quadruplet* $(u_h, p_h, f, 0) \in V_h \times Q_h \times V' \times Q'$ *satisfying* (6.5.8), *one has*

$$||u_h||_V \leq C||f||_{V_h'}, \tag{6.5.29}$$

then the operator T_V^h *is defined in all* V' *and the discrete ellipticity on the kernel* (6.5.9) *holds with* $\alpha = 1/(C^2 M_a)$, M_a *being the continuity constant of* $a(\cdot,\cdot)$ *(see* (6.5.1)*).* □

Problems of the form (0, g**):** In the same way, we have from Proposition 3.5.5 the following result.

Proposition 6.5.4. *Assume that the following* weak discrete *inf-sup* condition *holds: for every* $h > 0$, *there exists a constant* $\beta_h > 0$ *such that*

$$\inf_{q_h \in Q_h} \sup_{v_h \in V_h} \frac{b(v_h, q_h)}{||v_h||_V ||q_h||_Q} \geq \beta_h. \tag{6.5.30}$$

Then, for every $g \in V'$ *and* $f = 0$, *problem* (6.5.8) *has at least one solution* (u_h, p_h) *and* p_h *is uniquely determined by* g. □

Proposition 6.5.5. *Assume that there exists a constant* $C > 0$ *such that for every* $h > 0$ *and for every quadruplet* $(u_h, p_h, 0, g) \in V_h \times Q_h \times V' \times Q'$ *satisfying* (6.5.8), *one has*

$$||p_h||_V \leq C||g||_{V_h'}. \tag{6.5.31}$$

Then, the operator T_Q^h *is defined in all* Q' *and the weak discrete inf-sup condition* (6.5.30) *holds. In general,* (6.5.31) *does not imply the discrete inf-sup condition* (6.5.10). □

Proof. As in Proposition 6.5.5, the assumption (6.5.31) implies that, with obvious notation, B_h^t is injective, therefore B_h will be surjective and this implies (6.5.30).

However, (6.5.10) cannot be deduced in general: consider the case when $a(\cdot,\cdot) \equiv 0$, $V_h = Q_h$ and $b(\cdot,\cdot)$ is h times the scalar product in V_h. □

Proposition 6.5.6. *Assume that there exists a constant* $C > 0$ *such that for every* $h > 0$ *and for every quadruplet* $(u_h, p_h, 0, g) \in V_h \times Q_h \times V' \times Q'$ *satisfying* (6.5.8), *one has*

$$||u_h||_V + ||p_h||_Q \leq C||g||_{Q_h'}, \tag{6.5.32}$$

then both T_Q^h *and* $C_V^* \circ S_h \circ C_Q$ *are defined on* Q' *and* (6.5.10) *holds with* $\beta = 1/C$. □

Proof. This results directly from Proposition 6.5.6. □

Moreover, we have the following proposition.

Proposition 6.5.7. *If there exists* $C > 0$ *such that*

$$||C_Q^* \circ S_h \circ C_V||_{\mathcal{L}(V_h', V_h)} \leq C \tag{6.5.33}$$

for every $h > 0$*, then* (6.5.10) *holds with* $\beta = 1/C$. □

Proof. The proof can be done as we did for Lemma 3.5.2. □

We therefore see from Propositions 6.5.3 and 6.5.7, that for problems of the type $(f, 0)$, the estimate (6.5.29) on u_h implies (6.5.9) and the estimate (6.5.33) on p_h implies (6.5.10). Analogue properties do not entirely hold for problems of the type $(0, g)$.

6.5.4 Eigenvalue Problems of the Type $(f, 0)$

In this section, together with (6.5.1) and (6.5.2), we assume that ellipticity on the kernel (6.5.6) and the *inf-sup* condition (6.5.5) are verified. We also assume that we are given a Hilbert space H_V (that we shall identify with its dual space H_V') such that

$$V \subseteq H_V \subseteq V' \tag{6.5.34}$$

with continuous and dense embeddings. We consider the eigenvalue problem: *find* (λ, u) *in* $\mathbb{R} \times V$*, with* $u \neq 0$ *such that there exists* $p \in V$ *verifying*

$$\begin{aligned} &a(u, v) + b(v, p) = \lambda (u, v)_{H_V} \quad && \forall v \in V, \\ &b(u, q) = 0 && \forall q \in Q. \end{aligned} \tag{6.5.35}$$

In the formalism of Sect. 6.5.2, this can be written as

$$\lambda\, T_V u = u. \tag{6.5.36}$$

We assume that the operator T_V is compact from H_V to V.

