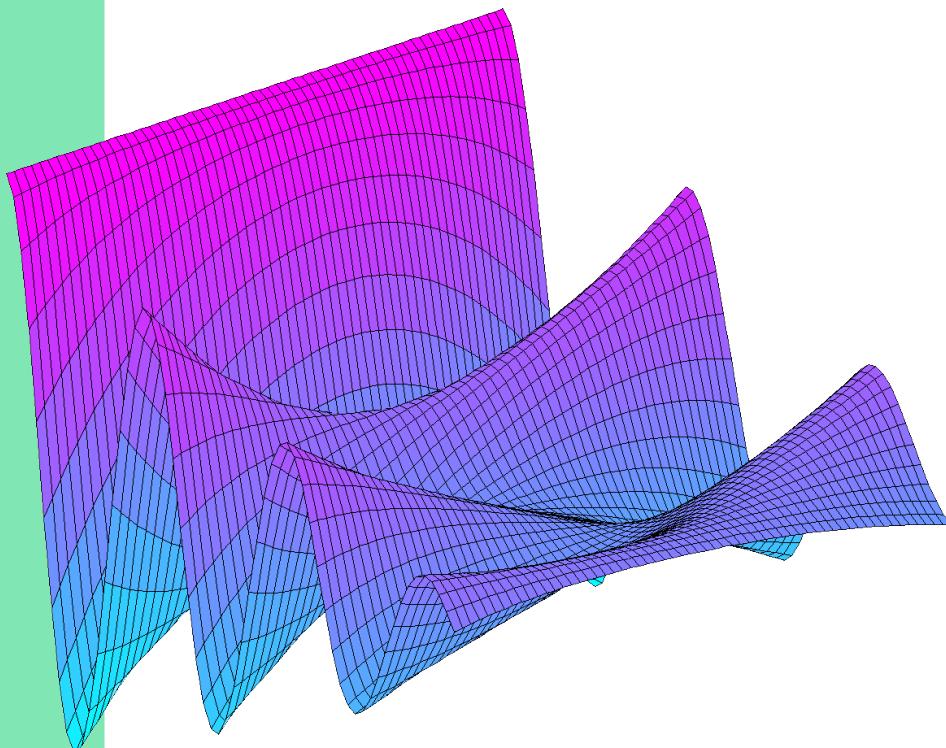


# PREMIÈRE APPROCHE DE LA NON-LINÉARITÉ

Averil PROST  
5<sup>e</sup> année Génie Mathématique

Projet de Fin d'Études  
*sous la direction de*  
Antoine TONNOIR

Mémoire MFA 2021-2022  
*sous la direction de*  
Nicolas FORCADEL



# Table des matières

<b>Introduction</b>	<b>3</b>
<b>1 Équations paraboliques non linéaires monotones</b>	<b>4</b>
1.1 Problèmes bien posés ? . . . . .	4
1.1.1 Choix des espaces fonctionnels . . . . .	5
1.1.2 Existence et unicité . . . . .	6
1.2 Prendre le contrôle . . . . .	10
1.2.1 Existence d'un contrôle optimal . . . . .	11
1.2.2 Différentiabilité . . . . .	12
1.2.3 Système d'optimalité . . . . .	14
<b>2 Système de Navier-Stokes incompressible</b>	<b>18</b>
2.1 Les fameuses équations de Navier-Stokes . . . . .	18
2.1.1 Aperçu de la modélisation . . . . .	19
2.1.2 Formulation variationnelle . . . . .	20
2.1.3 Aperçu du problème bien posé . . . . .	21
2.1.4 Problème adjoint . . . . .	22
2.1.5 Problème inhomogène au bord . . . . .	25
2.2 Du modèle aux schémas . . . . .	27
2.2.1 Une autre formulation . . . . .	27
2.2.2 Discrétisation . . . . .	28
2.2.3 Condition LBB . . . . .	29
2.3 Qualités numériques . . . . .	31
2.3.1 Stabilité . . . . .	31
2.3.2 Convergence . . . . .	33
<b>3 Applications</b>	<b>39</b>
3.1 Implémentation de Navier-Stokes . . . . .	39
3.1.1 Systèmes linéaires . . . . .	39
3.1.2 Validation du code . . . . .	42
3.1.3 Explorations . . . . .	44
3.2 Méthodes numériques pour le contrôle . . . . .	45
3.2.1 Descente de gradient par l'adjoint . . . . .	45
3.2.2 Sequential Quadratic Programming (SQP) . . . . .	50
<b>Conclusion</b>	<b>56</b>
<b>Bibliographie</b>	<b>57</b>

# Introduction

Ce document rend compte du travail effectué dans deux cadres distincts, mais fortement liés.

D'une part, il constitue le rapport de Projet de Fin d'Étude (PFE) mené en GM5, sous la direction de M. Antoine Tonnoir, et dont l'objectif annoncé était d'introduire les équations de Navier-Stokes et leur résolution numérique (avec pour ambition de justifier le vol des oiseaux ou des superhéros). Le chapitre 2 est dédié à l'étude de ce modèle, d'abord dans les espaces continus physiques, puis dans les espaces discrets de la simulation. Le chapitre 3 présente les résultats de l'implémentation. On s'est particulièrement appuyé sur le cours de master pour la recherche de J. F. Scheid [Sch19], dispensé à l'Université de Lorraine.

D'autre part, ce document tient lieu de mémoire de master Mathématiques Fondamentales et Appliquées, sous la direction de M. Nicolas Forcadel, sur le thème du contrôle des équations non linéaires. Le chapitre 1 s'intéresse à une sous-classe de tels problèmes, en suivant l'ouvrage de F. Tröltzsch [Trö10]. Le lien entre ces deux travaux est fait au chapitre 3, où une méthode de contrôle de l'équation de Navier-Stokes est exposée, grâce aux travaux de M. Günzburger [GM00].

La rédaction est volontairement synthétique. Les définitions les plus importantes sont rappelées, ainsi que les énoncés des théorèmes fondateurs. De fréquentes références sont faites pour compléter les démonstrations, et certaines parties sont traitées formellement. L'exposé complet des thématiques abordées a déjà rempli la vie d'éménents chercheurs, et on a - heureusement - échoué à le réduire à 57 pages.

L'auteure remercie chaleureusement les professeurs qui ont bien voulu l'accompagner dans ce travail. En particulier, M. Tonnoir a bien voulu proposer un sujet sur la mécanique des fluides, et l'a encadré avec un enthousiasme communicatif. Merci également à M. Forcadel pour ses précieuses ressources et sa disponibilité, et à Mme Zidani pour nos échanges structurants.

# Chapitre 1

## Équations paraboliques non linéaires monotones

### 1.1 Problèmes bien posés ?

Dans toute cette partie, on se place dans les différents domaines suivants.

**Définition 1 – Domaines** Soit  $\Omega \subset \mathbb{R}^n$  un ouvert que l'on considérera toujours borné et au moins Lipschitzien, et  $T > 0$  un temps final. On notera

- $Q = \Omega \times ]0, T[$  le domaine spatio-temporel,
- $\Sigma = \partial\Omega \times ]0, T[$  la frontière en espace,
- $\Pi = \Omega \times \{0, T\}$  la frontière en temps. On se placera fréquemment sur la composante connexe  $\Omega \times \{0\}$ .

On notera  $(\cdot, \cdot)_Q := (\cdot, \cdot)$  le produit scalaire de  $L^2(Q)$ , et  $(\cdot, \cdot)_\Sigma$  (resp.  $(\cdot, \cdot)_\Pi$ ) celui de  $L^2(\Sigma)$  (resp.  $L^2(\Pi)$ ).

Notons  $S$  un domaine quelconque dans  $\{Q, \Sigma\}$ .

**Définition 2 – Opérateur de Nemytskii** Soit  $\xi : S \times \mathbb{R} \rightarrow \mathbb{R}$ , et  $\mathcal{U}$  un espace de fonctions de  $S \rightarrow \mathbb{R}$ . La composée

$$\begin{aligned}\Xi : S \times \mathcal{U} &\rightarrow \mathbb{R} \\ (x, u) &\mapsto \xi(x, u(x))\end{aligned}$$

est appelée opérateur de superposition, ou opérateur de Nemytskii, associé à  $\xi$ .

En particulier, un tel opérateur ne contient pas de dérivées ni d'intégrales de  $u$ , ou de termes non locaux. Par la suite, on adoptera la même notation pour désigner l'opérateur  $\Xi$  et sa fonction génératrice  $\xi$ . Soit alors  $d$  associé au domaine  $Q$ , et  $b$  associé au bord  $\Sigma$  de tels opérateurs.

**Définition 3 – Problème de référence** Soient  $f$ ,  $g$  et  $u_0$  trois fonctions données. Nous cherchons  $u$  solution du problème semilinéaire parabolique

$$\partial_t u(x, t) - \Delta u(x, t) + d(x, t, u(x, t)) = f(x, t) \quad (x, t) \text{ dans } Q \tag{1.1a}$$

$$\partial_\nu u(\sigma, t) + b(\sigma, t, u(\sigma, t)) = g(\sigma, t) \quad (\sigma, t) \text{ dans } \Sigma \tag{1.1b}$$

$$u(x, 0) = u_0(x) \quad x \text{ dans } \Omega \times \{0\} \tag{1.1c}$$

**Remarque 1** L'opérateur  $\Delta$  peut être remplacé par  $\sum_{i=1}^n \sum_{j=1}^n \frac{\partial}{\partial x_i} \left( a_{i,j} \frac{\partial u}{\partial x_j} \right)$ , où les fonctions  $a_{i,j} \in L^\infty(\Omega)$  sont telles que la matrice  $\mathcal{A} = (a_{i,j})_{i,j}$  soit presque partout symétrique définie positive. Nous n'aurons besoin que du cas où  $a_{i,j} \equiv \nu \delta_{i,j}$ , où  $\nu > 0$  : le seul changement à apporter aux démonstrations est de multiplier (1.1b) par  $\nu$  lors des utilisations de la formule de Green. Par la suite, on se restreint (en perdant un peu de généralité) au Laplacien.

**Remarque 2** Par la suite, on désignera indifféremment  $u : Q \rightarrow \mathbb{R}$  la fonction à deux variables, et  $u : [0, T] \rightarrow F(\Omega, \mathbb{R})$  la fonction à valeurs dans un espace de fonctions. On confondra également leurs évaluations  $u(x, t) = u(t)(x)$  pour presque tout  $x, t$ . Dans les espaces que nous utiliserons, les deux objets peuvent être identifiés (voir [Dre12], section 7.2). De même, pour alléger les écritures, on désignera par  $d(t, u(t))$  la fonction  $x \rightarrow d(t, u(t))(x) = d(x, t, u(x, t))$ .

Nous commençons par établir formellement une formulation variationnelle de (1.1), qui va nous conduire à choisir les espaces fonctionnels de nos variables, ainsi que le sens à donner aux différents opérateurs différentiels.

### 1.1.1 Choix des espaces fonctionnels

Supposons dans un premier temps que toutes nos données ont la régularité demandée pour justifier les opérations. En particulier, on s'autorise à prendre le produit scalaire  $L^2(\Omega)$  (noté  $(\cdot, \cdot)$ ) de (1.1a) avec une fonction test  $v : \Omega \rightarrow \mathbb{R}$ . Appliquons naïvement le théorème de Green :

$$\begin{aligned} -(\Delta u(t), v) &= -(\nabla u(t) \cdot n, v)_\Sigma + (\nabla u, \nabla v) = (b(t, u(t)) - g(t), v)_\Sigma + (\nabla u, \nabla v) \\ (\partial_t u(t), v) + (\nabla u, \nabla v) + (b(t, u(t)), v)_\Sigma + (d(t, u(t)), v) &= (f(t), v) + (g(t), v)_\Sigma \end{aligned}$$

et nous pouvons déjà préciser les exigences de cette formulation.

**Définition 4** On notera

$$V = H^1(\Omega), \quad H = L^2(\Omega), \quad W(0, T) = \{u \in L^2(0, T; V) \mid \partial_t u \in L^2(0, T : V') \text{ au sens de } \mathscr{D}([0, T])\}$$

Le dernier espace sera fondamental, et il convient de le préciser.  $W(0, T)$  est l'espace maximal qui donne un sens à  $(\partial_t u(t), v)$ , de la manière suivante : la fonction  $t \rightarrow \partial_t(u(t), v)$  est un élément de  $\mathscr{D}'([0, T])$ , défini de manière unique par

$$\langle \partial_t(u(t), v), \phi \rangle = -\langle (u(t), v), \partial_t \phi \rangle \quad \forall \phi \in \mathscr{D}([0, T])$$

La seconde partie de la définition restreint  $\partial_t u$  à  $L^2(0, T; V')$ . Comme  $[0, T]$  est compact,  $1 \in L^2(0, T)$ , et on pourra par exemple utiliser  $\langle (\partial_t u, v), 1 \rangle_{L^2(0, T)}$  pour établir des égalités d'énergie. De plus,  $W(0, T)$  est un espace de Hilbert muni de la norme

$$\|u\|_{W(0, T)} = \left( \|u\|_{L^2(0, T; V)}^2 + \|\partial_t u\|_{L^2(0, T; V')}^2 \right)^{1/2}$$

On s'appuie sur [Trö10] pour énoncer l'utile résultat suivant :

**Proposition 1** L'espace  $W(0, T)$  s'injecte compactement dans  $C([0, T], H)$ , au sens où la classe d'équivalence de  $u \in W(0, T)$  contient un élément de  $C([0, T], H)$ .

Cette injection permet de définir correctement  $u(0)$ , et place  $u_0$  dans l'espace  $H$ . La formulation variationnelle requiert au moins l'espace  $L^1(0, T; H')$  pour  $f$ , et son analogue pour  $g$  : les futurs raisonnements sur les inégalités d'énergie demandent à ce que l'on se place dans  $L^2(Q)$  et  $L^2(\Sigma)$ . Ces choix sont résumés dans le théorème principal de cette section, que l'on peut maintenant énoncer.

### 1.1.2 Existence et unicité

On fait les hypothèses suivantes sur  $d$  et  $b$  :

**Hypothèse 1** La fonction  $d : Q \times \mathbb{R} \rightarrow \mathbb{R}$  satisfait

1.  $d$  est mesurable par rapport à  $(x, t)$  pour tout  $y$  fixé,
2.  $d$  est uniformément bornée en  $u = 0$  :  $\exists M > 0, \forall x, t \in Q \quad |b(x, t, 0)| \leq M$ ,
3.  $d$  est globalement lipschitzienne :  $\exists L > 0, \forall (u, v) \in \mathbb{R}^2 \quad |b(x, t, u) - b(x, t, v)| \leq L|u - v| \text{ p.p. } (x, t) \in Q$ ,
4.  $d$  est monotone croissante : p.p.  $(x, t \in Q), \forall (u, v) \in \mathbb{R}^2, u \leq v \implies b(x, t, u) \leq b(x, t, v)$ .

De plus,  $b : \Sigma \times \mathbb{R} \rightarrow \mathbb{R}$  satisfait les mêmes hypothèses relativement à  $\Sigma$ .

**THÉORÈME 1 – EXISTENCE ET UNICITÉ** Supposons l'hypothèse (1) satisfaite. Alors pour tout  $f \in L^2(Q)$ ,  $g \in L^2(\Sigma)$  et  $u_0 \in L^2(\Omega)$ , il existe un unique élément  $u$  de  $W(0, T)$  satisfaisant la formulation variationnelle

$$\begin{aligned} \langle \partial_t u(t), v \rangle + (\nabla u, \nabla v) + (b(t, u(t)), v)_\Sigma + (d(t, u(t)), v) &= (f(t), v) + (g(t), v)_\Sigma \\ u(0) &= u_0 \end{aligned}$$

et cette solution est appelée **solution faible** du problème (1.1). De plus,  $u$  vérifie

$$\|u\|_{W(0, T)} \leq C \left( \|f - d(\cdot, 0)\|_{L^2(Q)} + \|g - b(\cdot, 0)\|_{L^2(\Sigma)} + \|u_0\|_{C(\bar{\Omega})} \right) \quad (1.2)$$

La démonstration de ce théorème suit le plan suivant :

- On se ramène en dimension finie grâce à une suite de Faedo-Galerkin dans une base bien choisie. On peut y appliquer le théorème de Carathéodory.
- Grâce à des estimations uniformes, on exhume une sous-suite faiblement convergente dans un espace de Hilbert.
- La propriété de monotonie permet de montrer que la limite faible satisfait bien le problème, et donne l'unicité.

#### Étape 1 (Suite de Faedo-Galerkin)

Choisissons  $(e_m)_{m \in \mathbb{N}}$  une base hilbertienne de  $V$ . On construit une suite  $(u_m)_{m \in \mathbb{N}}$  en posant  $u_m(t) = \sum_{i=1}^m g_i(t)e_i$ , où les  $g_i \in W^{1,1}(]0, T[)$  sont déterminés par l'équation

$$\begin{aligned} (\partial_t u_m(t), e_j) + (\nabla u_m(t), \nabla e_j) + (d(t, u_m(t)), e_j) + (b(t, u_m(t)), e_j)_\Sigma &= (f(t), e_j) + (g(t), e_j) \\ g(0) &= (u_0, e_j) \end{aligned} \quad (1.3)$$

L'orthogonalité de la base permet d'écrire le problème satisfait par le vecteur  $g := (g_i)_{i \in \llbracket 1, m \rrbracket}$  comme un système d'EDO de la forme  $g'(t) = F_j(t, g)$ , où

$$F_j(t, x) := (f(t), e_j) + (g(t), e_j) - \sum_{i=1}^m x_i (\nabla e_i, \nabla e_j) + \left( d \left( t, \sum_{i=1}^m x_i e_i \right), e_j \right) + \left( b \left( t, \sum_{i=1}^m x_i e_i \right), e_j \right)_\Sigma$$

Énonçons le théorème de Carathéodory, que nous appliquerons immédiatement.

**THÉORÈME 2 – CARATHÉODORY** Supposons que  $F : [0, T] \times \mathbb{R}^m \rightarrow \mathbb{R}^m$  satisfasse

- $t \rightarrow F(t, \xi)$  est mesurable pour tout  $\xi \in \mathbb{R}^m$ ,
- $\xi \rightarrow F(t, \xi)$  est continue pour presque tout  $t \in [0, T]$ ,
- pour toute boule  $B(0, R) \subset \mathbb{R}^m$ , il existe  $h_R \in L^1(0, T)$  telle que  $|F(t, \xi)| \leq h_R(t)$  pour tout  $\xi \in B(0, R)$ .

Alors pour tout  $\xi_0$ , l'équation  $\xi'(t) = F(t, \xi(t))$  admet au moins une solution **locale** absolument continue telle que  $\xi(0) = \xi_0$ . De plus, si toutes les solutions issues d'un même  $\xi_0$  sont uniformément bornées, i.e.  $\exists C > 0$  telle que  $\forall \xi$  solution issue de  $\xi_0$ ,  $|\xi(t)| \leq C$  p.p.  $t \in [0, T]$ , il existe au moins une solution  $\xi$  globale (définie sur tout  $[0, T]$ ).

Une preuve de ce théorème utilisant le point fixe de Schauder peut être trouvée dans [Dre12].

Vérifions les hypothèses. La mesurabilité de  $F$  est héritée de celles de  $f$ ,  $g$ ,  $d$  et  $b$ . Grâce au caractère Lipschitzien de  $d$  et  $b$ ,  $F$  est également continue. Soit  $M > 0$  : on majore

$$\begin{aligned} |F_i(t, \xi)| &\leq |(f(t), e_j)| + |(g(t), e_j)| + \alpha K R + (L_d + L_d) \left| \sum_{i=1}^m x_i e_i \right| + M_d + M_b \\ &\leq |(f(t), e_j)| + |(g(t), e_j)| + \alpha(K + L_d + L_d) \alpha R + M_d + M_b \end{aligned}$$

où  $\alpha$  est la constante d'équivalence de la norme 1 vers la norme euclidienne de  $\mathbb{R}^m$ ,  $K = \max_{i,j} |(\nabla e_i, \nabla e_j)|$  et  $L_{d,b}, M_{d,b}$  sont les constantes de Lipschitz et de majoration uniforme de  $d$  et  $b$ . Comme  $[0, T]$  est compact,  $f \in L^2(0, T; V)$  implique que  $(f(t), e_j) \in L^1(0, T)$  : le même raisonnement pour  $g$  achève de montrer la troisième hypothèse.

Les  $u_m(t)$  existent donc localement. Pour appliquer la deuxième partie du théorème de Carathéodory, nous devons mettre à jour des majorations uniformes en temps. Soit alors  $\xi(\cdot)$  une solution issue de  $\xi_0$ , et  $u_m := \sum_{i=1}^m \xi_i e_i$ . En prenant la combinaison linéaire des problèmes (1.3) correspondant aux coordonnées de  $\xi$ , on forme

$$(\partial_t u_m, u_m) + (\nabla u_m, \nabla u_m) + (d(t, u_m), u_m) + (b(t, u_m), u_m)_\Sigma = (f, u_m) + (g, u_m)_\Sigma$$

En raison de la monotonie, on peut écrire  $(d(t, u_m(t)) - d(t, 0), u_m(t) - 0) \geq 0$ , et l'analogue pour  $b$ , d'où

$$(\partial_t u_m(t), u_m(t)) + |u_m(t)|_1^2 \leq (f(t) - d(t, 0), u_m(t)) + (g(t) - b(t, 0), u_m)_\Sigma \quad (1.4)$$

Tout est intégrable en temps, et on a l'identité  $\int_0^\tau (\partial_t u_m(t), u_m(t)) dt = \frac{1}{2} \int_0^\tau \frac{d}{dt} \|u_m(t)\|^2 dt = \|u_m(\tau)\|^2 - \|u_m(0)\|^2$ . En minorant le membre de droite grâce à la positivité de  $|u_m(t)|$ , on peut écrire

$$\begin{aligned} \|u_m(\tau)\|^2 - \|u_m(0)\|^2 &\leq \int_0^\tau (f(t) - d(t, 0), u_m(t)) + (g(t) - b(t, 0), u_m)_\Sigma dt \\ &\leq \int_0^\tau \|f(t) - d(t, 0)\| \|u_m(t)\| + \|g(t) - b(t, 0)\| \|u_m(t)\|_\Sigma dt \end{aligned}$$

Par continuité de la trace de  $H^1 \subset H^{1/2}$  dans  $L^2(\Sigma)$ , il existe une constante telle que  $\|u_m(t)\|_\Sigma \leq C \|u_m(t)\|$ . Grâce à Hölder, l'inégalité est (enfin !) sous la forme adéquate pour l'application du lemme de Grönwall à la fonction  $\|u_m(\cdot)\|^2$  :

$$\|u_m(\tau)\|^2 \leq \|u_m(0)\|^2 + \frac{1}{2} \int_0^\tau (\|f(t) - d(t, 0)\|^2 + \|g(t) - b(t, 0)\|_\Sigma^2) dt + \frac{1+C}{2} \int_0^\tau \|u_m(t)\|^2 dt$$

et l'on a

$$\|u_m(\tau)\|^2 \leq (\|u_m(0)\|^2 + \frac{1}{2} \|f - d(\cdot, 0)\|_{L^2(0, T; V)}^2 + \|g - b(\cdot, 0)\|_{L^2(0, T; V_\Sigma)}^2) e^{t(1+C)/2}$$

ce qui est borné si  $T$  est fini. Comme on a  $\|u_m(t)\|^2 = \sum_{i=1}^m g_i^2(t)$ , chaque  $g_i$  est borné indépendamment de  $t$ , et le théorème de Carathéodory nous assure que  $u_m$  existe bien globalement sur l'intervalle  $[0, T]$ .

### Étape 2 (extraction d'une sous-suite)

Il se trouve que par Pythagore dans la base hilbertienne  $(e_m)_m$ ,  $\|u_m(0)\|^2 \leq \|u_0\|^2$ . Ainsi, sans aucun effort supplémentaire, nous obtenons une majoration en norme  $C([0, T], H)$  indépendante de  $t$  et de  $m$  :

$$\forall m \in \mathbb{N}, \quad \sup_{t \in [0, T]} \|u_m(t)\| \leq (\|u_0\|^2 + \frac{1}{2} \|f - d(\cdot, 0)\|_{L^2(0, T; V)}^2 + \|g - b(\cdot, 0)\|_{L^2(0, T; V_\Sigma)}^2)^{1/2} e^{T(1+C)/4} := K_1$$

On peut obtenir une autre majoration sans trop d'efforts. Revenons à (1.4), et intégrons encore sur le temps. Le même raisonnement conduit à

$$\|u_m(\tau)\|^2 + \int_0^\tau |u_m(t)|_1^2 dt \leq \|u_m(0)\|^2 + \int_0^\tau \|f(t) - d(t, 0)\| K_1 + \|g(t) - b(t, 0)\| C K_1 dt$$

ce qui amène directement

$$\|u_m\|_{L^2(0, T; V)}^2 \leq \|u_m(0)\|^2 + \int_0^T (\|f(t) - d(t, 0)\| K_1 + \|g(t) - b(t, 0)\| C K_1 + K_1^2) dt \leq K_2$$

Par le théorème de Banach-Alaoglu, toute suite uniformément bornée d'un espace de Hilbert contient une sous-suite faiblement convergente. L'espace de Hilbert naturellement adapté pour notre cas est  $W(0, T)$  : on doit donc établir une estimation de la norme duale  $|\partial_t u_m(t)|_{V'}$ . Soit  $t \in [0, T]$  tel que  $\|f(t)\|$  et  $\|g(t)\|$  soient définies, et  $v \in V$ . On a

$$\begin{aligned} (\partial_t u_m, v) &= (f, v) + (g, v) - ((\nabla u_m, v) + (d(t, u_m), v) + (b(t, u_m), v)_\Sigma) \\ &\leq (\|f(t)\| + \|g(t)\| + K_2 + (M_d + L_{d, K_2}) + (M_b + L_{b, K_2})) \|v\| \end{aligned}$$

d'où  $\|\partial_t u_m(t)\|_{V'} \leq \|f(t)\| + \|g(t)\| + K_2 + (M_d + L_{d, K_2}) + (M_b + L_{b, K_2})$ , et en temps  $T$  fini, la suite  $(u_m)_{m \in \mathbb{N}}$  est uniformément bornée en norme  $W(0, T)$ . Notons  $u \in W(0, T)$  la limite faible d'une sous-suite extraite (que l'on notera également  $(u_m)_m$ ). À ce stade, on sait que cet élément existe, mais il reste à prouver qu'il satisfait le problème variationnel.

### Étape 3 (Satisfaction du problème initial)

La convergence faible  $u_m \rightharpoonup u$  permet de passer à la limite dans tous les termes linéaires continus par rapport à  $u_m$ . Cependant, la limite  $d(x, t, u_m(x, t)) \rightarrow d(x, t, u)$  n'est pas évidente. Une manière de procéder consiste à chercher un espace dans lequel la convergence de  $u_m$  vers  $u$  soit forte, et c'est l'approche que nous évoquerons dans le cas de Navier-Stokes (voir [Lio69], [Dre12]). Pour le cas présent, nous allons nous appuyer sur la monotonie des opérateurs non linéaires.

La suite  $(u_m(t))_m$  étant bornée dans  $V$ , les suites  $(d(\cdot, \cdot, u_m(\cdot)))_m$  et  $(b(\cdot, \cdot, u_m(\cdot)))_m$  sont en particulier bornées dans  $L^2(Q)$  et  $L^2(\Sigma)$  (ces deux ensembles étant de mesure finie). Par Banach-Alaoglu, il existe des éléments  $D \in L^2(Q)$  et  $B \in L^2(\Sigma)$  tels que, quitte à prendre une sous-suite,

$$d(\cdot, \cdot, u_m(\cdot)) \rightharpoonup D \text{ dans } L^2(Q), \quad b(\cdot, \cdot, u_m(\cdot)) \rightharpoonup B \text{ dans } L^2(\Sigma)$$

Nous cherchons à établir que  $D = d(\cdot, \cdot, u(\cdot))$ , et  $B = b(\cdot, \cdot, u(\cdot))$ . Pour cela, nous introduisons un résultat sur les opérateurs monotones, extrait de [Zei89].

**Lemme 1** Soit  $X$  un espace de Banach réel réflexif. Soit  $A : X \rightarrow X'$  un opérateur

- semicontinu, i.e. tel que l'application  $t \rightarrow (A(u + tv), w)$  soit continue pour  $t \in [0, 1]$ ,  $\forall u, v, w \in X$ ,

- monotone, i.e. tel que  $\langle Au - Av, u - v \rangle \geq 0 \forall u, v \in X$ .

Alors les conditions suivantes

$$u_m \rightharpoonup u \text{ dans } X, \quad Au_m \rightarrow b \text{ dans } X', \quad \lim_{m \rightarrow \infty} \langle Au_m, u_m \rangle \leq \langle b, u \rangle$$

impliquent  $Au = b$ .

La preuve est d'une charmante simplicité, et mérite d'être reproduite avant application.

