Inf-sup condition and locking: Understanding and circumventing.

Stokes, Laplacian, bi-Laplacian, Kirchhoff–Love and Mindlin–Reissner locking type, boundary conditions

Gilles Leborgne
Isima, University of Clermont-Ferrand, France

April 22, 2020

Abstract

The inf-sup condition, also called the Ladyzhenskaya–Babuška–Brezzi (LBB) condition, ensures the well-posedness of a saddle point problem, relative to a partial differential equation. Discretization by the finite element method gives the discrete problem which must satisfy the discrete inf-sup condition. But, depending on the choice of finite elements, the discrete condition may fail. This paper attempts to explain why it fails from an engineer's perspective, and reviews current methods to work around this failure. The last part recalls the mathematical bases.

Contents

Part I
Introduction

1 The inf-sup condition, what is that ?

1.1 The constrained problem under concern

Let $(V, (\cdot,\cdot)_V)$ be a Hilbert space and $(Q, ||.||_Q)$ be a Banach space. Let $V' = \mathcal{L}(V;\mathbb{R})$ and $Q' = \mathcal{L}(Q;\mathbb{R})$ be the associated dual spaces (spaces of the linear continuous forms). Let $f \in V'$ and $g \in Q'$ be linear continuous forms, and let $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ and $b(\cdot,\cdot) : V \times Q \to \mathbb{R}$ be bilinear continuous forms.

The problem under concern is: Find $(u,p) \in V \times Q$ s.t.

$$\begin{cases} a(u,v) + b(v,p) = \langle f, v\rangle_{V',V}, & \forall v \in V, \\ b(u,q) \qquad\qquad\quad = \langle g, q\rangle_{Q',Q}, & \forall q \in Q. \end{cases} \tag{1.1}$$

E.g., with Ω an open regular bounded set in $\mathbb{R}^n$ and $L^2_0(\Omega) \simeq L^2(\Omega)/\mathbb{R}$ (that is the space of L^2 functions defined up to a constant), find $(\vec{u},p) \in H^1_0(\Omega)^n \times L^2_0(\Omega)$ s.t.

$$\begin{cases} (\mathrm{grad}\vec{u}, \mathrm{grad}\vec{v})_{L^2} - (p, \mathrm{div}\vec{v})_{L^2} = \langle \vec{f}, \vec{v}\rangle_{H^{-1},H^1_0}, \quad \forall \vec{v} \in H^1_0(\Omega)^n \\ -(\mathrm{div}\vec{u}, q)_{L^2} = 0, \quad \forall q \in L^2_0(\Omega). \end{cases} \tag{1.2}$$

Let $A \in \mathcal{L}(V;V')$, $B \in \mathcal{L}(V;Q')$ and $B^t \in \mathcal{L}(Q,V')$ be the associated linear continuous mapping (bounded operators) given by

$$\langle Au, v\rangle_{V',V} = a(u,v), \quad \langle Bv, p\rangle_{Q',Q} = b(v,p) = \langle B^t p, v\rangle_{V',V}. \tag{1.3}$$

Then (1.1) also reads

$$\begin{cases} Au + B^t p = f \in V', \\ Bu \qquad\quad = g \in Q', \end{cases} \tag{1.4}$$

the equation $Bu = g$ being the constraint. E.g., find $(\vec{u},p) \in H^1_0(\Omega)^n \times L^2_0(\Omega)$ s.t.

$$\begin{cases} -\Delta\vec{u} + \vec{\mathrm{grad}}p = \vec{f}, \\ \mathrm{div}\vec{u} \qquad\quad = 0. \end{cases} \tag{1.5}$$

1.2 The control on p to get a well-posed problem, and inf-sup

The simplest numerical finite element simulations show non admissible results for p (the pressure) in (1.2). And p being only present in $(1.2)_1$, we need to study B^t, cf. (1.4). The needed result will be the closure of $\mathrm{Im}(B^t)$ in V': In that case the open mapping theorem gives the control on p thanks to:

$$\exists\beta > 0,\ \forall p \in Q, \quad ||B^t p||_{V'} \geq \beta ||p||_{Q/\mathrm{Ker}(B^t)}, \tag{1.6}$$

also written as the "inf-sup condition": $\exists\beta > 0$, $\inf_{p\in Q} \sup_{v\in V} \frac{|b(v,p)|}{||v||_V ||p||_{Q/\mathrm{Ker}(B^t)}} \geq \beta$, since $||B^t p||_{V'} = \sup_{v\in V} \frac{|\langle B^t p, v\rangle_{V',V}|}{||v||_V} = \sup_{v\in V} \frac{|b(v,p)|}{||v||_V}$.

E.g. for (1.2), with $B^t = \vec{\mathrm{grad}} : L^2_0(\Omega) \to H^{-1}(\Omega)^n$, we have $\mathrm{Im}(B^t)$ closed in $H^{-1}(\Omega)^n$ (this is "the" difficult theorem to establish, see next §), thus

$$\exists\beta > 0,\ \forall p \in L^2(\Omega),\ ||\vec{\mathrm{grad}}p||_{H^{-1}} \geq \beta ||p||_{L^2_0}. \tag{1.7}$$

Which can be written as the "inf-sup condition": $\exists\beta > 0$, $\inf_{p\in L^2(\Omega)} \sup_{v\in H^1_0(\Omega)^n} \frac{|(\mathrm{div}\vec{v}, p)_{L^2}|}{||v||_{H^1_0} ||p||_{L^2_0}} \geq \beta$.

And we get the theorem: The problem (1.2) is well-posed, that is, the solution (u,p) exists, is unique, and depends continuously on f and g. See (12.14).

Remark: Thanks to the closed range theorem 11.2, the closure of $\mathrm{Im}(B^t)$ is equivalent to the closure of $\mathrm{Im}(B)$ (under usual hypotheses). This result is needed to get the existence of u.

1.3 The loss of control on p for the discrete problem

Let $V_h \subset V$ and $Q_h \subset Q$ be finite dimensional subspaces (conform finite elements to simplify). The discretization of (1.1) (for computation purposes) reads: Find $(u_h, p_h) \in V_h \times Q_h$ s.t.

$$\begin{cases} a(u_h, v_h) + b(v_h, p_h) = \langle f, v_h \rangle_{V',V}, & \forall v_h \in V_h, \\ b(u_h, q_h) \qquad\qquad\quad = \langle g, q_h \rangle_{Q',Q}, & \forall q_h \in Q_h. \end{cases} \tag{1.8}$$

E.g., with $V_h \subset H_0^1(\Omega)^n$, $Q_h \subset L_0^2(\Omega)$, and $\vec{f} \in L^2(\Omega)^n$,

$$\begin{cases} (\text{grad}\vec{u}_h, \text{grad}\vec{v}_h)_{L^2} - (p_h, \text{div}\vec{v}_h)_{L^2} = (\vec{f}, \vec{v}_h)_{L^2}, & \forall \vec{v}_h \in V_h, \\ -(\text{div}\vec{u}_h, q_h)_{L^2} = 0, & \forall q_h \in Q_h. \end{cases} \tag{1.9}$$

The h-discrete inf-sup condition is

$$\exists \beta_h > 0,\ \forall p_h \in Q_h, \quad ||B_h^t p_h||_{V'} \geq \beta_h ||p_h||_{Q_h/\text{Ker}(B^t)}. \tag{1.10}$$

And β_h should satisfy $\beta_h > \gamma$ is satisfied for some $\gamma > 0$: we get the so-called discrete inf-sup condition:

$$\exists \gamma > 0,\ \forall h > 0, \forall p_h \in Q_h, \quad ||B_h^t p_h||_{V'} \geq \gamma ||p_h||_{Q_h/\text{Ker}(B^t)}. \tag{1.11}$$

Fortin [17] gives a general useful method to check if the discrete inf-sup condition is satisfied.

Unfortunately, in many situations the stability condition (1.11) is not satisfied. E.g. P_1-continuous finite elements for both the velocity and pressure.

The associated matrix problem relative to (1.8) reads: Find $(U_h, P_h) \in \mathbb{R}^{n_V} \times \mathbb{R}^{n_Q}$ s.t. :

$$\begin{pmatrix} [A_h] & [B_h]^T \\ [B_h] & 0 \end{pmatrix} \begin{pmatrix} [U_h] \\ [P_h] \end{pmatrix} = \begin{pmatrix} [F_h] \\ [G_h] \end{pmatrix}, \tag{1.12}$$

$[B_h]^T$ being $[B_h]$ transposed. And if (1.11) is not satisfied then the matrix $\begin{pmatrix} [A_h] & [B_h]^T \\ [B_h] & [0] \end{pmatrix}$ is non invertible for some h.

1.4 Where is the problem?

E.g., with (1.9) and continuous P_1-continuous finite elements for $\vec{v}_h$ and p_h we have

$$b(\vec{v}_h, p_h) = (\vec{\text{grad}} p_h, \vec{v}_h)_{L^2} = (\Pi_{V_h} \vec{\text{grad}} p_h, \vec{v}_h)_{L^2}, \tag{1.13}$$

with $\Pi_{V_h} : L^2(\Omega)^n \to V_h$ the $(\cdot,\cdot)_{L^2}$-orthogonal projection on V_h; Here $\vec{\text{grad}} p_h$ is constant by element, and Π_{V_h} is the projection on continuous P_1 functions.

This projection Π_{V_h}, as any projection, looses information: Here we would like to consider $\vec{\text{grad}} p_h$ (to control p_h), but (1.13) tell us that only $\Pi_{V_h} \vec{\text{grad}} p_h$ is taken into account (is computed): Since $\vec{\text{grad}} p_h = \Pi_{V_h} \vec{\text{grad}} p_h + (\vec{\text{grad}} p_h - \Pi_{V_h} \vec{\text{grad}} p_h)$, we have lost $\vec{\text{grad}} p_h - \Pi_{V_h} \vec{\text{grad}} p_h$. And, e.g. with P_1-continuous finite elements for $\vec{v}_h$ and p_h, if nothing is done then the computation fails to give a good result (and it get worse as $h \to 0$).

To recover a well-posed problem, the missing term $\vec{\text{grad}} p_h - \Pi_{V_h} \vec{\text{grad}} p_h$ can be reintroduced, see (3.13), and we then get an optimal result (e.g., order h for convergence for P_1-continuous finite elements).

Part II
Examples and preventions

2 Stokes model

2.1 A first model

Let Ω be a bounded open set. Let $\text{div} : H_0^1(\Omega)^n \to L_0^2(\Omega)$ with $(\cdot,\cdot)_{H_0^1} = (\text{grad}(.), \text{grad}(.))_{L^2}$, and let

$$V = \{\vec{v} \in H_0^1(\Omega)^n : \text{div}\vec{v} = 0\}. \tag{2.1}$$

Problem (homogeneous Dirichlet type): for $\vec{f} \in H^{-1}(\Omega)^n$, find $\vec{u} \in H_0^1(\Omega)$ s.t.

$$-\Delta\vec{u} = \vec{f}. \tag{2.2}$$

Associated weak problem : find $\vec{u} \in V$ s.t.

$$(\text{grad}\vec{u}, \text{grad}\vec{v})_{L^2} = (\vec{f}, \vec{v})_{L^2}, \quad \forall \vec{v} \in H_0^1(\Omega). \tag{2.3}$$

The Lax–Milgram gives the well-posedness in $(H_0^1(\Omega), (\cdot,\cdot)_{H_0^1})$.

The optimized associated problem is: Find $\vec{u} \in H_0^1(\Omega)$ realizing the minimum of

$$J(\vec{v}) := \frac{1}{2}||\vec{\text{grad}}\vec{v}||_{L^2}^2 - (\vec{f}, \vec{v})_{L^2}. \tag{2.4}$$

2.2 Constrained associated problem

The constraint $\text{div}\vec{u} = 0$ is imposed with a Lagrangian multiplier p: The problem (2.2) is transformed into: Find $(\vec{u}, p) \in H_0^1(\Omega) \times L_0^2(\Omega)$ s.t.

$$\begin{cases} (\text{grad}\vec{u}, \text{grad}\vec{v})_{L^2} - (p, \text{div}\vec{v})_{L^2} = (\vec{f}, \vec{v})_{L^2}, & \forall \vec{v} \in H_0^1(\Omega)^n, \\ -(\text{div}\vec{u}, q)_{L^2} = 0, & \forall q \in L_0^2(\Omega). \end{cases} \tag{2.5}$$

We have obtained (1.1) with $V = H_0^1(\Omega)^n$, $Q = L_0^2(\Omega)$, $g = 0$, $B = \text{div} : H_0^1(\Omega)^n \to L_0^2(\Omega)$ and

$$\begin{cases} a(\vec{u}, \vec{v}) = (\text{grad}\vec{u}, \text{grad}\vec{v})_{L^2} \quad \text{sur } H_0^1(\Omega)^n \times H_0^1(\Omega)^n, \\ b(\vec{v}, q) = -(\text{div}\vec{v}, q)_{L^2} \quad \text{sur } H_0^1(\Omega)^n \times L_0^2(\Omega). \end{cases} \tag{2.6}$$

Since $B = \text{div} : H_0^1(\Omega) \to L_0^2(\Omega)$, $\text{Ker}B = V$, $a(\cdot,\cdot)$ is coercive on $\text{Ker}(B)$ (it is on $H_0^1(\Omega)^n$), and $B^t = \vec{\text{grad}} : L^2(\Omega) \to H^{-1}(\Omega)^n$ is surjective, cf. theorem 10.1, the theorem 12.1 applies, and the problem (2.5) is well-posed.

The associated weak problem reads: Find $(\vec{u}, p) \in H_0^1(\Omega) \times L_0^2(\Omega)$ s.t.

$$\begin{cases} -\Delta\vec{u} + \vec{\text{grad}}p = \vec{f} \in H^{-1}(\Omega), \\ \text{div}\vec{u} = 0. \end{cases} \tag{2.7}$$

The associated Lagrangian reads, find the saddle point in $H_0^1(\Omega)^n \times L_0^2(\Omega)$ of

$$\mathcal{L}(\vec{v}, q) = \frac{1}{2}||\text{grad}\vec{v}||_{L^2}^2 - (q, \text{div}\vec{v})_{L^2} - (\vec{f}, \vec{v})_{L^2}. \tag{2.8}$$

3 Numerical approximation of the Stokes model

3.1 Approximation

Let $V_h \subset H_0^1(\Omega)$ and $Q_h \subset L_0^2(\Omega)$ (conform approximation to simplify) be finite dimension subspaces. The discretization of (2.5) is: Find $\vec{u}_h \in (V_h)^n$ and $p_h \in Q_h$ s.t.

$$\begin{cases} (\text{grad}\vec{u}_h, \text{grad}\vec{v}_h)_{L^2} - (p_h, \text{div}\vec{v}_h)_{L^2} = (\vec{f}, \vec{v}_h)_{L^2}, & \forall \vec{v}_h \in (V_h)^n, \\ (\text{div}\vec{u}_h, q_h)_{L^2} = 0, & \forall q_h \in Q_h. \end{cases} \tag{3.1}$$

3.2 Projections (finite element method)

If X_h is a subspace in $L^2(\Omega)$, let $\Pi_{X_h} : \begin{Bmatrix} L^2(\Omega) & \to X_h \\ f & \to \Pi_{X_h} f \end{Bmatrix}$ be the $(\cdot,\cdot)_{L^2}$-orthogonal projection on X_h, that is,

$$\forall f \in L^2(\Omega), \quad (\Pi_{X_h} f, x_h)_{L^2} = (f, x_h)_{L^2}, \quad \forall x_h \in X_h. \tag{3.2}$$

E.g., if $X_h = P_1$ then $\Pi_{P_1} f \in P_1$ is the best approximation P_1 of f for the $(\cdot,\cdot)_{L^2}$ inner product. Similar notation for X_h a subspace in $L^2(\Omega)^n$.

Let

$$\vec{\text{grad}}_h \stackrel{\text{def}}{=} \Pi_{V_h} \circ \vec{\text{grad}} : \begin{cases} L^2(\Omega) & \to V_h \\ \quad p & \to \vec{\text{grad}}_h p = \Pi_{V_h}(\vec{\text{grad}} p). \end{cases} \tag{3.3}$$

So, $\vec{\text{grad}}_h p$ is characterized by $(\vec{\text{grad}}_h p, \vec{v}_h)_{L^2} = (\vec{\text{grad}} p, \vec{v}_h)_{L^2}$ pour tout $\vec{v}_h \in V_h$. And (3.1) reads

$$\begin{cases} (\text{grad}\vec{u}_h, \text{grad}\vec{v}_h)_{L^2} + (\vec{\text{grad}}_h p_h, \vec{v}_h)_{L^2} = (\vec{f}, \vec{v}_h)_{L^2}, \quad \forall \vec{v}_h \in V_h, \\ (\vec{u}_h, \vec{\text{grad}} q_h)_{L^2} = 0, \quad \forall q_h \in Q_h., \end{cases} \tag{3.4}$$

3.3 Matrix representation

With given bases in V_h and Q_h, (3.1) become

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} . \begin{pmatrix} \vec{x} \\ \vec{y} \end{pmatrix} = \begin{pmatrix} F \\ 0 \end{pmatrix}. \tag{3.5}$$

3.4 The problematic pressure

In many cases there is no problem with the computation of u_h (the A matrix i (3.5) is well conditioned since $a(\cdot, \cdot)$ is continuous and coercive).

But the results obtained for p_h can be absurd. To see why, suppose that u_h is known, let $(g, \vec{v}_h)_{L^2} := (\vec{\text{grad}}\vec{u}_h, \vec{\text{grad}}\vec{v}_h)_{L^2} - (\vec{f}, \vec{v}_h)_{L^2}$, and try to find $p_h \in Q_h$ s.t.

$$(\vec{\text{grad}} p_h, \vec{v}_h)_{H^{-1},H_0^1} = -(g, \vec{v}_h)_{L^2}, \quad \forall \vec{v}_h \in V_h, \tag{3.6}$$

that is, e.g. with continuous finite elements where $\langle \vec{\text{grad}} p_h, \vec{v}_h \rangle_{H^{-1},H_0^1} = (\vec{\text{grad}} p_h, \vec{v}_h)_{L^2}$,

$$(\vec{\text{grad}}_h p_h, \vec{v}_h)_{L^2} = -(g, \vec{v}_h)_{L^2}, \quad \forall \vec{v}_h \in V_h. \tag{3.7}$$

1- Nice case: $\vec{\text{grad}}_h = \Pi_{V_h} \circ \vec{\text{grad}} : V_h \to Q_h$ is surjective (onto) with a constant independent of h, cf. (10.3), that is,

$$\exists k > 0,\ \forall h > 0,\ \forall p_h \in Q_h, \quad ||\vec{\text{grad}}_h p_h||_{H^{-1}} \geq k\, ||p_h||_{L_0^2}. \tag{3.8}$$

And (3.8) is called the "discrete inf-sup condition".

Then the problem (3.4) is well-posed, i.e. the matrix i (3.5) is well-conditioned, cf. theorem 12.1. See Fortin [17] for V_h and Q_h finite element spaces that can satisfy (3.8).

(Remark: the problem (3.7) cannot be solved on its own in general, since it is surjective but not bijective. But (3.4) can be solved, the matrix $\begin{pmatrix} A & B^t \\ B & 0 \end{pmatrix}$ being invertible and well-conditioned if (3.8) is satisfied.)

2- Bad case: In (3.8), $k > 0$ does not exists, e.g.

$$\exists k_h > 0, \quad \inf_{p_h \in Q_h} \sup_{\vec{v}_h \in V_h} \frac{(\text{div}\vec{v}_h, p_h)_{L^2(\Omega)}}{||\vec{v}_h||_{H_0^1} ||p_h||_{L^2}} \geq k_h, \quad \text{but} \quad k_h \underset{h\to 0}{\longrightarrow} 0. \tag{3.9}$$

Then $\begin{pmatrix} A & B^t \\ B & 0 \end{pmatrix}$ is not invertible (at least not numerically invertible as $h \to 0$: bad conditioning).

Example 3.1 A useful criteria to check the discrete inf-sup condition (3.8) is given by Fortin [17]. E.g., the discrete inf-sup condition is satisfied with the classical:

P_2,P_1 (velocity-pressure) Taylor–Hood finite elements (see e.g. Bercovier–Pironneau [4]).

P_1-bubble,P_1 (velocity-pressure) finite elements, named the mini-elements, see Arnold–Brezzi–Fortin [1].

P_2,P_0 (velocity-pressure) finite elements, see Crouzeix–Raviart [13].

(And for non conformity, the P_1-discontinuous velocity, P_0-pression, see Crouzeix–Raviart [13].) ▪

Example 3.2 No convergence e.g. for the P_1,P_1 continuous finite elements, or the P_1-continuous,P_0 elements (checkerboard instability). ▪

3.5 What has been lost...

(3.4) reads

$$(\Pi_{V_h}\vec{\text{grad}}p_h, \vec{v}_h)_{L^2} = -(g, \vec{v}_b), \quad \forall \vec{v}_h \in V_h. \tag{3.10}$$

So we want $\vec{\text{grad}}p_h$, but we can only compute $\vec{\text{grad}}_h p_h = \Pi_{V_h}\vec{\text{grad}}p_h$, which in many cases is different from $\vec{\text{grad}}p_h$. Since

$$\vec{\text{grad}}p_h = \Pi_{V_h}\vec{\text{grad}}p_h + \big(\vec{\text{grad}}p_h - \Pi_{V_h}\vec{\text{grad}}p_h\big), \tag{3.11}$$

we have lost

$$\text{loss} = (\vec{\text{grad}}p_h - \Pi_{V_h}\vec{\text{grad}}p_h) = (\vec{\text{grad}}p_h - \vec{\text{grad}}_h p_h). \tag{3.12}$$

This can be an admissible loss, see e.g. example 3.1, or not, see e.g. example 3.2.

3.6 ... and a reintroduction of the loss

To recover the loss (3.12), we modify (3.1) to get the new problem: Find $\vec{u}_h \in V_h$ et $p_h \in Q_h$ s.t.

$$\begin{cases} (\vec{\text{grad}}\vec{u}_h, \vec{\text{grad}}\vec{v}_h)_{L^2} - (p_h, \text{div}\vec{v}_h)_{L^2} = (\vec{f}, \vec{v}_h)_{L^2}, \quad \forall \vec{v}_h \in V_h, \\ -(\text{div}\vec{u}_h, q_h)_{L^2} - \displaystyle\sum_{K=1}^{n_K} h_K^2 (\vec{\text{grad}}p_h - \vec{\text{grad}}_h p_h, \vec{\text{grad}}q_h - \vec{\text{grad}}_h q_h)_{L^2(K)} = 0, \quad \forall q_h \in Q_h. \end{cases} \tag{3.13}$$

where n_K is the number of elements constituting the mesh, h is "the size of an element", and the h_K^2 coefficient to get optimal results, see Leborgne [23] (we are interested in p_h and, for quasi-uniform meshes, p_h is of the same order than $h\,\vec{\text{grad}}p_h$). E.g. for P_1,P_1 continuous finite elements for both the velocity and the pressure, we get order 1 convergence results (classic for P_1 finite elements).

(3.13) can also be written

$$\begin{cases} (\vec{\text{grad}}\vec{u}_h, \vec{\text{grad}}\vec{v}_h)_{L^2} - (p_h, \text{div}\vec{v}_h)_{L^2} = (\vec{f}, \vec{v}_h)_{L^2}, \quad \forall \vec{v}_h \in V_h, \\ -(\text{div}\vec{u}_h, q_h)_{L^2} - \displaystyle\sum_{K=1}^{n_K} h_K^2 (\vec{\text{grad}}p_h - \Pi_{V_h}\vec{\text{grad}}p_h, \vec{\text{grad}}q_h)_{L^2(K)} = 0, \quad \forall q_h \in Q_h, \end{cases} \tag{3.14}$$

since $(\vec{\text{grad}}p_h - \Pi_{V_h}\vec{\text{grad}}p_h, \vec{w}_h) = 0$ for $\vec{w}_h \in V_h$ (definition of Π_{V_h}).

Computation: we have to compute a new unknown $\vec{z}_h = \Pi_{V_h}\vec{\text{grad}}p_h \in V_h$ (luckily very cheap for P_1 finite elements): Find $\vec{u}_h, \vec{z}_h \in V_h$ et $p_h \in Q_h$ s.t.

$$\begin{cases} (\vec{\text{grad}}\vec{u}_h, \vec{\text{grad}}\vec{v}_h)_{L^2} - (p_h, \text{div}\vec{v}_h)_{L^2} = (\vec{f}, \vec{v}_h)_{L^2}, \quad \forall \vec{v}_h \in V_h, \\ -(\text{div}\vec{u}_h, q_h)_{L^2} - \displaystyle\sum_{K=1}^{n_K} h_K^2 (\vec{\text{grad}}p_h, \vec{\text{grad}}q_h)_{L^2} + h^2(\vec{z}_h, \vec{\text{grad}}q_h)_{L^2} = 0, \quad \forall q_h \in Q_h, \\ (\vec{\text{grad}}p_h, \vec{z}'_h)_{L^2(K)} - (\vec{z}_h, \vec{z}'_h)_{L^2(K)} = 0, \quad \forall \vec{z}'_h \in V_h,\ \forall K. \end{cases} \tag{3.15}$$

E.g.with P_1 finite elements, the $(\vec{z}_h, \vec{z}'_h)_{L^2}$ associated matrix can be made diagonal thanks to the "mass lumping" technique: Thus the last equation (in $\vec{z}'_h$) gives $\vec{z}_h$ explicitly as a function of $\vec{\text{grad}}_h p_h$ (order 1 precision).

Remark 3.3 The associated Lagrangian, cf. (2.8), is now:

$$L_h(\vec{v}_h, p_h) = \frac{1}{2}||\text{grad}\vec{v}_h||_{L^2}^2 - (p_h, \text{div}\vec{v}_h)_{L^2} - (f, v_h)_{L^2} - \frac{1}{2}\sum_{K=1}^{n_K} h_K^2 ||\vec{\text{grad}}p_h - \Pi_{V_h}\vec{\text{grad}}p_h||_{L^2(K)}^2. \tag{3.16}$$

■

3.7 Brezzi and Pitkäranta's method

A previous method proposed by Brezzi and Pitkäranta [9] consists in penalizing the initial problem with the Laplacian of the pressure (to "control the oscillations" of p_h): Find $\vec{u}_h \in V_h$ and $p_h \in Q_h$ s.t.

$$\begin{cases} (\vec{\mathrm{grad}}\vec{u}_h, \vec{\mathrm{grad}}\vec{v}_h)_{L^2} - (p_h, \mathrm{div}\vec{v}_h)_{L^2} = (\vec{f}, \vec{v}_h)_{L^2}, \quad \forall \vec{v}_h \in V_h, \\ -(\mathrm{div}\vec{u}_h, q_h)_{L^2} - \varepsilon \sum_{K=1}^{n_K} h_K^2 (\vec{\mathrm{grad}} p_h, \vec{\mathrm{grad}} q_h)_{L^2(K)} = 0, \quad \forall q_h \in Q_h, \end{cases} \tag{3.17}$$

with some $\varepsilon > 0$. We however get a spurious limit condition $\frac{\partial p}{\partial n} = 0$ independent of ε (by integration by parts). (This spurious limit condition is lessen with (3.14).)

Remark 3.4 The associated Lagrangian is now:

$$L_h(\vec{v}_h, p_h) = \frac{1}{2}||\mathrm{grad}\vec{v}_h||_{L^2}^2 - (p_h, \mathrm{div}\vec{v}_h)_{L^2} - (f, v_h)_{L^2} - \frac{1}{2}\varepsilon \sum_{K=1}^{n_K} h_K^2 ||\vec{\mathrm{grad}} p_h||_{L^2(K)}^2, \tag{3.18}$$

to compare with (3.16). ■

3.8 Hughes, Franca and Balestra's method

Hughes, Franca and Balestra [22] proposed a "Galerkin Least-squares" method: The pressure is stabilized "with the solution". The problem reads, with the associated Lagrangian,

$$L(\vec{v}, p) = \frac{1}{2}||\mathrm{grad}\vec{v}||_{L^2(\Omega)}^2 - (p, \mathrm{div}\vec{v})_{L^2(\Omega)} - (f, v)_{L^2(\Omega)} - \frac{\varepsilon}{2} \sum_{K=1}^{n_K} h_K^2 || - \Delta u + \vec{\mathrm{grad}} p - f||_{L^2(K)}^2. \tag{3.19}$$

(For P_1 finite elements, this method is similar to Brezzi and Pitkäranta's method.)

Here ε has to be small enough not to destroy the coercivity in u, see the term $(\vec{\mathrm{grad}} u, \vec{\mathrm{grad}} v)_{L^2(\Omega)} - \varepsilon \sum_k h^2 (\Delta u, \Delta v)_{L^2(K)}$, the control being done thanks to the inverse inequality (quasi-uniform mesh)

$$||\Delta u_h||_{L^2(K)} \le Ch ||\vec{\mathrm{grad}} u_h||_{L^2(K)}, \quad \forall u_h \in V_h.$$

(So $0 < \varepsilon < \frac{1}{\sqrt{C}}$.)