Suppose now that we are given two finite dimensional subspaces V_h and Q_h of V and Q, respectively. Then, the approximation of (6.5.35) is: *find* (λ_h, u_h) *in* $\mathbb{R} \times V_h$*, with* $u_h \neq 0$ *such that there exists* $p_h \in Q_h$ *verifying*

$$\begin{aligned} &a(u_h, v_h) + b(v_h, p_h) = \lambda_h (u_h, v_h)_{H_V} \quad && \forall v_h \in V_h, \\ &b(u_h, q_h) = 0 && \forall q_h \in Q_h, \end{aligned} \tag{6.5.37}$$

which can be written as

$$\lambda_h\, T_V^h u_h = u_h. \tag{6.5.38}$$

We are now looking for necessary and sufficient conditions that ensure the uniform convergence of T_V^h to T_V in $\mathcal{L}(H_V, V)$ which, as we have seen, implies the convergence of eigenvalues and eigenvectors (see (6.5.18)).

To start with, we look for sufficient conditions and for this, we introduce some notation. Let V_0^H and Q_0^H be the subspaces of V and Q, respectively, containing all the solutions $u \in V$ and $p \in V$, respectively, of problem (6.5.4) when $g = 0$; that is, with the formalism of the Sect. 6.5.2,

$$\begin{aligned} V_0^H &= C_V^* \circ S \circ C_V(H_V) = T_V(H_V) \\ Q_0^H &= C_Q^* \circ S \circ C_V(H_V). \end{aligned} \tag{6.5.39}$$

Notice that the following inclusion holds true:

$$V_0^H \subseteq \mathrm{Ker} B.$$

The spaces V_0^H and Q_0^H will be endowed with the *natural* norm: that is, for instance,

$$\begin{aligned} ||v||_{V_0^H} &:= \inf\{||\eta||_{H_V},\ T_V\eta = v\}; \\ ||q||_{Q_0^H} &:= \inf\{||\eta||_{H_V},\ C_Q^* \circ S \circ C_V\eta = q\}. \end{aligned} \tag{6.5.40}$$

Definition 6.5.1. We say that the **weak approximability** of Q_0^H is verified if there exists $\omega_1(h)$, tending to zero as h tends to zero, such that for every $p \in Q_0^H$,

$$\sup_{v_h \in \mathrm{Ker} B_h} \frac{b(v_h, p)}{||v_h||_V} \leq \omega_1(h)||p||_{Q_0^H}. \tag{6.5.41}$$

Notice that, in spite of its appearance, (6.5.41) is indeed an approximability property. Actually, as $v_h \in \mathrm{Ker} B_h$, we have $b(v_h, p) = b(v_h, p - p_I)$ for every $p_I \in Q_h$, which has, usually, to be used to verify (6.5.41).

Definition 6.5.2. We say that the **strong approximability** of V_0^H is verified if there exists $\omega_2(h)$, tending to zero as h tends to zero, such that for every $u \in V_0^H$, there exists $u_I \in \mathrm{Ker} B_h$ such that

$$||u - u_I||_V \leq \omega_2(h)||u||_{V_0^H}. \tag{6.5.42}$$

Theorem 6.5.1. *Let us assume that the discrete ellipticity on the kernel* (6.5.9) *is verified. Assume moreover the* weak approximability *of* Q_0^H *and the* strong approximability *of* V_0^H. *Then, the sequence* T_V^h *converges uniformly to* T_V *in* $\mathcal{L}(H_V, V)$*, that is, there exists* $\omega_3(h)$*, tending to zero as h tends to zero, such that*

$$||T_V f - T_V^h f||_V \leq \omega_3(h)||f||_{H_V}, \quad \textit{for all } f \in H_V. \tag{6.5.43}$$