### Démonstration

Par la monotonie de  $A$ , pour tout  $v \in X$ ,

$$\langle Au_m, u_m \rangle - \langle Av, u_m \rangle - \langle Au_m - Av, v \rangle = \langle Au_m - Av, u_m - v \rangle \geq 0$$

Par hypothèse, la limite  $m \rightarrow \infty$  amène

$$\langle b, u \rangle - \langle Av, u \rangle - \langle b - Av, v \rangle = \langle b - Av, u - v \rangle \geq 0$$

Or,  $A$  est semicontinu : posons  $v = u - tw$  pour  $t > 0$ . Il vient

$$t \langle b - A(u - tw), w \rangle \geq 0 \implies \langle b - Au, w \rangle + t \langle Aw, w \rangle \geq 0$$

et par limite en  $t \downarrow 0$ ,  $\langle b - Au, w \rangle \geq 0 \forall w \in X$ , ce qui, par réflexivité, est équivalent à  $Au = b$ .  $\square$

Construisons un opérateur  $A : W(0, T) \rightarrow (W(0, T))'$  qui satisfasse la monotonie et semicontinuité demandées : l'idée directrice est d'exploiter l'égalité satisfaite par  $u_m$  pour travailler avec  $f$ ,  $g$  et  $u_0$  plutôt qu'avec les opérateurs non linéaires. Définissons

$$\begin{aligned} \langle Au_m, v \rangle &:= \int_0^T (d(t, u_m(t)), v(t)) + (b(t, u_m(t)), v(t))_\Sigma + (\nabla u_m(t), \nabla v(t))_\Sigma dt \\ \langle w, v \rangle &:= \int_0^T (D(t), v(t)) + (B(t), v(t))_\Sigma + (\nabla u(t), \nabla v(t))_\Sigma dt \end{aligned}$$

L'opérateur  $A$  hérite directement sa monotonie et sa continuité de celles de  $d$ ,  $b$  et de la semi-norme  $H^1(\Omega)$ . La formulation variationnelle nous ramène à

$$\begin{aligned} \langle Au_m, u_m \rangle &= \int_0^T (f(t), u_m(t)) + (g(t), u_m(t))_\Sigma dt + \frac{1}{2} (\|u_m(0)\|^2 - \|u_m(T)\|^2) \\ \langle w, u \rangle &= \int_0^T (f(t), v(t)) + (g(t), v(t))_\Sigma dt + \frac{1}{2} (\|u(0)\|^2 - \|u(T)\|^2) \end{aligned}$$

où tout est bien défini grâce à l'injection  $W(0, T) \hookrightarrow C(0, T; H)$ . Remarquons que  $W(0, T) \subset L^2(Q) \equiv (L^2(Q))' \subset (W(0, T))'$ , donc la convergence faible de  $u_m$  implique la convergence de  $(f(t), u_m(t))$  vers  $(f(t), u(t))$ , et l'analogique pour  $b$ . La définition de  $u_m(0) = \sum_{i=1}^m (u_0, e_i) e_i$  implique la convergence forte  $u_m(0) \rightarrow u_0$  dans  $H$ , donc  $\|u_m(0)\|^2 \rightarrow \|u_0\|^2$ .

Le seul terme délicat est  $\|u_m(T)\|^2$ . La projection  $u_m \rightarrow u_m(T)$  est linéaire continue, donc  $u_m(T) \rightharpoonup u(T)$ . De plus, l'adhérence faible de la sphère unité est la boule unité, donc  $\|u(T)\| \leq \liminf_{m \rightarrow \infty} \|u_m(T)\|$ . Enfin, la fonction  $x \rightarrow -x^2$  est décroissante, d'où  $\lim_{m \rightarrow \infty} -\|u_m(T)\|^2 \leq -\|u(T)\|^2$ , ce qui achève de vérifier la troisième hypothèse.

On conclut que  $Au = w$  dans  $(W(0, T))'$ , ce qui montre exactement que  $w$  est solution faible.

#### Étape 4 (Unicité)

La monotonie permet d'appliquer le même type de raisonnement que dans le cas linéaire. Soient  $u_1$  et  $u_2$  deux solutions présumées de (3). Leur différence  $w := u_1 - u_2$  satisfait le problème

$$\begin{aligned} \langle \partial_t w(t), v(t) \rangle + (\nabla w(t), \nabla v(t)) + (d(t, u_1(t)) - d(t, u_2(t)), w(t)) + (b(t, u_1(t)) - b(t, u_2(t)), w(t))_\Sigma &= 0 \\ w(0) &= 0 \end{aligned}$$

En évaluant en  $v = w$ , les termes non linéaires peuvent être minorés par 0, et l'intégration en temps amène

$$\frac{1}{2} \|w(\tau)\|^2 - 0 + \int_0^\tau |w(t)|_1^2 dt \leq 0$$

d'où  $w \equiv 0$ , et la solution est unique dans  $W(0, T)$ . L'inégalité d'énergie est obtenue par le même raisonnement que dans le cas des suites  $(u_m)_m$ . Ceci achève la preuve du théorème (1).  $\square$

#### Relaxation et extension

**Proposition 2 – Relaxation de l'hypothèse Lipschitz** Les résultats sont identiques si l'on suppose seulement  $d$  et  $b$  localement Lipschitziennes, i.e. telles que pour tout  $M > 0$ ,  $\exists L(M) > 0$  t.q.

$$|d(x, t, u_1) - d(x, t, u_2)| \leq L(M)|u_1 - u_2| \quad \forall (u_1, u_2) \in B(0, M)$$

Cette relaxation se base sur l'indépendance de l'inégalité d'énergie (1.2) par rapport à  $d$  et  $b$ . Supposons  $d$  seulement localement Lipschitzienne : sa troncature  $d_M := \min(M, \max(-M, d(\cdot, \cdot, \cdot)))$  est alors globalement Lipschitzienne. La solution obtenue par résolution de l'équation associée à  $d_M$  est bornée en norme  $\mathcal{C}(\bar{Q})$  par  $K$  indépendante de  $M$  : ainsi, une seconde troncature  $d_{K+1}$  sera également globalement Lipschitzienne, et la solution associée à son équation coïncidera avec celle issue de  $d$ . L'argument est symétrique pour  $b$ , et nous permettra en particulier d'explorer numériquement la non-linéarité  $b(x, t, u) = u|u|^3$ , qui n'est que localement Lipschitzienne.

On se contente d'énoncer le résultat plus fort suivant, dont la preuve requiert des espaces plus élaborés ([Gri07]).

**THÉORÈME 3 – EXISTENCE D'UNE SOLUTION CONTINUE** Soient  $r > n/2 + 1$  et  $s > n + 1$ , ainsi que  $f \in L^r(Q)$ ,  $g \in L^s(\Sigma)$  et  $u_0 \in \mathcal{C}(\bar{\Omega})$ . Alors l'unique solution faible du problème parabolique linéaire

$$\begin{aligned} \partial_t u + \Delta u &= f && \text{dans } Q \\ \partial_\nu u &= g && \text{dans } \Sigma \\ u(\cdot, T) &= u_0 && \text{dans } \Omega \end{aligned}$$

appartient à  $W(0, T) \cap \mathcal{C}(\bar{Q})$ , et il existe une constante  $C(s, r) > 0$  telle que

$$\|u\|_{W(0, T)} + \|u\|_{\mathcal{C}(\bar{Q})} \leq C(s, r) \left( \|f\|_{L^r(Q)} + \|g\|_{L^s(\Sigma)} + \|u_0\|_{\mathcal{C}(\bar{\Omega})} \right) \quad (1.5)$$

## 1.2 Prendre le contrôle

Introduisons maintenant le problème de contrôle associé à notre système parabolique semilinéaire.

**Définition 5 – Espace des contrôles admissibles** Soient  $(\alpha_{i,\min})_{i \in \llbracket 1, 2 \rrbracket}$  et  $(\alpha_{i,\max})_{i \in \llbracket 1, 2 \rrbracket}$  4 fonctions essentiellement bornées sur leur domaines respectifs. Soient  $s > n + 1$  et  $r > n/2 + 1$

deux entiers. On notera

$$A = \{\alpha \in L^s(Q) \times L^r(\Sigma) \mid \alpha_{i,\min} \leq \alpha_i \leq \alpha_{i,\max} \text{ p.p}\}$$

**Définition 6 – Fonctionnelle  $J$**  Soient  $\varphi : Q \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ,  $\psi : \Sigma \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  et  $\phi : \Omega \times \mathbb{R} \rightarrow \mathbb{R}$  trois fonctions données. On notera  $\alpha := (\alpha_1, \alpha_2)$  un élément quelconque de  $A$ . Définissons la fonctionnelle de coût

$$J(u, \alpha) := \iint_Q \varphi(q, u(q), \alpha_1(q)) dq + \iint_\Sigma \psi(\sigma, t, u(\sigma, t), \alpha_2(\sigma, t)) d\sigma dt + \int_\Omega \phi(x, u(x, T)) dx$$

**Définition 7 – Problème de contrôle** Soit  $u_0 \in \mathcal{C}(\bar{\Omega})$ . Cherchons  $(u, \alpha) \in W(0, T) \cap \mathcal{C}(\bar{Q}) \times A$  minimisant  $J$  et tel que

$$\begin{aligned} \partial_t u - \Delta u + d(\cdot, u(\cdot)) &= \alpha_1 && \text{dans } Q \\ \partial_\nu u + b(\cdot, u(\cdot)) &= \alpha_2 && \text{dans } \Sigma \\ u(\cdot, 0) &= u_0 && \text{dans } \Omega \end{aligned}$$

### 1.2.1 Existence d'un contrôle optimal

**Hypothèse 2** Supposons que  $\varphi$ ,  $\psi$  et  $\phi$  soient

1. mesurables sur leur domaines respectifs,
2. convexes par rapport à  $\alpha$ ,
3. localement lipschitziennes par rapport à  $u$  et  $\alpha$ ,
4. uniformément bornées en  $(u, \alpha) = 0$ , i.e.  $\exists M_\varphi > 0$  t.q.  $|\varphi(x, t, 0, 0)| \leq M_\varphi$  p.p.  $(x, t) \in Q$ , et les équivalents pour  $\psi$  et  $\phi$ .

**THÉORÈME 4 – EXISTENCE DU CONTRÔLE** Sous les hypothèses (1) and (2), éventuellement relaxées en vertu de la proposition (2), le problème de contrôle admet au moins une solution  $(u, \alpha)$ .

#### Démonstration

Notons  $G : A \rightarrow W(0, T)$  l'application entrée-sortie qui, à  $\alpha \in A$ , associe  $G(\alpha) = u$  solution du problème de référence de second membre  $\alpha$ . Par le théorème (1)),  $G$  est bien définie.

**$G(A)$  est borné.** Grâce à l'injection continue  $L^\infty(K) \hookrightarrow L^p(K)$  pour  $K$  compact,  $A$  est borné dans  $L^s(Q) \times L^r(\Sigma)$ . Grâce à l'inégalité d'énergie (1.2), et à l'hypothèse de bornitude de  $b(\cdot, 0)$  et  $d(\cdot, 0)$ , l'image  $G(A)$  est bornée en norme  $W(0, T)$ .

**$J$  est inférieurement bornée.** Grâce au caractère Lipschitzien de  $\varphi$ ,  $\|\varphi(\cdot, G(\alpha), \alpha)\|_{L^\infty} < \infty$  sur  $A$ :

$$\begin{aligned} |\varphi(x, t, G(\alpha)(x, t), \alpha_1(x, t))| &\leq |\varphi(x, t, G(\alpha)(x, t), \alpha_1(x, t)) - \varphi(x, t, 0, 0)| + |\varphi(x, t, 0, 0)| \\ &\leq L_\varphi \left( \|G(\alpha)\|_{\times C(\bar{\Omega})C(\bar{Q})} + \|\alpha_1\|_{L^\infty(Q)} \right) \end{aligned}$$

et l'intégrale de  $\varphi$  sur  $Q$  compact sera inférieurement bornée. Le même raisonnement s'applique à  $\psi$  et  $\phi$ , en utilisant la continuité de la trace de  $C(\bar{Q})$  sur  $\mathcal{C}(\bar{\Omega} \times \{T\})$  pour cette dernière fonction.

Les deux points précédents impliquent l'existence d'une suite minimisante  $(G(\alpha_m), \alpha_m)_m$  dans  $W(0, T) \times L^s(Q) \times L^r(\Sigma)$ . Grâce aux bornes uniformes et à la réflexivité de l'espace considéré, par le théorème de

Banach-Alaoglu, on peut extraire une sous-suite (encore notée  $(G(\alpha_m), \alpha_m)_m$ ) qui converge faiblement vers un élément  $(u, \alpha) \in W(0, T) \times L^s(Q) \times L^r(\Sigma)$ . Comme  $A$  est convexe, il est faiblement fermé, et  $\alpha \in A$ .

**Convergence forte de  $(G(\alpha_m))_m$**  On admettra le résultat suivant, extrait de [Gri07] :

**Proposition 3 – Régularité de  $G$**  Pour le choix de  $s > n + 1$  et  $r > n/2 + 1$ , la restriction de  $G$  à  $u_0 = 0$  est continue de  $L^r(Q) \times L^s(Q)$  dans l'espace des fonctions Hölder-continues  $C^{0,\kappa}(\overline{Q})$ , pour un certain  $\kappa \in ]0, 1[$ .

Décomposons alors notre suite en  $G(\alpha_m) = G((\alpha_{m,1}, \alpha_{m,2}, 0)) + G_l(u_0)$ , où  $G_l$  est l'application entrée-sortie du problème parabolique linéaire avec condition de Neumann homogène

$$\begin{aligned} \partial_t u - \Delta u &= 0 && \text{dans } Q \\ \partial_\nu u &= 0 && \text{dans } \Sigma \\ u(\cdot, 0) &= u_0 && \text{dans } \Omega \end{aligned}$$

Par la proposition (3), la suite  $(u_m - G_l(u_0))_m$  converge faiblement dans  $C^{0,\kappa}(\overline{Q})$ . Par le théorème d'Arzelà-Ascoli, cela implique (à une sous-suite près) une convergence **forte** dans  $C(\overline{Q})$ . Comme  $G_l(u_0) \in C(\overline{\Omega})$  par le théorème (3), il existe un élément  $u \in C(\overline{\Omega})$  tel que  $G(\alpha_m) \rightarrow u$  quand  $m \rightarrow +\infty$ .

**Passage à la limite** La convergence forte permet de passer à la limite en  $m \rightarrow \infty$  dans les problèmes satisfaisants par  $(G(\alpha_m), \alpha_m)$ , et d'en déduire que la limite  $(u, \alpha)$  vérifie bien  $G(\alpha) = u$ . D'autre part, les fonctions  $\varphi$  et  $\psi$  sont lipschitziennes et convexes, ce qui implique que  $J$  est convexe et continue : elle est donc faiblement semi-continue inférieurement (ce qui est nécessaire, car la suite des  $\alpha_m$  ne converge que faiblement). Ainsi,

$$J(u, \alpha) \leq \liminf_{m \rightarrow \infty} J(G(\alpha_m), \alpha_m) = \inf_{\alpha \in A} J(G(\alpha), \alpha)$$

et le candidat  $(u, \alpha)$  est bien un contrôle optimal. □

### 1.2.2 Différentiabilité

Sous les hypothèses (1) and (2), nous savons maintenant qu'il existe un unique contrôle optimal. L'étape suivante est naturellement de le déterminer, et pour adapter les algorithmes de minimisation classiques, il faut s'assurer de pouvoir dériver la fonctionnelle  $J$  le long de la courbe  $(\alpha, G(\alpha))$  des solutions. En particulier, l'opérateur  $G$  doit être différentiable, et c'est le sujet de cette section.

**Hypothèse 3** Soit  $\xi : S \times \mathbb{R} \rightarrow \mathbb{R}$  un opérateur de superposition, et  $m \in \mathbb{N}^*$  un ordre.

1.  $\xi(\cdot, u)$  est mesurable pour tout  $u \in \mathbb{R}$ , et  $\xi(x, \cdot)$  est  $m$  fois différentiable (de  $\mathbb{R}$  dans  $\mathbb{R}$ ) pour presque tout  $x \in S$ .
  2.  $\xi$  est uniformément borné à l'ordre  $m$  : il existe  $M_i > 0$ ,  $i \in \llbracket 1, m \rrbracket$  des constantes telles que  $|D_u^i \xi(x, 0)| \leq M_i$  p.p.  $x \in S$
  3.  $\xi$  est localement lipschitzien à l'ordre  $m$  : pour tout  $M > 0$ , il existe  $L_i(M) > 0$  telles que
- $$|D_u^i \xi(x, u_1) - D_u^i \xi(x, u_2)| \leq L_i(M) |u_1 - u_2| \quad \text{p.p. } x \in S$$
4.  $\xi$  est monotone croissant, ainsi que  $D_u^i \xi$  pour  $i \in \llbracket 1, m-1 \rrbracket$ .

**Lemme 2 – Différentiabilité d'un opérateur de Nemytskii** Sous les hypothèses (3) à l'ordre 1, l'opérateur  $\xi(x, \cdot)$  est différentiable de  $L^\infty(S)$  dans lui-même, et pour  $h \in L^\infty(S)$  une direction,

$$\langle D_u \xi(\cdot, u), h \rangle(x) = \partial_u \xi(x, u(x)) h(x)$$

### Démonstration

Par hypothèse, pour  $u, h \in S \times \mathbb{R}$ , on a

$$\xi(x, u + h) = \xi(x, u) + h\partial_u \xi(x, u) + \int_0^1 (\partial_u \xi(x, u + th)h - h\partial_u \xi(x, u)) dt$$

L'opérateur de superposition  $h \in L^\infty(S) \mapsto h(x)\partial_u \xi(x, u(x))$  est linéaire. De plus,

$$\sup_{x \in S} |h(x)\partial_u \xi(x, u(x))| \leq \|h\|_{L^\infty(S)} \sup_{x \in S} |\partial_u \xi(x, u(x))| \leq \|h\|_{L^\infty(S)} (L_1(\|u\|_{L^\infty(S)}) \|u\|_{L^\infty(S)} + M_1)$$

donc il est également à valeurs dans  $L^\infty(S)$ , et continu par rapport à  $h$  dans cet espace. Enfin, grâce au caractère lipschitzien local, le reste intégral se majore en norme  $L^\infty(S)$  par

$$\left| \int_0^1 (\partial_u \xi(x, u(x) + th(x))h(x) - h(x)\partial_u \xi(x, u(x))) dt \right| \leq \|h\|_{L^\infty(S)} \int_0^1 L_1(\|u\|_{L^\infty(S)}) |th(x)| dt \leq C_u \|h\|_{L^\infty(S)}^2$$

d'où la Fréchet-différentiabilité de  $\xi$  dans  $L^\infty(S)$ .  $\square$

Supposons maintenant que  $d$  et  $b$  satisfassent les hypothèses (3) relativement à  $Q$  et  $\Sigma$ . On en déduit

**Lemme 3** L'opérateur entrée-sortie  $G$  est globalement lipschitzien de  $L^\infty(Q) \times L^\infty(\Sigma)$  dans  $W(0, T) \cap \mathcal{C}(\overline{Q})$ , i.e. il existe  $L > 0$  une constante telle que

$$\|u - \tilde{u}\|_{W(0, T)} + \|u - \tilde{u}\|_{\mathcal{C}(\overline{Q})} \leq L \left( \|\alpha_1 - \tilde{\alpha}_1\|_{L^\infty(Q)} + \|\alpha_2 - \tilde{\alpha}_2\|_{L^\infty(\Sigma)} \right)$$

pour tout  $u = G(\alpha)$  et  $\tilde{u} = G(\tilde{\alpha})$ .

### Démonstration

Étudions le problème satisfait par la différence  $v := u - \tilde{u}$  :

$$\begin{aligned} \partial_t v - \Delta v + d(\cdot, u) - d(\cdot, \tilde{u}) &= \alpha_1 - \tilde{\alpha}_1 && \text{dans } Q \\ \partial_\nu v + b(\cdot, u) - b(\cdot, \tilde{u}) &= \alpha_2 - \tilde{\alpha}_2 && \text{dans } \Sigma \\ v_0 &= 0 && \text{dans } \Omega \end{aligned}$$

Les termes  $\alpha_i - \tilde{\alpha}_i$  sont essentiellement bornés sur des domaines compacts, donc dans tous les  $L^p$ . Grâce aux hypothèses, on peut écrire

$$\begin{aligned} |d(x, t, u(x, t)) - d(x, t, \tilde{u}(x, t))| &= \left| \int_0^1 d_u(x, t, \theta u(x, t) + (1-\theta)\tilde{u}(x, t)) d\theta \right| |u(x, t) - \tilde{u}(x, t)| \\ &\leq C_{\|u\|_\infty, \|\tilde{u}\|_\infty} \|u - \tilde{u}\|_{L^\infty(Q)} \end{aligned}$$

ce qui implique que à  $u, \tilde{u}$  fixés, la fonction  $x, t \mapsto d(x, t, u(x, t)) - d(x, t, \tilde{u}(x, t))$  est mesurable par rapport à  $x, t$  et uniformément bornée. Le même argument s'applique à  $b$ . Ainsi,  $v$  satisfait un problème linéaire dont les seconds membres sont suffisamment réguliers pour appliquer le théorème général (3), ce qui amène directement l'inégalité demandée.  $\square$

Le lemme précédent permet de démontrer le résultat suivant, qui est à la base des méthodes numériques.

**Proposition 4 – Différentiabilité de l'opérateur entrée-sortie** Soit  $\bar{\alpha} \in A$  un contrôle localement optimal. Supposons que  $d$  et  $b$  satisfont les hypothèses (3) à l'ordre 2. Alors l'opérateur  $G$  est Fréchet-différentiable de  $L^r(Q) \times L^s(\Sigma)$  dans  $W(0, T) \cap \mathcal{C}(\overline{Q})$ , et la dérivée directionnelle selon

$\alpha$  au point  $\bar{u}$  est donnée par la solution faible du problème

$$\partial_t u - \Delta u + d_u(\cdot, \bar{u})u = \alpha_1 \quad \text{dans } Q \quad (1.6a)$$

$$\partial_\nu u + b_u(\cdot, \bar{u})u = \alpha_2 \quad \text{dans } \Sigma \quad (1.6b)$$

$$u_0 = 0 \quad \text{dans } \Omega \quad (1.6c)$$

### Démonstration

En vertu des hypothèses à l'ordre 1 sur  $d$  et  $b$ , les opérateurs  $d_u$  et  $b_u$  sont suffisamment réguliers pour appliquer le théorème (1) au problème linéarisé (1.6). Ainsi, l'opérateur  $G'(\alpha) \rightarrow u$  associé est linéaire et continu de  $A$  dans  $W(0, T) \cap \mathcal{C}(\bar{Q})$ . Notons  $u_{\bar{\alpha}} := G(\bar{\alpha})$  ainsi que  $u := G'(\alpha)$ , et montrons maintenant que  $v := \bar{u}_{\bar{\alpha}+\alpha} - \bar{u}_{\bar{\alpha}} - u$  est un  $o(\|\alpha\|^2)$ .

Le problème satisfait par  $v$  s'écrit

$$\begin{aligned} \partial_t v - \Delta v + (d(\cdot, \bar{u}_{\bar{\alpha}+\alpha}) - d(\cdot, \bar{u}_{\bar{\alpha}}) - d_u(\cdot, \bar{u}_{\bar{\alpha}})u) &= (\bar{\alpha}_1 + \alpha_1) - \bar{\alpha}_1 - \alpha_1 = 0 \\ \partial_\nu v + (b(\cdot, \bar{u}_{\bar{\alpha}+\alpha}) - b(\cdot, \bar{u}_{\bar{\alpha}}) - b_u(\cdot, \bar{u}_{\bar{\alpha}})u) &= (\bar{\alpha}_2 + \alpha_2) - \bar{\alpha}_2 - \alpha_2 = 0 \\ v_0 &= 0 \end{aligned}$$

ou encore

$$\begin{aligned} \partial_t v - \Delta v + d_u(\cdot, \bar{u}_{\bar{\alpha}})v &= -(d(\cdot, \bar{u}_{\bar{\alpha}+\alpha}) - d(\cdot, \bar{u}_{\bar{\alpha}})) + d_u(\cdot, \bar{u}_{\bar{\alpha}})(\bar{u}_{\bar{\alpha}+\alpha} - \bar{u}_{\bar{\alpha}}) \\ \partial_\nu v + b_u(\cdot, \bar{u}_{\bar{\alpha}})v &= -(b(\cdot, \bar{u}_{\bar{\alpha}+\alpha}) - b(\cdot, \bar{u}_{\bar{\alpha}})) + b_u(\cdot, \bar{u}_{\bar{\alpha}})(\bar{u}_{\bar{\alpha}+\alpha} - \bar{u}_{\bar{\alpha}}) \end{aligned}$$

Comme  $d_u$  et  $b_u$  satisfont les hypothèses (3) à l'ordre 1, l'opérateur entrée-sortie  $G'$  associé au problème linéarisé (1.6) est globalement lipschitzien. Il est également linéaire, donc  $G'(0) = 0$ , et l'on peut majorer

$$\begin{aligned} \|v\|_{W(0, T)} + \|v\|_{\mathcal{C}(\bar{Q})} &\leq L'(\|d(\cdot, \bar{u}_{\bar{\alpha}+\alpha}) - d(\cdot, \bar{u}_{\bar{\alpha}}) - d_u(\cdot, \bar{u}_{\bar{\alpha}})(\bar{u}_{\bar{\alpha}+\alpha} - \bar{u}_{\bar{\alpha}})\|_{L^\infty(Q)} \\ &\quad + \|b(\cdot, \bar{u}_{\bar{\alpha}+\alpha}) - b(\cdot, \bar{u}_{\bar{\alpha}}) - b_u(\cdot, \bar{u}_{\bar{\alpha}})(\bar{u}_{\bar{\alpha}+\alpha} - \bar{u}_{\bar{\alpha}})\|_{L^\infty(Q)}) \end{aligned}$$

Développons les termes en  $d$  : grâce au caractère lipschitzien de  $d_u$ ,

$$\begin{aligned} &|d(\cdot, \bar{u}_{\bar{\alpha}+\alpha}) - d(\cdot, \bar{u}_{\bar{\alpha}}) - d_u(\cdot, \bar{u}_{\bar{\alpha}})(\bar{u}_{\bar{\alpha}+\alpha} - \bar{u}_{\bar{\alpha}})| \\ &\leq \int_0^1 |d_u(\cdot, \theta \bar{u}_{\bar{\alpha}+\alpha} + (1-\theta)\bar{u}_{\bar{\alpha}}) - d_u(\cdot, \bar{u}_{\bar{\alpha}})| d\theta |\bar{u}_{\bar{\alpha}+\alpha} - \bar{u}_{\bar{\alpha}}| \leq \int_0^1 \theta d\theta |\bar{u}_{\bar{\alpha}+\alpha} - \bar{u}_{\bar{\alpha}}|^2 \end{aligned}$$

Par passage au sup sur  $(x, t) \in Q$ , on obtient l'inégalité en norme  $L^\infty(Q)$ . Le même argument s'applique à  $b$ . Enfin, en utilisant le caractère lipschitzien de  $G$  donné par le lemme (3), on obtient  $\|\bar{u}_{\bar{\alpha}+\alpha} - \bar{u}_{\bar{\alpha}}\| \leq L \|\alpha\|_{L^\infty(Q) \times L^\infty(\Sigma)}$ . Nous avons démontré que

$$\|G(\bar{\alpha} + \alpha) - G(\bar{\alpha}) - G'(\alpha)\|_{W(0, T)} + \|G(\bar{\alpha} + \alpha) - G(\bar{\alpha}) - G'(\alpha)\|_{\mathcal{C}(\bar{Q})} \leq C \|\alpha\|_{L^\infty(Q) \times L^\infty(\Sigma)}^2$$

ce qui prouve la Fréchet-différentiabilité de  $G$ . □

### 1.2.3 Système d'optimalité

Cette section s'écarte du cadre du problème de référence (1.1), pour pouvoir être plus tard appliquée au problème de Navier-Stokes. En conséquence, on adopte des notations plus générales, qui pourront inclure des conditions au bord de Dirichlet, ou des non-linéarités dépendantes de  $\nabla u$ . Les questions de différentiabilité et d'existence d'un adjoint ne sont pas abordées dans le cas général : pour les applications présentées, on le construira explicitement. Nous nous appuyons sur [Gun03] et [MAS18].

**Définition 8 – Notations** Posons  $E := L^2(Q) \times L^2(\Sigma) \times L^2(\Pi)$  un espace de Hilbert muni du produit scalaire

$$(\cdot, \cdot)_E := (\cdot, \cdot)_{L^2(Q)} + (\cdot, \cdot)_{L^2(\Sigma)} + (\cdot, \cdot)_{L^2(\Pi)}$$

Soit  $A \subset E$  l'espace des contrôles admissibles, et  $V$  l'espace d'état. On définit  $J : V \times A \rightarrow \mathbb{R}$  par

$$J(u, \alpha) := (\Phi(u, \alpha), 1_E)_E, \quad \Phi(u, \alpha) := (\varphi(u, \alpha_1), \psi(u, \alpha_2), \phi(u, \alpha_3))^t$$

que l'on cherche à minimiser sous la condition  $\mathcal{T}(u) = \alpha$ , au sens

$$\begin{aligned} \mathcal{T}_1(u) &= \alpha_1 && \text{dans } Q \\ \mathcal{T}_2(\tau_\Sigma(u)) &= \alpha_2 && \text{dans } \Sigma \\ \mathcal{T}_3(\tau_\Pi(u)) &= \alpha_3 && \text{dans } \Pi \end{aligned}$$

où  $\mathcal{T} : V \rightarrow E$  est non linéaire, et  $\tau_K : V \subset H^{1/2}(Q) \mapsto L^2(K)$  désigne l'opérateur trace sur  $K$ .

**Remarque 3** Afin de ne pas alourdir l'énoncé, on a pris la liberté de considérer l'état initial  $\alpha_3$  comme un contrôle, là où nos résultats d'existence ne portent que sur  $\alpha_{1,2}$ . Les applications se restreindront naturellement à ce cas. De plus, il est important de bien garder en tête que  $\alpha \in E$  est un triplet, et toutes les opérations de dérivation sont à prendre au sens vectoriel.

