The T-coercivity approach for mixed problems

Mathieu Barré, Patrick Ciarlet

Comptes Rendus
Mathématique
Draft

The T-coercivity approach for mixed problems

T-*coercivité et problèmes mixtes*

Mathieu Barré [a, b] **and Patrick Ciarlet** [c]

[a] Inria, 1 Rue Honoré d'Estienne d'Orves, 91120 Palaiseau, France

[b] LMS, École Polytechnique, CNRS, Institut Polytechnique de Paris, Route de Saclay, 91120 Palaiseau, France

[c] POEMS, CNRS, Inria, ENSTA Paris, Institut Polytechnique de Paris, 828 Boulevard des Maréchaux, 91120 Palaiseau, France

E-mails: mathieu.a.barre@inria.fr (M. Barré), patrick.ciarlet@ensta-paris.fr (P. Ciarlet)

Abstract. Classically, the well-posedness of variational formulations of mixed linear problems is achieved through the inf-sup condition. In this note, we propose an alternative framework to study such problems by using the T-coercivity approach. This is a constructive approach that leads to the design of suitable approximations in a simple way. In general, the derivation of the uniform discrete inf-sup condition for the approximate problems stems straightforwardly from the study of the original problem. To support our view, we solve a series of classical mixed problems with the T-coercivity approach. Among others, the celebrated Fortin Lemma appears naturally in the numerical analysis of the approximate problems.

Résumé. Classiquement, le caractère bien posé des formulations variationnelles de problèmes linéaires mixtes est obtenu à l'aide de la condition inf-sup. Dans cette note, nous proposons un cadre alternatif pour étudier de tels problèmes en utilisant la notion de T-coercivité. Il s'agit d'une approche constructive qui permet en outre de concevoir simplement des approximations numériques adaptées car la dérivation de la condition inf-sup discrète uniforme découle en général directement de l'étude du problème continu. Pour appuyer notre propos, nous résolvons une série de problèmes mixtes classiques grâce à la notion de T-coercivité. Entre autres, le lemme de Fortin apparaît naturellement dans l'analyse numérique des problèmes discrets.

2020 Mathematics Subject Classification. 65N30, 35J57, 76D07, 78M10.

This article is a draft (not yet accepted!)

1. Introduction

Traditionally, the well-posedness of variational formulations of mixed linear problems is achieved through the inf-sup condition, also called stability condition [2, 11, 34]. As a matter of fact, proving this condition allows to derive existence and uniqueness of the solution, and continuous dependence with respect to the data. On the other hand, the way this condition is established depends on the problem to be solved. The analysis of such problems can be performed either following a *monolithic* approach, namely studying the *all-in-one* bilinear form incorporating the constraint, or by studying the constrained part of the problem separately.

In this note, we focus on the monolithic approach and investigate the mixed problem's well-posedness based on the T-coercivity framework. The principle of this framework is to find an *explicit realization* of the inf-sup condition for the all-in-one bilinear form. Of equal importance, in the T-coercivity framework, is the *design of suitable approximations* of the original problem. Indeed, with the help of the explicit realization of the condition for the original problem, one can get useful insight on how to derive the so-called uniform discrete inf-sup condition for the approximate, or discrete, problems set in finite-dimensional vector spaces. Thus, convergence of the approximate solutions to the exact one follows under well-known principles in numerical analysis, such as Céa's Lemma (or a variant), and a basic approximability property of elements of the original space of solutions. To summarize, although the T-coercivity approach may not bring new result to the theory of variational formulations, it proposes an entirely constructive way to study them theoretically and also on how to approximate them efficiently.

So far, the T-coercivity approach has been mainly applied to two categories of linear problems. First, for problems involving an invertible operator and a compact perturbation, see eg. [12–14, 20, 30, 37]. Then, for problems with sign-changing coefficients, cf. [5–10, 15–19, 21, 23, 29, 36]. For the second category, we observe that well-posedness and (efficient) approximation of the variational formulations has actually been achieved with the help of the T-coercivity approach. Up to the authors' knowledge, this approach was only applied to mixed problems in [31, 33]: in the first reference, it is applied to the specific case of neutron diffusion, whereas the second one focuses on perturbed saddle-point problems.

In this note, we apply the T-coercivity approach to general mixed problems, including unperturbed and perturbed saddle-point problems. In particular, we will explain the connections with the classical theory, for which we use [4] as the reference textbook. Among those connections, we note that the celebrated Fortin Lemma will appear naturally in the (numerical) analysis of the discrete problems.

Let us introduce some notations. Given a Hilbert space V, we denote by $(\cdot,\cdot)_V$ and $\|\cdot\|_V$ the scalar product and the norm on V, and by V' its dual space. In a product space $V\times W$, we use the norm

$$\|(v,w)\|_{V\times W}=\left(\|v\|_V^2+\|w\|_W^2\right)^{1/2},$$

and similarly for the scalar product. Vector-valued function spaces are written in boldface character. A connected, bounded, open subset of $\mathbb{R}^d$ with a Lipschitz boundary is called a *domain*.

Let Ω be a domain with boundary $\partial\Omega$. We denote by $\boldsymbol{n}$ the unit outward normal vector field to $\partial\Omega$. Let $L^2(\Omega)$ and $\boldsymbol{L}^2(\Omega)$ be the set of square-integrable real-valued and $\mathbb{R}^d$-valued functions on Ω. The natural norm in $L^2(\Omega)$ or $\boldsymbol{L}^2(\Omega)$ is denoted by $\|\cdot\|$, and we let

$$L^2_0(\Omega)=\left\{v\in L^2(\Omega)\,,\ \int_\Omega v\,\mathrm{d}x=0\right\}.$$

In what follows, unless otherwise stated, the standard Sobolev space $H^1_0(\Omega)$ is endowed with the norm $v\mapsto\|\nabla v\|$, that defines a norm that is equivalent to $\|\cdot\|_{H^1(\Omega)}$ thanks to Poincaré's inequality. The dual space of $H^1_0(\Omega)$ is denoted by $H^{-1}(\Omega)$. Similarly, $\boldsymbol{H}^1_0(\Omega)$ is endowed with the norm $\boldsymbol{v}\mapsto\left(\sum_{i=1,d}\|\nabla v_i\|^2\right)^{1/2}$, that defines a norm that is equivalent to $\|\cdot\|_{\boldsymbol{H}^1(\Omega)}$, and its dual space

is denoted by $\boldsymbol{H}^{-1}(\Omega)$. We introduce the usual Sobolev spaces for vector-valued fields [1]

$$\begin{aligned}
\boldsymbol{H}(\mathrm{div};\Omega) &= \left\{\boldsymbol{v}\in\boldsymbol{L}^2(\Omega),\, \mathrm{div}\,\boldsymbol{v}\in L^2(\Omega)\right\},\\
\boldsymbol{H}_0(\mathrm{div};\Omega) &= \left\{\boldsymbol{v}\in\boldsymbol{H}(\mathrm{div};\Omega),\, \boldsymbol{v}\cdot\boldsymbol{n}=0 \text{ on } \partial\Omega\right\},\\
\boldsymbol{H}(\mathrm{div}0;\Omega) &= \left\{\boldsymbol{v}\in\boldsymbol{H}(\mathrm{div};\Omega),\, \mathrm{div}\,\boldsymbol{v}=0\right\},\\
\boldsymbol{H}(\mathbf{curl};\Omega) &= \left\{\boldsymbol{v}\in\boldsymbol{L}^2(\Omega),\, \mathbf{curl}\,\boldsymbol{v}\in\boldsymbol{L}^2(\Omega)\right\}, \qquad \text{for } d=3,\\
\boldsymbol{H}_0(\mathbf{curl};\Omega) &= \left\{\boldsymbol{v}\in\boldsymbol{H}(\mathbf{curl};\Omega),\, \boldsymbol{v}\times\boldsymbol{n}=0 \text{ on } \partial\Omega\right\}, \qquad \text{for } d=3.
\end{aligned}$$

Unless otherwise specified, $\boldsymbol{H}(\mathrm{div};\Omega)$ is endowed with the norm $\boldsymbol{v}\mapsto\left(\|\boldsymbol{v}\|^2+\|\mathrm{div}\,\boldsymbol{v}\|^2\right)^{1/2}$ and $\boldsymbol{H}(\mathbf{curl};\Omega)$ with the norm $\boldsymbol{v}\mapsto\left(\|\boldsymbol{v}\|^2+\|\mathbf{curl}\,\boldsymbol{v}\|^2\right)^{1/2}$.

The outline is as follows. In Section 2, we introduce the T-coercivity approach, and explain how it can be applied to solve the Stokes problem theoretically. Then, in Section 3, we develop the abstract framework underlying the approach for mixed problems, including saddle-point, augmented and perturbed ones. In Sections 4, 5 and 6, we propose some applications, respectively to electromagnetism, nearly-incompressible elasticity, and diffusion. Then, in Section 7, we propose the *natural* extension of the T-coercivity approach for the conforming approximation of mixed problems. As before, we begin by the Stokes problem, then we consider the numerical analysis for mixed problems in general, before describing how the approach can be applied to electromagnetism, nearly-incompressible elasticity, and diffusion. We conclude by a list of further extensions and recent applications of the T-coercivity approach.

2. T-coercivity for the Stokes problem

The starting point of our study is to propose a T-coercivity approach to solve Stokes problem. Let $\Omega\subset\mathbb{R}^d$ be a domain. We consider the Stokes problem with homogeneous Dirichlet boundary conditions: given a prescribed body force $\boldsymbol{f}\in\boldsymbol{H}^{-1}(\Omega)$, find the velocity $\boldsymbol{u}\in\boldsymbol{H}^1(\Omega)$ and the pressure $p\in L^2_0(\Omega)$ such that

$$\begin{aligned}
-\nu\Delta\boldsymbol{u}+\nabla p &= \boldsymbol{f}, \quad \text{in } \Omega,\\
\mathrm{div}\,\boldsymbol{u} &= 0, \quad \text{in } \Omega,\\
\boldsymbol{u} &= 0, \quad \text{on } \partial\Omega,
\end{aligned} \tag{1}$$

where $\nu>0$ denotes the fluid's viscosity.

The standard method to solve Problem (1) – see [27] – consists in a *one-plus-one* approach. The problem is split into a coercive part

$$a(\boldsymbol{u},\boldsymbol{v})=\nu\int_\Omega\nabla\boldsymbol{u}:\nabla\boldsymbol{v}\,\mathrm{d}x$$

and divergence constraint terms of the form

$$b(\boldsymbol{v},q)=-\int_\Omega q\,\mathrm{div}\,\boldsymbol{v}\,\mathrm{d}x,$$

so that the weak formulation of Problem (1) reads: find $(\boldsymbol{u},p)\in\boldsymbol{H}^1_0(\Omega)\times L^2_0(\Omega)$ such that

$$\begin{aligned}
a(\boldsymbol{u},\boldsymbol{v})+b(\boldsymbol{v},p) &= \langle\boldsymbol{f},\boldsymbol{v}\rangle, \quad \forall\boldsymbol{v}\in\boldsymbol{H}^1_0(\Omega),\\
b(\boldsymbol{u},q) &= 0, \quad \forall q\in L^2_0(\Omega),
\end{aligned} \tag{2}$$

where $\langle\cdot,\cdot\rangle$ denotes the duality product in $\boldsymbol{H}^1_0(\Omega)$. The well-posedness of Problem (2) then follows from Ladyzhenskaya–Babuška–Brezzi's theory [2, 11, 34] since the bilinear form a is coercive on $\boldsymbol{H}^1_0(\Omega)$ and the bilinear form b satisfies the *inf-sup condition*

$$\inf_{q\in L^2_0(\Omega)\setminus\{0\}}\sup_{\boldsymbol{v}\in\boldsymbol{H}^1_0(\Omega)\setminus\{0\}}\frac{b(\boldsymbol{v},q)}{\|\nabla\boldsymbol{v}\|\,\|q\|}\geq\underline{\beta} \tag{3}$$

for some constant $\beta > 0$.

Here, we are going to give an alternative proof that Problem (1) is well-posed by analysing the *all-in-one* bilinear form defined on $\boldsymbol{H}^1_0(\Omega) \times L^2_0(\Omega)$

$$\mathscr{A}\big((\boldsymbol{u},p),(\boldsymbol{v},q)\big) = \nu \int_\Omega \nabla \boldsymbol{u} : \nabla \boldsymbol{v} \,\mathrm{d}x - \int_\Omega p \,\mathrm{div}\, \boldsymbol{v} \,\mathrm{d}x - \int_\Omega q \,\mathrm{div}\, \boldsymbol{u} \,\mathrm{d}x$$

instead of splitting it into two bilinear forms a and b as in (2). This bilinear form is not coercive since

$$\mathscr{A}\big((0,p),(0,p)\big) = 0, \quad \forall\, p \in L^2_0(\Omega).$$

For this reason, we use the notion of T-coercivity [19,20], which can be seen as a reformulation of Banach-Nečas-Babuška's theory. The definition and the main property of T-coercivity are recalled below.

Definition 1. *Let W be a Hilbert space and let $\mathscr{A}(\cdot,\cdot)$ be a continuous bilinear form over $W \times W$. We say that $\mathscr{A}$ is* T-coercive *if there exists a bijective operator $\mathtt{T} \in \mathscr{L}(W)$ and $\underline{\alpha} > 0$ such that*

$$|\mathscr{A}(u, \mathtt{T}u)| \geq \underline{\alpha} \|u\|^2_W, \quad \forall\, u \in W.$$

Proposition 2. *Let W be a Hilbert space. Let $\ell(\cdot)$ be a continuous linear form over W and $\mathscr{A}(\cdot,\cdot)$ be a continuous bilinear form over $W \times W$. The problem*

$$\begin{cases} \text{Find } u \in W \quad \text{such that} \\ \forall\, v \in W, \quad \mathscr{A}(u,v) = \ell(v) \end{cases}$$

is well-posed if and only if $\mathscr{A}$ is T*-coercive. If so, it holds that*

$$\|u\|_W \leq \frac{C_\ell}{\underline{\alpha}} |||\mathtt{T}|||, \tag{4}$$

with C_ℓ the continuity constant of the linear form ℓ. When the bilinear form $\mathscr{A}(\cdot,\cdot)$ is in addition symmetric, the requirement that the operator T *is bijective can be dropped.*

2.1. *Proving well-posedness with* T*-coercivity*

With the T-coercivity tool in mind, we are now ready to establish the main result of this section.

Theorem 3. *The problem*

$$\begin{cases} \text{Find } (\boldsymbol{u},p) \in \boldsymbol{H}^1_0(\Omega) \times L^2_0(\Omega) \quad \text{such that} \\ \forall (\boldsymbol{v},q) \in \boldsymbol{H}^1_0(\Omega) \times L^2_0(\Omega), \quad \mathscr{A}\big((\boldsymbol{u},p),(\boldsymbol{v},q)\big) = \langle \boldsymbol{f}, \boldsymbol{v} \rangle \end{cases} \tag{5}$$

is well-posed and

$$\|(\boldsymbol{u},p)\|_{\boldsymbol{H}^1_0(\Omega)\times L^2_0(\Omega)} \leq \frac{2\max\Big(\sqrt{2}\nu C^2_{\mathrm{div}}, C_{\mathrm{div}}\big(2+\nu^2 C^2_{\mathrm{div}}\big)^{1/2}\Big)}{\min(\nu^2 C^2_{\mathrm{div}}, 1)} \|\boldsymbol{f}\|_{\boldsymbol{H}^{-1}(\Omega)}. \tag{6}$$

Proof. The linear form defined by

$$\ell\big((\boldsymbol{v},q)\big) = \langle \boldsymbol{f}, \boldsymbol{v} \rangle, \quad \forall (\boldsymbol{v},q) \in \boldsymbol{H}^1_0(\Omega), \times L^2_0(\Omega)$$

is continuous over $\boldsymbol{H}^1_0(\Omega) \times L^2_0(\Omega)$ in view of the inequality

$$\ell\big((\boldsymbol{v},q)\big) \leq \|\boldsymbol{f}\|_{\boldsymbol{H}^{-1}(\Omega)} \|(\boldsymbol{v},q)\|_{\boldsymbol{H}^1_0(\Omega)\times L^2_0(\Omega)}. \tag{7}$$

The bilinear form $\mathscr{A}$ is continuous over $\big(\boldsymbol{H}^1_0(\Omega) \times L^2_0(\Omega)\big)^2$ and we observe that it is also symmetric.

Then, from Proposition 2, it is sufficient to show that the bilinear form $\mathscr{A}$ is T-coercive. For a given $(\boldsymbol{u},p) \in \boldsymbol{H}^1_0(\Omega) \times L^2_0(\Omega)$, we look for an element $(\boldsymbol{v}^*, q^*)$ of $\boldsymbol{H}^1_0(\Omega) \times L^2_0(\Omega)$ depending continously on $(\boldsymbol{u},p)$ and such that

$$\mathscr{A}\big((\boldsymbol{u},p),(\boldsymbol{v}^*,q^*)\big) \geq \underline{\alpha} \|(\boldsymbol{u},p)\|^2_{\boldsymbol{H}^1_0(\Omega)\times L^2_0(\Omega)}$$

for some constant $\underline{\alpha} > 0$. In order to get an intuitive idea of the construction of $(\boldsymbol{v}^*, q^*)$, let us start with specific elements $(\boldsymbol{u}, p)$.

- If $p = 0$, then $\|(\boldsymbol{u}, p)\|^2_{\boldsymbol{H}^1_0(\Omega)\times L^2_0(\Omega)} = \|\nabla \boldsymbol{u}\|^2$ and
$$\mathscr{A}\big((\boldsymbol{u}, p), (\boldsymbol{v}^*, q^*)\big) = \nu \int_\Omega \nabla \boldsymbol{u} : \nabla \boldsymbol{v}^* \,\mathrm{d}x - \int_\Omega \operatorname{div} \boldsymbol{u}\, q^* \,\mathrm{d}x,$$
so that we can take $\boldsymbol{v}^* = \boldsymbol{u}$ and $q^* = p = 0$.

- If $\boldsymbol{u} = 0$, then $\|(\boldsymbol{u}, p)\|^2_{\boldsymbol{H}^1_0(\Omega)\times L^2_0(\Omega)} = \|p\|^2$ and
$$\mathscr{A}\big((\boldsymbol{u}, p), (\boldsymbol{v}^*, q^*)\big) = -\int_\Omega p \operatorname{div} \boldsymbol{v}^* \,\mathrm{d}x.$$
In order to recover the expected term $\|p\|^2$ in the above expression, we have to choose $\boldsymbol{v}^*$, the divergence of which is "as close as possible" to p. To that aim, we use the result below, see for instance [27, Corollary I.2.4]. Let $p \in L^2_0(\Omega)$. Then, there exists $\boldsymbol{v}_p \in \boldsymbol{H}^1_0(\Omega)$ satisfying
$$-\operatorname{div} \boldsymbol{v}_p = p. \tag{8}$$
In addition, there exists a constant $C_{\mathrm{div}} > 0$ independent of p such that
$$\|\nabla \boldsymbol{v}_p\| \le C_{\mathrm{div}} \|p\|. \tag{9}$$
The idea is now to choose $\boldsymbol{v}^* = \boldsymbol{v}_p$, where $\boldsymbol{v}_p$ is as in (8)-(9). Hence, taking $q^* = 0$, we find
$$\mathscr{A}\big((\boldsymbol{u}, p), (\boldsymbol{v}^*, q^*)\big) = \|p\|^2,$$
and (9) ensures that the pair $(\boldsymbol{v}_p, 0)$ depends continously on $(0, p)$ in $\boldsymbol{H}^1_0(\Omega) \times L^2_0(\Omega)$.