Proof. Let $f \in H_V$ and let $(u, p) \in V_0^H \times Q_0^H$ be solution of (6.5.4): $(u, p) = S(f, 0)$. As we assumed (6.5.9), Proposition 6.5.2 ensures that T_V^h is well defined on V'. Recall that $u := T_V(f)$. Let $u_h := T_V^h(f)$ and let p_I be such that (u_h, p_I) is a solution of (6.5.8) (such a p_I might not be unique). In order to prove the uniform convergence of T_V^h to T_V, we have to estimate the difference $||u - u_h||_V$. We do it by bounding the term $||u_I - u_h||_V$, where u_I is given by (6.5.42), and then by using the triangular inequality. We have

$$
\begin{aligned}
\alpha||u_I - u_h||_V^2 &\le a(u_I - u_h, u_I - u_h) \\
&= a(u_I - u, u_I - u_h) + a(u - u_h, u_I - u_h) \\
&\le M_a||u_I - u||_V||u_I - u_h||_V - b(u_I - u_h, p - p_h) \\
&\le \left(M_a||u_I - u||_V + \sup_{v_h \in \mathrm{Ker} B_h} \frac{b(v_h, p - p_h)}{||v_h||_V} \right) ||u_I - u_h||_V \\
&= \left(M_a||u_I - u||_V + \sup_{v_h \in \mathrm{Ker} B_h} \frac{b(v_h, p)}{||v_h||_V} \right) ||u_I - u_h||_V .
\end{aligned}
\tag{6.5.44}
$$

The result then follows immediately from the strong approximability of V_0^H and the *weak approximability* of Q_0^H. In particular, we can take $\omega_3(h) = (1 + M_a/\alpha)\omega_2(h) + \omega_1(h)/\alpha$. □

In the following theorem, we shall see that the assumptions of Theorem 6.5.1 are also, in a sense, necessary for the uniform convergence of T_V^h to T_V in $\mathcal{L}(H_V, V)$.

Theorem 6.5.2. *Assume that the sequence T_V^h is bounded in $\mathcal{L}(V', V)$, and converges uniformly to T_V in $\mathcal{L}(H_V, V)$ (see (6.5.43)). Then, the ellipticity in the kernel property (6.5.9) holds true. Moreover, both the* strong approximability *of V_0^H and the* weak approximability *of Q_0^H are satisfied.*

Proof. Condition (6.5.9) can be obtained applying Proposition 6.5.3. Let u be an element of V_0^H. Then, by definition of V_0^H, there is $f \in H_V$ such that $u = T_V f$. Define $u_I := T_V^h f$. Uniform convergence implies the *strong approximability* of V_0^H.

In a similar way, let p be an element of Q_0^H. Then, by definition of Q_0^H, $p = C_Q^* \circ S \circ C_V f$ for some $f \in H_V$. There might be more than one such f. We choose $\overline{f}$ such that $||\overline{f}||_{H_V} \le \frac{3}{2} \inf f\{||f||_{H_V} : C_Q^* \circ S \circ C_V f = p\} = \frac{3}{2}||p||_{Q_0^H}$. Let $u := T_V \overline{f}$. Correspondingly, let $u_h := T_V^h \overline{f}$ and let p_h be such that (u_h, p_h) is a solution of (6.5.8) with the same right-hand side (such a p_h might not be unique). Then, we obtain

$$
\sup_{v_h \in \mathrm{Ker} B_h} \frac{b(v_h, p)}{||v_h||_V} = \sup_{v_h \in \mathrm{Ker} B_h} \frac{b(v_h, p - p_h)}{||v_h||_V} = \sup_{v_h \in \mathrm{Ker} B_h} \frac{a(u - u_h, v_h)}{||v_h||_V}
$$

$$\leq M_a||u - u_h||_V \leq M_a\omega_3(h)||f||_{H_V} \leq \frac{3}{2}M_a\omega_3(h)||p||_{Q_0^H},$$

which gives (6.5.41) with $\omega_1(h) = \frac{3}{2}M_a\omega_3(h)$, that is, the *weak approximability* of Q_0^H. □

Remark 6.5.1. We shall present examples of eigenvalue problems of type $(f, 0)$ for the Stokes problem in Sect. 8.11. □

6.5.5 *Eigenvalue Problems of the Form* $(0, g)$

In this section, together with (6.5.1) and (6.5.2), we assume that, for every given $g \in Q'$ and $f = 0$, problem (6.5.4) has a unique solution (u, p) and that there exists a constant C (independent of g) such that

$$||u||_V + ||p||_Q \leq C||g||_{Q'}. \tag{6.5.45}$$

It is easy to see that this implies the *inf-sup* condition (6.5.5) but not the ellipticity on the kernel (6.5.6).