**Définition 9 – Adjoint** Soit  $L : E \rightarrow F$  un opérateur linéaire entre deux espaces de Hilbert.  $L^* : F^* \rightarrow E^*$  sera dit adjoint à  $E$  si  $\forall u, v \in E \times F^*$ ,

$$(Lu, v)_F = (u, L^*v)_E$$

**Proposition 5 – Système d'optimalité** Supposons que  $\partial_u \mathcal{T}$  admette un adjoint. Si  $(\bar{u}, \bar{\alpha})$  est une solution du problème de minimisation, alors il existe un élément  $w \in E^*$  tel que le système suivant soit satisfait :

$$\begin{aligned} \mathcal{T}(\bar{u}) - \bar{\alpha} &= 0 && \text{dans } E && \text{l'équation d'état} \\ (\partial_u \mathcal{T}(\bar{u}))^* w - \partial_u \Phi(\bar{u}, \bar{\alpha}) &= 0 && \text{dans } E^* && \text{l'équation adjointe} \\ \partial_\alpha \Phi(\bar{u}, \bar{\alpha}) + w &= 0 && \text{dans } E^* && \text{les conditions d'optimalité} \end{aligned}$$

### Démonstration

On utilisera le raccourci d'écriture  $(w, \cdot)_E$  pour désigner  $((w, \tau_\Sigma(w), \tau_\Pi(w))^t, \cdot)_E$ . On forme le Lagrangien

$$\mathcal{L}(u, \alpha, w) := J(u, \alpha) - (w, \mathcal{T}(u) - \alpha)_E$$

La minimisation sans contrainte du Lagrangien produit (formellement) le système d'équations suivant :

$$\partial_u J(\bar{u}, \bar{\alpha}) \cdot - (w, \partial_u \mathcal{T}(\bar{u}) \cdot)_E = 0 \quad \text{dans } L^2(Q) \quad (1.7)$$

$$\partial_{\alpha_1} \varphi(\bar{u}, \bar{\alpha}_1) + w = 0 \quad \text{dans } (L^2(Q))^* \quad (1.8)$$

$$\partial_{\alpha_2} \psi(\bar{u}, \bar{\alpha}_2) + \tau_\Sigma w = 0 \quad \text{dans } (L^2(\Sigma))^* \quad (1.9)$$

$$\partial_{\alpha_3} \phi(\bar{u}, \bar{\alpha}_3) + \tau_\Pi w = 0 \quad \text{dans } (L^2(\Pi))^* \quad (1.10)$$

$$(\cdot, \mathcal{T}(\bar{u}) - \bar{\alpha})_E = 0 \quad \text{dans } (L^2(Q))^* \quad (1.11)$$

L'équation (1.11) exprime exactement l'équation d'état, et (1.8) to (1.10) correspondent aux conditions

d'optimalité. Enfin, l'équation (1.7) nous donne l'équation adjointe :

$$\left( \begin{pmatrix} \partial_u \varphi(\bar{u}, \bar{\alpha}_1) \cdot \\ \partial_u \psi(\bar{u}, \bar{\alpha}_2) \cdot \\ \partial_u \phi(\bar{u}, \bar{\alpha}_3) \cdot \end{pmatrix}, 1_E \right)_E - ((\partial_u \mathcal{T}(\bar{u}))^* w, \cdot)_E = 0 \implies \begin{pmatrix} \partial_u \varphi(\bar{u}, \bar{\alpha}_1) \\ \partial_u \psi(\bar{u}, \bar{\alpha}_2) \\ \partial_u \phi(\bar{u}, \bar{\alpha}_3) \end{pmatrix} = (\partial_u \mathcal{T}(\bar{u}))^* w$$

qui est à comprendre dans  $E^*$ , et que l'on résume en  $\partial_u \Phi(\bar{u}, \bar{\alpha}) = (\partial_u \mathcal{T}(\bar{u}))^* w$ .  $\square$

### Exemple d'expression de l'adjoint

On cherche à préciser le sens de  $(\partial_u \mathcal{T}(\bar{u}))^* w$  dans le cas particulier du problème de référence. L'opérateur  $\mathcal{T}$  et sa dérivée directionnelle en  $\bar{u}$  dans la direction  $u$  sont respectivement donnés par

$$\mathcal{T}(u) = \begin{pmatrix} \partial_t u - \Delta u + d(\cdot, u) \\ \partial_\nu u + b(\cdot, \tau_\Sigma(u)) \\ \tau_{\Omega \times \{0\}} u \end{pmatrix}, \quad (\partial_u \mathcal{T}(\bar{u}), u) = \begin{pmatrix} \partial_t u - \Delta u + d_u(\cdot, \bar{u})u \\ \partial_\nu u + b_u(\cdot, \tau_\Sigma(\bar{u}))u \\ \tau_{\Omega \times \{0\}} u \end{pmatrix}$$

avec Soit  $w \in E^*$ , que l'on identifie à  $E$ . Le produit scalaire  $(w, \mathcal{T}(u))_E$  est égal à

$$(w, \partial_u \mathcal{T}(\bar{u})u)_E = (w, \partial_t u + \Delta u + d_u(\cdot, \bar{u})u)_Q + (\tau_\Sigma w, \partial_\nu u + b_u(\cdot, \tau_\Sigma \bar{u})u)_\Sigma + (\tau_\Pi w, \tau_{\Omega \times \{0\}} u)_\Pi$$

où, par la formule de Green,

$$\begin{aligned} (w, \partial_t u)_Q &= (\tau_\Pi w, \tau_{\Omega \times \{T\}} u)_\Pi - (\tau_\Pi w, \tau_{\Omega \times \{0\}} u)_\Pi - (\partial_t w, u)_Q \\ (w, -\Delta u)_Q &= -(\tau_\Sigma w, \partial_\nu u)_\Sigma + (\partial_\nu w, \tau_\Sigma u)_\Sigma - (\Delta w, u)_Q \end{aligned}$$

et d'autre part,  $(w, a(\cdot)u)_K = (a(\cdot)w, u)_K$  pour  $a$  ne dépendant ni de  $w$ , ni de  $u$ . Remarquons enfin que  $(\tau_\Pi w, \tau_{\Omega \times \{T\}} u)_\Pi = (\tau_{\Omega \times \{T\}} w, \tau_\Pi u)_\Pi$ , par simple discussion sur l'ensemble d'intégration. Le produit scalaire devient

$$(w, \partial_u \mathcal{T}(\bar{u})u)_E = (-\partial_t w - \Delta w + d_u(\cdot, \bar{u})w, u)_Q + (\partial_\nu w, +b_u(\cdot, \tau_\Sigma \bar{u})\tau_\Sigma w, u)_\Sigma + (\tau_{\Omega \times \{T\}} w, \tau_\Pi u)_\Pi$$

D'où l'expression recherchée :

$$(\partial_u \mathcal{T}(\bar{u}))^* w = \begin{pmatrix} -\partial_t w - \Delta w + d_u(\cdot, \bar{u})w \\ \partial_\nu w + b_u(\cdot, \tau_\Sigma \bar{u})\tau_\Sigma w \\ \tau_{\Omega \times \{T\}} w \end{pmatrix}$$

### Gradient de la fonctionnelle de coût

L'introduction du système d'optimalité n'est pas la seule utilité de l'adjoint. L'étape suivante prolonge le raisonnement précédent en exprimant le gradient de la fonctionnelle réduite  $J(u(\alpha), \alpha)$  grâce à  $w$ .

**Proposition 6** La forme linéaire  $D_\alpha J(u(\alpha), \alpha)$  est donnée par

$$D_\alpha J(u(\alpha), \alpha)(\cdot) = (w, \cdot)_E + (\partial_\alpha \Phi(u(\alpha), \alpha), \cdot)_E$$

### Démonstration

Nous suivons le raisonnement développé dans [GM00]. Notons par  $u := u(\alpha)$  l'élément de  $E$  solution de l'équation d'état. Par définition, on a  $0_E = \mathcal{T}(u(\alpha)) - \alpha$ , et la différentielle d'une quantité nulle reste nulle.

Par la règle de la chaîne, on en déduit une inégalité de sensibilité

$$\partial_u \mathcal{T}(u(\alpha)) \begin{pmatrix} \partial_\alpha u \cdot \\ \partial_\alpha \tau_\Sigma u \cdot \\ \partial_\alpha \tau_\Pi u \cdot \end{pmatrix} - id_\alpha(\cdot) = 0 \quad (1.12)$$

où l'on s'autorisera l'abus d'écriture  $\partial_\alpha u \cdot$  pour l'accroissement de  $(u, \tau_\Sigma u, \tau_\Pi u)$  dans la direction  $\alpha$ . D'autre part,

$$D_\alpha J(u(\alpha), \alpha)(\cdot) = (\partial_u \Phi(u, \alpha), \partial_\alpha u \cdot, 1_E)_E + (\partial_\alpha \Phi(u, \alpha), \cdot, 1_E)_E = (\partial_u \Phi(u, \alpha), \partial_\alpha u \cdot)_E + (\partial_\alpha \Phi(u, \alpha), \cdot)_E$$

Or, en utilisant le problème adjoint, le premier terme devient

$$(\partial_u \Phi(u, \alpha), \partial_\alpha u \cdot)_E = ((\partial_u \mathcal{T}(\bar{u}))^* w, \partial_\alpha u \cdot)_E = (w, \partial_u \mathcal{T}(\bar{u}) \partial_\alpha u \cdot)_E$$

Par substitution avec l'inégalité de sensibilité (1.12), il vient

$$D_\alpha J(u(\alpha), \alpha)(\cdot) = (w, \cdot)_E + (\partial_\alpha \Phi(u, \alpha), \cdot)_E$$

et l'on retrouve exactement les conditions d'optimalité pour  $D_\alpha J(\bar{u}, \bar{\alpha}) \equiv 0$ .  $\square$

### Dérivée seconde de la fonctionnelle de coût

On souhaite poursuivre le raisonnement à l'ordre 2.

**Proposition 7** La dérivée seconde de la fonctionnelle de coût  $J(u, \alpha)$  le long de la courbe  $(u(\alpha), \alpha)$  peut être exprimée comme une dérivée partielle du Lagrangien en variables  $(u, \alpha)$  :

$$D_{\alpha\alpha}^2 J(u(\alpha), \alpha)(\cdot, \cdot) = \partial_{(u,\alpha),(u,\alpha)}^2 \mathcal{L}(u, \alpha, w)(\partial_\alpha u \cdot, \cdot)$$

où  $w$  satisfait l'équation adjointe de la proposition (5).

### Démonstration

Le gradient de  $J$  le long de la courbe paramétrée  $(u(\alpha), \alpha)$  s'écrit

$$\begin{aligned} D_\alpha J(u(\alpha), \alpha)(\cdot) &= (\partial_u \Phi(u, \alpha), \partial_\alpha u \cdot)_E + (\partial_\alpha \Phi(u, \alpha), \cdot)_E \\ D_{\alpha\alpha}^2 J(u(\alpha), \alpha)(\cdot, \cdot) &= (\partial_{uu}^2 \Phi(u, \alpha), \partial_\alpha u \cdot, \partial_\alpha u \cdot)_E + (\partial_{u\alpha}^2 \Phi(u, \alpha), \partial_\alpha u \cdot, \cdot)_E + (\partial_u \Phi(u, \alpha), (\partial_{\alpha\alpha}^2 u, \cdot, \cdot))_E \\ &\quad + (\partial_{\alpha u}^2 \Phi(u, \alpha), \cdot, \partial_\alpha u \cdot)_E + (\partial_{\alpha\alpha}^2 \Phi(u, \alpha), \cdot, \cdot)_E \end{aligned}$$

Le problème adjoint nous donne

$$(\partial_u \Phi(u, \alpha), (\partial_{\alpha\alpha}^2 u, \cdot, \cdot))_E = ((\partial_u \mathcal{T}(\bar{u}))^* w, (\partial_{\alpha\alpha}^2 u, \cdot, \cdot))_E = (w, \partial_u \mathcal{T}(\bar{u}) (\partial_{\alpha\alpha}^2 u, \cdot, \cdot))_E$$

Calculons cette expression en dérivant une nouvelle fois l'inégalité de sensibilité (1.12) par rapport à  $\alpha$  :

$$(\partial_{uu}^2 \mathcal{T}(\bar{u}), \partial_\alpha u \cdot, \partial_\alpha u \cdot) + (\partial_u \mathcal{T}(\bar{u}), (\partial_{\alpha\alpha}^2 u, \cdot, \cdot)) = 0$$

Ainsi

$$\begin{aligned} D_\alpha^2 J(u(\alpha), \alpha)(\cdot, \cdot) &= (\partial_{uu}^2 \Phi(u, \alpha), \partial_\alpha u \cdot, \partial_\alpha u \cdot)_E + (\partial_{u\alpha}^2 \Phi(u, \alpha), \partial_\alpha u \cdot, \cdot)_E - (w, (\partial_{uu}^2 \mathcal{T}(\bar{u}), \partial_\alpha u \cdot, \partial_\alpha u \cdot))_E \\ &\quad + (\partial_{\alpha u}^2 \Phi(u, \alpha), \cdot, \partial_\alpha u \cdot)_E + (\partial_{\alpha\alpha}^2 \Phi(u, \alpha), \cdot, \cdot)_E \end{aligned}$$

et on reconnaît la différentielle seconde du Lagrangien  $\mathcal{L}(u, \alpha, w) = J(u, \alpha) - (w, \mathcal{T}(u) - \alpha)_E$  par rapport aux variables  $(u, \alpha)$ , évaluée dans la direction  $(\partial_\alpha u \cdot, \cdot)$  (où  $\partial_\alpha u \cdot$  est bien, moralement, un accroissement de  $u$ ) :

$$D_\alpha^2 J(u(\alpha), \alpha)(\cdot, \cdot) = \partial_{(u,\alpha),(u,\alpha)}^2 \mathcal{L}(u, \alpha, w)(\partial_\alpha u \cdot, \cdot)$$

et cette expression fait disparaître le terme  $\partial_\alpha^2 u$ .  $\square$

## Chapitre 2

# Système de Navier-Stokes incompressible

Ce chapitre a pour but d'introduire le modèle de Navier-Stokes en dimension 2. On donne une esquisse de la modélisation et des démonstrations des résultats principaux, puis on discute le passage à une formulation numérique et quelques propriétés des schémas proposés.

Contrairement au chapitre précédent, les variables utilisées sont des fonctions à valeurs vectorielles. Clarifions dès à présent les notations employées.

**Définition 10 – Notations** Soit  $\Omega \subset \mathbb{R}^n$  un ouvert à frontière Lipschitzienne.

- Soit  $u : \Omega \rightarrow \mathbb{R}^n$  une fonction à valeurs vectorielles, et  $H$  un espace de fonctions réelles. L'espace  $[H]^2$  est l'ensemble des fonctions  $u$  telles que  $\forall i \in \llbracket 1, n \rrbracket$ ,  $u_i \in H$ .
- Soit  $([H]^2, \|\cdot\|_{[H]^2})$  un espace métrique. On notera également par  $\|\cdot\|_H$  la norme définie sur l'espace des fonctions à valeurs vectorielles par

$$\|u\|_{[H]^2} := \sum_{i=1}^n \|u_i\|_H$$

On emploiera sans la mentionner l'équivalence des normes en dimension finie.

- Les opérateurs de dérivation  $\partial$ ,  $\Delta$ ,  $\nabla$  sont définis coordonnées par coordonnée : on écrira

$$\partial_t u = (\partial_t u_i)_{i \in \llbracket 1, n \rrbracket}^t \in \mathbb{R}^n, \quad \Delta u = (\Delta u_i)_{i \in \llbracket 1, n \rrbracket}^t \in \mathbb{R}^n, \quad \nabla u = \left( \frac{\partial u_i}{\partial x_j} \right)_{i, j \in \llbracket 1, n \rrbracket^2} \in \mathbb{M}_{n,n}(\mathbb{R})$$

- Un produit scalaire  $(\cdot, \cdot)$  défini sur des fonctions réelles est étendu aux fonctions vectorielles par

$$(u, v) := \sum_{i=1}^n (u_i, v_i) \quad \forall u, v : \Omega \rightarrow \mathbb{R}^n$$

On conserve les notations du premier chapitre pour les ensembles  $Q = \Omega \times ]0, T[$ ,  $\Sigma = \partial\Omega \times ]0, T[$  et  $\Pi = \Omega \times \{0, T\}$ .

### 2.1 Les fameuses équations de Navier-Stokes

Le modèle que nous allons étudier s'intéresse au mouvement des fluides incompressibles. Les étapes de la modélisation physique sont amplement détaillées dans une multitude d'ouvrages, dont par exemple [VM07] : on se contente d'en donner une intuition.

### 2.1.1 Aperçu de la modélisation

**Loi de Newton** Représentons-nous un fluide comme un ensemble de particules de fluide se déplaçant le long de lignes de courant. Si l'on repère par  $x(t) \in \mathbb{R}^n$  la position de l'une d'entre elle, la deuxième loi de Newton nous dit que sa trajectoire satisfait  $\rho \ddot{x}(t) = g(x(t), t)$ , où  $\rho$  est la masse volumique du fluide, considérée constante, et  $g$  est la somme des forces par unité de volume. Nous allons nous intéresser à une vision Eulérienne du fluide, où notre inconnue sera le champ de vitesse du fluide : plus précisément, on note  $u : \mathbb{R}^n \times [0, T] \rightarrow T\mathbb{R}^n$  l'espace tangent à  $\mathbb{R}^n$  (identifié à  $\mathbb{R}^n$ ) le champ tel que pour toute ligne de courant,

$$\dot{x}(t) = u(x(t), t) \quad \forall t \in [0, T]$$

Par la règle de la chaîne, on obtient la dérivée particulaire

$$\ddot{x}(t) = \partial_t u(x(t), t) + \nabla u(x(t), t) \dot{x}(t) = \partial_t u(x(t), t) + \nabla u(x(t), t) \cdot u(x(t), t)$$

La première version de notre modèle a pour inconnue un tel champ  $u$ , de valeur initiale  $u_0$  et se propageant selon la loi de Newton :

$$\begin{aligned} \rho \partial_t u + \rho \nabla u \cdot u &= g && \text{dans } Q \\ u(\cdot, 0) &= u_0 && \text{dans } \Omega \end{aligned} \tag{2.1}$$

**Incompressibilité** On s'intéresse aux fluides incompressibles, c'est-à-dire tels qu'en l'absence de source, le flux entrant dans n'importe quel volume soit égal au flux sortant simultanément dudit volume. L'application du théorème de Green sur un volume  $V$  fermé quelconque rend cette contrainte équivalente à  $\operatorname{div} u \equiv 0$ . L'introduction dans le modèle se fait au moyen d'un multiplicateur de Lagrange : supposons (très formellement) que  $J(u, g)$  soit telle que

$$\nabla_u J(u, g) \cdot v = (\rho \partial_t u + \rho \nabla u \cdot u - g, v)$$

L'équation (2.1) impose que  $u$  soit une extrémale de  $J$ . Soit  $p : \mathbb{R}^n \times [0, T] \rightarrow \mathbb{R}$  un champ de multiplicateurs de Lagrange. Le Lagrangien associé à l'incompressibilité s'écrit  $\mathcal{L}(u, g, p) := J(u, g) + (p, \operatorname{div} u)$ . On choisit d'annuler  $u$  au bord (dans un sens à justifier) pour que grâce à Green,  $(p, \operatorname{div} u) = (\nabla p, u)$  : ceci permet d'écrire

$$\begin{aligned} \nabla_u \mathcal{L}(u, g, p) \cdot v &= (\rho \partial_t u + \rho \nabla u \cdot u - g, v) + (\nabla p, v) = 0 \\ \nabla_p \mathcal{L}(u, g, p) \cdot v &= (v, \operatorname{div} u) = 0 \end{aligned}$$

et si  $v(\cdot, t)$  appartient à un espace dans lequel  $[\mathcal{D}(\Omega)]^2$  est dense, l'évolution du couple  $(u, p)$  satisfait

$$\begin{aligned} \rho \partial_t u + \rho \nabla u \cdot u + \nabla p &= g && \text{dans } Q \\ \operatorname{div} u &= 0 && \text{dans } Q \end{aligned}$$

Le lecteur aura reconnu dans le terme  $p$  la pression physique.

**Viscosité** La dernière étape consiste à classer les forces appliquées en deux catégories :

- les forces internes de viscosité, qui agissent par l'intermédiaire de  $\gamma \Delta u$ , où  $\gamma > 0$  pondère l'importance du frottement visqueux,
- "les autres", regroupées dans  $f$  et supposées indépendantes de  $u$ .

On termine en normalisant chacun des termes par  $\rho > 0$ , et en notant  $\nu := \frac{\gamma}{\rho}$  ce que l'on appellera la viscosité. Finalement, nous avons obtenu le modèle suivant :

### Définition 11 – Équation de Navier-Stokes

$$\partial_t u + \nabla u \cdot u - \nu \Delta u + \nabla p = f \quad \text{dans } Q \quad (2.2a)$$

$$\operatorname{div} u = 0 \quad \text{dans } Q \quad (2.2b)$$

$$\tau_\Sigma u = 0 \quad \text{dans } \Sigma \quad (2.2c)$$

$$u(\cdot, 0) = u_0 \quad \text{dans } \Omega \quad (2.2d)$$

### 2.1.2 Formulation variationnelle

Dans toute la suite, on se place dans un domaine spatial  $\Omega$  supposé suffisamment régulier (de classe  $C^1$  ou Lipschitz). Une fonction test aura la forme  $(v, q)$ , où la vitesse  $v : \Omega \rightarrow \mathbb{R}^n$  et la pression  $q : \Omega \rightarrow \mathbb{R}$  sont, pour l'instant, aussi régulières que désiré. Supposons  $v$  nulle au bord : l'application (formelle) de Green amène à la formulation

$$\begin{aligned} \partial_t(u(t), v) + (\nabla u \cdot u(t), v) + \nu(\nabla u(t), \nabla v) + (p(t), \operatorname{div} v) &= (f(t), v) \\ (\operatorname{div} u, q) &= 0 \end{aligned}$$

Les espaces choisis pour  $u, v, p, q$  doivent donner un sens à chacun des termes de cette formulation. Les dérivées partielles en espace de  $u(t)$  et  $v(t)$  doivent être définies et dans  $L^2(\Omega)$ , avec  $u, v$  nulles au bord. La vitesse  $u$  est à divergence nulle, et le terme  $(\partial_t u(t), v)$  doit avoir un sens. Pour satisfaire à ces exigences, on définit

### Définition 12 – Espaces spatiaux pour la vitesse

$$\mathcal{D} = \{w \in [\mathcal{D}(\Omega)]^2 \mid \operatorname{div} w = 0\},$$

$$\mathbb{H} = \text{adhérence de } \mathcal{D} \text{ dans } [L^2(\Omega)]^2,$$

$$\mathbb{V} = \text{adhérence de } \mathcal{D} \text{ dans } [H^1(\Omega)]^2$$

En particulier,  $\mathbb{H}$  et  $\mathbb{V}$  sont des sous-espaces vectoriels fermés de  $[H_0^1(\Omega)]^2$ , donc des espaces de Hilbert. La dérivée en temps est traitée au sens des distributions : on impose  $u \in \mathcal{C}([0, T]; \mathbb{H})$ , et on définit

$$\int_0^T \partial_t(u(t), v) \phi(t) dt = - \int_0^T (u(t), v) \partial_t \phi(t) dt \quad \forall \phi \in [\mathcal{D}(]0, T[)]^2$$

La continuité en temps est une condition forte, qui assure au passage que la condition initiale  $u_0 \in \mathbb{H}$  puisse être satisfaite, et qui amène également  $u \in L^\infty(0, T; \mathbb{H})$  si  $T$  est fini. Cette condition n'est d'ailleurs satisfaite qu'en dimension 2. On choisit le terme source  $f \in L^2(0, T; [H^{-1}(\Omega)]^2)$ , et l'on remplace le produit scalaire  $(f(t), v)$  par le crochet de dualité  $\langle f(t), v \rangle$ .

Dans la suite, on emploiera les notations classiques suivantes.

### Définition 13 – Formes multilinéaires

On introduit les formes

$$a(u, v) = \nu \int_{\Omega} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial u_i}{\partial x_j} \frac{\partial v_i}{\partial x_j} dx = \nu(\nabla u, \nabla v) \quad \forall u, v \in [H^1(\Omega)]^2$$

$$b(u, p) = - \int_{\Omega} \operatorname{div}(u) p dx \quad \forall u \in [H^1(\Omega)]^2, p \in [L^2(\Omega)]^2$$

$$c(u, v, w) = \int_{\Omega} \sum_{i=1}^n \sum_{j=1}^n u_i \frac{\partial v_j}{\partial x_i} w_j dx = ((u \cdot \nabla)v, w) \quad \forall u, v, w \text{ bien choisis}$$

La forme  $b$  apparaîtra dans la formulation que nous discréterons, mais n'a pas d'intérêt pour l'étude théorique, car  $b(u, p) \equiv 0$  pour  $u \in \mathbb{H}$ .

**Définition 14 – Formulation variationnelle de Navier-Stokes** Soient  $u_0 \in \mathbb{H}$  et  $f \in L^2(0, T; [H^{-1}(\Omega)]^2)$ . Le champ  $u \in X_t := L^2(0, T; \mathbb{V}) \cap C([0, T], \mathbb{H})$  est une solution faible de Navier-Stokes s'il satisfait

$$\begin{aligned} \partial_t(u(t), v) + a(u(t), v) + c(u(t), u(t), v) &= \langle f(t), v \rangle & \forall v \in \mathbb{V}, \text{ au sens de } [\mathcal{D}'(0, T)]^2 \\ u(\cdot, 0) &= u_0 & \text{p.p dans } \Omega \end{aligned}$$

Le lecteur attentif n'aura pas manqué de s'offusquer de l'apparent oubli du terme non linéaire  $c(u(t), u(t), v)$  dans la construction des espaces de fonctions. La proposition suivante justifie ce silence.

**Proposition 8** Soient  $u, v, w \in [H^1(\Omega)]^2$ . En dimension  $n = 2$ , il existe une constante  $C > 0$  telle que

$$|c(u, v, w)| \leq C \|u\|_{H^1(\Omega)} \|\nabla v\|_{L^2(\Omega)} \|w\|_{H^1(\Omega)}$$

### Démonstration

D'après les injections de Sobolev, en dimension  $n = 2$ ,  $H^1(\Omega)$  s'injecte continûment dans  $L^p(\Omega)$  pour  $p \in [1, \infty[$ . En particulier, il existe une constante  $C > 0$  telle que  $\|u\|_{L^4(\Omega)} \leq C \|u\|_{H^1(\Omega)}$ . On a alors

$$\begin{aligned} |c(u, v, w)| &= \left| \sum_{i=1}^n \sum_{j=1}^n \int_{\Omega} u_i \frac{\partial v_j}{\partial x_i} w_j dx \right| \leq \sum_{i=1}^n \sum_{j=1}^n \left( \int_{\Omega} u_i^4 dx \right)^{1/4} \left( \int_{\Omega} \left( \frac{\partial v_j}{\partial x_i} \right)^2 dx \right)^{1/2} \left( \int_{\Omega} w_j^4 dx \right)^{1/4} \\ &\leq C \|u\|_{L^4(\Omega)} \|\nabla v\|_{L^2(\Omega)} \|w\|_{L^4(\Omega)} \leq C \|u\|_{H^1(\Omega)} \|\nabla v\|_{L^2(\Omega)} \|w\|_{H^1(\Omega)} \end{aligned}$$

et cette estimation, bien que grossière, suffit à justifier l'emploi de  $c$  dans la formulation variationnelle.  $\square$

### 2.1.3 Aperçu du problème bien posé

Le résultat suivant provient des travaux Jacques-Louis Lions ([Lio69]). On en suit en partie l'exposé fait dans [Dre12].

**THÉORÈME 5 – EXISTENCE ET UNICITÉ EN DIMENSION 2** En dimension  $n = 2$ , la formulation variationnelle (14) admet une unique solution, qui satisfait l'égalité d'énergie

$$\frac{1}{2} \|u(t)\|_{L^2(\Omega)}^2 - \frac{1}{2} \|u_0\|_{L^2(\Omega)}^2 + \int_0^t \nu \|\nabla u(t)\|_{L^2(\Omega)}^2 dt = \int_0^t \langle f(t), u(t) \rangle dt$$

Cette section a pour but de résumer les étapes du raisonnement, sans ambition d'être ni exhaustif ni précis, mais plutôt de permettre au lecteur (et surtout à l'auteure) d'avoir une idée des arguments nécessaires.

#### Étape 1 (Carathéodory en dimension finie)

Soit  $\mathbb{V}_2 := \{\text{l'adhérence de } \mathcal{V} \text{ dans } [H^2(\Omega)]^2\}$  et  $A : \mathbb{H} \rightarrow \mathbb{V}_2$  défini par  $(Au, v)_{\mathbb{V}_2} = (u, v)$ . L'opérateur  $A$  est compact autoadjoint, donc admet une base hilbertienne  $\{e_i\}_{i \in \mathbb{N}}$  de fonctions propres. Posons alors  $u_m(\cdot, t) = \sum_{i=1}^m g_i(t) e_i$  une famille de fonctions telles que  $g_i$  soit absolument continu en temps, et  $u_m$

satisfasse

$$\begin{aligned} \partial_t(u_m(t), e_j) + a(u_m(t), e_j) + c(u_m(t), u_m(t), e_j) &= \langle f(t), e_j \rangle \quad \forall j \in \llbracket 1, m \rrbracket, \text{ dans } [\mathcal{D}'(]0, T[)]^2 \\ g_i(0) &= (u_0, e_i) \quad \forall i \in \llbracket 1, m \rrbracket \end{aligned} \quad (2.3)$$

Le système d'équation obtenu s'écrit  $g'(t) = F(t, g(t))$  pour

$$F(t, x) := \langle f(t), e_i \rangle - c \left( \sum_{i=1}^n x_i e_i, \sum_{i=1}^n x_i e_i, e_i \right) - a \left( \sum_{i=1}^n x_i e_i, e_j \right) \quad (2.4)$$

$F$  est mesurable, continue et intégrable en temps sur tout compact en espace : le théorème de Carathéodory (2) s'applique et nous donne une solution locale. Par combinaison linéaire des problèmes (2.3), on montre que toutes les solutions issues d'un même  $u_0$  sont uniformément bornées.

**Remarque 4** *Dans le premier chapitre, la non-linéarité était contournée grâce à l'hypothèse de monotonie : ici, c'est simplement la propriété  $b(u, v, v) = 0$  qui permet de se ramener à un problème linéaire dans toute estimation d'énergie ou de norme.*

### Étape 2 (Borne sur la dérivée en temps)

La bornitude uniforme de la suite  $(u_m)_m$  permet d'en extraire une sous-suite convergeant faiblement vers un élément  $u$ . Le "point fondamental" de la démonstration (selon J.L. Lions) est de montrer que les termes  $\partial_t u_m$  sont uniformément bornés dans  $L^2(0, T; \mathbb{V}')$ . Le choix de la base  $(e_i)$  intervient pour assurer que  $\|P_m\|_{\mathcal{L}(\mathbb{V}_2, \mathbb{V}_2)} \leq 1$ , où  $P_m$  est le projecteur sur la base  $(e_i)_i$ . Grâce aux estimations uniformes précédentes,  $\partial_t u_m$  reste dans un borné de  $L^2(0, T; \mathbb{V}'_2)$ .

### Étape 3 (Convergence forte dans $L^2(0, T; \mathbb{H})$ )

La convergence faible ne permet pas de passer à la limite dans les termes non linéaires de la formulation variationnelle : l'argument utilisé est le suivant.