- If $\operatorname{div} \boldsymbol{u} = 0$, then
$$\mathscr{A}\big((\boldsymbol{u}, p), (\boldsymbol{v}^*, q^*)\big) = \nu \int_\Omega \nabla \boldsymbol{u} : \nabla \boldsymbol{v}^* \,\mathrm{d}x - \int_\Omega p \operatorname{div} \boldsymbol{v}^* \,\mathrm{d}x.$$
Since we need to get a term of the form $\|\nabla \boldsymbol{u}\|^2$ but also of the form $\|p\|^2$, we combine the previous two cases by setting $\boldsymbol{v}^* = \lambda \boldsymbol{u} + \boldsymbol{v}_p$, where λ is a positive coefficient to be adjusted and $\boldsymbol{v}_p$ is the divergence lifting from (8) – (9). Now, we compute
$$\begin{aligned}\mathscr{A}\big((\boldsymbol{u}, p), (\boldsymbol{v}^*, q^*)\big) &= \nu\lambda \int_\Omega \nabla \boldsymbol{u} : \nabla \boldsymbol{u} \,\mathrm{d}x + \nu \int_\Omega \nabla \boldsymbol{u} : \nabla \boldsymbol{v}_p \,\mathrm{d}x - \lambda \int_\Omega p \operatorname{div} \boldsymbol{u} \,\mathrm{d}x - \int_\Omega p \operatorname{div} \boldsymbol{v}_p \,\mathrm{d}x \\ &= \nu\lambda \|\nabla \boldsymbol{u}\|^2 + \nu \int_\Omega \nabla \boldsymbol{u} : \nabla \boldsymbol{v}_p \,\mathrm{d}x + \|p\|^2\end{aligned}$$
since $\operatorname{div} \boldsymbol{u} = 0$ and $-\operatorname{div} \boldsymbol{v}_p = p$. For all $\eta > 0$, Young's inequality implies that
$$\begin{aligned}\int_\Omega \nabla \boldsymbol{u} : \nabla \boldsymbol{v}_p \,\mathrm{d}x &\ge -\frac{\eta}{2} \|\nabla \boldsymbol{u}\|^2 - \frac{1}{2\eta} \|\nabla \boldsymbol{v}_p\|^2 \\ &\ge -\frac{\eta}{2} \|\nabla \boldsymbol{u}\|^2 - \frac{C^2_{\mathrm{div}}}{2\eta} \|p\|^2 \quad \text{in virtue of (9)},\end{aligned}$$
and thus
$$\mathscr{A}\big((\boldsymbol{u}, p), (\boldsymbol{v}^*, q^*)\big) \ge \nu \left(\lambda - \frac{\eta}{2}\right) \|\nabla \boldsymbol{u}\|^2 + \left(1 - \frac{\nu C^2_{\mathrm{div}}}{2\eta}\right) \|p\|^2.$$
Hence, by setting $\eta = \lambda = \nu C^2_{\mathrm{div}}$, we obtain
$$\mathscr{A}\big((\boldsymbol{u}, p), (\boldsymbol{v}^*, q^*)\big) \ge \frac{\nu^2 C^2_{\mathrm{div}}}{2} \|\nabla \boldsymbol{u}\|^2 + \frac{1}{2} \|p\|^2.$$
Note that this result holds for any $q^* \in L^2_0(\Omega)$ depending continously on p.

In the general case, we choose $\boldsymbol{v}^* = \lambda \boldsymbol{u} + \boldsymbol{v}_p$ with $\lambda = \nu C_{\mathrm{div}}^2$ and $q^* = -\lambda p$ so that, even if $\mathrm{div}\,\boldsymbol{u} \neq 0$, the term $-\lambda \int_\Omega p\,\mathrm{div}\,\boldsymbol{u}\,\mathrm{d}x$ cancels with the term $-\int_\Omega \mathrm{div}\,\boldsymbol{u}\,q^*\,\mathrm{d}x$ and we get the same results as in the case $\mathrm{div}\,\boldsymbol{u} = 0$. Namely, the bilinear form $\mathscr{A}$ is T-coercive for the mapping

$$\begin{aligned} \mathtt{T} : \boldsymbol{H}_0^1(\Omega) \times L_0^2(\Omega) &\longrightarrow \boldsymbol{H}_0^1(\Omega) \times L_0^2(\Omega) \\ (\boldsymbol{u}, p) &\longmapsto \big(\nu C_{\mathrm{div}}^2 \boldsymbol{u} + \boldsymbol{v}_p, -\nu C_{\mathrm{div}}^2 p\big), \end{aligned}$$

where $\boldsymbol{v}_p$ is defined by (8) with estimate (9), and it holds that

$$\mathscr{A}\big((\boldsymbol{u}, p), \mathtt{T}(\boldsymbol{u}, p)\big) \geq \frac{\nu^2 C_{\mathrm{div}}^2}{2} \|\nabla \boldsymbol{u}\|^2 + \frac{1}{2}\|p\|^2 \geq \frac{1}{2}\min(\nu^2 C_{\mathrm{div}}^2, 1)\,\|(\boldsymbol{u}, p)\|^2_{\boldsymbol{H}_0^1(\Omega)\times L_0^2(\Omega)}. \tag{10}$$

Thanks to (8)-(9), T belongs to $\mathscr{L}(\boldsymbol{H}_0^1(\Omega) \times L_0^2(\Omega))$. More precisely, we have

$$\begin{aligned} \|\mathtt{T}(\boldsymbol{u}, p)\|^2_{\boldsymbol{H}_0^1(\Omega)\times L_0^2(\Omega)} &= \|\nu C_{\mathrm{div}}^2 \boldsymbol{u} + \boldsymbol{v}_p\|^2_{\boldsymbol{H}_0^1(\Omega)} + \|\nu C_{\mathrm{div}}^2 p\|^2 \\ &\leq 2(\nu C_{\mathrm{div}}^2)^2 \|\nabla \boldsymbol{u}\|^2 + 2\|\nabla \boldsymbol{v}_p\|^2 + (\nu C_{\mathrm{div}}^2)^2 \|p\|^2 \\ &\leq 2(\nu C_{\mathrm{div}}^2)^2 \|\nabla \boldsymbol{u}\|^2 + \big(2C_{\mathrm{div}}^2 + (\nu C_{\mathrm{div}}^2)^2\big)\|p\|^2 \end{aligned}$$

and thus

$$|||\mathtt{T}||| \leq \max\Big(\sqrt{2}\nu C_{\mathrm{div}}^2, C_{\mathrm{div}}\big(2 + \nu^2 C_{\mathrm{div}}^2\big)^{1/2}\Big). \tag{11}$$

Using (7), (10) and (11) in the stability estimate (4), we finally obtain (6). □

Remark 4. The previous result readily extends to the case of a non-null divergence constraint

$$\begin{aligned} -\nu\Delta \boldsymbol{u} + \nabla p &= \boldsymbol{f}, \quad \text{in } \Omega, \\ \mathrm{div}\,\boldsymbol{u} &= g, \quad \text{in } \Omega, \\ \boldsymbol{u} &= 0, \quad \text{on } \partial\Omega, \end{aligned}$$

with $g \in L_0^2(\Omega)$, leading to the stability estimate

$$\|(\boldsymbol{u}, p)\|_{\boldsymbol{H}_0^1(\Omega)\times L_0^2(\Omega)} \leq \frac{2\max\Big(\sqrt{2}\nu C_{\mathrm{div}}^2, C_{\mathrm{div}}\big(2 + \nu^2 C_{\mathrm{div}}^2\big)^{1/2}\Big)}{\min(\nu^2 C_{\mathrm{div}}^2, 1)} \|(\boldsymbol{f}, g)\|_{\boldsymbol{H}^{-1}(\Omega)\times L_0^2(\Omega)}. \tag{12}$$

2.2. *Comments*

The stability estimates (6) and (12) are valid for all C_{div} that fulfills (9). On the other hand, one has

$$\lim_{C_{\mathrm{div}}\to\infty} \frac{2\max\Big(\sqrt{2}\nu C_{\mathrm{div}}^2, C_{\mathrm{div}}\big(2 + \nu^2 C_{\mathrm{div}}^2\big)^{1/2}\Big)}{\min(\nu^2 C_{\mathrm{div}}^2, 1)} = +\infty,$$

i.e. the stability estimates become meaningless for large C_{div}.

Going through the proof of Theorem 3, we observe that the constant obtained in (6) and (12) is just one of the many bounds one can achieve with T-coercivity for the Stokes problem. Indeed, one can choose any positive value of λ: hence, there exists a family of admissible operators T, in the sense of Definition 1, which shows the flexibility of the approach.

Let us provide an illustration. For small viscosity ν (the domain Ω being fixed), it is well-known that the stability constant appearing in the estimate

$$\|(\boldsymbol{u}, p)\|_{\boldsymbol{H}_0^1(\Omega)\times L_0^2(\Omega)} \leq C(\nu)\,\|(\boldsymbol{f}, g)\|_{\boldsymbol{H}^{-1}(\Omega)\times L_0^2(\Omega)}$$

behaves likes $O(\nu^{-1})$. For instance, for the velocity $\boldsymbol{u}$, the result is elementarily obtained by taking the test field $(\boldsymbol{v}, q) = (\boldsymbol{u}, p)$ in (2). On the other hand, in (6) and (12), we find a behavior in $O(\nu^{-2})$. But, if one is interested in obtaining a less severe blowup, one can simply choose

$$\eta = \frac{\nu C_{\mathrm{div}}^2}{2(1 - \frac{\nu}{2})} \quad \text{and} \quad \lambda = \frac{1}{2}(1 + \eta)$$

in the above proof, for all $0 < \nu \leq 1$. Then, one finds that

$$\underline{\alpha} = \frac{\nu}{2} \quad \text{and} \quad |||\mathtt{T}||| \leq \max\Big(\frac{1}{\sqrt{2}}(1+C_{\mathrm{div}}^2), \big(2C_{\mathrm{div}}^2 + \frac{1}{4}(1+C_{\mathrm{div}}^2)^2\big)^{1/2}\Big),$$

so that (4) actually yields a stability constant in $O(\nu^{-1})$.

Theorem 3 provides a fully constructive proof for the well-posedness of Stokes problem, which is an emblematic example of mixed problem. In the next section, we show that the T-coercivity approach employed here is in fact very general and can be extended to a large class of saddle-point problems.

3. Abstract framework

We start with the classical statements regarding the definition of saddle-point problems, and the equivalent conditions to ensure an inf-sup condition. Then, we proceed with the design of abstract operators T to ensure well-posedness for saddle-problems, and for augmented saddle-point problems.

3.1. *Saddle-point problems in Hilbert spaces*

Let V and Q be two Hilbert spaces. In the Hilbert space Q, we introduce the canonical isomorphism $\mathbb{1}_{Q\to Q'} : Q \to Q'$ defined by

$$\langle \mathbb{1}_{Q\to Q'} p, q\rangle_{Q',Q} = (p,q)_Q, \quad \forall p \in Q,\ \forall q \in Q,$$

which is a bijective isometry according to Riesz Theorem. As a matter of fact, its inverse $\mathbb{1}_{Q'\to Q}$ is also a bijective isometry, and

$$(\mathbb{1}_{Q'\to Q} g, q)_Q = \langle g, q\rangle_{Q',Q}, \quad \forall g \in Q',\ \forall q \in Q.$$

We then introduce two bilinear forms $a(\cdot,\cdot)$ on $V\times V$ and $b(\cdot,\cdot)$ on $V\times Q$ that are assumed to be continuous, *i.e.* there exist $C_a > 0$ and $C_b > 0$ such that

$$a(u,v) \leq C_a \|u\|_V \|v\|_V, \quad \forall u \in V, \forall v \in V, \tag{13}$$

$$b(v,q) \leq C_b \|v\|_V \|q\|_Q, \quad \forall v \in V, \forall q \in Q. \tag{14}$$

We denote by A and B the linear continuous operators associated with a and b, defined by

$$A \in \mathscr{L}(V,V'), \quad \langle Au, v\rangle_{V',V} = a(u,v), \quad \forall u \in V, \forall v \in V,$$
$$B \in \mathscr{L}(V,Q'), \quad \langle Bv, q\rangle_{Q',Q} = b(v,q), \quad \forall v \in V, \forall q \in Q.$$

The adjoint operator of B is given by

$$B^* \in \mathscr{L}(Q,V'), \quad \langle B^* q, v\rangle_{V',V} = \langle Bv, q\rangle_{Q',Q} = b(v,q), \quad \forall v \in V, \forall q \in Q.$$

Given $f \in V'$ and $g \in Q'$, we consider the saddle-point problem: find $(u,p) \in V\times Q$ such that

$$\begin{aligned} Au + B^* p &= f, \quad \text{in } V', \\ Bu &= g, \quad \text{in } Q'. \end{aligned} \tag{15}$$

Or, equivalently, in variational form:

$$\begin{cases} \text{Find } (u,p) \in V\times Q \quad \text{such that} \\ \forall v \in V, \quad a(u,v) + b(v,p) = \langle f, v\rangle_{V',V}, \\ \forall q \in Q, \quad b(u,q) = \langle g, q\rangle_{Q',Q}. \end{cases} \tag{16}$$

As for the Stokes problem, we write Problem (15) as an *all-in-one* variational formulation

$$\begin{cases} \text{Find } (u,p) \in V \times Q \quad \text{such that} \\ \forall (v,q) \in V \times Q, \quad \mathscr{A}\big((u,p),(v,q)\big) = \langle f, v\rangle_{V',V} + \langle g, q\rangle_{Q',Q}, \end{cases} \tag{17}$$

where

$$\mathscr{A}\big((u,p),(v,q)\big) = a(u,v) + b(v,p) + b(u,q).$$

In what follows, we show that Problem (17) is well-posed using the notion of $\mathtt{T}$-coercivity, with slightly different techniques depending on the assumptions made on the bilinear form a.

Regarding the form $b(\cdot,\cdot)$ and the operator B, one has the well-known result below, see for instance [27, Lemma I.4.1][1], which can be viewed as a reformulation of Banach's Closed Range Theorem.

Theorem 5. *The following three statements are equivalent:*

(i) *There exists* $\underline{\beta} > 0$ *such that*

$$\inf_{q \in Q\setminus\{0\}} \sup_{v \in V\setminus\{0\}} \frac{b(v,q)}{\|v\|_V \|q\|_Q} \geq \underline{\beta}. \tag{18}$$

(ii) $B : (\operatorname{Ker} B)^\perp \to Q'$ *is an* isomorphism, *and*

$$\|Bv\|_{Q'} \geq \underline{\beta} \|v\|_V, \quad \forall v \in (\operatorname{Ker} B)^\perp.$$

(iii) *There exists an isomorphic operator* $L_B : Q' \to (\operatorname{Ker} B)^\perp$ *such that*

$$B(L_B g) = g \quad \textit{and} \quad \|g\|_{Q'} \geq \underline{\beta} \|L_B g\|_V, \quad \forall g \in Q'.$$

Since our aim is to build operators $\mathtt{T}$ from $V \times Q$ to itself, we first introduce the operator

$$\mathtt{B} = \mathbb{1}_{Q' \to Q} \circ B : V \to Q.$$

For all $v \in V$, $\|Bv\|_{Q'} = \|\mathbb{1}_{Q' \to Q}(Bv)\|_Q = \|\mathtt{B}v\|_Q$ and, for all $(v,q) \in V \times Q$,

$$b(v,q) = \langle Bv, q\rangle_{Q',Q} = \langle \mathbb{1}_{Q \to Q'}(\mathtt{B}v), q\rangle_{Q',Q} = (\mathtt{B}v, q)_Q.$$

Whenever applicable, we also introduce its *right-inverse*

$$\mathtt{L}_B = L_B \circ \mathbb{1}_{Q \to Q'} : Q \to (\operatorname{Ker}\mathtt{B})^\perp.$$

Observe that

$$b(\mathtt{L}_B p, q) = \langle B\mathtt{L}_B p, q\rangle_{Q',Q} = \langle \mathbb{1}_{Q \to Q'} p, q\rangle_{Q',Q} = (p,q)_Q, \quad \forall p \in Q,\ \forall q \in Q. \tag{19}$$

Under these notations, items (ii)-(iii) of Theorem 5 now write

(ii) $\mathtt{B} : (\operatorname{Ker}\mathtt{B})^\perp \to Q$ is an *isomorphism*, and

$$\|\mathtt{B}v\|_Q \geq \underline{\beta} \|v\|_V, \quad \forall v \in (\operatorname{Ker}\mathtt{B})^\perp. \tag{20}$$

(iii) There exists an isomorphic operator $\mathtt{L}_B : Q \to (\operatorname{Ker}\mathtt{B})^\perp$ such that

$$\mathtt{B}(\mathtt{L}_B q) = q \quad \text{and} \quad \|q\|_Q \geq \underline{\beta} \|\mathtt{L}_B q\|_V, \quad \forall q \in Q. \tag{21}$$

For convenience, we often use $\beta = \underline{\beta}^{-1}$, so that

$$\|\mathtt{L}_B q\|_V \leq \beta \|q\|_Q, \quad \forall q \in Q.$$

[1] Item (iii) below is a rephrasing of the original statement, because it is better suited for our purposes. For details, see the proof of Lemma I.4.1. of [27] p. 59, item 2°. The operator L_B is a *right-inverse* of the operator B.

3.2. *How to achieve* T*-coercivity for saddle-point problems?*

If a is coercive on the whole space V, we can extend the proof of Theorem 3 in the following way.

Theorem 6. *Assume that (18) holds true and that the form a is symmetric and positive. If there exists a constant $\alpha > 0$ such that*

$$a(u,u) \geq \alpha \|u\|_V^2, \quad \forall u \in V, \tag{22}$$

then there exists a unique solution to Problem (17) *and*

$$\|(u,p)\|_{V\times Q} \leq \frac{2\max\left(\sqrt{2}C_a\beta^2, \beta\left(2+C_a^2\beta^2\right)^{1/2}\right)}{\min(\alpha C_a\beta^2, 1)} \|(f,g)\|_{V'\times Q'}. \tag{23}$$

Proof. First, we note that the symmetry of the bilinear form a implies that $\mathscr{A}$ is also symmetric. Then, we follow the same ideas as in the proof of Theorem 3, replacing $\boldsymbol{v}_p$ by $\mathrm{L}_B p$. We introduce the mapping

$$\begin{aligned} \mathrm{T} : V\times Q &\longrightarrow V\times Q \\ (u,p) &\longmapsto \left(\lambda u + \mathrm{L}_B p, -\lambda p\right) \end{aligned}$$

and we compute

$$\begin{aligned} \mathscr{A}\big((u,p),\mathrm{T}(u,p)\big) &= a(u,\lambda u) + a(u,\mathrm{L}_B p) + b(\lambda u,p) + b(\mathrm{L}_B p,p) - b(u,\lambda p) \\ &= \lambda a(u,u) + a(u,\mathrm{L}_B p) + \|p\|_Q^2, \end{aligned}$$

in view of (19).

Because the form a is symmetric and positive, we can apply Young's inequality: for any $\eta > 0$,

$$a(u,\mathrm{L}_B p) \geq -\frac{\eta}{2} a(u,u) - \frac{1}{2\eta} a(\mathrm{L}_B p, \mathrm{L}_B p).$$

Taking into account (13) and (21), the latter being equivalent to (18), we get

$$a(\mathrm{L}_B p, \mathrm{L}_B p) \leq C_a \|\mathrm{L}_B p\|_V^2 \leq C_a\beta^2 \|p\|_Q^2$$

and thus

$$a(u,\mathrm{L}_B p) \geq -\frac{\eta}{2} a(u,u) - \frac{C_a\beta^2}{2\eta} \|p\|_Q^2.$$

Hence, recalling (22), if $\lambda - \frac{\eta}{2} > 0$ it follows that

$$\mathscr{A}\big((u,p),\mathrm{T}(u,p)\big) \geq \alpha\left(\lambda - \frac{\eta}{2}\right)\|u\|_V^2 + \left(1 - \frac{C_a\beta^2}{2\eta}\right)\|p\|_Q^2.$$

Setting in particular $\eta = \lambda = C_a\beta^2$, we infer that

$$\mathscr{A}\big((u,p),\mathrm{T}(u,p)\big) \geq \alpha\frac{C_a\beta^2}{2}\|u\|_V^2 + \frac{1}{2}\|p\|_Q^2 \geq \frac{1}{2}\min(\alpha C_a\beta^2, 1)\|(u,p)\|_{V\times Q}^2, \tag{24}$$

which proves that $\mathscr{A}$ is T-coercive.

Since $\mathrm{T}(u,p) = (C_a\beta^2 u + \mathrm{L}_B p, -C_a\beta^2 p)$, it holds that

$$\begin{aligned} \|\mathrm{T}(u,p)\|_{V\times Q}^2 &= \|C_a\beta^2 u + \mathrm{L}_B p\|_V^2 + \|C_a\beta^2 p\|_Q^2 \\ &\leq 2(C_a\beta^2)^2\|u\|_V^2 + 2\|\mathrm{L}_B p\|_V^2 + (C_a\beta^2)^2\|p\|_Q^2 \\ &\leq 2(C_a\beta^2)^2\|u\|_V^2 + \left(2\beta^2 + (C_a\beta^2)^2\right)\|p\|_Q^2, \end{aligned}$$

which yields

$$|||\mathrm{T}||| \leq \max\left(\sqrt{2}C_a\beta^2, \beta\left(2+C_a^2\beta^2\right)^{1/2}\right). \tag{25}$$

Lastly, we observe that

$$\langle f,v\rangle_{V',V} + \langle g,q\rangle_{Q',Q} \leq \|(f,g)\|_{V'\times Q'}\|(v,q)\|_{V\times Q}. \tag{26}$$

Combining (24), (25) and (26), the stability estimate (4) furnishes exactly (23). □

Remark 7. By applying Theorem 6 to Stokes problem, we recover stability estimates (6) and (12) from the correspondence $\alpha = \nu$, $C_a = \nu$ and $\beta = C_{\mathrm{div}}$.

In Ladyzhenskaya–Babuška–Brezzi's theory and in many applications, the bilinear form a is not coercive on the whole space V but only on the kernel of the operator $\mathtt{B}$. This is for instance the case in electromagnetism, which will be detailed in Section 4. The next result shows how to address this situation in the $\mathtt{T}$-coercivity framework (provided that the form a is symmetric and positive), thus establishing the equivalence between the two theories.