Remark 6.5.2. An example of this situation can be found in Sect. 10.1.1 for the $\psi - \omega$ formulation of the biharmonic problem. □

In the following, we assume that we are given a Hilbert space H_Q (that we shall identify with its dual space H_Q') such that

$$Q \subseteq H_Q \subseteq Q' \tag{6.5.46}$$

with continuous and dense embeddings. For simplicity, we assume that for every $q \in Q$, we have $||q||_{H_Q} \leq ||q||_Q$ (with constant equal to 1).

We consider the eigenvalue problem: *find* (λ, p) *in* $\mathbb{R} \times V$, *with* $p \neq 0$ *such that there exists* $u \in V$ *verifying*

$$\begin{cases} a(u, v) + b(v, p) = 0 & \forall v \in V \\ b(u, q) = -\lambda(p, q)_{H_Q} & \forall q \in Q \end{cases} \tag{6.5.47}$$

which in the formalism of Sect. 6.5.2 can be written as

$$\lambda\, T_Q p = -p. \tag{6.5.48}$$

As we shall see, problems of the type $(0, g)$ are more closely related to the abstract theory of [188] than problems of the previous type $(f, 0)$.

From now on, we assume that the operator T_Q is compact from H_Q into Q.

We introduce two finite dimensional subspaces V_h and Q_h of V and Q, respectively. Then, the approximation of (6.5.47) reads: *find* (λ_h, p_h) *in* $\mathbb{R} \times Q_h$, *with* $p_h \neq 0$ *such that there exists* $u_\in V_h$ *verifying*

$$\begin{aligned} &a(u_h, v_h) + b(v_h, p_h) = 0 \quad \forall v_h \in V_h \\ &b(u_h, q_h) = -\lambda_h (p_h, q_h)_{H_Q} \quad \forall q_h \in Q_h, \end{aligned} \tag{6.5.49}$$

that is,

$$\lambda_h T_Q^h p_h = -p_h. \tag{6.5.50}$$

We are now looking for necessary and sufficient conditions that ensure the uniform convergence of T_Q^h to T_Q in $\mathcal{L}(H_Q, Q)$, which implies the convergence of eigenvalues and eigenvectors (see (6.5.18)).

To start with, we look for sufficient conditions.

We introduce some notation. Let V_H^0 and Q_H^0 be the subspaces of V and Q respectively, containing all the solutions $u \in V$ and $p \in Q$, respectively, of problem (6.5.4) when $f = 0$; that is, with the formalism of Sect. 6.5.2,

$$\begin{aligned} &V_H^0 = C_V^* \circ S \circ C_Q(H_Q) \\ &Q_H^0 = C_Q^* \circ S \circ C_Q(H_Q) = T_Q(H_Q). \end{aligned} \tag{6.5.51}$$

It will also be useful to define the space $V_{Q'}^0$ as the image of $C_V^* \circ S \circ C_Q$ (from Q' to V).

As before, the spaces V_H^0, Q_H^0 and $V_{Q'}^0$ will be endowed with their natural norms (see for instance (6.5.40)).

Definition 6.5.3. We say that the *weak approximability of* Q_H^0 *with respect to* $a(\cdot,\cdot)$ is verified if there exists $\omega_4(h)$, tending to zero as h goes to zero, such that for every $p \in Q_0^H$ and for every $v_h \in \mathrm{Ker} B_h$,

$$b(v_h, p) \leq \omega_4(h) ||p||_{Q_H^0} ||v_h||_a. \tag{6.5.52}$$

Notice that (6.5.52) is indeed an approximation property, as we already pointed out for its counterpart (6.5.41).

Definition 6.5.4. We say that the *strong approximability* of Q_H^0 is verified if there exists $\omega_5(h)$, tending to zero as h goes to zero, such that for every $p \in Q_H^0$ there exists $p^I \in Q_h$ such that

$$||p - p^I||_V \leq \omega_5(h) ||p||_{Q_H^0}. \tag{6.5.53}$$

Notice that (6.5.52) and (6.5.53) are (much) weaker forms of assumption H.7 of [188].