**Proposition 9 – Injection compacte** Soient  $B_0$ ,  $B$  et  $B_1$  trois espaces de Banach,  $B_i$  réflexif pour  $i = 0, 1$ , et tels que l'injection  $B_0 \hookrightarrow B$  soit compacte. On définit

$$W := \{v \in L^{p_0}(0, T; B_0) \mid \partial_t v \in L^{p_1}(0, T; B_1)\}$$

où  $T$  est fini et  $1 < p_0, p_1 < \infty$ . Alors l'injection  $W \hookrightarrow L^{p_0}(0, T; B)$  est compacte.

L'application de ce résultat à  $p_0 = p_1 = 2$ ,  $B_0 = \mathbb{V}$ ,  $B = \mathbb{H}$  et  $B_1 = \mathbb{V}'_2$  permet de récupérer une convergence forte de la suite  $(u_m)_m$ , et ainsi de passer à la limite dans les problèmes d'ordre  $m$ . La continuité en temps est obtenue en exhibant un représentant continu de la classe d'équivalence d'égalité presque partout de la limite obtenue.

### Étape 4 (Estimation d'énergie et unicité)

Comme dans le cas linéaire, l'unicité provient d'une inégalité d'énergie, qui s'obtient facilement en multipliant l'équation (1.1a) par  $u(t)$ , puis en intégrant sur le temps. La difficulté est de montrer que chaque terme est correctement défini, ce qui est fait en dimension 2, où  $\partial_t u \in L^2(0, T; \mathbb{V}')$  (voir [LM68]). Dans ce cas, la démonstration par l'absurde du cas linéaire s'adapte sans problème.  $\square$

#### 2.1.4 Problème adjoint

Supposons que l'on puisse simuler des écoulements par Navier-Stokes. L'étape suivante serait de contrôler ces écoulements, et comme préliminaire, nous nous intéressons au problème adjoint à l'équation des fluides.

## Différentiabilité

L'opérateur non linéaire de l'équation de Navier-Stokes ne rentre pas dans le cadre des opérateurs de Nemytskii, et sa différentiabilité doit faire l'objet d'une étude à part.

**Définition 15 – Notations** On se donne les applications à valeur dans les espaces duals

$$\begin{aligned} C : \mathbb{H} &\rightarrow \mathbb{H}', & u \rightsquigarrow (C(u)u, v) &= c(u, u, v) && \forall v \\ C' : \mathbb{H}^2 &\rightarrow \mathbb{H}', & u, b \rightsquigarrow (C'(u)b, v) &= c(u, b, v) + c(b, u, v) && \forall v \end{aligned}$$

et on définit la fonctionnelle entrée-sortie  $G : f \in L^2(0, T; [L^2(\Omega)]^2) \mapsto G(f) = u$ , qui sera parfois notée  $G(f) = u(f)$ .

Établissons un petit résultat préliminaire.

**Lemme 4** Soit  $u, h \in \mathbb{V}$ . L'opérateur  $C$  est différentiable de  $\mathbb{V}$  dans  $\mathbb{V}'$ , et sa dérivée directionnelle au point  $\bar{u}$  dans la direction  $h$  est donnée par  $C'(\bar{u})h$ .

### Démonstration

Soit  $h \in \mathbb{V} \setminus \{0\}$  et  $v \in \mathbb{V}$ . On a

$$\begin{aligned} (C(\bar{u} + h)(\bar{u} + h), v) &= c(\bar{u} + h, \bar{u} + h, v) = c(\bar{u}, \bar{u}, v) + c(\bar{u}, h, v) + c(h, \bar{u}, v) + c(h, h, v) \\ &= (C(\bar{u})\bar{u}, v) + (C'(\bar{u})h, v) + c(h, h, v) \end{aligned}$$

En invoquant  $|c(h, h, v)| \leq \|h\|_{\mathbb{V}}^2 \|v\|_{\mathbb{V}}$ , on obtient directement

$$\frac{\|C(\bar{u} + h)(\bar{u} + h) - C(\bar{u})\bar{u} - (C'(\bar{u})h, v)\|_{\mathbb{V}^*}}{\|h\|_{\mathbb{V}}} \leq \|h\|_{\mathbb{V}} \rightarrow 0$$

□

Nous pouvons maintenant énoncer la différentiabilité de l'application entrée-sortie elle-même.

**Proposition 10 – Différentiabilité de  $G$**   $G$  est Fréchet-différentiable dans  $Z := L^2(0, T; \mathbb{V})$ , et pour toute direction  $h \in Z$ , la dérivée directionnelle  $u := G'(\bar{u})h$  au point  $\bar{u}$  satisfait le problème linéarisé

$$\begin{aligned} \partial_t u - \nu \Delta u + C'(\bar{u})u &= h && \text{dans } Q \\ \tau_{\Sigma} u &= 0 && \text{dans } \Sigma \\ u(0, \cdot) &= 0 && \text{dans } \Omega \end{aligned}$$

On donne une preuve de ce résultat dans le cas où la viscosité satisfait une condition de la forme  $\nu - K\|\nabla G(f)\| > 0$  (voir [Hei98]). Une autre démonstration peut être trouvée dans [GHS91], qui n'impose pas de condition sur la viscosité : cependant, elle est plus difficile d'accès, et l'on réserve son étude pour de futurs travaux.

### Démonstration

Soient  $f$  et  $h$  dans  $Z := L^2([0, T], \mathbb{H}(\Omega))$ . On notera  $\bar{u}_f := G(f)$ .  $G$  est Fréchet-différentiable si

$$\lim_{\|h\| \rightarrow 0} \left( \frac{\|\bar{u}_{f+h} - \bar{u}_f - u\|_Z}{\|h\|_Z} \right) = 0$$

Soit  $v := \bar{u}_{f+h} - \bar{u}_f - u$ . Par différence,  $v$  satisfait le problème

$$\begin{aligned}\partial_t v - \nu \Delta v + C(\bar{u}_{f+h})\bar{u}_{f+h} - C(\bar{u}_f)\bar{u}_f - C'(\bar{u})u &= f + h - f - h = 0 && \text{dans } Q \\ \tau_\Sigma v &= 0 && \text{dans } \Sigma \\ v(0, \cdot) &= 0 && \text{dans } \Omega\end{aligned}$$

où la première ligne est équivalente à

$$\partial_t v - \nu \Delta v + C'(\bar{u})v = -(C(\bar{u}_{f+h})\bar{u}_{f+h} - C(\bar{u}_f)\bar{u}_f - C'(\bar{u})(\bar{u}_{f+h} - \bar{u}_f)) := g \in L^2(0, T; \mathbb{V}')$$

Par produit scalaire avec  $v$ , puis intégration sur  $t \in [0, \tau]$ , il vient

$$\frac{1}{2}\|v(\cdot, \tau)\|^2 + \nu \int_0^\tau \|v(\cdot, t)\|_1^2 + (C'(\bar{u}(\cdot, t))v(\cdot, t), v(\cdot, t)) dt = \int_0^\tau \langle g(\cdot, t), v(\cdot, t) \rangle dt$$

Remarquons que d'après la proposition (8),  $|C'(\bar{u})v, v| = |c(\bar{u}, v, v) + c(v, \bar{u}, v)| = |c(v, \bar{u}, v)| \leq K\|v\|_1^2\|\nabla \bar{u}\|$ . Supposons alors que  $\nu - K\|\nabla \bar{u}\| = \tilde{\nu} > 0$ . Dès lors,

$$\frac{1}{2}\|v(\cdot, \tau)\|^2 + \tilde{\nu} \int_0^\tau \|v(\cdot, t)\|_1^2 dt \leq \int_0^\tau \|g(\cdot, t)\|_{\mathbb{V}'} \|v(\cdot, t)\|_{\mathbb{V}} dt \leq \int_0^\tau \left( K_{\tilde{\nu}} \|g(\cdot, t)\|_{\mathbb{V}'}^2 + \frac{\tilde{\nu}}{2} \|v(\cdot, t)\|_{\mathbb{V}} \right) dt$$

et on en déduit les deux estimations

$$\begin{aligned}\int_0^\tau \|v(\cdot, t)\|_1^2 dt &\leq K_{\tilde{\nu}} \int_0^\tau \|g(\cdot, t)\|_{\mathbb{V}'}^2 dt && \text{grâce à l'hypothèse sur } \nu \\ \|v(\cdot, \tau)\|^2 &\leq K_{\tilde{\nu}} \|g\|_{Z'}^2 e^{\tilde{\nu}T/2} && \text{p.p } \tau \text{ par Grönwall, avec } T \text{ fini}\end{aligned}$$

ce qui permet de majorer la norme  $\|v\|_Z$  par  $K_{\tilde{\nu}, T}\|g\|_{Z'}$ . Or, par la différentiabilité de  $C$  établie au lemme (4), la norme de  $g$  dans l'espace  $Z = L^2(0, T; \mathbb{V}')$  est un  $o(\|\bar{u}_{f+h} - \bar{u}_f\|_{\mathbb{V}})$ . Par soustraction, le terme  $w := \bar{u}_{f+h} - \bar{u}_f$  satisfait

$$\begin{aligned}\partial_t w - \nu \Delta w + C(\bar{u}_{f+h})\bar{u}_{f+h} - C(\bar{u}_f)\bar{u}_f &= h \\ \frac{1}{2} \frac{d}{dt} \|w(\cdot, t)\|^2 + \nu \|w(\cdot, t)\|_1^2 + c(\bar{u}_{f+h}, \bar{u}_{f+h}, w) - c(\bar{u}_f, \bar{u}_f, w) &= (h, w)\end{aligned}$$

or  $b^2(b-a) - a^2(b-a) = (b-a)(b+a)(b-a)$ , d'où les termes en  $c$  sont égaux à  $c(w, \bar{u}_f, w)$ . Par l'exact même raisonnement que pour  $v$ , on déduit de l'hypothèse sur la viscosité (ou, de manière équivalente, sur la "variation lente" du champ de vitesse), de la finitude de  $T$  et du lemme de Grönwall que  $\|w\|_Z \leq C_{\tilde{\nu}, T}\|h\|_Z$ . Nous avons démontré que

$$\|v\|_Z = \|\bar{u}_{f+h} - \bar{u}_f - u\|_Z \leq K_{\tilde{\nu}, T}\|g\|_{Z'}^2 = o(\|\bar{u}_{f+h} - \bar{u}_f\|_{\mathbb{V}}) = o(\|h\|_Z)$$

d'où la Fréchet-différentiabilité de l'opérateur  $G$  dans  $Z$ . □

### Expression de l'adjoint

Le problème de Navier-Stokes peut s'écrire  $\mathcal{T}(u) = (f, 0, 0, u_0)$  pour un certain opérateur  $\mathcal{T}$  à valeurs dans  $L^2(0, T, [H^{-1}(\Omega)]^2) \times L^2(Q) \times L^2(\Sigma) \times L^2(\Pi)$ .

**Proposition 11 – Adjoint de Navier-Stokes** Soit  $(\bar{u}, \bar{p})$  dans l'espace des solutions. L'opérateur linéaire  $\partial_{(u,p)}\mathcal{T}(\bar{u}, \bar{p})$  admet un adjoint sur  $X_t \times L^2(0, T; [L_0^2(\Omega)]^2)$ , donné par

$$(\partial_{(u,p)}\mathcal{T}(\bar{u}, \bar{p}))^*(w, r) = \begin{pmatrix} -\partial_t w - \nu \Delta w + (\nabla \bar{u})^* w - \nabla w \cdot \bar{u} + \nabla r \\ -\operatorname{div} w \\ \tau_\Sigma w \\ w(T) \end{pmatrix}$$

## Démonstration

Le problème linéarisé au point  $(\bar{u}, \bar{p})$  dans la direction  $(u, p)$  s'écrit

$$\begin{aligned} \partial_t u - \nu \Delta u + \nabla \bar{u} \cdot u + \nabla u \cdot \bar{u} + \nabla p &= 0 && \text{dans } Q \\ \operatorname{div}(u) &= 0 && \text{dans } Q \\ \tau_\Sigma u &= 0 && \text{dans } \Sigma \\ u(\cdot, 0) &= 0 && \text{dans } \Omega \end{aligned}$$

Désignons la variable adjointe par  $(w, r)$ . Traitons premièrement les termes hérités de la non-linéarité :

$$\begin{aligned} (\nabla \bar{u} \cdot u + \nabla u \cdot \bar{u}, w)_Q &= (u, (\nabla \bar{u})^* w)_Q + \left( \sum_{i=1}^n \sum_{j=1}^n \bar{u}_i \frac{\partial u_j}{\partial x_i}, w_j \right)_Q \\ &= (u, (\nabla \bar{u})^* w)_Q + \sum_{i=1}^n \sum_{j=1}^n \left[ (\bar{u}_i u_j n_i, w_j)_\Sigma - (u_j, \frac{\partial \bar{u}_i}{\partial x_i} w_j + \frac{\partial w_j}{\partial x_i} \bar{u}_i)_Q \right] \\ &= (u, (\nabla \bar{u})^* w)_Q + \underbrace{(u, \bar{u} \cdot n w)_\Sigma - (u, \operatorname{div} \bar{u} w)_Q}_{=0 \text{ par propriété de } \bar{u}} - (u, \nabla w \cdot \bar{u})_Q \end{aligned}$$

Le choix de  $w$  à trace nulle permet de faire s'évanouir certains termes de bords, qui n'auraient pas pu être exprimés sous forme de produits scalaires avec  $u$ . Pour des raisons esthétiques, la condition de divergence nulle  $\operatorname{div} u = 0$  est comprise comme  $-\operatorname{div} u = 0$  : cela revient à changer le signe de  $r$ , et permet d'avoir une équation adjointe plus proche de celle d'origine. Poursuivons avec

$$\begin{aligned} (-\nu \Delta u, w)_Q &= -\nu (\nabla u \cdot n, w)_\Sigma + \nu (u, \nabla w \cdot n)_\Sigma - (u, \Delta w)_Q = -(u, \Delta w)_Q \\ (-\operatorname{div} u, r)_Q &= -(u \cdot n, r)_\Sigma + (u, \nabla r)_Q = (u, \nabla r)_Q \\ (\nabla p, w)_Q &= (p, w \cdot n)_\Sigma - (p, \operatorname{div} w)_Q = -(p, \operatorname{div} w)_Q \end{aligned}$$

et munis de ces petits calculs préliminaires, il est un peu fastidieux mais simple de voir que

$$\begin{aligned} &\left( \begin{pmatrix} \partial_t u - \nu \Delta u + \nabla \bar{u} \cdot u + \nabla u \cdot \bar{u} + \nabla p \\ -\operatorname{div} u \end{pmatrix}, \begin{pmatrix} w \\ r \end{pmatrix} \right)_Q + \left( \begin{pmatrix} \tau_\Sigma u \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_\Sigma w \\ \tau_\Sigma(r) \end{pmatrix} \right)_\Sigma + \left( \begin{pmatrix} \tau_{\{0\} \times \Omega} u \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_\Pi w \\ \tau_\Pi r \end{pmatrix} \right)_\Pi \\ &= \left( \begin{pmatrix} u \\ p \end{pmatrix}, \begin{pmatrix} -\partial_t w - \nu \Delta w + (\nabla \bar{u})^* w - \nabla w \cdot \bar{u} + \nabla r \\ -\operatorname{div} w \end{pmatrix} \right)_Q + \left( \begin{pmatrix} \tau_\Sigma u \\ \tau_\Sigma p \end{pmatrix}, \begin{pmatrix} \tau_\Sigma w \\ 0 \end{pmatrix} \right)_\Sigma + \left( \begin{pmatrix} \tau_\Pi u \\ \tau_\Pi p \end{pmatrix}, \begin{pmatrix} \tau_{\{T\} \times \Omega} w \\ 0 \end{pmatrix} \right)_\Pi \end{aligned}$$

ce qui nous donne le résultat attendu.  $\square$

### 2.1.5 Problème inhomogène au bord

Pour l'instant, une réalisation physique de notre problème ressemble à une boîte rigide fermée, contenant un seul fluide incompressible, et soumise à une force interne agissant sans média physique. On peut penser à une poche de magma sous l'influence de la chaleur, une expérience sur un fluide conducteur soumis à un champ magnétique, ou le chargement d'un camion-citerne soumis aux variations de vitesse du véhicule. Malgré ces exemples, il est naturel d'imaginer un flot non nul au bord du domaine. Considérons alors le

**Définition 16 – Problème inhomogène** Cherchons  $u \in X_t$  tel que

$$\begin{aligned} \partial_t u - \nu \Delta u + (u \cdot \nabla) u + \nabla p &= f && \text{dans } Q \\ \operatorname{div} u &= 0 && \text{dans } Q \\ \tau_\Sigma u &= g && \text{dans } \Sigma \\ u(\cdot, 0) &= u_0 && \text{dans } \Omega \end{aligned}$$

On s'appuie sur les arguments de [Lio69] pour le cas stationnaire. Remarquons d'abord que la condition au bord doit être compatible avec l'incompressibilité : par le théorème de Stokes,

$$\int_{\Omega} \operatorname{div} u(x) dx = \int_{\partial\Omega} \tau_{\Sigma} u(\sigma) \cdot n(\sigma) d\sigma = \int_{\partial\Omega} g \cdot n(\sigma) d\sigma = 0$$

Cette condition est satisfaite si  $g$  est la trace d'un rotationnel, c'est-à-dire s'il existe une fonction  $\Lambda \in \mathcal{C}([H^2(\Omega)]^2)$  telle que  $g(t) = \tau_{\Sigma} \operatorname{rot} \Lambda(t)$  : dès lors,  $\operatorname{div} \operatorname{rot} \Lambda = 0$ . Comme dans le cas linéaire, on se ramène à un problème homogène au bord en introduisant un relèvement  $G$  de  $g$ .

**Définition 17 – Relèvement** Nous cherchons  $G \in \mathcal{C}([0, T]; [H^1(\Omega)]^2)$  telle que

$$\operatorname{div} G = 0, \quad \tau_{\Sigma} G = g, \quad \partial_t G - \nu \Delta G + (G \cdot \nabla) G \in L^2(0, T; [H^{-1}(\Omega)]^2)$$

**Remarque 5** Nous pourrions prendre  $G = \operatorname{rot} \Lambda$ , et vérifier chacune des conditions au sens défini pour les solutions faibles de Navier-Stokes. Cependant, nous avons besoin de pouvoir contrôler la perturbation non linéaire induite par  $G$ , et ce n'est pas possible avec ce choix.

**Définition 18 – Problème perturbé** Posons  $u = y + G$ . Alors le problème consiste à chercher  $y \in X_t$  vérifiant

$$\begin{aligned} \partial_t \langle y(t), v \rangle + a(y(t), v) + c(y(t), y(t), v) + c(G(t), y(t), v) + c(y(t), G(t), v) &= \langle \tilde{f}(t), v \rangle \quad \forall v \in H_0^1(\Omega) \\ y(\cdot, 0) &= \tilde{u}_0 \end{aligned}$$

où  $\tilde{f} := f - (\partial_t G - \nu \Delta G + (G \cdot \nabla) G) \in L^2(0, T; [H^{-1}(\Omega)]^2)$ , et  $\tilde{u}_0 := u_0 - G(\cdot, 0) \in [L^2(\Omega)]^2$ .

Cette équation diffère de Navier-Stokes par l'ajout d'une forme linéaire continue  $\mu(u, v) := c(G(t), u, v) + c(u, G(t), v)$ . Les étapes de la démonstration de l'existence d'une solution peuvent être reprises, mais la perturbation va gêner les estimations uniformes sur la norme des solutions locales  $u_m(t)$  : plus précisément, évaluons la formulation variationnelle en  $v = y(t)$ . Il vient

$$\partial_t \langle y(t), y(t) \rangle + a(y(t), y(t)) = \langle \tilde{f}(t), y(t) \rangle - c(y(t), G(t), y(t))$$

Pour pouvoir se ramener aux estimations classiques, on cherche à contrôler  $c(y(t), G(t), y(t))$  par  $C(\varepsilon) \|y(t)\|_1^2$ , où  $C(\varepsilon)$  est arbitrairement petit. L'idée de Jacques-Louis Lions est la suivante.

**Proposition 12 – Choix du relèvement** Il existe un relèvement  $G$  admissible, au sens où

$$\exists C \in [0, 1[ \quad |c(v, G(t), v)| \leq C \|v\|_1^2 \quad \forall v \in [H_0^1(\Omega)]^2, t \in [0, T]$$

### Démonstration – (Idée)

Soit  $d(\Sigma, \cdot) : \Omega \rightarrow \mathbb{R}^+$  la distance au bord. Pour tout  $\varepsilon > 0$ , on introduit

$$\theta_{\varepsilon} \in \mathcal{C}^2(\Omega, \mathbb{R}), \quad \theta_{\varepsilon}(x) = \begin{cases} 1 & \text{dans un voisinage de } \Sigma \\ 0 & \text{si } d(\Sigma, x) \geq \exp(-1/\varepsilon) \end{cases}, \quad |\partial_{x_k} \theta_{\varepsilon}(x)| \leq \frac{\varepsilon}{d(\Sigma, x)} \quad \forall k \in \llbracket 1, n \rrbracket$$

Une telle fonction existe, et peut être construite par régularisation. On pose  $G = \operatorname{rot}(\theta_{\varepsilon} \Lambda)$ . Dès lors, pour

tout  $u \in [H^1(\Omega)]^2$ ,

$$|c(u, G(t), u)| = |c(u, u, G(t))| = \left| \int_{\Omega} (\nabla u G(t)) \cdot u \, dx \right| \leq \| \nabla u \| \sum_{i,j=1}^n \| u_i F_j(t) \|$$

Par hypothèse, là où  $F_i(t)$  est non nulle, on a l'estimation

$$|F_i(x, t)| \leq C (\| \nabla \theta_\varepsilon(x) \| |\Lambda_i(x, t)| + |\nabla \Lambda(x, t)|) \leq C \left( \frac{\varepsilon}{d(\Sigma, x)} |\Lambda_i(x, t)| + |\nabla \Lambda(x, t)| \right)$$

ce qui permet de majorer

$$\| u_i F_j(t) \| \leq C \left( \varepsilon \left\| \frac{u_i}{d(\Sigma, \cdot)} \right\| + \left( \int_{\text{supp } G} u_i^2 |\nabla \Lambda(t)|^2 \, dx \right)^{1/2} \right)$$

ce qui, en vertu des espaces choisis pour  $u$  et  $\Lambda$ , est une fonction continue par rapport à  $\varepsilon$ , tendant vers 0 quand  $\varepsilon \rightarrow 0$ . On peut donc rendre  $|c(u, G(t), u)|$  aussi petit que désiré.  $\square$

La preuve rigoureuse nécessite plusieurs lemmes, que l'on laisse le lecteur curieux découvrir dans [Lio69], chapitre 1, section 7. Pour ce travail, on admettra sans plus de justification le caractère bien posé du problème inhomogène. Dans la suite, on s'intéresse à l'étude numérique des solutions.

## 2.2 Du modèle aux schémas

### 2.2.1 Une autre formulation

Les espaces  $\mathbb{V}$  et  $\mathbb{H}$  sont complexes à discréteriser, en raison de la condition de divergence nulle. Pour contourner ce problème, on s'appuie sur la formulation complète de Navier-Stokes, qui emploie un multiplicateur de Lagrange "naturel" pour imposer l'incompressibilité. Il va cependant falloir s'assurer que le problème discret est bien posé, ce qui entraîne une condition dite de Ladyzhenskaya-Babuška-Brezzi (LBB) sur les espaces discrets.

**Définition 19 – Espaces** On définit

$$\begin{aligned} X &:= [H_0^1(\Omega)]^2 & X_t &:= L^2(0, T; X) \cap C([0, T]; [L^2(\Omega)]^2) \\ Y &:= [L_0^2(\Omega)]^2 = \left\{ p \in L^2(\Omega) \mid \int_{\Omega} p(x) \, dx = 0 \right\} & Y_t &:= L^2(0, T; Y) \end{aligned}$$

**Remarque 6** L'espace défini pour la pression permet de fixer le multiplicateur  $p$ , qui est défini à une constante près, puisque

$$b(v, p + \xi) = b(v, p) - \int_{\Omega} \operatorname{div}(v)(x) \xi \, dx = b(v, p) - \xi \int_{\Omega} \operatorname{div}(v)(x) \, dx = b(v, p) \quad \forall \xi \in \mathbb{R}$$

où, grâce au théorème de Stokes,  $\int_{\Omega} \operatorname{div}(v)(x) \, dx = \int_{\partial \Omega} v(x) \cdot n \, dx = 0$  pour tout  $v \in [H_0^1(\Omega)]^2$ .

Enfin, le choix de  $X$  fait perdre une propriété importante de  $c$  : l'antisymétrie par permutation des deux derniers arguments, ou  $c(u, v, w) = -c(u, w, v)$ . Pour conserver cette propriété dans le cadre numérique, on substitue à  $c$  la forme trilinéaire

$$\tilde{c}(u, v, w) := \frac{1}{2} (c(u, v, w) - c(u, w, v))$$

Naturellement,  $c$  et  $\tilde{c}$  coïncident sur l'espace de la solution exacte.

**Définition 20 – Formulation variationnelle pour la discréétisation** Le problème consiste à chercher un couple  $(u, p) \in X_t \times Y_t$  tel que

$$\begin{aligned} \partial_t(u(t), v) + a(u(t), v) + \tilde{c}(u(t), u(t), v) + b(v, p(t)) &= (f(t), v) & \forall v \in X \\ b(u(t), q) &= 0 & \forall q \in Y \\ u(0) &= u_0 \end{aligned}$$

## 2.2.2 Discréétisation

On choisit de discréétiser par différences finies en temps, et éléments finis en espaces. Nos schémas seront des propagateurs, au sens la suite  $(u^n, p^n)_n$  sera calculée par récurrence à partir de l'estimation de  $u_0$ . On verra une méthode directe en section (3.2.2), où toutes les approximations seront calculées simultanément.

### Semi-discréétisation en temps

Soit  $T$  un temps final fixé, et  $(\Delta t, N)$  tels que  $N\Delta t = T$ ,  $\Delta t > 0$  et  $N \in \mathbb{N}^*$ . Soit  $(u^n, p^n)_n \subset X \times Y$  une suite d'approximations de  $(u(n\Delta t), p(n\Delta t))$ . Le choix d'une formulation par récurrence n'est pas unique, et nous en considérerons deux, extraites de [Sch19].

**Définition 21 – Semi-schéma plutôt explicite** Pour tout  $n \in \llbracket 1, N \rrbracket$ ,

$$m\left(\frac{u^n - u^{n-1}}{\Delta t}, v\right) + a(u^n, v) + \tilde{c}(u^{n-1}, u^n, v) + b(v, p^n) = m(f^n, v) \quad \forall v \in X \quad (2.5a)$$

$$b(u^n, q) = 0 \quad \forall q \in Y \quad (2.5b)$$

$$u^0 = u_0 \quad (2.5c)$$

L'intérêt de ce schéma est que le terme non linéaire est constamment au second membre de l'équation, contrairement au second problème :

**Définition 22 – Semi-schéma plutôt implicite** Pour tout  $n \in \llbracket 1, N \rrbracket$ ,

$$m\left(\frac{u^n - u^{n-1}}{\Delta t}, v\right) + a(u^n, v) + \tilde{c}(u^{n-1}, u^n, v) + b(v, p^n) = m(f^n, v) \quad \forall v \in X \quad (2.6a)$$

$$b(u^n, q) = 0 \quad \forall q \in Y \quad (2.6b)$$

$$u^0 = u_0 \quad (2.6c)$$

Cette seconde proposition reste une équation linéaire en fonction de  $u^n$ , mais l'opérateur linéaire dépend de  $u^{n-1}$ . Remarquons que dans les deux schémas, la pression est une inconnue : c'est la raison pour laquelle la pression initiale n'est pas fixée. Le multiplicateur de Lagrange est un sous-produit du schéma numérique, et sa valeur n'est pas utilisée une fois calculée.

La mention "plutôt explicite" ou "plutôt implicite" est abusive, et ne sert qu'à différentier les deux schémas : malheureusement, au cours de la propagation, les deux nécessiteront la résolution d'un système linéaire. Pour l'obtenir, on poursuit la discréétisation.

### Discréétisation en espace

Le lecteur familier avec la méthode des éléments finis de Lagrange reconnaîtra les espaces classiques dans les définitions suivantes.

**Définition 23 – Maillage** On définit un maillage de  $\Omega$  comme une suite  $(K_i)_i$  finie vérifiant

$$K_i \text{ ouvert de } \Omega \quad \forall i, \quad \bigcup_i \overline{K_i} = \Omega, \quad K_i \cap K_j = \emptyset \quad \forall i \neq j$$

On se restreint au cas où le maillage est uniforme, au sens où il existe  $0 < \delta \leq \rho < \infty$  tels que  $B(x_i, \delta) \subset K_i \subset B(y_i, \rho)$  pour certains  $(x_i, y_i)$ . Cette condition assure une non-dégénérescence uniforme des éléments  $K_i$ , et permettra d'obtenir une inégalité "réciproque" à celle de Poincaré sur les espaces discrets.

**Définition 24 – Espaces éléments finis** On pose

$$\begin{aligned} X_h &= \{u \in X \mid (u_j)_{|K_i} \in \mathbb{P}^k(K_i) \quad \forall j \in \llbracket 1, n \rrbracket, \quad u \in [\mathcal{C}(\overline{\Omega})]^2\} \\ Y_h &= \{p \in Y \mid p_{|K_i} \in \mathbb{P}^l(K_i), \quad u \in [\mathcal{C}(\overline{\Omega})]^2\} \end{aligned}$$

On notera que nous considérons des espaces construits sur le même maillage, mais de degrés  $(k, l)$  *a priori* différents.