Theorem 8. *Assume that the form a is symmetric and positive.*
1. If (18) holds true, and if there exists a constant $\alpha_0 > 0$ such that

$$a(u_0, u_0) \geq \alpha_0 \|u_0\|_V^2, \quad \forall\, u_0 \in \mathrm{Ker}\,\mathtt{B}, \tag{27}$$

then the form $\mathscr{A}$ is $\mathtt{T}$-coercive. In other words, Problem (17) *is well-posed and*

$$\|(u,p)\|_{V\times Q} \leq C \|(f,g)\|_{V'\times Q'}, \tag{28}$$

with C a constant depending only on α_0, β, C_a and C_b.
2. Conversely, if Problem (17) *is well-posed, that is, if the form $\mathscr{A}$ is $\mathtt{T}$-coercive, then (18) and* (27) *both hold.*

Proof. 1. We consider the mapping

$$\begin{aligned} \mathtt{T} : V \times Q &\longrightarrow V \times Q \\ (u,p) &\longmapsto \big(\lambda u + \mathtt{L}_B p, -\lambda p + \lambda\mu \mathtt{B}u\big). \end{aligned}$$

This is almost the same mapping as the one used in the proof of Theorem 6. The only difference is the term $\lambda\mu\mathtt{B}u$, which is going to help us handling the extra terms that do not belong to the kernel of $\mathtt{B}$ by adjusting the value of the constant μ. We get

$$\begin{aligned} \mathscr{A}\big((u,p),\mathtt{T}(u,p)\big) &= a(u,\lambda u) + a(u,\mathtt{L}_B p) + b(\lambda u, p) + b(\mathtt{L}_B p, p) - b(u,\lambda p) + b(u,\lambda\mu\mathtt{B}u) \\ &= \lambda a(u,u) + a(u,\mathtt{L}_B p) + \|p\|_Q^2 + \lambda\mu\|\mathtt{B}u\|_Q^2 \end{aligned}$$

because $b(\mathtt{L}_B p, p) = \|p\|_Q^2$ as previously, and

$$b(u,\mathtt{B}u) = \big\langle Bu, \mathtt{B}u\big\rangle_{Q',Q} = \big(\mathbb{1}_{Q'\to Q}(Bu), \mathtt{B}u\big)_Q = \|\mathtt{B}u\|_Q^2.$$

Since the form a is symmetric and positive, one may use Young's inequality. By proceeding as in the proof of Theorem 6 and after setting $\lambda = C_a\beta^2$, we know that

$$\lambda a(u,u) + a(u,\mathtt{L}_B p) + \|p\|_Q^2 \geq \frac{C_a\beta^2}{2} a(u,u) + \frac{1}{2}\|p\|_Q^2,$$

from which we deduce

$$\mathscr{A}\big((u,p),\mathtt{T}(u,p)\big) \geq \frac{C_a\beta^2}{2}\big(a(u,u) + 2\mu\|\mathtt{B}u\|_Q^2\big) + \frac{1}{2}\|p\|_Q^2.$$

To compensate the lack of coercivity of a outside $\mathrm{Ker}\,\mathtt{B}$, we use the decomposition $u = u_0 + \bar{u}$ with $u_0 \in \mathrm{Ker}\,\mathtt{B}$ and $\bar{u} \in (\mathrm{Ker}\,\mathtt{B})^{\perp}$. Following [4, p. 254], Young's inequality yields

$$\begin{aligned} a(u,u) &= a(u_0,u_0) + 2a(u_0,\bar{u}) + a(\bar{u},\bar{u}) \\ &\geq (1-\theta)a(u_0,u_0) + \Big(1 - \frac{1}{\theta}\Big)a(\bar{u},\bar{u}) \\ &\geq (1-\theta)a(u_0,u_0) + \Big(C_a - \frac{C_a}{\theta}\Big)\|\bar{u}\|_V^2 \end{aligned}$$

for all $0 < \theta < 1$. Since $u_0 \in \operatorname{Ker}\mathtt{B}$, we have $\|\mathtt{B}u\|_Q^2 = \|\mathtt{B}\bar{u}\|_Q^2$. Moreover, using (20) yields $\|\mathtt{B}\bar{u}\|_Q^2 \geq \beta^{-2}\|\bar{u}\|_V^2$. Thus

$$a(u,u) + 2\mu\|\mathtt{B}u\|_Q^2 \geq (1-\theta)a(u_0,u_0) + \Big(C_a - \frac{C_a}{\theta} + \frac{2\mu}{\beta^2}\Big)\|\bar{u}\|_V^2. \tag{29}$$

Choosing $\theta = \frac{1}{2}$ and $\mu = \frac{3}{4}C_a\beta^2$, it holds that

$$a(u,u) + 2\mu\|\mathtt{B}u\|_Q^2 \geq \frac{1}{2}a(u_0,u_0) + \frac{C_a}{2}\|\bar{u}\|_V^2.$$

Hence, recalling (27) and using the inequality $C_a \geq \alpha_0$, we obtain

$$a(u,u) + 2\mu\|\mathtt{B}u\|_Q^2 \geq \frac{\alpha_0}{2}\|u_0\|_V^2 + \frac{\alpha_0}{2}\|\bar{u}\|_V^2 = \frac{\alpha_0}{2}\|u\|_V^2$$

and we conclude that

$$\mathscr{A}\big((u,p),\mathtt{T}(u,p)\big) \geq \alpha_0\frac{C_a\beta^2}{4}\|u\|_V^2 + \frac{1}{2}\|p\|_Q^2. \tag{30}$$

From the above, we have

$$\mathtt{T}(u,p) = \big(C_a\beta^2 u + \mathtt{L}_B p, -C_a\beta^2 p + \frac{3}{4}(C_a\beta^2)^2\mathtt{B}u\big).$$

Finally, $\mathtt{T}$ belongs to $\mathscr{L}(V\times Q)$ since $\|\mathtt{L}_B p\|_V \leq \beta\|p\|_Q$ (see (21)) and[2]

$$\|\mathtt{B}u\|_Q \leq C_b\|u\|_V. \tag{31}$$

The stability estimate (28) is then given by (4).

2. Conversely, suppose that there exist $\alpha_V > 0$, $\alpha_Q > 0$ and $\mathtt{T} \in \mathscr{L}(V\times Q)$ such that

$$\mathscr{A}\big((u,p),\mathtt{T}(u,p)\big) \geq \alpha_V\|u\|_V^2 + \alpha_Q\|p\|_Q^2, \quad \forall (u,p) \in V\times Q. \tag{32}$$

Noting $\mathtt{T} : (u,p) \mapsto \big(\mathtt{T}_V(u,p), \mathtt{T}_Q(u,p)\big)$, we have

$$\mathscr{A}\big((u,p),\mathtt{T}(u,p)\big) = a\big(u,\mathtt{T}_V(u,p)\big) + b\big(\mathtt{T}_V(u,p),p\big) + b\big(u,\mathtt{T}_Q(u,p)\big)$$

and, since $\mathtt{T}$ is bounded,

$$\|\mathtt{T}_V(u,p)\|_V^2 + \|\mathtt{T}_Q(u,p)\|_Q^2 \leq |||\mathtt{T}|||^2\big(\|u\|_V^2 + \|p\|_Q^2\big). \tag{33}$$

Now, choosing $u = 0$ in (32) and (33) yields

$$b\big(\mathtt{T}_V(0,p),p\big) \geq \alpha_Q\|p\|_Q^2 \quad \text{and} \quad \|\mathtt{T}_V(0,p)\|_V \leq |||\mathtt{T}|||\cdot\|p\|_Q, \quad \forall p \in Q.$$

Thus, for $p \in Q\setminus\{0\}$, $\mathtt{T}_V(0,p) \neq 0$, otherwise $b\big(\mathtt{T}_V(0,p),p\big) = 0$, which contradicts $b\big(\mathtt{T}_V(0,p),p\big) > 0$. Then it follows that

$$\sup_{v\in V\setminus\{0\}}\frac{b(v,p)}{\|v\|_V} \geq \frac{b\big(\mathtt{T}_V(0,p),p\big)}{\|\mathtt{T}_V(0,p)\|_V} \geq \frac{\alpha_Q}{|||\mathtt{T}|||}\|p\|_Q, \quad \forall p \in Q\setminus\{0\},$$

which shows that the inf-sup condition (18) is fulfilled. Likewise, taking $p = 0$ and $u \in \operatorname{Ker}\mathtt{B}$ in (32) and (33), we get

$$a\big(u,\mathtt{T}_V(u,0)\big) \geq \alpha_V\|u\|_V^2 \quad \text{and} \quad \|\mathtt{T}_V(u,0)\|_V \leq |||\mathtt{T}|||\,\|u\|_V, \quad \forall u \in \operatorname{Ker}\mathtt{B}.$$

By symmetry and positivity of a, it holds that

$$a\big(u,\mathtt{T}_V(u,0)\big) \leq \big(a(u,u)\big)^{1/2}a\big(\mathtt{T}_V(u,0),\mathtt{T}_V(u,0)\big)^{1/2}.$$

[2] Classically,

$$\begin{aligned}\|\mathtt{B}u\|_Q^2 &= (\mathtt{B}u,\mathtt{B}u)_Q = \langle \mathbb{1}_{Q\to Q'}(\mathtt{B}u),\mathtt{B}u\rangle_{Q',Q} \quad \text{by definition of } \mathbb{1}_{Q\to Q'},\\ &= \langle \mathbb{1}_{Q\to Q'}\circ\mathbb{1}_{Q'\to Q}(Bu),\mathtt{B}u\rangle_{Q',Q} = \langle Bu,\mathtt{B}u\rangle_{Q',Q} \quad \text{since } \mathbb{1}_{Q\to Q'}\circ\mathbb{1}_{Q'\to Q} = \mathrm{Id}_{Q'},\\ &= b(u,\mathtt{B}u) \leq C_b\|u\|_V\|\mathtt{B}u\|_Q \quad \text{by definition and continuity of } b \text{ (14)}.\end{aligned}$$

Thus

$$\alpha_V \|u\|_V^2 \le a\big(u, \mathtt{T}_V(u,0)\big) \le \big(a(u,u)\big)^{1/2} \big(C_a |||\mathtt{T}|||^2 \|u\|_V^2\big)^{1/2}$$

and hence $a(u,u) \ge \frac{\alpha_V^2}{C_a |||\mathtt{T}|||^2} \|u\|_V^2$ for all $u \in \mathrm{Ker}\,\mathtt{B}$, which proves (27). □

Remark 9. The T-coercivity estimate (30) is very close to the case where a is coercive on the whole space V. As a matter of fact, the only difference compared to (24) is that the constant before the term $\|u\|_V^2$ is twice as small, with $\alpha_0 = \alpha$.

3.3. *Augmented saddle-point problems*

Let $c(\cdot,\cdot)$ be a positive and continuous bilinear form defined on $Q \times Q$, namely

$$c(p,p) \ge 0, \quad \forall p \in Q \quad \text{and} \quad \exists C_c > 0, \quad c(p,q) \le C_c \|p\|_Q \|q\|_Q, \quad \forall p \in Q, \forall q \in Q. \tag{34}$$

We denote by C the linear operator associated with the bilinear form c, defined by

$$C \in \mathscr{L}(Q, Q'), \quad \langle Cp, q\rangle_{Q',Q} = c(p,q), \quad \forall p \in Q, \forall q \in Q.$$

The *all-in-one* approach developed previously also enables us to deal with augmented saddle-point problems: given $f \in V'$ and $g \in Q'$, find $(u,p) \in V \times Q$ such that

$$\begin{aligned} Au + B^* p &= f, \quad \text{in } V', \\ Bu - Cp &= g, \quad \text{in } Q', \end{aligned} \tag{35}$$

where the operator C possibly acts as a *small perturbation* of the original saddle-point problem (15). The weak formulation of (35) reads:

$$\begin{cases} \text{Find } (u,p) \in V \times Q \quad \text{such that} \\ \forall (v,q) \in V \times Q, \quad \mathscr{A}_c\big((u,p),(v,q)\big) = \langle f, v\rangle_{V',V} + \langle g, q\rangle_{Q',Q}, \end{cases} \tag{36}$$

with

$$\mathscr{A}_c\big((u,p),(v,q)\big) = a(u,v) + b(v,p) + b(u,q) - c(p,q).$$

As before, the bilinear form a is supposed to be symmetric and positive.

3.4. *How to achieve* T*-coercivity for augmented saddle-point problems?*

Once again, we distinguish the case where the form a is coercive on V or only on $\mathrm{Ker}\,\mathtt{B}$. If the form a is coercive on V, the results from the un-augmented case allow straightforwardly to handle the augmented one.

Theorem 10. *Assume that (18) holds true and that the form a is symmetric and positive. If there exists a constant $\alpha > 0$ such that*

$$a(u,u) \ge \alpha \|u\|_V^2, \quad \forall u \in V,$$

then there exists a unique solution to Problem (36).

Proof. With the same operator T as for the un-augmented problem, namely

$$\begin{aligned} \mathtt{T} : V \times Q &\longrightarrow V \times Q \\ (u,p) &\longmapsto \big(C_a \beta^2 u + \mathtt{L}_B p, -C_a \beta^2 p\big), \end{aligned}$$

it holds that

$$\mathscr{A}_c\big((u,p), \mathtt{T}(u,p)\big) = C_a \beta^2 a(u,u) + a(u, \mathtt{L}_B p) + \|p\|_Q^2 + C_a \beta^2 c(p,p).$$

Therefore, a similar argument as in Theorem 6 furnishes

$$\mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big) \geq \alpha \frac{C_a\beta^2}{2}\|u\|_V^2 + \frac{1}{2}\|p\|_Q^2 + C_a\beta^2 c(p,p),$$

which shows that $\mathscr{A}_c$ is T-coercive since c is positive. □

Remark 11. A particular case that appears in many applications – see Section 5 for the example of nearly-incompressible elasticity – is when c has the form

$$c(p,q) = \varepsilon(p,q)_Q, \quad \varepsilon \geq 0.$$

In this case, we obtain the estimate

$$\mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big) \geq \alpha \frac{C_a\beta^2}{2}\|u\|_V^2 + \left(\frac{1}{2} + \varepsilon C_a\beta^2\right)\|p\|_Q^2, \tag{37}$$

so that the augmentation c improves the constant before the term $\|p\|_Q^2$ and thus stabilizes the bilinear form $\mathscr{A}_c$. Moreover, the above estimate is robust for small values of ε. Besides, it even allows to take negative values of ε. Indeed, if $\varepsilon < 0$, we have

$$\mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big) \geq \alpha \frac{C_a\beta^2}{2}\|u\|_V^2 + \left(\frac{1}{2} - |\varepsilon|\, C_a\beta^2\right)\|p\|_Q^2.$$

Hence, the bilinear form $\mathscr{A}_c$ remains T-coercive whenever $|\varepsilon| < \frac{1}{2C_a\beta^2}$.

Let us now suppose that a is not coercive on the whole space V but only on the kernel of B. Then, two different situations occur. Either the form c can be viewed as a *small perturbation*, and we shall look for a solution of (35) that is *close* to the solution of the original problem (15). Or this is not the case, and the form c is viewed as a "fixed" augmentation, and there is no obvious connection *a priori* between the solutions of the augmented and un-augmented problems.

3.5. *Additional results for small perturbations*

We say that c is a small perturbation if it can be written as

$$c(p,q) = \varepsilon c_0(p,q), \quad \varepsilon > 0, \tag{38}$$

with ε a small parameter and c_0 a symmetric, positive and continuous form on Q. We start with the simple case

$$c(p,q) = \varepsilon(p,q)_Q, \quad \varepsilon > 0, \tag{39}$$

for which the T-coercivity approach yields a shorter proof than the corresponding result stated in Ladyzhenskaya–Babuška–Brezzi's framework, see [4, pages 247-252].

Theorem 12. *Assume that (18) holds true, that the form a is symmetric and positive, and that c takes the simple form of* (39)*. If there exists a constant $\alpha_0 > 0$ such that*

$$a(u_0,u_0) \geq \alpha_0\|u_0\|_V^2, \quad \forall u_0 \in \operatorname{Ker}\mathtt{B}, \tag{40}$$

and if ε is small enough*, namely*

$$\varepsilon \leq \frac{1}{2C_a\beta^4C_b^2}\left(2 - \frac{\alpha_0}{C_a}\right), \tag{41}$$

then Problem (36) *is well-posed and*

$$\|(u,p)\|_{V\times Q} \leq C\|(f,g)\|_{V'\times Q'}, \tag{42}$$

with C a constant depending only on α_0, β, C_a and C_b.

Proof. Here again, we consider the mapping

$$\begin{aligned} \mathtt{T} : V \times Q &\longrightarrow V \times Q \\ (u,p) &\longmapsto \big(\lambda u + \mathtt{L}_B p, -\lambda p + \lambda\mu \mathtt{B} u\big). \end{aligned}$$

The beginning of the proof is the same as in Theorem 8. Taking into account the extra terms coming from the perturbation, we get

$$\mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big) = \lambda a(u,u) + a(u,\mathtt{L}_B p) + \|p\|_Q^2 + \lambda\mu\|Bu\|_{Q'}^2 + \lambda c(p,p) - \lambda\mu c(p,\mathtt{B}u).$$

Using Young's inequality and setting $\lambda = C_a\beta^2$, it follows that

$$\mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big) \geq \frac{C_a\beta^2}{2}\big(a(u,u) + 2\mu\|Bu\|_{Q'}^2\big) + \frac{1}{2}\|p\|_Q^2 + C_a\beta^2 c(p,p) - C_a\beta^2\mu c(p,\mathtt{B}u). \tag{43}$$

Now, as in (29), it holds that

$$a(u,u) + 2\mu\|Bu\|_{Q'}^2 \geq (1-\theta)\,a(u_0,u_0) + \Big(C_a - \frac{C_a}{\theta} + \frac{2\mu}{\beta^2}\Big)\|\bar{u}\|_V^2 \tag{44}$$

for all $0<\theta<1$, where $u = u_0 + \bar{u}$ with $u_0 \in \mathrm{Ker}\,\mathtt{B}$ and $\bar{u} \in (\mathrm{Ker}\,\mathtt{B})^{\perp}$.

Knowing that $c(p,q) = \varepsilon(p,q)_Q$ for all p and q in Q, Young's inequality implies that, for all $\delta > 0$,

$$\begin{aligned} -c(p,\mathtt{B}u) = -c(p,\mathtt{B}\bar{u}) = -\varepsilon(p,\mathtt{B}\bar{u})_Q &\geq -\varepsilon\frac{\delta}{2}\|p\|_Q^2 - \frac{\varepsilon}{2\delta}\|\mathtt{B}\bar{u}\|_Q^2 \\ &\geq -\varepsilon\frac{\delta}{2}\|p\|_Q^2 - \varepsilon\frac{C_b^2}{2\delta}\|\bar{u}\|_V^2 \quad \text{in view of (31).} \end{aligned}$$

Putting (43), (44) and the above inequality together, we find that

$$\begin{aligned} \mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big) \geq \frac{C_a\beta^2}{2}\Big((1-\theta)\,a(u_0,u_0) + \Big(C_a - \frac{C_a}{\theta} + \frac{2\mu}{\beta^2} - \mu\varepsilon\frac{C_b^2}{\delta}\Big)\|\bar{u}\|_V^2\Big) \\ + \frac{1}{2}\|p\|_Q^2 + \varepsilon C_a\beta^2\Big(1 - \mu\frac{\delta}{2}\Big)\|p\|_Q^2. \end{aligned}$$

Hence, choosing $\theta = \frac{1}{2}$, $\mu = C_a\beta^2$ and recalling (40), it holds that

$$\mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big) \geq \frac{C_a\beta^2}{2}\Big(\frac{\alpha_0}{2}\|u_0\|_V^2 + C_a\Big(1 - \varepsilon\frac{\beta^2 C_b^2}{\delta}\Big)\|\bar{u}\|_V^2\Big) + \frac{1}{2}\|p\|_Q^2 + \varepsilon C_a\beta^2\Big(1 - C_a\beta^2\frac{\delta}{2}\Big)\|p\|_Q^2. \tag{45}$$

Finally, we set $\delta = \frac{1}{C_a\beta^2}$ so that

$$1 - C_a\beta^2\frac{\delta}{2} = \frac{1}{2} \qquad \text{and} \qquad 1 - \varepsilon\frac{\beta^2 C_b^2}{\delta} = 1 - \varepsilon C_a\beta^4 C_b^2 \geq \frac{1}{2}\cdot\frac{\alpha_0}{C_a}$$

in virtue of (41). Thus

$$\mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big) \geq \alpha_0\frac{C_a\beta^2}{4}\|u\|_V^2 + \Big(\frac{1}{2} + \varepsilon\frac{C_a\beta^2}{2}\Big)\|p\|_Q^2, \tag{46}$$

where we used that $\|u\|_V^2 = \|u_0\|_V^2 + \|\bar{u}\|_V^2$. All in all, we have chosen

$$\mathtt{T}(u,p) = \big(C_a\beta^2 u + \mathtt{L}_B p, -C_a\beta^2 p + (C_a\beta^2)^2\mathtt{B}u\big).$$

Then, estimate (42) follows from (4) with a stability constant independent of ε since (46) is robust for vanishing ε and since $|||\mathtt{T}|||$ does not depend on ε either. □

Remark 13. The final estimate (46) is very close to (37). The only difference between these two estimates is a factor of 2 between the constants multiplying the norms of u and p, with $\alpha_0 = \alpha$.