Definition 6.5.5. Following Sect. 5.4.3, we say that an operator Π_h from V (or from a subspace of it) into V_h is a B-compatible operator with respect to the bilinear form $b(\cdot,\cdot)$ and the subspace $Q_h \subset Q$ if it verifies, for all v in its domain,

$$b(v - \Pi_h v, q_h) = 0 \quad \forall q_h \in Q_h, \tag{6.5.54}$$

and there exists a constant C_Π, independent of h, such that:

$$||\Pi_h||_{\mathcal{L}(V^0_{Q'},V)} \le C_\Pi. \tag{6.5.55}$$

We now introduce a stronger form of (6.5.55).

Definition 6.5.6. *B-Id-compatible operator*
We shall say that the operator Π_h is B-Id-compatible if it satisfies (6.5.54), (6.5.55) and if moreover it converges to the Identity operator in norm, that is, if there exists $\omega_6(h)$, tending to zero as h tends to zero, such that for every $v \in V^0_H$, we have

$$||v - \Pi_h v||_a \le \omega_6(h)||v||_{V^0_H}. \tag{6.5.56}$$

Remark 6.5.3. In Sect. 5.4.3, we have seen that (6.5.54) and (6.5.55) imply the discrete *inf-sup* condition (6.5.10). Notice that (6.5.56) is strongly related to assumption H.5 of [188]. As we shall see, the condition that $\omega_6(h)$ goes to 0 with h is actually necessary for the convergence of eigenvalues. In [188], it is only assumed to be bounded, that is, essentially (6.5.55). Indeed, their interest was in a priori bounds (and not on necessity) and, moreover, they were dealing with direct problems (and not with eigenvalues). In particular, (6.5.56) is not necessary to obtain point-wise convergence of T^h_Q to T_Q where the discrete ellipticity on the kernel (6.5.9) and the discrete *inf-sup* condition (6.5.10) are sufficient. Notice that, from Remark 5.4.3, the *inf-sup* (6.5.5) and its discrete counterpart (6.5.10) imply (6.5.55), but not (6.5.56). □

We can now prove the following result.

Theorem 6.5.3. *Let us assume that there exists a B-Id-compatible operator Π_h : $V^0_{Q'} \to V_h$, that is satisfying* (6.5.54)–(6.5.56)*. Assume moreover that the* strong approximability *of Q^0_H is verified (see* (6.5.53)*) as well as the* weak approximability of Q^0_H with respect to *a (see* (6.5.52)*). Then, the sequence T^h_Q converges to T_Q uniformly from H_Q into Q, that is, there exists $\omega_7(h)$, tending to zero as h goes to zero, such that*

$$||T_Q g - T^h_Q g||_V \le \omega_7(h)||g||_{H_Q}, \quad \textit{for all } g \in H_Q. \tag{6.5.57}$$

Proof. As we recalled above (6.5.5) and (6.5.54)–(6.5.55) imply the discrete *inf-sup* condition (6.5.10). Thanks to Proposition 6.5.4, T_Q^h is then well defined.

Let $g \in H_Q$ and let $(u, p) \in V_H^0 \times Q_H^0$ be the solution of (6.5.4) with $f = 0$. Recall that $p = T_Q g$. Let $p_h := T_Q^h g$ and let u_h be such that (u_h, p_h) is a solution of (6.5.8) (such a u_h might be not unique). In order to prove the uniform convergence of T_Q^h to T_Q, we have to find a priori estimates for the error $||p - p_h||_Q$. Let $\tilde{g} \in Q'$ be such that $\langle \tilde{g}, p - p_h \rangle = ||p - p_h||_Q$ and $||\tilde{g}||_{Q'} = 1$. Take $vt := C_V^* \circ S \circ C_Q \tilde{g}$, hence $||vt||_{V^0_{Q'}} \leq ||\tilde{g}||_{Q'} = 1$ (see (6.5.40)). Then, we have

$$
\begin{aligned}
||p - p_h||_Q &= \langle \tilde{g}, p - p_h \rangle = b(vt, p - p_h) \\
&= b(vt - \Pi_h vt, p - p_h) + b(\Pi_h vt, p - p_h) \\
&= b(vt - \Pi_h vt, p - pI) - a(u - u_h, \Pi_h vt).
\end{aligned}
\tag{6.5.58}
$$

Let us estimate separately the two terms in the right-hand side:

$$
\begin{aligned}
b(vt - \Pi_h vt, p - pI) &\leq M_b ||vt - \Pi_h vt||_V ||p - pI||_Q \\
&\leq M_b \left(||vt||_V + ||\Pi_h vt||_V\right) ||p - pI||_Q \\
a(u - u_h, \Pi_h vt) &\leq ||\Pi_h vt||_a ||u - u_h||_a.
\end{aligned}
\tag{6.5.59}
$$