Nos variables complètement discrétisées seront la suite  $(u_h^n, p_h^n)_n \subset X_h \times Y_h$ , qui satisfont les équivalents discrets des schémas semi-discrétisés : pour  $k \in \{n, n-1\}$ ,

**Définition 25 – Schémas** Pour tout  $n \in \llbracket 1, N \rrbracket$ ,

$$m\left(\frac{u_h^n - u_h^{n-1}}{\Delta t}, v_h\right) + a(u_h^n, v_h) + \tilde{c}(u_h^{n-1}, u_h^n, v_h) + b(v_h, p_h^n) = m(f^n, v_h) \quad \forall v_h \in X_h \quad (2.7a)$$

$$b(u_h^n, q_h) = 0 \quad \forall q_h \in Y_h \quad (2.7b)$$

$$u_h^0 = \Pi_h(u_0) \quad (2.7c)$$

### 2.2.3 Condition LBB

On a vu que la pression est définie à une constante près, ce qui est imposé pour assurer l'unicité de la solution. Seulement, rien ne garantit que la discrétisation préserve cette propriété. Dans le cadre des problèmes où l'on cherche  $u$  tel que

$$\mu(u, v) = L(v) \quad \forall v$$

pour  $\mu, L$  multilinéaires continues sur un espace de Hilbert, l'unicité de la solution  $u$  vient de la coercivité de  $\mu$ , i.e. l'existence de  $\alpha > 0$  tel que  $\mu(u, u) \geq \alpha \|u\|^2$ . Le problème jouet similaire pour  $b$  consisterait à chercher  $p$  tel que

$$b(u, p) = L(u) \quad \forall u$$

mais la condition de coercivité n'a pas de sens, puisque l'on ne peut pas évaluer  $b(p, p)$  ou  $b(u, u)$  ! On introduit une autre condition.

**Définition 26 – Condition de Ladyzhenskaya-Babuška-Brezzi** On dira que  $b$  satisfait la condition LBB, ou condition inf-sup, s'il existe  $\beta > 0$  tel que

$$\inf_{p \in Y} \sup_{u \in X \setminus \{0\}} \frac{b(u, p)}{\|u\|_X \|p\|_Y} \geq \beta$$

**Remarque 7** Si  $\mu(\cdot, \cdot)$  est une forme bilinéaire continue sur un espace de Hilbert  $Z$ , la coercivité de  $\mu$  implique la condition LBB. En effet,

$$\sup_{z \in Z \setminus \{0\}} \frac{\mu(u, z)}{\|u\|_Z \|z\|_Z} \geq \frac{\mu(u, u)}{\|u\|_Z^2} \geq \alpha$$

et, par passage à l'inf sur  $u$ , on a directement  $\alpha = \beta$ . La réciproque est fausse : pour  $Z = \mathbb{R}^2$  muni du produit scalaire naturel, la forme bilinéaire

$$\mu(u, z) := u^t \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} z$$

satisfait bien la condition inf-sup pour  $\beta = 1$ , mais n'est pas coercive, puisque  $\mu(u, u) = 0$ .

Dans le cadre du problème de Navier-Stokes,  $b$  est fixée, et on dira plutôt que ce sont les espaces  $X_h$  et  $Y_h$  qui satisfont la condition. Remarquons que pour  $p_h \in Y_h$  fixé, l'application linéaire  $u_h \rightarrow b(u_h, p_h)$  est continue dans l'espace de Hilbert  $X_h \subset H^1(\Omega)$  : en effet,

$$|b(u_h, p_h)| = \left| \int_{\Omega} \operatorname{div} u_h(x) p_h(x) dx \right| \leq \|\nabla u_h\|_{L^2(\Omega)} \|p_h\|_{L^2(\Omega)} \leq C_p \|u_h\|_X$$

et, par le théorème de Riesz, il existe un unique  $Lp_h \in X_h$  tel que  $b(u_h, p_h) = (u_h, Lp_h)_X$ . On attire l'attention du lecteur sur le produit scalaire employé, qui est celui de  $[H^1(\Omega)]^2$ , et non celui de  $[L^2(\Omega)]^2$ .

**Définition 27 – Espace  $V_h$**  On définit  $V_h = \{v_h \in X_h \mid b(v_h, q_h) = 0 \quad \forall q_h \in Y_h\}$  un sous-espace vectoriel fermé de  $X_h$ .

Par définition,  $u_h^n \in V_h$  pour tout  $n$  : cet espace est un équivalent discret de  $\mathbb{V}$ .  $Y_h$  étant de dimension finie, c'est l'intersection (dénombrable) des noyaux des applications  $b(\cdot, q_h)$ , donc un fermé.

**Lemme 5** Si les espaces  $X_h$  et  $Y_h$  satisfont la condition LBB, l'application

$$\begin{aligned} L : Y_h &\rightarrow X_h \\ p_h &\mapsto Lp_h \quad \text{t.q. } b(u_h, p_h) = (u_h, Lp_h)_X \quad \forall u_h \in X_h \end{aligned}$$

est bijective de  $Y_h$  sur  $V_h^\perp := \{u_h \in X_h \mid (u_h, v_h)_X = 0 \quad \forall v_h \in V_h\}$ .

### Démonstration

$L$  est à valeurs dans  $V_h^\perp$ , car  $(v_h, Lp_h)_X = b(v_h, p_h) = 0 \quad \forall v_h \in V_h$ . Notons  $N$  la dimension de  $X_h$ , et  $M$  celle de  $Y_h$ . Par définition, on a

$$\dim(V_h^\perp) = N - \dim(V_h) = N - (N - \dim(Im(L))) = \dim(Im(L))$$

donc par linéarité, il suffit de montrer que  $L$  est injective. Supposons que  $Lp_h = L\tilde{p}_h$  : la condition LBB impose

$$\beta \|p_h - \tilde{p}_h\|_Y \leq \sup_{u \in X_h \setminus \{0\}} \frac{b(u, p_h - \tilde{p}_h)}{\|u\|_X} = \|Lp_h - L\tilde{p}_h\|_X = 0$$

par l'isométrie de Riesz. Donc  $p_h = \tilde{p}_h$ , et la bijectivité est démontrée.  $\square$

Sous cette condition, nos méthodes numériques ont un sens :

**Proposition 13 – Caractère bien posé des formulations discrétisées** Si  $X_h$  et  $Y_h$  satisfont la condition inf-sup, les schémas (2.7) admettent une unique solution  $(u_h^n, p_h^n) \subset X_h \times Y_h$  à chaque itération.

### Démonstration

Supposons  $u_h^{n-1}$  connu, et montrons que l'on peut en déduire  $(u_h^n, p_h^n)$ . Restreignons pour un moment l'espace des fonctions test  $v_h$  à  $V_h$ , qui est bien un espace de Hilbert. Les deux formulations variationnelles se réduisent à

$$m\left(\frac{u_h^n - u_h^{n-1}}{\Delta t}, v_h\right) + a(u_h^n, v_h) + \tilde{c}(u_h^{n-1}, u_h^k, v_h) = m(f^n, v_h) \quad \forall v_h \in V_h$$

On définit la forme bilinéaire continue sur  $V_h$

$$\mu(\cdot, \cdot) := \begin{cases} \frac{1}{\Delta t}m(\cdot, \cdot) + a(\cdot, \cdot) & \text{si } k = n-1 \\ \frac{1}{\Delta t}m(\cdot, \cdot) + a(\cdot, \cdot) + \tilde{c}(u_h^{n-1,h}, \cdot, \cdot) & \text{si } k = n \end{cases}$$

et grâce à l'antisymétrie de  $\tilde{c}$  par rapport à ses deux derniers arguments, dans les deux cas,  $\mu(v_h, v_h) = \frac{1}{\Delta t}m(v_h, v_h) + a(v_h, v_h) \geq \alpha \|v_h\|_X^2$ . Par Lax-Milgram, il existe une unique solution  $u_h^n \in V_h$ , et qui dépend continûment des données.

Revenons maintenant à la formulation complète pour  $v_h \in X_h$ . À  $u_h$  maintenant fixé, le problème est un système linéaire de la forme

$$b(v_h, p_h) = R(v_h) \quad \forall v_h \in X_h, \quad \text{où } R(v_h) := m(f^n, v_h) - \left[ m\left(\frac{u_h^n - u_h^{n-1}}{\Delta t}, v_h\right) + a(u_h^n, v_h) + \tilde{c}(u_h^{n-1}, u_h^k, v_h) \right]$$

$R$  est linéaire continue dans  $X$ , donc admet une représentation sous la forme  $R(v_h) = (v_h, r)_X$  pour  $r \in X$ . Remarquons que  $\forall v_h \in V_h$ ,  $(v_h, r)_X = b(v_h, r) = 0$ , donc  $r \in V_h^\perp$ . Le problème est donc équivalent à  $Lp_h = r$ , et par le lemme (5),  $L$  est bijective de  $Y_h$  sur  $V_h^\perp$ , et la solution  $p_h^n \in Y_h$  existe et est unique.  $\square$

## 2.3 Qualités numériques

### 2.3.1 Stabilité

**Proposition 14 – Stabilité conditionnelle du schéma plutôt explicite** Supposons  $f \in L^\infty(0, T; [L^2(\Omega)]^2)$ . Alors il existe une constante  $C > 0$  indépendante de  $h$  et  $\Delta t$  telle que la condition CFL  $\Delta t \leq C_0 h^2$  garantisse la stabilité du schéma  $k = n-1$ , au sens où il existe une constante  $K$  dépendant de  $u^0$ ,  $T$ ,  $\nu$ ,  $f$ , et  $C_0$  telle que

$$\|u_h^n\|_0^2 + \nu \Delta t \|\nabla u_h^n\|_0^2 \leq K \quad \forall n \Delta t \in ]0, T]$$

### Démonstration

Remarquons que si  $f \equiv u_0 \equiv 0$ , l'unique solution est la fonction nulle. Pour les cas restants, évaluons la formulation variationnelle (2.7) en  $k = n-1$  et  $v_h = u_h^n$ . Il vient

$$\begin{aligned} m\left(\frac{u_h^n - u_h^{n-1}}{\Delta t}, u_h^n\right) + a(u_h^n, u_h^n) + \tilde{c}(u_h^{n-1}, u_h^{n-1}, u_h^n) + b(u_h^n, p_h^n) &= m(f^n, u_h^n) \\ b(u_h^n, q_h) &= 0 \quad \forall q_h \in Y_h \\ u_h^0 &= \Pi_h(u_0) \end{aligned}$$

En utilisant l'identité  $(a - b)a = \frac{1}{2}((a - b)^2 + a^2 - b^2)$ , on obtient

$$m(u_h^n - u_h^{n-1}, u_h^n) = \frac{1}{2}\|u_h^n - u_h^{n-1}\|_0^2 + \frac{1}{2}\|u_h^n\|_0^2 - \frac{1}{2}\|u_h^{n-1}\|_0^2$$

On emploie maintenant la propriété d'antisymétrie de  $\tilde{c}$  : on a  $\tilde{c}(u_h^{n-1}, u_h^{n-1}, u_h^n) = \tilde{c}(u_h^{n-1}, u_h^{n-1}, u_h^n - u_h^{n-1})$ . Nous en sommes donc à

$$\frac{1}{2}\|u_h^n - u_h^{n-1}\|_0^2 + \frac{1}{2}\|u_h^n\|_0^2 - \frac{1}{2}\|u_h^{n-1}\|_0^2 + \nu\Delta t\|\nabla u_h^n\|_0^2 = \Delta t m(f^n, u_h^n) - \Delta t \tilde{c}(u_h^{n-1}, u_h^{n-1}, u_h^n - u_h^{n-1})$$

Quittons les égalités pour des majorations. Dans le terme  $m(f^n, u_h^n)$ , on applique l'inégalité d'Hölder  $ab \leq \frac{\varepsilon a^2}{2} + \frac{b^2}{2\varepsilon}$  en choisissant  $\varepsilon$  de manière à absorber  $\|u_h^n\|_0^2$  par le membre de gauche de l'équation. Pour le terme  $\tilde{c}$ , on aura besoin du lemme suivant (tiré de [Sch19]) :

**Lemme 6 – une autre majoration pour  $\tilde{c}$**  Supposons que l'espace  $X_h$  permette de satisfaire l'inégalité "inverse"  $\|\nabla u_h\|_0 \leq C/h\|u_h\|_0$  (ce qui est le cas pour un espace élément fini basé sur un maillage uniformément régulier). Alors

$$\tilde{c}(u_h, v_h, w_h) \leq C\|u_h\|_0^{1/2}\|\nabla u_h\|_0^{1/2}\|\nabla v_h\|_0\|w_h\|_0^{1/2}\|\nabla w_h\|_0^{1/2} \leq \frac{C}{h}\|u_h\|_0\|\nabla v_h\|_0\|w_h\|_0$$

Grâce à Poincaré sur le terme  $u_h^n - u_h^{n-1}$ , on en déduit

$$\|u_h^n - u_h^{n-1}\|_0^2 + \|u_h^n\|_0^2 - \|u_h^{n-1}\|_0^2 + (2-1)\nu\Delta t\|\nabla u_h^n\|_0^2 \leq \frac{\Delta t}{\nu}\|f^n\|_0^2 + 2\frac{\Delta t C}{h}\|u_h^{n-1}\|_0\|\nabla u_h^{n-1}\|_0\|u_h^n - u_h^{n-1}\|_0$$

Par la même inégalité d'Hölder, le terme  $\|u_h^n - u_h^{n-1}\|_0$  peut être absorbé par le membre de gauche : nous en sommes donc à

$$\|u_h^n\|_0^2 - \|u_h^{n-1}\|_0^2 + \nu\Delta t\|\nabla u_h^n\|_0^2 \leq \frac{\Delta t}{\nu}\|f^n\|_0^2 + C\left(\frac{\Delta t}{h}\right)^2\|u_h^{n-1}\|_0^2\|\nabla u_h^{n-1}\|_0^2$$

Par analogie avec le cas continu où l'on intègre en temps, sommons sur  $i \in \llbracket 1, n \rrbracket$ . On note  $\|f\|_{\infty,0}^2 := \sup_{n \geq 1} \|f^n\|_0^2$  une quantité bien définie par hypothèse. Dès lors,

$$\|u_h^n\|_0^2 - \|u_h^0\|_0^2 + \sum_{i=1}^n \nu\Delta t\|\nabla u_h^i\|_0^2 \leq \frac{T}{\nu}\|f\|_{\infty,0}^2 + 2\sum_{i=1}^n C\left(\frac{\Delta t}{h}\right)^2\|u_h^{i-1}\|_0^2\|\nabla u_h^{i-1}\|_0^2 \quad (2.8)$$

On conclut par récurrence, ce que l'on va formuler comme un lemme pour éclaircir les notations.

**Lemme 7** Soient  $(a_i)_i, (b_i)_i$  deux suites positives et  $\alpha, \beta, C$  trois constantes positives satisfaisant

$$a_n + \alpha \sum_{i=1}^n b_i \leq C + \beta \sum_{i=0}^{n-1} a_i b_i$$

Notons  $\lambda = C + a_0 b_0$ , et supposons que  $\alpha - \beta\lambda \geq 0$ . Alors pour tout  $n$ , on a  $a_n + \alpha b_n \leq \lambda$ .

### Démonstration

Le cas  $n = 1$  donne directement l'inégalité. Supposons que  $a_i + \alpha b_i \leq \lambda$  pour tout  $i \in \llbracket 0, n-1 \rrbracket$ . En particulier, on a  $a_i \leq \lambda$ . Ceci amène

$$a_n^2 + \alpha b_n + \sum_{i=1}^{n-1} b_i(\alpha - \beta\lambda) \leq a_n + \alpha b_n + \sum_{i=1}^{n-1} b_i(\alpha - \beta a_i) \leq C + a_0 b_0 = \lambda$$

et par la condition de positivité de l'énoncé,  $a_n + \alpha b_n \leq \lambda$ . □

Dans notre cas, pour  $a_n = \|u_h^n\|_0^2$ ,  $b_n = \|\nabla u_h^n\|_0^2$  et  $\lambda = \|u_h^0\|_0^2 + \frac{T}{\nu} \|f\|_{\infty,0}^2 + 2C \frac{\Delta t^2}{h^2} \|u_h^0\|_0^2 \|\nabla u_h^0\|_0^2$ , la condition de positivité donne la condition CFL

$$\nu \Delta t - 2C \left( \frac{\Delta t}{h} \right)^2 \lambda \geq 0 \implies \nu \frac{h^2}{2C\lambda} \geq \Delta t$$

sous laquelle le schéma est stable. Si  $f$  et  $u^0$  ne sont pas simultanément nulles,  $\lambda > 0$  et cette condition est correctement définie.  $\square$

**Remarque 8** La condition CFL, ou plus clairement l'équation (2.8), mettent en valeur le rôle joué par la viscosité  $\nu$ . Quand celle-ci tend vers 0, le terme  $T/\nu$  explode, et l'influence de  $\|\nabla u_h^i\|$  s'évanouit. Cela correspond à la perte de la régularité due au Laplacien, et n'est pas en désaccord avec l'interprétation physique des fluides très peu visqueux, sujets à un comportement chaotique (ou, à un changement d'échelle près, à des fluides soumis à de très grandes vitesses relativement à un obstacle).

Ce premier résultat est encourageant pour la simulation. Cependant, la proposition suivante a le double avantage d'être plus intéressante, et beaucoup plus légère à démontrer :

**Proposition 15 – Stabilité inconditionnelle du schéma plutôt implicite** Supposons  $f \in L^\infty(0, T; [L^2(\Omega)]^2)$ . Alors le schéma donné par  $k = n$  est inconditionnellement stable.

### Démonstration

Tout l'argument repose sur la propriété  $\tilde{c}(u_h^{n-1}, u_h^n, u_h^n) = 0$ . En suivant les mêmes étapes que la preuve précédente, sans terme non linéaire, on arrive à l'équation (2.8) :

$$\|u_h^n\|_0^2 + \sum_{i=1}^n \nu \Delta t \|\nabla u_h^i\|_0^2 \leq \frac{T}{\nu} \|f\|_{\infty,0}^2 + \|u_h^0\|_0^2$$

d'où le résultat.  $\square$

Pour terminer cette section, remarquons qu'un schéma pour l'équation de Stokes instationnaire (dans laquelle  $b$  est considérée nulle) vérifierait exactement les mêmes propriétés de stabilité que le schéma plutôt implicite. Nous pouvons maintenant nous pencher sur le véritable intérêt d'un schéma numérique.

### 2.3.2 Convergence

Cette section procède par étape. On s'intéresse d'abord au problème de Stokes stationnaire, pour lequel on établit des estimations d'erreur sous l'hypothèse que les espaces discrétisés  $X_h$  et  $Y_h$  permettent d'approcher  $X$  et  $Y$ . On poursuit en détaillant les étapes de la démonstration de convergence pour le schéma le plus implicite ( $k = n$ ).

#### Stokes stationnaire

Soit  $X$  l'espace de Hilbert choisi pour la vitesse, et  $Y$  celui pour la pression. On s'intéresse en premier lieu à la formulation variationnelle

$$\begin{aligned} a(u, v) + b(v, p) &= (f, v) & \forall v \in X \\ b(u, q) &= 0 & \forall q \in Y \end{aligned} \tag{2.9}$$

vérifiée par  $(u, p) \in X \times Y$  la solution exacte. Soient  $X_h \subset X$  et  $Y_h \subset Y$  deux espaces de Hilbert de dimension finie discrétilisant respectivement la vitesse et la pression. La solution approchée  $(u_h, p_h)$  est

définie par l'équation

$$\begin{aligned} a(u_h, v_h) + b(v_h, p_h) &= (f, v_h) \quad \forall v_h \in X_h \\ b(u_h, q_h) &= 0 \quad \forall q_h \in Y_h \end{aligned} \tag{2.10}$$

**Proposition 16** Notons  $\mathcal{P}_{X_h} u$  le projeté orthogonal de  $u$  sur  $X_h$ . Il existe une constante  $C > 0$  telle que

$$\|u - u_h\|_X + \|p - p_h\|_Y \leq C(\|u - \mathcal{P}_{X_h} u\|_X + \|p - \mathcal{P}_{Y_h} p\|_Y)$$

### Démonstration

Évaluons (2.9) en  $v = v_h$  et soustrayons les deux formulations variationnelles :

$$a(u - u_h, v_h) + b(v_h, p - p_h) = 0 \quad \forall v_h \in X_h \tag{2.11}$$

$$b(u - u_h, q_h) = 0 \quad \forall q_h \in Y_h \tag{2.12}$$

On procède en trois étapes.

**Majoration de l'erreur en vitesse par  $V_h$**  Soit  $V_h = \{v_h \in X_h \mid b(v_h, q_h) = 0 \forall q_h \in Y_h\}$  l'intersection des noyaux des applications  $v_h \mapsto b(v_h, q_h)$ . C'est un sous-espace vectoriel fermé de  $X_h$ , et on a les inclusions  $V_h \subset X_h \subset X$ . Par définition,  $u_h \in V_h$ . Commençons par majorer  $\|\mathcal{P}_{V_h} u - u_h\|_X$ . On évalue (2.11) en  $v_h = \mathcal{P}_{V_h} u - u_h$  :

$$a(u - \mathcal{P}_{V_h} u + \mathcal{P}_{V_h} u - u_h, \mathcal{P}_{V_h} u - u_h) + b(\mathcal{P}_{V_h} u - u_h, p - p_h) = 0$$

On a alors

$$\begin{aligned} \gamma_0 \|\mathcal{P}_{V_h} u - u_h\|_X^2 &\leq a(\mathcal{P}_{V_h} u - u_h, \mathcal{P}_{V_h} u - u_h) \quad \text{par coercivité de } a \\ &= -a(u - \mathcal{P}_{V_h} u, \mathcal{P}_{V_h} u - u_h) - b(\mathcal{P}_{V_h} u - u_h, p - p_h) \\ &= -a(u - \mathcal{P}_{V_h} u, \mathcal{P}_{V_h} u - u_h) - b(\mathcal{P}_{V_h} u - u_h, p - \mathcal{P}_{Y_h} p) - \underbrace{b(\mathcal{P}_{V_h} u - u_h, \mathcal{P}_{Y_h} p - p_h)}_{=0} \\ &\leq C (\|u - \mathcal{P}_{V_h} u\|_X \|\mathcal{P}_{V_h} u - u_h\|_X + \|\mathcal{P}_{V_h} u - u_h\|_X \|p - \mathcal{P}_{Y_h} p\|_Y) \quad \text{par continuité} \end{aligned}$$

d'où la première inégalité :

$$\|u - u_h\|_X \leq \|u - \mathcal{P}_{V_h} u\|_X + \|\mathcal{P}_{V_h} u - u_h\|_X \leq C (\|u - \mathcal{P}_{V_h} u\|_X + \|p - \mathcal{P}_{Y_h} p\|_Y)$$

L'utilisation du projeté sur  $V_h$ , et non sur  $X_h$ , est l'argument qui a permis de majorer par  $\|p - \mathcal{P}_{Y_h} p\|_Y$  (au lieu de  $\|p - p_h\|_Y$ ). Remarquons que les propriétés de la projection orthogonale ne sont pas employées : les arguments se tiennent pour n'importe quel élément de  $V_h \times Y_h$ .

**Majoration de l'erreur en vitesse par  $X_h$**  Décomposons  $X_h$  en  $V_h \oplus V_h^\perp$  au sens du produit scalaire de  $X$ . Par le lemme (5), l'isométrie de Riesz  $L : Y_h \rightarrow V_h^\perp$  est bijective. Soit alors  $\mathcal{P}_{X_h} u$  le projeté orthogonal de  $u$  sur  $X_h$  : on peut écrire  $\mathcal{P}_{X_h} u = \mathcal{P}_{V_h} u + \mathcal{P}_{V_h^\perp} u$ . Par bijectivité de  $L$ , il existe  $q_h \in Y_h$  tel que  $Lq_h = \mathcal{P}_{V_h^\perp} u$  : pour ce  $q_h$  particulier, on a

$$\sup_{v_h \in V_h \setminus \{0\}} \frac{b(v_h, q_h)}{\|v_h\|_X} = \sup_{v_h \in V_h \setminus \{0\}} \frac{(v_h, Lq_h)_X}{\|v_h\|_X} = \frac{(\mathcal{P}_{V_h^\perp} u, Lq_h)_X}{\|\mathcal{P}_{V_h^\perp} u\|_X} \geq \beta \|q_h\|_Y$$

d'où

$$\beta \|q_h\|_Y \|\mathcal{P}_{V_h^\perp} u\|_X \leq b(\mathcal{P}_{V_h^\perp} u, q_h) = b(\mathcal{P}_{V_h^\perp} u + \mathcal{P}_{V_h} u - u, q_h) = b(\mathcal{P}_{X_h} u - u, q_h) \leq C \|\mathcal{P}_{X_h} u - u\|_X \|q_h\|_Y$$

et la majoration

$$\|u - \mathcal{P}_{V_h} u\|_X = \|u - \mathcal{P}_{X_h} u + \mathcal{P}_{V_h^\perp} u\|_X \leq (1 + C) \|\mathcal{P}_{X_h} u - u\|_X$$

**Majoration de l'erreur en pression** Comme précédemment, on commence par majorer  $\|\mathcal{P}_{Y_h} p - p_h\|_Y$ . Par la condition inf-sup, et la compacité de la boule unité de  $V_h$ , il existe  $\xi_h \in X_h$  de norme 1 tel que

$$\beta \|\mathcal{P}_{Y_h} p - p_h\|_Y \leq b(\xi_h, \mathcal{P}_{Y_h} p - p_h) = -b(\xi_h, p - \mathcal{P}_{Y_h} p) - a(u - u_h, \xi_h) \leq C \|\xi_h\| (\|p - \mathcal{P}_{Y_h} p\|_Y + \|u - u_h\|_X)$$

En exploitant les résultats précédents, il vient

$$\begin{aligned} \|p - p_h\|_Y &\leq \|p - \mathcal{P}_{Y_h} p\|_Y + \|\mathcal{P}_{Y_h} p - p_h\|_Y \\ &\leq (1 + C) \|p - \mathcal{P}_{Y_h} p\|_Y + C \|u - u_h\|_X && \text{par l'estimation précédente} \\ &\leq C (\|u - \mathcal{P}_{X_h} u\|_X + \|p - \mathcal{P}_{Y_h} p\|_Y) && \text{par les deux premières étapes} \end{aligned}$$

et, par somme,  $\|u - u_h\|_X + \|p - p_h\|_Y \leq C (\|u - \mathcal{P}_{X_h} u\|_X + \|p - \mathcal{P}_{Y_h} p\|_Y)$ . On peut ainsi étudier l'erreur de convergence sous l'angle d'une erreur d'approximation de l'espace  $X \times Y$  par  $X_h \times Y_h$ .  $\square$

On fait maintenant l'hypothèse que nos espaces discrétilisés permettent d'approcher les espaces fonctionnels continus  $X$  et  $Y$ . Plus précisément, on fait l'hypothèse suivante.

**Hypothèse 4 (Approximation des espaces)** Il existe des projecteurs  $\Pi_h^1 \in \mathcal{L}([H^2(\Omega)]^2 \cap [H_0^1(\Omega)]^2, X_h)$  et  $\Pi_h^2 \in \mathcal{L}(H^1(\Omega) \cap L_0^2(\Omega), Y_h)$  tels que pour  $C_1 > 0, C_2 > 0$  deux constantes indépendantes de  $h$ ,

$$\begin{aligned} \|v - \Pi_h^1 v\|_1 &\leq C_1 h \|v\|_2 && \forall v \in [H^2(\Omega)]^2 \\ \|p - \Pi_h^2 p\|_0 &\leq C_2 h \|p\|_1 && \forall p \in H^1(\Omega) \end{aligned}$$

**Proposition 17** Supposons l'hypothèse (4) satisfaite. Alors il existe des constantes  $C > 0$  et  $C' > 0$  indépendantes de  $h$  telles que

$$\begin{aligned} \|u - u_h\|_1 + \|p - p_h\|_0 &\leq Ch(\|u\|_2 + \|p\|_1) \\ \|u - u_h\|_1 &\leq C'h^2(\|u\|_2 + \|p\|_1) \end{aligned}$$

### Démonstration

La première inégalité découle directement de l'hypothèse, et de la proposition (16). Pour la seconde, on introduit le problème dual consistant à trouver  $(\varphi_g, \xi_g) \in X \times Y$  tels que

$$\begin{aligned} a(v, \varphi_g) + b(v, \xi_g) &= (g, v) && \forall v \in X \\ b(\varphi_g, q) &= 0 && \forall q \in Y \end{aligned} \tag{2.13}$$

On se rappelle que

$$\begin{aligned} a(u - u_h, v_h) + b(v_h, p - p_h) &= 0 && \forall v_h \in X_h \\ b(u - u_h, q_h) &= 0 && \forall q_h \in Y_h \end{aligned} \tag{2.14}$$

On prend  $v = u - u_h$  dans (2.13). Par soustraction avec (2.14), il vient

$$\begin{aligned} a(u - u_h, \varphi_g - v_h) + b(u - u_h, \xi_g) - b(v_h, p - p_h) &= (g, u - u_h) && \forall v_h \in X_h \\ a(u - u_h, \varphi_g - v_h) + b(u - u_h, \xi_g - q_h) - b(\varphi_g - v_h, p - p_h) &= (g, u - u_h) && \forall v_h, q_h \in X_h \times Y_h \end{aligned}$$

où l'on a introduit les termes nuls  $b(u - u_h, q_h)$  et  $b(\varphi_g - v_h, p - p_h)$ . Afin de pouvoir majorer la norme de  $u - u_h$ , on se rappelle que  $\|u - u_h\|_X = \sup_{\|v\|_X=1} (v, u - u_h)_X$ . Pour pouvoir majorer sur la sphère unité, on divise par

la norme de  $g$ , puis on prend le sup sur  $g \in [L^2(\Omega)]^2$  : il vient

$$\begin{aligned} \sup_g \left( \frac{g}{\|g\|_X}, u - u_h \right) &= \sup_g \frac{1}{\|g\|_X} (a(u - u_h, \varphi_g - v_h) + b(u - u_h, \xi_g - q_h) - b(\varphi_g - v_h, p - p_h)) \\ &\leq \sup_g \frac{1}{\|g\|_X} (C \|u - u_h\|_X \|\varphi_g - v_h\|_X + C' \|u - u_h\|_X \|\xi_g - q_h\|_Y + C' \|\varphi_g - v_h\|_X \|p - p_h\|_Y) \\ &\leq C (\|u - u_h\|_X + \|p - p_h\|_Y) \times \sup_g \frac{1}{\|g\|_X} (\|\varphi_g - v_h\|_X + \|\xi_g - q_h\|_Y) \quad \forall v_h, q_h \in X_h \times Y_h \end{aligned}$$

d'où, en particulier,

$$\|u - u_h\|_X \leq C (\|u - u_h\|_X + \|p - p_h\|_Y) \times \sup_g \frac{1}{\|g\|_X} \left( \inf_{v_h \in X_h} \|\varphi_g - v_h\|_X + \inf_{q_h \in Y_h} \|\xi_g - q_h\|_Y \right)$$

Le problème dual correspond à un problème de Stokes, qui, sous l'hypothèse que  $\Omega$  est borné polygonal convexe, permet d'écrire  $\varphi_g \in [H^2(\Omega)]^2 \cap [H_0^1(\Omega)]^2$ ,  $\xi_g \in H^1(\Omega) \cap L_0^2(\Omega)$  et la continuité de la solution par rapport à  $g$ , i.e.

$$\|\varphi_g\|_2 + \|\xi_g\|_1 \leq C \|g\|_0$$

Les hypothèses d'approximation des espaces donnent alors

$$\frac{1}{\|g\|_X} \left( \inf_{v_h \in X_h} \|\varphi_g - v_h\|_X + \inf_{q_h \in Y_h} \|\xi_g - q_h\|_Y \right) \leq Ch \frac{1}{\|g\|_X} (\|\varphi_g\|_2 + \|\xi_g\|_1) \leq Ch$$

et la première estimation permet de conclure que

$$\|u - u_h\|_X \leq Ch^2 (\|u\|_2 + \|p\|_1)$$

□

Enfin, nous pouvons énoncer le résultat clef de cette section.