Remark 14. In Ladyzhenskaya–Babuška–Brezzi's framework, it is commonly assumed that $\varepsilon \leq 1$. On the other hand, in (41), we find a smallness condition that depends *explicitly* on the various constants of the problem.

Remark 15. The inf-sup condition (18) and the continuity of b imply that $\underline{\beta} \leq C_b$, *i.e.* $C_b\beta \geq 1$. Therefore, (41) yields in particular

$$\varepsilon \leq \frac{1}{C_a\beta^2},$$

which corresponds to the condition found in Remark 11 for negative values of ε. As a matter of fact, the non-coercivity of a on the whole space V calls for the introduction of a term $\mathtt{B}u$ in the mapping $\mathtt{T}$. This term induces an additional term of the form $c(p,\mathtt{B}u)$ in the expression of $\mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big)$, that can be interpreted as a "negative perturbation" of the bilinear form $\mathscr{A}$.

Now, we move to the case where c is given by (38). Let us denote by C_{c_0} the continuity constant of the bilinear form c_0. The next theorem establishes the well-posedness of the perturbed problem for a very general form c_0.

Theorem 16. *Assume that (18) holds true, and that the bilinear forms a and c_0 are both symmetric and positive. Suppose in addition that there exist $\alpha_0 > 0$ and $\gamma_0 > 0$ such that*

$$a(u_0,u_0) \geq \alpha_0 \|u_0\|_V^2, \quad \forall\, u_0 \in \mathrm{Ker}\,\mathtt{B},$$

and

$$c_0(p_0,p_0) \geq \gamma_0 \|p_0\|_Q^2, \quad \forall\, p_0 \in \mathrm{Ker}\,\mathtt{B}^*. \tag{47}$$

If ε is small enough*, namely*

$$\varepsilon \leq \frac{1}{2C_{c_0}C_a\beta^4C_b^2}, \tag{48}$$

then Problem (36) *is well-posed and*

$$\|(u,p)\|_{V\times Q} \leq C\|(f,g)\|_{V'\times Q'},$$

with C a constant depending only on α_0, β, γ_0, C_a and C_b.

Proof. First, we adapt the beginning of the proof of Theorem 12 to take into consideration the bilinear form c_0. Since c_0 is symmetric and positive, we can use Young's inequality to obtain

$$\begin{aligned}
-c(p,\mathtt{B}u) = -\varepsilon c_0(p,\mathtt{B}\bar{u})_Q &\geq -\varepsilon\frac{\delta}{2}c_0(p,p) - \frac{\varepsilon}{2\delta}c_0(\mathtt{B}\bar{u},\mathtt{B}\bar{u}) \\
&\geq -\varepsilon\frac{\delta}{2}c_0(p,p) - \varepsilon\frac{C_{c_0}C_b^2}{2\delta}\|\bar{u}\|_V^2 \quad \text{since } \|\mathtt{B}\bar{u}\|_Q^2 \leq C_b^2\|\bar{u}\|_V^2,
\end{aligned}$$

and thus (45) becomes

$$\begin{aligned}
\mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big) \geq \frac{C_a\beta^2}{2}&\left(\frac{\alpha_0}{2}\|u_0\|_V^2 + C_a\left(1-\varepsilon\frac{C_{c_0}\beta^2C_b^2}{\delta}\right)\|\bar{u}\|_V^2\right) \\
&+ \frac{1}{2}\|p\|_Q^2 + \varepsilon C_a\beta^2\left(1 - C_a\beta^2\frac{\delta}{2}\right)c_0(p,p),
\end{aligned}$$

where $\mathtt{T}$ is the mapping

$$\begin{aligned}
\mathtt{T} : V\times Q &\longrightarrow V\times Q \\
(u,p) &\longmapsto \big(C_a\beta^2u + \mathtt{L}_Bp, -C_a\beta^2p + (C_a\beta^2)^2\mathtt{B}u\big).
\end{aligned}$$

Setting $\delta = \frac{1}{C_a\beta^2}$ as before, we get the estimate

$$\mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big) \geq \alpha_0\frac{C_a\beta^2}{4}\|u\|_V^2 + \frac{1}{2}\|p\|_Q^2 + \varepsilon\frac{C_a\beta^2}{2}c_0(p,p),$$

as long as $\varepsilon \leq \frac{1}{2C_{c_0}C_a\beta^4C_b^2}(2-\frac{\alpha_0}{C_a})$, which is the case under the assumption (48) since $\alpha_0 \leq C_a$.

Then, as the bilinear form c_0 is not necessarily coercive on the whole space Q, we use the decomposition $p = p_0 + \bar{p}$ with $p_0 \in \mathrm{Ker}\,\mathtt{B}^*$ and $\bar{p} \in (\mathrm{Ker}\,\mathtt{B}^*)^\perp$. From Young's inequality, we have

$$c_0(p,p) \geq (1-\theta)c_0(p_0,p_0) + \left(C_{c_0} - \frac{C_{c_0}}{\theta}\right)\|\bar{p}\|_Q^2$$

for all $0<\theta<1$. Setting $\theta=\frac{1}{2}$ and using (47), it follows that

$$\mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big)\geq\alpha_0\frac{C_a\beta^2}{4}\|u\|_V^2+\frac{1}{2}\|p\|_Q^2+\varepsilon\frac{C_a\beta^2}{2}\left(\frac{\gamma_0}{2}\|p_0\|_Q^2-C_{c_0}\|\bar{p}\|_Q^2\right).$$

Now, we notice that

$$\frac{1}{2}\|p\|_Q^2\geq\frac{1}{8}\|p\|_Q^2+\frac{3}{8}\|\bar{p}\|_Q^2$$

and that, thanks to (48),

$$\frac{3}{8}\|\bar{p}\|_Q^2=\frac{1}{2}\cdot\frac{3}{4}\|\bar{p}\|_Q^2\geq\varepsilon C_{c_0}C_a\beta^4C_b^2\cdot\frac{3}{4}\|\bar{p}\|_Q^2\geq\varepsilon\frac{C_a\beta^2}{2}\cdot\frac{3}{2}C_{c_0}\|\bar{p}\|_Q^2$$

because $C_b\beta\geq1$. Hence

$$\begin{aligned}\mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big)&\geq\alpha_0\frac{C_a\beta^2}{4}\|u\|_V^2+\frac{1}{8}\|p\|_Q^2+\varepsilon\frac{C_a\beta^2}{2}\left(\frac{\gamma_0}{2}\|p_0\|_Q^2+\left(\frac{3}{2}C_{c_0}-C_{c_0}\right)\|\bar{p}\|_Q^2\right)\\&\geq\alpha_0\frac{C_a\beta^2}{4}\|u\|_V^2+\left(\frac{1}{8}+\varepsilon\gamma_0\frac{C_a\beta^2}{4}\right)\|p\|_Q^2\quad\text{since } C_{c_0}\geq\gamma_0,\end{aligned}$$

which shows that $\mathscr{A}_c$ is $\mathtt{T}$-coercive. □

Lastly, we mention that an important consequence of the previous result is to estimate the distance between the solution $(u_\varepsilon,p_\varepsilon)$ of the perturbed problem

$$\begin{aligned}Au_\varepsilon+B^*p_\varepsilon&=f,\quad\text{in } V',\\Bu_\varepsilon-\varepsilon C_0p_\varepsilon&=g,\quad\text{in } Q',\end{aligned}\tag{49}$$

and the solution (u,p) of the original saddle-point problem (15) as a function of the penalty parameter ε.

Corollary 17. *Assume that (18) holds true, that the form a is symmetric and positive, and that c takes the form of* (38)*. If there exist $\alpha_0>0$ and $\gamma_0>0$ such that*

$$a(u_0,u_0)\geq\alpha_0\|u_0\|_V^2,\quad\forall u_0\in\operatorname{Ker}\mathtt{B},\qquad c_0(p_0,p_0)\geq\gamma_0\|p_0\|_Q^2,\quad\forall p_0\in\operatorname{Ker}\mathtt{B}^*,$$

and if

$$\varepsilon\leq\frac{1}{2C_{c_0}C_a\beta^4C_b^2},$$

then we have

$$\|u-u_\varepsilon\|_V+\|p-p_\varepsilon\|_Q\leq C\varepsilon,\tag{50}$$

with C a constant depending only on α_0, β, γ_0, C_a, C_b and C_{c_0}.

Proof. Subtracting (49) from (15), we find that $(u-u_\varepsilon,p-p_\varepsilon)$ solves the system

$$\begin{aligned}A(u-u_\varepsilon)+B^*(p-p_\varepsilon)&=0,&&\text{in } V',\\B(u-u_\varepsilon)-\varepsilon C_0(p-p_\varepsilon)&=-\varepsilon C_0p,&&\text{in } Q'.\end{aligned}$$

From Theorem 16, we infer that

$$\|(u-u_\varepsilon,p-p_\varepsilon)\|_{V\times Q}\leq C\|(0,-\varepsilon C_0p)\|_{V'\times Q'}$$

with C depending only on α_0, β, γ_0, C_a and C_b. Thus

$$\|(u-u_\varepsilon,p-p_\varepsilon)\|_{V\times Q}\leq CC_{c_0}\varepsilon\|p\|_Q,$$

which proves (50). □

3.6. *Case of a "fixed" augmentation*

If the bilinear form c is given, the extra terms of the form $c(p,\mathrm{B}u)$ arising from the previously considered $\mathtt{T}$-coercivity operator can not be controlled as before, because there is no factor ε to adjust. Below, we assume that c is coercive on Q, namely that there exists $\gamma>0$ such that

$$c(p,p)\geq\gamma\|p\|_Q^2,\quad\forall p\in Q. \tag{51}$$

So, to control these extra terms, we introduce an operator C^{-1} in the expression of $\mathtt{T}$, where $C^{-1}\in\mathscr{L}(Q',Q)$ is defined by

$$c(C^{-1}g,q)=\langle g,q\rangle_{Q',Q},\quad\forall g\in Q',\forall q\in Q.$$

One can easily check that the operator C^{-1} satisfies

$$(C_c)^{-1}\|g\|_{Q'}\leq\|C^{-1}g\|_Q\leq\gamma^{-1}\|g\|_{Q'},\quad\forall g\in Q',$$

and

$$\big\langle g,C^{-1}g\big\rangle_{Q',Q}\geq\frac{\gamma}{C_c^2}\|g\|_{Q'}^2,\quad\forall g\in Q'. \tag{52}$$

Theorem 18. *Assume that* (51) *holds true and that the bilinear forms a and c are both symmetric and positive. Suppose in addition that there exists a constant $\alpha_B>0$ such that*

$$a(u,u)+\frac{\gamma}{2C_c^2}\|Bu\|_{Q'}^2\geq\alpha_B\|u\|_V^2,\quad\forall u\in V, \tag{53}$$

then Problem (36) *is well-posed and*

$$\|(u,p)\|_{V\times Q}\leq C\|(f,g)\|_{V'\times Q'},$$

with C a constant depending only on α_B, γ and C_b.

Proof. For $\eta,\mu>0$, we consider the mapping

$$\begin{aligned}\mathtt{T}:V\times Q&\longrightarrow V\times Q\\(u,p)&\longmapsto\big(u,-\eta p+\mu C^{-1}(Bu)\big).\end{aligned}$$

Then, using the definitions of C^{-1} and B, we compute

$$\begin{aligned}\mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big)&=a(u,u)+b(u,p)-\eta b(u,p)+\mu b(u,C^{-1}Bu)+\eta c(p,p)-\mu c(p,C^{-1}Bu)\\&=a(u,u)+(1-\eta)b(u,p)+\mu\langle Bu,C^{-1}Bu\rangle_{Q',Q}+\eta c(p,p)-\mu\langle Bu,p\rangle_{Q',Q}\\&=a(u,u)+(1-\eta-\mu)b(u,p)+\mu\langle Bu,C^{-1}Bu\rangle_{Q',Q}+\eta c(p,p).\end{aligned}$$

Let us choose $\eta,\mu>0$ such that $\eta+\mu=1$ to cancel the second term above. To fix ideas, let $\eta=\mu=1/2$, so that

$$\mathtt{T}(u,p)=\Big(u,-\frac{1}{2}p+\frac{1}{2}C^{-1}(Bu)\Big) \tag{54}$$

and

$$\mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big)=a(u,u)+\frac{1}{2}\langle Bu,C^{-1}Bu\rangle_{Q',Q}+\frac{1}{2}c(p,p).$$

Owing to (52) and (51), we deduce that

$$\mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big)\geq a(u,u)+\frac{\gamma}{2C_c^2}\|Bu\|_{Q'}^2+\frac{\gamma}{2}\|p\|_Q^2,$$

and the result follows. □

Remark 19. The $\mathtt{T}$-coercivity estimate reads

$$\mathscr{A}_c\big((u,p),\mathtt{T}(u,p)\big)\geq\alpha_B\|u\|_V^2+\frac{\gamma}{2}\|p\|_Q^2, \tag{55}$$

so that it depends on γ, whereas it was independent of ε in the small perturbation case. Moreover, because of the term $C^{-1}(Bu)$ in (54), $|||\mathtt{T}|||$ behaves as γ^{-1}. Nevertheless, the final stability estimate is robust because the value of the constant γ is fixed.

Remark 20. Note that Theorem 18 does not require the inf-sup condition (18) to be true. However, if (18) holds, then (53) is automatically satisfied. As a matter of fact, for any $u \in V$, using the decomposition $u = u_0 + \bar{u}$ with $u_0 \in \mathrm{Ker}\,\mathtt{B}$ and $\bar{u} \in (\mathrm{Ker}\,\mathtt{B})^{\perp}$, we have seen in the proof of Theorem 8 that, for all $0 < \theta < 1$, it holds

$$a(u,u) \geq (1-\theta)a(u_0,u_0) + \Big(C_a - \frac{C_a}{\theta}\Big)\|\bar{u}\|_V^2 \quad \text{and} \quad \|Bu\|_{Q'}^2 = \|\mathtt{B}\bar{u}\|_Q^2 \geq \beta^{-2}\|\bar{u}\|_V^2.$$

Hence,

$$a(u,u) + \frac{\gamma}{2C_c^2}\|Bu\|_{Q'}^2 \geq (1-\theta)a(u_0,u_0) + \Big(C_a - \frac{C_a}{\theta} + \frac{\gamma}{2C_c^2}\beta^{-2}\Big)\|\bar{u}\|_V^2.$$

We then observe that

$$\Big(C_a - \frac{C_a}{\theta} + \frac{\gamma}{2C_c^2}\beta^{-2}\Big) > 0, \quad \forall \theta \in \left(\Big(1 + \frac{\gamma}{2C_c^2 C_a}\beta^{-2}\Big)^{-1}, 1\right),$$

so (53) is obtained by choosing some $\theta = \theta(C_a, \beta, C_c, \gamma)$ in the above interval.

In addition to the Stokes problem, let us see next how other typical examples of mixed formulations fall within the T-coercivity framework.

4. Application to electromagnetism

Our goal is to solve the so-called quasi-static magnetic problem set in a anisotropic medium, surrounded by a perfect conductor (see [1, Section 6.4]). The medium is characterized by its dielectric permittivity $\underline{\varepsilon}$ and its magnetic permeability $\underline{\mu}$.

Let Ω be the domain of $\mathbb{R}^3$ in which the problem is set. For simplicity, we assume that Ω is simply connected, with a connected boundary. Moreover, we assume that $\xi \in \{\underline{\varepsilon}, \underline{\mu}\}$ satisfy the following assumption:

$$\begin{cases} \xi \text{ is a real-valued, symmetric, measurable tensor field on } \Omega, \\ \exists \xi_-, \xi_+ > 0,\ \forall \boldsymbol{z} \in \mathbb{R}^3,\ \xi_-|\boldsymbol{z}|^2 \leq \xi \boldsymbol{z}\cdot\boldsymbol{z} \leq \xi_+|\boldsymbol{z}|^2 \text{ a.e. in } \Omega. \end{cases} \tag{56}$$

Because one is dealing with symmetric tensors, if ξ fulfills (56), so does ξ^{-1}, with $(\xi^{-1})_+ = (\xi_-)^{-1}$ and $(\xi^{-1})_- = (\xi_+)^{-1}$.

Given $\boldsymbol{H}^\star \in \boldsymbol{L}^2(\Omega)$, such that $\underline{\mu}\boldsymbol{H}^\star \in \boldsymbol{H}_0(\mathrm{div};\Omega) \cap \boldsymbol{H}(\mathrm{div}0;\Omega)$ and $\rho \in H^{-1}(\Omega)$, the quasi-static magnetic problem amounts to finding $\boldsymbol{E} \in \boldsymbol{L}^2(\Omega)$ such that

$$\begin{aligned} \underline{\mu}^{-1}\mathbf{curl}\,\boldsymbol{E} &= \boldsymbol{H}^\star, && \text{in } \Omega, \\ \mathrm{div}(\underline{\varepsilon}\boldsymbol{E}) &= \rho, && \text{in } \Omega, \\ \boldsymbol{E}\times\boldsymbol{n} &= 0, && \text{on } \partial\Omega. \end{aligned} \tag{57}$$

Under the assumptions on $\underline{\varepsilon}$ and $\underline{\mu}$, on the one hand we note that $\boldsymbol{E} \in \boldsymbol{H}_0(\mathbf{curl};\Omega)$. On the other hand, it is known that the problem (57) is well-posed, see for instance [1, Theorem 6.1.4]. Below, we propose to recover well-posedness using the T-coercivity approach.

4.1. *Proving well-posedness with* T*-coercivity*

Classically, the electromagnetic energy is equal to $(\underline{\varepsilon}\boldsymbol{E},\boldsymbol{E})_{\boldsymbol{L}^2(\Omega)} + (\underline{\mu}\boldsymbol{H},\boldsymbol{H})_{\boldsymbol{L}^2(\Omega)}$, where $\boldsymbol{H}$ is the magnetic field. For our problem with the electric field $\boldsymbol{E}$ as the unknown, it can be expressed as $(\underline{\varepsilon}\boldsymbol{E},\boldsymbol{E})_{\boldsymbol{L}^2(\Omega)} + (\underline{\mu}^{-1}\mathbf{curl}\,\boldsymbol{E},\mathbf{curl}\,\boldsymbol{E})_{\boldsymbol{L}^2(\Omega)}$ because $\boldsymbol{H}^\star \sim -\partial_t\boldsymbol{H}$. Indeed, under the assumption (56) made on $\underline{\varepsilon}$ and $\underline{\mu}$, we note that we can endow $\boldsymbol{H}_0(\mathbf{curl};\Omega)$ with the scalar product $(\cdot,\cdot)_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}} : (\boldsymbol{u},\boldsymbol{v}) \mapsto (\underline{\varepsilon}\boldsymbol{u},\boldsymbol{v})_{\boldsymbol{L}^2(\Omega)} + (\underline{\mu}^{-1}\mathbf{curl}\,\boldsymbol{u},\mathbf{curl}\,\boldsymbol{v})_{\boldsymbol{L}^2(\Omega)}$, and the associated scaled norm

$$\|\boldsymbol{u}\|_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}} = \big((\underline{\varepsilon}\boldsymbol{u},\boldsymbol{u})_{\boldsymbol{L}^2(\Omega)} + (\underline{\mu}^{-1}\mathbf{curl}\,\boldsymbol{u},\mathbf{curl}\,\boldsymbol{u})_{\boldsymbol{L}^2(\Omega)}\big)^{1/2}$$

is equivalent to the "natural" norm. We endow $H^1_0(\Omega)$ with the scalar product $(\cdot,\cdot)_{1,\underline{\varepsilon}} : (p,q) \mapsto (\underline{\varepsilon}\nabla p, \nabla q)_{\boldsymbol{L}^2(\Omega)}$, and the associated scaled norm

$$\|q\|_{1,\underline{\varepsilon}} = \big((\underline{\varepsilon}\nabla q, \nabla q)_{\boldsymbol{L}^2(\Omega)}\big)^{1/2}$$

is equivalent to $\|\cdot\|_{H^1(\Omega)}$ according to Poincaré inequality. In this setting, $\mathbb{1}_{H^1_0(\Omega)\to H^{-1}(\Omega)}$ is the isomorphism defined by

$$\langle \mathbb{1}_{H^1_0(\Omega)\to H^{-1}(\Omega)} p, q\rangle_{H^{-1}(\Omega),H^1_0(\Omega)} = (p,q)_{1,\underline{\varepsilon}} = (\underline{\varepsilon}\nabla p, \nabla q)_{\boldsymbol{L}^2(\Omega)}, \quad \forall p,q \in H^1_0(\Omega),$$

while the norm in $H^{-1}(\Omega)$ is

$$\|g\|_{-1,\underline{\varepsilon}^{-1}} = \sup_{q\in H^1_0(\Omega)\setminus\{0\}} \frac{\langle g, q\rangle_{H^{-1}(\Omega),H^1_0(\Omega)}}{\|q\|_{1,\underline{\varepsilon}}}, \quad \forall g \in H^{-1}(\Omega).$$

Bearing in mind that $\mathbf{curl}\,(\nabla p) = 0$, it follows that

$$\|q\|_{1,\underline{\varepsilon}} = \big((\underline{\varepsilon}\nabla q, \nabla q)_{\boldsymbol{L}^2(\Omega)}\big)^{1/2} = \|\nabla q\|_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}}, \quad \forall q \in H^1_0(\Omega).$$

Finally, for $\xi \in \{\underline{\varepsilon}, \underline{\varepsilon}^{-1}, \underline{\mu}, \underline{\mu}^{-1}\}$, we use the scalar product $(\cdot,\cdot)_\xi : (\boldsymbol{u},\boldsymbol{v}) \mapsto (\xi\boldsymbol{u},\boldsymbol{v})_{\boldsymbol{L}^2(\Omega)}$, and the associated scaled norm $\|\cdot\|_\xi$ in $\boldsymbol{L}^2(\Omega)$.

