Algorithms for optimization and control of free interface - Application to the industrial aluminium production

Antonin Orriols

THÈSE

présentée par

Antonin Orriols

pour l'obtention du titre de

DOCTEUR DE L'ÉCOLE NATIONALE DES PONTS ET CHAUSSÉES

Spécialité :

Mathématiques et Informatique

Sujet :

Algorithmes d'optimisation et de contrôle d'interface libre

Application à la production industrielle d'aluminium

Soutenance le 15 décembre 2006 devant le jury composé de :

Président :	Yvon Maday
Rapporteurs :	Michel Bercovier Bertrand Maury
Examinateurs :	Tony Lelièvre Thierry Tomasino
Invité :	Jean-Frédéric Gerbeau
Directeur de thèse :	Claude Le Bris

Remerciements

Merci à Claude Le Bris (ENPC) qui m'a proposé cette thèse, ainsi qu'à Jean-Frédéric Gerbeau (INRIA), Thierry Tomasino (Alcan) et Tony Lelièvre (ENPC) qui l'ont encadrée.

Je tiens à exprimer ma reconnaissance à Bertrand Maury (Université Paris Sud) et Michel Bercovier (The Hebrew University of Jerusalem), qui ont accepté de rapporter le présent mémoire et autorisé sa soutenance.

Cette thèse n'aurait pu aboutir sans la présence d'Yvon Maday (Université Pierre et Marie Curie), qui m'a initié au calcul scientifique et a accepté de présider le jury de soutenance. Je tiens à lui témoigner ma gratitude.

Merci également aux collègues, qui contribué à l'excellente ambiance du CERMICS, tant d'un point de vue humain que professionnel : Lætitia Andrieu, Maxime Barrault, Guy Bencteux, Adel Ben Haj Yedder, Sylvie Berte, Adrien Blanchet, Éric Cancès, Jean-Philippe Chancelier, Amélie Deleurence, Alexandre Ern, Antoine Gloria, Julien Guyon, Bernard Lapeyre, Frédéric Legoll, Mazyar Mirrahimi, Régis Monneau, Serge Piperno, Annette Stephansen, Gabriel Stoltz, Pierre Tardif d'Hamonville et Gabriel Turinici.

Enfin, je tiens à remercier mes collègues et amis, Anas Dallagi et Yousra Gati, ainsi que ma famille, sans lesquels ces trois années auraient été encore plus difficiles.

À la mémoire de mon père.

AO,
Paris, 2006.

Résumé : La production industrielle d'aluminium met en jeu plusieurs aspects physiques, à la fois chimiques, thermiques et magnétohydrodynamiques (MHD). L'une de ses particularités est la coexistence dans une cuve de deux fluides non miscibles, ce qui conduit à la présence d'une interface libre. Ce procédé consomme près de 2% de l'électricité mondiale, la moitié étant perdue par effet Joule. L'enjeu est de réduire ce coût sans déstabiliser le procédé : il s'agit typiquement d'un problème de contrôle optimal, que nous traitons en considérant une modélisation MHD de la cuve. Deux approches sont utilisées pour effectuer cette optimisation, à savoir considérer une contrainte d'état non linéaire basée sur un couplage entre les équations de Maxwell et de Navier-Stokes multifluides, et une contrainte d'état linéaire résultant d'une approximation *shallow water* de la précédente. Après une courte introduction à la modélisation du procédé et aux concepts du contrôle optimal basé sur le principe de Pontryagin, nous décrivons dans un premier temps le contrôle de l'évolution de l'interface modélisée par l'approximation *shallow water*. S'ensuivent un travail de parallélisation du logiciel de simulation du procédé dans le cadre non linéaire et la recherche numérique d'actionneurs acceptables pour son contrôle. Enfin, un algorithme d'optimisation de la forme de l'interface est proposé sous une contrainte d'état non linéaire simplifiée, à savoir les équations de Navier-Stokes bifluides en dimension deux.