Using (6.5.55), we obtain the following estimate for $\Pi_h vt$

$$
||\Pi_h vt||_V \leq C_\Pi ||vt||_{V^0_{Q'}} \leq C_\Pi. \tag{6.5.60}
$$

Putting together (6.5.58)–(6.5.60) and using (6.5.53), we obtain

$$
\begin{aligned}
||p - p_h||_Q &\leq M_b(1 + C_\Pi)||p - pI||_Q + C_\Pi ||u - u_h||_a \\
&\leq M_b(1 + C_\Pi)\omega_5(h)||p||_{Q_H^0} + C_\Pi ||u - u_h||_a.
\end{aligned}
\tag{6.5.61}
$$

To conclude the proof, there remains to estimate $||u - u_h||_a$. Thanks to the triangular inequality and to (6.5.56), we bound only $||\Pi_h u - u_h||_a$ using also (6.5.52) and (6.5.54). Notice that $\Pi_h u - u_h$ belongs to $\mathrm{Ker} B_h$

$$
\begin{aligned}
||\Pi_h u - u_h||_a^2 &= a(\Pi_h u - u, \Pi_h u - u_h) + a(u - u_h, \Pi_h u - u_h) \\
&\leq ||u - \Pi_h u||_a ||\Pi_h u - u_h||_a - b(\Pi_h u - u_h, p - p_h) \\
&= ||u - \Pi_h u||_a ||\Pi_h u - u_h||_a - b(\Pi_h u - u_h, p) \\
&\leq ||\Pi_h u - u_h||_a \left(||u - \Pi_h u||_a + \omega_4(h)||p||_{Q_H^0} \right),
\end{aligned}
\tag{6.5.62}
$$

which, due to (6.5.56), gives

$$||u-u_h||_a \le 2||u-\Pi_h u||_a + \omega_4(h)||p||_{Q_H^0} \le 2\omega_6(h)||u||_{V_H^0} + \omega_4(h)||p||_{Q_H^0} \tag{6.5.63}$$

and (6.5.57) holds with $\omega_7(h) = M_b(1+C_\Pi)\omega_5(h) + 2C_\Pi\omega_6(h) + C_\Pi\omega_4(h)$. □

Remark 6.5.4. In Theorem 6.5.3, we have proved the uniform convergence of T_Q^h to T_Q in $\mathcal{L}(H_Q, Q)$. However, in Sect. 6.5.2, we have seen that the convergence of the spectrum is equivalent to the uniform convergence of T_Q^h to T_Q in $\mathcal{L}(H_Q)$. Indeed, the latter holds under the weaker assumption that there exists a B-compatible operator satisfying only (6.5.56) as we shall see in the following theorem. □

Theorem 6.5.4. *Let us assume that there exists a B-compatible operator (see (6.5.54)) $\Pi_h : V_{Q'}^0 \to V_h$ satisfying (6.5.56). Assume moreover that both the* strong approximability *of Q_H^0 (see (6.5.53)) and the* weak approximability of Q_H^0 with respect to $a(\cdot,\cdot)$ *(see (6.5.52)) are verified. Then, the sequence T_Q^h converges uniformly to T_Q in H_Q.*

Proof. We observe that (6.5.5) and (6.5.54) imply the weak discrete *inf-sup* condition (6.5.30). Thanks to Proposition 6.5.4, T_Q^h is then well defined.

Let $g \in H_Q$ and let $(u,p) \in V_H^0 \times Q_H^0$ be the solution of (6.5.4) with $f = 0$. Recall that $p = T_Q g$. Let $p_h := T_Q^h g$ and let u_h be such that (u_h, p_h) is a solution of (6.5.8) with right-hand side $(0, g)$ (such a u_h might be not unique). We estimate $||p-p_h||_{H_Q}$. Using a duality argument, let $(ut, pt) \in V \times V$ be defined by $(ut, pt) := S(0, p-p_h)$. Due to the definition (6.5.51), ut belongs to V_H^0 with the following estimate $||ut||_{V_H^0} \le ||p-p_h||_{H_Q}$ (see (6.5.40))