3.9 Douglas and Wang's method

To avoid the eventual destruction of the coercivity for $\vec{u}$, Douglas et Wang consider

$$\underbrace{(\mathrm{grad}\vec{u}, \mathrm{grad}\vec{v})_{L^2} - (p, \mathrm{div}\vec{v}) + (q, \mathrm{div}\vec{u}) + \varepsilon \sum_{K=1}^{n_K} h_K (-\Delta \vec{u} + \vec{\mathrm{grad}} p - \vec{f}, -\Delta \vec{v} + \vec{\mathrm{grad}} q)_{L^2(K)}}_{c((\vec{u},p),(\vec{v},q))} = \underbrace{(f, v)_{L^2}}_{\ell(\vec{v},q)}. \tag{3.20}$$

This preserves the stability since $c(\cdot, \cdot)$ is coercive, but the symmetry is lost. So this method is adapted to the generalization of the Stokes equations to the Navier–Stokes equations.

4 Laplacian (harmonic problem)

The linear spaces needed are described in § 9.

4.1 A dirichlet problem

Let $f \in H^{-1}(\Omega)$. Problem: Find $p \in H_0^1(\Omega)$ s.t.

$$-\Delta p = f. \tag{4.1}$$

That is,

$$(\vec{\mathrm{grad}} p, \vec{\mathrm{grad}} q)_{L^2} = \langle f, q \rangle, \quad \forall q \in H_0^1(\Omega). \tag{4.2}$$

The associated minimum problem is: Find the minimum of $J(p) = \min_{q \in H_0^1(\Omega)} J(q)$, where

$$J(q) := \frac{1}{2} ||\vec{\mathrm{grad}} q||_{L^2}^2 - \langle f, q \rangle. \tag{4.3}$$

To get $\vec{\mathrm{grad}} p$ during the computation, introduce

$$\vec{u} = \vec{\mathrm{grad}} p \ \in L^2(\Omega), \quad \text{and then} \quad -\mathrm{div}\vec{u} = f. \tag{4.4}$$

And (4.1) becomes: Find $(\vec{u}, p) \in L^2(\Omega) \times H_0^1(\Omega)$ s.t.

$$\begin{cases} (\vec{u}, \vec{v})_{L^2} - (\vec{\mathrm{grad}} p, \vec{v})_{L^2} = 0, \quad \forall \vec{v} \in L^2(\Omega), \\ -(\vec{u}, \vec{\mathrm{grad}} q)_{L^2} = -\langle f, q \rangle_{H^{-1}, H_0^1}, \quad \forall q \in H_0^1(\Omega). \end{cases} \tag{4.5}$$

And p is now the Lagrangian multiplier for the constraint $\mathrm{div}\vec{u} = -f$, cf. the integration by parts.

And if $(\vec{u}, p) \in L^2(\Omega) \times H_0^1(\Omega)$ is a solution, then $\vec{u} = \vec{\mathrm{grad}} p \in L^2(\Omega)^n$ in Ω, and $\mathrm{div}\vec{u} = -f$ in $H^{-1}(\Omega)$. So $\Delta p = f$ in $H^{-1}(\Omega)$ with $p \in H_0^1(\Omega)$: This is (4.1).

With

$$\begin{cases} a(\vec{u}, \vec{v}) = (\vec{u}, \vec{v})_{L^2} \quad \text{sur } L^2(\Omega)^n \times L^2(\Omega)^n, \\ b(\vec{v}, q) = -(\vec{v}, \vec{\mathrm{grad}} q)_{L^2} \quad \text{sur } L^2(\Omega)^n \times H_0^1(\Omega), \end{cases} \tag{4.6}$$

(4.5) has the appearance of (1.1) with $V = L^2(\Omega)^n$ and $Q = H_0^1(\Omega)$.

Here $b(\vec{v}, p) = \langle B\vec{v}, p \rangle_{H^{-1}, H^1(\Omega)} = (B^t p, \vec{v})_{L^2(\Omega)}$, so $B = \mathrm{div} : \begin{Bmatrix} L^2(\Omega)^n & \to H^{-1}(\Omega) \\ \vec{v} & \to B\vec{v} = \mathrm{div}(\vec{v}) \end{Bmatrix}$ and $B^t = -\vec{\mathrm{grad}} : \begin{Bmatrix} H_0^1(\Omega) & \to L^2(\Omega) \\ p & \to B^t p = -\vec{\mathrm{grad}} p \end{Bmatrix}$.

Thus $\mathrm{Ker}(B) = \mathrm{Ker}(\mathrm{div}) = \{\vec{v} \in L^2(\Omega)^n : \mathrm{div}\vec{v} = 0\}$, and thanks to the Helmholtz decomposition (9.31) $L^2(\Omega)^n = \vec{\mathrm{grad}}(H_0^1(\Omega)) \oplus^{\perp_{L^2}} \mathrm{Ker}(\mathrm{div})$, the bilinear form $a(\cdot, \cdot)$ is $(\cdot, \cdot)_{L^2}$ coercive on $\mathrm{Ker}(B)$.

And B^t is surjective since B is, cf. (13.5) and the closed range theorem 11.2.

Thus (4.5) is well-posed, see theorem 12.1.

4.2 A Neumann problem

Let $f \in L^2(\Omega)$. Problem: Find $p \in H^1(\Omega)$ s.t.

$$-\Delta p = f, \quad \text{and} \quad \frac{\partial p}{\partial n}_{|\Gamma} = 0. \tag{4.7}$$

That is,

$$(\vec{\mathrm{grad}} p, \vec{\mathrm{grad}} q)_{L^2} = (f, q)_{L^2}, \quad \forall q \in H_0^1(\Omega). \tag{4.8}$$

The associated minimum problem is: Find the minimum of $J(p) = \min_{q \in H^1(\Omega)} J(q)$, where

$$J(q) := \frac{1}{2} ||\vec{\mathrm{grad}} q||_{L^2}^2 - (f, q)_{L^2}. \tag{4.9}$$

The mixed associated problem is: Find $(\vec{u},p) \in H^{\text{div}}(\Omega) \times L^2(\Omega)$ s.t.

$$\begin{cases} (\vec{u},\vec{v})_{L^2} + (p, \text{div}\vec{v})_{L^2} = 0, \quad \forall \vec{v} \in H^{\text{div}}(\Omega), \\ (\text{div}\vec{u}, q)_{L^2} = (f,q)_{L^2}, \quad \forall q \in L^2(\Omega). \end{cases} \tag{4.10}$$

Indeed, if $(\vec{u},p) \in H^{\text{div}}(\Omega) \times L^2(\Omega)$ is a solution, then $\text{div}\vec{u} = f \in L^2(\Omega)$, $\vec{u} = -\vec{\text{grad}}p \in H^{-1}(\Omega)$, thus $-\Delta p = f \in L^2(\Omega)$, with $\frac{\partial p}{\partial n}_{|\Gamma} = 0$ (since $\text{Im}(\gamma_n) = H^{-\frac{1}{2}}(\Gamma)$).

With

$$\begin{cases} a(\vec{u},\vec{v}) = (\vec{u},\vec{v})_{L^2} \quad \text{sur } H^{\text{div}}(\Omega) \times H^{\text{div}}(\Omega), \\ b(\vec{v},q) = (\text{div}\vec{v}, q)_{L^2} \quad \text{sur } H^{\text{div}}(\Omega) \times L^2(\Omega). \end{cases} \tag{4.11}$$

(4.5) has the appearance of (1.1) with $V = H^{\text{div}}(\Omega)$ and $Q = L^2(\Omega)$.

Here $B : \vec{v} \in H^{\text{div}}(\Omega) \to B\vec{v} = \text{div}\vec{v} \in L^2(\Omega)$ is surjective, cf. (13.2), and $a(\cdot,\cdot)$ est H^{div}-coercive on $\text{Ker}(B) = \{\vec{v} \in H^{\text{div}}(\Omega) : \text{div}\vec{v} = 0\}$. And $\text{Im}(B)$ being closed (since it is surjective), so is $\text{Im}(B^t)$ (closed range theorem11.2). Thus (4.10) is well-posed, see theorem 12.1.

5 Bilaplacian (biharmonic problem)

The linear spaces needed are described in § 9.

5.1 Problem

We look here at the Dirichlet problem: If $f \in H^{-2}(\Omega) = (H_0^2(\Omega))'$, then find $p \in H_0^2(\Omega)$ s.t.

$$\Delta(\Delta p) = f, \quad \text{so with} \quad p_{|\Gamma} = 0 \quad \text{and} \quad \frac{\partial p}{\partial n}_{|\Gamma} = 0. \tag{5.1}$$

Weak form: Find $p \in H_0^2(\Omega)$ s.t.

$$(\Delta p, \Delta q)_{L^2} = \langle f, q \rangle_{H^{-2}, H_0^2}, \quad \forall q \in H_0^2(\Omega). \tag{5.2}$$

The Lax–Milgram theorem indicates that (5.2) is well-posed.

The associated minimum problem is: Find the minimum of $J(p) = \min_{q \in H_0^1(\Omega)} J(q)$, where

$$J(q) := \frac{1}{2}||\Delta q||_{L^2}^2 - \langle f, q \rangle_{H^{-2}, H_0^2}. \tag{5.3}$$

5.2 Introduction of Δp

(Not conclusive.) A function in $\in H^2(\Omega)$, s.t. Δp and Δq in (5.2), is cumbersome to approximate, cf. the C^1 Argyris finite elements. Let

$$\phi = \Delta p \tag{5.4}$$

Then problem (5.1) is rewritten as: Find $(\phi, p) \in L^2(\Omega) \times H_0^2(\Omega)$ s.t.

$$\begin{cases} \phi = \Delta p, \\ \Delta \phi = f. \end{cases} \tag{5.5}$$

And the weak form is, if $f \in H^{-1}(\Omega)$: Find $(\phi, p) \in H^1(\Omega) \times H_0^1(\Omega)$ s.t.

$$\begin{cases} (\phi, \psi)_{L^2} + (\vec{\text{grad}} p, \vec{\text{grad}} \psi)_{L^2} = 0, \quad \forall \psi \in H^1(\Omega), \\ (\vec{\text{grad}} \phi, \vec{\text{grad}} q)_{L^2} = -\langle f, q \rangle_{H^{-1}, H_0^1}, \quad \forall q \in H_0^1(\Omega). \end{cases} \tag{5.6}$$

Indeed, if $(\phi, p) \in H^1(\Omega) \times H_0^1(\Omega)$ is as solution of (5.6), then $\phi - \Delta p = 0$ (thus $\Delta p \in L^2(\Omega)$), and $\frac{\partial p}{\partial n}_{|\Gamma} = 0$. And $p \in H_0^1(\Omega)$ with $\Delta \phi = f \in H^{-1}(\Omega)$, thus $\Delta^2 p = f$.

With

$$\begin{cases} a(\phi, \psi) = (\phi, \psi)_{L^2} \quad \text{sur } H^1(\Omega) \times H^1(\Omega), \\ b(\phi, v) = (\vec{\text{grad}} \phi, \vec{\text{grad}} v)_{L^2} \quad \text{sur } H^1(\Omega) \times H_0^1(\Omega), \end{cases} \tag{5.7}$$

(4.5) has the appearance of (5.6) with $V = H^1(\Omega)$ and $Q = H_0^1(\Omega)$.

Here $b(\phi, v) = -\langle \Delta \phi, v \rangle_{H^{-1}, H_0^1}$, thus $B = -\Delta : H^1(\Omega) \to H^{-1}(\Omega)$.

Thus $B : \phi \in H^1(\Omega) \to B\phi = -\Delta \phi \in H^{-1}(\Omega)$ is surjective: Apply Lax–Milgram theorem for $g \in H^{-1}(\Omega)$ and $(\vec{\text{grad}} \phi, \vec{\text{grad}} \psi)_{L^2} = \langle g, \psi \rangle_{H^{-1}, H_0^1}$.

And $\text{Ker} B = \{\phi \in H^1(\Omega) : \Delta \phi = 0\}$ (harmonic functions). So $\phi \in \text{Ker} B$ iff $(\vec{\text{grad}} \phi, \vec{\text{grad}} v)_{L^2} = 0$ for all $v \in H_0^1(\Omega)$ (this is not $(\vec{\text{grad}} \phi, \vec{\text{grad}} v)_{L^2} = 0$ for all $v \in H^1(\Omega)$). Thus $a(\cdot, \cdot)$ is not $(\cdot, \cdot)_{H^1}$-coercive on $\text{Ker} B$, but only $(\cdot, \cdot)_{L^2}$ coercive, and the usual theorem is not applicable: A loss of precision (precision$||.||_{L^2}$ instead of precision $||.||_{H^1}$ for ϕ) is to be expected.

5.3 Introduction of $\vec{\text{grad}} p$

5.3.1 Weak form

In (5.2) let us introduce

$$\vec{u} = \vec{\text{grad}} p, \quad \text{thus} \quad \Delta p = \text{div} \vec{u}. \tag{5.8}$$

Notation:

$$\text{if } \vec{v} = \vec{\text{grad}} q \in \vec{\text{grad}}(H_0^1(\Omega)), \quad \text{then} \quad \vec{v} \overset{\text{denoted}}{=} \vec{v}_q. \tag{5.9}$$

(That is, $\vec{v}_q$ derives from a potential $q \in H_0^1(\Omega)$.)

Then (5.2) becomes: Find $\vec{u}_p = \vec{\text{grad}} p \in \vec{\text{grad}}(H_0^1(\Omega))$ s.t.

$$(\text{div}\vec{u}_p, \text{div}\vec{v}_q)_{L^2} = \langle f, q\rangle_{H^{-2},H_0^2}, \quad \forall q \in H_0^2(\Omega). \tag{5.10}$$

5.3.2 A first constrained formulation

(Not conclusive.) To avoid working with the "small space" $\vec{\text{grad}}(H_0^1(\Omega)) \ni \vec{u}$, we consider the whole space $H_0^1(\Omega) \ni \vec{u}$ and we add the constraint $\vec{u} - \vec{\text{grad}} p = 0$, cf. (5.8) (and the associated Lagrangian multiplier $\vec{\lambda}$).

And $\vec{u} \in H^1(\Omega)$ and $\vec{u} = \vec{\text{grad}} p$ for some $p \in H_0^2(\Omega)$ solution of (5.2), give $\text{div}\vec{u} = \Delta p \in L^2(\Omega)$ with $0 = \frac{\partial p}{\partial n}_{|\Gamma} = \vec{u}.\vec{n}_{|\Gamma}$, so $\vec{u} \in H_0^{\text{div}}(\Omega)$. Then let

$$X = H_0^{\text{div}}(\Omega) \times H_0^1(\Omega), \tag{5.11}$$

provided with the inner product

$$((\vec{u}, p), (\vec{v}, q))_X = (\text{div}\vec{u}, \text{div}\vec{v})_{L^2} + (\vec{\text{grad}} p, \vec{\text{grad}} q)_{L^2(\Omega)}, \tag{5.12}$$

so that X is a Hilbert space.

Then, if $f \in H^{-1}(\Omega)$, (5.10) is turned into: Find $((\vec{u}, p), \vec{\lambda}) \in X \times L^2(\Omega)^n$ s.t.

$$\begin{cases} (\text{div}\vec{u}, \text{div}\vec{v})_{L^2} + (\vec{\lambda}, \vec{v} - \vec{\text{grad}} q)_{L^2} = \langle f, q\rangle_{H^{-1},H_0^1}, \quad \forall (\vec{v}, q) \in X, \\ (\vec{u} - \vec{\text{grad}} p, \vec{\mu})_{L^2} = 0, \quad \forall \vec{\mu} \in L^2(\Omega)^n, \end{cases} \tag{5.13}$$

that is,

$$\begin{cases} (\text{div}\vec{u}, \text{div}\vec{v})_{L^2} + (\vec{\lambda}, \vec{v})_{L^2} = 0, \quad \forall \vec{v} \in H_0^{\text{div}}(\Omega), \\ -(\vec{\lambda}, \vec{\text{grad}} q)_{L^2} = \langle f, q\rangle_{H^{-1},H_0^1}, \quad \forall q \in H_0^1(\Omega), \\ (\vec{u}, \vec{\mu})_{L^2} - (\vec{\text{grad}} p, \vec{\mu})_{L^2} = 0, \quad \forall \vec{\mu} \in L^2(\Omega)^n. \end{cases} \tag{5.14}$$

Check: If $((\vec{u}, p), \vec{\lambda}) \in X \times L^2(\Omega)^n$ is a solution of (5.13) or (5.14), then $\vec{\lambda} = \vec{\text{grad}}(\text{div}\vec{u})$, $\text{div}\vec{\lambda} = f$, $\vec{u} = \vec{\text{grad}} p$, thus $\text{div}\vec{u} = \Delta p$ and $\text{div}(\vec{\text{grad}}(\Delta p)) = f$, i.e. $\Delta^2 p = f$. And $\vec{u} \in H_0^{\text{div}}(\Omega)$ gives $\vec{u}.\vec{n} = 0$, so $\vec{\text{grad}} p.\vec{n} = 0$, and with $p \in H_0^1(\Omega)$ we get $p \in H_0^2(\Omega)$.

With

$$\begin{cases} a((\vec{u}, p), (\vec{v}, q)) = (\text{div}\vec{u}, \text{div}\vec{v})_{L^2} \quad \text{sur } X \times X, \\ b((\vec{v}, q), \vec{\mu}) = (\vec{v} - \vec{\text{grad}} q, \vec{\mu})_{L^2} \quad \text{sur } X \times L^2(\Omega)^n, \end{cases} \tag{5.15}$$

(5.13) has the appearance of (1.1) with $V = X$ and $Q = L^2(\Omega)^n$.

Here $B : \begin{Bmatrix} H_0^{\text{div}}(\Omega) \times H_0^1(\Omega) & \to L^2(\Omega)^n \\ (\vec{v}, q) & \to B(\vec{v}, q) = \vec{v} - \vec{\text{grad}} q \end{Bmatrix}$.

And $\text{Ker}(B) = \{(\vec{v}, q) \in H_0^{\text{div}}(\Omega) \times H_0^1(\Omega) : \vec{v} = \vec{\text{grad}} q\}$. Thus is $(\vec{v}, q) \in \text{Ker}(B)$ then $\Delta p \in L^2(\Omega)$ and $a((\vec{u}, p), (\vec{v}, q)) = \frac{1}{2}(\text{div}\vec{u}, \text{div}\vec{v})_{L^2} + \frac{1}{2}(\Delta p, \Delta q)_{L^2}$. And when $p \in H^2(\Omega) \cap H_0^1(\Omega)$ we have $||\Delta p||_{L^2} \geq C\,||p||_{H^2} \geq C\,||p||_{H^1}$, cf. (9.29). Thus $a(\cdot, \cdot)$ is coercive on $(\text{Ker}(B), (\cdot, \cdot)_X)$.

But B is not surjective: If $\vec{\ell} \in L^2(\Omega)^n$ we should find $(\vec{v}, q) \in H_0^{\text{div}}(\Omega) \times H_0^1(\Omega)$ s.t. $\vec{\ell} = \vec{v} - \vec{\text{grad}} q$, but we only have (9.31). So $\vec{\lambda}$ has a priori no $||.||_{L^2(\Omega)}$ control, and for the discretization we expect a loss of precision. Here $B^t : \begin{Bmatrix} D \subset L^2(\Omega)^n & \to H_0^{\text{div}}(\Omega)' \times H^{-1}(\Omega) \\ \mu & \to B^t\mu : B^t\mu(\vec{v}, q) = \langle \vec{v}, \mu\rangle_{H_0^{\text{div}\prime},H^{\text{div}}} + \langle q, \text{div}\mu\rangle_{H^{-1},H_0^1} \end{Bmatrix}$ where $D = H_0^{\text{div}}(\Omega)$ (the domain of definition) is not closed in $L^2(\Omega)$ (its closure is $L^2(\Omega)$).

5.3.3 A second constrained formulation

(Not conclusive.) Let

$$X_+ = H_0^1(\Omega)^n \times H_0^1(\Omega), \tag{5.16}$$

provided with the inner product

$$((\vec{u}, p), (\vec{v}, q))_{X_+} = (\text{grad}\vec{u}, \text{grad}\vec{v})_{L^2} + (\vec{\text{grad}} p, \vec{\text{grad}} q)_{L^2(\Omega)}, \tag{5.17}$$

so that X_+ is a Hilbert space.

With a Cartesian basis, we notice that $(\Delta p, \Delta q)_{L^2} = \sum_{ij} \int_\Omega \frac{\partial^2 p}{\partial x_i^2} \frac{\partial^2 q}{\partial x_j^2}\, d\Omega$, and, for $p, q \in H_0^2(\Omega)$,

$$\int_\Omega \frac{\partial^2 p}{\partial x_i^2} \frac{\partial^2 q}{\partial x_j^2}\, d\Omega = -\int_\Omega \frac{\partial^3 p}{\partial x_i^2 \partial x_j} \frac{\partial q}{\partial x_j}\, d\Omega = \int_\Omega \frac{\partial^2 p}{\partial x_i \partial x_j} \frac{\partial^2 q}{\partial x_i \partial x_j}\, d\Omega. \tag{5.18}$$

Thus, with (5.9) and $\vec{v}_p, \vec{v}_q \in \vec{\text{grad}}(H_0^1(\Omega))$ cf., we get

$$(\text{div}\vec{v}_p, \text{div}\vec{v}_q)_{L^2} = (\Delta p, \Delta q)_{L^2} = (\text{grad}(\vec{\text{grad}}p), \text{grad}(\vec{\text{grad}}q))_{L^2} = (\text{grad}\vec{v}_p, \text{grad}\vec{v}_q)_{L^2}. \tag{5.19}$$

(Nota Bene: Here $\vec{v}_p$ and $\vec{v}_q$ derives from a potential.) Thus (5.10) reads: Find $\vec{u} = \vec{v}_p \in \vec{\text{grad}}(H_0^1(\Omega))$ s.t.

$$(\text{grad}\vec{u}, \text{grad}\vec{v}_q)_{L^2} = (f, q)_{L^2}, \quad \forall q \in H_0^1(\Omega). \tag{5.20}$$

So, if $f \in H^{-1}(\Omega)$, then (5.10) is transformed into: Find $((\vec{u}, p), \vec{\lambda}) \in X_+ \times L^2(\Omega)^n$ s.t.

$$\begin{cases} (\text{grad}\vec{u}, \text{grad}\vec{v})_{L^2} + (\vec{\lambda}, \vec{v} - \vec{\text{grad}}q)_{L^2} = \langle f, q \rangle_{H^{-1}, H_0^1}, \quad \forall (\vec{v}, q) \in X_+, \\ (\vec{u} - \vec{\text{grad}}p, \vec{\mu})_{L^2} = 0, \quad \forall \vec{\mu} \in L^2(\Omega)^n, \end{cases} \tag{5.21}$$

$\vec{\lambda}$ being the Lagrangian multiplier of the constraint (5.8). That is,

$$\begin{cases} (\text{grad}\vec{u}, \text{grad}\vec{v})_{L^2} + (\vec{\lambda}, \vec{v})_{L^2} = 0, \quad \forall \vec{v} \in H_0^1(\Omega)^n, \\ -(\vec{\lambda}, \vec{\text{grad}}q)_{L^2} = \langle f, q \rangle_{H^{-1}, H_0^1}, \quad \forall q \in H_0^1(\Omega), \\ (\vec{u}, \vec{\mu})_{L^2} - (\vec{\text{grad}}p, \vec{\mu})_{L^2} = 0, \quad \forall \vec{\mu} \in L^2(\Omega)^n. \end{cases} \tag{5.22}$$

With

$$\begin{cases} a((\vec{u}, p), (\vec{v}, q)) = (\text{grad}\vec{u}, \text{grad}\vec{v})_{L^2} \quad \text{sur } X_+ \times X_+, \\ b((\vec{v}, q), \vec{\mu}) = (\vec{v}, \vec{\mu})_{L^2} - (\vec{\text{grad}}p, \vec{\mu})_{L^2} \quad \text{sur } X_+ \times L^2(\Omega)^n, \end{cases} \tag{5.23}$$

(5.21) has the appearance of (1.1) with $V = X_+$ and $Q = L^2(\Omega)^n$.

Et $B : \left\{ \begin{array}{ll} H_0^1(\Omega)^n \times H_0^1(\Omega) & \to L^2(\Omega)^n \\ (\vec{v}, q) & \to B(\vec{v}, q) = \vec{v} - \vec{\text{grad}}q \end{array} \right\}$. And $(\vec{v}, q) \in \text{Ker}(B)$ iff $\vec{\text{grad}}q = \vec{v} \in H_0^1(\Omega)^n$, thus $a(\cdot, \cdot)$ is coercive on $(\text{Ker}(B), (\cdot, \cdot)_{X_+})$. (Compared to (5.15), here we a $||.||_{H^1}$ for $\vec{u}$, not only a $||.||_{H^{\text{div}}}$ control.)

However B is not surjective. And $\vec{\lambda}$ is not controlled the classical way.

5.3.4 A third constrained formulation

("A good one".) Let

$$Y = H_0^{\text{div}}(\Omega) \times H^1(\Omega), \tag{5.24}$$

provided with the inner product

$$((\vec{u}, p), (\vec{v}, q))_Y = (\text{div}\vec{u}, \text{div}\vec{v})_{L^2} + (p, q)_{H^1} \tag{5.25}$$

so that Y is a Hilbert space.

If $f \in L^2(\Omega)$ (or in $(H^1(\Omega))'$), (5.13) is transformed into: Find $((\vec{u}, p), \vec{\lambda}) \in Y \times H^{\text{div}}(\Omega)$ s.t.

$$\begin{cases} (\text{div}\vec{u}, \text{div}\vec{v})_{L^2} + (\vec{\lambda}, \vec{v})_{L^2} + (q, \text{div}\vec{\lambda})_{L^2} = (f, q)_{L^2}, \quad \forall (\vec{v}, q) \in Y, \\ (\vec{u}, \vec{\mu})_{L^2} + (\vec{\lambda}, \text{div}\vec{\mu}) = 0, \quad \forall \vec{\mu} \in H^{\text{div}}(\Omega), \end{cases} \tag{5.26}$$

that is,

$$\begin{cases} (\text{div}\vec{u}, \text{div}\vec{v})_{L^2} + (\vec{\lambda}, \vec{v})_{L^2} = 0, \quad \forall \vec{v} \in H_0^{\text{div}}(\Omega), \\ (\text{div}\vec{\lambda}, q)_{L^2} = (f, q)_{L^2}, \quad \forall q \in H^1(\Omega), \\ (\vec{u}, \vec{\mu})_{L^2} + (p, \text{div}\vec{\mu})_{L^2} = 0, \quad \forall \vec{\mu} \in H^{\text{div}}(\Omega). \end{cases} \tag{5.27}$$

So $\vec\lambda = \vec{\text{grad}}(\text{div}\vec u)$, $\text{div}\vec\lambda = f$, $\vec u = \vec{\text{grad}} p$, thus $\text{div}(\vec{\text{grad}}(\text{div}\vec{\text{grad}} p)) = f$, i.e. $\Delta^2 p = f$. And $\int_\Gamma p\,\vec\mu.\vec n\,d\Gamma = 0 = \langle p, \vec\mu.\vec n\rangle_{H^{\frac12}(\Gamma),H^{-\frac12}(\Gamma)}$ for all $\vec\mu \in H^{\text{div}}(\Omega)$, thus $p_{|\Gamma} = 0$ in $H^{\frac12}(\Gamma)$ (the trace operator $\vec v \in H^{\text{div}}(\Omega) \to \vec v.\vec n \in H^{-\frac12}(\Gamma)$ is surjective). Thus $p \in H^1_0(\Omega)$. With $\vec u \in H^{\text{div}}_0(\Omega)$, thus $\vec{\text{grad}} p.\vec n = \vec u.\vec n = 0$, so $p \in H^2_0(\Omega)$.

With

$$\begin{cases} a((\vec u, p), (\vec v, q)) = (\text{div}\vec u, \text{div}\vec v)_{L^2} \quad \text{sur } Y \times Y, \\ b((\vec v, q), \vec\mu) = (\vec v, \vec\mu)_{L^2} + (q, \text{div}\vec\mu)_{L^2} \quad \text{sur } Y \times H^{\text{div}}(\Omega), \end{cases} \tag{5.28}$$

(5.26) has the appearance of (1.1) with $V = Y$ and $Q = H^{\text{div}}(\Omega)$.

And $B : \begin{Bmatrix} H^{\text{div}}_0(\Omega) \times H^1(\Omega) & \to H^{\text{div}}(\Omega)' \\ (\vec v, q) & \to B(\vec v, q), \end{Bmatrix}$ with $\langle B(\vec v, q), \vec\mu\rangle = (\vec v, \vec\mu)_{L^2} + (q, \text{div}\vec\mu)_{L^2}$. So B is surjective, cf. (9.25).