Mots-clés : EDP paraboliques non linéaires, contrôle optimal, magnétohydrodynamique, surface libre, méthodes d'éléments finis, calcul parallèle

Sommaire

Keywords : nonlinear parabolic PDEs, optimal control, magnetohydrodynamics, free surface, finite element methods, parallel computing

Summary

Table des matières

Chapitre 1

Problème étudié

1.1 Généralités

De la vie courante à l'industrie lourde, en passant par les équipements de laboratoires, de nombreux produits sont composés d'aluminium. On pourra citer notamment des objets des plus usuels comme les boîtes de conserve prévues pour le conditionnement des produits alimentaires, les feuilles d'aluminium destinées à l'emballage des médicaments, mais aussi des récipients en tous genres utilisés en chimie ou en biologie, les câbles électriques, les pièces et modules à base d'aluminium pour l'automobile et l'aéronautique, ou encore nombre de matériaux composites rencontrés en architecture, imprimerie, produits aquatiques, etc. Cette omniprésence nécessite une production de masse, essentiellement à partir d'*alumine* (oxyde d'aluminium), seule forme sous laquelle se trouve le métal dans la nature, en abondance du reste. L'alumine est contenue dans diverses roches, comme les kaolins, les schistes et la bauxite (nom provenant de la ville des Baux-de-Provence où fut exploitée pour la première fois une mine de bauxite). Cette dernière est la seule roche utilisée par l'industrie de l'aluminium en raison de sa forte teneur en alumine et sa faible concentration en silice. Aujourd'hui, on la trouve surtout dans des mines en Guinée, Australie et Amérique Latine. On en extrait l'alumine par un enchaînement complexe de traitements chimiques, le *procédé Bayer* (Fig. 1.1).

Une fois l'alumine extraite de la bauxite, il reste à la dissoudre dans une solution liquide constituée majoritairement de fluor, la *cryolite*, afin de disposer d'un *électrolyte* ou *bain* adapté à la réaction d'*électrolyse* par laquelle on obtient l'aluminium pur. Notons, pour finir avec les procédés se situant en amont de l'électrolyse, qu'un ingrédient supplémentaire est nécessaire à cette réaction, le carbone, qui n'est autre que le matériau dont sont faites les anodes, et qu'il faut donc également extraire et transformer. En aval de l'électrolyse, l'aluminium produit à l'état liquide subit lui-même diverses transformations en fonderie, pour finalement être stocké à l'état solide sous forme de feuilles, billettes ou encore lingots servant à la fabrication des produits présentés au début. Notre sujet d'étude principal portant sur l'électrolyse de l'alumine, nous nous concentrons à présent sur cette dernière, connue sous le nom de *procédé Hall-Héroult* (Fig. 1.2).