First, for $\boldsymbol{H}^\star \in \boldsymbol{H}_0(\mathrm{div};\Omega) \cap \boldsymbol{H}(\mathrm{div}0;\Omega)$ and $\rho \in H^{-1}(\Omega)$, we observe that the equivalent weak formulation of Problem (57) reads: find $\boldsymbol{E} \in \boldsymbol{H}_0(\mathbf{curl};\Omega)$ such that

$$\begin{aligned} (\underline{\mu}^{-1}\mathbf{curl}\,\boldsymbol{E}, \mathbf{curl}\,\boldsymbol{v})_{\boldsymbol{L}^2(\Omega)} &= (\boldsymbol{H}^\star, \mathbf{curl}\,\boldsymbol{v})_{\boldsymbol{L}^2(\Omega)}, \qquad \forall \boldsymbol{v} \in \boldsymbol{H}_0(\mathbf{curl};\Omega), \\ (\underline{\varepsilon}\boldsymbol{E}, \nabla q)_{\boldsymbol{L}^2(\Omega)} &= -\langle \rho, q\rangle_{H^{-1}(\Omega),H^1_0(\Omega)}, \quad \forall q \in H^1_0(\Omega). \end{aligned}$$

Second, in order to fit (57) into the abstract framework (15), we introduce an artifical pressure unknown $\tilde{p}$ by adding a term $(\underline{\varepsilon}\boldsymbol{v}, \nabla\tilde{p})_{\boldsymbol{L}^2(\Omega)}$ in the first equation. The previous formulation becomes: find $(\boldsymbol{E},\tilde{p}) \in \boldsymbol{H}_0(\mathbf{curl};\Omega) \times H^1_0(\Omega)$ such that

$$\begin{aligned} (\underline{\mu}^{-1}\mathbf{curl}\,\boldsymbol{E}, \mathbf{curl}\,\boldsymbol{v})_{\boldsymbol{L}^2(\Omega)} + (\underline{\varepsilon}\boldsymbol{v}, \nabla\tilde{p})_{\boldsymbol{L}^2(\Omega)} &= (\boldsymbol{H}^\star, \mathbf{curl}\,\boldsymbol{v})_{\boldsymbol{L}^2(\Omega)}, \qquad \forall \boldsymbol{v} \in \boldsymbol{H}_0(\mathbf{curl};\Omega), \\ (\underline{\varepsilon}\boldsymbol{E}, \nabla q)_{\boldsymbol{L}^2(\Omega)} &= -\langle \rho, q\rangle_{H^{-1}(\Omega),H^1_0(\Omega)}, \quad \forall q \in H^1_0(\Omega). \end{aligned} \tag{58}$$

Indeed, one can easily check that $(\boldsymbol{E},\tilde{p})$ is solution of (58) if and only if $\tilde{p} = 0$ and $\boldsymbol{E}$ is solution of (57). So, defining the bilinear forms

$$\begin{aligned} a_{\underline{\mu}^{-1}}(\boldsymbol{u},\boldsymbol{v}) &= (\underline{\mu}^{-1}\mathbf{curl}\,\boldsymbol{u}, \mathbf{curl}\,\boldsymbol{v})_{\boldsymbol{L}^2(\Omega)}, \quad \forall \boldsymbol{u} \in \boldsymbol{H}_0(\mathbf{curl};\Omega), \forall \boldsymbol{v} \in \boldsymbol{H}_0(\mathbf{curl};\Omega), \\ b_{\underline{\varepsilon}}(\boldsymbol{v},q) &= (\underline{\varepsilon}\boldsymbol{v}, \nabla q)_{\boldsymbol{L}^2(\Omega)}, \qquad \forall \boldsymbol{v} \in \boldsymbol{H}_0(\mathbf{curl};\Omega), \forall q \in H^1_0(\Omega), \end{aligned}$$

the *all-in-one* bilinear form of Maxwell problem is given by

$$\mathscr{A}\big((\boldsymbol{E},\tilde{p}),(\boldsymbol{v},q)\big) = a_{\underline{\mu}^{-1}}(\boldsymbol{E},\boldsymbol{v}) + b_{\underline{\varepsilon}}(\boldsymbol{v},\tilde{p}) + b_{\underline{\varepsilon}}(\boldsymbol{E},q). \tag{59}$$

Thanks to the introduction of scaled norms, we find that the bilinear form $a_{\underline{\mu}^{-1}}$ is continuous on $\boldsymbol{H}_0(\mathbf{curl};\Omega) \times \boldsymbol{H}_0(\mathbf{curl};\Omega)$ with a continuity constant $C_a = 1$, while the bilinear form $b_{\underline{\varepsilon}}$ is continuous on $\boldsymbol{H}_0(\mathbf{curl};\Omega) \times H^1_0(\Omega)$ with a continuity constant $C_b = 1$. Besides, we have

$$\big|(\boldsymbol{H}^\star, \mathbf{curl}\,\boldsymbol{v})_{\boldsymbol{L}^2(\Omega)} - \langle \rho, q\rangle_{H^{-1}(\Omega),H^1_0(\Omega)}\big| \le \|\boldsymbol{H}^\star\|_{\underline{\mu}}\|\mathbf{curl}\,\boldsymbol{v}\|_{\underline{\mu}^{-1}} + \|\rho\|_{-1,\underline{\varepsilon}^{-1}}\|q\|_{1,\underline{\varepsilon}},$$

so that $C_\ell \le \big(\|\boldsymbol{H}^\star\|^2_{\underline{\mu}} + \|\rho\|^2_{-1,\underline{\varepsilon}^{-1}}\big)^{1/2}$.

Let us give an explicit expression of the abstract operators

$$\mathtt{B}_{\underline{\varepsilon}} \in \mathscr{L}\big(\boldsymbol{H}_0(\mathbf{curl};\Omega), H^1_0(\Omega)\big), \qquad \mathtt{L}_{B_{\underline{\varepsilon}}} \in \mathscr{L}\big(H^1_0(\Omega), (\mathrm{Ker}\,\mathtt{B}_{\underline{\varepsilon}})^\perp\big)$$

corresponding to this problem. Note that for $\boldsymbol{u} \in \boldsymbol{H}_0(\mathbf{curl};\Omega)$,

$$\mathtt{B}_{\underline{\varepsilon}}\boldsymbol{u} = 0 \Longleftrightarrow (\underline{\varepsilon}\boldsymbol{u}, \nabla q)_{\boldsymbol{L}^2(\Omega)} = 0,\ \forall q \in H^1_0(\Omega) \Longleftrightarrow \mathrm{div}\,(\underline{\varepsilon}\boldsymbol{u}) = 0.$$

Hence,

$$\mathrm{Ker}\,\mathtt{B}_{\underline{\varepsilon}} = \boldsymbol{K}_N(\Omega;\underline{\varepsilon}), \quad \text{where} \quad \boldsymbol{K}_N(\Omega;\underline{\varepsilon}) = \big\{\boldsymbol{v} \in \boldsymbol{H}_0(\mathbf{curl};\Omega),\ \mathrm{div}\,(\underline{\varepsilon}\boldsymbol{v}) = 0\big\}. \tag{60}$$

In addition, one knows that (see (6.16) in [1])

$$(\mathrm{Ker}\,\mathtt{B}_{\underline{\varepsilon}})^{\perp} = \big\{\boldsymbol{v} \in \boldsymbol{H}_0(\mathbf{curl};\Omega),\, \exists q \in H^1_0(\Omega),\ \boldsymbol{v} = \nabla q\big\}. \tag{61}$$

With those results, we can explicit $\mathtt{L}_{B_{\underline{\varepsilon}}}$. On the one hand, by definition of $b_{\underline{\varepsilon}}$, we observe that

$$b_{\underline{\varepsilon}}(\mathtt{L}_{B_{\underline{\varepsilon}}} p, q) = \big(\underline{\varepsilon}(\mathtt{L}_{B_{\underline{\varepsilon}}} p), \nabla q\big)_{\boldsymbol{L}^2(\Omega)}, \quad \forall p, q \in H^1_0(\Omega). \tag{62}$$

On the other hand, according to (19), one has

$$b_{\underline{\varepsilon}}(\mathtt{L}_{B_{\underline{\varepsilon}}} p, q) = (p, q)_{1,\underline{\varepsilon}} = (\underline{\varepsilon}\nabla p, \nabla q)_{\boldsymbol{L}^2(\Omega)}, \quad \forall p, q \in H^1_0(\Omega). \tag{63}$$

Putting (61), (62) and (63) together, we deduce that

$$\mathtt{L}_{B_{\underline{\varepsilon}}} p = \nabla p.$$

Then, for all $p \in H^1_0(\Omega)$, it follows that

$$\|\mathtt{L}_{B_{\underline{\varepsilon}}} p\|^2_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}} = \|\nabla p\|^2_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}} = \|p\|^2_{1,\underline{\varepsilon}},$$

hence $\mathtt{L}_{B_{\underline{\varepsilon}}}$ is an isometry, so $\mathtt{L}_{B_{\underline{\varepsilon}}}$ satisfies (21) with $\beta = 1$, *i.e.* the inf-sup condition (18) holds.

Going back to $\mathrm{Ker}\,\mathtt{B}_{\underline{\varepsilon}}$ (cf. the characterization (60)), we recall Weber inequality [38] (see also [1, Theorem 6.1.4]): there exists $C_K > 1$ such that

$$\|\boldsymbol{k}\|_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}} \leq C_K \|\mathbf{curl}\,\boldsymbol{k}\|_{\underline{\mu}^{-1}}, \quad \forall \boldsymbol{k} \in \boldsymbol{K}_N(\Omega;\underline{\varepsilon}). \tag{64}$$

The fact that $C_K > 1$ stems from the definition of the scaled norms. Hence, Weber inequality (64) says exactly that the form $a_{\underline{\mu}^{-1}}$ is coercive on $\mathrm{Ker}\,\mathtt{B}_{\underline{\varepsilon}}$, so that all the conditions of Theorem 8 are fulfilled, with $\alpha_0 = (C_K)^{-2} < 1$. Precisely, Theorem 8 states that the bilinear form $\mathscr{A}$ is $\mathtt{T}$-coercive for the mapping

$$\begin{aligned} \mathtt{T} : \boldsymbol{H}_0(\mathbf{curl};\Omega) \times H^1_0(\Omega) &\longrightarrow \boldsymbol{H}_0(\mathbf{curl};\Omega) \times H^1_0(\Omega) \\ (\boldsymbol{E}, \tilde{p}) &\longmapsto \Big(\boldsymbol{E} + \nabla\tilde{p}, -\tilde{p} + \frac{3}{4}\phi_{\boldsymbol{E}}\Big), \end{aligned}$$

where $\phi_{\boldsymbol{E}} = \mathtt{B}_{\underline{\varepsilon}}\boldsymbol{E} \in H^1_0(\Omega)$. Note that, for any $\boldsymbol{v} \in \boldsymbol{H}_0(\mathbf{curl};\Omega)$, we have

$$(\mathtt{B}_{\underline{\varepsilon}}\boldsymbol{v}, q)_{1,\underline{\varepsilon}} = b_{\underline{\varepsilon}}(\boldsymbol{v}, q) = (\underline{\varepsilon}\boldsymbol{v}, \nabla q)_{\boldsymbol{L}^2(\Omega)}, \quad \forall q \in H^1_0(\Omega).$$

Therefore, $\mathtt{B}_{\underline{\varepsilon}}\boldsymbol{v}$ is the unique $\phi_{\boldsymbol{v}} \in H^1_0(\Omega)$ satisfying

$$(\underline{\varepsilon}\nabla\phi_{\boldsymbol{v}}, \nabla q)_{\boldsymbol{L}^2(\Omega)} = (\underline{\varepsilon}\boldsymbol{v}, \nabla q)_{\boldsymbol{L}^2(\Omega)}, \quad \forall q \in H^1_0(\Omega). \tag{65}$$

Furthermore, following (30), it holds that

$$\mathscr{A}\big((\boldsymbol{E}, \tilde{p}), \mathtt{T}(\boldsymbol{E}, \tilde{p})\big) \geq \frac{(C_K)^{-2}}{4}\|\boldsymbol{E}\|^2_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}} + \frac{1}{2}\|\tilde{p}\|^2_{1,\underline{\varepsilon}} \geq \underline{\alpha}\big(\|\boldsymbol{E}\|^2_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}} + \|\tilde{p}\|^2_{1,\underline{\varepsilon}}\big),$$

with $\underline{\alpha} = \frac{(C_K)^{-2}}{4}$.

To get the stability constant, we need to compute $|||\mathtt{T}|||$, that is, bound $\|\mathtt{T}(\boldsymbol{E}, \tilde{p})\|_{\boldsymbol{H}_0(\mathbf{curl};\Omega)\times H^1_0(\Omega)}$ for $(\boldsymbol{E}, \tilde{p}) \in \boldsymbol{H}_0(\mathbf{curl};\Omega) \times H^1_0(\Omega)$. We find that

$$\begin{aligned} \|\mathtt{T}(\boldsymbol{E}, \tilde{p})\|^2_{\boldsymbol{H}_0(\mathbf{curl};\Omega)\times H^1_0(\Omega)} &= \|\boldsymbol{E} + \nabla\tilde{p}\|^2_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}} + \Big\|-\tilde{p} + \frac{3}{4}\phi_{\boldsymbol{E}}\Big\|^2_{1,\underline{\varepsilon}} \\ &\leq 2\|\boldsymbol{E}\|^2_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}} + 2\|\nabla\tilde{p}\|^2_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}} + 2\|\tilde{p}\|^2_{1,\underline{\varepsilon}} + 2\cdot\Big(\frac{3}{4}\Big)^2\|\phi_{\boldsymbol{E}}\|^2_{1,\underline{\varepsilon}} \\ &\leq 2\|\boldsymbol{E}\|^2_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}} + 4\|\tilde{p}\|^2_{1,\underline{\varepsilon}} + 2\cdot\Big(\frac{3}{4}\Big)^2\|\boldsymbol{E}\|^2_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}} \\ &\leq 4\big(\|\boldsymbol{E}\|^2_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}} + \|\tilde{p}\|^2_{1,\underline{\varepsilon}}\big), \end{aligned}$$

where we used that $\|\phi_{\boldsymbol{E}}\|_{1,\underline{\varepsilon}} = \|\mathtt{B}_{\underline{\varepsilon}}\boldsymbol{E}\|_{1,\underline{\varepsilon}} \leq \|\boldsymbol{E}\|_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}}$ thanks to (31). Therefore, $|||\mathtt{T}||| \leq 2$.

Applying (4), we conclude that

$$\|\boldsymbol{E}\|_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}} \leq 8C_K^2\big(\|\boldsymbol{H}^\star\|_{\underline{\mu}}^2 + \|\rho\|_{-1,\underline{\varepsilon}^{-1}}^2\big)^{1/2}. \tag{66}$$

4.2. *Optimized bounds*

To achieve T-coercivity, the abstract theory does not take into account the so-called double orthogonality property (or Helmholtz decomposition), which states that for all $\boldsymbol{k} \in \boldsymbol{K}_N(\Omega;\underline{\varepsilon})$ and all $q \in H_0^1(\Omega)$, one has $(\underline{\varepsilon}\boldsymbol{k}, \nabla q)_{\boldsymbol{L}^2(\Omega)} = (\underline{\mu}^{-1}\mathbf{curl}\,\boldsymbol{k}, \mathbf{curl}\,(\nabla q))_{\boldsymbol{L}^2(\Omega)} = 0$, so that

$$\|\boldsymbol{k} + \nabla q\|_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}}^2 = \|\boldsymbol{k}\|_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}}^2 + \|q\|_{1,\underline{\varepsilon}}^2.$$

Indeed, given $\boldsymbol{E} \in \boldsymbol{H}_0(\mathbf{curl};\Omega)$, we note that, with the help of $\phi_{\boldsymbol{E}} \in H_0^1(\Omega)$ solving (65), it holds that $\boldsymbol{k}_{\boldsymbol{E}} = \boldsymbol{E} - \nabla\phi_{\boldsymbol{E}} \in \boldsymbol{K}_N(\Omega;\underline{\varepsilon})$.

We sketch below how one can improve the estimates, see [22] for further details. Let us choose

$$\begin{aligned} \mathtt{T}_{opt} : \boldsymbol{H}_0(\mathbf{curl};\Omega) \times H_0^1(\Omega) &\longrightarrow \boldsymbol{H}_0(\mathbf{curl};\Omega) \times H_0^1(\Omega) \\ (\boldsymbol{E}, \tilde{p}) &\longmapsto \big(\boldsymbol{k}_{\boldsymbol{E}} + \nabla\tilde{p}, \phi_{\boldsymbol{E}}\big). \end{aligned}$$

Thanks to the double orthogonality property, one finds easily that $\mathtt{T}_{opt}$ is an isometry and that

$$\begin{aligned} \mathscr{A}\big((\boldsymbol{E},\tilde{p}), \mathtt{T}_{opt}(\boldsymbol{E},\tilde{p})\big) &= \|\mathbf{curl}\,\boldsymbol{k}_{\boldsymbol{E}}\|_{\underline{\mu}^{-1}}^2 + \|\tilde{p}\|_{1,\underline{\varepsilon}}^2 + \|\phi_{\boldsymbol{E}}\|_{1,\underline{\varepsilon}}^2 \\ &\geq (C_K)^{-2}\big(\|\boldsymbol{E}\|_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}}^2 + \|\tilde{p}\|_{1,\underline{\varepsilon}}^2\big). \end{aligned}$$

Applying (4), we have the optimized stability estimate

$$\|\boldsymbol{E}\|_{\underline{\varepsilon},\underline{\mu}^{-1}\mathbf{curl}} \leq C_K^2\big(\|\boldsymbol{H}^\star\|_{\underline{\mu}}^2 + \|\rho\|_{-1,\underline{\varepsilon}^{-1}}^2\big)^{1/2}. \tag{67}$$

We conclude that, for all possible choices of coefficients $\underline{\varepsilon}$ and $\underline{\mu}$, there is only a factor 8 difference between the stability constant obtained via the abstract T-coercivity approach, see (66), and the optimized stability constant which relies explicitly on the double orthogonality property, see (67). This shows the robustness of the abstract theory.

Remark 21. One can obtain similar results in more general geometries, such as a non-simply-connected domain, or a non-connected boundary.