And $b((\vec v, q), \vec\mu) = (\vec v, \vec\mu)_{L^2(\Omega)} - (\vec{\text{grad}} q, \vec\mu)_{L^2(\Omega)} + (q, \vec\mu.\vec n)_{L^2(\Gamma)}$ gives $(\vec v, q) \in \text{Ker}(B)$ iff $(\vec v, q) \in H^{\text{div}}_0(\Omega) \times H^1_0(\Omega)$ with $\vec v = \vec{\text{grad}} q$. Thus $a(\cdot,\cdot)$ is coercive on $(\text{Ker}(B), (\cdot,\cdot)_Y)$, and the classical theorem 12.1 apply.

Remark 5.1 (Neumann.) With $Z = H^{\text{div}}(\Omega) \times H^1_0(\Omega)$, (5.26) is transformed into: Find $((\vec u, p), \vec\lambda) \in Z \times H^{\text{div}}_0(\Omega)$ s.t.

$$\begin{cases} (\text{div}\vec u, \text{div}\vec v)_{L^2} + (\vec\lambda, \vec v)_{L^2} = 0, \quad \forall \vec v \in H^{\text{div}}(\Omega), \\ (\text{div}\vec\lambda, q)_{L^2} = (f, q)_{L^2}, \quad \forall q \in H^1_0(\Omega), \\ (\vec u, \vec\mu)_{L^2} + (p, \text{div}\vec\mu)_{L^2} = 0, \quad \forall \vec\mu \in H^{\text{div}}_0(\Omega). \end{cases} \tag{5.29}$$

So, in Ω, $\vec\lambda = \vec{\text{grad}}(\text{div}\vec u)$, $\text{div}\vec\lambda = f$, $\vec u = \vec{\text{grad}} p$, thus $\text{div}(\vec{\text{grad}}(\text{div}\vec{\text{grad}} p)) = f$, i.e. $\Delta^2 p = f$. And on Γ, $\vec{\text{grad}}(\text{div}\vec u).\vec n = 0$, thus $\vec{\text{grad}}(\Delta p).\vec n = 0$ (Neumann limit condition), with $p \in H^1_0(\Omega)$. ▪

5.3.5 A fourth constrained formulation

Let

$$Y_+ = H^1_0(\Omega) \times H^1(\Omega), \tag{5.30}$$

provided with the inner product

$$((\vec u, p), (\vec v, q))_{Y_+} = (\vec{\text{grad}}\vec u, \vec{\text{grad}}\vec v)_{L^2} + (p, q)_{H^1}. \tag{5.31}$$

And (5.27) is replaced with

$$\begin{cases} (\text{grad}\vec u, \text{grad}\vec v)_{L^2} + (\vec\lambda, \vec v)_{L^2} = 0, \quad \forall \vec v \in H^{\text{div}}_0(\Omega), \\ (\text{div}\vec\lambda, q)_{L^2} = (f, q)_{L^2}, \quad \forall q \in H^1(\Omega), \\ (\vec u, \vec\mu)_{L^2} + (p, \text{div}\vec\mu)_{L^2} = 0, \quad \forall \vec\mu \in H^{\text{div}}(\Omega). \end{cases} \tag{5.32}$$

And (5.28) is replaced with

$$\begin{cases} a((\vec u, p), (\vec v, q)) = (\text{grad}\vec u, \text{grad}\vec v)_{L^2} \quad \text{sur } Y_+ \times Y_+, \\ b((\vec v, q), \vec\mu) = (\vec v, \vec\mu)_{L^2} + (q, \text{div}\vec\mu)_{L^2} \quad \text{sur } Y_+ \times H^{\text{div}}(\Omega) \end{cases} \tag{5.33}$$

6 Locking

The locking phenomenon appears when the coercivity of the approximated problem increases much faster than the coercivity of the continuous problem. Thus the numerical solution is close to zero, which is absurd in general.

Let Ω be bounded open set in $\mathbb{R}^n$.

6.1 A typical situation

Let $\lambda \in \mathbb{R}$ so that $\lambda >> 1$ (a "large" given real). We look for $\vec{u} \in H_0^1(\Omega)^n$ and $p \in H_0^1(\Omega)$ that minimize

$$M(\vec{v},q) = \frac{1}{2}||\mathrm{grad}\vec{v}||^2_{L^2(\Omega)} + \frac{\lambda}{2}||\vec{v} - \vec{\mathrm{grad}}q||_{L^2(\Omega)} - (\vec{f},\vec{v})_{L^2(\Omega)} - (g,q)_{L^2(\Omega)}. \tag{6.1}$$

Example 6.1 For the Mindlin–Reissner problem, $||\mathrm{grad}\vec{v}||^2_{L^2(\Omega)}$ is replaced by $|a(\vec{v},\vec{v})|$ where $a(\cdot,\cdot)$ is a bilinear form that is continuous and coercive on $H_0^1(\Omega)$. ▪

Let

$$X = H_0^1(\Omega)^n \times H_0^1(\Omega) \tag{6.2}$$

provided with the inner product associated to the norm

$$||(\vec{v},q)||_X = (||\mathrm{grad}\vec{v}||^2_{L^2} + ||\vec{\mathrm{grad}}q||^2_{L^2})^{\frac{1}{2}} \tag{6.3}$$

so that X is a Hilbert space.

A solution $(\vec{u},p) \in X$ realizing the min of M satisfies

$$\begin{cases} (\mathrm{grad}\vec{u},\mathrm{grad}\vec{v})_{L^2} + \lambda(\vec{u} - \vec{\mathrm{grad}}p,\vec{v})_{L^2} = (\vec{f},\vec{v}), & \forall \vec{v} \in H_0^1(\Omega)^n, \\ \lambda(\vec{u} - \vec{\mathrm{grad}}p, \vec{\mathrm{grad}}q)_{L^2} = (g,q), & \forall q \in H^1(\Omega). \end{cases} \tag{6.4}$$

Let

$$\Phi((\vec{u},p),(\vec{v},q)) = (\mathrm{grad}\vec{u},\mathrm{grad}\vec{v})_{L^2} + \lambda(\vec{u} - \vec{\mathrm{grad}}p, \vec{v} - \vec{\mathrm{grad}}q)_{L^2}. \tag{6.5}$$

Thus (6.4) reads: Find $(\vec{u},p) \in X$ s.t.

$$\Phi((\vec{u},p),(\vec{v},q)) = (\vec{f},\vec{v})_{L^2} + (g,q)_{L^2}, \quad \forall (\vec{v},q) \in X. \tag{6.6}$$

Proposition 6.2 *The bilinear form $\Phi : X \times X \to \mathbb{R}$ is coercive and continuous on X: with the Poincaré inequality (9.28) we have*

$$\begin{cases} \exists \alpha_\Phi > 0, \quad \forall (\vec{v},q) \in X, \quad \Phi((\vec{v},q),(\vec{v},q)) \geq \alpha_\Phi ||(\vec{v},q)||^2_X, \quad et \quad \alpha_\Phi \underset{\lambda\to\infty}{\sim} \frac{1}{c_\Omega}, \\ \exists C > 0, \quad \forall (\vec{u},p),(\vec{v},q) \in X, \quad \Phi((\vec{u},p),(\vec{v},q)) \leq C||(\vec{u},p)||_X||(\vec{v},q)||_X, \quad et \quad C \underset{\lambda\to\infty}{=} O(\lambda). \end{cases} \tag{6.7}$$

And problem (6.6) is well posed.

Proof. Bi-linearity. Since Φ is symmetric (trivial), that is $\Phi((\vec{u},p),(\vec{v},q)) = \Phi((\vec{v},q),(\vec{u},p))$, we have to prove that $\Phi((\vec{u}_1,p_1) + \alpha(\vec{u}_2,p_2),(\vec{v},q)) = \Phi((\vec{u}_1,p_1),(\vec{v},q)) + \alpha\Phi((\vec{u}_2,p_2),(\vec{v},q))$: trivial.

Coercivity. If $\kappa > 0$ then, with (9.28) :

$$\begin{aligned} \Phi((\vec{v},q),(\vec{v},q)) &= ||\mathrm{grad}\vec{v}||^2_{L^2} + \lambda||\vec{v} - \vec{\mathrm{grad}}q||^2_{L^2} \\ &\geq [(1-\kappa)||\mathrm{grad}\vec{v}||_{L^2} + c_\Omega\kappa||\vec{v}||^2_{L^2}] + \lambda[||\vec{v}||^2_{L^2} + ||\vec{\mathrm{grad}}q||^2_{L^2} - 2||\vec{v}||_{L^2}||q||_{L^2}]. \end{aligned}$$

Let $x = ||\vec{v}||_{L^2}$ and $y = ||\vec{\mathrm{grad}}q||_{L^2}$. We have

$$c_\Omega\kappa x^2 + \lambda(x-y)^2 \geq \frac{\lambda c_\Omega\kappa}{\lambda + c_\Omega\kappa}y^2,$$

with $c_{\max} = \frac{\lambda c_\Omega\kappa}{\lambda + c_\Omega\kappa}$ the largest constant possible ($c_{\max}$ is largest c s.t. $c_\Omega\kappa x^2 + \lambda(x-y)^2 \geq cy^2$: easy check). Thus

$$\Phi((\vec{v},q),(\vec{v},q)) \geq (1-\kappa)||\mathrm{grad}\vec{v}||_{L^2} + \frac{\lambda c_\Omega\kappa}{\lambda + c_\Omega\kappa}||\vec{\mathrm{grad}}q||_{L^2}.$$

And the "best" α_Φ (the largest possible) is obtained by choosing κ s.t. $1-\kappa = \frac{\lambda c_\Omega \kappa}{\lambda + c_\Omega \kappa}$, i.e. κ solution of $\kappa^2 + b\kappa - \frac{\lambda}{c_\Omega} = 0$ where $b = \lambda \frac{c_\Omega + 1}{c_\Omega} - 1$. The discriminant is $b^2 + 4\frac{\lambda}{c_\Omega}$, and the positive root is $\kappa = \frac{b}{2}(-1 + \sqrt{1 + 4\frac{\lambda}{b^2 c_\Omega}})$. And for $\lambda >> 1$, we have $b \simeq \lambda$, thus $4\frac{\lambda}{b^2 c_\Omega} \simeq 4\frac{1}{\lambda c_\Omega}$, thus $-1 + \sqrt{1 + 4\frac{\lambda}{b^2 c_\Omega}} \simeq \frac{2}{\lambda c_\Omega}$, so $\kappa \simeq \frac{1}{c_\Omega}$ in the vicinity of $\lambda = \infty$.

Continuity: Easy check. ∎

Remark 6.3 For the associated numerical approximation with finite elements, it λ is "large" then difficulties are expected, since $C = O(\lambda)$. Indeed, the conditioning of the associated matrix is $\simeq \frac{C}{\alpha_\Phi} = 0(\lambda)$, and this conditioning explodes with λ. However, a bad choice of the discrete spaces leads to a much faster explosion than expected, see proposition 6.5. ∎

6.2 The coercivity constant for p

For the analysis of the locking phenomenon (due to a "bad choice" of the discrete spaces), let us look at the coercivity constant for p (where λ appears):

Proposition 6.4 *If $(\vec{v}, q) \in X$ then, with (9.28),*

$$\Phi((\vec{v}, q), (\vec{v}, q)) \ge c_\Omega \frac{\lambda}{\lambda + c_\Omega} ||\vec{\text{grad}} q||^2_{L^2}, \tag{6.8}$$

and

$$c_\Omega \frac{\lambda}{\lambda + c_\Omega} \simeq c_\Omega \quad \text{as} \quad \lambda \to \infty. \tag{6.9}$$

Proof. Modification or the previous proof:

$$\Phi((\vec{v}, q), (\vec{v}, q)) \ge c_\Omega ||\vec{v}||^2_{L^2} + \lambda[||\vec{v}||^2_{L^2} + ||\vec{\text{grad}} q||^2_{L^2} - 2||\vec{v}||_{L^2} ||q||_{L^2}] \ge \alpha_p ||\vec{\text{grad}} q||^2_{L^2},$$

and the largest α_p possible is $\alpha_p = c_\Omega \frac{\lambda}{\lambda + c_\Omega}$ (has to satisfy "$c_\Omega x^2 + \lambda(x-y)^2 \ge \alpha_p y^2$). ∎

6.3 The discrete problem

Let $V_h \subset H^1_0(\Omega)^n$ be a finite dimension subspace. Let $\Pi_{V_h} : L^2(\Omega)^n \to V_h$ be the $(\cdot, \cdot)_{L^2}$ projection onto V_h, that is,

$$\forall \vec{v} \in H^1_0(\Omega)^n, \quad \forall \vec{w}_h \in V_h, \quad (\Pi_{V_h} \vec{v}, \vec{w}_h)_{L^2(\Omega)} = (\vec{v}, \vec{w}_h)_{L^2(\Omega)}.$$

Let $Q_h \subset H^1_0(\Omega)$ be a finite dimension subspace.

Let $X_h = V_h \times Q_h$. The discrete problem associated to (6.6) is: Find $(\vec{u}_h, p_h) \in X_h$ s.t.

$$\Phi((\vec{u}_h, p_h), (\vec{v}_h, q_h)) = (\vec{f}, \vec{v}_h) + (g, q_h), \qquad \forall (\vec{v}_h, q_h) \in X_h. \tag{6.10}$$

Proposition 6.5 *If $(\vec{v}_h, q_h) \in X_h$ then*

$$\Phi((\vec{v}_h, q_h), (\vec{v}_h, q_h)) \ge c_\Omega \frac{\lambda}{\lambda + c_\Omega} ||\vec{\text{grad}} q_h||^2_{L^2} + \lambda \frac{\lambda}{\lambda + c_\Omega} ||\vec{\text{grad}} q_h - \Pi_{V_h} \vec{\text{grad}} q_h||^2_{L^2(\Omega)}, \tag{6.11}$$

to be compared with (6.8).

Illustration: If V_h is "small relatively to Q_h" so that for some q_h the real $||\vec{\text{grad}} q_h - \Pi_{V_h} \vec{\text{grad}} q_h||_{L^2(\Omega)}$ does not vanish (fast enough with h) then the right hand side of (6.11) increases with λ, to compare with (6.9). And the solution $(u_n, p_h) \in X_h$ is bounded by the inverse constant that decreases with λ, thus (u_h, p_h) decreases to zero as λ increases: We get the "locking" phenomenon.

Proof. Let $(\vec{u}_h, p_h)$ and $(\vec{v}_h, q_h) \in X_h$. Then

$$\begin{aligned}\Phi((\vec{u}_h, p_h), (\vec{v}_h, q_h)) &= (\text{grad}\vec{u}_h, \text{grad}\vec{v}_h)_{L^2} + \lambda(\vec{\text{grad}} p_h - \vec{u}_h, \vec{\text{grad}} q_h - \vec{v}_h)_{L^2} \\ &= (\text{grad}\vec{u}_h, \text{grad}\vec{v}_h)_{L^2} + \lambda(\vec{\text{grad}} p_h - \Pi_{V_h} \vec{\text{grad}} p_h, \vec{\text{grad}} q_h - \Pi_{V_h} \vec{\text{grad}} q_h)_{L^2} \\ &\quad + \lambda(\Pi_{V_h} \vec{\text{grad}} p_h - \vec{u}_h, \Pi_{V_h} \vec{\text{grad}} q_h - \vec{v}_h)_{L^2}.\end{aligned}$$

Thus

$$\Phi((\vec{v}_h, q_h), (\vec{v}_h, q_h)) \ge c_\Omega \frac{\lambda}{\lambda + c_\Omega} ||\Pi_{V_h} \vec{\text{grad}} q_h||^2_{L^2} + \lambda ||\vec{\text{grad}} q_h - \Pi_{V_h} \vec{\text{grad}} q_h||^2_{L^2},$$

see previous §computation. And Pythagoras give (6.5). ∎

Remark 6.6 The term $\vec{\text{grad}} q_h - \Pi_{V_h} \vec{\text{grad}} q_h$ is also a problem for the Stokes equations, see § 3.6. ∎

6.4 An optimal correction

Due to the choice of V_h and Q_h, we eventually have too much of coercivity, cf. (6.5), so we decide to get rid of it. That is, we modify (6.1) to get: Find $(\vec{u}, p) \in V_h \times Q_h$ realizing the minimum of

$$\begin{aligned} M_h(\vec{v}, q) = \frac{1}{2}||\text{grad}\vec{v}_h||_{L^2}^2 + \frac{\lambda}{2}(||\vec{v}_h - \vec{\text{grad}}q_h||_{L^2}^2 - \lambda\frac{\lambda}{\lambda + c_\Omega}||\vec{\text{grad}}q_h - \Pi_{V_h}\vec{\text{grad}}q_h||_{L^2}^2) \\ - (\vec{f}, \vec{v}_h)_{L^2} - (g, q_h)_{L^2}. \end{aligned} \tag{6.12}$$

Thus Φ_h has been transformed into

$$\begin{aligned} \Phi_h((\vec{u}_h, p_h), (\vec{v}_h, q_h)) = (\text{grad}\vec{u}_h, \text{grad}\vec{v}_h)_{L^2} + \lambda(\vec{u}_h - \vec{\text{grad}}p_h, \vec{v} - \vec{\text{grad}}q_h)_{L^2} \\ - \frac{\lambda^2}{\lambda + c_\Omega}(\vec{\text{grad}}p_h - \Pi_{V_h}\vec{\text{grad}}p_h, \vec{\text{grad}}q_h - \Pi_{V_h}\vec{\text{grad}}q_h)_{L^2}, \end{aligned}$$

and the problem reads: Find $(\vec{u}_h, p_h) \in V_h \times Q_h$ s.t., for all $(\vec{v}_h, q_h) \in V_h \times Q_h$,

$$\Phi_h((\vec{u}_h, p_h), (\vec{v}_h, q_h)) = (\vec{f}, \vec{v}_h)_{L^2} + (g, q_h)_{L^2}.$$

To solve this problem, we need to compute $\Pi_{V_h}\vec{\text{grad}}p_h$: If the $V_h = P_1$-continuous finite elements is made, the computation amounts to inverse a diagonal matrix, thanks to the mass-lumping technique, thus is costless.

Computation: we have to compute $(\vec{u}_h, p_h) \in V_h \times Q_h$ s.t., for all $(\vec{v}_h, q_h) \in V_h \times Q_h$,

$$\begin{cases} (\text{grad}\vec{u}_h, \text{grad}\vec{v}_h)_{L^2} + \lambda(\vec{u}_h - \vec{\text{grad}}p_h, \vec{v}_h)_{L^2} = (\vec{f}, \vec{v}_h)_{L^2}, \\ -\lambda(\vec{u}_h - \vec{\text{grad}}p_h, \vec{\text{grad}}q_h)_{L^2} - \lambda\frac{\lambda}{\lambda + c_\Omega}(\vec{\text{grad}}p_h - \Pi_{V_h}\vec{\text{grad}}p_h, \vec{\text{grad}}q_h)_{L^2} = (g, q_h)_{L^2}. \end{cases}$$

Introducing $\vec{w}_h = \Pi_{V_h}\vec{\text{grad}}p_h$, we have to find $(\vec{u}_h, p_h, \vec{w}_h) \in V_h \times Q_h \times V_h$ s.t., for all $(\vec{v}_h, q_h) \in V_h \times Q_h$,

$$\begin{cases} (\text{grad}\vec{u}_h, \text{grad}\vec{v}_h)_{L^2} + \lambda(\vec{u}_h, \vec{v}_h)_{L^2} - \lambda(\vec{\text{grad}}p_h, \vec{v}_h)_{L^2} = (\vec{f}, \vec{v}_h)_{L^2}, \\ -\lambda(\vec{u}_h, \vec{\text{grad}}q_h)_{L^2} + \frac{c_\Omega\lambda}{\lambda + c_\Omega}(\vec{\text{grad}}p_h, \vec{\text{grad}}q_h)_{L^2} + \lambda\frac{\lambda}{\lambda + c_\Omega}(\vec{w}_h, \vec{\text{grad}}q_h)_{L^2} = (g, q_h)_{L^2}, \\ (\vec{\text{grad}}p_h, \vec{w}_h')_{L^2} - (\vec{w}_h, \vec{w}_h')_{L^2} = 0. \end{cases} \tag{6.13}$$

This method gives optimal convergence results. E.g., for P_1-continuous finite elements of $\vec{u}_h$ and p_h, we get an $O(h)$ convergence.

6.5 Classical treatment of the locking

6.5.1 Initial problem

See e.g. Chapelle [10], Brezzi and Fortin [8]. The variable

$$\vec{\gamma} = \lambda(\vec{u} - \vec{\text{grad}}p) \tag{6.14}$$

is introduced. Then problem (6.4) becomes: Find $(\vec{u}, p, \vec{\gamma}) \in H_0^1(\Omega)^n \times H_0^1(\Omega) \times Y$ s.t.

$$\begin{cases} a((\vec{u}, p), (\vec{v}, q)) + b((\vec{v}, q), \vec{\gamma}) = (\vec{f}, \vec{v})_{L^2} + (g, q)_{L^2}, & \forall(\vec{v}, q) \in X, \\ b((\vec{u}, p), \vec{\delta}) - \frac{1}{\lambda}\langle\vec{\delta}, \vec{\gamma}\rangle_{H^{-1},H_0^1} = 0, & \forall\vec{\delta} \in Y, \end{cases} \tag{6.15}$$

with

$$\begin{cases} Y = \{\vec{\delta} \in (H^{-1}(\Omega))^n : \text{div}\vec{\delta} \in H^{-1}(\Omega)\}, \\ a(\cdot,\cdot) : X \times X \to \mathbb{R} : \quad a((\vec{u}, p), (\vec{v}, q)) = (\text{grad}\vec{u}, \text{grad}\vec{v})_{L^2}, \\ b(\cdot,\cdot) : X \times Y \to \mathbb{R} : \quad b((\vec{u}, p), \vec{\delta}) = \langle\vec{\delta}, \vec{u}\rangle_{H^{-1},H_0^1} - \langle\text{div}\vec{\delta}, p\rangle_{H^{-1},H_0^1}. \end{cases} \tag{6.16}$$

(And $(6.15)_2$ gives $\vec{u} - \vec{\text{grad}}p - \frac{1}{\lambda}\vec{\gamma} = 0$.) And it is shown that Y is a Banach space for the norm

$$||\vec{\delta}||_Y \stackrel{\text{def}}{=} ||\vec{\delta}||_{H^{-1}(\Omega)^n} + ||\text{div}\vec{\delta}||_{H^{-1}(\Omega)}.$$

Remark 6.7 Y is the space corresponding to $\frac{1}{\lambda} = 0$ (i.e. λ "infinitely large"), cf. Kirchhoff–Love shell model:

$$\begin{cases} a((\vec{u},p),(\vec{v},q)) + b((\vec{v},q),\vec{\gamma}) = (\vec{f},\vec{v})_{L^2} + (g,q)_{L^2}, & \forall(\vec{v},q) \in X, \\ b((\vec{u},p),\vec{\delta}) = 0, \quad \forall\vec{\delta} \in Y, \end{cases} \tag{6.17}$$

to be compared with (6.15).

For a discretization with finite elements, finite dimensional spaces are often chosen s.t. $V_h \subset L^2(\Omega)^n$, $Y_h \subset L^2(\Omega)^n$ and $Q_h \subset C^0(\Omega;\mathbb{R})$; And then (6.15) is meaningful in $V_h \times Q_h \times Y_h$ with $b((\vec{u}_h,p_h),\vec{\delta}_h) = (\vec{\delta}_h, \vec{u}_h - \vec{\text{grad}} p_h)_{L^2(\Omega)}$. ▪

Let $B : X \to Y'$ be the operator associated to $b(\cdot,\cdot)$, that is, $\langle B(\vec{v},q),\vec{\delta}\rangle_{H_0^1,H^{-1}} := b((\vec{v},q),\vec{\delta})$, i.e.,

$$B(\vec{v},q) = \vec{v} - \vec{\text{grad}} q.$$

Then, with Poincaré inequality, it is easy to check that $a(\cdot,\cdot)$ is coercive on $\text{Ker}(B) = \{(\vec{v},q) \in X : \vec{v} = \vec{\text{grad}} q\}$.

Then is shown that B is surjective (inf-sup condition), that is,

$$\exists k > 0, \quad \forall\vec{\delta} \in Y, \quad \sup_{(\vec{v},q)\in X} \frac{b((\vec{v},q),\vec{\delta})}{||(\vec{v},q)||_X ||\vec{\delta}||_Y} \geq k.$$

See e.g. Chapelle [10], Brezzi et Fortin [8].

6.5.2 discrete problem

The discrete inf-sup condition has to be satisfied: this lead to numerous articles. There are two difficulties:

1- An adequat choice of finite element spaces to satisfy the inf-sup condition (it the stabilization is not used), la stabilisation de $\vec{\gamma}_h$, ou le choix adequat d'éléments finis compatibles pour satisfaire la condition inf-sup,

2- An adequat choice of finite element spaces to satisfy the coercivity of $a(\cdot,\cdot)$ on the kernel $\text{Ker}(B_h)$ (with B_h the discrete operator). But this problem can be easily fixed by modiying (6.1) into

$$\tilde{M}(\vec{v},q) = \frac{1}{2}||\text{grad}\vec{v}||^2_{L^2(\Omega)} + \frac{1}{2}||\vec{v} - \vec{\text{grad}} q||_{L^2(\Omega)} + \frac{\lambda - 1}{2}||\vec{v} - \vec{\text{grad}} q||_{L^2(\Omega)} - (\vec{f},\vec{v})_{L^2(\Omega)} - (g,q)_{L^2(\Omega)}, \tag{6.18}$$

that is, by replacing $a(\cdot,\cdot)$, cf. (6.16), with

$$\tilde{a}((\vec{u},p),(\vec{v},q)) = (\text{grad}\vec{u}, \text{grad}\vec{v})_{L^2} + (\vec{u} - \vec{\text{grad}} p, \vec{v} - \vec{\text{grad}} q)_{L^2}.$$

And we consider (6.15) with $\tilde{a}(\cdot,\cdot)$ instead of $a(\cdot,\cdot)$.

Now $\tilde{a}(\cdot,\cdot)$ is coercive sur X (thanks to (9.28)), thus on $\text{Ker}(B) = \{(\vec{v},q) : \vec{v} = \vec{\text{grad}} q\}$, and we get a similar problem to the Stokes problem (choice of adequat finite element spaces, or choice of a stabilization).

7 Weak Dirichlet condition

(See Babuška [2] for the initial manuscript.)

7.1 Initial problem

Let $f \in L^2(\Omega)$ and $d \in H^{\frac{1}{2}}(\Gamma)$. Let $u_d \in H^1(\Omega)$ s.t. $u_{d|\Gamma} = d$; Such a function u_d exists since the trace operator $\Gamma_0 : H^1(\Omega) \to L^2(\Gamma)$ is surjective onto Γ, cf. (9.8). Let $d + H_0^1(\Omega) := u_d + H_0^1(\Omega) = \{u_d + v,\ v \in H_0^1(\Omega)\}$, affine space in $H^1(\Omega)$ independent of the choice of u_d the reverse image of d by γ_0 (trivial).

Consider the problem: Find $u \in d + H_0^1(\Omega)$ s.t.

$$\begin{cases} -\Delta u + u = f \quad \text{dans } \Omega, \\ u_{|\Gamma} = d \quad \text{sur } \Gamma. \end{cases} \tag{7.1}$$

Thanks to the Lax–Milgram theorem, this problem is well-posed.

7.2 Mixed problem

The aim is to impose the Dirichlet condition with a Lagrangian multiplier. So the problem becomes: Find $(u, \lambda) \in H^1(\Omega) \times H^{-\frac{1}{2}}(\Gamma)$ s.t.

$$\begin{cases} (u, v)_{H^1(\Omega)} + \langle \lambda, v \rangle_{H^{-\frac{1}{2}}(\Gamma), H^{\frac{1}{2}}(\Gamma)} = (f, v)_{L^2(\Omega)}, \quad \forall v \in H^1(\Omega), \\ \langle u, \mu \rangle_{H^{\frac{1}{2}}(\Gamma), H^{-\frac{1}{2}}(\Gamma)} = \langle d, \mu \rangle_{H^{\frac{1}{2}}(\Gamma), H^{-\frac{1}{2}}(\Gamma)}, \quad \forall \mu \in H^{-\frac{1}{2}}(\Gamma). \end{cases} \tag{7.2}$$

If (u, λ) exists in $H^1(\Omega) \times H^{-\frac{1}{2}}(\Gamma)$, then we get:

$$\begin{cases} -\Delta u + u = f \ \in L^2(\Omega), \\ u = d \ \in H^{\frac{1}{2}}(\Gamma), \\ \lambda = -\dfrac{\partial u}{\partial n} \ \in H^{\frac{1}{2}}(\Gamma). \end{cases} \tag{7.3}$$

Interpretation of the Lagrangian multiplier: λ is, up to the sign, the force $\vec{\text{grad}} u.\vec{n}$ needed on Γ for u to stay equal to d on Γ.