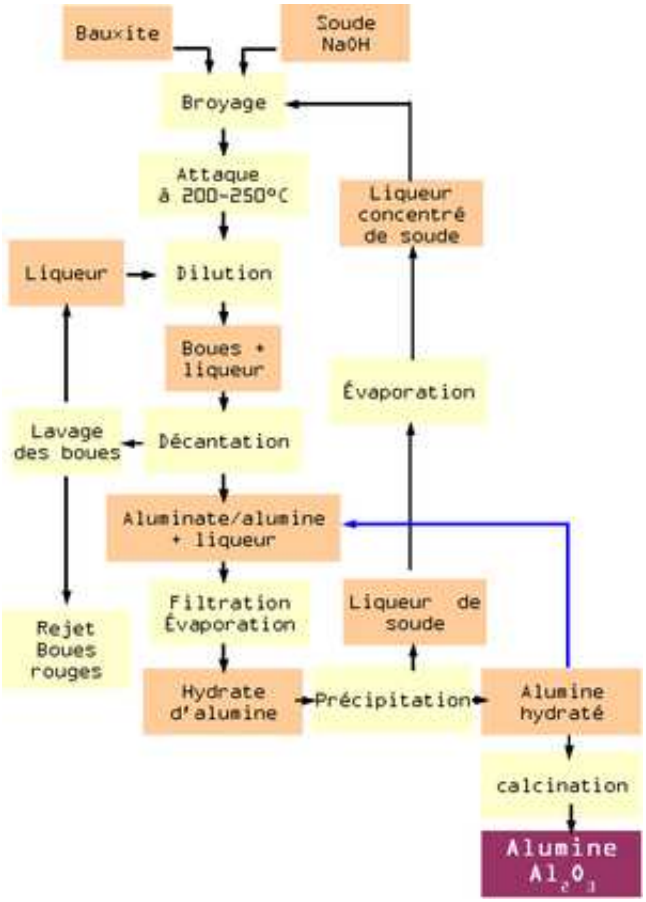


FIG. 1.1 – Extraction de l'alumine : procédé Bayer

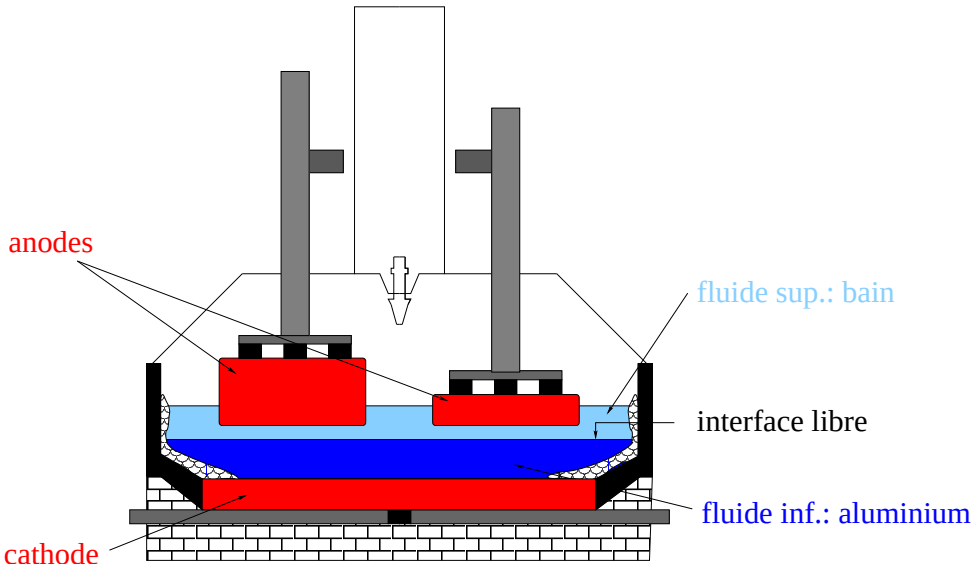


FIG. 1.2 – Réduction de l'alumine en aluminium : procédé Hall-Héroult

1.2 Le procédé Hall-Héroult

En 1886, les ingénieurs américain Hall et français Héroult ont indépendamment mis au point une technique consistant à fabriquer l'aluminium à l'état liquide par une réaction d'électrolyse. Dans ce procédé, encore le seul utilisé de nos jours à l'échelle industrielle, l'alumine et le carbone font office de réactifs, le premier étant réduit en aluminium et le deuxième oxydé en dioxyde de carbone, suivant le bilan global :

$$2\,Al_2O_3 + 3\,C \rightarrow 4\,Al + 3\,CO_2\,. \tag{1.1}$$

Cette équation-bilan est très synthétique et mérite d'être quelque peu détaillée, ne serait-ce que pour mettre en valeur la réduction de l'alumine et l'oxydation du carbone ; ainsi, on peut sommairement décomposer ce bilan en les deux sous-réactions :

$$\left\{\begin{array}{rcl} 4\,Al^{3+} + 12\,e^- & \rightarrow & 4\,Al\,, \\ 6\,O^{2-} + 3\,C & \rightarrow & 3\,CO_2 + 12\,e^-\,. \end{array}\right.$$

En fait, la réaction est encore plus complexe, la cryolite (Na_3AlF_6) servant non seulement de solvant, mais aussi de vecteur pour le transport des électrons par l'intermédiaire des ions sodium : elle se décompose en ions fluorures F^-, fluoroaluminates AlF_4^- et sodium Na^+. À noter qu'il existe une réaction concurrente génératrice de monoxyde de carbone :

$$2\,Al + 3\,CO_2 \rightarrow Al_2O_3 + 3\,CO\,, \tag{1.2}$$

qui est donc contre-productive puisque qu'elle resynthétise l'alumine.

1.2.1 Aspects thermiques

Pour le bon déroulement du procédé, l'aluminium et l'électrolyte doivent se trouver tous deux à l'état liquide. Il en résulte que la température dans la cuve doit se situer au-dessus de la plus haute température de fusion des deux composants, qui est celle de l'électrolyte (960 ^{o}C contre 660 ^{o}C pour l'aluminium). Ainsi, le bain et l'aluminium sont portés à 970 ^{o}C, de sorte qu'au voisinage direct des zones liquides (au-dessus du bain et sur les parois de la cuve) se trouvent des morceaux d'électrolyte solidifié à 940 ^{o}C environ. Cette phase solide joue son propre rôle dans l'équilibre de la cuve, car elle constitue un isolant thermique au niveau de la surface du bain et des parois de la cuves, améliore le trajet des lignes de courant (en servant d'isolant électrique autour des électrodes), et protège les parois contre l'attaque de la cryolite. Ce dernier point est peut-être le plus important, car la fonte du bain solidifié provoquée par une surchauffe peut sérieusement endommager la cuve en la perçant, et stopper ainsi son fonctionnement.