**Hypothèse 5** *La solution exacte est régulière au sens*

$$\begin{aligned} u &\in \mathcal{C}^0([0, T]; [W^{1,\infty}(\Omega)]^2 \cap X) \cap \mathcal{C}^1([0, T]; [L^2(\Omega)]^2 \cap H_0^1(\Omega)) \cap H^2(0, T; X) \\ p &\in H^1(0, T; H^1(\Omega) \cap Y) \end{aligned}$$

**Proposition 18 – Convergence du schéma implicite** Supposons les hypothèses (4) et (5) satisfaites. Alors il existe  $\bar{\Delta t} > 0$  et  $\bar{h} > 0$ , dépendant linéairement de  $\nu$ , tels que  $\forall \Delta t \leq \bar{\Delta t}$  et tout  $h \leq \bar{h}$ ,

$$\sup_n \|u(t^n) - u_h^n\|_1 \leq C (\|\Pi_h^1 u_0 - u_h^0\|_1 + h^2 + \Delta t)$$

Le lecteur suffisamment patient pour en arriver jusqu'ici appréciera peut-être notre choix de ne pas relater en détail la démonstration, plutôt longue et majoritairement technique, si bien développée dans [Sch19]. On en donne les étapes et les arguments, sans s'étendre sur les calculs.

### Étape 1 (Distinction des sources d'erreur)

La première étape consiste à décomposer les termes  $u - u_h$  et  $p - p_h$  en une composante d'approximation de l'espace  $u - \Pi_h^1 u$  et  $p - \Pi_h^2 p$ , et un terme d'écart à la projection  $\Pi_h^1 u - u_h$  et  $\Pi_h^2 p - p_h$ . Les premiers termes sont dépendants de la méthode numérique choisie, et c'est l'hypothèse (4) qui tient lieu de résultat. Concentrons-nous sur les termes  $w_h := \Pi_h^1 u - u_h$  et  $r_h := \Pi_h^2 p - p_h$ . Au terme de quelques manipulations,

on parvient à écrire

$$\|w_h^n - w_h^{n-1}\|^2 + \|w_h^n\|^2 - \|w_h^{n-1}\|^2 + \nu \Delta t \|\nabla w_h^2\|^2 \leq \|w_1^n\| (S_1 + S_2) + S_3 \quad (2.15)$$

où les trois sources d'erreur méritent chacune une étape.

#### Étape 2 ( $S_1$ : variation temporelle de l'erreur de projection)

Le premier terme est  $S_1 = \|\Pi_h^1(u(t^n) - u(t^{n-1})) - (u(t^n) - u(t^{n-1}))\|$ . Sous l'hypothèse de régularité  $u(\cdot, t) \in H^1(\Omega)$ , on s'attend à ce que cette norme diminue quand  $\Delta t$  diminue. Seulement pour ceci, il faut pouvoir écrire

$$\Pi_h \partial_t u = \partial_t \Pi_h u$$

ce qui est faux en général, même sur un exemple en dimension 1, pour un espace élément fini à un seul degré de liberté, un projecteur orthogonal et une fonction très régulière. Cette propriété découle d'un problème satisfait par la différence  $\Pi_h^1 u - u$ , dont la solution est unique : en dérivant ce problème, on conclut nécessairement que la dérivée temporelle et la projection spatiale commutent pour la solution exacte. Une fois ceci établi, la régularité  $C^1$  dans l'espace  $[H^2(\Omega)]^2$  demandée permet de faire des développements de Taylor et d'en évaluer les normes. Grâce à un lemme similaire à la proposition (16), on obtient la majoration

$$S_1 \leq C_{1,n} \sqrt{\Delta t} h^2$$

#### Étape 3 ( $S_2$ : convergence du schéma d'Euler)

Ce terme est donné par  $S_2 = \|\Delta t \frac{\partial u}{\partial t}(t^n) - (u(t^n) - u(t^{n-1}))\|_0$ . L'hypothèse de régularité  $C^1$  sur  $u$  permet, par développement de Taylor, d'obtenir

$$S_2 \leq \Delta t^{3/2} C_{2,n}$$

#### Étape 4 ( $S_3$ : convergence de $\tilde{c}$ )

Le troisième terme est  $S_3 = \Delta t (\tilde{c}(u(t^n), u(t^n), w_h^n) - \tilde{c}(u_h^{n-1}, u_h^n, w_h^n))$ . Moralement, si  $u_h^n$  tend bien vers  $u(t^n)$ , les deux derniers arguments convergent bien l'un vers l'autre : on a de plus l'intuition que  $u_h^{n-1}$  va se rapprocher de  $u(t^n)$  quand  $\Delta t$  diminue, grâce à l'hypothèse de continuité en temps dans l'espace des fonctions lipschitziennes. Cette dernière condition permet de mettre à jour une nouvelle estimation pour  $\tilde{c}$ , qui permet de majorer

$$S_3 \leq C \Delta t \left( \|w_h^n\|_0 + h^2 + \|w_h^n - w_h^{n-1}\|_0 + C_{3,n} \sqrt{\Delta t} + h \|w_h^{n-1}\|_1 \right) \|w_h^n\|_1$$

Pour nous résumer, l'estimation de (2.15) est maintenant réduite à

$$\begin{aligned} \|w_h^n - w_h^{n-1}\|^2 + \|w_h^n\|^2 - \|w_h^{n-1}\|^2 + \nu \Delta t \|\nabla w_h^2\|^2 &\leq \|w_1^n\| \left( C_{1,n} \sqrt{\Delta t} h^2 + \Delta t^{3/2} C_{2,n} \right) \\ &+ C \Delta t \left( \|w_h^n\|_0 + h^2 + \|w_h^n - w_h^{n-1}\|_0 + C_{3,n} \sqrt{\Delta t} + h \|w_h^{n-1}\|_1 \right) \|w_h^n\|_1 \end{aligned}$$

où les  $C_{i,n}$  sont dans  $l^2(\llbracket 0, N \rrbracket)$ . Tout repose donc sur l'estimation de  $w_h^n$  en normes  $\|\cdot\|_0 = \|\cdot\|$  et  $\|\cdot\|_1$ .

#### Étape 5 (Conclusion par lemme de Grönwall)

Le membre de droite de l'expression précédente coïncide avec la formulation variationnelle du problème de Stokes instationnaire, évaluée en  $w_h^n$ . Suivant cette idée, au terme de manipulations de type Hölder, Poincaré et sommation sur  $n$ , on en vient à l'inégalité

$$K_1 \|w_h^n\|_0^2 + K_2 \sum_{k=1}^n \|\nabla w_h^k\|^2 \leq C_0 + C(h^2 + \Delta t)^2 + K_3 \sum_{k=1}^{n-1} \|\nabla w_h^k\|^2$$

Les constantes  $K_1$  et  $K_2$  sont positives si  $\Delta t \leq C_1\nu$ , et  $h \leq C_2\nu$ . Sous cette hypothèse, on conclut que

$$\|w_h^n\|_0^2 + \delta\Delta t \sum_{k=1}^n \|\nabla w_h^k\|^2 \leq (\|w_h^0\|_1^2 + (h^2 + \Delta t)^2)$$

ce qui, si  $w_h^0 = \Pi_h^1 u_0 - u_h^0 = 0$ , donne la convergence du schéma.

**Remarque 9** *Comme dans le cas de la stabilité, la viscosité vient limiter les qualités numériques du schéma. La condition de convergence implique en particulier que quand  $\nu \rightarrow 0$ , les pas  $\Delta t$  et  $h$  doivent tendre vers 0 : ces schémas sont donc inapplicables à une équation limite.*

## Chapitre 3

# Applications

Cette section se divise en deux parties. Premièrement, on propose une implémentation des schémas pour Navier-Stokes, basée sur la librairie LIB obligamment fournie par M. Tonnoir. On s'intéresse ensuite à deux méthodes numériques pour le contrôle, l'une appliquée à Navier-Stokes, et l'autre développée en Matlab.

### 3.1 Implémentation de Navier-Stokes

La structure du code appelé `Insta` (instationnaire) qui accompagne ce rapport est relativement simple, et l'exécution suit les étapes suivantes :

#### Algorithme 1 – Déroulement d'une simulation

```
// En amont, l'utilisateur a fixé  $f$ ,  $g$  et  $u_0$ 
Construire les matrices du schéma
Projeter les conditions initiales
// Itérations
Pour chaque  $n$ 
    Projeter  $f^n$  et  $g^n$ 
    Résoudre le système linéaire associé au schéma
    Mesures d'erreurs, enregistrement des résultats
```

Bien que le code laisse libre le choix des degrés des éléments finis employés pour la vitesse et la pression, il est raisonnable de travailler avec  $\mathbb{P}^2/\mathbb{P}^1$  : ce couple d'espace minimise le degré (dans la famille  $\mathbb{P}^k/\mathbb{P}^l$  sans ajout de bulles ou autres subtilités) des choix qui respectent la condition inf-sup (voir [Sch19]). Le code est en C++, les maillages sont générés par GMSH, et la visualisation par Paraview.

Nous avons implémenté les deux schémas (2.7), ainsi que la possibilité de simuler un écoulement de Stokes instationnaire en annulant  $\tilde{c}$ .

#### 3.1.1 Systèmes linéaires

Cette section a pour but de guider le lecteur pas à pas dans la construction des systèmes linéaires effectivement implémentés. Dans un premier temps, on donne le système naturel associé à la formulation variationnelle sans terme de bord. Puis, on inclut la contrainte de moyenne nulle pour la pression, au moyen d'un multiplicateur de Lagrange. Enfin, on traite le cas inhomogène par la technique de pseudo-élimination. L'exposé complet des subtilités du code prendrait un volume considérable relativement à son intérêt : le courageux lecteur intéressé par une reprise de notre implémentation est invité à nous contacter directement.

### Cas homogène sans traitement de la pression

On se place dans le cas d'une triangulation  $(K_i)_i$  du domaine  $\Omega$ . Soit  $(x_i)_i$  (resp.  $(y_j)_j$ ) une famille de noeuds associée à un maillage éléments finis d'ordre  $d_u \in \mathbb{N}^*$  (resp.  $d_p \in \mathbb{N}^*$ ). On considère les bases Lagrangiennes

$$\begin{aligned}\Phi &= \{\phi_i, i \in \llbracket 1, N \rrbracket\}, & \phi_i(x_r) &= \delta_{i,r} \\ \Psi &= \{\psi_j, j \in \llbracket 1, M \rrbracket\}, & \psi_j(y_s) &= \delta_{j,s}\end{aligned}$$

Les inconnues  $u_h^n$  et  $p_h^n$  s'écrivent sous la forme

$$(u_h^n)_k = \sum_{i=1}^N U_{k,i}^n \phi_i, \quad k \in \{1, 2\}, \quad p_h^n = \sum_{j=1}^M P_j^n \psi_j$$

En évaluant les formulations variationnelles du schéma pour  $v_h$  parcourant la base  $\Phi$ , puis  $q_h$  parcourant la base  $\Psi$ , on déduit un système linéaire de taille  $2N + M$ . La disposition des coefficients dépend du choix de l'ordre des variables : pour clarifier l'exposé, on choisit ici de concaténer les abscisses des vecteurs  $u^n$ , puis leurs ordonnées, puis les variables de pression. Dans l'implémentation, les variables de la vitesse alternent dans l'espoir d'augmenter la stabilité : c'est ici parfaitement équivalent.

On prend pour exemple la formulation du schéma plutôt implicite ( $k = n$ ). Une première formulation serait

$$\begin{pmatrix} \frac{1}{\Delta t} \mathbb{M} + \mathbb{A} + \tilde{\mathbb{C}}_1^{n-1} & 0 & \mathbb{B}_1 \\ 0 & \frac{1}{\Delta t} \mathbb{M} + \mathbb{A} + \tilde{\mathbb{C}}_2^{n-1} & \mathbb{B}_2 \\ \mathbb{B}_1^t & \mathbb{B}_2^t & 0 \end{pmatrix} \begin{pmatrix} U_1^n \\ U_2^n \\ P^n \end{pmatrix} = \begin{pmatrix} \mathbb{M}(f_1^n + \frac{1}{\Delta t} U_1^{n-1}) \\ \mathbb{M}(f_2^n + \frac{1}{\Delta t} U_2^{n-1}) \\ 0 \end{pmatrix} \quad (3.1)$$

où

$$\begin{aligned}\mathbb{M}_{i,j} &= m\left(\begin{pmatrix} \phi_i \\ 0 \end{pmatrix}, \begin{pmatrix} \phi_j \\ 0 \end{pmatrix}\right) = m\left(\begin{pmatrix} 0 \\ \phi_i \end{pmatrix}, \begin{pmatrix} 0 \\ \phi_j \end{pmatrix}\right), & \mathbb{A}_{i,j} &= a\left(\begin{pmatrix} \phi_i \\ 0 \end{pmatrix}, \begin{pmatrix} \phi_j \\ 0 \end{pmatrix}\right) = a\left(\begin{pmatrix} 0 \\ \phi_i \end{pmatrix}, \begin{pmatrix} 0 \\ \phi_j \end{pmatrix}\right) \\ (\mathbb{B}_1)_{i,j} &= b\left(\begin{pmatrix} \phi_i \\ 0 \end{pmatrix}, \psi_j\right), & (\mathbb{B}_2)_{i,j} &= b\left(\begin{pmatrix} 0 \\ \phi_i \end{pmatrix}, \psi_j\right) \\ (\tilde{\mathbb{C}}_1^{n-1})_{i,j} &= \tilde{c}\left(u_h^{n-1}, \begin{pmatrix} \phi_j \\ 0 \end{pmatrix}, \begin{pmatrix} \phi_i \\ 0 \end{pmatrix}\right), & (\tilde{\mathbb{C}}_2^{n-1})_{i,j} &= \tilde{c}\left(u_h^{n-1}, \begin{pmatrix} 0 \\ \phi_j \end{pmatrix}, \begin{pmatrix} 0 \\ \phi_i \end{pmatrix}\right)\end{aligned}$$

**Remarque 10** L'assemblage des matrices est facilité par le choix d'une triangulation : chaque élément  $K_i$  est alors affine-équivalent à un élément de référence  $\hat{K}$ . On pourra procéder en plusieurs étapes :

- Calculer les formes multilinéaires sur les restrictions des fonctions de base à l'élément de référence. Pour  $\tilde{c}$ , cela implique de calculer toutes les combinaisons non nulles d'arguments  $(\phi_i \ 0)^t$  et  $(0 \ \hat{\phi}_i)^t$ , ce qui, pour une base de  $N$  éléments, s'élève à  $2N^2(N - 1)$  coefficients.
- En déduire les valeurs des formes multilinéaires sur les restrictions des fonctions de base à chaque élément  $K_i$ , grâce à un changement de variable affine. On prendra garde aux formes d'ordre 1, qui demandent de pouvoir appliquer la règle de la chaîne, donc de disposer séparément des dérivées par rapport à chaque variable spatiale des fonctions de référence.
- Assembler les matrices, ce qui se fait classiquement en sommant les contributions de chaque élément  $K_i$  aux formes bilinéaires des noeuds adjacents.

### Traitement de la moyenne nulle

On impose la nullité de l'intégrale de la pression par un paramètre de Lagrange  $\lambda$ . Notons  $Z = (U, P)^t$ , et  $\mathbb{H}Z = F$  le système linéaire (3.1). Résoudre ce système revient à minimiser la forme quadratique

$$J := \frac{1}{2} Z^t \mathbb{H}Z - F^t Z$$

Soit  $S$  le vecteur tel que  $S^t P = \int_{\Omega} p_h(x) dx$ , de coordonnées  $S_j = \int_{\Omega} \psi_j(x) dx$ . On forme le Lagrangien associé à la contrainte  $S^t P = 0$  :

$$\mathcal{L} := J - \lambda S^t P = \frac{1}{2} Z^t \mathbb{H} Z - F^t Z - \lambda S^t P$$

La minimisation sans contrainte du Lagrangien produit le système

$$\begin{cases} HZ - F - \lambda(0, S)^t = 0 \\ -S^t P = 0 \end{cases} \iff \begin{pmatrix} \mathbb{H} & \begin{pmatrix} 0 \\ S \end{pmatrix} \\ (0, S^t) & 0 \end{pmatrix} \begin{pmatrix} Z \\ \lambda \end{pmatrix} = \begin{pmatrix} F \\ 0 \end{pmatrix}$$

et c'est ce nouveau système linéaire, augmenté d'une variable, que l'on résoudra pendant les itérations.

**Remarque 11** *Cette proposition n'est pas la seule manière d'imposer une intégrale nulle à la pression. On peut se donner un noeud fictif, dont la valeur compensera l'intégrale totale : on peut également résoudre le système linéaire sans imposer de contrainte, et centrer le résultat obtenu. Pour notre implémentation, le petit minimiseur s'est avéré tout à fait satisfaisant.*

### Pseudo-élimination

Soit  $\Lambda \in \mathcal{C}([0, T]; [H^2(\Omega)]^2)$ . Sous notations de la section (2.1.5), nous cherchons un relèvement  $G$  sous la forme  $G = \text{rot}(\theta_\varepsilon \Lambda)$ .

**Proposition 19 – Discrétisation du relèvement** Le vecteur  $\mathbf{g} = (\mathbf{g}_1, \mathbf{g}_2)^t$  défini par

$$(\mathbf{g}_k)_i := \begin{cases} (\text{rot } \Lambda)_k(x_i) & \text{si } x_i \text{ est un noeud du bord} \\ 0 & \text{si } x_i \text{ est un noeud interne} \end{cases}$$

est l'interpolation d'un relèvement  $G$  admissible.

### Démonstration

En effet, soit  $\varepsilon_0 > 0$  tel que  $F_0 := \text{rot}(\theta_{\varepsilon_0} \Lambda)$  soit admissible. Le maillage de  $\Omega$  étant uniforme, il existe un  $\delta > 0$  tel que pour tout  $x_i$  un noeud interne du maillage,  $d(\Sigma, x_i) \geq \delta$ . Soit  $\varepsilon_1$  tel que  $\exp(-1/\varepsilon_1) \leq \delta/2$  : le relèvement  $F_{\varepsilon_1}$  associé est nul en chaque noeud interne. On considère le relèvement associé à  $\min(\varepsilon_0, \varepsilon_1)$  : il est admissible, et son interpolation est bien  $\mathbf{g}$ .  $\square$

On considère maintenant la décomposition  $u_h = y + \mathbf{g}$ , où  $u$  est notre véritable inconnue, non homogène,  $\mathbf{g}$  est fixé, et  $y \in X_t$  est la solution d'un problème homogène au bord qui sera effectivement résolu. La pseudo-élimination consiste à séparer les variables en deux classes : les coordonnées  $y_i$  associées à des noeuds internes, et celles  $y_b$  associées à des noeuds du bord.

En substituant  $u = y + \mathbf{g}$  dans les schémas (2.7), on obtient le système satisfait par les coordonnées internes  $y_i$ , posé pour  $(v_h, q_h) \in X_h \times Y_h$  :

$$\begin{aligned} m\left(\frac{y_i^n + \mathbf{g}^n - (y_i^{n-1} + \mathbf{g}^{n-1})}{\Delta t}, v_h\right) + a(y_i^n + \mathbf{g}^n, v_h) + \tilde{c}(y_i^{n-1} + \mathbf{g}^{n-1}, y_i^k + \mathbf{g}^k, v_h) + b(v_h, p_h^n) &= m(f^n, v_h) \\ b(y_i^n + \mathbf{g}^n, q_h) &= 0 \\ y_i^0 + \mathbf{g}^0 &= \Pi_h u_0 \end{aligned}$$

Il n'est techniquement pas nécessaire d'assembler les lignes du système linéaire pour les coordonnées  $y_b$ . Cependant, c'est plus simple. Cet argument imparable fait que sur ces coordonnées, on impose  $\alpha \mathbb{I} y_b = 0$ ,

c'est-à-dire que le système linéaire s'écrira (à une permutation des variables près, ce qui rend la pseudo-élimination obscure quand non explicitée)

$$\begin{pmatrix} \mathbb{H} & 0 \\ 0 & \alpha\mathbb{I} \end{pmatrix} \begin{pmatrix} y_i \\ y_b \end{pmatrix} = \begin{pmatrix} F \\ 0 \end{pmatrix}, \quad \alpha \text{ un réel strictement positif}$$

Une méthode non optimale, mais aisée, est d'assembler le système linéaire complet, puis de remplacer toutes les lignes / colonnes associées à des coordonnées du bord par  $\alpha\mathbb{I}$ , où  $\alpha$  est choisi de manière à ne pas perturber le conditionnement du système. On se donne pour ceci le petit lemme suivant :

**Lemme 8 – Conditionnement** Supposons  $\mathbb{H}$  inversible. Alors

$$\alpha = \sqrt{\frac{\text{tr}(\mathbb{H}^t \mathbb{H})}{\text{card}(X_h)}} \implies \text{cond} \begin{pmatrix} \mathbb{H} & 0 \\ 0 & \alpha\mathbb{I} \end{pmatrix} = \text{cond}(\mathbb{H})$$

où le conditionnement considéré est subordonné à la norme spectrale, i.e.  $\text{cond}(\mathbb{H}) = \|\mathbb{H}\|_2 \|\mathbb{H}^{-1}\|_2$ .

### Démonstration

Pour  $\mathbb{H}$  inversible, le conditionnement est donné par  $\text{cond}(\mathbb{H}) = \sqrt{\frac{\sigma_n}{\sigma_1}}$ , où  $\sigma_1$  est la plus petite valeur propre de la matrice symétrique définie positive  $\mathbb{H}^t \mathbb{H}$ , et  $\sigma_n$  la plus grande. D'autre part, le spectre de la matrice  $\begin{pmatrix} \mathbb{H}^t \mathbb{H} & 0 \\ 0 & \alpha^2 \mathbb{I} \end{pmatrix}$  est donné par  $\text{sp}(\mathbb{H}^t \mathbb{H}) \cup \{\alpha^2\}$ , donc choisir  $\alpha^2$  comme une combinaison convexe des valeurs propres de  $\mathbb{H}^t \mathbb{H}$  permet de ne pas modifier  $\sigma_1$  et  $\sigma_n$ . Le choix de  $\alpha$  proposé est donc simplement la moyenne des valeurs propres.  $\square$

Une fois ceci fait, la propagation est conduite en variables  $y$ , et le résultat final est  $y + \mathbf{g}$ .

### 3.1.2 Validation du code

Cette rapide section présente les cas tests sur lesquels le code **Insta** a été validé. On considère  $\Omega = [0, 1]^2$ , et on construit une famille d'exemples en imposant une solution de la forme suivante : soit  $\phi(x, y, t) = (p(x) + q(y))\tau(t)$ . On pose

$$\begin{aligned} u(x, y, t) &:= \text{rot } \phi(x, y, t) = \begin{pmatrix} -q'(y)\tau(t) \\ p'(x)\tau(t) \end{pmatrix} & \partial_t u(x, y, t) &= \begin{pmatrix} -q'(y)\tau'(t) \\ p'(x)\tau'(t) \end{pmatrix} \\ (u \cdot \nabla)u(x, y, t) &= \begin{pmatrix} -q''(y)p'(x)\tau^2(t) \\ -q'(y)p''(x)\tau^2(t) \end{pmatrix} & \Delta u(x, y, t) &= \begin{pmatrix} -q^{(3)}(y)\tau(t) \\ p^{(3)}(x)\tau(t) \end{pmatrix} \end{aligned}$$

La solution  $u$  est à divergence nulle par construction. L'équation de Navier-Stokes donnera

$$\begin{pmatrix} -q'(y)\tau'(t) \\ p'(x)\tau'(t) \end{pmatrix} + \begin{pmatrix} -q''(y)p'(x)\tau^2(t) \\ -q'(y)p''(x)\tau^2(t) \end{pmatrix} - \nu \begin{pmatrix} -q^{(3)}(y)\tau(t) \\ p^{(3)}(x)\tau(t) \end{pmatrix} + \nabla p = f$$

Le choix de  $\nabla p$  et  $f$  est libre, sous la contrainte que  $\nabla p$  corresponde effectivement au gradient d'une quantité. Dans les exemples de test suivants, on prendra par défaut  $\nu = 0.1$ , des éléments finis  $\mathbb{P}^2/\mathbb{P}^1$ , un temps final de  $T = 1$ , 30 itérations, et une même gamme de maillages sur  $\Omega = [0, 1]^2$ , correspondant à respectivement 41, 74, 132, 185 et 220 noeuds du maillage fourni par GMSH. On a effectué les mesures d'erreur avec une approximation de la norme  $L^1([0, T]; [L^2(\Omega)]^2)$ , discrétisée par différences finies en temps et quadrature en espace.

### Test 1 - Poiseuille stationnaire

On choisit

$$p(x) = 0, \quad q(y) = \frac{y^3}{3} - \frac{y^2}{2}, \quad \tau(t) = 1, \quad p(x, y) = -2\nu x + \nu, \quad f = 0$$

La solution exacte correspond au profil parabolique de Poiseuille pour un écoulement horizontal :

$$u(x, y) = \begin{pmatrix} y(1-y) \\ 0 \end{pmatrix}$$

La composante non linéaire s'annule, ce qui fait que les résultats de Stokes et Navier-Stokes sont identiques. On a pu constater que pour  $u_0 = u(0, \cdot)$ , le schéma conserve la solution exacte (stationnaire). De plus, la solution se décompose exactement dans la base élément finis employée : l'erreur est constante par rapport au pas  $h$ . La solution est donnée par la figure (3.1).

### Test 2 - Stationnaire, $(u \cdot \nabla)u$ non nulle

On choisit

$$p(x) = x, \quad q(y) = y^3, \quad \tau(t) = 1, \quad p(x, y) = (y - 6\nu x) - (1/2 - 3\nu), \quad f(x, y) = \begin{pmatrix} -6y\mathbf{1}_{\{\text{Navier-Stokes}\}} \\ 1 \end{pmatrix}$$

où le choix de  $f$  conserve le même gradient de pression entre les modèles de Stokes et Navier-Stokes. Encore une fois, les solutions exactes se décomposent dans la base choisie, et l'erreur est constante en fonction du pas.

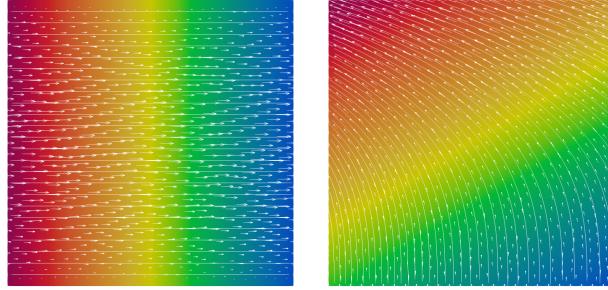


Figure 3.1: Solutions des exemples 1 et 2. Les flèches indiquent  $u$ , et la couleur de fond représente  $p$ .

### Test 3 - Instationnaire

On cherche à construire une rotation du flux. On considère  $T = 3$  et les données de Poiseuille stationnaire pour  $t \in [0, 1]$ , puis une rotation de  $i$  sur  $t \in [1, 2]$ , et les données d'un Poiseuille "vertical" stationnaire sur  $t \in [2, 3]$ . On veillera à ce que le champ de pression suive (même si c'est artificiel) la rotation imposée au champ de vitesse.

Les données sont les suivantes :

$$\begin{aligned} u(x, y, t) &= \begin{pmatrix} y(1-y) \\ 0 \end{pmatrix} \mathbf{1}_{t \in [0, 1]} + \begin{pmatrix} \cos(\frac{\pi}{2}(t-1))y(1-y) \\ \sin(\frac{\pi}{2}(t-1))x(1-x) \end{pmatrix} \mathbf{1}_{t \in [1, 2]} + \begin{pmatrix} 0 \\ x(1-x) \end{pmatrix} \mathbf{1}_{t \in [2, 3]} \\ p(x, y, t) &= (-2\nu x + \nu) \left( \mathbf{1}_{t \in [0, 1]} + \cos\left(\frac{\pi}{2}(t-1)\right) \mathbf{1}_{t \in [1, 2]} \right) + (-2\nu y + \nu) \left( \sin\left(\frac{\pi}{2}(t-1)\right) \mathbf{1}_{t \in [2, 3]} + \mathbf{1}_{t \in [2, 3]} \right) \\ f(x, y, t) &= \begin{pmatrix} -\frac{\pi}{2} \sin(\frac{\pi}{2}(t-1))y(1-y) \\ \frac{\pi}{2} \cos(\frac{\pi}{2}(t-1))x(1-x) \end{pmatrix} \mathbf{1}_{t \in [1, 2]} + \frac{1}{2} \sin(\pi(t-1)) \begin{pmatrix} (1-2y)x(1-x) \\ (1-2x)y(1-y) \end{pmatrix} \mathbf{1}_{t \in [1, 2] \cap \text{Navier-Stokes}} \end{aligned}$$

Comme précédemment, le terme source s'adapte au modèle pour conserver un même gradient de pression. Le lecteur est invité à consulter les vidéos `test3_Stokes` et `test3_NavierStokes` pour observer les résultats.