With

$$\begin{cases} a(u, v) = (u, v)_{H^1(\Omega)}, \\ b(v, \lambda) = \langle v, \lambda \rangle_{H^{\frac{1}{2}}(\Gamma), H^{-\frac{1}{2}}(\Gamma)}, \end{cases} \tag{7.4}$$

(7.2) has the appearance of (1.1) with $V = H^1(\Omega)$ and $Q = H^{-\frac{1}{2}}(\Gamma)$. And $a(\cdot, \cdot)$ is bilinear (trivial) continuous and coercive (it is the $V = H^1(\Omega)$-inner product), and $b(\cdot, \cdot)$ is bilinear (trivial) continuous since $|b(v, \lambda)| \le ||v||_{H^{\frac{1}{2}}(\Gamma)} ||\lambda||_{H^{-\frac{1}{2}}(\Gamma)}$ and γ_0 is continuous (so $||v||_{H^{\frac{1}{2}}(\Gamma)} \le ||\gamma_0|| \, ||v||_{H^1(\Omega)}$).

We have $Q' = H^{\frac{1}{2}}(\Gamma)$, so $B : \left\{ \begin{array}{rl} H^1(\Omega) & \to H^{\frac{1}{2}}(\Gamma) \\ v & \to Bv = \gamma_0(v) \end{array} \right\}$ is linear continuous (since $b(\cdot, \cdot)$ is bilinear continuous) and surjective (definition of $H^{\frac{1}{2}}(\Gamma)$), thus $\text{Im}(B^t)$ is closed in $V' = H^1(\Omega)'$, with B^t : $\left\{ \begin{array}{rl} H^{-\frac{1}{2}}(\Gamma) & \to H^1(\Omega)' \\ \lambda & \to B^t \lambda \end{array} \right\}$ defined by $\langle B^t \lambda, v \rangle_{H^1(\Omega)', H^1(\Omega)} = \langle v, \lambda \rangle_{H^{\frac{1}{2}}(\Gamma), H^{-\frac{1}{2}}(\Gamma)}$. Thus

$$\exists k > 0,\ \forall \lambda \in H^{-\frac{1}{2}}(\Gamma),\ ||B^t \lambda||_{H^1(\Omega)'} \ge k \, ||\lambda||_{H^{-\frac{1}{2}}(\Gamma)/\text{Ker} B^t}. \tag{7.5}$$

(That is, $\exists k > 0, \ \inf_{\lambda \in H^{-\frac{1}{2}}(\Gamma)} \sup_{v \in H^1(\Omega)} |b(\frac{v}{||v||_{H^1(\Omega)}||}, \frac{\lambda}{\lambda||_{H^{-\frac{1}{2}}(\Gamma)}})| \ge k$.)

Remark 7.1 The computation of λ may give disappointing results since the control for λ is done with the $||.||_{H^{-\frac{1}{2}}(\Gamma)}$-norm, cf. (7.5) (not even a $||.||_{L^2(\Gamma)}$ control). So numerical problems are expected. ▟

Remark 7.2 This mixed problem leads to "transmission problems" (or hybrid problems) with "mortar finite elements", see Bernardi, Maday, Patera. ▟

The associated Lagrangian is (saddle point problem)

$$L(u, \lambda) = \frac{1}{2}(||\vec{\text{grad}} u||^2_{L^2(\Omega)} + ||u||^2_{L^2(\Omega)}) + \langle u - d, \lambda \rangle_{H^{\frac{1}{2}}(\Gamma), H^{-\frac{1}{2}}(\Gamma)} - (f, v)_{L^2(\Omega)}, \tag{7.6}$$

7.3 Discrete problem

Let $V_h \subset H^1(\Omega)$ and $\Lambda_h \subset H^{-\frac{1}{2}}(\Gamma)$ be finite dimension spaces. The discrete problem relative to (7.2) is: Find $(u_h, \lambda_h) \in V_h \times \Lambda_h$ s.t.

$$\left\{ \begin{array}{lll} (u_h, v_h)_{H^1(\Omega)} + (v_h, \lambda_h)_{L^2(\Gamma)} = (f, v_h)_{L^2(\Omega)}, & & \forall v_h \in V_h, \\ (u_h, \mu_h)_{L^2(\Gamma)} \qquad\qquad\qquad\quad = (d, \mu_h)_{L^2(\Gamma)}, & & \forall \mu_h \in \Lambda_h. \end{array} \right. \tag{7.7}$$

7.4 Finite elements P_k–C^0: unstable

7.4.1 The discrete inf-sup condition

Consider the Lagrange finite elements $V_h = P_k$–C^0 in Ω and $\Lambda_h = \gamma_0(V_h)$ P_k–C^0 on Γ. The discrete (trace) operator $B_h = \gamma_{0|\Gamma} : \left\{ \begin{array}{rl} (V_h, ||.||_{H^1(\Omega)}) & \to (\Lambda_h, ||.||_{H^{\frac{1}{2}}(\Gamma)}) \\ v_h & \to \gamma_0(v_h) = v_{h|\Gamma} \end{array} \right\}$ is continuous and surjective (trivial here with P_k-C^0 finite elements for both V_h and Λ_h), thus (7.5) holds with some k_h instead of k, and Q_h instead of Q. But the control on λ_h is very weak (a $H^{-\frac{1}{2}}(\Gamma)$ control), and k_h a priori depends on h.

7.4.2 Barbosa et Hughes

(See Barbosa and Hughes [3], Pitkäranta [27], Stenberg [31]).

A finite element mesh $\mathcal{T}_h$ is defined in Ω, and the trace of this mesh on Γ will be used as a mesh on Γ.

Barbosa and Hughes stabilize the Lagrangian multiplier λ with its value $\lambda = -\frac{\partial u}{\partial n}$, cf. (7.3). Thus the problem now reads: Find $(u_h, \lambda_h) \in V_h \times \Lambda_h$ s.t., for all $(v_h, \mu_h) \in V_h \times \Lambda_h$,

$$\left\{ \begin{array}{l} (u_h, v_h)_{H^1(\Omega)} + \displaystyle\int_\Gamma v_h \lambda_h \, d\Gamma - \alpha h \int_\Gamma (\lambda_h + \frac{\partial u_h}{\partial n}) \frac{\partial v_h}{\partial n} \, d\Gamma = (f, v_h)_{L^2(\Omega)}, \\ \displaystyle\int_\Gamma u_h \mu_h \, d\Gamma - \alpha h \int_\Gamma (\lambda_h + \frac{\partial u_h}{\partial n}) \mu_h \, d\Gamma = \int_\Gamma d \mu_h \, d\Gamma, \end{array} \right. \tag{7.8}$$

with α a constant to be chosen, corresponding to the saddle point of the modified Lagrangian

$$L_h(u, \lambda) = L(u, \lambda) - \alpha \, h \, ||\lambda + \frac{\partial u}{\partial n}||^2_{L^2(\Gamma)}, \tag{7.9}$$

cf. (7.6). See Stenberg [31].

We then get the "penalized" problem, written here as the matrix problem

$$\begin{pmatrix} A & B^t \\ B & -\alpha C \end{pmatrix} . \begin{pmatrix} \vec{u} \\ p \end{pmatrix} = \begin{pmatrix} \vec{f} \\ \vec{g} \end{pmatrix}. \tag{7.10}$$

Theorem 7.3 *(Barbosa and Hughes [3], Pitkäranta [27].) The mesh $\mathcal{T}_h$ is supposed to be quasi-uniform, that is, the following inverse inequality is true:*

$$\exists C_i > 0, \quad \forall v_h \in V_h, \quad h^{\frac{1}{2}} \, ||\frac{\partial v_h}{\partial n}||_{L^2(\Gamma)} \leq C_i \, ||\vec{\text{grad}} v_h||_{L^2(\Omega)}. \tag{7.11}$$

And α is supposed small enough (not to destroy the coercivity for u), namely:

$$0 < \alpha < \frac{1}{C_i}. \tag{7.12}$$

Then the stabilized problem (7.8) is well posed, and for P_k-C^0 finite elements, as soon as the exact solution u is in $H^{k+1}(\Omega)$, and we get the usual a priori estimate

$$||u - u_h||_{H^1(\Omega)} \leq C h^k ||u||_{H^{k+1}(\Omega)},$$

C being a constant independent of h.

Proof. See Barbosa–Hughes [3] and Stenberg [31]. ■

Remark 7.4 The inequality (7.11) also reads

$$h \int_\Gamma (\vec{\text{grad}} v . \vec{n})^2 \, d\Gamma \leq C_i^2 \int_\Omega ||\vec{\text{grad}} v||^2_{\mathbb{R}^n} \, d\Omega,$$

where h on the left hand side is expected: $(h \int_\Gamma)$ has a volume dimension, same dimension as $(\int_\Omega)$, for a quasi-uniform mesh. ■

7.4.3 Multiplier elimination: Nitsche method

Stenberg [31] has shown that Barbosa and Hughes [3] method is equivalent to Nitsche [26] method when $\Lambda_h = P_0(\Gamma_h)$ and $V_h = P_1(\Omega_h)$ (when the mesh on Γ is the trace of the mesh in Ω): Find $u_h \in V_h$ s.t., for all $v_h \in V_h$,

$$(u_h, v_h)_{H^1(\Omega)} - \langle \frac{\partial u_h}{\partial n}, v_h \rangle_\Gamma - \langle \frac{\partial v_h}{\partial n}, u_h - d \rangle_\Gamma + \gamma \sum_{E \in \mathcal{E}_h} \frac{1}{h_E} \langle u_h - d, v_h \rangle_E = (f, v_h)_{L^2(\Omega)},$$

for some $\gamma > 0$, i.e., find $u_h \in V_h$ s.t., for all $v_h \in V_h$,

$$\left\{ \begin{aligned} & (u_h, v_h)_{H^1(\Omega)} - \langle \frac{\partial u_h}{\partial n}, v_h \rangle_\Gamma - \langle \frac{\partial v_h}{\partial n}, u_h \rangle_\Gamma + \gamma \sum_{E \in \mathcal{E}_h} \frac{1}{h_E} \langle u_h, v_h \rangle_E \\ & \qquad = (f, v_h)_{L^2(\Omega)} - \langle \frac{\partial v_h}{\partial n}, d \rangle_\Gamma + \gamma \sum_{E \in \mathcal{E}_h} \frac{1}{h_E} \langle d, v_h \rangle_E. \end{aligned} \right. \tag{7.13}$$

We then get $u_\Gamma = d$. This method is simpler to compute since no Lagrangian multiplier intervenes.

Proposition 7.5 *If (7.11), if $\gamma > C_i$, if $V_h = P_k$-C^0 and $u \in H^{k+1}(\Omega)$, then (7.13) gives the usual result:*

$$||u - u_h||_{H^1(\Omega)} \le C h^k ||u||_{H^{k+1}(\Omega)}.$$

Proof. See Stenberg [31]. ∎

Comparison of the method of Nitsche with the method of Barbosa and Hughes : $(7.8)_2$ gives

$$\lambda_h = -\Pi_{\Lambda_h}(\frac{\partial u_h}{\partial n}) + \frac{1}{\alpha h}(u_h - d).$$

Thus $(7.8)_1$ becomes

$$\begin{aligned} (u_h, v_h)_{H^1(\Omega)} - \int_\Gamma \Pi_{\Lambda_h}(\frac{\partial u_h}{\partial n}) v_h \, d\Gamma - \int_\Gamma \Pi_{\Lambda_h}(\frac{\partial v_h}{\partial n}) u_h \, d\Gamma + \frac{1}{\alpha h} \int_\Gamma u_h v_h \, d\Gamma \\ - \alpha h (\int_\Gamma \frac{\partial u_h}{\partial n} \frac{\partial v_h}{\partial n} - \Pi_{\Lambda_h} \frac{\partial u_h}{\partial n} \Pi_{\Lambda_h} \frac{\partial v_h}{\partial n} \, d\Gamma) \\ = (f, v_h)_{L^2(\Omega)} - \int_\Gamma g \, \Pi_{\Lambda_h} \frac{\partial v_h}{\partial n} \, d\Gamma + \frac{1}{\alpha h} \int_\Gamma g v_h \, d\Gamma, \end{aligned}$$

With $\Lambda_h = P_0$ and $V_h = P_1$ we then get $\Pi_{\Lambda_h}(\frac{\partial u_h}{\partial n}) = \frac{\partial u_h}{\partial n}$, and then (7.13).

Part III
Theory

Most of the results can be found in Brézis [6].

8 The open mapping theorem

8.1 Notations

If E and F are linear spaces, a map $T : E \to F$ is linear iff $T(x_1 + \lambda x_2) = T(x_1) + \lambda T(x_2)$ for all $x_1, x_2 \in E$ and all $\lambda \in \mathbb{R}$; And then $T(x)$ is denoted $T.x$ or Tx.

Let $(E, ||.||_E)$ be a normed space. Let $B_E(x, \rho) = \{x' \in E; ||x' - x||_E < \rho\}$ the ball of radius $\rho > 0$ centered at $x \in E$.

If $(E, ||.||_E)$ and $(F, ||.||_F)$ are two normed spaces, if $T : \left\{ \begin{array}{ll} E & \to F \\ x & \to T(x) \end{array} \right\}$ is a linear map, then T is said to be continuous (or bounded) iff

$$\exists c > 0, \quad \forall x \in E, \quad ||T.x||_F \le c\,||x||_E. \quad \text{and then} \quad ||T|| := \sup_{x \in B_E(0,1)} ||T.x||_F. \tag{8.1}$$

In the sequel, the space E and F will be Banach spaces (complete for the norm in use).

Let $\mathcal{L}(E;F)$ be the set of linear continuous mapping from E to F. Then

$$||.||_{\mathcal{L}(E;F)} : \left\{ \begin{array}{ll} \mathcal{L}(E;F) & \to \mathbb{R} \\ T & \to ||T||_{\mathcal{L}(E;F)} := \sup\limits_{x \in B_E(0,1)} ||T.x||_F \overset{\text{denoted}}{=} ||T|| \end{array} \right. \tag{8.2}$$

define a norm in $\mathcal{L}(E;F)$ (easy check), and $(\mathcal{L}(E;F), ||.||)$ is a Banach space (check: If $(T_n)_{\mathbb{N}^*}$ is a Cauchy sequence, that is $||T_n - T_m|| \longrightarrow_{n,m\to\infty} 0$, then, for any $x \in E$, $||(T_n - T_m)(x)||_F \longrightarrow_{n,m\to\infty} 0$, thus $(T_n(x))_{\mathbb{N}^*}$ is a Cauchy sequence in F complete, thus converge to a $y_x \in F$; Then define $T : x \in E \to T(x) = y_x$: It is easy to check that T is linear and continuous with $||T - T_n|| \longrightarrow_{n\to\infty} 0$.)

Let $E' := \mathcal{L}(E;\mathbb{R})$, called the dual of E (the set of linear continuous real valued functions, $\mathbb{R}$ being provided with its usual norm). For $\ell \in E'$ and $x \in E$ denote:

$$\ell(x) = \ell.x = \langle \ell, x \rangle_{E',E} \quad \in \mathbb{R}. \tag{8.3}$$

So, cf. (8.1),

$$||\ell||_{E'} = \sup_{x \in B_E(0,1)} |\ell.x| = \sup_{x \in B_E(0,1)} |\langle \ell, x \rangle_{E',E}| \tag{8.4}$$

defines a norm in E' s.t. $(E', ||.||_{E'})$ is a Banach space.

If $T \in \mathcal{L}(E;F)$ (linear and continuous) then its adjoint is the linear map $T' : F' \to E'$ characterized by:

$$T' : \left\{ \begin{array}{ll} F' & \to E' \\ \ell & \to T'(\ell) \overset{\text{denoted}}{=} T'.\ell, \quad \text{where} \quad \langle T'.\ell, x \rangle_{E',E} := \langle \ell, T.x \rangle_{F',F}, \quad \forall x \in E. \end{array} \right. \tag{8.5}$$

Proposition 8.1 *T' is continuous with*

$$||T'|| = ||T||. \tag{8.6}$$

Thus $T' \in \mathcal{L}(F', E')$.

Proof. $||T'.\ell||_{E'} = \sup_{||x||_E \le 1} |\langle T'.\ell, x \rangle_{E',E}| = \sup_{||x||_E \le 1} |\langle \ell, T.x \rangle_{F',F}| \le \sup_{||x||_E \le 1} ||\ell||_{F'} ||T.x||_F = \sup_{||x||_E \le 1} ||T||\,||x||_E ||\ell||_{F'} = ||T||\,||\ell||_{F'}$, thus $||T'|| \le ||T||$; And similarly $||T.x||_F \le ||T'||\,||\ell||_{F'}$, thus $||T|| \le ||T'||$. ∎

$E'' = (E')' = \mathcal{L}(E';\mathbb{R})$ is a Banach space (since $\mathbb{R}$ is complete). Let

$$J : \left\{ \begin{array}{ll} E & \to E'' = \mathcal{L}(E';\mathbb{R}) \\ x & \to J(x), \quad \text{where} \quad J(x)(\ell) := \ell.x, \ \forall x \in E. \end{array} \right. \tag{8.7}$$

J is linear (trivial), is continuous, with $||J|| = \sup_{x\in B_E(0,1)} |J(x)| = \sup_{x\in B_E(0,1)}(\sup_{\ell\in B_{E'}(0,1)} |J(x)(\ell)|) = \sup_{x\in B_E(0,1)}(\sup_{\ell\in B_{E'}(0,1)} |\ell.x|) = \sup_{x\in B_E(0,1)}(||x||_E) = 1$, and injective (one-to-one) since $J(x) = 0$ implies $\ell.x = 0$ for all $\ell \in E'$ that implies $x = 0$. Thus J is a "canonical injection".

Thus $J(E) = \text{Im}(E)$, the range or image of E by J, can be identified to a subspace of E''.

Definition 8.2 A Banach space E is reflexive iff J is bijective (= one-to-one and onto), and then is identified with E, denoted $E'' \simeq E$, and $J(x)$ is denoted x.

(Remark: A Hilbert space is always reflexive, and a reflexive Banach space "almost" behaves like a Hilbert space for computation purposes (with the use of the bracket $\langle ., .\rangle_{E',E}$ similar to the use of a inner product). There are however some substantial differences: e.g. in a reflexive Banach space there exist closed subspaces without any complement, whereas in a Hilbert space any closed subspace has a complement (even an orthogonal one); And this causes some theoretical difficulties treated in the sequel.)

8.2 The open mapping theorem

Theorem 8.3 (Open mapping theorem) *Let E and F be Banach spaces. If $T \in \mathcal{L}(E; F)$ (linear and continuous) is surjective (= onto, i.e. $\text{Im}(T) = F$), then*

$$\exists\gamma > 0 \quad s.t. \quad T(B_E(0, 1)) \supset B_F(0, \gamma). \tag{8.8}$$

That is, if T is linear continuous and surjective, then any open set in E is transformed by T into an open set in F. So $T(B_E(0, 1))$ is not "flat" (it contains an open set).

And the converse is true: if (8.8) then T is surjective.

Proof. See Brézis [6]. Steps : 1- T being onto, we have $\bigcup_{n\in\mathbb{N}^*} \overline{T(B_E(0, n))} = F$, and Baire's Theorem gives the existence of a closed space $\overline{T(B_E(0, n))}$ containing an open set; 2- The linearity of T then implies that $\overline{T(B_E(0, 1))}$ contains an open set $B_F(0, 2\gamma)$ for some $\gamma > 0$. 3- And T being continuous and E being complete we get $T(B_E(0, 1)) \supset B_F(0, \gamma)$.

Converse: $T(B_E(0, 1)) \supset B_F(0, \gamma)$, and T is linear, so $T(E) = F$. ∎

Corollary 8.4 *If $T \in \mathcal{L}(E; F)$ is bijective, i.e. injective (= one-to-one) and surjective (= onto), then the linear map $T^{-1} : F \to E$ is continuous, that is,*

$$\exists\gamma > 0,\ \forall y \in F, \quad ||T^{-1}.y||_E \le \frac{1}{\gamma}||y||_F. \tag{8.9}$$

Thus

$$\exists\gamma > 0,\ \forall x \in E, \quad ||T.x||_F \ge \gamma||x||_E. \tag{8.10}$$

Proof. Then T bijective gives $T^{-1}(B_F(0, \gamma)) \subset B_E(0, 1)$. And T^{-1} is linear since T is, thus $T^{-1}(B_F(0, 1)) \subset B_E(0, \frac{1}{\gamma})$. So $y \in B_F(0, 1)$ gives $||T^{-1}.y||_E \le \frac{1}{\gamma}||y||_F$, i.e. (8.9). Then $y = T.x$ gives (8.10) (bijectivity). ∎

Remark 8.5 If T is bijective between Banach spaces, then the problem: For $b \in F$ find $x \in E$ s.t. $T.x = b$ is well-posed, that is, has a unique solution $x = T^{-1}.b$ s.t. $\exists c > 0$ (independent of b), $||x||_E \le c||b||_F$ (the inverse T^{-1} is continuous). Indeed, the bijectivity of T gives a unique solution $x = T^{-1}.b$, and (8.9) gives $||x||_E = ||T^{-1}.b||_E \le \frac{1}{\gamma}||b||_F$. ∎

Remark 8.6 A linear continuous bijective mapping between two infinite dimensional Banach spaces behaves like in finite dimension, e.g. like $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by its matrix $\begin{pmatrix} -2 & 0 \\ 0 & 3 \end{pmatrix}$. Here $||T|| = 3$, $T^{-1} = \begin{pmatrix} -\frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}$, and $||T^{-1}|| = \frac{1}{2} = \frac{1}{\gamma}$. ∎

Remark 8.7 The bijectivity between Banach spaces (complete spaces) is required:

Let $\ell^2 = \{(x_n)_{\mathbb{N}^*} \in \mathbb{R}^{\mathbb{N}^*} : \sum_{n\in\mathbb{N}^*} x_n^2 < \infty\}$ (the space of finite energy sequences), let $E = F = \ell^2$, and let $T : \ell^2 \to \ell^2$ be given by $T((x_n)_{\mathbb{N}^*}) = (\frac{x_n}{n})_{\mathbb{N}^*}$ for any $(x_n) \in \ell^2$, that is, with $(e_n)_{\mathbb{N}^*}$ the canonical basis in ℓ^2, $T.e_n = \frac{1}{n}e_n$ (the associated generalized matrix is the infinite diagonal matrix $\text{diag}(1, \frac{1}{2}, ..., \frac{1}{n}, ...)$.) Here T is injective since $\text{Ker}(T) = \{0\}$ (trivial), but not surjective since $(\frac{1}{n})_{\mathbb{N}^*} \in \ell^2$

has no counterimage in ℓ^2 (it would be the constant sequence $(1)_{\mathbb{N}^*} \notin \ell^2$). And its range $\mathrm{Im}(T)$ is dense in ℓ^2: Indeed, if $(y_n)_{\mathbb{N}^*} \in \ell^2$ then let $x_n = ny_n$, so that $(x_n)_{\mathbb{N}^*} \in \mathbb{R}^{\mathbb{N}^*}$, and for $N \in \mathbb{N}^*$, define the truncated sequence $(x_n^N)_{\mathbb{N}^*}$ by $x_n^N = x_n$ if $n \le N$ and $x_n^N = 0$ otherwise; then $(x_n^N) \in \ell^2$ (trivial) and $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}^*$, $||y_n - Tx_n^N||_{\ell^2}^2 = \sum_{n=N+1}^\infty y_n^2 < \varepsilon$). Here $\mathrm{Im}(T)$ is not closed in ℓ^2 (since dense and closed would imply $\mathrm{Im}(T) = \ell^2$), and $\mathrm{Im}(T)$ is "flat" in ℓ^2, that is, $T(B_{\ell^2}(0,1))$ does not contain any open ball: In this example it can be seen with the canonical basis $(e_n)_{\mathbb{N}^*}$ that verifies $T^{-1}.e_n = ne_n$, so that $T^{-1}(B_{\ell^2}(0,1))$ is not bounded (if one prefers, $T^{-1}(\frac{1}{2}\gamma e_n) = n\frac{1}{2}\gamma e_n \notin T(B_{\ell^2}(0,1))$ as soon as $n > \frac{2}{\gamma}$ although $\frac{1}{2}\gamma e_n \in B_{\ell^2}(0,\gamma)$). ∎

Corollary 8.8 *If $T \in \mathcal{L}(E;F)$ (linear and continuous) is injective (=one-to-one), and if $\mathrm{Im}(T)$ is closed in F, then*

$$\exists \gamma > 0,\ \forall x \in E, \quad ||T.x||_F \ge \gamma ||x||_E. \tag{8.11}$$

Proof. For $y \in \mathrm{Im}(T)$ denote $||y||_{\mathrm{Im}(T)} := ||y||_F$ for all $y \in \mathrm{Im}(T)$. So $||.||_{\mathrm{Im}(T)}$ is a norm in $\mathrm{Im}(T)$ (trivial). Then $\mathrm{Im}(T)$ closed in F implies $(\mathrm{Im}(T), ||.||_{\mathrm{Im}(T)}) = (\mathrm{Im}(T), ||.||_F)$ is a Banach space denoted $\mathrm{Im}(T)$. Let

$$T_R : \begin{cases} E & \to \mathrm{Im}(T) \\ x & \to T_R(x) = T(x). \end{cases} \tag{8.12}$$

Then T_R is linear continuous bijective between Banach spaces. Thus (8.10) gives $\exists \gamma > 0$, $\forall x \in E$, $||T_R.x||_{\mathrm{Im}(T)} \ge \gamma ||x||_E$, i.e. (8.11). ∎

8.3 Quotient space $E/\mathrm{Ker}(T)$, and open mapping theorem

Let E and F be banach spaces. Let $T \in \mathcal{L}(E;F)$ (linear and continuous). Then $K = \mathrm{Ker}(T) = T^{-1}(\{0\})$ (the kernel of T) is linear subspace that is closed (since T is continuous).

Consider the relation in E defined by: $x \sim y$ iff $x - y \in K$. This is an equivalence relation (easy check). Let $E/K = \{Z \subset E : \exists x \in E,\ Z = x + K\} = \{x + K : x \in E\}$ be the set of the equivalence classes (quotient space). An element of E/K is denoted $\dot{x} = x + K$. In particular $\dot{0} = K$.

The (usual) operators $+$ in E/K and $.$ on E/K are defined by, if $\dot{x} = x + K$, $\dot{y} = y + K$ and $\lambda \in \mathbb{R}$,

$$\dot{x} + \dot{y} = x + y + K, \quad \text{and} \quad \lambda.\dot{x} = \lambda x + K, \tag{8.13}$$

definition independent of the $x' \in \dot{x}$ and $y' \in \dot{y}$ (easy check). Then $(E/K, +, .)$ is a linear space (easy check) with $\dot{0}$ the zero in E/K.

Lemma 8.9 *The canonical map* $\pi : \left\{ \begin{array}{ll} E & \to E/K \\ x & \to \pi(x) := x + K = \dot{x} \end{array} \right\}$ *is linear and surjective.*

Proof. Linearity: $\pi(x + \lambda y) = (x + \lambda y) + K = x + K + \lambda y + \lambda K = \pi(x) + \lambda \pi(y)$ since K is a linear space (so $K = K + \lambda K$).

Surjectivity: If $\dot{x} \in E/K$ then $\exists x \in E$ s.t. $\dot{x} = x + K = \pi(x)$ (definition of E/K). ∎

For $\dot{x} \in E/K$, define $||.||_{E/K} : E/K \to \mathbb{R}$ by

$$||\dot{x}||_{E/K} = ||\pi(x)||_{E/K} := \inf_{x_0 \in K} ||x + x_0||_E \overset{\text{denoted}}{=} ||x||_{E/K}. \tag{8.14}$$

Lemma 8.10 *$||.||_{E/K}$ is a norm in E/K, and $(E/K, ||.||_{E/K})$ is a Banach space.*
And π is continuous with $||\pi|| \le 1$.

Proof. $||\dot{x}||_{E/K} = 0 \Leftrightarrow \inf_{x_0 \in K} ||x + x_0||_E = 0 \Leftrightarrow ||x||_E \le 0$ since $0 \in K \Leftrightarrow x = 0 \Leftrightarrow \pi(x) = 0$ (since π is linear) $\Leftrightarrow \dot{x} = K = \dot{0}$.

$||\lambda \dot{x}||_{E/K} = \inf_{x_0 \in K} ||\lambda x + x_0||_E = \inf_{x_0 \in K} ||\lambda x + \lambda x_0||_E = \inf_{x_0 \in K} |\lambda|\,||x + x_0||_E |\lambda|\,||\dot{x}||_{E/K}$.

$||\dot{x} + \dot{y}||_{E/K} = \inf_{x_0, y_0 \in K} ||x + y + x_0 + y_0||_E \le \inf_{x_0, y_0 \in K} ||x + x_0||_E + ||y + y_0||_E \le ||\dot{x}||_{E/K} + ||\dot{x} + \dot{y}||_{E/K}$.

Thus $||.||_{E/K}$ is a norm in E/K.

Let $(\dot{x}_n)_{\mathbb{N}^*}$ be a Cauchy sequence in E/K, that is, $||\pi(x_n - x_m)||_{E/K} = ||\dot{x}_n - \dot{x}_m||_{E/K} \longrightarrow_{n,m \to \infty} 0$. Let a subsequence, still denoted (dx_n) s.t. $||\pi(x_{k+1} - x_k)||_{E/K} < \frac{1}{2^k}$ for all $k \in \mathbb{N}^*$. Thus $\exists (y_k)_{\mathbb{N}^*} \in K$ s.t.