5. Application to nearly-incompressible elasticity

In this section, we apply the T-coercivity framework to the equations of elasticity, assuming homogeneous Dirichlet boundary conditions. Let $\Omega \subset \mathbb{R}^d$ be a domain, where $2 \leq d \leq 3$. For a prescribed body force $\boldsymbol{f} \in \boldsymbol{H}^{-1}(\Omega)$, we look for the displacement $\boldsymbol{u} \in \boldsymbol{H}^1(\Omega)$ such that

$$\begin{aligned} -\mathrm{div}\big(\sigma(\boldsymbol{u})\big) &= \boldsymbol{f}, \quad \text{in } \Omega, \\ \boldsymbol{u} &= 0, \quad \text{on } \partial\Omega, \end{aligned} \tag{68}$$

where $\sigma(\boldsymbol{u})$ denotes the stress tensor. We assume that it is given by Hooke's law

$$\sigma(\boldsymbol{u}) = 2\mu\,\varepsilon(\boldsymbol{u}) + \lambda(\mathrm{div}\,\boldsymbol{u})\,I,$$

where $\lambda, \mu > 0$ are the Lamé coefficients of the material and $\varepsilon(\boldsymbol{u}) = \frac{1}{2}\big(\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\big)$ is the linearized strain tensor. Thanks to Korn inequality [25], the space $\boldsymbol{H}_0^1(\Omega)$ is here endowed with the scalar product

$$(\boldsymbol{u},\boldsymbol{v}) \longmapsto \int_\Omega \varepsilon(\boldsymbol{u}) : \varepsilon(\boldsymbol{v})\,\mathrm{d}x,$$

whose associated norm $\boldsymbol{u} \mapsto \|\varepsilon(\boldsymbol{u})\|$ is equivalent to the $\boldsymbol{H}^1(\Omega)$-norm in $\boldsymbol{H}^1_0(\Omega)$. Introducing the new unknown $p = \lambda \operatorname{div} \boldsymbol{u}$, the elasticity system (68) can be written in mixed form as follows: find $\boldsymbol{u} \in \boldsymbol{H}^1_0(\Omega)$ and $p \in L^2_0(\Omega)$ such that

$$\begin{aligned} -2\mu \operatorname{div}\big(\varepsilon(\boldsymbol{u})\big) - \nabla p &= \boldsymbol{f}, \quad \text{in } \Omega, \\ \operatorname{div} \boldsymbol{u} - \frac{1}{\lambda} p &= 0, \quad \text{in } \Omega. \end{aligned}$$

Or equivalently, in variational form: find $\boldsymbol{u} \in \boldsymbol{H}^1_0(\Omega)$ and $p \in L^2_0(\Omega)$ such that

$$\begin{aligned} a(\boldsymbol{u}, \boldsymbol{v}) + b(\boldsymbol{v}, p) &= \langle \boldsymbol{f}, \boldsymbol{v} \rangle, \quad \forall \boldsymbol{v} \in \boldsymbol{H}^1_0(\Omega), \\ b(\boldsymbol{u}, q) - \frac{1}{\lambda} c_0(p, q) &= 0, \qquad \forall q \in L^2_0(\Omega), \end{aligned} \tag{69}$$

with

$$a(\boldsymbol{u}, \boldsymbol{v}) = 2\mu \int_\Omega \varepsilon(\boldsymbol{u}) : \varepsilon(\boldsymbol{v}) \, \mathrm{d}x, \quad b(\boldsymbol{v}, q) = \int_\Omega q \operatorname{div} \boldsymbol{v} \, \mathrm{d}x \quad \text{and} \quad c_0(p, q) = \int_\Omega pq \, \mathrm{d}x.$$

For nearly-incompressible materials, the first Lamé coefficient λ goes to infinity, so that λ^{-1} goes to zero. Therefore, (69) can be seen as a small perturbation of Stokes system.

Since the bilinear form a is coercive on the whole space $\boldsymbol{H}^1_0(\Omega)$, we can directly apply Theorem 10 in the special case of Remark 11. The bilinear form a is continuous and coercive, with $C_a = \alpha = 2\mu$. In addition, the bilinear form b is continuous and satisfies the inf-sup condition (18) with $\beta = C_{\mathrm{div}}$ since b is the same form – except to the sign – as for Stokes problem. Then, Theorem 10 furnishes that the *all-in-one* bilinear form $\mathscr{A}_c$ defined by

$$\mathscr{A}_c\big((\boldsymbol{u}, p), (\boldsymbol{v}, q)\big) = 2\mu \int_\Omega \varepsilon(\boldsymbol{u}) : \varepsilon(\boldsymbol{v}) \, \mathrm{d}x + \int_\Omega p \operatorname{div} \boldsymbol{v} \, \mathrm{d}x + \int_\Omega q \operatorname{div} \boldsymbol{u} \, \mathrm{d}x - \frac{1}{\lambda} \int_\Omega pq \, \mathrm{d}x \tag{70}$$

is $\mathtt{T}$-coercive for the mapping

$$\begin{aligned} \mathtt{T} : \boldsymbol{H}^1_0(\Omega) \times L^2_0(\Omega) &\longrightarrow \boldsymbol{H}^1_0(\Omega) \times L^2_0(\Omega) \\ (\boldsymbol{u}, p) &\longmapsto \big(2\mu C^2_{\mathrm{div}} \boldsymbol{u} + \boldsymbol{v}_{-p}, -2\mu C^2_{\mathrm{div}} p\big), \end{aligned} \tag{71}$$

and (37) implies that

$$\mathscr{A}_c\big((\boldsymbol{u}, p), \mathtt{T}(\boldsymbol{u}, p)\big) \geq 2\mu^2 C^2_{\mathrm{div}} \|\varepsilon(\boldsymbol{u})\|^2 + \left(\frac{1}{2} + \frac{2\mu}{\lambda} C^2_{\mathrm{div}}\right) \|p\|^2.$$

Note that this estimate is robust in the incompressible limit, namely for large values of λ.

Finally, replacing ν by 2μ in (11) and using (4), we get that the unique solution of (69) satisfies

$$\|(\boldsymbol{u}, p)\|_{\boldsymbol{H}^1_0(\Omega) \times L^2_0(\Omega)} \leq \frac{2 \max\Big(2\sqrt{2}\mu C^2_{\mathrm{div}}, C_{\mathrm{div}}\big(2 + 4\mu^2 C^2_{\mathrm{div}}\big)^{1/2}\Big)}{\min(4\mu^2 C^2_{\mathrm{div}}, 1 + 4\mu\lambda^{-1} C^2_{\mathrm{div}})} \|\boldsymbol{f}\|_{(\boldsymbol{H}^1_0(\Omega))'},$$

where $\|\cdot\|_{(\boldsymbol{H}^1_0(\Omega))'}$ denotes the dual norm of $\|\varepsilon(\cdot)\|$.

6. Application to neutron diffusion

Let $\Omega \subset \mathbb{R}^d$ be a domain, where $2 \leq d \leq 3$. We consider the neutron diffusion equation with zero flux boundary condition: given a prescribed fission source $S_f \in L^2(\Omega)$, find $u \in H^1(\Omega)$ such that

$$\begin{aligned} -\operatorname{div}(D \nabla u) + \sigma u &= S_f, \quad \text{in } \Omega, \\ u &= 0, \quad \text{on } \partial\Omega, \end{aligned} \tag{72}$$

where u, D, and σ denote respectively the neutron flux, the diffusion coefficient and the macroscopic absorption cross section. It is assumed that the diffusion coefficient D fulfills (56), and that the macroscopic absorption cross section is such that

$$\begin{cases} \sigma \text{ is a real-valued measurable scalar field on } \Omega, \\ \exists \sigma_-, \sigma_+ > 0, \ \sigma_- \leq \sigma \leq \sigma_+ \text{ a.e. in } \Omega. \end{cases} \tag{73}$$

Because $S_f \in L^2(\Omega)$, one has $D\nabla u \in \boldsymbol{H}(\mathrm{div};\Omega)$. This problem can be recast equivalently in mixed form, introducing the auxiliary unknown $\boldsymbol{p} = -D\nabla u$, called the neutron current. It reads: find $(u,\boldsymbol{p}) \in H^1_0(\Omega) \times \boldsymbol{H}(\mathrm{div};\Omega)$ such that

$$\begin{aligned} \mathrm{div}\,\boldsymbol{p} + \sigma u &= S_f, \quad \text{in } \Omega, \\ D^{-1}\boldsymbol{p} + \nabla u &= 0, \quad \text{in } \Omega. \end{aligned} \tag{74}$$

It can be shown that equivalent weak form is: find $(u,\boldsymbol{p}) \in L^2(\Omega) \times \boldsymbol{H}(\mathrm{div};\Omega)$ such that

$$\int_\Omega \left(v\,\mathrm{div}\,\boldsymbol{p} + \sigma u v - D^{-1}\boldsymbol{p}\cdot\boldsymbol{q} + u\,\mathrm{div}\,\boldsymbol{q}\right) \mathrm{d}x = \int_\Omega S_f v\,\mathrm{d}x \quad \forall (v,\boldsymbol{q}) \in L^2(\Omega)\times \boldsymbol{H}(\mathrm{div};\Omega). \tag{75}$$

Remark 22. Among other things, one can recover that the solution $u \in L^2(\Omega)$ from the weak form (75) is such that $u \in H^1(\Omega)$, and that $u = 0$ on $\partial\Omega$.

6.1. *Proving well-posedness with* T*-coercivity*

Defining the bilinear forms

$$\begin{aligned} a_{D^{-1}}(\boldsymbol{p},\boldsymbol{q}) &= (D^{-1}\boldsymbol{p},\boldsymbol{q})_{\boldsymbol{L}^2(\Omega)}, \quad \forall \boldsymbol{p} \in \boldsymbol{H}(\mathrm{div};\Omega), \forall \boldsymbol{q} \in \boldsymbol{H}(\mathrm{div};\Omega), \\ b(\boldsymbol{q},v) &= -(\mathrm{div}\,\boldsymbol{q},v)_{L^2(\Omega)}, \quad \forall \boldsymbol{q} \in \boldsymbol{H}(\mathrm{div};\Omega), \forall v \in L^2(\Omega), \\ c_\sigma(u,v) &= (\sigma u, v)_{L^2(\Omega)}, \quad \forall u \in L^2(\Omega), \forall v \in L^2(\Omega), \end{aligned}$$

the *all-in-one* bilinear form of the diffusion problem is given by

$$\mathscr{A}_c\big((\boldsymbol{p},u),(\boldsymbol{q},v)\big) = a_{D^{-1}}(\boldsymbol{p},\boldsymbol{q}) + b(\boldsymbol{q},u) + b(\boldsymbol{p},v) - c_\sigma(u,v). \tag{76}$$

Here, we are in the case of a "fixed" augmentation, as treated in Section 3.6.

Let us check below that all the conditions of Theorem 18 are fulfilled. First, c_σ is coercive on $L^2(\Omega)$ with $\gamma = \sigma_-$. Then, $a_{D^{-1}}$ fulfills (13) with $C_a = (D_-)^{-1}$, whereas b fulfills (14) with $C_b = 1$. Finally, we look for the condition (53). It is straightforward to check that, for all $\boldsymbol{p} \in \boldsymbol{H}(\mathrm{div};\Omega)$, $B\boldsymbol{p} = \mathtt{B}\boldsymbol{p} = -\mathrm{div}\,\boldsymbol{p}$. Hence

$$\begin{aligned} a_{D^{-1}}(\boldsymbol{p},\boldsymbol{p}) + \frac{\gamma}{2C_c^2}\|B\boldsymbol{p}\|^2 &= (D^{-1}\boldsymbol{p},\boldsymbol{p}) + \frac{\sigma_-}{2\sigma_+^2}\|\mathrm{div}\,\boldsymbol{p}\|^2 \\ &\geq \min\Big((D_+)^{-1}, \frac{\sigma_-}{2\sigma_+^2}\Big)\|\boldsymbol{p}\|^2_{\boldsymbol{H}(\mathrm{div};\Omega)}. \end{aligned}$$

Then, Theorem 18 establishes that the bilinear form $\mathscr{A}_c$ is T-coercive for the mapping (54)

$$\begin{aligned} \mathtt{T} : \boldsymbol{H}(\mathrm{div};\Omega)\times L^2(\Omega) &\longrightarrow \boldsymbol{H}(\mathrm{div};\Omega)\times L^2(\Omega) \\ (\boldsymbol{p},u) &\longmapsto \Big(\boldsymbol{p}, \frac{1}{2}(-u - \sigma^{-1}\mathrm{div}\,\boldsymbol{p})\Big). \end{aligned}$$

Furthermore, using the estimate (55), it holds that

$$\mathscr{A}_c\big((\boldsymbol{p},u),\mathtt{T}(\boldsymbol{p},u)\big) \geq \min\Big((D_+)^{-1}, \frac{\sigma_-}{2\sigma_+^2}\Big)\|\boldsymbol{p}\|^2_{\boldsymbol{H}(\mathrm{div};\Omega)} + \frac{\sigma_-}{2}\|u\|^2 \geq \underline{\alpha}\,\|(\boldsymbol{p},u)\|^2_{\boldsymbol{H}(\mathrm{div};\Omega)\times L^2(\Omega)}, \tag{77}$$

with $\underline{\alpha} = \frac{1}{2}\min\big(2(D_+)^{-1}, \sigma_-(\sigma_+)^{-2}, \sigma_-\big)$.

There remains to estimate $|||\mathtt{T}|||$. One has

$$\begin{aligned} \|\mathtt{T}(\boldsymbol{p},u)\|^2_{\boldsymbol{H}(\mathrm{div};\Omega)\times L^2(\Omega)} &= \|\boldsymbol{p}\|^2_{\boldsymbol{H}(\mathrm{div};\Omega)} + \frac{1}{4}\|-u-\sigma^{-1}\mathrm{div}\,\boldsymbol{p}\|^2_{L^2(\Omega)} \\ &\leq \|\boldsymbol{p}\|^2_{\boldsymbol{H}(\mathrm{div};\Omega)} + \frac{1}{4}\left((1+3)\|u\|^2_{L^2(\Omega)} + \Big(1+\frac{1}{3}\Big)(\sigma_-)^{-2}\|\mathrm{div}\,\boldsymbol{p}\|^2_{L^2(\Omega)}\right) \\ &\leq \Big(1+\frac{1}{3}(\sigma_-)^{-2}\Big)\|(\boldsymbol{p},u)\|^2_{\boldsymbol{H}(\mathrm{div};\Omega)\times L^2(\Omega)}, \end{aligned}$$

so that $|||\mathtt{T}||| \leq \left(1+\frac{1}{3}(\sigma_-)^{-2}\right)^{1/2}$. Applying (4), we conclude that

$$\|(\boldsymbol{p},u)\|_{\boldsymbol{H}(\mathrm{div};\Omega)\times L^2(\Omega)} \leq \frac{2\left(1+\frac{1}{3}(\sigma_-)^{-2}\right)^{1/2}}{\min\left(2(D_+)^{-1},\sigma_-(\sigma_+)^{-2},\sigma_-\right)}\|S_f\|_{L^2(\Omega)}.$$

Remark 23. Some of those computations can be found in [24, 33]. Here, we see them as a consequence of the general result stated in Theorem 18. Note that in [24, 33], the T-coercivity estimate (77) is obtained with a constant $\underline{\alpha}' = \frac{1}{2}\min\left(2(D_+)^{-1},(\sigma_+)^{-1},\sigma_-\right)$, which is very close to $\underline{\alpha}$ since $\sigma_-(\sigma_+)^{-2} = (\sigma_+)^{-1}\cdot\frac{\sigma_-}{\sigma_+}$ and $\frac{\sigma_-}{\sigma_+}\leq 1$.

Remark 24. If one wants to obtain estimates without the bounding factors $\sigma_\pm$ and $D_\pm$, a standard path is to imbed the parameters D and σ into the definition of the norms, like it is done in Section 4. Namely, one chooses the norms:

$$\begin{aligned}
&\|v\|_\sigma = \left((\sigma v,v)_{L^2(\Omega)}\right)^{1/2},\\
&\|\boldsymbol{q}\|_{D^{-1},\sigma^{-1}\mathrm{div}} = \left((D^{-1}\boldsymbol{q},\boldsymbol{q})_{\boldsymbol{L}^2(\Omega)} + (\sigma^{-1}\mathrm{div}\,\boldsymbol{q},\mathrm{div}\,\boldsymbol{q})_{L^2(\Omega)}\right)^{1/2},\\
&\|(\boldsymbol{q},v)\|_V = \left(\|v\|_\sigma^2 + \|\boldsymbol{q}\|^2_{D^{-1},\sigma^{-1}\mathrm{div}}\right)^{1/2}.
\end{aligned}$$

On the one hand, all norms are "fixed" once the parameters are given. On the other hand, one can easily check that the stability constant is now independent of the bounding factors, by using the same mapping T as before.

7. T-coercivity at the discrete level

Previously, we demonstrated the robustness and the flexbility of the T-coercivity approach to study mixed problems at the continuous level. In this section, we are going to see how T-coercivity also enables us to provide a stable discretization of such problems with mixed finite elements. Let us recall the simple results below [19, 20].

Definition 25. *Let W be a Hilbert space, $\mathscr{A}(\cdot,\cdot)$ be a continuous bilinear form over $W\times W$ and $(W_h)_h$ be conforming approximations of W. We say that $\mathscr{A}$ is* uniformly $\mathtt{T}_h$-coercive *if*

$$\exists\alpha^*,\beta^*>0,\ \forall h>0,\ \exists\mathtt{T}_h\in\mathscr{L}(W_h),\quad |\mathscr{A}(u_h,\mathtt{T}_h u_h)|\geq\alpha^*\|u_h\|_W^2,\quad \forall u_h\in W_h,\quad \textit{and}\quad |||\mathtt{T}_h|||\leq\beta^*.$$

Proposition 26. *Let W be a Hilbert space, f be an element of W', $\mathscr{A}(\cdot,\cdot)$ be a continuous bilinear form over $W\times W$ and $(W_h)_h$ be conforming approximations of W. Denote by $\boldsymbol{A}_h\in\mathscr{L}(W_h,W_h')$ the discrete operator associated to $\mathscr{A}_{|W_h}$. The problem*

$$\begin{cases}\textit{Find } u_h\in W_h \quad \textit{such that}\\ \forall v_h\in W_h,\quad \mathscr{A}(u_h,v_h)=\langle f,v_h\rangle\end{cases}$$

is well-posed and $(\boldsymbol{A}_h^{-1})_h$ is uniformly bounded if and only if $\mathscr{A}$ is uniformly $\mathtt{T}_h$-coercive. In that case, denoting by $C_{\mathscr{A}}$ the continuity constant of the bilinear form $\mathscr{A}$, it holds that

$$\|u-u_h\|_W \leq C\inf_{v_h\in W_h}\|u-v_h\|_W, \tag{78}$$

with $C=1+\frac{C_{\mathscr{A}}\beta^*}{\alpha^*}$ independent of h.

Remark 27. Proposition 26 can be extended to the case where the discrete forms $\mathscr{A}_h$ and f_h differs from the continuous forms $\mathscr{A}$ and f. In that case, Céa's lemma (78) becomes

$$\|u-u_h\|_W \leq C\inf_{v_h\in W_h}\left(\|u-v_h\|_W + \mathrm{Cons}_{f,h} + \mathrm{Cons}_{\mathscr{A},h}(v_h)\right),$$

with

$$\mathrm{Cons}_{f,h} = \sup_{v_h\in W_h\setminus\{0\}}\frac{\left|\langle f-f_h,v_h\rangle\right|}{\|v_h\|_W}\quad\text{and}\quad \mathrm{Cons}_{\mathscr{A},h}(v_h) = \sup_{w_h\in W_h\setminus\{0\}}\frac{|(\mathscr{A}-\mathscr{A}_h)(v_h,w_h)|}{\|w_h\|_W},\ \forall v_h\in W_h.$$

As before, we start with the leading example of Stokes problem.

7.1. *Stokes problem*

For a given h, the natural discretization of Problem (5) reads:

$$\begin{cases} \text{Find } (\boldsymbol{u}_h, p_h) \in \boldsymbol{V}_h \times Q_h \quad \text{such that} \\ \forall (\boldsymbol{v}_h, q_h) \in \boldsymbol{V}_h \times Q_h, \quad \mathscr{A}\big((\boldsymbol{u}_h, p_h), (\boldsymbol{v}_h, q_h)\big) = \langle \boldsymbol{f}, \boldsymbol{v}_h \rangle, \end{cases} \tag{79}$$

where $\boldsymbol{V}_h \subset \boldsymbol{H}^1_0(\Omega)$ and $Q_h \subset L^2_0(\Omega)$ are two finite dimensional spaces constituting a *conforming* approximation of $\boldsymbol{H}^1_0(\Omega) \times L^2_0(\Omega)$.