### 3.1.3 Explorations

La solution exacte n'est pas connue de l'auteure pour les résultats suivants : on les présente sans autre garantie que la confiance en la robustesse du code. On considère un domaine rectangulaire dans lequel circule un courant avec un profil de Poiseuille. On introduit un obstacle muni de conditions de Dirichlet homogène sur son bord.

**Obstacle** Nous avons testé les schémas de Stokes et Navier-Stokes pour des paramètres identiques. On a pris  $T = 1$  et  $\nu = 0.1$ . Les deux simulations convergent rapidement vers les deux premiers profils de la figure (3.2).

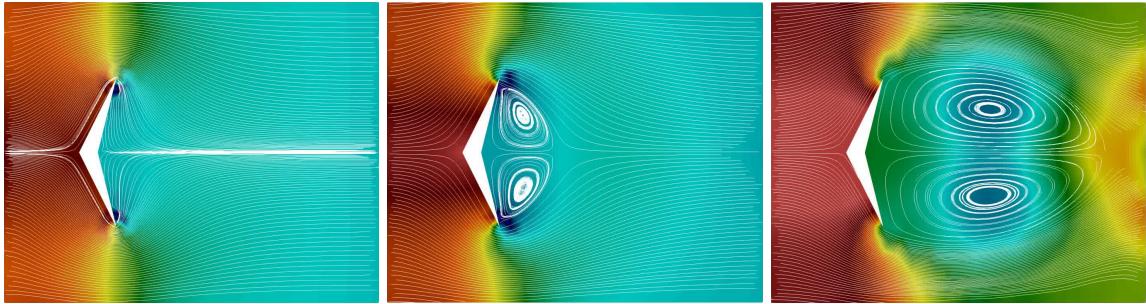


Figure 3.2: De gauche à droite : Stokes et Navier-Stokes pour  $\nu = 0.1$ , puis Navier-Stokes pour  $\nu = 0.01$

Les lignes de courant sont calculées par Paraview, et leur précision ne reflète pas la finesse du maillage. On constate que le modèle de Stokes ne crée pas de tourbillons. La même simulation, pour le modèle de Navier-Stokes avec une viscosité de  $\nu = 0.01$ , produit le troisième résultat de la figure (3.2). Le sillage est beaucoup plus allongé, et la condition de Dirichlet au bord droit de l'image (qui demande un profil de Poiseuille stationnaire sortant) entre en conflit avec la rotation du fluide, augmentant la pression dans cette zone.

Le lecteur curieux pourra observer la formation des tourbillons dans les vidéos commençant par `obstacle`.

**Allée de Von Kàrmàn** Le second exemple présenté est plus ambitieux. Il existe un phénomène physique appelé Allée de Bénard-Von Kàrmàn (Vortex Street), mis en lumière de manière spectaculaire dans An Album of Fluid Motion ([van82]). Le sillage d'un objet "de petite dimension" voit s'alterner des tourbillons de sens contraire. La formation des tourbillons ne dépend que du nombre de Reynolds, qui se déduit d'une longueur caractéristique, d'une vitesse caractéristique et de la viscosité : aussi, le phénomène expérimental à l'échelle du mètre pour un liquide s'observe dans le sillage des îles californiennes au travers des nappes nuageuses.

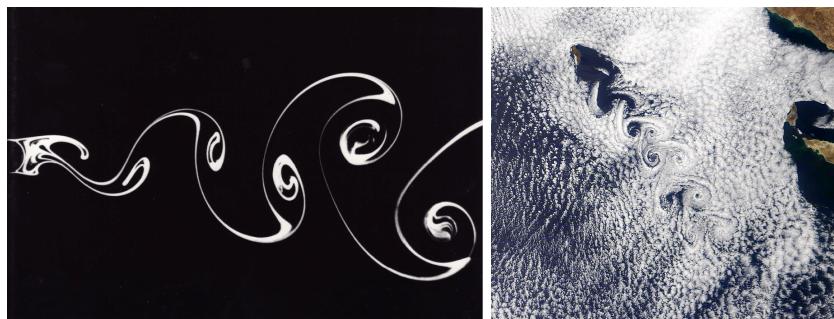


Figure 3.3: Allée de Bénard-Von Kàrmàn (sources : [gauche](#) et [droite](#))

On demande au lecteur un effort d'abstraction pour juger de la simulation que nous avons pu obtenir. Nous avons introduit une discréétisation relativement grossière d'un cylindre dans un domaine rectangulaire, et raffiné (un peu) le maillage dans la zone de sillage pour atteindre 327 noeuds (donc 1254 valeurs de vitesse, ce qui est déjà conséquent pour notre implémentation). Comme précédemment, les conditions au bord correspondent au profil de Poiseuille, et de Dirichlet homogène sur le bord de l'obstacle.

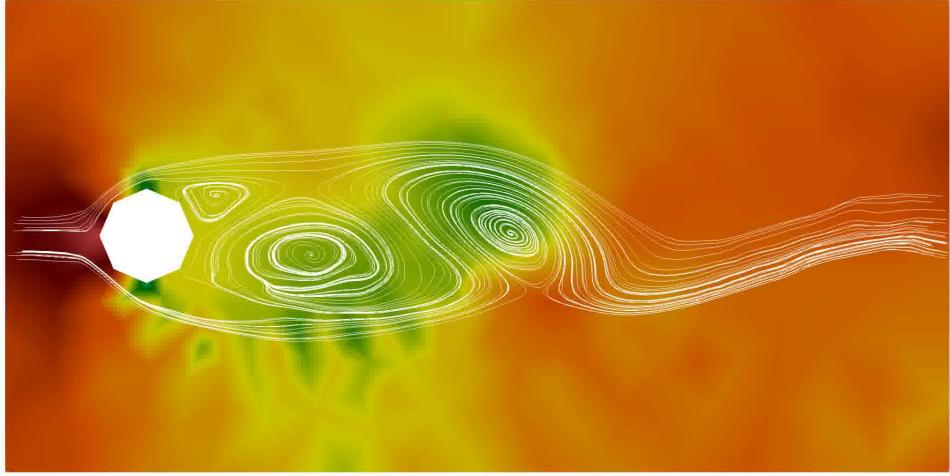


Figure 3.4: Un instantané de la simulation du sillage d'un boulon

Nous avons choisi  $T = 200$  et  $\nu = 5 \times 10^{-5}$ , ainsi qu'un pas de temps de 0.1. Après une phase d'installation, la solution semble périodique et relativement stable : les tourbillons se créent, se détachent de l'obstacle et se dissolvent contre le profil de Poiseuille de sortie de manière alternée, symétrique par rapport à un axe horizontal central. Les vidéos labellisées `karman` montrent bien la formation de la périodicité, et les grandes variations de pression proche du bord droit du domaine, où la condition de Dirichlet n'est pas adaptée au mouvement du fluide.

## 3.2 Méthodes numériques pour le contrôle

### 3.2.1 Descente de gradient par l'adjoint

Dans le cadre d'un problème de contrôle

$$\begin{aligned} & \text{minimiser } J(u, \alpha) \\ & \mathcal{T}(u) = \alpha \end{aligned}$$

dans un espace  $E$  réflexif, dès que  $\mathcal{T}$  admet un adjoint, on peut écrire un algorithme de descente dérivé du modèle suivant. Soit  $s > 0$  un pas de descente, et  $c(u^n, \alpha^n, n) \in \{\text{True}, \text{False}\}$  un critère d'arrêt.

#### Algorithme 2 – Gradient

```

Choisir  $(u^0, \alpha^0)$ 
Tant que  $c(u^n, \alpha^n, n)$ 
    Déterminer  $w^n$  à partir de  $(u^n, \alpha^n)$ 
    Déterminer  $\alpha^{n+1} := \alpha^n - s(w^n + \partial_\alpha \Phi(u^n, \alpha^n))$ 
    Déterminer  $u^{n+1}$  à partir de  $\alpha^{n+1}$ 
    Ajuster  $s$ 

```

On s'appuie sur les travaux de M. Gunzburger *et al.* ([GM00], [Gun03]) pour l'équation de Navier-Stokes. Dans toute cette section, les contrôles  $f$  seront pris dans l'espace  $L^2(0, T; [H_0^1(\Omega)]^2)$ .

**Définition 28 – Contrôle de la vitesse** Soit  $U \in \mathcal{C}([0, T], [H^2(\Omega)]^2 \cap \mathbb{V})$ . On se propose de minimiser

$$J(u, f) := \frac{1}{2} \iint_Q |u(t, x) - U(t, x)|^2 dx dt + \frac{\beta}{2} \iint_Q |f(t, x)|^2 dx dt + \frac{\gamma}{2} \int_{\Omega} |u(T, x) - U(T, x)|^2 dx$$

où  $u, p$  satisfait l'équation de Navier-Stokes (2.2) de second membre  $f$ , et  $\beta > 0, \gamma > 0$  sont constants.

On introduit de plus  $(w, r)$  l'état adjoint. Grâce à la formulation de l'adjoint obtenue proposition (11), et à l'équation adjointe (5),  $(w, r)$  satisfait

$$\begin{aligned} -\partial_t w - \nu \Delta w + (\nabla \bar{u})^* w - \nabla w \cdot \bar{u} + \nabla r &= \bar{u} - U && \text{dans } Q \\ -\operatorname{div} w &= 0 && \text{dans } Q \\ \tau_{\Sigma} w &= 0 && \text{dans } \Sigma \\ w(T) &= \gamma(\bar{u}(T) - U(T)) && \text{dans } \Pi \end{aligned}$$

On se place dans les espaces  $X_h, Y_h$  discrets définis précédemment.

**Proposition 20 – Convergence locale** Soient  $(u_h^n, p_h^n, w_h^n, r_h^n, f_h^n)$  respectivement la vitesse et pression approchées à l'étape  $n$ , la vitesse adjointe et pression adjointe à l'étape  $n$ , et le contrôle à l'étape  $n$ , ainsi que la solution exacte  $(u, p, w, r, f)$ . Pour un  $\Delta t$  convenablement choisi, il existe un bassin de convergence  $B$  dépendant de  $\beta, \gamma$  tel que si  $f_h^0 \in B$ , la suite  $(f_h^n)_n$  converge vers  $f$ .

On se donne un petit lemme, dont nous tâcherons par la suite de remplir les conditions.

**Lemme 9** Soit  $X$  un espace de Hilbert de produit scalaire  $\langle \cdot, \cdot \rangle$  et de norme  $\|\cdot\|$ . Supposons que  $J$  est une fonctionnelle de classe  $C^2(X, \mathbb{R})$  admettant un minimum en  $f$ , dont la Hessienne est continue et coercive : pour  $\alpha > 0$  et  $M > 0$ ,

$$J''(f)(\delta f, \delta f) \geq \alpha \|\delta f\|^2 \quad J''(f)(\delta f_1, \delta f_2) \leq M \|\delta f_1\| \|\delta f_2\| \quad \forall \delta f, \delta f_1, \delta f_2 \in X \quad (3.2)$$

Alors l'algorithme du gradient converge localement autour de  $f$ .

### Démonstration

En effet, le choix de  $f^{n+1} = f^n - s \nabla J(f^n)$  entraîne

$$J(f^{n+1}) = J(f^n - s \nabla J(f^n)) = J(f^n) - s \langle \nabla J(f^n), \nabla J(f^n) \rangle + \frac{s^2}{2} J''(f)(\nabla J(f^n), \nabla J(f^n)) + o(s^2)$$

et l'on peut choisir  $s > 0$  suffisamment petit pour que  $J(f^{n+1}) \leq J(f^n)$ , avec inégalité stricte si le gradient n'est pas nul. De plus, la hessienne en un minimiseur  $f$  agit localement comme une norme : pour  $m$  arbitraire,

$$J(f + (f^m - f)) = J(f) + \underbrace{\langle \nabla J(f), f^m - f \rangle}_{=0} + \frac{1}{2} \langle J''(f), f^m - f, f^m - f \rangle + O(\|f^m - f\|^3)$$

et pour tout  $\beta > 0$ , il existe une boule  $B(f, \varepsilon)$  dans laquelle  $O(\|f^m - f\|^3)$  soit négligeable devant  $\beta \|f^m - f\|^2$ . Il suffit de choisir  $\frac{\alpha}{2} - \beta > 0$  pour obtenir la convergence : pour  $f^n, f^{n+1}$  dans  $B(0, \varepsilon)$ ,

$$\left(\frac{\alpha}{2} - \beta\right) \|f^{n+1} - f\|^2 \leq J(f^{n+1}) - J(f) < J(f^n) - J(f) \leq \left(\frac{M}{2} + \beta\right) \|f^n - f\|^2$$

□

Pour éclaircir les notations, on se place dans le cas suivant.

**Lemme 10** Supposons maintenant que le problème discréétisé s'écrive sous forme variationnelle

$$\frac{1}{\Delta t}(u_h^n - u_h^{n-1}, v_h) + a(u_h^n, v_h) + (d(u_h^n), v_h) = (f^n, v_h) \quad \forall n \in \llbracket 1, N \rrbracket$$

$$u_h^0 = u_0$$

où  $d$  satisfait les hypothèses de monotonie et Lipschitz du premier ordre. Alors pour  $\Delta t$  suffisamment petit, il existe un bassin de convergence pour l'algorithme du gradient.

### Démonstration

Par des arguments classiques de Hölder, Poincaré et somme télescopique, la monotonie de  $d$  permet d'obtenir l'estimation

$$\|u_h^n\|^2 + 2\Delta t\nu |u_h^n|_1^2 \leq C(\|f\|^2 + \|u_h^0\|^2)$$

Le problème linéarisé satisfait par  $w_{h,i}$  s'écrit

$$\frac{1}{\Delta t}(w_{h,i}^n - w_{h,i}^{n-1}, v_h) + a(w_{h,i}^n, v_h) + (d'(u_h^n)w_{h,i}, v_h) = (f_i^n, v_h) \quad \forall n \in \llbracket 1, N \rrbracket$$

$$w_{h,i}^0 = 0$$

Par évaluation en  $v_h = w_{h,i}$ , il vient après quelques manipulations

$$\frac{1}{2\Delta t}(\|w_{h,i}^n\|^2 - \|w_{h,i}^{n-1}\|^2) + (\nu - \varepsilon) |w_{h,i}^n|_1^2 + (d'(u_h^n)w_{h,i}, w_{h,i}) \leq C\|f_i^n\|^2$$

Par hypothèse,  $|(d'(u_h^n)w_{h,i}, w_{h,i})| \leq M(u_h^n)\|w_{h,i}^n\|^2$ , d'où

$$(1 - 2\Delta t M(u_h^n))\|w_{h,i}^n\|^2 - \|w_{h,i}^{n-1}\|^2 + 2\Delta t(\nu - \varepsilon) |w_{h,i}^n|_1^2 \leq C\Delta t\|f_i^n\|^2$$

Notons  $C^n$  la quantité  $(1 - 2\Delta t M(u_h^n))$ , et multiplions l'équation par  $\prod_{j=0}^{n-1} C^j$ , de manière à faire apparaître une somme télescopique de la suite  $(\prod_{j=0}^m C^j \|w_{h,i}^j\|^2)_m$ . En sommant sur  $m \in \llbracket 1, N \rrbracket$ , il vient

$$\prod_{j=0}^n C^j \|w_{h,i}^n\|^2 - \underbrace{\|w_{h,i}^0\|^2}_{=0} + 2\Delta t \sum_{m=0}^n \prod_{j=0}^m C^j (\nu - \varepsilon) |w_{h,i}^m|_1^2 \leq C\Delta t \sum_{m=0}^n \prod_{j=0}^m C^j \|f_i^m\|^2$$

Supposons maintenant que  $2\Delta t \sum_{n=0}^N M(u_h^n) < e^{-\Delta t \|u_h^n\|^2}$ , de manière à ce que  $e^{-\Delta t \|u_h^j\|^2} < C^j \leq 1$  pour tout  $j$ . Il vient alors

$$2e^{-\Delta t \sum_{m=1}^N \|u_h^m\|^2} \Delta t \sum_{m=0}^n (\nu - \varepsilon) |w_{h,i}^m|_1^2 \leq C\Delta t \sum_{m=0}^n \|f_i^m\|^2$$

$$|w_{h,i}|_1^2 \leq \frac{e^{\|u_h\|^2} C}{\nu - \varepsilon} \|f_i\|^2$$

et Poincaré nous permet d'en déduire que  $\|w_{h,i}\| \leq K_i(\|u_h\|)\|f_i\|$ ,  $K_i$  une fonction continue dépendante de  $\nu$  et  $\Delta t$ . On procède de manière analogue pour le problème adjoint, à la seule différence que le second membre se trouve être  $u_h^n - U^n$ , ce qui amène la majoration  $\|w_h\| \leq K(\|u_h\|)\|u_h - U\|$  pour  $K$  une fonction continue. Cette continuité nous permet d'affirmer que pour  $u_h \in B(U, r)$ , les fonctions  $K_1$ ,  $K_2$  et  $K$  sont bornées par un certain  $\kappa > 0$ . De plus, le produit  $P(u_h) := K_1(u_h)K_2(u_h)K(u_h)\|u_h - U\|$  est continu et nul en  $u_h = U$  : ainsi, il existe un  $r_\beta > 0$  tel que  $P(u_h) \leq \frac{\beta}{2}$  pour tout  $u_h \in B(U, r_\beta)$ . On peut ainsi conclure par

$$|\langle D_{ff}J(u_h, f), f_1, f_2 \rangle| \leq (\kappa^2(\alpha + \kappa) + \beta)\|f_1\|\|f_2\|$$

$$\langle D_{ff}J(u_h, f), f_1, f_1 \rangle \geq (\beta - K_1 K_2 K \|u_h - U\|)\|f_1\|^2 \geq \frac{\beta}{2}\|f_1\|^2$$

ce qui, par le lemme (3.2.1), prouve la convergence locale de l'algorithme du gradient.  $\square$

## Résultats numériques

**Exemple de M. Gunzburger** On s'appuie sur les données de [GM00]. Soit  $\Omega = [0, 1]^2$ . L'objectif est de stabiliser le champ de vitesse du fluide autour de la valeur

$$U_Q = 10 \begin{pmatrix} \partial_y(\phi(x)\phi(y)) \\ -\partial_x(\phi(x)\phi(y)) \end{pmatrix} \quad \text{où} \quad \phi(z) := (1-z)^2(1-\cos(0.8\pi z))$$

représentée dans la figure (3.5). En particulier,  $U_Q$  est à divergence et trace nulle.

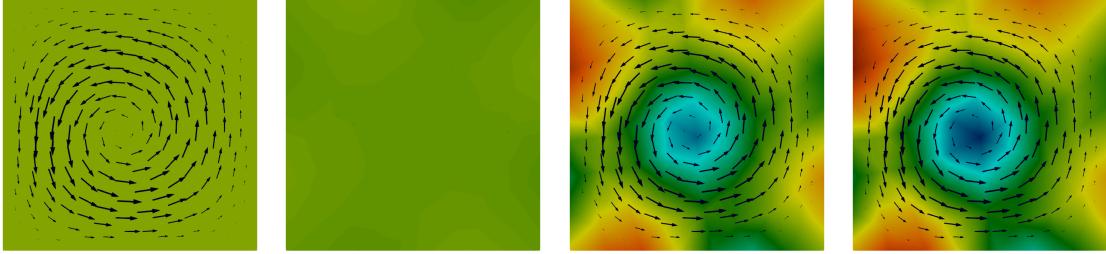


Figure 3.5: Cible  $U_Q$ , et solution à  $t = T$  pour  $\beta \in \{1, 0.01, 0.0001\}$

On considère  $\nu = 0.1$ ,  $\gamma = 0.5$ ,  $T = 1.0$ , un maillage de 74 noeuds (265 points de calcul de la vitesse), et 20 itérations en temps. L'état initial est pris égal à  $u_0 = -10U_Q$ . On emploie l'algorithme du gradient (2), implémenté en C++ en utilisant pour solveur de Navier-Stokes le code développé pour le chapitre 2. Le lecteur est invité à consulter les vidéos de l'évolution de l'état, qui commencent par `gunz`.

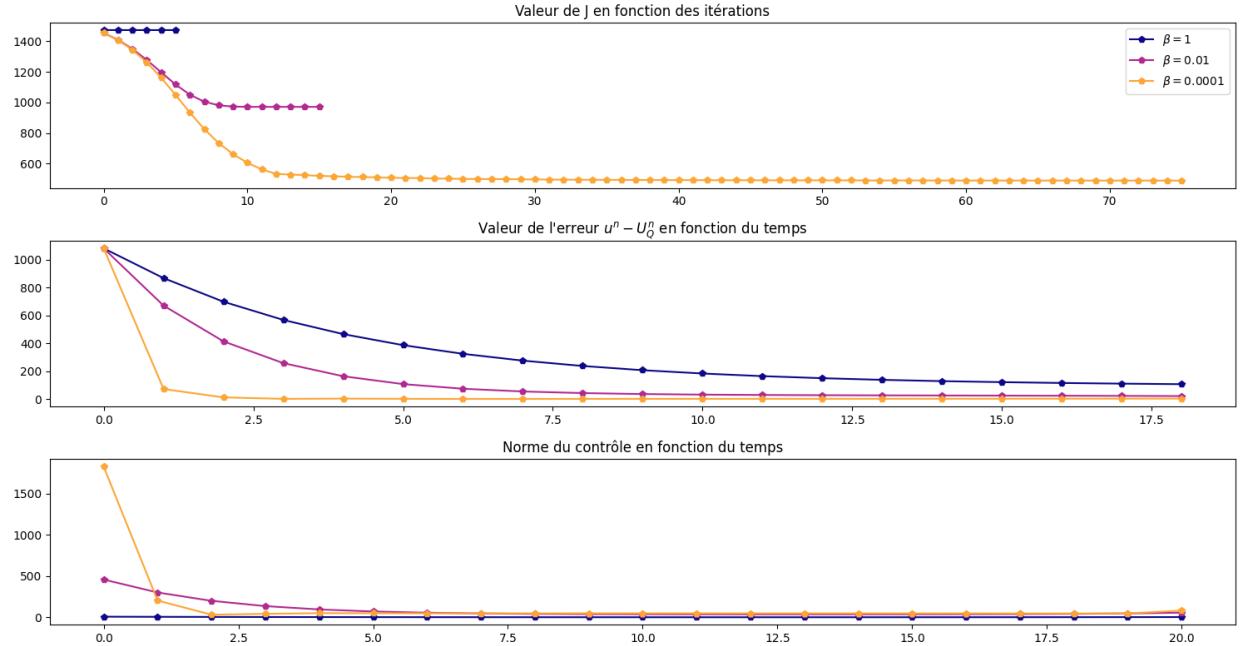


Figure 3.6: Mesures d'exécution - exemple de M. Gunzburger

La figure (3.6) présente les résultats pour ces paramètres. On constate qu'une grande valeur de  $\beta$  n'est pas d'une grande utilité pour contrôler le système : la norme de  $f$  reste faible, mais l'état  $u$  tend vers un champ nul. Lorsque  $\beta$  décroît, le contrôle devient efficace, et l'état tend rapidement vers l'objectif  $U_Q$ . La

norme du contrôle est très grande dans les premières itérations pour inverser le sens de rotation du champ et en diminuer l'amplitude, puis se stabilise durant les itérations. Les résultats sont conformes à ceux de l'article dont est issu cet exemple.

**Réduction de la vorticité** Pour deuxième exemple, on considère un domaine polygonal au bord duquel un champ de vitesse non nul et constant est imposé. On introduit un obstacle à l'intérieur du domaine, ce qui a pour effet de créer des tourbillons dans la traînée de l'objet. L'objectif est de contrer ces tourbillons, et de trouver le contrôle optimal permettant de stabiliser le système autour de la solution de Stokes stationnaire.

On conserve les mêmes paramètres  $\nu = 0.1$ ,  $\gamma = 0.5$ ,  $T = 1$  et 20 itérations. Le maillage compte 174 noeuds, soit 643 valeurs de  $u$  calculées. La solution initiale sera prise nulle à l'intérieur du domaine. Comme précédemment, on étudie l'influence du paramètre  $\beta$ . Les états à  $t = T$  sont donnés par la figure (3.7). Le lecteur curieux pourra observer l'évolution du champ dans les animations labellisées f1y.

On observe bien une diminution des tourbillons dans le sillage du hobereau. Le contrôle agit de manière inégale au fil des itérations : comme le montre la figure (3.8), la norme de  $f$  est croissante le long des itérations, et augmente brusquement au temps final. Comme dans le premier exemple, une valeur plus faible de  $\beta$  permet de mieux contrôler le système, et la solution se stabilise autour de  $U_Q$ .

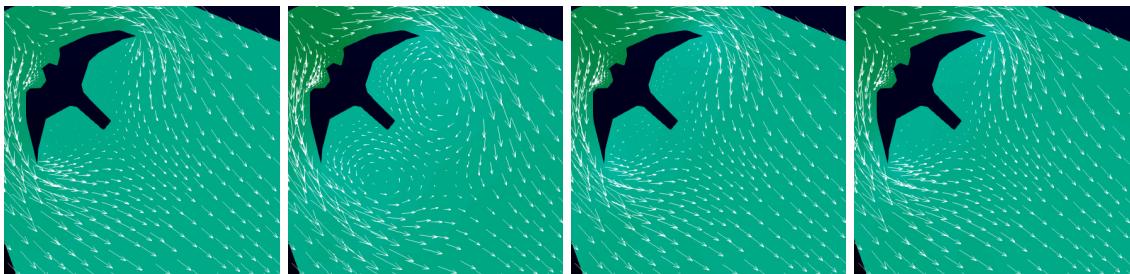


Figure 3.7: De gauche à droite : objectif, puis état final pour  $\beta \in \{1, 0.01, 0.0001\}$

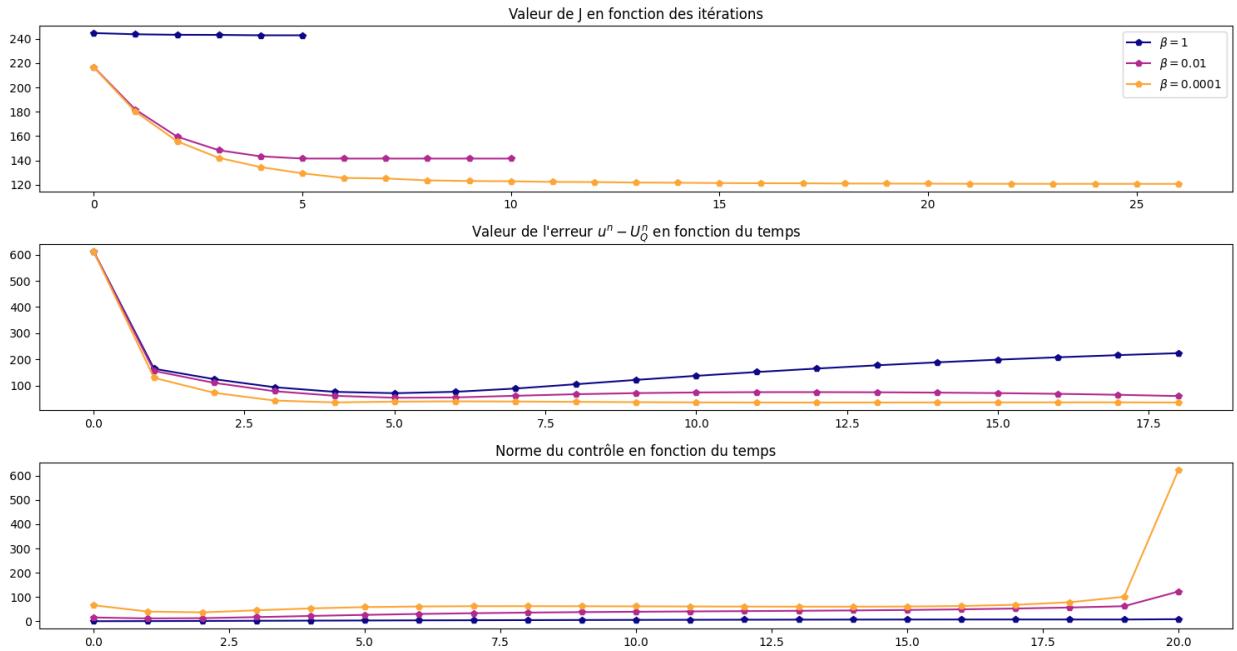


Figure 3.8: Mesures d'exécution - exemple du hobereau

### 3.2.2 Sequential Quadratic Programming (SQP)

#### Base : méthode de Newton

Nous utilisons les notations de la section (1.2.3). L'idée directrice est d'appliquer la méthode de Newton à la fonctionnelle d'une seule variable  $J(u(\alpha), \alpha)$ . Dans le cas général, si  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  est une fonction de classe  $\mathcal{C}^1$  dont on cherche un zéro, la méthode de Newton consiste à itérer le système  $f(u^n) + f'(u^n)(u - u^n) = 0$  et à poser  $u^{n+1} = u$ , ou, de manière équivalente, à chercher  $u^{n+1}$  comme la solution du problème de minimisation

$$\min_u \langle f(u^n), u - u^n \rangle + \frac{1}{2} \langle f'(u^n)(u - u^n), u - u^n \rangle$$

où  $\langle \cdot, \cdot \rangle$  désigne ici le produit scalaire de  $\mathbb{R}^n$ .

La méthode SQP applique cet algorithme à  $f := J'(u(\alpha), \alpha)$ . Soit  $(\alpha^n)_n \subset A$  une suite de contrôles destinée à tendre vers  $\bar{\alpha}$  un point de minimum de  $J$ , ainsi que  $(u(\alpha^n))_n$  et  $(w(u(\alpha^n), \alpha^n))_n$  la suite des états et adjoints exacts associés à  $(\alpha^n)_n$ . En vertu de la proposition (7), pour tout  $\alpha \in A$ , on peut écrire

$$\frac{1}{2} (J''(u(\alpha^n), \alpha^n), (\alpha - \alpha^n)^2) = \frac{1}{2} \left( \partial_{u,\alpha}^2 \mathcal{L}(u(\alpha^n), \alpha^n, w(u(\alpha^n), \alpha^n)), ((\partial_\alpha u, \alpha - \alpha^n), \alpha - \alpha^n)^2 \right)$$

d'où le problème de minimisation de la méthode de Newton (où l'on omet les arguments) :

$$\min_{\alpha \in A} (\partial_u J, (\partial_\alpha u, \alpha - \alpha^n)) + (\partial_\alpha J, \alpha - \alpha^n) + \frac{1}{2} \left( \partial_{u,\alpha}^2 \mathcal{L}, ((\partial_\alpha u, \alpha - \alpha^n), \alpha - \alpha^n)^2 \right) \quad (3.3)$$

Remarquons qu'en linéarisant l'équation d'état  $\mathcal{T}(u(\alpha)) = \alpha$  en  $\alpha^n$  dans la direction  $\alpha - \alpha^n$ , il vient

$$\mathcal{T}(u(\alpha^n)) + (\partial_u \mathcal{T}(u^n), (\partial_\alpha u(\alpha^n), \alpha - \alpha^n)) = \alpha^n + (\alpha - \alpha^n) = \alpha$$

une expression implicite de l'accroissement  $(\partial_\alpha u(\alpha^n), \alpha - \alpha^n)$ .