$||x_{k+1}-x_k-y_k||_E < \frac{2}{2^k}$ for all $k\in\mathbb{N}^*$, see (8.14). Then let $(z_k)_{\mathbb{N}^*}$ be defined by $z_1=0$ ad $y_k=z_{k+1}-z_k$. Thus $||x_{k+1}-z_{k+1}-(x_k-y_k)||_E < \frac{2}{2^k}$ for all $k\in\mathbb{N}^*$, thus $||x_{n+1}-z_{n+1}-(x_m-y_m)||_E \longrightarrow_{n,m\to\infty} 0$, thus $((x_n-z_n)_{\mathbb{N}^*}$ is a Cauchy sequence in E, thus converges to a limit $w\in E$. Thus $\pi(x_n-z_n)=\pi(x_n)-0$ converges to $\pi(w)\in E/K$, and E/K is closed.

$||\pi(x)||_{E/K}=\min_{x_0\in K}||x+x_0||_E$, and $0\in K$ (linear subspace), thus $||\pi(x)||_{E/K}\le||x||_E$. ∎

Let

$$\widetilde{T}:\begin{cases} E/K \ \to F \\ \quad \dot{x} \ \to \widetilde{T}(\dot{x}):=T(x) \quad \text{when} \quad x\in\dot{x}, \end{cases} \tag{8.15}$$

definition independent of $x\in\dot{x}$ since $T(x+x_0)=T(x)$ for all $x_0\in K$ $(=\mathrm{Ker}(T))$. In other words, $\widetilde{T}$ is characterized by $\widetilde{T}\circ\pi=T$.

Lemma 8.11 *$\widetilde{T}$ is linear, injective and continuous with $||\widetilde{T}||=||T||$.*

Proof. With $\dot{x}=x+K$ and $\dot{y}=y+K$ we get $\dot{y}+\lambda\dot{x}=x+\lambda y+K$ since K is a linear space, thus $\widetilde{T}(\dot{x}+\lambda\dot{y})=T(x+\lambda y)=T(x)+\lambda T(y)=\widetilde{T}(\dot{x})+\lambda\widetilde{T}(\dot{y})$, and $\widetilde{T}$ is linear.

$\widetilde{T}.\dot{x}=0 \Rightarrow T(x+x_0)=0$ for all $x_0\in K \Rightarrow x+x_0\in K \Rightarrow x\in K\Rightarrow \dot{x}=\dot{0}$, thus $\widetilde{T}$ is injective.

Let $\dot{x}\in E/K$. We have $||\widetilde{T}(\dot{x})||_F=||T.(x+x_0)||_F\le||T||\,||x+x_0||_E$ for all $x_0\in K$, thus $||\widetilde{T}(\dot{x})||_F\le ||T||\,\min_{x_0\in K}||x+x_0||_E=||T||\,||\dot{x}||_{E/K}$. Thus $\widetilde{T}$ is continuous, with $||\widetilde{T}||\le||T||$.

Let $x\in E$. We have $||T.x||_F=||\widetilde{T}.\dot{x}||_F\le||\widetilde{T}||\,||\dot{x}||_{E/K}$, thus $||T.x||_F\le||\widetilde{T}||\,||x+x_0||_E$ for all $x_0\in K$, with $T.x=T(x+x_0)$ for all $x_0\in K$, thus $||T(x+x_0)||_F\le||\widetilde{T}||\,||x+x_0||_E$ for all $x_0\in K$, $||T.x||_F\le||\widetilde{T}||\,||x||_E$. Thus $||T||\le||\widetilde{T}||$. ∎

Corollary 8.12 *Let E and F be banach spaces. If $T\in\mathcal{L}(E;F)$ (linear and continuous), and if* $\mathrm{Im}(T)$ *is closed in F, then*

$$\exists\gamma>0,\ \forall x\in E,\quad ||T.x||_F\ge\gamma||x||_{E/\mathrm{Ker}(T)}\quad (=\gamma\inf_{x_0\in\mathrm{Ker}(T)}||x+x_0||_E). \tag{8.16}$$

Proof. $K:=\mathrm{Ker}(T)=T^{-1}(\{0\})$ is closed since T is continuous.

$\mathrm{Im}(T)$ is closed in F, therefore $(\mathrm{Im}(T),||.||_F)$ is a Banach space denoted $\mathrm{Im}(T)$. Then $\widetilde{T}_R:\dot{x}\in E/K\to\widetilde{T}_R(\dot{x})=T(x)\in\mathrm{Im}(T)$ is linear continuous and bijective between Banach spaces. Thus (8.10) gives $\exists\gamma>0$, $\forall\dot{x}\in E/K$, $||\widetilde{T}_R.\dot{x}||_F\ge\gamma||\dot{x}||_{E/K}$, i.e. (8.16). ∎

8.4 The inf-sup condition

(8.16) is rewritten

$$\exists\gamma>0,\ \inf_{x\in E}\frac{||T.x||_F}{||x||_{E/\mathrm{Ker}(T)}}\ge\gamma. \tag{8.17}$$

(Light writing of $\exists\gamma>0,\ \inf_{x\in E-\{0\}}\frac{||T.x||_F}{||x||_{E/\mathrm{Ker}(T)}}\ge\gamma$.)

Consider $B\in\mathcal{L}(E;F')$ (linear and continuous). Then (8.17) gives

$$\exists\gamma>0,\ \inf_{x\in E}\frac{||B.x||_{F'}}{||x||_{E/\mathrm{Ker}(T)}}\ge\gamma, \tag{8.18}$$

Let $b(\cdot,\cdot):E\times F\to\mathbb{R}$ be the bilinear form defined by, for all $(x,y)\in E\times F$,

$$b(x,y)=\langle B.x,y\rangle_{F',F}. \tag{8.19}$$

Since $||B.x||_{F'}=\sup_{y\in F}\frac{|\langle B.x,y\rangle_{F',F}|}{||y||_F}$, (8.18) gives

$$\exists\gamma>0,\ \inf_{x\in E}(\sup_{y\in F}\frac{b(x,y)}{||x||_{E/\mathrm{Ker}(T)}||y||_F})\ge\gamma, \tag{8.20}$$

named the inf-sup condition satisfied by $b(\cdot,\cdot)$

9 Some spaces and their duals

9.1 Divergence, Gradient, Rotationnal

Let $(\vec{e}_i)$ be a Euclidean basis in $\mathbb{R}^n$, let $(\cdot,\cdot)_{\mathbb{R}^n}$ be the associated inner product, and let $\vec{v}.\vec{w} := (\vec{v},\vec{w})_{\mathbb{R}^n}$. Let Ω be an open set in $\mathbb{R}^n$. Let $(\vec{e}_i)$ be a basis in $\mathbb{R}^n$ and $\frac{\partial f}{\partial x^i} := df.\vec{e}_i$.

The divergence operator is formally given by

$$\text{div} : \begin{cases} \mathcal{F}(\Omega;\mathbb{R}^n) & \to \mathbb{R} \\ \vec{v} = \sum_{i=1}^n v^i \vec{e}_i & \to \text{div}\vec{v} = \sum_{i=1}^n \frac{\partial v^i}{\partial x^i}. \end{cases} \tag{9.1}$$

(The real value $\text{div}\vec{v}$ does not depend on the choice of the basis).

The gradient operator is formally given by

$$\vec{\text{grad}} : \begin{cases} \mathcal{F}(\Omega;\mathbb{R}) & \to \mathbb{R}^n \\ f & \to \vec{\text{grad}} f = \sum_{i=1}^n \frac{\partial f}{\partial x^i}\vec{e}_i. \end{cases} \tag{9.2}$$

The rotationnal operator is formally given by

$$\vec{\text{curl}} : \begin{cases} \mathcal{F}(\Omega;\mathbb{R}^3) & \to \mathbb{R}^3 \\ \vec{v} = \sum_{i=1}^n v^i \vec{e}_i & \to \vec{\text{curl}}\vec{v} = (\frac{\partial v_3}{\partial x_2} - \frac{\partial v_2}{\partial x_3})\vec{e}_1 + (\frac{\partial v_1}{\partial x_3} - \frac{\partial v_3}{\partial x_1})\vec{e}_2 + (\frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2})\vec{e}_3. \end{cases} \tag{9.3}$$

(In $\mathbb{R}^2$, $\text{curl} : \vec{v} \in \mathbb{R}^2 \to \text{curl}\vec{v} = \frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2} \in \mathbb{R}$.)

9.2 Some Hilbert spaces

Let Ω be an open set in $\mathbb{R}^n$, $n = 1,2,3$, and $\Gamma = \partial\Omega$ be its boundary.

$$L^2(\Omega) = \{f : \Omega \to \mathbb{R} : \int_\Omega f^2\,d\Omega < \infty\}, \qquad (f,g)_{L^2} = \int_\Omega fg\,d\Omega.$$

$$H^1(\Omega) = \{f \in L^2(\Omega) : \vec{\text{grad}} f \in L^2(\Omega)^n\}, \qquad (f,g)_{H^1} = (f,g)_{L^2} + (\vec{\text{grad}} f, \vec{\text{grad}} g)_{L^2}.$$

$$H^2(\Omega) = \{f \in H^1(\Omega) : d^2 f \in L^2(\Omega)^{n^2}\}, \qquad (f,g)_{H^2} = (f,g)_{H^1} + (d^2 f, d^2 g)_{L^2}.$$

$$H^{\text{div}}(\Omega) = \{\vec{v} \in L^2(\Omega)^n : \text{div}\vec{v} \in L^2(\Omega)\}, \qquad (\vec{u},\vec{v})_{H^{\text{div}}} = (\vec{u},\vec{v})_{L^2} + (\text{div}\vec{u}, \text{div}\vec{v})_{L^2}.$$

$$H^{\text{curl}}(\Omega) = \{\vec{v} \in L^2(\Omega)^3 : \vec{\text{curl}}\vec{v} \in L^2(\Omega)^3\}, \qquad (\vec{u},\vec{v})_{H^{\text{curl}}} = (\vec{u},\vec{v})_{L^2} + (\vec{\text{curl}}\vec{u}, \vec{\text{curl}}\vec{v})_{L^2}.$$

Integrations by parts : If $f \in H^1(\Omega)$ and $\vec{v} \in H^{\text{div}}(\Omega)$ then

$$\int_\Omega \vec{\text{grad}} f.\vec{v}\,d\Omega = -\int_\Omega f\,\text{div}\vec{v}\,d\Omega + \int_\Gamma f\,\vec{v}.\vec{n}\,d\Gamma. \tag{9.4}$$

If $\vec{v} \in H^1(\Omega)^n$ and $\vec{w} \in H^{\text{curl}}(\Omega)$ then

$$\int_\Omega \vec{\text{curl}}\vec{v}.\vec{w}\,d\Omega = +\int_\Omega \vec{v}.\vec{\text{curl}}\vec{w}\,d\Omega + \int_\Gamma \vec{v}.(\vec{w} \wedge \vec{n})\,d\Gamma. \tag{9.5}$$

9.3 Some sup-spaces

Closures $\mathcal{D}(\Omega) = C_c^\infty(\Omega)$ (space of C^∞ functions with compact support):

$$H_0^1(\Omega) = \{f \in H^1(\Omega) : f_{|\Gamma} = 0\} = \overline{\mathcal{D}(\Omega)}^{H^1}, \qquad (f,g)_{H_0^1} = (\vec{\text{grad}} f, \vec{\text{grad}} g)_{L^2}.$$

$$H_0^2(\Omega) = \{f \in H^2(\Omega) : f_{|\Gamma} = 0 \text{ et } \vec{\text{grad}} f.\vec{n}_{|\Gamma} = 0\} = \overline{\mathcal{D}(\Omega)}^{H^2}, \qquad (f,g)_{H_0^2} = (d^2 f, d^2 g)_{L^2}.$$

$$H_0^{\text{div}}(\Omega) = \{\vec{v} \in H^{\text{div}}(\Omega) : (\vec{v}.\vec{n})_{|\Gamma} = 0\} = \overline{\mathcal{D}(\Omega)^n}^{H^{\text{div}}}, \qquad (\vec{u},\vec{v})_{H_0^{\text{div}}} = (\text{div}\vec{u}, \text{div}\vec{v})_{L^2}.$$

$$H_0^{\text{curl}}(\Omega) = \{\vec{v} \in H^{\text{curl}}(\Omega) : (\vec{v} \wedge \vec{n})_{|\Gamma} = 0\} = \overline{\mathcal{D}(\Omega)^3}^{H^{\text{curl}}}, \qquad (\vec{u},\vec{v})_{H_0^{\text{curl}}} = (\vec{\text{curl}}\vec{u}, \vec{\text{curl}}\vec{v})_{L^2}.$$

When Ω is bounded, the given semi-inner products are equivalent to the inner products of the embedding spaces.

9.4 Trace operator γ_0 and the Hilbert space $H^{\frac{1}{2}}(\Gamma)$

The trace mapping is

$$\gamma_0 : \begin{cases} H^1(\Omega) & \to L^2(\Gamma), \\ \quad f & \to \gamma_0(f) = f_{|\Gamma}. \end{cases} \tag{9.6}$$

(Same notation ifor $\gamma_0 : \vec{v} \in H^1(\Omega)^n \to \gamma_0(\vec{v}) = \vec{v}_{|\Gamma} \in L^2(\Gamma)^n$.) If Ω is regular, then

$$H_0^1(\Omega) = \text{Ker}(\gamma_0). \tag{9.7}$$

With $\text{Im}(T)$ the range of a mapping T, let

$$\text{Im}(\gamma_0) = H^{\frac{1}{2}}(\Gamma), \quad \text{and} \quad ||d||_{H^{\frac{1}{2}}(\Gamma)} = \inf_{u \in H^1(\Omega): u_{|\Gamma} = d} ||u||_{H^1}. \tag{9.8}$$

Proposition 9.1 *Let $d \in H^{\frac{1}{2}}(\Gamma)$ and let $u_d \in H^1(\Omega)$ be the solution of the Dirichlet problem: Find $u \in H^1(\Omega)$ s.t.*

$$\begin{cases} -\Delta u + u = 0 \quad \text{in } \Omega, \\ u_{|\Gamma} = d \quad \text{on } \Gamma, \end{cases} \tag{9.9}$$

then

$$||d||_{H^{\frac{1}{2}}(\Gamma)} = ||u_d||_{H^1}, \tag{9.10}$$

and $||.||_{H^{\frac{1}{2}}(\Gamma)}$ is the norm of the inner product

$$(c, d)_{H^{\frac{1}{2}}(\Gamma)} := (u_c, u_d)_{H^1(\Omega)}. \tag{9.11}$$

Moreover $(H^{\frac{1}{2}}(\Gamma), (\cdot, \cdot)_{H^{\frac{1}{2}}(\Gamma)})$ is a Hilbert space.

Moreover $\gamma_0 : v \in (H^1(\Omega), ||.||_{H^1(\Omega)}) \to v_{|\gamma} \in (H^{\frac{1}{2}}(\Gamma), ||.||_{H^{\frac{1}{2}}(\Gamma)})$ is (linear) continuous.

Proof. Let $z_d \in H^1(\Omega)$ be an counter image of $d \in H^{\frac{1}{2}}(\Gamma)$ (exists by definition of $H^{\frac{1}{2}}(\Gamma)$, cf. (9.8)). So $\gamma_0(z_d) = d = z_{d|\Gamma}$. Let $u = u_0 + z_d \in H_0^1(\Omega) + z_d$. With (9.9) we get

$$\left\{ \begin{array}{l} -\Delta u_0 + u_0 = \Delta z_d - z_d \quad \text{dans } H^{-1}(\Omega), \\ u_{0|\Gamma} = 0 \quad \text{dans } H^{\frac{1}{2}}(\Gamma). \end{array} \right\} \tag{9.12}$$

Thus $u_0 \in H_0^1(\Omega)$ satisfies

$$(u_0, v_0)_{H^1(\Omega)} = -(z_d, v_0)_{H^1(\Omega)}, \qquad \forall v_0 \in H_0^1(\Omega), \tag{9.13}$$

and the Lax–Milgram theorem gives the existence of a solution $u_0 \in H_0^1(\Omega)$. Then we check that $u_d = u_0 + z_d$ is independent of the chosen z_d: If z'_d satisfies $\gamma(z'_d) = d$, if the associate solution is u'_0, if $u'_d = u'_0 + z'_d$, then $u_d - u'_d \in H_0^1(\Omega)$ and $(u_d - u'_d, v_0)_{H^1(\Omega)} = 0$ for any $v_0 \in H_0^1(\Omega)$, so $u_d - u'_d = 0$ (Lax–Milgram theorem). Moreover (9.13) tells that $u_0 + z_d = u_d \perp_{H^1} H_0^1(\Omega)$. Thus we get, for any $v_0 \in H_0^1(\Omega)$,

$$||u_d + v_0||^2_{H^1(\Omega)} = ||u_d||^2_{H^1(\Omega)} + ||v_0||^2_{H^1(\Omega)} + 2(u_d, v_0)_{H^1(\Omega)} = ||u_d||^2_{H^1(\Omega)} + ||v_0||^2_{H^1(\Omega)} + 0.$$

So $\inf_{\substack{w \in H^1(\Omega) \\ w_\Gamma = d}} ||w||^2_{H^1(\Omega)} = \inf_{v_0 \in H_0^1(\Omega)} ||u_d + v_0||^2_{H^1(\Omega)} = ||u_d||^2_{H^1(\Omega)}$, denoted $||d||_{H^{\frac{1}{2}}(\Gamma)}$. Then we define $(c, d)_{H^{\frac{1}{2}}(\Gamma)}$ as in (9.11), and $(\cdot, \cdot)_{H^{\frac{1}{2}}}$ is trivially bilinear symmetric positive (is an inner product).

Then we check that $(H^{\frac{1}{2}}(\Gamma), ||.||_{H^{\frac{1}{2}}})$ is complete: If $(d_n)_{\mathbb{N}^*} \in H^{\frac{1}{2}}(\Gamma)$ is a Cauchy sequel in $H^{\frac{1}{2}}(\Gamma)$ and if u_{d_n} is the solution of (9.9), then $(u_{d_n})_{\mathbb{N}^*}$ is a Cauchy sequel in $(H^1(\Omega), ||.||_{H^1})$, cf. (9.10), so converges in $H^1(\Omega)$ (since $H^1(\Omega)$ is complete) toward some $u \in H^1(\Omega)$. Then let $d := \gamma_0(u) \in H^{\frac{1}{2}}(\Gamma)$. Since $(u_{d_n}, v_0)_{H^1(\Omega)} = 0$ for any $v_0 \in H_0^1(\Omega)$, cf. (9.9), we get $(u_d, v_0)_{H^1(\Omega)} = 0$ for any $v_0 \in H_0^1(\Omega)$ (continuity of an inner product relatively to itself). Thus u_d is the solution of (9.9), and $||d - d_n||_{H^{\frac{1}{2}}} = ||u - u_{d_n}||_{H^1} \longrightarrow_{n \to \infty} 0$.

And $\gamma_0 : (H^1(\Omega), ||.||_{H^1}) \to (H^{\frac{1}{2}}(\Gamma), ||.||_{H^{\frac{1}{2}}(\Gamma)})$ (linear) satisfies, for $u \in H^1(\Omega)$, with $d := \gamma_0(u)$ and u_d solution of (9.9), $||\gamma_0(u)||_{H^{\frac{1}{2}}(\Gamma)} = ||u_d||_{H^1(\Omega)} \leq ||u_d + u_0||_{H^1(\Omega)}$ for any $u_0 \in H_0^1(\Omega)$, thus with $u_0 = u - u_d$ we get $||\gamma_0(u)||_{H^{\frac{1}{2}}(\Gamma)} = ||u||_{H^1(\Omega)}$, so γ_0 is bounded ($||\gamma_0|| \leq 1$). ∎

9.5 Some other trace operators

$$\gamma_1 : \begin{cases} H^2(\Omega) & \to L^2(\Gamma), \\ \quad f & \to \gamma_1(f) = (\vec{\text{grad}} f)_{|\Gamma}.\vec{n} \overset{\text{denoted}}{=} \dfrac{\partial f}{\partial \vec{n}}_{|\Gamma}. \end{cases} \tag{9.14}$$

$$\gamma_n : \begin{cases} H^{\text{div}}(\Omega) & \to H^{-\frac{1}{2}}(\Gamma), \\ \quad \vec{v} & \to \gamma_n(\vec{v}) = \gamma_0(\vec{v}).\vec{n} \overset{\text{denoted}}{=} (\vec{v}.\vec{n})_{|\Gamma} \end{cases} \tag{9.15}$$

(and the divergence operator enables the control of $\vec{v}.\vec{n}$ on Γ),

$$\vec{\gamma}_t : \begin{cases} H^{\text{curl}}(\Omega) & \to H^{-\frac{1}{2}}(\Gamma)^3, \\ \quad \vec{v} & \to \vec{\gamma}_t(\vec{v}) = \gamma_0(\vec{v}) \wedge \vec{n} \overset{\text{denoted}}{=} (\vec{v} \wedge \vec{n})_{|\Gamma} \end{cases} \tag{9.16}$$

(and the rotational operator enables the control $\vec{v} \wedge \vec{n}$ on Γ.

9.6 Some Dual spaces

Banach spaces:

$$\begin{aligned}
&(L^2(\Omega))' \simeq L^2(\Omega) \quad \text{(usual identification)} && ||f||_{L^2(\Omega)'} = ||f||_{L^2(\Omega)}. \\
&L^2_0(\Omega) = \{f \in L^2(\Omega) : \int_\Omega f\, d\Omega = 0\} \simeq L^2(\Omega)/\mathbb{R}, && ||f||_{L^2_0} = \inf_{c\in\mathbb{R}} ||f+c||_{L^2}. \\
&H^{-1}(\Omega) = H^1_0(\Omega)', && ||f||_{H^{-1}} = \sup_{v \in H^1_0(\Omega)} \frac{\langle f, v\rangle}{||v||_{H^1_0}}. \\
&H^{-\frac{1}{2}}(\Gamma) = (H^{\frac{1}{2}}(\Gamma))', && ||\mu||_{H^{-\frac{1}{2}}(\Gamma)} = \sup_{\lambda \in H^{\frac{1}{2}}(\Gamma)} \frac{|\langle \mu, \lambda\rangle|}{||\lambda||_{H^{\frac{1}{2}}(\Gamma)}}.
\end{aligned} \tag{9.17}$$

We have identified $(L^2(\Omega))'$ with $L^2(\Omega)$ thanks to the Riesz representation theorem in $(L^2(\Omega), (\cdot,\cdot)_{L^2})$, that is,

$$\forall \ell \in L^2(\Omega)',\ \exists! f \in L^2(\Omega),\ \forall g \in L^2(\Omega),\ \langle \ell, g\rangle_{L^{2\prime},L^2} = (f,g)_{L^2(\Omega)}, \quad \text{and} \quad ||f||_{L^2(\Omega)} = ||\ell||_{L^2(\Omega)'}. \tag{9.18}$$

Thus $H^1_0(\Omega) \subset H^1(\Omega) \subset L^2(\Omega) \simeq L^2(\Omega)' \subset (H^1(\Omega))' \subset H^{-1}(\Omega)$. And $L^2(\Omega)$ is named the "pivot space" (a central space in distribution theory of Schwartz [29]).

Proposition 9.2 *Let $\lambda \in H^{-\frac{1}{2}}(\Gamma)$ and let $w_\lambda \in H^1(\Omega)$ be the solution of the Nenmann problem: Find $w \in H^1(\Omega)$ s.t.*

$$\begin{cases} -\Delta w + w = 0 \quad \textit{dans } \Omega, \\ \dfrac{\partial w}{\partial n} = \lambda \quad \textit{sur } \Gamma, \end{cases} \tag{9.19}$$

then

$$||\lambda||_{H^{-\frac{1}{2}}(\Gamma)} = ||w_\lambda||_{H^1(\Omega)}. \tag{9.20}$$

Proof. (9.19) reads $(w,v)_{H^1(\Omega)} = \langle \lambda, v\rangle_{H^{-\frac{1}{2}}(\Gamma),H^{\frac{1}{2}}(\Gamma)}$ for all $v \in H^1(\Omega)$, thus (9.19) has a unique solution w_λ (Lax–Milgram theorem: The bilinear form given by $a(u,v) = (u,v)_{H^1(\Omega)}$ is trivially $H^1(\Omega)$-continuous and coercive, and the linear form given by $\ell(v) = \langle \lambda, v\rangle_{H^{-\frac{1}{2}}(\Gamma),H^{\frac{1}{2}}(\Gamma)}$ is continuous since

$|\ell(v)| \le ||\lambda||_{H^{-\frac12}(\Gamma)} ||\gamma_0(v)||_{H^{\frac12}(\Gamma)} \le ||\lambda||_{H^{-\frac12}(\Gamma)} ||v||_{H^1(\Omega)}$, cf. prop. 9.1). And :

$$
\begin{aligned}
||\lambda||_{H^{-\frac12}(\Gamma)} &= \sup_{d\in H^{\frac12}(\Gamma)} \frac{|\langle\lambda, d\rangle_{H^{-\frac12}(\Gamma),H^{\frac12}(\Gamma)}|}{||d||_{H^{\frac12}(\Gamma)}} \quad \text{(definition)}\\
&= \sup_{d\in H^{\frac12}(\Gamma)} \frac{|\langle\lambda, u_d\rangle_{H^{-\frac12}(\Gamma),H^{\frac12}(\Gamma)}|}{||u_d||_{H^1(\Omega)}} \quad \text{(cf. (9.11))}\\
&= \sup_{d\in H^{\frac12}(\Gamma)} \frac{|(w_\lambda, u_d)_{H^1(\Omega)}|}{||u_d||_{H^1(\Omega)}} \quad \text{(cf. (9.19))}\\
&\le ||w_\lambda||_{H^1(\Omega)}, \quad \text{(Cauchy–Schwarz in } H^1(\Omega)).
\end{aligned}
\tag{9.21}
$$

Et $\gamma_0(w_\lambda) \in H^{\frac12}(\Gamma)$ gives

$$
\begin{aligned}
||\lambda||_{H^{-\frac12}(\Gamma)} &\ge \frac{|\langle\lambda, \gamma_0(w_\lambda)\rangle_{H^{-\frac12}(\Gamma),H^{\frac12}(\Gamma)}|}{||\gamma_0(w_\lambda)||_{H^{\frac12}(\Gamma)}} \quad \text{(by definition of the sup)}\\
&\ge \frac{|(w_\lambda, w_\lambda)_{H^1(\Omega)}|}{||\gamma_0(w_\lambda)||_{H^{\frac12}(\Gamma)}} \quad \text{(cf. (9.19))}\\
&\ge ||w_\lambda||_{H^1(\Omega)} \quad \text{(since } ||\gamma_0(w_\lambda)||_{H^{\frac12}(\Gamma)} \le ||w_\lambda||_{H^1(\Omega)} \text{ cf. (9.8))}.
\end{aligned}
\tag{9.22}
$$

Thus (9.20). ∎

9.7 Dual of $H^1(\Omega)$ and $H^{\mathrm{div}}(\Omega)$ (characterizations)

Theorem 9.3 Dual of $H^1(\Omega)$.

$$
\ell \in (H^1(\Omega))' \quad \Rightarrow \quad \exists (f,\vec u) \in L^2(\Omega)\times L^2(\Omega)^n : \langle \ell, \psi\rangle = (f,\psi)_{L^2} + (\vec u, \vec{\mathrm{grad}}\psi)_{L^2} \quad \forall \psi \in H^1(\Omega). \tag{9.23}
$$

Dual of $H_0^1(\Omega)$.

$$
\ell \in H^{-1}(\Omega) \quad \Rightarrow \quad \exists (f,\vec u) \in L^2(\Omega)\times L^2(\Omega)^n \quad t.q. \quad \ell = f - \mathrm{div}\vec u. \tag{9.24}
$$

And if Ω is bounded then we can choose $f = 0$ (with $(H_0^1(\Omega), (\cdot,\cdot)_{H_0^1})$).