Cet apport calorifique, obtenu par effet Joule lorsque le courant passe dans l'électrolyte (cf. *infra*), doit être bien maîtrisé, car si une bonne isolation thermique permet *a priori* de diminuer les pertes d'énergie, une surchauffe à l'intérieur de la cuve peut conduire à une dilution de l'aluminium pur dans la cryolite, ce qui occasionne des pertes de rendement (cf. [37]). De plus, une quantité trop importante de bain solidifié peut dégrader le passage du courant électrique, à l'inverse de l'effet bénéfique (car stabilisateur, voir le paragraphe suivant) qu'il peut avoir en diminuant les courants horizontaux (cf. [11], [87]). Ainsi, le maintien de la température des fluides autour de 970 ^{o}C est d'une grande importance, mais est en partie autorégulé par la quantité d'électrolyte solidifié : par exemple, s'il y a augmentation de la température, l'épaisseur de la phase solide diminuera, ce qui aura comme effet d'augmenter le flux de chaleur à travers la paroi, et de permettre une stabilisation de la température interne. On pourra consulter [11] pour plus de précisions sur l'influence de l'électrolyte gelé dans les cuves.

1.2.2 Aspects magnétohydrodynamiques

Comme dans toute réaction d'électrolyse, un apport de courant électrique est nécessaire au transfert des électrons du réducteur à l'oxydant, c'est à dire ici du carbone à l'alumine. Dans les cuves de production d'aluminium, l'intensité du courant est extrêmement élevée (plusieurs centaines de kiloampères), car la quantité d'aluminium produit est directement proportionnelle à cette intensité. Le passage du courant et son convoyage jusqu'à la cuve par des conducteurs extérieurs génèrent un champ magnétique très important, qui donne naissance à des forces électromagnétiques de type Laplace-Lorentz en se combinant avec le courant électrique :

$$\vec{F_{\mathrm{L}}} = \vec{j} \times \vec{B}\,, \quad (\ \vec{j}\ : \text{densité de courant électrique, } \vec{B}\ : \text{champ magnétique}). \tag{1.3}$$

Ces forces agissent à leur tour sur les fluides conducteurs que sont l'aluminium et l'électrolyte, et génèrent ainsi un mouvement résultant notamment de l'interaction entre des courants horizontaux et le champ magnétique ambiant. Ces courants sont en général dus à des différences d'altitude de l'interface électrolyte-aluminium d'un endroit à l'autre de la cuve : l'électrolyte étant moins bon conducteur que l'aluminium, le courant passe préférentiellement aux endroits où l'interface est haute, ce qui aboutit à des redistributions horizontales de courant dans l'aluminium (conducteur "parfait") :