From Proposition 26, we know that Problem (79) is well-posed if and only if $\mathscr{A}$ is uniformly $\mathtt{T}_h$-coercive. To build a suitable mapping $\mathtt{T}_h \in \mathscr{L}(\boldsymbol{V}_h \times Q_h)$, a natural idea is to reproduce the continuous mapping from the proof of Theorem 3

$$\begin{aligned} \mathtt{T} : \boldsymbol{H}^1_0(\Omega) \times L^2_0(\Omega) &\longrightarrow \boldsymbol{H}^1_0(\Omega) \times L^2_0(\Omega) \\ (\boldsymbol{u}, p) &\longmapsto \big(\lambda \boldsymbol{u} + \boldsymbol{v}_p, -\lambda p\big) \end{aligned}$$

at the discrete level. The operator $\mathtt{T}$ above depends on the divergence lifting $\boldsymbol{v}_p \in \boldsymbol{H}^1_0(\Omega)$ of the pressure $p \in L^2_0(\Omega)$ defined by, see (8)-(9),

$$-\operatorname{div} \boldsymbol{v}_p = p \quad \text{and} \quad \|\nabla \boldsymbol{v}_p\| \leq C_{\mathrm{div}} \|p\|.$$

To obtain a similar lifting in the discrete setting, we consider the continuous lifting of the discrete pressure $p_h \in Q_h \subset L^2_0(\Omega)$, namely $\boldsymbol{v}_{p_h} \in \boldsymbol{H}^1_0(\Omega)$ such that

$$-\operatorname{div} \boldsymbol{v}_{p_h} = p_h \quad \text{and} \quad \|\nabla \boldsymbol{v}_{p_h}\| \leq C_{\mathrm{div}} \|p_h\|. \tag{80}$$

This lifting $\boldsymbol{v}_{p_h}$ does not necessarily belong to the discrete space $\boldsymbol{V}_h \subset \boldsymbol{H}^1_0(\Omega)$, so we need an operator $\Pi_h : \boldsymbol{H}^1_0(\Omega) \longrightarrow \boldsymbol{V}_h$ to project it on $\boldsymbol{V}_h$. Therefore, we consider a discrete mapping of the form

$$\begin{aligned} \mathtt{T}_h : \boldsymbol{V}_h \times Q_h &\longrightarrow \boldsymbol{V}_h \times Q_h \\ (\boldsymbol{u}_h, p_h) &\longmapsto \big(\lambda \boldsymbol{u}_h + \Pi_h(\boldsymbol{v}_{p_h}), -\lambda p_h\big). \end{aligned} \tag{81}$$

Now, let us precise under which conditions the bilinear form $\mathscr{A}$ is uniformly $\mathtt{T}_h$-coercive by mimicking the proof of Theorem 3. We compute

$$\mathscr{A}\big((\boldsymbol{u}_h, p_h), \mathtt{T}_h(\boldsymbol{u}_h, p_h)\big) = \nu\lambda \|\nabla \boldsymbol{u}_h\|^2 + \nu \int_\Omega \nabla \boldsymbol{u}_h : \nabla\big(\Pi_h(\boldsymbol{v}_{p_h})\big)\,\mathrm{d}x - \int_\Omega p_h \operatorname{div}(\Pi_h \boldsymbol{v}_{p_h})\,\mathrm{d}x.$$

In order to get a term of the form $\|p_h\|^2$, we assume that

$$\int_\Omega p_h \operatorname{div}(\Pi_h \boldsymbol{v}_{p_h})\,\mathrm{d}x = \int_\Omega p_h \operatorname{div} \boldsymbol{v}_{p_h}\,\mathrm{d}x, \tag{82}$$

so that

$$\mathscr{A}\big((\boldsymbol{u}_h, p_h), \mathtt{T}_h(\boldsymbol{u}_h, p_h)\big) = \nu\lambda \|\nabla \boldsymbol{u}_h\|^2 + \nu \int_\Omega \nabla \boldsymbol{u}_h : \nabla\big(\Pi_h(\boldsymbol{v}_{p_h})\big)\,\mathrm{d}x + \|p_h\|^2$$

in view of (80). Then, for any $\eta > 0$, Young inequality yields

$$\begin{aligned} \int_\Omega \nabla \boldsymbol{u}_h : \nabla\big(\Pi_h(\boldsymbol{v}_{p_h})\big)\,\mathrm{d}x &\geq -\frac{\eta}{2}\|\nabla \boldsymbol{u}_h\|^2 - \frac{1}{2\eta}\|\nabla\big(\Pi_h(\boldsymbol{v}_{p_h})\big)\|^2 \\ &\geq -\frac{\eta}{2}\|\nabla \boldsymbol{u}_h\|^2 - \frac{C^2_{\mathrm{div}} C^2_\pi}{2\eta}\|p_h\|^2 \end{aligned}$$

provided that there exists a constant $C_\pi > 0$, independent of h and of p_h, such that

$$\|\nabla\big(\Pi_h(\boldsymbol{v}_{p_h})\big)\| \leq C_\pi \|\nabla \boldsymbol{v}_{p_h}\|. \tag{83}$$

Hence, it holds that

$$\mathscr{A}\big((\boldsymbol{u}_h, p_h), \mathtt{T}_h(\boldsymbol{u}_h, p_h)\big) \geq \nu\left(\lambda - \frac{\eta}{2}\right)\|\nabla \boldsymbol{u}_h\|^2 + \left(1 - \frac{\nu C^2_{\mathrm{div}} C^2_\pi}{2\eta}\right)\|p_h\|^2.$$

Setting $\eta = \lambda = \nu C_{\mathrm{div}}^2 C_\pi^2$, we obtain

$$\begin{aligned}\mathscr{A}\big((\boldsymbol{u}_h,p_h),\mathtt{T}_h(\boldsymbol{v}_h,p_h)\big) &\geq \frac{\nu^2 C_{\mathrm{div}}^2 C_\pi^2}{2}\|\nabla u_h\|^2 + \frac{1}{2}\|p_h\|^2 \\ &\geq \frac{1}{2}\min(\nu^2 C_{\mathrm{div}}^2 C_\pi^2, 1)\|(u_h,p_h)\|^2_{\boldsymbol{H}_0^1(\Omega)\times L_0^2(\Omega)}. \end{aligned} \tag{84}$$

Moreover, taking into account (83) and mimicking the continuous case (see (11)), we have

$$|||\mathtt{T}_h||| \leq \max\Big(\sqrt{2}\nu C_{\mathrm{div}}^2 C_\pi^2, C_{\mathrm{div}} C_\pi\big(2+\nu^2 C_{\mathrm{div}}^2 C_\pi^2\big)^{1/2}\Big). \tag{85}$$

So, with the help of the operator $\Pi_h : \boldsymbol{H}_0^1(\Omega) \longrightarrow \boldsymbol{V}_h$, we have proven the following result.

Theorem 28. *If there exist a family of operators $(\Pi_h)_h$ and a constant $C_\pi > 0$ such that, for all h,*

$$\int_\Omega q_h \operatorname{div}(\Pi_h \boldsymbol{v})\,\mathrm{d}x = \int_\Omega q_h \operatorname{div}\boldsymbol{v}\,\mathrm{d}x, \quad \forall \boldsymbol{v}\in\boldsymbol{H}_0^1(\Omega), \forall q_h \in Q_h, \tag{86}$$

$$\|\nabla\big(\Pi_h(\boldsymbol{v})\big)\| \leq C_\pi \|\nabla \boldsymbol{v}\|, \quad \forall \boldsymbol{v}\in\boldsymbol{H}_0^1(\Omega), \tag{87}$$

then Problem (79) *is well-posed for all h and*

$$\|(u-u_h,p-p_h)\|_{\boldsymbol{H}_0^1(\Omega)\times L_0^2(\Omega)} \leq C \inf_{(v_h,q_h)\in \boldsymbol{V}_h\times Q_h} \|(u-v_h,p-q_h)\|_{\boldsymbol{H}_0^1(\Omega)\times L_0^2(\Omega)}, \tag{88}$$

with

$$C = 1 + \frac{2\max(\nu,2)\max\Big(\sqrt{2}\nu C_{\mathrm{div}}^2 C_\pi^2, C_{\mathrm{div}} C_\pi\big(2+\nu^2 C_{\mathrm{div}}^2 C_\pi^2\big)^{1/2}\Big)}{\min(\nu^2 C_{\mathrm{div}}^2 C_\pi^2, 1)}.$$

Proof. The previous reasoning shows that the bilinear form $\mathscr{A}$ is uniformly $\mathtt{T}_h$-coercive for the mapping

$$\begin{aligned}\mathtt{T}_h : \boldsymbol{V}_h\times Q_h &\longrightarrow \boldsymbol{V}_h\times Q_h \\ (\boldsymbol{u}_h,p_h) &\longmapsto \big(\nu C_{\mathrm{div}}^2 C_\pi^2 \boldsymbol{u}_h + \Pi_h(\boldsymbol{v}_{p_h}), -\nu C_{\mathrm{div}}^2 C_\pi^2 p_h\big)\end{aligned}$$

as long as the two conditions (82) and (83) are fulfilled for all $p_h \in Q_h$, which is the case if (86) and (87) hold true. The stability estimate (88) then follows by using (84) and (85) in (78). □

The conditions (86) and (87) correspond exactly to the assumptions of an abstract result known as Fortin's lemma [26]. Above, the T-coercivity approach allowed us to recover these two conditions in a fully constructive way. Moreover, we recall that, since the form b fulfills an inf-sup condition (3), those conditions (86)-(87) are equivalent to the so-called *uniform discrete inf-sup condition*

$$\exists \underline{\beta}' > 0, \quad \forall h, \quad \inf_{q_h\in Q_h\setminus\{0\}} \sup_{\boldsymbol{v}_h\in\boldsymbol{V}_h\setminus\{0\}} \frac{\int_\Omega q_h \operatorname{div}\boldsymbol{v}_h\,\mathrm{d}x}{\|\nabla \boldsymbol{v}_h\|\,\|q_h\|} \geq \underline{\beta}',$$

see for instance [27, Lemma II.1.1].
Finally, we recall that, provided there is a basic approximability property (*i.e.* any element of $\boldsymbol{V}\times Q$ can be approximated by a sequence of elements of $(\boldsymbol{V}_h\times Q_h)_h$), the convergence of the discrete solutions to the exact one is a consequence of (88).

7.2. *Approximation of saddle-point problems*

We now derive a *conforming* approximation of the abstract problem (15), starting from the variational expressions (16) or (17), the latter with the form

$$\mathscr{A}\big((u,p),(v,q)\big) = a(u,v) + b(v,p) + b(u,q).$$

So, let $(V_h)_h$, resp. $(Q_h)_h$, be two families of finite dimensional subspaces of V, resp. Q. Starting from (16), the discrete variational formulation writes

$$\begin{cases} \text{Find } (u_h, p_h) \in V_h \times Q_h \quad \text{such that} \\ \forall v_h \in V_h, \quad a(u_h, v_h) + b(v_h, p_h) = \langle f, v_h \rangle_{V',V} \\ \forall q_h \in Q_h, \quad b(u_h, q_h) = \langle g, q_h \rangle_{Q',Q}. \end{cases}$$

while, starting from (17), the *all-in-one* discrete variational formulation writes

$$\begin{cases} \text{Find } (u_h, p_h) \in V_h \times Q_h \quad \text{such that} \\ \forall (v_h, q_h) \in V_h \times Q_h, \quad \mathscr{A}\big((u_h, p_h), (v_h, q_h)\big) = \langle f, v_h \rangle_{V',V} + \langle g, q_h \rangle_{Q',Q}. \end{cases}$$

In abstract form, the *uniform discrete inf-sup condition* writes

$$\exists \underline{\beta}' > 0, \quad \forall h, \quad \inf_{q_h \in Q_h \setminus \{0\}} \sup_{v_h \in V_h \setminus \{0\}} \frac{b(v_h, q_h)}{\|v_h\|_V \|q_h\|_Q} \geq \underline{\beta}'. \tag{89}$$

We suppose that the discrete version of the operator B is the restriction of B to V_h, namely

$$B(V_h) \subset Q_h'. \tag{90}$$

Introducing the discrete operators $\mathtt{B}_h : V_h \to Q_h$ such that for all h,

$$(\mathtt{B}_h v_h, q_h)_Q = b(v_h, q_h), \quad \forall (v_h, q_h) \in V_h \times Q_h,$$

the straightforward discrete counterpart of Theorem 5 is

Theorem 29. *The following three statements are equivalent:*

(i) *There exists $\underline{\beta}' > 0$ such that $b(\cdot,\cdot)$ fulfills the uniform discrete inf-sup condition (89).*

(ii) *For all h, $\mathtt{B}_h : (\mathrm{Ker}\,\mathtt{B}_h)^\perp \to Q_h$ is an* isomorphism*, and*

$$\|\mathtt{B}_h v_h\|_Q \geq \underline{\beta}' \|v_h\|_V, \quad \forall v_h \in (\mathrm{Ker}\,\mathtt{B}_h)^\perp. \tag{91}$$

(iii) *For all h, there exists an isomorphic operator $\mathtt{L}_{B,h} : Q_h \to (\mathrm{Ker}\,\mathtt{B}_h)^\perp$ such that*

$$\mathtt{B}_h(\mathtt{L}_{B,h} q_h) = q_h \quad \textit{and} \quad \|q_h\|_Q \geq \underline{\beta}' \|\mathtt{L}_{B,h} q_h\|_V, \quad \forall q_h \in Q_h. \tag{92}$$

Remark 30. Obviously, this result also holds if the value of the constant in the discrete inf-sup condition depends on h, *i.e.* for each h it holds for some $\underline{\beta}'(h) > 0$, with $\lim_{h\to 0} \underline{\beta}'(h) = 0$. In this case however, getting error estimates can be more intricate.

As mentioned above for the Stokes system, one has the Fortin lemma (cf. [27, Lemma II.1.1]).

Theorem 31. *Assume that the form b fulfills an inf-sup condition (18). The uniform discrete inf-sup condition (89) holds if, and only if, there exist a family of operators $(\Pi_h)_h$, with $\Pi_h : V \longrightarrow V_h$, and a constant $C_\pi > 0$ such that, for all h,*

$$\begin{aligned} & b(\Pi_h v, q_h) = b(v, q_h), \quad \forall v \in V, \ \forall q_h \in Q_h, \\ & \sup_h |||\Pi_h||| \leq C_\pi. \end{aligned} \tag{93}$$

Let us now proceed with the derivation of conditions to ensure that the form $\mathscr{A}$ is uniformly $\mathtt{T}_h$-coercive. As a general rule, the proofs of the results follow very closely the proofs that were given in the exact case. The straightforwardness of the procedure when going from the continuous to the discrete level is one of the main features of the T-coercivity approach. We give next the discrete counterparts of Theorems 6 and 8.

Theorem 32. *Assume that the form a is symmetric and positive, that there exists a constant $\alpha' > 0$ such that*

$$a(u_h, u_h) \geq \alpha' \|u_h\|_V^2, \quad \forall u_h \in V_h, \tag{94}$$

and that the uniform discrete inf-sup condition (89) on the form b holds true.
Then the form $\mathscr{A}$ is uniformly $\mathtt{T}_h$-coercive.

The property (94) is sometimes called the the uniform discrete coercivity property.

Proof. Let h be given. We introduce the mapping

$$\begin{aligned} \mathtt{T}_h : V_h \times Q_h &\longrightarrow V_h \times Q_h \\ (u_h, p_h) &\longmapsto \big(\lambda u_h + \mathtt{L}_{B,h} p_h, -\lambda p_h\big). \end{aligned}$$

We then compute

$$\begin{aligned} \mathscr{A}\big((u_h,p_h),\mathtt{T}_h(u_h,p_h)\big) &= a(u_h,\lambda u_h) + a(u_h,\mathtt{L}_{B,h}p_h) + b(\lambda u_h,p_h) + b(\mathtt{L}_{B,h}p_h,p_h) - b(u_h,\lambda p_h) \\ &= \lambda a(u_h,u_h) + a(u_h,\mathtt{L}_{B,h}p_h) + \|p_h\|_Q^2, \text{ according to (92)-left.} \end{aligned}$$

Because the form a is symmetric and positive, we can apply Young's inequality: for any $\eta > 0$,

$$a(u_h,\mathtt{L}_{B,h}p_h) \geq -\frac{\eta}{2}a(u_h,u_h) - \frac{1}{2\eta}a(\mathtt{L}_{B,h}p_h,\mathtt{L}_{B,h}p_h).$$

According now to (92)-right, we find

$$a(\mathtt{L}_{B,h}p_h,\mathtt{L}_{B,h}p_h) \leq C_a\|\mathtt{L}_{B,h}p_h\|_V^2 \leq C_a(\underline{\beta}')^{-2}\|p_h\|_Q^2.$$

Using assumption (94), if $\lambda - \frac{\eta}{2} > 0$, it follows that

$$\mathscr{A}\big((u_h,p_h),\mathtt{T}_h(u_h,p_h)\big) \geq \alpha'\left(\lambda - \frac{\eta}{2}\right)\|u_h\|_V^2 + \left(1 - \frac{C_a(\underline{\beta}')^{-2}}{2\eta}\right)\|p_h\|_Q^2.$$

Setting $\eta = \lambda = C_a(\underline{\beta}')^{-2}$ as in the exact case, we infer that

$$\mathscr{A}\big((u_h,p_h),\mathtt{T}_h(u_h,p_h)\big) \geq \frac{1}{2}\min(\alpha' C_a(\underline{\beta}')^{-2},1)\|(u_h,p_h)\|_{V\times Q}^2$$

which proves that $\mathscr{A}$ is $\mathtt{T}_h$-coercive, with a $\mathtt{T}$-coercivity constant $\frac{1}{2}\min(\alpha' C_a(\underline{\beta}')^{-2},1) > 0$ that is independent of h.

Since $\mathtt{T}_h(u_h,p_h) = (C_a(\underline{\beta}')^{-2}u_h + \mathtt{L}_{B,h}p_h, -C_a(\underline{\beta}')^{-2}p_h)$, one finds that

$$\begin{aligned} \|\mathtt{T}_h(u_h,p_h)\|_{V\times Q}^2 &\leq 2(C_a(\underline{\beta}')^{-2})^2\|u_h\|_V^2 + 2\|\mathtt{L}_{B,h}p_h\|_V^2 + (C_a(\underline{\beta}')^{-2})^2\|p_h\|_Q^2 \\ &\leq 2(C_a(\underline{\beta}')^{-2})^2\|u_h\|_V^2 + \big(2(\underline{\beta}')^{-2} + (C_a(\underline{\beta}')^{-2})^2\big)\|p_h\|_Q^2, \end{aligned}$$

where the last inequality follows from (92)-right. The bound is valid for all h, which yields

$$\sup_h |||\mathtt{T}_h||| \leq \max\Big(\sqrt{2}C_a(\underline{\beta}')^{-2}, \beta\big(2 + C_a^2(\underline{\beta}')^{-2}\big)^{1/2}\Big),$$

so the form $\mathscr{A}$ is uniformly $\mathtt{T}_h$-coercive. □

Remark 33. As for the Stokes problem, the discrete right-inverse $\mathtt{L}_{B,h}$ is connected to the Fortin operator Π_h. As a matter of fact, if there exists a family of discrete projectors $(\Pi_h)_h$ verifying (93), the operator defined by $\mathtt{L}_{B,h} = \Pi_h(\mathtt{L}_B)$ satisfies (92) with $\underline{\beta}' = (C_\pi\beta)^{-1}$ since for all $q_h \in Q_h$

$$\|\Pi_h(\mathtt{L}_B q_h)\|_V \leq C_\pi\|\mathtt{L}_B q_h\|_V \leq C_\pi\beta\|q_h\|_Q,$$

according to (21). As a consequence, to perform stability estimates at the discrete level using $\mathtt{T}_h$-coercivity, one has only to replace β by $C_\pi\beta$ in the computations done at the continuous level.

Theorem 34. *Assume that the form a is symmetric and positive, that there exists a constant $\alpha_0' > 0$ such that*

$$a(u_{0,h},u_{0,h}) \geq \alpha_0'\|_{0,h}\|_V^2, \quad \forall u_{0,h} \in \mathrm{Ker}\,\mathtt{B}_h, \tag{95}$$

and that the uniform discrete inf-sup condition (89) on the form b holds true.
Then the form $\mathscr{A}$ is uniformly $\mathtt{T}_h$-coercive.

The property (95) is sometimes called the uniform discrete coercivity property on the kernels.