Proof. (from Brézis [6].) Characterization of $H^1(\Omega)^n$. Define $Z := L^2(\Omega)\times L^2(\Omega)^n$ provided with the inner product $((\phi,\vec u),(\psi,\vec v))_Z = (\phi,\psi)_{L^2} + (\vec u,\vec v)_{L^2}$ so that Z is a Hilbert space. Define $T : \begin{Bmatrix} H^1(\Omega) \to Z \\ \psi \to T\psi = (\psi, \vec{\mathrm{grad}}\psi) \end{Bmatrix}$. So $||\psi||_{H^1} = ||T\psi||_Z = ||(\psi,\vec{\mathrm{grad}}\psi)||_Z$, and $T : (H_0^1(\Omega),(\cdot,\cdot)_{H_0^1}) \to (\mathrm{Im}(T), ||.||_Z)$ is an isometry. Let $\ell \in H^1(\Omega)'$. Define

$$
\Phi_{\mathrm{Im}(T)} : \begin{cases} \mathrm{Im}(T) \to \mathbb{R} \\ (\psi, \vec v{=}\vec{\mathrm{grad}}\psi) \to \langle \Phi_{\mathrm{Im}(T)}, (\psi,\vec v)\rangle_{Z',Z} = \langle \ell, T^{-1}(\psi,\vec v)\rangle_{H^{1\prime},H^1} = \langle \ell,\psi\rangle_{H^{1\prime},H^1}. \end{cases}
$$

$\Phi_{\mathrm{Im}(T)}$ is linear (trivial) and continuous since ℓ and T^{-1} are. With Hahn–Banach theorem, extend $\Phi_{\mathrm{Im}(T)}$ to Z, so that we get a linear countinous form $\Phi_Z : \begin{Bmatrix} Z \to \mathbb{R} \\ (\psi,\vec v) \to \langle \Phi_Z, (\psi,\vec v)\rangle \end{Bmatrix}$. Then the Riesz representation theorem gives: $\exists(\phi,\vec u) \in Z$ s.t. $\langle \Phi_Z,(\psi,\vec v)\rangle = ((\phi,\vec u),(\psi,\vec v))_Z = \int_\Omega \phi\,\psi\,d\Omega + \int_\Omega \vec u.\vec v\,d\Omega$ for all $(\psi,\vec v)\in Z$. Then take $(\psi, \vec v{=}\vec{\mathrm{grad}}\psi) \in \mathrm{Im}(T)$ to get (9.23).

Similar proof for (9.24). ∎

Theorem 9.4 Dual de $H^{\mathrm{div}}(\Omega)$.

$$
F \in (H^{\mathrm{div}}(\Omega))' \quad \Rightarrow \quad \exists (\vec f,\phi) \in L^2(\Omega)^n \times L^2(\Omega) \text{ s.t. } \langle F,\vec v\rangle = (\vec f,\vec v)_{L^2} + (\phi, \mathrm{div}\vec v)_{L^2}, \quad \forall \vec v \in H^{\mathrm{div}}(\Omega). \tag{9.25}
$$

Dual de $H_0^{\mathrm{div}}(\Omega)$. *In particular*

$$F \in H_0^{\mathrm{div}}(\Omega)' \quad\Rightarrow\quad \exists(\vec{f},\phi) \in L^2(\Omega)^n \times L^2(\Omega) \text{ s.t. } F = \vec{f} - \vec{\mathrm{grad}}\phi. \tag{9.26}$$

And if Ω is bounded we can choose $\vec{f}=0$ (with $(H_0^{\mathrm{div}}(\Omega),(\cdot,\cdot)_{H_0^{\mathrm{div}}})$).

Proof. (Similar to the proof of (9.23).) Define $Z = L^2(\Omega)^n \times L^2(\Omega)$ provided with the inner product $((\vec{u},p),(\vec{v},q))_Z = (\vec{u},\vec{v})_{L^2}+(p,q)_{L^2}$ so that $(Z,(\cdot,\cdot)_Z)$ is a Hilbert space. Define $T : \left\{\begin{array}{ll} H^{\mathrm{div}}(\Omega) & \to Z \\ \vec{v} & \to T\vec{v} = (\vec{v},\mathrm{div}\vec{v}) \end{array}\right\}$. So $||\vec{v}||_{H^{\mathrm{div}}} = ||T\vec{v}||_Z = ||(\vec{v},\mathrm{div}\vec{v})||_Z$, and $T : (H^{\mathrm{div}}(\Omega),||.||_{H^{\mathrm{div}}}) \to (\mathrm{Im}(T),||.||_Z)$ is an isometry. Let $F \in H^{\mathrm{div}}(\Omega)'$. The mapping $(\vec{v},q{=}\mathrm{div}\vec{v}) \in \mathrm{Im}(T) \to \langle F, T^{-1}(\vec{v},q)\rangle_{H^{\mathrm{div}\prime},H^{\mathrm{div}}} = \langle F,\vec{v}\rangle_{H^{\mathrm{div}\prime},H^{\mathrm{div}}}$ is a linear form (trivial) that is continuous since F and T^{-1} are. With Hahn–Banach theorem, extend it to Z to get a linear continuous form named $\Phi : (\vec{v},q) \in Z \to \langle \Phi,(\vec{v},q)\rangle$. Then the Riesz representation theorem gives: $\exists(\vec{u},p) \in Z$ s.t. $\langle \Phi,(\vec{v},q)\rangle = ((\vec{u},p),(\vec{v},q))_Z = \int_\Omega \vec{u}.\vec{v}\,d\Omega + \int_\Omega pq\,d\Omega$ for all $(\vec{v},q) \in Z$. An choose $(\vec{v},q{=}\mathrm{div}\vec{v}) \in \mathrm{Im}(T)$ to get (9.25).

Similar proof for (9.26). ∎

9.8 Kernel of the trace operators

Ω is supposed to be a regular open set.

$$\begin{array}{ll} \mathrm{Ker}(\gamma_0) = H_0^1(\Omega), & \mathrm{Im}(\gamma_0) = H^{\frac12}(\Gamma) \text{ dense in } L^2(\Gamma). \\ \mathrm{Ker}(\gamma_1) \cap \mathrm{Ker}(\gamma_0) = H_0^2(\Omega), & \mathrm{Im}(\gamma_1) = H^{\frac12}(\Gamma). \\ \mathrm{Ker}(\gamma_n) = H_0^{\mathrm{div}}(\Omega), & \mathrm{Im}(\gamma_n) = H^{-\frac12}(\Gamma). \\ \mathrm{Ker}(\vec{\gamma}_t) = H_0^{\mathrm{curl}}(\Omega), & \mathrm{Im}(\vec{\gamma}_t) = H^{-\frac12}(\Gamma)^3. \end{array} \tag{9.27}$$

9.9 Poincaré–Friedrichs

(See e.g. manuscript "Eléments finis", or Raviart–Thomas [28], or Ciarlet [12]...

If Ω is bounded (at least in one direction), then we have Poincaré's inequality in $H_0^1(\Omega)$: There exists $c_\Omega > 0$ s.t

$$\forall v \in H_0^1(\Omega), \quad ||v||_{L^2} \le c_\Omega ||\vec{\mathrm{grad}}v||_{L^2}, \tag{9.28}$$

and the norms $||v||_{H^1(\Omega)}$ and $||\vec{\mathrm{grad}}v||_{L^2(\Omega)}$ are equivalent in $H_0^1(\Omega)$ (this space is closed in $H^1(\Omega)$ it is the closure of $\mathcal{D}(\Omega)$ in $H^1(\Omega)$).

And if Ω is bounded then there exists $c_\Omega > 0$ s.t:

$$\forall v \in H_0^1(\Omega) \bigcap H^2(\Omega), \qquad ||v||_{H^2(\Omega)} \le c_\Omega ||\Delta v||_{L^2(\Omega)}. \tag{9.29}$$

and the norms $||v||_{H^2(\Omega)}$ and $||\Delta v||_{L^2(\Omega)}$ are equivalent in $H_0^1(\Omega) \bigcap H^2(\Omega)$ (this space is not closed in $H^1(\Omega)$).

9.10 $L^2(\Omega)^n$ Decomposition (Helmholtz)

Let $\Omega \subset \mathbb{R}^n$ an open regular bounded set. Let $\mathrm{div} : H^{\mathrm{div}}(\Omega) \to L^2(\Omega)$, so $\mathrm{Ker}(\mathrm{div}) = \{\vec{v} \in H^{\mathrm{div}}(\Omega) : \mathrm{div}\vec{v} = 0\}$. And let

$$\mathrm{Ker}(\mathrm{div})_0 = \mathrm{Ker}(\mathrm{div}) \cap H_0^{\mathrm{div}}(\Omega) = \{\vec{v} \in \mathrm{Ker}(\mathrm{div}) : (\vec{v}.\vec{n})_{|\Gamma} = 0\}, \tag{9.30}$$

the subspace of incompressible functions with Γ impervious.

Theorem 9.5

$$\left\{\begin{array}{l} L^2(\Omega)^n = \vec{\mathrm{grad}}(H_0^1(\Omega)) \oplus^{\perp_{L^2}} \mathrm{Ker}(\mathrm{div}), \\ L^2(\Omega)^n = \vec{\mathrm{grad}}(H^1(\Omega)) \oplus^{\perp_{L^2}} \mathrm{Ker}(\mathrm{div})_0, \end{array}\right. \tag{9.31}$$

i.e, for any $\vec{f} \in L^2(\Omega)^n$ there exists $(\phi,\vec{w}) \in \vec{\mathrm{grad}}(H_0^1(\Omega)) \times \mathrm{Ker}(\mathrm{div})$ for (9.31)$_1$, and there exists $(\phi,\vec{w}) \in \vec{\mathrm{grad}}(H^1(\Omega)) \times \mathrm{Ker}(\mathrm{div})_0$ for (9.31)$_2$, s.t.

$$\vec{f} = \vec{\mathrm{grad}}\phi + \vec{w}, \quad \textit{with} \quad (\vec{\mathrm{grad}}\phi,\vec{w})_{L^2} = 0. \tag{9.32}$$

Proof. Let $\vec{f} \in L^2(\Omega)^n$.

For $(9.31)_1$, consider the solution of the homogenous Dirichlet problem: Find $\phi \in H_0^1(\Omega)$ s.t. $\Delta\phi = \mathrm{div}\vec{f}$ (distribution), meaning, find $\phi \in H_0^1(\Omega)$ s.t. $(\vec{\mathrm{grad}}\phi, \vec{\mathrm{grad}}\psi)_{L^2} = (\vec{f}, \vec{\mathrm{grad}}\psi)_{L^2}$ for all $\psi \in H_0^1(\Omega)$. The Lax–Milgram theorem gives a unique solution $\phi \in H_0^1(\Omega)$.

Let $\vec{w} = \vec{f} - \vec{\mathrm{grad}}\phi \in L^2(\Omega)^n$. So $(\vec{w}, \vec{\mathrm{grad}}\psi)_{L^2} = 0$ for all $\psi \in H_0^1(\Omega)$, by definition of ϕ, thus $\vec{w} \perp_{L^2} \vec{\mathrm{grad}}(H_0^1(\Omega))$ and $\mathrm{div}\vec{w} = 0 \in H^{-1}(\Omega)$, and $0 \in L^2(\Omega)$, thus $\vec{w} \in H^{\mathrm{div}}(\Omega)$ and $\vec{w} \in \mathrm{Ker}(\mathrm{div})$; Thus $\vec{f} = \vec{w} + \vec{\mathrm{grad}}\phi \in \mathrm{Ker}(\mathrm{div}) \oplus^{\perp_{L^2}} \vec{\mathrm{grad}}(H_0^1(\Omega))$, thus $(9.31)_1$.

For $(9.31)_2$, consider the solution of the homogenous Neumann problem: Find $\phi \in H^1(\Omega)$ s.t. $\int_\Omega \vec{\mathrm{grad}}\phi.\vec{\mathrm{grad}}\psi \, d\Omega = \int_\Omega \vec{f}.\vec{\mathrm{grad}}\psi \, d\Omega$ for all $\psi \in H^1(\Omega)$. The Lax–Milgram theorem gives a unique solution $\phi \in H^1(\Omega)/\mathbb{R}$ (i.e. up to a constant), moreover with $\phi \in H^2(\Omega)$ (regularity result thanks to $\vec{f} \in L^2(\Omega)$), so that $-\langle \Delta\phi, \psi \rangle_{(H^1(\Omega))', H^1(\Omega)} + \int_\Gamma \vec{\mathrm{grad}}\phi(x).\vec{n}(x)\psi(x) \, d\Gamma = -\langle \mathrm{div}\vec{f}, \psi \rangle_{(H^1(\Omega))', H^1(\Omega)}$ for all $\psi \in H^1(\Omega)$. In particular $\psi \in H_0^1(\Omega)$ gives $\Delta\phi = \mathrm{div}\vec{f} \in (H^1(\Omega))'$, and we are left with $\int_\Gamma \vec{\mathrm{grad}}\phi(x).\vec{n}(x)\psi(x) \, d\Gamma$ for all $\psi \in H^1(\Omega)$, thus for all $\psi_{|\Gamma} \in H^{\frac12}(\Gamma)$, thus $\vec{\mathrm{grad}}\phi.\vec{n}_{|\Gamma} = 0$.

Let $\vec{w} = \vec{f} - \vec{\mathrm{grad}}\phi \in L^2(\Omega)^n$. Thus $(\vec{w}, \vec{\mathrm{grad}}\psi)_{L^2} = 0$ for all $\psi \in H^1(\Omega)$, by definition of ϕ, thus $\vec{w} \perp \vec{\mathrm{grad}}(H^1(\Omega))$. And $\mathrm{div}\vec{w} = \mathrm{div}\vec{f} - \Delta\phi = 0$, thus $\mathrm{div}\vec{w} \in L^2(\Omega)$ and $\vec{w} \in \mathrm{Ker}(\mathrm{div})$. With $\int_\Omega \vec{w}.\vec{\mathrm{grad}}\psi \, d\Omega = 0$ for all $\psi \in H^1(\Omega)$, thus $\int_\Omega \vec{w}.\vec{n} \, \psi \, d\Gamma = 0$ for all $\psi \in H^1(\Omega)$, and $\vec{w}.\vec{n} = 0 \in H^{-\frac12}(\Gamma)$ (since $H^{\frac12}(\Gamma)$ is dense in $L^2(\Gamma)$), thus $\vec{w} \in \mathrm{Ker}(\mathrm{div})_0$, thus $\vec{f} = \vec{\mathrm{grad}}\phi + \vec{w} \in \vec{\mathrm{grad}}(H^1(\Omega)) \oplus^{\perp_{L^2}} \mathrm{Ker}(\mathrm{div})_0$, thus $(9.31)_2$. ▪

10 A surjectivity of the gradient operator

See e.g. Girault–Raviart [18]. We deal here with infinite dimensional spaces. The surjectivity of $\vec{\mathrm{grad}}$ is need for a Stokes like problem, see (1.2).

10.1 The theorem

Let Ω be an open regular set in $\mathbb{R}^n$. Let $(\vec{e}_i)$ be a given Cartesian $\mathbb{R}^n$, and $\vec{n}(x) = \sum_{i=1}^n n_i(x)\vec{e}_i$ be the outer normal unit to Γ at x.

$H^{-1}(\Omega) = (H_0^1(\Omega))' = \mathcal{L}(H_0^1(\Omega); \mathbb{R})$ is the set of continuous linear forms defined on $H_0^1(\Omega)$, cf. (9.17).

With (9.18), $L^2(\Omega)'$ is identified to $L^2(\Omega)$, and $H^{-1}(\Omega) \supset L^2(\Omega)' = L^2(\Omega) \supset H_0^1(\Omega)$.

And if $g \in L^2(\Omega)$, then $\frac{\partial g}{\partial x_i} \in H^{-1}(\Omega)$, and, for all $\phi \in H_0^1(\Omega)$,

$$\langle \frac{\partial g}{\partial x_i}, \phi \rangle_{H^{-1}, H_0^1} := -\int_\Omega g(x) \frac{\partial \phi}{\partial x_i}(x) \, dx, \tag{10.1}$$

see Schwartz [29]. In particular, if $p \in L^2(\Omega)^n$ then $\vec{\mathrm{grad}}p \in H^{-1}(\Omega)^n$ and, for all $\vec{v} \in H_0^1(\Omega)^n$,

$$\langle \vec{\mathrm{grad}}p, \vec{v} \rangle_{H^{-1}, H_0^1} = -\int_\Omega p(x) \mathrm{div}\vec{v}(x) \, dx = -(p, \mathrm{div}\vec{v})_{L^2}. \tag{10.2}$$

Theorem 10.1 *The range of the gradient operator* $\vec{\mathrm{grad}} : \left\{ \begin{array}{ll} L^2(\Omega) & \to H^{-1}(\Omega) \\ p & \to \vec{\mathrm{grad}}p \end{array} \right\}$ *is closed, and its kernel* $\mathrm{Ker}(\vec{\mathrm{grad}})$ *is the set of constant functions.*

Proof. The proof of this quite difficult theorem is given in the next §. ▪

And the open mapping theorem, cf. (8.10), then gives the needed result for the Stokes like problem, cf. (1.2):

Corollary 10.2

$$\exists \beta > 0, \ \forall p \in L^2(\Omega), \quad ||\vec{\mathrm{grad}}p||_{H^{-1}} \geq \beta ||p||_{L^2(\Omega)/\mathbb{R}}, \tag{10.3}$$

that is $\exists \beta > 0, \ \inf_{p \in L^2(\Omega)} \sup_{v \in H_0^1(\Omega)} \frac{|(\mathrm{div}\vec{v}, p)_{L^2}|}{||\vec{v}||_{H_0^1(\Omega)/\mathrm{Ker}(\mathrm{div})} ||p||_{L_0^2(\Omega)}} \geq \beta$ *(inf-sup condition).*

10.2 Steps for the proof

10.2.1 Equivalent norms in $H^{-1}(\Omega)$

Ω being bounded, the Poincaré inequality gives:

$$\exists c_\Omega \in \mathbb{R},\ \forall q \in H_0^1(\Omega),\ ||q||_{L^2} \le c_\Omega ||q||_{H_0^1}. \tag{10.4}$$

Let $q \in L^2(\Omega)$, and let $\ell_q \in H^{-1}(\Omega)$ be defined on $H_0^1(\Omega)$ by

$$\forall \psi \in H_0^1(\Omega), \quad \langle \ell_q, \psi \rangle_{H^{-1},H_0^1} := (q, \psi)_{L^2(\Omega)}. \tag{10.5}$$

Thus ℓ_q is trivially linear, and, with (10.4),

$$\forall \psi \in H_0^1(\Omega), \quad |\langle \ell_q, \psi \rangle_{H^{-1},H_0^1}| = |(q, \psi)_{L^2(\Omega)}| \le ||q||_{L^2} ||\psi||_{L^2} \le c_\Omega ||q||_{L^2} ||\psi||_{H_0^1}. \tag{10.6}$$

Thus ℓ_q is continuous, thus $\ell_q \in H^{-1}(\Omega)$, $L^2(\Omega)$ is considered to be a subspace in $H^{-1}(\Omega)$.

Proposition 10.3 *If $q \in L^2(\Omega)$, then*

$$||\ell_q||_{H^{-1}} \le c_\Omega ||q||_{L^2}, \quad ||\vec{\text{grad}} q||_{H^{-1}} \le ||q||_{L^2}. \tag{10.7}$$

Thus the injection $\left\{ \begin{array}{ll} L^2(\Omega) & \to H^{-1}(\Omega) \\ q & \to \ell_q \end{array} \right\}$ *and the gradient operator* $\left\{ \begin{array}{ll} L^2(\Omega) & \to H^{-1}(\Omega)^n \\ q & \to \vec{\text{grad}} q \end{array} \right\}$ *are continuous. In particular, with (10.6),*

$$\text{if } q \in L^2(\Omega) \text{ then } \ell_q \overset{\text{denoted}}{=} q. \tag{10.8}$$

(The space $L^2(\Omega)$ is the pivot space.)

Proof. $(10.7)_1$ is given by (10.6). Let $q \in L^2(\Omega)$, with (10.2) we get $\vec{\text{grad}} q \in H^{-1}(\Omega)^n$. We have, for all $\vec{\phi} \in H_0^1(\Omega)^n$, cf. (10.2),

$$|\langle \vec{\text{grad}} q, \vec{\phi} \rangle_{H^{-1},H_0^1}| = |-(q, \text{div}\vec{\phi})_{L^2}| \le ||q||_{L^2} ||\text{div}\vec{\phi}||_{L^2} \le ||q||_{L^2} ||\text{grad}\vec{\phi}||_{(L^2)^{n^2}} \le ||q||_{L^2} ||\vec{\phi}||_{(H_0^1)^n}.$$

Thus $(10.7)_2$. ∎

Let :

$$||.||_+ : \left\{ \begin{array}{ll} L^2(\Omega) & \to \mathbb{R} \\ v & \to ||v||_+ = ||v||_{H^{-1}} + ||\vec{\text{grad}} v||_{H^{-1}}. \end{array} \right. \tag{10.9}$$

Corollary 10.4 *In $L^2(\Omega)$ the norms $||.||_{L^2}$ and $||.||_+$ are equivalent norms:*

$$\exists c_1, c_2 > 0,\ \forall v \in L^2(\Omega), \quad c_1 ||v||_+ \le ||v||_{L^2} \le c_2 ||v||_+. \tag{10.10}$$

Proof. (10.9) trivially defines a norm in $L^2(\Omega)$, and (10.7) gives $c_1 = \frac{1}{1+c_\Omega}$.

Let $Z = (L^2(\Omega), ||.||_+)$. Thanks to $\frac{1}{c_1}$, Z is a Banach space. Then consider the canonical injection $I_+ : v \in (L^2(\Omega), ||.||_{L^2(\Omega)}) \to I_+(v) = v \in (L^2(\Omega), ||.||_Z)$: it is the algebraic identity and thus is bijective. And I_+ is continuous (thanks to $\frac{1}{c_1}$). Thus ${I_+}^{-1} : v \in (L^2(\Omega), ||.||_Z) \to I_+(v) = v \in (L^2(\Omega), ||.||_{L^2(\Omega)})$ is continuous (open mapping theorem 8.3). Then let $c_2 = ||{I_+}^{-1}||$. ∎

10.2.2 Rellich theorem $L^2(\Omega) \to H^{-1}(\Omega)$

Reminder: An operator $\kappa \in \mathcal{L}(E; F)$ is compact iff $\overline{\kappa(B_E(0,1))}$ is compact in F.

Lemma 10.5 *Let E and F be Banach spaces. If $\kappa \in \mathcal{L}(E; F)$ is compact, then it dual $\kappa' : F' \to E'$ is compact.*

Proof. Let $(\ell_n) \in B_{F'}(0,1)$. We have to prove that the sequence $(T'.\ell_n) \in T'(B_{F'}) \subset E'$ has a converging subsequence. Let $K = \overline{T(B_E(0,1))}$. K is a compact in F since T is compact. Then consider the restriction $\phi_n = \ell_{n|K} : K \to \mathbb{R}$. So $(\phi_n)_{\mathbb{N}^*}$ is a sequence in $C^0(K;\mathbb{R})$, and $(\phi_n)_{\mathbb{N}^*} \subset B_{F'}(0,1)$ is a bounded set in F'. Moreover $(\phi_n)_{\mathbb{N}^*}$ is equicontinuous since ℓ_n is linear continuous $||\ell_n(y)|| \le ||\ell_n||\,||y||_F \le ||y||_F$). Thus the set $(\phi_n)_{\mathbb{N}^*}$ is relatively compact in $C^0(K;\mathbb{R})$ (Ascoli theorem, see Brézis [6]). Thus we can extract a convergent subsequence $(\phi_{n_k})_{k\in\mathbb{N}^*}$ in $C^0(K;\mathbb{R})$. Thus, $T(B_E)$ being relatively compact and thus bounded, we have

$$\sup_{x\in B_E} |\langle \ell_{n_k} - \ell_{n_m}, T.x\rangle| \underset{k,m\to\infty}{\longrightarrow} 0.$$

Thus $||T'.\ell_{n_k} - T'.\ell_{n_m}||_{E'} \to 0$. Thus E' being a Banach space, since E is, $(T'.\ell_{n_k})_{k\in\mathbb{N}^*}$ converges in E'. Thus the set $\overline{(T'.\ell_{n_k})_{k\in\mathbb{N}^*}}$ is compact, thus T' is compact. ∎

Theorem 10.6 *(Rellich) The canonical injection* $T : v \in L^2(\Omega) \to v \in H^{-1}(\Omega)$ *is compact.*

Proof. $I_{10} : v \in H_0^1(\Omega) \to v \in L^2(\Omega)$ is compact, Rellich theorem see Brézis [6], thus $I_{10}' : v \in L^2(\Omega) \to v \in H^{-1}(\Omega)$ is compact, cf. Lemma 10.5. ∎

10.2.3 Petree–Tartar compactness theorem

Let E and F be two Banach spaces, and $T \in \mathcal{L}(E;F)$ (linear and continuous). The purpose is to prove that the range of T is eventually closed. But to use theorem 8.3 and (8.8) to prove it, can be difficult. It can be easier to find a compact operator $\kappa : E \to G$, where G is a Banach space, s.t.

$$\exists\gamma > 0,\ \forall x \in E, \quad ||T.x||_F + ||\kappa.x||_G \ge \gamma||x||_E. \tag{10.11}$$

Theorem 10.7 *Let E, F and G be three Banach spaces, let $T \in \mathcal{L}(E;F)$ be injective (one-to-one), and $\kappa \in \mathcal{L}(E;G)$ be compact. If (10.11) holds then (8.8) holds, and thus* $\mathrm{Im}(T)$ *is closed.*

(If T is not injective, consider $E/\mathrm{Ker}(T)$.)

Proof. Suppose (8.8) is false. Thus there exists a sequence $(x_n)_{n\in\mathbb{N}^*}$ in E s.t. $||x_n||_E = 1$ and $||T.x_n|| \longrightarrow_{n\to\infty} 0$, cf. (8.10). And κ being compact and $(x_n)_{n\in\mathbb{N}^*}$ being bounded, the sequence $(\kappa.x_n)_{n\in\mathbb{N}^*}$ has a convergent subsequence $(\kappa.x_{n_k})_{k\in\mathbb{N}^*}$ that converges in the Banach space G. With T continuous, κ compact and the hypothesis (10.11), we get

$$\gamma||x_{n_i} - x_{n_j}||_E \le ||T.x_{n_i} - T.x_{n_j}||_F + ||\kappa.x_{n_i} - \kappa.x_{n_j}||_G \underset{i,j\to\infty}{\longrightarrow} 0 + 0 = 0.$$

Thus $(x_{n_k})_{k\in\mathbb{N}^*}$ is a Cauchy sequence in the Banach space E, so converges to a limit $x \in E$. Since $||T.x_{n_k}|| \longrightarrow_{k\to\infty} 0$ and T is continuous, we get $||T.x|| = 0$, thus $x = 0$ since T is injective. But $||x_{n_k}||_E = 1$ implies $||x||_E = 1$. Absurd, thus (8.8) is true. ∎

10.2.4 The range of $\vec{\mathrm{grad}} : L^2(\Omega) \to H^{-1}(\Omega)^n$ is closed

We can now prove theorem 10.1. Let $T = \vec{\mathrm{grad}} : L^2(\Omega) \to H^{-1}(\Omega)^n$, and κ the canonical injection $L^2(\Omega) \to H^{-1}(\Omega)$. Since T is linear continuous, cf. (10.7), and κ is compact, cf. Rellich theorem 10.6, the Petree–Tartar theorem 10.7 implies that the range of T is closed. ∎

11 The closed range theorem

The results and full proofs can be found e.g. in Brézis [6].

11.0.1 The closed range theorem

Let $T \in L(E;F)$, so $T' \in \mathcal{L}(F';E')$, cf. (8.5). We have

$$\mathrm{Ker}(T') = \{m \in F' \text{ s.t. } T'.m = 0\} = \{m \in F' \text{ s.t. } m.T = 0\} \subset F', \tag{11.1}$$

since $m \in \mathrm{Ker}(T') \Leftrightarrow T'.m = 0 \Leftrightarrow \langle T'.m, x\rangle_{E',E} = 0 = \langle m, T.x\rangle_{F',F} = m(T.x) = (m \circ T)(x)$ for all $x \in E$ $\Leftrightarrow m.T = 0$, with $m.T$ the notation of $m \circ T$ when the maps are linear maps.

If $M \subset E$ is a linear subspace in E then the dual orthogonal of M is

$$M^o := \{\ell \in E' : \langle \ell, x\rangle_{E',E} = 0,\ \forall x \in M\} \quad (\subset E') \tag{11.2}$$

(the subspace of E' of linear forms vanishing on M). M^o is a linear subspace in E' (trivial). And M^o is closed in E': Indeed if $(\ell_n)_{\mathbb{N}^*}$ is a Cauchy sequence in M^o, so $||\ell_n - \ell_m||_{E'} \longrightarrow_{n,m\to\infty} 0$, then, for $x \in E$ the sequence $(\ell_n(x))$ is a Cauchy sequence in $\mathbb{R}$, thus convergence toward a real named $\ell(x)$; This defines a function $\ell : E \to \mathbb{R}$. And $\ell(x_1 + \lambda x_2) = \lim_{n\to\infty} \ell_n(x_1 + \lambda x_2) = \lim_{n\to\infty} \ell_n(x_1) + \lambda \lim_{n\to\infty}(x_2) = \ell(x_1) + \lambda \ell(x_2)$, thus ℓ is linear, and, for $x \in E$, ℓ is continuous at x since $|\ell.x' - \ell.x| \le |(\ell - \ell_n).x' + (\ell - \ell_n).x| + |\ell_n.x' - \ell_n.x| \le (||\ell - \ell_n|| + ||\ell_n||)||x' - x||_E$ with $||\ell_n|| \le ||\ell_N|| + 1$ for N large enough and $n \ge N$.