$$\begin{aligned}
||p-p_h||^2_{H_Q} &= (p-p_h, p-p_h) = b(ut, p-p_h)\\
&= b(ut-\Pi_h ut, p) + b(\Pi_h ut, p-p_h)\\
&= -a(u, ut-\Pi_h ut) - a(u-u_h, \Pi_h ut)\\
&\le ||u||_a||ut-\Pi_h ut||_a + ||u-u_h||_a||\Pi_h ut||_a\\
&\le ||u||_a\omega_6(h)||ut||_{V_H^0} + 2||u-u_h||_a||ut||_{V_H^0}\\
&\le (\omega_6(h)||u||_a + 2||u-u_h||_a)\,||p-p_h||_{H_Q},
\end{aligned}$$

having assumed $\omega_6(h) \le 1$. Hence,

$$||p-p_h||_{H_Q} \le \omega_6(h)||u||_a + 2||u-u_h||_a.$$

The rest of the proof follows the same lines as the one of Theorem 6.5.3, using (6.5.52) and (6.5.56) (see (6.5.62) and (6.5.63)). □

The remaining part of this section is devoted to see what one can deduce from the uniform convergence of T_Q^h to T_Q.

Theorem 6.5.5. *Assume that the sequence T_Q^h is bounded in $\mathcal{L}(Q', Q)$. Then, there exists a B-compatible operator (see (6.5.54)) $\Pi_h : V^0_{Q'} \to V_h$ such that*

$$||u - \Pi_h u||_a \le C||u||_{V^0_{Q'}}. \tag{6.5.64}$$

Proof. Let u belong to $V^0_{Q'}$. Then, by definition, $u = C_V^* \circ S \circ C_Q g$ for some $g \in Q'$. There is only one g in this condition, and therefore, by definition, $||u||_{V^0_{Q'}} = ||g||_{Q'}$ (see (6.5.40)). Let $p \in Q$ be such that $(u, p) = S(0, g)$. Let $p_h := T_Q^h g$; notice that, by assumption, $||p_h||_Q \le C||g||_{Q'}$. By Propositions 6.5.5 and 6.5.4, there exists at least one u_h such that $(u_h, p_h) \in V_h \times Q_h$ is a corresponding discrete solution of (6.5.8). If such a u_h is unique, we define $\Pi_h u := u_h$. Otherwise, we still define $\Pi_h u$ as the u_h having minimum norm in V. By construction, we have (6.5.54) and

$$||\Pi_h u||_a^2 = \langle g, p_h\rangle \le ||g||_{Q'}||T_Q^h g||_Q \le C||g||^2_{Q'} = C||u||^2_{V^0_{Q'}}. \tag{6.5.65}$$

Let us bound $||u - \Pi_h u||_a$:

$$\begin{aligned} ||u - \Pi_h u||_a^2 &= a(u - \Pi_h u, u - \Pi_h u) \\ &= a(u, u - \Pi_h u) - a(\Pi_h u, u - \Pi_h u) \\ &= -b(u - \Pi_h u, p) - a(u - \Pi_h u, \Pi_h u). \end{aligned} \tag{6.5.66}$$

The first term in the right-hand side can be handled as follows:

$$\begin{aligned} b(u - \Pi_h u, p) &= b(u - \Pi_h u, p - p_h) \\ &= \langle g, p - p_h\rangle - b(\Pi_h u, p - p_h) \\ &= \langle g, p - p_h\rangle + a(u - \Pi_h u, \Pi_h u). \end{aligned} \tag{6.5.67}$$

Inserting (6.5.67) in (6.5.66), we obtain

$$\begin{aligned} ||u - \Pi_h u||_a^2 &= -\langle g, p - p_h\rangle - 2a(u - \Pi_h u, \Pi_h u) \\ &\le ||g||_{Q'}||p - p_h||_Q + 2||u - \Pi_h u||_a||\Pi_h u||_a \\ &\le ||g||_{Q'}\big(||p||_Q + ||p_h||_Q\big) + 2||u - \Pi_h u||_a||\Pi_h u||_a, \end{aligned} \tag{6.5.68}$$

and then the boundedness of T_Q^h and (6.5.65) imply (6.5.64). □

Theorem 6.5.6. *Assume that the sequence T_Q^h converges to T_Q uniformly in $\mathcal{L}(H_Q, Q)$. Then, for all $p \in Q^0_H$, there exists $p^I \in Q_h$ such that (6.5.53) holds true.*