Proof. Let h be given. We consider the mapping

$$\begin{aligned}\mathtt{T}_h : V_h \times Q_h &\longrightarrow V_h \times Q_h\\ (u_h, p_h) &\longmapsto \big(\lambda u_h + \mathtt{L}_{B,h} p_h, -\lambda p_h + \lambda\mu \mathtt{B}_h u_h\big).\end{aligned}$$

As in the proof of Theorem 8, we can compute

$$\mathscr{A}\big((u_h,p_h),\mathtt{T}_h(u_h,p_h)\big) = \lambda a(u_h,u_h) + a(u_h,\mathtt{L}_{B,h}p_h) + \|p_h\|_Q^2 + \lambda\mu\|\mathtt{B}_h u_h\|_Q^2$$

because $b(\mathtt{L}_{B,h}p_h, p_h) = \|p_h\|_Q^2$. Since the form a is symmetric and positive, one may use Young's inequality. By proceeding as in the proof of Theorem 32 and after setting $\lambda = C_a(\underline{\beta}')^{-2}$, we find that

$$\lambda a(u_h,u_h) + a(u_h,\mathtt{L}_{B,h},p_h) + \|p_h\|_Q^2 \geq \frac{1}{2}C_a(\underline{\beta}')^{-2}a(u_h,u_h) + \frac{1}{2}\|p_h\|_Q^2,$$

and

$$\mathscr{A}\big((u_h,p_h),\mathtt{T}_h(u_h,p_h)\big) \geq \frac{1}{2}C_a(\underline{\beta}')^{-2}\big(a(u_h,u_h) + 2\mu\|\mathtt{B}_h u_h\|_Q^2\big) + \frac{1}{2}\|p_h\|_Q^2.$$

Then, we use the decomposition $u_h = u_{0,h} + \bar{u}_h$ with $u_{0,h} \in \operatorname{Ker}\mathtt{B}_h$ and $\bar{u}_h \in (\operatorname{Ker}\mathtt{B}_h)^\perp$. As before, Young's inequality yields

$$a(u_h,u_h) \geq (1-\theta)a(u_{0,h},u_{0,h}) + \Big(C_a - \frac{C_a}{\theta}\Big)\|\bar{u}_h\|_V^2$$

for all $0<\theta<1$. Moreover, $\|\mathtt{B}_h u_h\|_Q^2 = \|\mathtt{B}_h\bar{u}_h\|_Q^2 \geq (\underline{\beta}')^2\|\bar{u}_h\|_V^2$ according to (91), so that

$$a(u_h,u_h) + 2\mu\|\mathtt{B}_h u_h\|_Q^2 \geq (1-\theta)a(u_{0,h},u_{0,h}) + \Big(C_a - \frac{C_a}{\theta} + 2\mu(\underline{\beta}')^2\Big)\|\bar{u}_h\|_V^2.$$

Choosing $\theta = \frac{1}{2}$ and $\mu = \frac{3}{4}C_a(\underline{\beta}')^{-2}$, it holds that

$$\begin{aligned}a(u_h,u_h) + 2\mu\|\mathtt{B}_h u_h\|_Q^2 &\geq \frac{1}{2}a(u_{0,h},u_{0,h}) + \frac{C_a}{2}\|\bar{u}_h\|_V^2\\ &\geq \frac{\alpha_0'}{2}\|u_{0,h}\|_V^2 + \frac{\alpha_0'}{2}\|\bar{u}_h\|_V^2 = \frac{\alpha_0'}{2}\|u_h\|_V^2,\end{aligned}$$

where we used assumption (95) and $C_a \geq \alpha_0'$ on the second line.

Finally, we conclude that

$$\mathscr{A}\big((u_h,p_h),\mathtt{T}_h(u_h,p_h)\big) \geq \frac{1}{4}\alpha_0' C_a(\underline{\beta}')^{-2}\|u_h\|_V^2 + \frac{1}{2}\|p_h\|_Q^2,$$

which yields that $\mathscr{A}$ is $\mathtt{T}_h$-coercive, with a $\mathtt{T}$-coercivity constant $\min(\frac{1}{4}\alpha_0' C_a(\underline{\beta}')^{-2}, \frac{1}{2}) > 0$ that is independent of h.

From the above, we have $\mathtt{T}_h(u_h,p_h) = \big(C_a(\underline{\beta}')^{-2}u_h + \mathtt{L}_{B,h}p_h, -C_a(\underline{\beta}')^{-2}p_h + \frac{3}{4}(C_a(\underline{\beta}')^{-2})^2\mathtt{B}_h u_h\big)$, and, noting that $\|\mathtt{B}_h u_h\|_Q \leq C_b\|u_h\|_V$, one concludes that

$$\sup_h |||\mathtt{T}_h||| \leq \infty,$$

so the form $\mathscr{A}$ is uniformly $\mathtt{T}_h$-coercive. □

Remark 35. Note that replacing $\mathtt{B}u$ by $\mathtt{B}_h u_h$ when going from the continuous operator $\mathtt{T}$ to the discrete operator $\mathtt{T}_h$ is possible because we assumed that B_h is the restriction of B to V_h, see (90). If (90) does not hold, one introduces $\Phi_h : Q \longrightarrow Q_h$ defined by

$$b(v_h, \Phi_h q) = b(v_h, q), \quad \forall q \in Q, \forall v_h \in V_h,$$

like in [4, Proposition 5.1.2]. Then, one has to replace $\mathtt{B}_h u_h$ by $\Phi_h(\mathtt{B}_h u_h)$ in the previous proof, *i.e.*

$$\begin{aligned}\mathtt{T}_h : V_h \times Q_h &\longrightarrow V_h \times Q_h\\ (u_h, p_h) &\longmapsto \Big(C_a(\underline{\beta}')^{-2}u_h + \mathtt{L}_{B,h}p_h, -C_a(\underline{\beta}')^{-2}p_h + \frac{3}{4}(C_a(\underline{\beta}')^{-2})^2\Phi_h(\mathtt{B}u_h)\Big).\end{aligned}$$

Again, provided a basic approximability property holds, that is, any element of $V \times Q$ can be approximated by a sequence of elements of $(V_h \times Q_h)_h$, convergence will follow under the assumptions of Theorem 32 or Theorem 34.

7.3. *Approximation of augmented saddle-point problems*

We now approximate the abstract problem (35), starting from the variational expression (36), with the form

$$\mathscr{A}_c\big((u,p),(v,q)\big) = a(u,v) + b(v,p) + b(u,q) - c(p,q),$$

where $c(\cdot,\cdot)$ is a form defined on $Q \times Q$ that fulfills (34); in particular, $c(\cdot,\cdot)$ is *positive*. So, let again $(V_h)_h$, resp. $(Q_h)_h$, be two families of finite dimensional subspaces of V, resp. Q. The discrete variational formulation writes

$$\begin{cases} \text{Find } (u_h,p_h) \in V_h \times Q_h \quad \text{such that} \\ \forall (v_h,q_h) \in V_h \times Q_h, \quad \mathscr{A}_c\big((u_h,p_h),(v_h,q_h)\big) = \langle f, v_h\rangle_{V',V} + \langle g, q_h\rangle_{Q',Q}, \end{cases}$$

To ensure that the form $\mathscr{A}_c$ is uniformly $\mathtt{T}_h$-coercive, the proofs once more follow very closely those that were given in the exact case. We give next the discrete counterparts of Theorems 10, 12 and 18.

Theorem 36. *Assume that the form a is symmetric, positive, fulfills the uniform discrete coercivity property (94), and that the uniform discrete inf-sup condition (89) on the form b holds true. Then the form $\mathscr{A}_c$ is uniformly $\mathtt{T}_h$-coercive.*

Theorem 37. *Assume that the form a is symmetric and positive, fulfills the uniform discrete coercivity property on the kernels (95), and that the uniform discrete inf-sup condition (89) on the form b holds true. If moreover the form c is like in* (39)*, where ε is* small enough*, namely*

$$\varepsilon \le \frac{1}{2C_a(C_\pi\beta)^4 C_b^2}\Big(2 - \frac{\alpha_0}{C_a}\Big),$$

then the form $\mathscr{A}_c$ is uniformly $\mathtt{T}_h$-coercive.

Theorem 38. *Assume that* (51) *holds true and that the bilinear forms a and c are both symmetric and positive. If there exists a constant $\alpha'_B > 0$ such that*

$$a(u_h,u_h) + \frac{\gamma}{2C_c^2}\|\mathtt{B}_h u_h\|_Q^2 \ge \alpha'_B \|u_h\|_V^2, \quad \forall\, u_h \in V_h, \tag{96}$$

then the form $\mathscr{A}_c$ is uniformly $\mathtt{T}_h$-coercive.

As before, provided a basic approximability property holds, that is, any element of $V \times Q$ can be approximated by a sequence of elements of $(V_h \times Q_h)_h$, convergence will follow under the assumptions of Theorem 36, Theorem 37 or Theorem 38.

7.4. *Applications*

Let us briefly see how the T-coercivity approach can be used to discretize the mixed problems, that is for Stokes, electromagnetism, nearly-incompressible elasticity and finally neutron diffusion. For each problem, we propose one or several possibilities. Note that, since there is a vast litterature on this topic, there is no need to devise new approximation techniques. On the contrary, the simple framework of the T-coercivity approach provides elementary guidelines to help us choose among existing techniques.

In each case, the first step is to choose a *conforming* finite element discretization adapted to the space V under consideration. We assume for simplicity that Ω is a polyhedron for $d = 3$, or a polygon for $d = 2$, so one can use meshes made of simplices for the discretization by finite elements. For $k \geq 1$, $\mathscr{P}_k$ stands for the Lagrange finite elements of order k.

For Stokes and elasticity, we note that the space $\boldsymbol{H}_0^1(\Omega)$ may be approximated using $(\mathscr{P}_k)^d$ finite elements with $k \geq 2$. For electromagnetism, we have to deal with the space $\boldsymbol{H}_0(\mathbf{curl};\Omega)$, which can be discretized using the (first-kind) Nédélec finite elements of order $k \geq 1$, denoted by $\mathscr{N}_k$. Lastly, for neutron diffusion, we have to deal with the space $\boldsymbol{H}(\mathrm{div};\Omega)$, discretized with the help of the Raviart-Thomas elements of order $k \geq 0$, denoted by $\mathscr{RT}_k$. We refer to [4] for details.

The next step is to choose the *conforming* finite element discretization in the space Q in such a way that convergence of the discrete solutions to the exact one is guaranteed.

First, for Stokes and elasticity, and for $k = 2$, setting $Q_h = \mathscr{P}_1$ leads to Fortin operators $\Pi_h^{\mathscr{P}} : \boldsymbol{H}_0^1(\Omega) \longrightarrow (\mathscr{P}_k)^d$ satisfying (86)-(87), or the abstract counterpart (93): the pair $((\mathscr{P}_2)^d, \mathscr{P}_1)$ is called the Taylor-Hood finite element. Then, for electromagnetism and for $k \geq 1$, setting $Q_h = \mathscr{P}_k$ leads to Fortin operators $\Pi_h^{\mathscr{N}} : \boldsymbol{H}_0(\mathbf{curl};\Omega) \longrightarrow \mathscr{N}_k$ satisfying (93). Last, for neutron diffusion and for $k \geq 1$, one only needs to select Q_h in such a way that (90) is fulfilled, namely $\mathrm{div}(\mathscr{RT}_k) \subset Q_h'$. To do so, we set $Q_h = \mathscr{P}_k^{pw}$, for $k \geq 0$, where the superscript ^{pw} stands for piecewise Lagrange finite elements of order k. We again refer to [4] for details and possible extensions, such as the generalized Taylor-Hood elements for Stokes or elasticity, for $k \geq 3$.

On the other hand, as demonstrated earlier, those properties may be also recovered straightforwardly thanks to the T-coercivity approach. As a matter of fact, we can build the discrete operators $\mathtt{T}_h$ similarly as in the continuous case but changing the constant β to take into account (when applicable) the influence of the projection operators, see Remark 33.
For Stokes, we refer to (81).
For electromagnetism, we infer from Theorem 34 that the bilinear form defined in (59) is uniformly $\mathtt{T}_h$-coercive for the mapping

$$\begin{aligned}\mathtt{T}_h : \mathscr{N}_k \times \mathscr{P}_k &\longrightarrow \mathscr{N}_k \times \mathscr{P}_k \\ (\boldsymbol{E}_h, \tilde{p}_h) &\longmapsto \Big((C_{\pi,\mathscr{N}})^2 \boldsymbol{E}_h + \Pi_h^{\mathscr{N}}(\nabla \tilde{p}_h), -(C_{\pi,\mathscr{N}})^2 \tilde{p}_h + \frac{3}{4}(C_{\pi,\mathscr{N}})^4 \phi_{\boldsymbol{E}_h} \Big),\end{aligned}$$

where $\phi_{\boldsymbol{E}_h} \in Q_h$ satisfies the discrete counterpart of (65), namely

$$(\underline{\varepsilon} \nabla \phi_{\boldsymbol{E}_h}, \nabla q_h)_{\boldsymbol{L}^2(\Omega)} = (\underline{\varepsilon} \boldsymbol{E}_h, \nabla q_h)_{\boldsymbol{L}^2(\Omega)}, \quad \forall q_h \in \mathscr{P}_k.$$

For nearly-incompressible elasticity, the bilinear form defined in (70) is uniformly $\mathtt{T}_h$-coercive for the mapping

$$\begin{aligned}\mathtt{T}_h : \mathscr{P}_2 \times \mathscr{P}_1 &\longrightarrow \mathscr{P}_2 \times \mathscr{P}_1 \\ (\boldsymbol{u}_h, p_h) &\longmapsto \big(2\mu (C_{\pi,\mathscr{P}} C_{\mathrm{div}})^2 \boldsymbol{u}_h + \Pi_h^{\mathscr{P}}(\boldsymbol{v}_{-p_h}), -2\mu (C_{\pi,\mathscr{P}} C_{\mathrm{div}})^2 p_h \big),\end{aligned}$$

according to Theorem 36 and (71).
Finally, for neutron diffusion, assuming for simplicity that σ restricted to any simplex is constant, we introduce the discrete mapping

$$\begin{aligned}\mathtt{T}_h : \mathscr{RT}_k \times \mathscr{P}_k^{pw} &\longrightarrow \mathscr{RT}_k \times \mathscr{P}_k^{pw} \\ (\boldsymbol{p}_h, u_h) &\longmapsto \Big(\boldsymbol{p}_h, \frac{1}{2}(-u_h + \sigma^{-1} \mathrm{div}\, \boldsymbol{p}_h) \Big),\end{aligned}$$

and the property $\mathrm{div}(\mathscr{RT}_k) \subset \mathscr{P}_k^{pw}$ guarantees the uniform $\mathtt{T}_h$-coercivity of the bilinear form (76) in virtue of Theorem 38.

All basic approximability properties are established in [4], which guarantees convergence in each case.

8. Conclusion and perspectives

We have demonstrated the flexibility of the T-coercivity approach, here applied to classical linear mixed problems, both for the theoretical study of the problems and for their numerical approximation by finite elements. Let us mention some possible extensions, such as nonconforming discretization methods for Stokes [32], multigroup diffusion [28] or DDM for diffusion [24].
It is our belief that numerous applications can be studied with the T-coercivity approach, both theoretically and numerically. Recent works include application in poromechanics [3], time-harmonic Maxwell's equations with impedance surfaces [35], and the applications listed in [31].

References

[1] F. Assous, P. Ciarlet Jr, S. Labrunie, *Mathematical foundations of computational electromagnetism*, Springer, 2018.
[2] I. Babuška, "The finite element method with Lagrangian multipliers", *Numerische Mathematik* **20** (1973), no. 3, p. 179-192.
[3] M. Barré, C. Grandmont, P. Moireau, "Analysis of a linearized poromechanics model for incompressible and nearly incompressible materials", Technical Report HAL, 2021, https://hal.inria.fr/hal-03501526.
[4] D. Boffi, F. Brezzi, M. Fortin *et al.*, *Mixed finite element methods and applications*, vol. 44, Springer, 2013.
[5] A.-S. Bonnet-Ben Dhia, C. Carvalho, P. Ciarlet, Jr, "Mesh requirements for the finite element approximation of problems with sign-changing coefficients", *Numer. Math.* **138** (2018), p. 801–838.
[6] A.-S. Bonnet-Ben Dhia, L. Chesnel, P. Ciarlet, Jr, "T-coercivity for scalar interface problems between dielectrics and metamaterials", *Math. Mod. Num. Anal.* **46** (2012), p. 1363-1387.
[7] ———, "T-coercivity for the Maxwell problem with sign-changing coefficients", *Communications in Partial Differential Equations* **39** (2014), p. 1007-1031.
[8] ———, "Two-dimensional Maxwell's equations with sign-changing coefficients", *Appl. Numer. Math.* **79** (2014), p. 29-41.
[9] A.-S. Bonnet-Ben Dhia, L. Chesnel, X. Claeys, "Radiation condition for a non-smooth interface between a dielectric and a metamaterial", *Mathematical Models and Methods in Applied Sciences* **23** (2013), no. 09, p. 1629-1662.
[10] A.-S. Bonnet-Ben Dhia, P. Ciarlet Jr, C. M. Zwölf, "Time harmonic wave diffraction problems in materials with sign-shifting coefficients", *Journal of Computational and Applied Mathematics* **234** (2010), no. 6, p. 1912-1919.
[11] F. Brezzi, "On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers", *Publications mathématiques et informatique de Rennes* (1974), no. S4, p. 1-26.
[12] A. Buffa, "Remarks on the discretization of some noncoercive operator with applications to heterogeneous Maxwell equations", *SIAM J. Numer. Anal.* **43** (2005), p. 1-18.
[13] A. Buffa, S. H. Christiansen, "The electric field integral equation on Lipschitz screens: definitions and numerical approximation", *Numerische Mathematik* **94** (2003), no. 2, p. 229-267.
[14] A. Buffa, M. Costabel, C. Schwab, "Boundary element methods for Maxwell's equations on non-smooth domains", *Numerische Mathematik* **92** (2002), no. 4, p. 679-710.
[15] R. Bunoiu, L. Chesnel, K. Ramdani, M. Rihani, "Homogenization of Maxwell's equations and related scalar problems with sign-changing coefficients", in *Annales de la Faculté des Sciences de Toulouse. Mathématiques.*, 2020.
[16] R. Bunoiu, K. Ramdani, "Homogenization of materials with sign changing coefficients", *Communications in Mathematical Sciences* **14** (2016), no. 4, p. 1137-1154.
[17] R. Bunoiu, K. Ramdani, C. Timofte, "T-coercivity for the asymptotic analysis of scalar problems with sign-changing coefficients in thin periodic domains", *Electronic Journal of Differential Equations* (2021), p. 1-22.
[18] L. Chesnel, "Bilaplacian problems with a sign-changing coefficient", *Mathematical Methods in the Applied Sciences* **39** (2016), no. 17, p. 4964-4979.
[19] L. Chesnel, P. Ciarlet Jr, "T-coercivity and continuous Galerkin methods: application to transmission problems with sign changing coefficients", *Numerische Mathematik* **124** (2013), no. 1, p. 1-29.
[20] P. Ciarlet Jr, "T-coercivity: Application to the discretization of Helmholtz-like problems", *Computers & Mathematics with Applications* **64** (2012), no. 1, p. 22-34.
[21] ———, "Mathematical and numerical analyses for the div-curl and div-curlcurl problems with a sign-changing coefficient", Technical Report HAL, 2020, https://hal.inria.fr/hal-02651682.

[22] ———, "Lecture notes on Maxwell's equations and their approximation (in French)", Master's degree Analysis, Modelling and Simulation from Paris-Saclay University and Institut Polytechnique de Paris, 2021, https://hal.inria.fr/hal-03153780.
[23] ———, "On the approximation of electromagnetic fields by edge finite elements – Part 4: analysis of the model with one sign-changing coefficient", *Numer. Math.* **152** (2022), p. 223-257.
[24] P. Ciarlet Jr, E. Jamelot, F. D. Kpadonou, "Domain Decomposition Methods for the diffusion equation with low-regularity solution", *Computers Math. Applic.* **74** (2017), p. 2369-2384.
[25] G. Duvaut, J. L. Lions, *Les inéquations en mécanique et en physique*, Dunod, 1972.
[26] M. Fortin, "An analysis of the convergence of mixed finite element methods", *RAIRO. Analyse numérique* **11** (1977), no. 4, p. 341-354.
[27] V. Girault, P.-A. Raviart, *Finite element methods for Navier-Stokes equations: theory and algorithms*, vol. 5, Springer Science & Business Media, 2012.
[28] L. Giret, "Numerical analysis of a non-conforming Domain Decomposition for the multigroup SPN equations", PhD Thesis, Paris-Saclay University, 2018, https://pastel.archives-ouvertes.fr/tel-01936967.
[29] M. Halla, "Galerkin approximation of holomorphic eigenvalue problems: weak T-coercivity and T-compatibility", *Numerische Mathematik* **148** (2021), no. 2, p. 387-407.
[30] R. Hiptmair, "Finite elements in computational electromagnetics", *Acta Numerica* (2002), p. 237-339.
[31] Q. Hong, J. Kraus, M. Lymbery, F. Philo, "A new framework for the stability analysis of perturbed saddle-point problems and applications in poromechanics", arXiv preprint, 2022, https://arxiv.org/pdf/2103.09357.pdf.
[32] E. Jamelot, "T-coercivity for solving Stokes problem with nonconforming finite elements", 2022.
[33] E. Jamelot, P. Ciarlet Jr, "Fast non-overlapping Schwarz domain decomposition methods for solving the neutron diffusion equation", *J. Comput. Phys.* **241** (2013), p. 445-463.
[34] O. A. Ladyzhenskaya, *The mathematical theory of viscous incompressible flow*, vol. 2, Gordon and Breach New York, 1969.
[35] D. P. Levadoux, "Analyse numérique de la formulation intégrodifférentielle d'un problème de Maxwell harmonique impliquant un diélectrique traversé de surfaces exfoliées métalliques et impédantes", Technical Report HAL, 2022, https://hal.archives-ouvertes.fr/hal-03644547.
[36] S. Nicaise, J. Venel, "A posteriori error estimates for a finite element approximation of transmission problems with sign changing coefficients", *J. Comput. Appl. Math.* **235** (2011), p. 4272-4282.
[37] F.-J. Sayas, T. S. Brown, M. E. Hassell, *Variational techniques for elliptic partial differential equations*, CRC Press, 2019.
[38] C. Weber, "A local compactness theorem for Maxwell's equations", *Mathematical Methods in the Applied Sciences* **2** (1980), no. 1, p. 12-25.