If $N \subset F'$ is a linear subspace in F' then let

$$N^\perp := \{y \in F : \langle m, y\rangle_{F',F} = 0,\ \forall m \in N\} \quad (\subset F). \tag{11.3}$$

Then $N^\perp$ is a linear subspace in F (trivial) that is closed in F (similar proof than for M^o). To be compared with, cf. (11.2),

$$N^o = \{y'' \in F'' : \langle y'', m\rangle_{F'',F'} = 0,\ \forall m \in N\} \quad (\subset F''). \tag{11.4}$$

Remark 11.1 If F is reflexive, that is $F'' \simeq F$ (identification) then $N^o \simeq N^\perp$ (identification). Indeed, with J the canonical isomorphism given in (8.7), if $y \in N^\perp$ then let $y'' = J(y) \in F$, so for all $m \in N$ we have $0 = m.y = y''.m$, and thus $y'' \in N^o$; And if $y'' \in N^o$ then let $y \in F$ s.t. $J(y) = y''$ (thanks to the reflexivity), then for all $m \in N$ we have $0 = y''.m = m.y$, and thus $y \in N^\perp$. ∎

Theorem 11.2 (Closed range theorem) *Let E and F be Banach spaces and $T \in \mathcal{L}(E;F)$ (linear and continuous). Then the following properties are equivalent:*
(i) $\mathrm{Im}(T)$ *is closed in* F,
(ii) $\mathrm{Im}(T')$ *is closed in* E',
(iii) $\mathrm{Im}(T) = \mathrm{Ker}(T')^\perp$,
(iv) $\mathrm{Im}(T') = \mathrm{Ker}(T)^o$.

We then deduce, with (8.16):

Corollary 11.3 *If* $\mathrm{Im}(T)$ *is closed in* F *then* $\mathrm{Im}(T')$ *is closed in* E'*, thus*

$$\exists \gamma' > 0, \quad \forall \ell \in F', \quad ||T'.\ell||_{E'} \ge \gamma' \, ||\ell||_{F'/\mathrm{Ker}(T')}. \tag{11.5}$$

Proof. The full proof of theorem 11.2 (even for unbounded operators with dense domain of definition) can be found e.g. in Brézis [6] or Yosida [34] (for locally convex spaces that are metrizable and complete). We give here the proof in the simplified case of T a linear continuous mapping between two Banach spaces (sufficient for our needs). We need some lemmas:

Lemma 11.4 *If E is a Banach space and M is a linear subspace in E, then*

$$\overline{M} = (M^o)^\perp. \tag{11.6}$$

Proof. If $x \in M$ then $\langle \ell, x\rangle_{E',E} = 0$ for all $\ell \in M^o$, thus $x \in (M^o)^\perp$, cf. (11.3). And $(M^o)^\perp$ being closed we get $\overline{M} \subset (M^o)^\perp$.

Conversely: Suppose $x_0 \in (M^o)^\perp$ and $x_0 \notin \overline{M}$; Then $\{x_0\}$ being compact and $\overline{M}$ being closed and convex (it is a linear subspace), there exists a hyperplane that strictly separates x_0 and $\overline{M}$ (geometric form or the Hahn–Banach theorem), that is there exists $\ell \in E'$ and $\alpha \in \mathbb{R}$ s.t. $\langle \ell, x\rangle_{E',E} < \alpha < \langle \ell, x_0\rangle_{E',E}$ for all $x \in M$. And M being a linear space, taking $-x \in M$, it follows that $\langle \ell, x\rangle_{E',E} = 0$ for all $x \in M$, thus $\ell \in M^o$. And $\langle \ell, x\rangle_{E',E} = 0$ for all $x \in M$ implies $\langle \ell, x_0\rangle_{E',E} > 0$ with $x_0 \notin \overline{M}$, thus $\ell \notin M^o$. Absurd, thus $(M^o)^\perp \subset \overline{M}$. ∎

Lemma 11.5 *If G and L are two closed subspaces in a Banach space, then*

$$G \cap L = (G^o + L^o)^\perp, \quad \text{and} \quad G^o \cap L^o = (G+L)^o. \tag{11.7}$$

Proof. $(11.7)_1$: If $x \in G \cap L$, and if $m = g + \ell \in G^o + L^o$, then $m.x = g.x + \ell.x = 0 + 0$, thus $x \in (G^o + L^o)^\perp$.

Conversely, we have $(G^o + L^o)^\perp \subset (G^o)^\perp$ (particular case of: If $Y \subset Z$ then $Z^\perp \subset Y^\perp$), and $(G^o)^\perp = G$ since G is closed,, thus if $x \in (G^o + L^o)^\perp$ then $x \in G$; And similarly $x \in L$,; Thus $x \in G \cap L$.

Similar proof for $(11.7)_2$. ∎

Let $E \times F$ be equipped with the (usual) norm $||(x,y)||_{E\times F} = \max(||x||_E, ||y||_F)$, so $E \times F$ is a Banach space.

Lemma 11.6 *If T is continuous, then its graph*

$$G(T) = \{(x,y) \in E \times F \text{ s.t. } \exists x \in E,\ y = T.x\} = \{(x, T.x) \in E \times F\} \tag{11.8}$$

is closed in $E \times F$.

Proof. If $((x_n, T.x_n)))_{\mathbb{N}^*}$ is a Cauchy sequence in $G(T)$, then, E being a Banach space, $(x_n)_{\mathbb{N}^*}$ converges toward a $x \in E$, thus, T being continuous, $T.x_n$ convergence toward $T.x \in F$, so $(x, T.x) \in G(T)$. ∎

We have

$$G(T') = \{(m, T'.m) \in F' \times E'\}. \tag{11.9}$$

Lemma 11.7 *If T is continuous, then $G(T')$ is closed in $F' \times E'$, and*

$$(m, \ell) \in G(T') \iff (-\ell, m) \in G(T)^o. \tag{11.10}$$

Proof. Let $((m_n, T'.m_n))_{\mathbb{N}^*}$ be a sequence in $G(T)$ s.t. $(m_n, T'.m_n) \longrightarrow_{n\to\infty} (m, z) \in F' \times E'$. Thus $m_n \longrightarrow m \in F'$ and $T'.m_n \longrightarrow_{n\to\infty} k \in E'$, that is $\langle T'.m_n, x\rangle_{E',E} \longrightarrow_{n\to\infty} \langle k, x\rangle_{E',E} \in \mathbb{R}$ for all $x \in E$. And we have to check that $k = T'.m$. For all $x \in E$, we have $\langle T'.m_n, x\rangle_{E',E} = \langle m_n, T.x\rangle_{F',F}$, thus $\langle m_n, T.x\rangle_{F',F} \longrightarrow_{n\to\infty} \langle k, x\rangle_{E',E}$, i.e. $\langle T' m_n, x\rangle_{F',F} \longrightarrow_{n\to\infty} \langle k, x\rangle_{E',E}$, thus $T' m_n, \longrightarrow_{n\to\infty} k \in F'$. So $G(T')$ is closed.

$(m,\ell) \in G(T') \Leftrightarrow \ell = T'.m \Leftrightarrow \langle \ell, x\rangle_{E',E} = \langle T'.m, x\rangle_{E',E} = \langle m, T.x\rangle_{E',E}$ for all $x \in E \Leftrightarrow \langle \ell, x\rangle_{E',E} - \langle m, T.x\rangle_{E',E} = 0$ for all $x \in E \Leftrightarrow \langle (\ell, -m), (x, T.x)\rangle_{E'\times F', E\times F} = 0$ for all $x \in E \Leftrightarrow (\ell, -m) \in G(T)^o$. ∎

Define

$$L := E \times \{0\}. \tag{11.11}$$

L is closed in $E \times F$ since E and $\{0\}$ are, and

$$L^o = \{0\} \times F'. \tag{11.12}$$

Indeed $L^o = \{(\ell, m) \in E' \times F' : \langle (\ell, m), (x, 0)\rangle_{E'\times F', E\times F} = 0, \forall x \in E\} = \{(\ell, m) \in E' \times F' : \langle \ell, x\rangle_{E',E} + 0 = 0, \forall x \in E\} = \{0\} \times F'$.

Lemma 11.8

$$\mathrm{Ker}(T) \times \{0\} = G(T) \cap L \tag{11.13}$$

$$E \times \mathrm{Im}(T) = G(T) + L \tag{11.14}$$

$$\{0\} \times \mathrm{Ker}(T') = G(T)^o \cap L^o \tag{11.15}$$

$$\mathrm{Im}(T') \times F' = G(T)^o + L^o \tag{11.16}$$

Proof. $(x,y) \in \mathrm{Ker}(T) \times \{0\}$ iff $Tx = 0$ and $y = 0$; And $(x,y) \in G(T) \cap L$ iff $y = Tx$ and $y = 0$, thus (11.13).

$(x_1, y_1) \in E \times \mathrm{Im}(T)$ iff $(x_1, y_1) = (x_1, T.x_1')$ for some $x_1' \in E$; And $(x_2, y_2) \in G(T) + L$ iff $\exists x_2' \in E$ and $\exists x_2'' \in E$ s.t. $(x_2, y_2) = (x_2', T.x_2') + (x_2'', 0) = (x_2' + x_2'', T.x_2') = (x_3, T.(x_3 - x_2'))$, thus (11.14).

$(\ell, m) \in \{0\} \times \mathrm{Ker}(T')$ iff $\ell = 0$ and $m \in \mathrm{Ker}T'$; And $(\ell, m) \in G(T)^o \cap L^o$ iff $(-m, \ell) \in G(T')$, cf. (11.10), and $(\ell, m) \in L^o$, i.e. iff $\ell = -T'.m$ and $\ell = 0$, i.e. iff $\ell = 0$ and $m \in \mathrm{Ker}T'$, thus (11.15).

$(\ell, m) \in \mathrm{Im}(T') \times F'$ iff $\exists k \in F'$ s.t. $\ell = T'.k$ and $m \in F'$; And $(\ell, m) \in G(T)^o + L^o$ iff $\exists (\ell_1, m_1) \in G(T)^o$ and $\exists (\ell_2, m_2) \in L^o$ s.t. $\ell = \ell_1 + \ell_2$ and $m = m_1 + m_2$, i.e., with (11.10) and (11.12), iff $\exists m_1 \in F'$ (and then $-\ell_1 = T'.m_1$) and $m_2 \in F'$ s.t. $\ell = -T.m_1 + 0$ and $m = m_1 + m_2$, i.e. iff $\ell \in \mathrm{Im}(T')$ and $m \in F'$, thus (11.16). ∎

Corollary 11.9

$$\mathrm{Ker}(T) = \mathrm{Im}(T')^{\perp}, \tag{11.17}$$

$$\mathrm{Ker}(T') = \mathrm{Im}(T)^{o}, \tag{11.18}$$

$$(\mathrm{Ker}(T))^{o} = \overline{\mathrm{Im}(T')}, \tag{11.19}$$

$$\mathrm{Ker}(T')^{\perp} = \overline{\mathrm{Im}(T)}. \tag{11.20}$$

Proof. (11.16) gives $R(T')^{\perp} \times \{0\} = (G(T)^o + L^o)^{\perp} = G(T) \cap L$, cf. (11.7), thus $= \mathrm{Ker}(T) \times \{0\}$, cf. (11.13), thus (11.17). Thus (11.19).

(11.14) gives $\{0\} \times \mathrm{Im}(T)^o = (G(T) + L)^o = G^o \cap L^o$, cf. (11.7), thus $= \{0\} \times \mathrm{Ker}(T')$, cf. (11.15), thus (11.18). Thus (11.20). ∎

Proof of theorem 11.2: apply corollary 11.9.

12 A well-posed mixed problem

12.1 Notations

Let V and Q be two Banach spaces. Let $b(\cdot,\cdot) : V \times Q \to \mathbb{R}$ be a bilinear form. $b(\cdot,\cdot)$ is said to be continuous (or bounded) iff

$$\exists c > 0, \quad \forall (v,q) \in B_V(0,1) \times B_Q(0,1), \quad |b(v,q)| \le c. \tag{12.1}$$

Then let

$$||b|| := \sup_{\substack{v \in B_V(0,1) \\ q \in B_Q(0,1)}} |b(v,q)|. \tag{12.2}$$

And let $\mathcal{L}(V,Q;\mathbb{R})$ be the space of bilinear and continuous forms with its (usual) norm given by (12.2).

If $b(\cdot,\cdot) \in \mathcal{L}(V,Q;\mathbb{R})$ (bilinear and continuous), then define

$$B : \left\{ \begin{array}{ll} V & \to Q' \\ v & \to Bv \end{array} \right\} \quad \text{and} \quad B^t : \left\{ \begin{array}{ll} Q & \to V' \\ q & \to B^t q \end{array} \right\} \tag{12.3}$$

by

$$b(v,q) = \langle Bv, q \rangle_{Q',Q} = \langle B^t q, v \rangle_{V',V}. \tag{12.4}$$

Thus B and B^t are linear (trivial) and continuous with

$$||B|| = ||B^t|| = ||b||. \tag{12.5}$$

Indeed $||Bv||_{V'} = \sup_{q \in B_Q(0,1)} |\langle Bv, q \rangle_{Q',Q}| = \sup_{q \in B_Q(0,1)} |b(v,q)| \le \sup_{q \in B_Q(0,1)} ||b|| \, ||v||_V ||q||_Q = ||b|| \, ||v||_V$ gives $||B|| \le ||b||$ (continuity), and $|b(x,y)| = |\langle Bv, q \rangle_{Q',Q}| \le ||Bx||_{Q'} ||y||_Q \le ||B|| \, ||x||_V ||y||_Q$ gives $||b|| \le ||B||_{V'}$. So $B \in \mathcal{L}(V;Q')$. Idem pour B^t.

Suppose Q reflexive, cf. definition 8.2, then the dual $B' \in \mathcal{L}(Q'';V')$ of $B \in \mathcal{L}(V;Q')$, defined by $\langle B'v, \ell \rangle_{V',V} = \langle v, B\ell \rangle_{Q'',Q}$ for all $v \in Q''$ and $\ell \in V'$ cf. (8.5), is identified to B^t:

$$\mathcal{L}(Q'';V') \ni B' \simeq B^t \in \mathcal{L}(Q;V'). \tag{12.6}$$

Suppose V reflexive, cf. definition 8.2, then the dual $(B^t)' \in \mathcal{L}(V'';Q')$ of $B^t \in \mathcal{L}(Q;V')$, defined by $\langle (B^t)'v, q \rangle_{Q',Q} = \langle v, B^t \ell \rangle_{V'',V'}$ for all $v \in Q''$ and $\ell \in V'$ cf. (8.5), is identified to B^t:

$$\mathcal{L}(V'';Q') \ni (B^t)' \simeq B \in \mathcal{L}(V;Q'). \tag{12.7}$$

12.2 The mixed problem

Let $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ and $b(\cdot,\cdot) : V \times Q \to \mathbb{R}$ be bilinear forms. Let $f \in V'$ and $g \in Q'$ (linear forms). A mixed problem is a problem of the type: Find $(u,p) \in V \times Q$ s.t.

$$\left\{ \begin{array}{lll} a(u,v) + b(v,p) & = \langle f, v \rangle_{V',V}, & \forall v \in V, \\ b(u,q) & = \langle g, q \rangle_{Q',Q}, & \forall q \in Q, \end{array} \right. \tag{12.8}$$

cf. (1.1).

12.3 The inf-sup conditions

For the existence (and control) of p, we suppose that the range $\text{Im}(B^t)$ of $B^t : Q \to V'$ is closed, that is, cf. (8.10),

$$\exists\beta > 0,\ \forall q \in Q,\ ||B^t q||_{V'} \geq \beta||q||_{Q/\text{Ker}(B^t)}, \tag{12.9}$$

i.e.,

$$\exists\beta > 0,\ \forall q \in Q,\ \sup_{v\in V} \frac{b(v,q)}{||v||_{V/\text{Ker}(B)}} \geq \beta||q||_{Q/\text{Ker}(B^t)}, \tag{12.10}$$

also written as the inf-sup condition $\inf_{q\in Q}\sup_{v\in V} \frac{b(v,q)}{||v||_{V/\text{Ker}(B)}||q||_{Q/\text{Ker}(B^t)}} \geq \beta$.

For the existence (and control) of u, we suppose the range $\text{Im}(B)$ of $B : V \to Q'$ is closed, that is, we suppose, cf. (8.10),

$$\exists\beta > 0,\ \forall v \in V,\ ||Bv||_{Q'} \geq \beta||v||_{V/\text{Ker}(B)}, \tag{12.11}$$

i.e.,

$$\exists\beta > 0,\ \forall v \in V,\ \sup_{q\in Q} \frac{b(v,q)}{||q||_{Q/\text{Ker}(B^t)}} \geq \beta||v||_{V/\text{Ker}(B)}, \tag{12.12}$$

also written as the inf-sup condition $\inf_{v\in V}\sup_{q\in Q} \frac{b(v,q)}{||v||_{V/\text{Ker}(B)}||q||_{Q/\text{Ker}(B^t)}} \geq \beta$.

Remark: With (12.6) or (12.7), the reflexivity of Q or V gives that (12.11) implies (12.9) or (12.9) implies (12.11).

12.4 The theorem for mixed problem

Theorem 12.1 . **Hypotheses:** *(i)* $(V, (\cdot,\cdot)_V)$ *is a Hilbert space,* $(Q, ||.||_Q)$ *is a reflexive Banach space,* $f \in V'$, *and* $g \in Q'$.

(ii) The bilinear form $a(\cdot,\cdot)$ *is continuous on* V, *cf. (12.1), and coercive on* $\text{Ker}(B)$, *that is,*

$$\exists\alpha > 0,\ \forall v \in \text{Ker}(B),\quad a(v,v) \geq \alpha||v||_V^2. \tag{12.13}$$

(iii) The bilinear form $b(\cdot,\cdot)$ *is continuous on* $V \times Q$, *cf. (12.1), and* B *is surjective (= onto), so we have (12.11) and then (12.9) since* Q *is reflexive.*

Conclusion: *Problem (12.8) has a unique solution* $(u,p) \in V \times Q/\text{Ker}B^t$ *that depends continuously on* f *and* g, *and more precisely, with* $C_a = (1 + \frac{||a||}{\alpha})$,

$$\begin{cases} ||u||_V \leq \dfrac{1}{\alpha}||f||_{V'} + \dfrac{C_a}{\beta}||g||_{Q'}, \\ ||p||_{Q/\text{Ker}B^t} \leq \dfrac{C_a}{\beta}\Big(||f||_{V'} + \dfrac{||a||}{\beta}||g||_{Q'}\Big). \end{cases} \tag{12.14}$$

Proof. Let $u_g \in V$ s.t. $B.u_g = g$, exists since B is surjective, and $||u_g||_{V/\text{Ker}(B)} \leq \frac{1}{\beta}||g||_{Q'}$, cf. (12.11).

Let $u_0 \in \text{Ker}(B)$ be the solution of the problem: Find $u_0 \in \text{Ker}(B)$ s.t.

$$a(u_0, v_0) = \langle f, v_0\rangle_{V',V} - a(u_g, v_0),\quad \forall v_0 \in \text{Ker}(B). \tag{12.15}$$

The Lax–Milgram theorem tells that (12.15) is well-posed: Indeed, $(\text{Ker}B, (\cdot,\cdot)_V)$ is a Hilbert space, $a(\cdot,\cdot)$ is bilinear continuous coercive, and $F : v_0 \in \text{Ker}(B) \to F(v_0) := \langle f, v_0\rangle_{V',V} - a(u_g, v_0)$ is linear (trivial) and continuous on $\text{Ker}(B)$, with $||F||_{V'} \leq ||f||_{V'} + ||a||\,||u_g||_V$ (easy check). So u_0 exists, is unique, and $||u_0||_V \leq \frac{1}{\alpha}||F||_{V'}$, that is, $||u_0||_V \leq \frac{1}{\alpha}(||f||_{V'} + ||a||\,||u_g||_V) \leq \frac{1}{\alpha}(||f||_{V'} + ||a||\,\frac{1}{\beta}||g||_{Q'})$.

Then let $u := u_0 + u_g$. So $a(u,v) = \langle f, v\rangle_{V',V}$, cf. (12.15), and $u_0 \in \text{Ker}(B)$ and $Bu_g = g$ give $b(u,q) = b(u_0,q) + b(u_g,q) = 0 + \langle g, q\rangle_{Q',Q}$, therefore u is as solution of (12.8).

Moreover u is independent of u_g: If si u'_g also satisfies $Bu'_g = g$, if $u'_0 \in \text{Ker}(B)$ is the associated solution, if $u' = u'_0 + u'_g$, then $u - u' = u_0 - u'_0 + u_g - u'_g \in \text{Ker}(B)$ (since $B(u_g - u'_g) = g - g = 0$) and $a(u - u', v_0) = 0$ for all $v_0 \in \text{Ker}(B)$, thus $u - u' = 0$ (coercitivy of $a(\cdot,\cdot)$ on $\text{Ker}(B)$), and $u = u'$. Thus $u = u_0 + u_g \in V$ exists and is unique.

And $||u||_V \leq ||u_0||_V + ||u_g||_V \leq \frac{1}{\alpha}(||f||_{V'} + ||a||\,\frac{1}{\beta}||g||_{Q'}) + \frac{1}{\beta}||g||_{Q'}$, that is $(12.14)_1$.

Then we look for p solution of $b(v,p) = a(u,v) - \langle f, v\rangle_{V',V}$ for all $v \in V$. Let $L(v) := a(u,v) - \langle f, v\rangle_{V',V}$. So if p exists then $L(v) = b(v,p)$, thus L vanishes on $\text{Ker}(B)$, i.e., $L \in (\text{Ker}(B))^\circ$. And $(\text{Ker}(B))^\circ = \overline{\text{Im}(B^t)}$, so $(\text{Ker}(B))^\circ = \text{Im}(B^t)$ (closed ranged theorem 11.2). Thus there exists $p \in Q$ s.t. $L = B^t p$. And $||B^t p||_{V'} \geq \beta||p||_{Q/\text{Ker}B^t}$, cf. (12.9). Then $(12.14)_2$. ▪

12.5 The saddle point problem

Let $L : V \times Q \to \mathbb{R}$ be defined by

$$\mathcal{L}(v,q) = \frac{1}{2}a(v,\vec{v}) + b(v,q) - (f,v)_{L^2} - (g,q)_{L^2}. \tag{12.16}$$

If $a(\cdot,\cdot)$ is symmetric then L is the Lagrangean bilinear form associated to the mixed problem (12.8). And the associated optimization problem is: Find $(u,p) \in V \times Q$ (saddle point) s.t.

$$\mathcal{L}(u,p) = \inf_{v\in V}(\sup_{q\in Q}\mathcal{L}(v,q)). \tag{12.17}$$

If (u,p) is a solution of (12.17), then, $a(\cdot,\cdot)$ being symmetric,

$$\begin{cases} \forall v \in V, & \dfrac{\partial \mathcal{L}}{\partial v}(u,p).v = \lim\limits_{h\to 0}\dfrac{\mathcal{L}(u+hv,p) - \mathcal{L}(u,,)}{h} = a(u,v) + b(u,q) - \langle f,v\rangle, \\ \forall q \in Q, & \dfrac{\partial \mathcal{L}}{\partial q}(u,p).q = \lim\limits_{h\to 0}\dfrac{\mathcal{L}(u,p+hq) - \mathcal{L}(u,p)}{h} = b(u,q) - \langle g,q\rangle. \end{cases} \tag{12.18}$$

So (u,p) is solution of (12.8).

13 The surjectivites of the divergence operator

Let Ω be an open bounded set in $\mathbb{R}^n$.

Let $b(\cdot,\cdot)$ be defined by $b : \begin{Bmatrix} V \times Q & \to \mathbb{R} \\ (\vec{v},q) & \to b(v,q) = \displaystyle\int_\Omega \mathrm{div}\vec{v}(x)q(x)\,d\Omega \end{Bmatrix}$ where V and Q are appropriate Banach spaces, see below ($b(\cdot,\cdot)$ is bilinear).

Let $B : V \to Q'$ be the associated operator defined by $\langle Bv,q\rangle_{Q',Q} = b(v,q)$, and B will be denoted div (notation of distribution of L. Schwartz).

Then the operator $B^t : Q \to V'$ is defined by $\langle B^t q,v\rangle_{V',V} = \langle Bv,q\rangle_{Q',Q} = b(v,q)$.

The integration by parts, if legitimate, gives

$$b(\vec{v},q) = \langle B\vec{v},q\rangle_{Q',Q} = \langle B^t q,\vec{v}\rangle_{V',V} = -\int_\Omega dq(x).\vec{v}(x)\,d\Omega + \int_\Gamma q(x)\,\vec{v}(x).\vec{n}(x)\,d\Gamma. \tag{13.1}$$

13.1 The divergence operator $\mathrm{div} : H^{\mathrm{div}}(\Omega) \to L^2(\Omega)$ is surjective

Here $b(\vec{v},q) = (\mathrm{div}\vec{v},q)_{L^2}$.

Theorem 13.1 *The linear mapping* $\mathrm{div} : \begin{Bmatrix} H^{\mathrm{div}}(\Omega) & \to L^2(\Omega) \\ \vec{v} & \to \mathrm{div}\vec{v}, \end{Bmatrix}$ *is continuous and surjective. And the open mapping theorem gives, cf. (8.16),*

$$\exists\beta > 0,\ \forall\vec{v} \in H^{\mathrm{div}}(\Omega),\quad ||\mathrm{div}(\vec{v})||_{L^2(\Omega)} \ge \beta||\vec{v}||_{H^{\mathrm{div}}/\mathrm{Ker(div)}}, \tag{13.2}$$

or

$$\exists\beta > 0,\ \forall\vec{v} \in H^{\mathrm{div}}(\Omega),\ \exists p \in L^2(\Omega),\quad (\mathrm{div}(\vec{v}),p)_{L^2(\Omega)} \ge \beta||\vec{v}||_{H^{\mathrm{div}}/\mathrm{Ker(div)}}||p||_{L^2(\Omega)}, \tag{13.3}$$

also written as the inf-sup inequality $\exists\beta > 0,\ \inf\limits_{v\in H^{\mathrm{div}}(\Omega)}\ \sup\limits_{p\in L^2(\Omega)} \dfrac{|(\mathrm{div}\vec{v},p)_{L^2}|}{||\vec{v}||_{H^{\mathrm{div}}/\mathrm{Ker(div)}}||p||_{L^2}} \ge \beta.$

Proof. Since $||\mathrm{div}\vec{v}||_{L^2} \le ||\vec{v}||_{H^{\mathrm{div}}}$, div is continuous. Let $f \in L^2(\Omega)$. Let $p \in H_0^1(\Omega)$ be the solution of $\Delta p = f$ (Lax–Milgram theorem). So $\mathrm{div}(\vec{\mathrm{grad}}p) = f \in L^2(\Omega)$, thus $\vec{\mathrm{grad}}p \in H^{\mathrm{div}}(\Omega)$; Then let $\vec{v} = \vec{\mathrm{grad}}p \in H^{\mathrm{div}}(\Omega)$. Thus $\mathrm{div}\vec{v} = f$, and div is surjective.

And (12.11) gives (13.2), thus (13.3). ∎

13.2 The divergence operator $\text{div} : H_0^{\text{div}}(\Omega) \to L_0^2(\Omega)$ is surjective

Here $b(\vec{v}, q) = (\text{div}\vec{v}, q)_{L^2}$.

Theorem 13.2 *The linear mapping* $\text{div} : \left\{ \begin{array}{ll} H_0^{\text{div}}(\Omega) & \to L_0^2(\Omega) \\ \vec{v} & \to \text{div}\vec{v}, \end{array} \right\}$ *is continuous and surjective. And the open mapping theorem gives, cf. (8.16),*

$$\exists \beta > 0,\ \forall \vec{v} \in H_0^{\text{div}}(\Omega), \quad ||\text{div}(\vec{v})||_{L^2(\Omega)} \geq \beta ||\vec{v}||_{H_0^{\text{div}}/\text{Ker}(\text{div})}, \tag{13.4}$$

also written as the inf-sup inequality $\exists \beta > 0, \quad \inf_{p \in L_0^2(\Omega)} \sup_{\vec{v} \in H_0^{\text{div}}(\Omega)} \dfrac{|(\text{div}\vec{v}, p)_{L^2}|}{||\vec{v}||_{H_0^{\text{div}}/\text{Ker}(\text{div})} ||p||_{L_0^2}} \geq \beta.$

Proof. Since $||\text{div}\vec{v}||_{L^2} \leq ||\vec{v}||_{H_0^{\text{div}}(\Omega)}$, div is continuous. Let $f \in L_0^2(\Omega)$. Let $p \in H^1(\Omega)/\mathbb{R}$ be the solution of $(\vec{\text{grad}}p, \vec{\text{grad}}q)_{L^2} = (f, q)_{L^2}$ for all $q \in H^1(\Omega)/\mathbb{R}$, cf. the Lax–Milgram Theorem in $H^1(\Omega)/\mathbb{R}$ (the hypothesis $f \in L_0^2(\Omega)$, that is $(f, 1_\Omega)_{L^2} = 0$ $(= (\vec{\text{grad}}p, \vec{\text{grad}}1_\Omega)_{L^2})$, is mandatory and is called the compatibility condition). And $(\vec{\text{grad}}p, \vec{\text{grad}}q)_{L^2} = (f, q)_{L^2}$ for all $q \in H^1(\Omega)/\mathbb{R}$ gives $(\vec{\text{grad}}p.\vec{n})_{|\Gamma} = 0$. Thus with $\vec{v} = \vec{\text{grad}}p$, we have $\vec{v} \in H_0^{\text{div}}(\Omega)$ and $\text{div}(\vec{\text{grad}}p) = f \in L^2(\Omega)$, so div is surjective from $H_0^{\text{div}}(\Omega)$ to $L^2(\Omega)$. So we get (13.4), cf. (12.11). ∎

13.3 The divergence operator $\text{div} : L^2(\Omega)^n \to H^{-1}(\Omega)$ is surjective

Here $b(\vec{v}, q) = \langle \text{div}\vec{v}, q \rangle_{H^{-1}, H_0^1} = -\langle \vec{v}, dq \rangle_{L^2(\Omega), L^2(\Omega)} := -\int_\Omega dq(x).\vec{v}(x)\, d\Omega$ (distributions of L. Schwartz) for all $q \in H_0^1(\Omega)$.