Proof. Let p belong to Q^0_H, then $p = T_Q g$ for a suitable g in H_Q. Let $p_h := T_Q^h g$ be the corresponding discrete solution, then we define $p^I := p_h$ and the inequality (6.5.53) is an easy consequence of the uniform convergence of $T_Q^h g$ to $T_Q g$ in Q. □

Theorem 6.5.7. *Let us assume that the sequence T_Q^h is bounded in $\mathcal{L}(Q', Q)$ and converges uniformly to T_Q in $\mathcal{L}(H_Q, Q)$. In addition, we assume that the following bound holds for the solutions of* (6.5.8) *with* $f = 0$

$$||u_h||_V \leq C||g||_{Q'}. \tag{6.5.69}$$

Then, there exists a B-compatible operator $\Pi_h : V^0_{Q'} \to V_h$ satisfying (6.5.55) *and* (6.5.56)*. Moreover, we have the discrete inf-sup condition* (6.5.10) *and the weak approximability of Q^0_H with respect to $a(\cdot,\cdot)$ (see* (6.5.52)*) holds.*

Proof. From Proposition 6.5.5, we have that $C_V^* \circ S \circ C_Q$ is also well defined and (6.5.10) holds. Let us check (6.5.55). For $u \in V^0_{Q'}$, there exists $g \in Q'$ and $p \in Q$ such that $(u, p) = S(0, g)$. We set $\Pi_h u := C_V^* \circ S_h \circ C_Q g$. As we have seen, (6.5.54) holds trivially, and now (6.5.55) also holds in virtue of (6.5.69), with $C_\Pi := C$.

Now, let us check (6.5.56). Let u belong to V^0_H; by definition $u = C_V^* \circ S \circ C_Q g$ for some $g \in H_Q$. As in the proof of Theorem 6.5.5, g is unique, and $||u||_{V^0_H} = ||g||_{H_Q}$. Let $p := T_Q g$; clearly, $p \in Q^0_H$. Let $p_h := T_Q^h g$. By construction, $(\Pi_h u, p_h)$ solves (6.5.8) with the right-hand side $(0, g)$. Moreover, by the same computations as above, we arrive at (see the first line in (6.5.68))

$$||u - \Pi_h u||_a^2 = -\langle g, p - p_h\rangle - 2a(u - \Pi_h u, \Pi_h u).$$

From this, we have

$$\begin{aligned} ||u - \Pi_h u||_a^2 &= -\langle g, p - p_h\rangle - 2b(\Pi_h u, p - p_h) \\ &\leq \big(||g||_{Q'} + 2M_b||\Pi_h u||_V\big)\, ||p - p_h||_Q \\ &\leq (1 + 2M_b C)||g||_{Q'}\omega_7(h)||g||_{H_Q} \\ &\leq (1 + 2M_b C)\omega_7(h)||g||_{H_Q}^2 \\ &= (1 + 2M_b C)\omega_7(h)||u||_{V^0_H}^2, \end{aligned} \tag{6.5.70}$$

where we used (6.5.69) and the uniform convergence of T_Q^h to T_Q in $\mathcal{L}(H_Q, V)$ (see (6.5.57)). The bound (6.5.70) gives (6.5.56) with:

$$\omega_6(h) = ((1 + 2M_b C)\omega_7(h))^{1/2}.$$

Now, let us check (6.5.52). If $p \in Q^0_H$, then $p = T_Q g$ for a suitable g in H_Q. Let u be such that $(u, p) = S(0, g)$ and set $p_h := T_Q^h g$ and $u_h := \Pi_h u$. Then we get, for every $v_h \in \mathrm{Ker} B_h$,

$$
\begin{aligned}
b(v_h, p) &= b(v_h, p - p_h) \\
&= a(\Pi_h u - u, v_h) \leq M_a ||\Pi_h u - u||_a ||v_h||_a
\end{aligned}
\tag{6.5.71}
$$

and (6.5.56) (already proved) ends the proof since, by definition:

$$
||u||_{V_H^0} = ||g||_{H_Q} = ||p||_{Q_H^0}. \qquad \square
$$

Examples for the mixed formulation of second order linear elliptic problems will be presented in Chap. 7, Sect. 7.1.3. We also refer to Sect. 10.1.2 for an example with the $\psi - \omega$ approximation of the biharmonic problems.