#### Choix algorithmiques

Notons  $(u^n)_n$  et  $(w^n)_n$  les approximations de  $(u(\alpha^n))_n$  et  $(w(u(\alpha^n), \alpha^n))_n$ . Si  $u^n$  et  $\alpha^n$  sont connus, on impose à l'adjoint  $w^n$  de satisfaire le système linéaire

$$(\partial_u \mathcal{T}(u^n))^* w^n = \partial_u \Phi(u^n, \alpha^n)$$

et la connaissance du triplet  $(u^n, w^n, \alpha^n)$  permet d'évaluer  $\partial_{(u,\alpha),(u,\alpha)}^2 \mathcal{L}$  à l'étape  $n$ . Pour  $\alpha \in A$  donné, notons  $u := u^n + (\partial_\alpha u(\alpha^n), \alpha - \alpha^n)$ . Les itérés  $(u^{n+1}, \alpha^{n+1})$  seront les solutions du problème de minimisation sous contrainte suivant :

$$\begin{aligned} \min_{u,\alpha} \tilde{J} &:= (\partial_u J(u^n, \alpha^n), u - u^n) + (\partial_\alpha J(u^n, \alpha^n), \alpha - \alpha^n) + \frac{1}{2} \left( \partial_{u,\alpha}^2 \mathcal{L}(u^n, \alpha^n, w^n), (u - u^n, \alpha - \alpha^n)^2 \right) \\ &\mathcal{T}(u^n) + (\partial_u \mathcal{T}(u^n), u - u^n) = \alpha, \quad \alpha \in A \end{aligned}$$

L'algorithme est alors le suivant :

#### Algorithme 3 – Méthode SQP

Choisir  $u^0, \alpha^0, w^0$

Pour chaque  $n$

Obtenir  $(u, \alpha)$  par résolution du problème quadratique

Poser  $u^{n+1} = u$  et  $\alpha^{n+1} = \alpha$

Déterminer  $w^{n+1}$  à partir de  $(u^{n+1}, \alpha^{n+1})$

### Cas particulier du problème de référence

La fonctionnelle  $\tilde{J}$  se développe en

$$\left( \frac{\partial_u J}{\partial_\alpha J} \right)^t \begin{pmatrix} u - u^n \\ \alpha - \alpha^n \end{pmatrix} + \frac{1}{2} \begin{pmatrix} u - u^n \\ \alpha - \alpha^n \end{pmatrix}^t \begin{pmatrix} \partial_u^2 \Phi & \partial_{u\alpha}^2 \Phi \\ \partial_{\alpha u}^2 \Phi & \partial_{\alpha\alpha}^2 \Phi \end{pmatrix} \begin{pmatrix} u - u^n \\ \alpha - \alpha^n \end{pmatrix} - (w^n, (\partial_{uu}^2 \mathcal{T}(\bar{u}), u - u^n, u - u^n))_E$$

et dans le cas particulier où

$$\Phi(u, \alpha) = \begin{pmatrix} \varphi(u, \alpha_1) \\ \psi(u, \alpha_2) \\ \phi(u, \alpha_3) \end{pmatrix}, \quad \mathcal{T}(u) = \begin{pmatrix} \partial_t u - \Delta u + d(\cdot, u) \\ \partial_\nu u + b(\cdot, u) \\ u(\cdot, 0) \end{pmatrix}$$

La forme quadratique  $\tilde{J}$  devient

$$\begin{aligned} & \iint_Q \left[ \begin{pmatrix} \varphi_u \\ \varphi_{\alpha_1} \end{pmatrix}^t \begin{pmatrix} u - u^n \\ \alpha - \alpha_1^n \end{pmatrix} + \frac{1}{2} \begin{pmatrix} u - u^n \\ \alpha - \alpha_1^n \end{pmatrix}^t \begin{pmatrix} \partial_{uu}^2 \varphi & \partial_{u\alpha_1}^2 \varphi \\ \partial_{\alpha_1 u}^2 \varphi & \partial_{\alpha_1}^2 \varphi \end{pmatrix} \begin{pmatrix} u - u^n \\ \alpha - \alpha_1^n \end{pmatrix} - \frac{1}{2} w^n (d_{uu}, u - u^n, u - u^n) \right] dq \\ & + \iint_\Sigma \left[ \begin{pmatrix} \psi_u \\ \psi_{\alpha_2} \end{pmatrix}^t \begin{pmatrix} u - u^n \\ \alpha - \alpha_2^n \end{pmatrix} + \frac{1}{2} \begin{pmatrix} u - u^n \\ \alpha - \alpha_2^n \end{pmatrix}^t \begin{pmatrix} \partial_{uu}^2 \psi & \partial_{u\alpha_2}^2 \psi \\ \partial_{\alpha_2 u}^2 \psi & \partial_{\alpha_2}^2 \psi \end{pmatrix} \begin{pmatrix} u - u^n \\ \alpha - \alpha_2^n \end{pmatrix} - \frac{1}{2} w^n (b_{uu}, u - u^n, u - u^n) \right] ds dt \\ & + \int_\Omega \left[ \begin{pmatrix} \phi_u \\ \phi_{\alpha_3} \end{pmatrix}^t \begin{pmatrix} u - u^n \\ \alpha - \alpha_3^n \end{pmatrix} + \frac{1}{2} \begin{pmatrix} u - u^n \\ \alpha - \alpha_3^n \end{pmatrix}^t \begin{pmatrix} \partial_{uu}^2 \phi & \partial_{u\alpha_3}^2 \phi \\ \partial_{\alpha_3 u}^2 \phi & \partial_{\alpha_3}^2 \phi \end{pmatrix} \begin{pmatrix} u - u^n \\ \alpha - \alpha_3^n \end{pmatrix} \right] dx \end{aligned}$$

ce que l'on peut discréteriser spatialement, par exemple par différences finies. Enfin, la contrainte de minimisation s'écrit

$$\begin{aligned} \partial_t u - \Delta u + d(\cdot, u^n) + d_u(\cdot, u^n)(u - u^n) &= \alpha_1 && \text{dans } Q \\ \partial_\nu u + b(\cdot, u^n) + b_u(\cdot, u^n)(u - u^n) &= \alpha_2 && \text{dans } \Sigma \\ u(\cdot, 0) &= \alpha_3 && \text{dans } \Omega \end{aligned}$$

### Application au problème de contrôle au bord

Cette section choisit de se placer dans un domaine  $\Omega = [0, l] \subset \mathbb{R}$ . Pour des raisons de diversité autant que de nombre de variables, on s'intéresse au cas où  $\alpha = (0, g, 0)$  est un contrôle sur le bord  $\Sigma$ . On se donne  $T > 0$  un temps final, et  $(x_i, t_j)_{(i,j) \in [\![1,N]\!] \times [\![1,M]\!]}$  une discréterisation régulière de  $\Omega \times [0, T]$  de pas  $(h, \Delta t)$ . On note  $u_{i,j}^n$  l'approximation de  $u(x_i, t_j)$  à l'étape  $n$  de la minimisation. Après plusieurs essais, on choisit d'approcher la dérivée temporelle par un schéma d'Euler backward, les traces normales par des schémas d'ordre 2 décentrés vers l'intérieur de  $\Omega$ , et le Laplacien par le schéma classique à 3 points. Ainsi, la contrainte sera discréterisée en

$$\begin{aligned} \frac{u_{i,j} - u_{i,j-1}}{\Delta t} - \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2} + d(x_i, t_j, u_{i,j}^n) + d_u(x_i, t_j, u_{i,j}^n)(u_{i,j} - u_{i,j}^n) &= 0 & \forall i \in [\![2, N-1]\!] \\ \frac{3u_{1,j} - 4u_{2,j} + u_{3,j}}{2h} + b(x_1, t_j, u_{1,j}^n) + b_u(x_1, t_j, u_{1,j}^n)(u_{1,j} - u_{1,j}^n) &= g_{1,j} & \forall j \in [\![2, M]\!] \\ \frac{3u_{N,j} - 4u_{N-1,j} + u_{N-2,j}}{2h} + b(x_N, t_j, u_{N,j}^n) + b_u(x_N, t_j, u_{N,j}^n)(u_{N,j} - u_{N,j}^n) &= g_{N,j} & \forall j \in [\![2, M]\!] \\ u_{i,1} - u_0(x_i) &= 0 & \forall i \in [\![1, N]\!] \end{aligned}$$

Le problème adjoint est traité de la même manière. Une discréterisation naïve de la fonctionnelle  $\tilde{J}$  nous a suffi à obtenir les résultats présentés. On prend deux critères d'arrêt, l'un observant le passage de la valeur de  $J$  sous un certain seuil, et l'autre basé sur la variation de la fonctionnelle, soit  $\frac{|J(u^{n+1}, \alpha^{n+1}) - J(u^n, \alpha^n)|}{|J(u^n, \alpha^n)|} < \varepsilon$  (où  $\varepsilon = 10^{-8}$  dans les exemples présentés).

**Exemples pour exploration** On commence par tester le code sur des exemples linéaires. On emploie le solveur `pdepe` de Matlab pour déterminer la solution exacte lorsque la fonctionnelle est de la forme  $J(u, g) = \|g - g_o\|_{L^2(\Sigma)}^2$ , où  $g_o = -\cos(9t)\mathbf{1}_{\{x=0\}} + 0.2 \sin(9t)\mathbf{1}_{\{x=l\}}$ . Pour l'équation de la chaleur ( $b = d = 0$ ), avec  $g_{\max} = -g_{\min}$  = un Très Grand Nombre, l'algorithme converge en une itération (ce qui est attendu d'un problème purement linéaire) et produit les résultats de la figure (3.9).

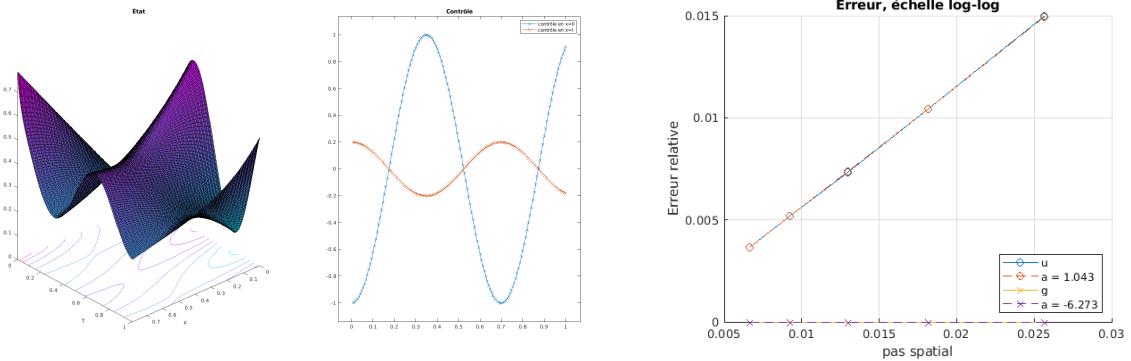


Figure 3.9: équation de la chaleur avec contrôle au bord,  $M = N = 151$

On observe ici une convergence à l'ordre 1. L'erreur commise par  $g_h$  sur  $g_o$  est de l'ordre de  $10^{-13}$ .

On poursuit en choisissant une fonctionnelle de la forme  $J(u, g) = \|u - u_o\|_{L^2(\Sigma)}^2$ , pour imposer une valeur de l'état au bord, et en choisissant une non-linéarité continue, croissante, localement lipschitzienne et uniformément bornée en  $u = 0$ , soit  $d(x, t, u) = 4e^u$ , et  $b = 0$ . Pour  $u_o = \cos(9t)$ , l'algorithme a produit en 5 itérations les résultats de la figure (3.10).

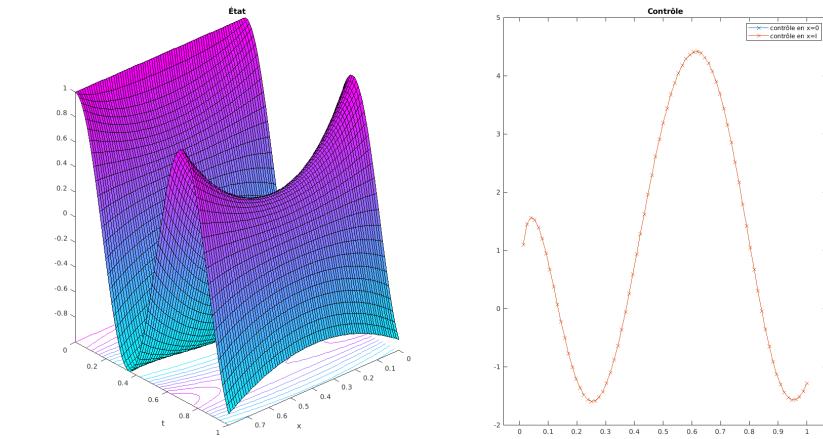


Figure 3.10: Non-linéarité exponentielle et objectif au bord,  $M = N = 77$

La forme de la fonctionnelle ne permet plus d'employer le solveur de Matlab, mais la valeur de  $J$  se trouve être une norme mesurant l'écart à l'objectif au bord. Son évolution au cours des itérations est la suivante :

itération	1	2	3	4	5
valeur de $J$	7.279e+00	1.593e-01	3.777e-03	5.505e-06	1.160e-11

Enfin, on cherche à observer l'impact de la condition  $g \in [g_{\min}, g_{\max}]$ . On considère cette fois un système non linéaire sur le bord et à l'intérieur du domaine, donné par

$$\begin{aligned} \partial_t u - \Delta u + u &= 0 && \text{dans } Q \\ \partial_\nu u + e^{tu} &= g && \text{dans } \Sigma \\ u(\cdot, 0) &= 0 && \text{dans } \Omega \end{aligned}$$

ainsi que le problème de contrôle

$$J(u, g) = \frac{1}{2} \int_0^T \left\{ \left( u(l, t) + \sqrt{\varepsilon + |\sin(5t)|} \right)^2 + |u(0, t)|^2 \right\} dt, \quad g_{\min} = -\infty, \quad g_{\max} = \begin{cases} \infty & \text{si } x = 0 \\ \alpha & \text{si } x = l \end{cases}$$

où  $\varepsilon$  est choisi à 0.01 pour assurer le caractère lipschitzien de  $\psi$ . Les résultats pour  $\alpha = \infty$  et  $\alpha = 0$  sont donnés par la figure (3.11).

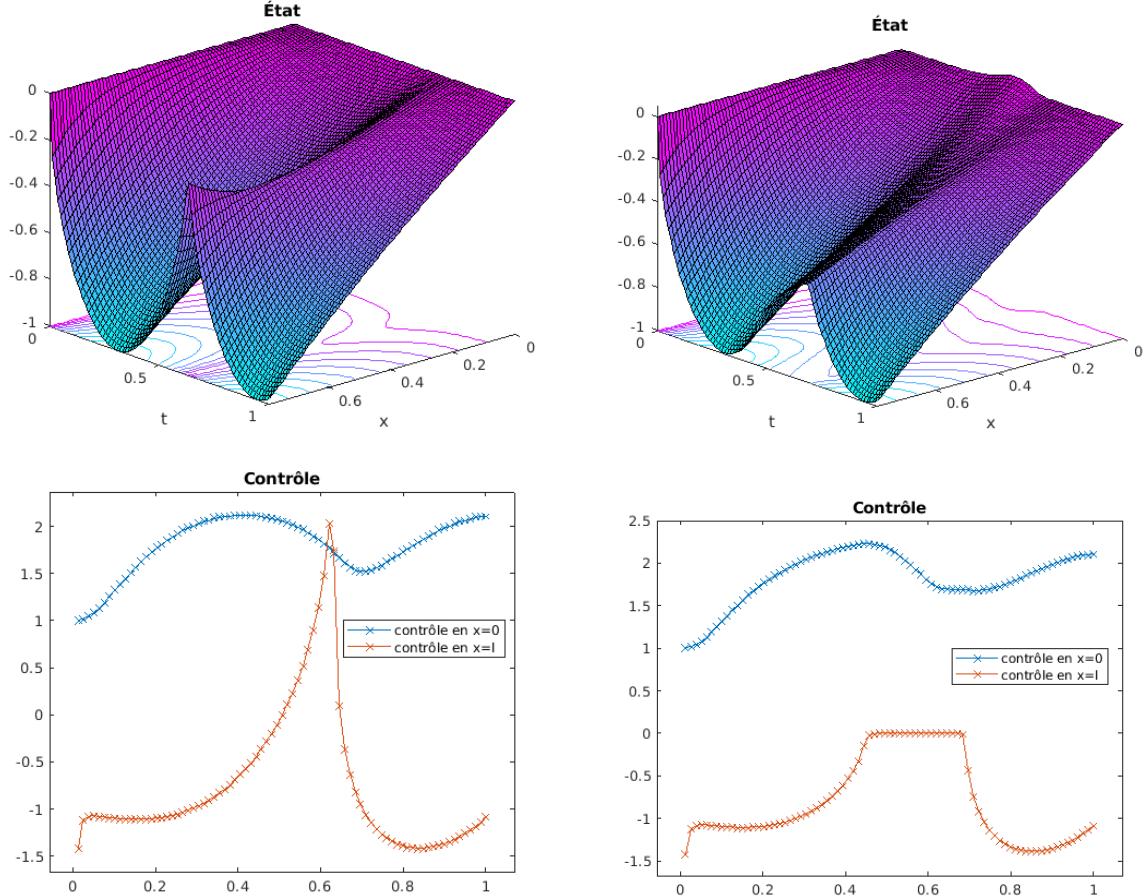


Figure 3.11: États optimaux et contrôles associés,  $\alpha \in \{\infty, 0\}$ ,  $M = N = 120$

La fonctionnelle atteint la valeur finale de 3.69e-13 pour  $\alpha = \infty$  en 4 itérations, et 3.64e-03 pour  $\alpha = 0$  en 6 itérations. C'est conforme à l'intuition, la valeur optimale n'étant plus accessible. On remarque que les contrôles pour  $\alpha = 0$  ne sont pas égaux aux simples troncatures de ceux non contraints. De plus, l'influence de la borne sur le bord  $x = l$  n'est pas uniquement locale : l'état est globalement modifié, et le contrôle sur la composante  $x = 0$  est également perturbé.

### Équation de la chaleur avec condition de Stefan-Boltzmann

Cette section reproduit l'exemple test de [Trö10], section 5.8 et 5.9.2. On choisit pour paramètres

$$l = \frac{\pi}{4}, \quad T = 1, \quad d \equiv \varphi \equiv 0, \quad b(\sigma, t, u) = u + u|u|^3 - e^{-4t}/4 + \min \left\{ 1, \max \left\{ 0, \frac{e^t - e^{1/3}}{e^{2/3} - e^{1/3}} \right\} \right\}$$

$$\psi(\sigma, t, u, g) = \left( -e^{-2t}u + \frac{e^{1/3}}{\sqrt{2}}g + \frac{(e^{2/3} - e^{1/3})}{2\sqrt{2}}g^2 \right) \mathbf{1}_{\{x=l\}}, \quad \phi(x, u) = \frac{1}{2}(u - (e + e^{-1})\cos(x))^2$$

$$g_{\min} = 0, \quad g_{\max} = 1, \quad u_0 = \cos(x), \quad u^0 \equiv w^0 \equiv g^0 = 0.5$$

La solution exacte se trouve être

$$u(x, t) = e^{-t} \cos(x), \quad w(x, t) = -e^t \cos(x) \quad g(x, t) = \min \left\{ 1, \max \left\{ 0, \frac{e^t - e^{1/3}}{e^{2/3} - e^{1/3}} \right\} \right\}$$

On a représenté en figure (3.12) l'état et le contrôle optimal. Les deux limites  $g_{\min}$  et  $g_{\max}$  sont atteintes : on observe numériquement une régularisation du contrôle aux points de discontinuité de la dérivée.

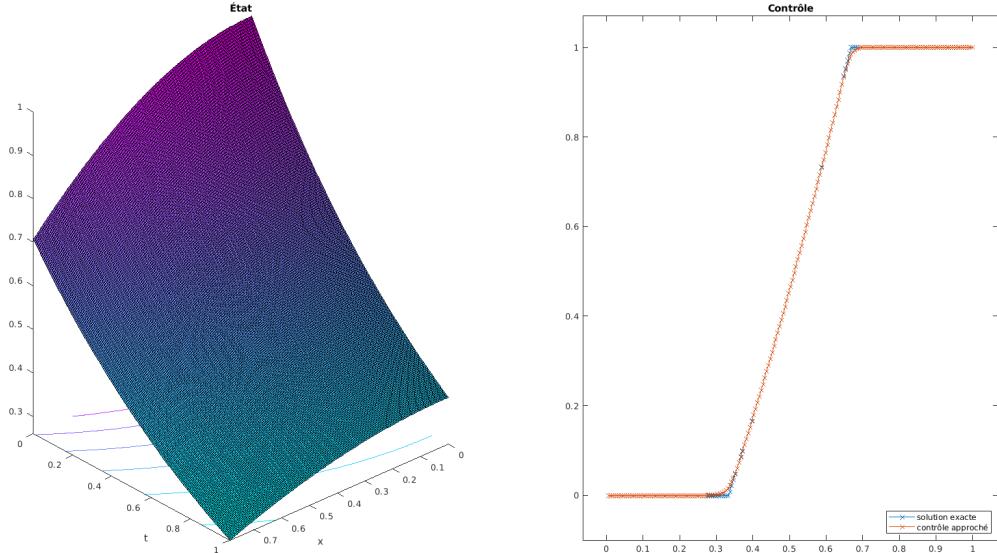
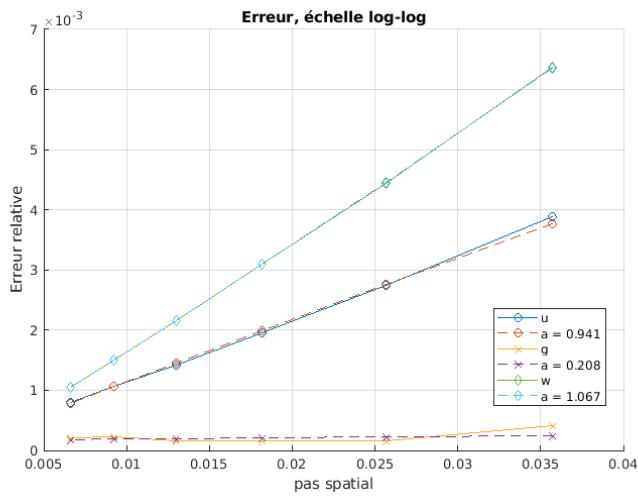


Figure 3.12: Résultats numériques pour  $N = M = 200$

Notre implémentation a convergé en 6 itérations pour le cas  $M = N = 200$ , pour les critères d'arrêt choisis. L'historique des erreurs est représenté figure (3.13). Nous avons essayé d'augmenter l'ordre des schémas employés pour atténuer la régularisation du contrôle, mais nos tentatives avec de simples schémas d'Euler n'ont pas abouti (les schémas downwind sont instables, et la prise en compte des conditions initiales affecte toute tentative de schéma d'ordre plus élevé décentré en amont). Pour aller plus loin, on pourrait tenter d'employer une méthode de Runge-Kutta sur le schéma discréteisé. Dans le cadre de l'algorithme SQP, toutes les variables sont calculées simultanément, et un schéma de Runge-Kutta n'est pas aussi simple à implémenter que dans le cas des propagateurs. Nous concluons donc ici les exemples numériques.



it.	erreur $u$	erreur $g$	erreur $w$
1	0.081635	0.024442	0.109115
2	0.012205	0.002622	0.019548
3	0.000818	0.000173	0.000495
4	0.000625	0.000169	0.000783
5	0.000625	0.000169	0.000784
6	0.000625	0.000169	0.000784

Figure 3.13: Ordre de convergence pour  $N \in \{28, 39, 55, 77, 108, 151\}$ , et erreurs pour  $N = 200$

# Conclusion

Synthétisons ce que nous avons exposé.

Dans une première partie, on s'est intéressé à une classe d'équations paraboliques non linéaires assez sympathiques. Deux propriétés justifient cet adjectif : premièrement, les non-linéarités sont des opérateurs de superposition, qui ne dépendent que de la valeur locale de la solution, et non de ses dérivées ou d'une valeur globale. Deuxièmement, ces mêmes non-linéarités sont considérées monotones croissantes : c'est l'argument qui nous a permis d'étendre les méthodes variationnelles au cas non linéaire. Pour cette famille de problèmes, on a d'abord considéré le caractère bien-posé des équations, puis leur contrôlabilité. Notamment, sous des conditions de régularité plus restrictives, on peut exhumer un système linéaire satisfait par la solution du problème de contrôle, grâce au problème adjoint.

Nous avons ensuite quitté le monde monotone pour nous intéresser au célèbre système de Navier-Stokes. Ce chapitre introduit le modèle simple des fluides incompressibles, et les questions associées d'existence et unicité de la solution. Le cœur du propos vise à discréteriser correctement le système, ce qui est fait en employant des schémas classiques sous une certaine condition sur les espaces discrets. Une fois cette étape franchie, on étudie la stabilité et la convergence des schémas proposés.

Pour finir, le chapitre 3 illustre les deux précédents. On a développé trois codes distincts : le premier implémente directement les schémas pour les fluides présentés dans le chapitre 2, et permet au deuxième code de contrôler ces équations, grâce au système adjoint. Le troisième s'intéresse directement à un problème de contrôle au bord d'une équation parabolique monotone.

L'étudiant.e souhaitant poursuivre ces travaux, par exemple dans un PFE, a l'embarras du choix dans les directions possibles. La condition de monotonie est assez restrictive, et on pourrait s'intéresser à des systèmes paraboliques comme le modèle de Gray-Scott, qui satisfait toutes les hypothèses de régularité demandées - hormis la croissance des opérateurs. De même, la condition inf-sup imposée sur les espaces de discréétisation pour Navier-Stokes est pénalisante pour la dimension des systèmes linéaires. On a considéré seulement les espaces classiques  $\mathbb{P}^k/\mathbb{P}^l$ , mais des choix plus exotiques comme  $\mathbb{P}^1$ -bulle pour la vitesse sont possibles, et permettraient de soulager le code. Enfin, on ne s'est pas attardé sur l'analyse de la méthode SQP, qui repose sur celle de Newton : pourtant, une telle base laisse espérer un ordre de convergence satisfaisant.

Merci au lecteur coriace d'être arrivé jusqu'ici. On lui souhaite d'avoir trouvé autant d'intérêt que nous à ces jolies histoires.

# Bibliographie

- [Dre12] Pierre Dreyfuss. *Introduction à l'analyse Des Équations de Navier-Stokes.* Références Sciences. Ellipses, Paris, 2012.
- [GHS91] M. D. Gunzburger, L. S. Hou, and Th. P. Svobodny. Analysis and finite element approximation of optimal control problems for the stationary Navier-Stokes equations with Dirichlet controls. *ESAIM: Mathematical Modelling and Numerical Analysis*, 25(6):711–748, 1991.
- [GM00] M. D. Gunzburger and S. Manservisi. Analysis and Approximation of the Velocity Tracking Problem for Navier-Stokes Flows with Distributed Control. *SIAM Journal on Numerical Analysis*, 37(5):1481–1512, 2000.
- [Gri07] Jens A Griepentrog. Maximal regularity for nonsmooth parabolic problems in Sobolev-Morrey spaces. *Advances in Differential Equations*, pages 1031–1078, 2007.
- [Gun03] Max D. Gunzburger. *Perspectives in Flow Control and Optimization.* Advances in Design and Control. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2003.
- [Hei98] Matthias Heinkenschloss. Formulation and Analysis of a Sequential Quadratic Programming Method for the Optimal Dirichlet Boundary Control of Navier-Stokes Flow. In William H. Hager and Panos M. Pardalos, editors, *Optimal Control: Theory, Algorithms, and Applications*, pages 178–203. Springer US, Boston, MA, 1998.
- [Lio69] Jacques-Louis Lions. *Quelques méthodes de résolution des problèmes aux limites non linéaires.* Les cours de référence. Dunod, Paris, 1969.
- [LM68] Jacques Louis Lions and Enrico Magenes. *Problèmes aux limites non homogènes et applications*, volume 1. Dunod, Paris, 1968.
- [MAS18] Guri I. Marchuk, Valeri I. Agoshkov, and Victor P. Shutyaev. *Adjoint Equations and Perturbation Algorithms in Nonlinear Problems.* CRC Press, first edition, April 2018.
- [Sch19] Jean-François Scheid. *Analyse Numérique Des Équations de Navier-Stokes.* Cours de Master 2 Mathématiques Pour La Recherche. Université de Lorraine, Nancy, 2019.
- [Trö10] Fredi Tröltzsch. *Optimal Control of Partial Differential Equations: Theory, Methods, and Applications.* Number v. 112 in Graduate Studies in Mathematics. American Mathematical Society, Providence, R.I, 2010.
- [van82] M. van Dyke. An album of fluid motion. *NASA STI/Recon Technical Report A*, 82:36549, June 1982.
- [VM07] H. K. Versteeg and W. Malalasekera. *An Introduction to Computational Fluid Dynamics: The Finite Volume Method.* Pearson Education Ltd, Harlow, England ; New York, 2nd ed edition, 2007.
- [Zei89] Eberhard Zeidler. *Nonlinear Functional Analysis and Its Applications. 2B: Nonlinear Monotone Operators.* Springer, New York Berlin Heidelberg, nachdr. edition, 1989.