Theorem 13.3 *The linear mapping* $\text{div} : \left\{ \begin{array}{ll} L^2(\Omega)^n & \to H^{-1}(\Omega) \\ \vec{v} & \to \text{div}\vec{v}, \end{array} \right\}$ *is continuous and surjective. And the open mapping theorem gives, cf. (8.16),*

$$\exists \beta > 0,\ \forall \vec{v} \in L^2(\Omega), \quad ||\text{div}(\vec{v})||_{H^{-1}} \geq \beta ||\vec{v}||_{L^2(\Omega)/\text{Ker}(\text{div})}, \tag{13.5}$$

also written as the inf-sup inequality $\exists \beta > 0, \quad \inf_{p \in H_0^1(\Omega)} \sup_{\vec{v} \in L^2(\Omega)} \dfrac{|b(\vec{v}, q)|}{||\vec{v}||_{L^2(\Omega)/\text{Ker}(\text{div})} ||p||_{H_0^1}} \geq \beta.$

Proof. $||\text{div}\vec{u}||_{H^{-1}} = \sup_{\phi \in H_0^1(\Omega)} \frac{|\langle \text{div}\vec{u}, \phi| \rangle}{||\phi||_{H_0^1}} = \sup_{\phi \in H_0^1(\Omega)} \frac{|(\vec{u}, \vec{\text{grad}}\phi)_{L^2}}{||\phi||_{H_0^1}} \leq ||\vec{u}||_{L^2}$ (Cauchy–Schwarz in $L^2(\Omega)$), therefore div is continuous. Let $\ell \in H^{-1}(\Omega)$. Thus there exists $f \in L^2(\Omega)$ and $\vec{u} \in L^2(\Omega)^n$ s.t. $\ell = f + \text{div}\vec{u}$, cf. (9.24). Let $\vec{w} \in H^{\text{div}}(\Omega)$ s.t. $\text{div}\vec{w} = f$, cf. thm. 13.1. So $\ell = \text{div}(\vec{w} + \vec{u})$ with $\vec{u} + \vec{w} \in L^2(\Omega)^n$, and div is continuous. So we get (13.5), cf. (12.11). ∎

13.4 The divergence operator $\text{div} : H_0^1(\Omega)^n \to L_0^2(\Omega)$ is surjective

Here $b(\vec{v}, q) = (\text{div}\vec{v}, q)_{L^2}$.

Theorem 13.4 *The linear mapping*

$$\text{div} : \left\{ \begin{array}{ll} H_0^1(\Omega)^n & \to L_0^2(\Omega) \\ \vec{v} & \to \text{div}\vec{v}, \end{array} \right\} \quad \textit{is continuous and surjective.} \tag{13.6}$$

And the open mapping theorem gives, cf. (8.16),

$$\exists \beta > 0,\ \forall \vec{v} \in H_0^1(\Omega)^n, \quad ||\text{div}(\vec{v})||_{L_0^2} \geq \beta ||\vec{v}||_{H_0^1(\Omega)^n/\text{Ker}(\text{div})}, \tag{13.7}$$

also written as the inf-sup inequality $\exists \beta > 0, \quad \inf_{p \in L_0^2(\Omega)} \sup_{\vec{v} \in H_0^1(\Omega)^n} \dfrac{|(\text{div}\vec{v}, q)_{L^2}|}{||\vec{v}||_{H_0^1/\text{Ker}(\text{div})} ||p||_{L_0^2)L^2}} \geq \beta.$

Proof. div $= \vec{\text{grad}}' : H_0^1(\Omega) \to L_0^2(\Omega)$ is the dual operator of the gradient operator $\vec{\text{grad}} : L^2(\Omega) \to H^{-1}(\Omega)^n$. Since the range of the $\vec{\text{grad}}$ is closed, cf. theorem 10.1, the range of the div operator is closed, cf. the closed range theorem 11.2. (Remark: For any $\vec{v} \in H_0^1(\Omega)^n$ we have $\int_\Omega \text{div}\vec{v}\, d\Omega = \int_\Gamma \vec{v}.\vec{n}\, d\Gamma = 0$, so $\text{Im}(\text{div}) \subset L_0^2(\Omega)$.) ▪

With div $: H_0^1(\Omega)^n \to L^2(\Omega)$ we have $\text{Ker}(\text{div}) = \{\vec{v} \in H_0^1(\Omega)^n : \text{div}\vec{v} = 0\}$, and

$$\text{Ker}(\text{div})^{\perp_{H_0^1}} := \{\vec{v} \in H_0^1(\Omega)^n : (\vec{v}, \vec{w})_{H_0^1} = 0,\ \forall \vec{w} \in H_0^1(\Omega)^n,\ \text{div}\vec{w} = 0\}. \tag{13.8}$$

Let $\Delta^{-1} : \left\{ \begin{array}{rl} H^{-1}(\Omega) & \to H_0^1(\Omega) \\ f & \to u = \Delta^{-1} f \end{array} \right\}$, that is, $u \in H_0^1(\Omega)$ solves the Dirichlet problem $\Delta u = f$.

Corollary 13.5

$$\text{Ker}(\text{div})^{\perp_{H_0^1}} = \{\vec{v} = \Delta^{-1}(\vec{\text{grad}} q),\ q \in L^2(\Omega)\} \quad (= \Delta^{-1}(\vec{\text{grad}}(L^2(\Omega)))), \tag{13.9}$$

that is $\vec{v} \in \text{Ker}(\text{div})^{\perp_{H_0^1}}$ iff $\Delta\vec{v}$ derives from a potential $q \in L^2(\Omega)$.

And $H_0^1(\Omega)^n = \text{Ker}(\text{div}) \oplus^{\perp_{H_0^1}} \text{Ker}(\text{div})^{\perp_{H_0^1}}$ give a decomposition of $H_0^1(\Omega)^n$.

Proof. Let $A := \{\vec{v} \in H_0^1(\Omega)^n : \vec{v} = \Delta^{-1}(\vec{\text{grad}} q),\ q \in L^2(\Omega)\}$. So $\vec{v} \in A$ iff $\vec{v} \in H_0^1(\Omega)^n$ and $\exists q \in L^2(\Omega)$, $\Delta\vec{v} = \vec{\text{grad}} q$, i.e. $(\vec{v}, \vec{w})_{H_0^1} = (\text{grad}\vec{v}, \text{grad}\vec{w})_{L^2} = (q, \text{div}\vec{w})_{L^2}$ for all $\vec{w} \in H_0^1(\Omega)^n$.

• $A \subset \text{Ker}(\text{div})^{\perp_{H_0^1}}$: Let $\vec{v} \in A$. Thus $\exists q \in L^2(\Omega)$ s.t. $(\vec{v}, \vec{w})_{H_0^1} = (q, \text{div}\vec{w})_{L^2}$ for all $\vec{w} \in H_0^1(\Omega)^n$. Thus $(\vec{v}, \vec{w})_{H_0^1} = 0$ for all $\vec{w} \in \text{Ker}(\text{div})$, thus $\vec{v} \in \text{Ker}(\text{div})^{\perp_{H_0^1}}$.

• $\text{Ker}(\text{div})^{\perp_{H_0^1}} \subset A$: Let $\vec{v} \in \text{Ker}(\text{div})^{\perp_{H_0^1}}$. We look for $q \in L_0^2(\Omega)$ s.t. $\Delta\vec{v} = \vec{\text{grad}} q$: thus we look for $q \in L_0^2(\Omega)$ s.t. $\Delta\vec{v} = \vec{\text{grad}} q$, that is $(q, \text{div}\vec{z})_{L^2} = -(\Delta\vec{v}, \vec{z})_{H^{-1}, H_0^1}$ for all $\vec{z} \in H_0^1(\Omega)$.

The operator $B = \text{div} : \vec{z} \in H_0^1(\Omega)/\text{Ker}(\text{div}) \to \text{div}\vec{z} \in L_0^2(\Omega)$ is linear continuous bijective, with $||\text{div}\vec{z}||_{L_0^2(\Omega)} \le ||B||\, ||\vec{z}||_{H_0^1(\Omega)/\text{Ker}(\text{div})}$.

Its inverse $B^{-1} : \psi \in L_0^2(\Omega) \to B^{-1}\psi \in H_0^1(\Omega)/\text{Ker}(\text{div})$ is linear continuous bijective, with $||B^{-1}\psi||_{H_0^1(\Omega)/\text{Ker}(\text{div})} \le ||B^{-1}||\, ||\psi||_{L_0^2}$.

Let $a(\cdot,\cdot) : (q, \psi) \in L_0^2(\Omega) \times L_0^2(\Omega) \to a(q, \psi) = (q, \psi)_{L^2}$: $a(\cdot,\cdot)$ is trivially bilinear continuous coercive in $(L^2(\Omega), (\cdot,\cdot)_{L^2})$.

Let $\ell : \psi \in L_0^2(\Omega) \to \ell(\psi) = -(\Delta\vec{v}, B^{-1}\psi)_{H^{-1}, H_0^1} = (\text{grad}\vec{v}, \text{grad}\psi)_{L^2} \in \mathbb{R}$: ℓ is trivially linear, and is continuous since $|\ell(\psi)| \le ||\Delta\vec{v}||_{H^{-1}} ||B^{-1}\psi||_{L^2} \le ||\Delta\vec{v}||_{H^{-1}} ||B^{-1}||\, ||\psi||_{L_0^2}$.

Thus the Lax-Milgram theorem gives the existence of q. And the first point shows that if $\Delta\vec{v} = \vec{\text{grad}} q$ then $\vec{v} \perp_{H_0^1} \text{Ker}(\text{div})$. ▪

A Singular value decomposition (SVD)

We want to estimate the β inf-sup constant, cf. (12.12). Consider a $m*n$ rectangular matrix B. We look for its singular value σ_i, that is, we look for a $m*n$ "diagonal" matrix σ, i.e. e.g. in the case $m<n$,

$$\Sigma = \text{diag}_{m,n}(\sigma_1, ...\sigma_p) = \begin{pmatrix} \sigma_1 & 0 & ... & & 0 & 0 & \dots & 0 \\ 0 & \sigma_1 & 0 & & 0 & \vdots & & \vdots \\ \vdots & & \ddots & \ddots & & & & \\ & & & & 0 & & & \\ 0 & & \dots & & \sigma_p & 0 & \dots & 0 \end{pmatrix},$$

and for two matrices U $m*m$ and V $n*n$ s.t.

$$\Sigma = U^T.B.V,$$

with U^T the U transposed matrix.

Proposition A.1 *Let B be a $m*n$ real matrix. If λ_i is an eigenvalue of the $n*n$ matrix $B^T.B$, then λ_i is positive and is an eigenvalue of the $m*m$ matrix $B.B^T$.*

If λ_i is an eigenvalue of the $m*m$ matrix $B.B^T$, then λ_i is positive and is an eigenvalue of the $n*n$ matrix $B^T.B$.

Let $\sigma_i = \sqrt{\lambda_i}$. Let $(\vec{v}_i)_{1,...,n}$ be an orthonormal basis of eigenvectors of $B^T.B$ associated to the eigenvalues λ_i, and let V be the (orthonormal) matrix whose j-th column is $\vec{v}_j$. Let $(\vec{u}_i)_{1,...,n}$ be an orthonormal basis of eigenvectors of $B.B^T$ associated to the eigenvalues λ_i, and let U be the (orthonormal) matrix whose j-th column is $\vec{u}_j$. And let $\Sigma = \text{diag}_{m,n}(\sigma_1, ..., \sigma_p)$ where $p = \min(m,n)$. Then the singular value decomposition of B is

$$\Sigma = U^T.B.V, \quad \text{i.e.} \quad B = U.\Sigma^T.V^T. \tag{A.1}$$

Thus, if $\text{rank}(B) = r$ and $\sigma_1 \geq ... \geq \sigma_r > 0$ (and $\sigma_i = 0$ pour $i > r$), then

$$B = \sum_{i=1}^{r} \sigma_i \, \vec{u}_i.\vec{v}_i^T. \tag{A.2}$$

Moreover, $\begin{pmatrix} \vec{u}_j \\ \vec{v}_j \end{pmatrix} \in \mathbb{R}^{m+n}$, $j = 1, ..., p$, is an eigenvector of $\begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$ associated to the eigenvalue σ_j.

Proof. $B^T.B$ is symmetric real, thus diagonalisable. Moreover $B^T.B$ is non negative since $\vec{x}^T.(B^T.B).\vec{x} = (B.\vec{x})^T.(B.\vec{x}) = ||B.\vec{x}||^2 \geq 0$. Let $\lambda_1, ..., \lambda_n$ be its eigenvalues, and $\lambda_1 \geq ... \geq \lambda_n (\geq 0)$, even if you have to renumber them. Let $\vec{v}_i$ be associated eigenvectors constituting an orthonormal basis in $\mathbb{R}^n$, and let V be the orthonormal matrix which columns are made of the $\vec{v}_i$'s. Suppose $(B) \leq p = \min(m,n)$, so that $\text{rank}(B^T.B) \leq p$ and $\lambda_{p+1} = ... = \lambda_n = 0$. Then

$$\text{diag}_{n,n}(\lambda_1, ..., \lambda_p, 0, ..., 0) = V^T.B^T.B.V \quad n*n \text{ matrix.}$$

Same steps for $B.B^T$ with eigenvalues $\mu_1 \geq .. \geq \mu_m \geq 0$ and the associated orthonormal matrix U:

$$\text{diag}_{m,m}(\mu_1, ..., \mu_p, 0, ..., 0) = U^T.B.B^T.U \quad m*m \text{ matrix.}$$

With $B.B^T.\vec{u}_i = \mu_i \vec{u}_i$ we get $B^T.B.B^T.\vec{u}_i = \mu_i B^T.\vec{u}_i$, thus $B^T.\vec{u}_i$ is an eigenvector for $B^T.B$ associated to the eigenvalue μ_i. And $B^T.B.\vec{v}_j = \lambda_j \vec{v}_j$ tells that μ_i is one the the λ_j.

Remark: If $\Sigma = \text{diag}_{m,n}(\sigma_1, ..., \sigma_p) = U^T.B.V$ then $\Sigma.\Sigma^T = \text{diag}_{m,m}(\sigma_1^2, ..., \sigma_p^2) = (U^T.B.V).(U^T.B.V)^T = U^T.(B.B^T).U$, and the $\lambda_i = \sigma_i^2$ are indeed the eigenvalues of $B.B^T$ associated to the eigenvectors of U. Idem for $B^T.B$. And the matrices U and V are the matrices made of the column vectors $\vec{u}_j$ and $\vec{v}_j$.

Existence of the decomposition: Let λ_i, $i = 1, ..., n$, be the eigenvalues of $B^T.B$. Suppose $\lambda_1 \geq ... \geq \lambda_r > 0$, and $\lambda_{r+1} = ... = \lambda_n = 0$. Let $\sigma_j = \sqrt{\lambda_j}$.

Then let

$$\vec{u}_j = \frac{B.\vec{v}_j}{\sigma_j} \in \mathbb{R}^m, \quad 1 \leq j \leq r. \tag{A.3}$$

The $\vec{u}_j$ are (orthonormal) eigenvectors of $B.B^T$: Indeed $(B.B^T).\vec{u}_j = \frac{B.(B^T.B).\vec{v}_j}{\sigma_j} = \frac{\lambda_j B.\vec{v}_j}{\sigma_j} = \lambda_j \vec{u}_j$. And $\vec{u}_i^T.\vec{u}_j = \frac{\vec{v}_i^T.(B^T.B\vec{v}_j)}{\sigma_i\sigma_j} = \lambda_j \frac{\vec{v}_i^T.\vec{v}_j}{\sigma_i\sigma_j} = \delta_{ij}$, and the $\vec{u}_j$ are the normalized vectors $B.\vec{v}_j$. We then complete $(\vec{u}_j)_{j=1,..,r}$ to get an orthonormal basis in $\mathbb{R}^m$ (e.g. with Gram–Schmidt method). Let U be the $m*m$ matrix made of the columns vector $\vec{u}_j$.

Let $\Sigma = U^T.B.V$. So $[\Sigma_{ij}] = [\vec{u}_i^T.B.\vec{v}_j] = [\sigma_j \vec{u}_i^T.\vec{u}_j] = [\sigma_j\delta_{ij}]$ if $j \le r$, and vanishes if $j > r$ (since $B.\vec{v}_j = 0$). Thus (A.1). And then (A.2).

And $B\vec{v}_j = \sigma_j\vec{u}_j$ for $j \le r$, cf. (A.3), thus $B^T.B.\vec{v}_j = \sigma_j B^T.\vec{u}_j = \lambda_j\vec{v}_j$ for $j \le r$, so $B^T.\vec{u}_j = \sigma_j\vec{v}_j$ for $j \le r$. And if $j > r$ then $B^T.\vec{u}_j = 0$ since $\vec{u}_j \in (\mathrm{Im}(B))^\perp = \mathrm{Ker}(B^T)$. Thus $\begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix} . \begin{pmatrix} \vec{u}_j \\ \vec{v}_j \end{pmatrix} = \sigma_j \begin{pmatrix} \vec{u}_j \\ \vec{v}_j \end{pmatrix}$.

And if B is a symmetric positive real matrix, then $B^T.B = B^2 = B.B^T$, so $B^T.B = B.B^T$; With $B \ge 0$ thus its eigenvalues are non negative, $\sigma_i = +\sqrt{\lambda_i}$, thus the σ_i are the singular values. ▪

Remark A.2 If $m > n$ and $j \ge m+1$ then the $\vec{u}_j$ are useless, and U is computed as a $m*n$ matrix (and U^T as a $n*m$ matrix). And Σ is then a $n*n$ matrix. This method is called the "Thin SVD". ▪

Corollary A.3 $\mathrm{Rang}(B) = r$, $\mathrm{Ker}(B) = \mathrm{Vect}\{\vec{v}_{r+1}, ...\vec{v}_n\}$ *and* $\mathrm{Im}(B) = \mathrm{Vect}\{\vec{u}_1, ...\vec{u}_r\}$.

Proof. Apply (A.2). ▪

Corollary A.4 *Let* $k \le r-1$ *and* $B_k = \sum_{i=1}^k \sigma_i \vec{u}_i.\vec{v}_i^T$. *Then*

$$\min_{Z:\mathrm{Rang}Z=k} ||B - Z|| = \sigma_{k+1} = ||B - B_k||,$$

where $||Z|| = \sup_{\vec{x}\ne\vec{0}} \frac{||Z.\vec{x}||_{\mathbb{R}^m}}{||\vec{x}||_{\mathbb{R}^n}}$ *is the usual norm.*

This gives a numerical measure of the rank of B*: If* σ_{k+1} *is of the precision order of the computer, then the numerical rank o* B *is* k.

Proof. We get $U^T.B_k.V = \mathrm{diag}_{m,n}(\sigma_1, ..., \sigma_k, 0, ...)$ (easy check).

Thus $U^T.(B - B_k).V = \mathrm{diag}_{m,n}(0, ..., 0, \sigma_{k+1}, ..., \sigma_r, 0, ...)$, thus $||B - B_k|| = \sigma_{k+1}$, and in particular $\min_{Z:\mathrm{Rang}Z=k} ||B - Z|| \le \sigma_{k+1}$.

Let Z be a $m*n$ matrix with rank k. Thus $\dim \mathrm{Ker}Z = n-k$. Let $E = \mathrm{Ker}Z \bigcap \mathrm{Vect}\{\vec{v}_1, ...\vec{v}_{k+1}\}$. So $\dim(E) \ge 1$ (intersection of a dimension $n-k$ space with a dimension $k+1$ space in $\mathbb{R}^n$).

Let $\vec{x} \in E$ s.t. $||\vec{x}||_{\mathbb{R}^n} = 1$; Then $||(B-Z).\vec{x}||^2_{\mathbb{R}^m} = ||B.\vec{x}||^2_{\mathbb{R}^m} = ||\sum_{i=1}^{k+1} \sigma_i(\vec{v}_i^T.\vec{x})\vec{u}_i||^2 = \sum_{i=1}^{k+1} \sigma_i^2(\vec{v}_i^T.\vec{x})^2$, the $\vec{u}_i$ being orthonormal vectors. Thus $||(B-Z).\vec{x}||^2_{\mathbb{R}^m} \ge \sigma_{k+1} \sum_{i=1}^{k+1}(\vec{v}_i^T.\vec{x})^2 = \sigma_{k+1}||\vec{x}||^2 = \sigma_{k+1}$. So $\min_{Z:\mathrm{Rang}Z=k} ||B - Z|| \ge \sigma_{k+1}$. ▪

B Application: The discrete inf-sup condition

Example of $\mathrm{div}\vec{u} = 0$, corresponding to B a rectangular $m*n$ matrix (computation of $b(\vec{v}_h, q_h) = 0$).

Here B^T stands for $[B_h]$, cf. (1.12), and we compute the singular values of B^T, i.e. the eigenvalues of $\begin{pmatrix} 0 & B^T \\ B & 0 \end{pmatrix}$. We get:

Proposition B.1 *Let* $\sigma_r > 0$ *be the smallest positive eigenvalue of* B. *We have (value of the inf-sup constant)*

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

Proof. $B = \sum_{i=1}^r \sigma_i\vec{u}_i.\vec{v}_i^T$ gives $B.\vec{x} = \sum_{i=1}^r \sigma_i(\vec{v}_i^T.\vec{x})\,\vec{u}_i$, so $\vec{y}^T.B.\vec{x} = \sum_{i=1}^r \sigma_i(\vec{v}_i^T.\vec{x})\,(\vec{y}^T.\vec{u}_i)$.

Let $\vec{x} = \sum_{j=1}^n x^j\vec{v}_j$ and $\vec{y} = \sum_{k=1}^n y^k\vec{u}_k$. Thus $\vec{y}^T.B.\vec{x} = \sum_{i=1}^r \sigma_i x^i\, y^i$.

Thus for $||x|| = 1$, and $\vec{y}$ being fixed with $||\vec{y}|| = 1$, the sup is given by $x^i = \frac{\sigma_i y^i}{(\sum_i \sigma_i^2 y_i^2)^{\frac{1}{2}}}$, and gives $\vec{y}^T.B.\vec{x} = \frac{1}{(\sum_i \sigma_i^2 y_i^2)^{\frac{1}{2}}} \sum_{i=1}^r \sigma_i^2 y_i^2 = (\sum_i \sigma_i^2 y_i^2)^{\frac{1}{2}}$. Thus the inf for $\vec{y}$ is given with $\vec{y} = \vec{u}_r$, and gives $\vec{y}^T.B.\vec{x} = \sigma_r$. ▪

References

1 Arnold D., Brezzi F., Fortin M.: *A stable finite element for the Stokes equations.* Calcolo, 21, pp. 337-344 (1984).

2 Babuška I.: *The Finite Element Method with Lagrangian Multipliers.* Numer. Math. 20, 179-192 (1973).

3 Barbosa H., Hughes T.: *The Finite Element Method with Lagrange Multipliers on the Boundary: circumventing the Babuška–Brezzi Condition.* Comp. Meths. in Appl. Mech. and Eng., 85, 109–128, 1991.

4 Bercovier M., Pironneau O.: *Error estimates for finite element method solution of the Stokes problem in the primitive variables.* Numer. Math., 33, pp. 211-224 (1977).

5 Brenner S.C., Scott L.R.: *The Mathematical Theory of Finite Element Methods.* Text in Applied Mathematics, Springer-Verlag.

6 Brézis H. : *Analyse fonctionnelle, théorie et applications.* Collection mathématiques appliquées pour la maîtrise, Masson (1983).

7 Brezzi F.: *On the uniqueness, existence and approximation of saddle point problem arising from lagrangian multipliers.* RAIRO Anal. Numer, 8, 129–151 (1974).

8 Brezzi F., Fortin M.: *Mixed and Hybrid Finite Element Methods.* Springer–Verlag, 1991.

9 Brezzi F., Pitkäranta J.: *On the stabilization of finite element approximations of the Stokes equations.* Efficient Solutions of Elliptic Systems, Notes on Numerical Fluid Mechanics, 10, Hackbush ed. (1984).

10 Chapelle D.: *Une formulation mixte de plaques où l'effort tranchant est approché dans son espace naturel.* INRIA, rapport de recherche 2248, novembre 1993.

11 Ciarlet P.G.: *Introduction à l'analyse numérique matricielle et à l'optimisation.* Masson, 1990.

12 Ciarlet P.G.: *The Finite Element Method for Elliptic Problems.* North-Holland (1977).

13 Crouzeix M., Raviart P.A.: *Conforming and non-conforming finite element methods for solving the stationary Stokes equations.* R.A.I.R.O. Anal. Numer., 7, pp. 33-76 (1973).

14 Douglas J. Jr, Wang J.: *An absolutely stabilized finite element method for the Stokes problem.* Math. Comp. 52, pp. 495-508, 1989.

15 Ern A., Guermond J.L.: *Eléments finis : théorie, applications, mise en oeuvre.* Mathématiques & Applications 36, Springer 2002.

16 Fortin M.: *Utilisation de la méthode des éléments finis en mécanique des fluides.* Calcalo, 12, pp. 405-441 (1975).

17 Fortin M.: *An analysis of the convergence of mixed finite element methods.* RAIRO analyse numérique, 11, n°4, pp. 341-354 (1977).

18 Girault, V., Raviart, P. A.: *Finite Element Methods for Navier-Stokes Equations.* Springer-Verlag (1986).

19 Golub H.G., Van Loan C.F.: *Matrix Computations.* John Hopkins, 3rd edition, 1996.

20 Goulaouic C.: *Analyse fonctionnelle et calcul différentiel.* Ed. de l'Ecole Polytechnique, 1982.

21 Goulaouic C., Meyer Y.: *Analyse fonctionnelle et calcul différentiel.* Ed. de l'Ecole Polytechnique, 1984.

22 Hughes T.J.R., Franca L.P., Balestra M.: *A new finite element formulation for computational fluid dynamics: V. Circumventing the Babuška–Brezzi condition: A stable Petrov–Galerkin formulation of the Stokes problem accomodating equal-order interpolations.* C.M.A.M.E. 59, pp. 85-99, 1986.

23 Leborgne G.: *An Optimally Consistent Stabilization of the Inf-Sup Condition.* Numer. Math., 91, pp. 35-56 (2002).

24 Lions J.L., Magenes E.: *Problèmes aux limites non homogènes et applications, Vol 1.* Dunod (1968).

25 Nédélec J.C. : *Méthode des éléments finis.* Mathématiques et applications n°7. Ellipses (1991).

26 Nitsche J.: *Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind.* Abh. Math. Sem. Univ. Hamburg, 36:9–15, 1970/1971.

27 Pitkäranta J.: *Boundary Subspaces for the Finite Element Method with Lagrange Multipliers.* Numer. Math. 33, pp. 273–189, 1979.

28 Raviart P.A., Thomas J.M. : *Introduction à l'analyse numérique des équations aux dérivées partielles.* Collection mathématiques appliquées pour la maîtrise, Masson (1983).

29 Schwartz L.: *Méthodes mathématiques pour les sciences physiques.* Collection enseignement des sciences, Hermann (1993).

30 Schweizer B.: www.mathematik.uni-dortmund.de/lsi/schweizer/Preprints/perfectcondrings-preprint.pdf

31 Stenberg R.: *On some techniques for approximating boundary conditions in the finite element method.* Journal of Computational and Applied Mathematics 63 (1995) 139-148. Et http://math.tkk.fi/∼rstenber/publications.htm.

32 Strang G.: *Linear Algebra and its Applications.* Harcourt Brace & Company, 3rd edition (1988).

33 Strang G., Fix J.F.: *An Analysis of the Finite Element Method.* Wellesley-Cambridge Press (1973).

34 Yosida K.: *Fonctional Analysis.* Sixth Edition, Springer–Verlag 1980.