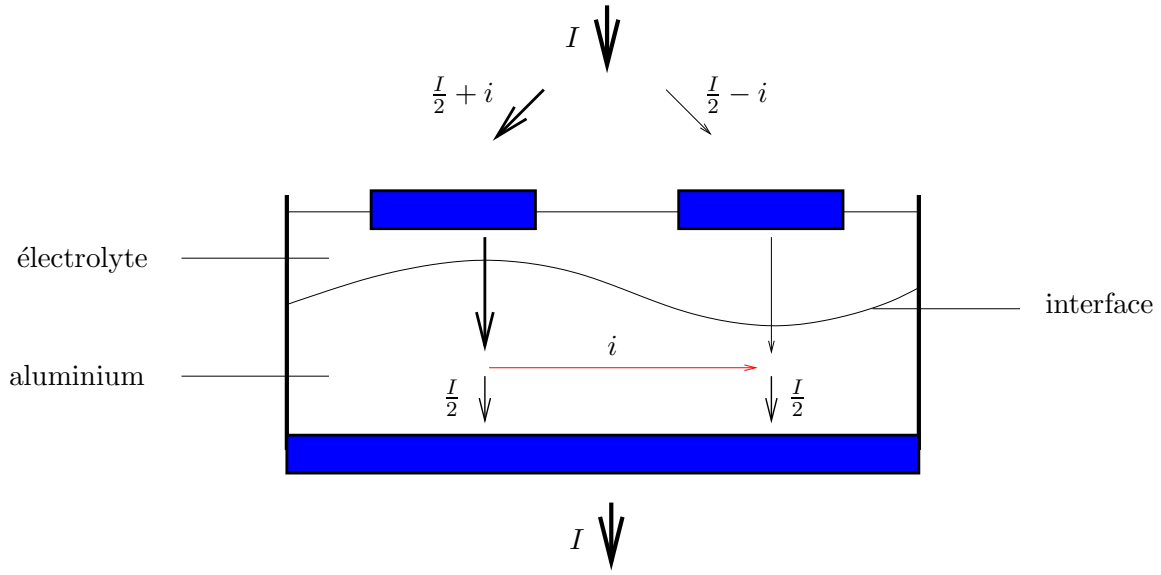


FIG. 1.3 – Influence de la forme de l'interface sur les courants électriques horizontaux (exemple)

On admet communément que les forces de Lorentz influençant le plus le mouvement sont dues à la seule composante verticale du champ magnétique, principalement créée par les conducteurs extérieurs et les autres cuves. Il en résulte par interaction avec les courants horizontaux une force horizontale perpendiculaire au plan de la figure 1.3. Nous aurons l'occasion de développer plus loin ce mécanisme qui est souvent considéré comme *la principale source des instabilités magnétohydrodynamiques* apparaissant dans les cuves. Notons enfin qu'une cuve est entourée d'un caisson en acier de quelques centimètres d'épaisseur qui possède des propriétés ferromagnétiques, et modifie ainsi l'influence des éléments extérieurs par un effet d'"écran magnétique".

1.3 Enjeux

1.3.1 Quelques chiffres

Les cuves d'électrolyse sont de forme approximativement parallélépipédique, de longueur $10m$, largeur $3m$ et hauteur $1m$ environ. La vue en coupe des figures 1.2 et 1.3 est en fait une vue dans le plan de la largeur et de la hauteur. Les anodes, d'une durée de vie de 20 jours, sont au nombre de plusieurs dizaines par cuve et disposées suivant deux rangées parallèles dans la longueur (d'où la présence de deux blocs sur les figures 1.2 et 1.3). Une usine standard compte plusieurs centaines de cuves (jusqu'à 300) connectées en série et agencées comme ci-dessous :

FIG. 1.4 – Photo d'une série de cuves (Alcan Primary Metal Group)

Ainsi, la différence de potentiel entre l'anode et la cathode étant de $4V$ environ (somme du potentiel électrochimique spécifique de la réaction : $1{,}2V$ et d'une surtension essentiellement due au passage du courant dans le mauvais conducteur qu'est l'électrolyte), la tension aux bornes d'une série de cuves peut dépasser $1000V$. L'intensité du courant, qui augmente sans cesse, atteint aujourd'hui $500kA$ sur la dernière génération de cuves, et génère ainsi des champs magnétiques 200 fois plus importants ($0.01T$) que le champ magnétique terrestre. Il est à noter que les dimensions horizontales des cuves augmentent continuellement elles aussi, de manière à maintenir des densités de courant assez basses pour limiter la vitesse des fluides, qui est de l'ordre de $0.2m/s$.

Au fil des décennies, la quantité d'aluminium produite chaque année dans le monde n'a cessé d'augmenter, passant de moins d'1 million de tonnes en 1945 à plus de 30 millions aujourd'hui. Le principal pays producteur est la Chine avec près de 6 millions de tonnes annuelles, loin devant la Russie, le Canada et les États-Unis qui sortent chacun de leurs usines environ 3 millions de tonnes par an. D'un autre côté, la quantité d'énergie (principalement électrique) nécessaire à la production d'une tonne d'aluminium a diminué de 30% au cours des 35 dernières années pour atteindre aujourd'hui environ $13\,MWh$. Il n'en demeure pas moins que cette industrie consomme près de 2% de l'électricité mondiale ; ainsi, pour donner un ordre de grandeur, le fonctionnement d'une usine moyenne nécessite la puissance d'une demi-tranche de centrale nucléaire. Cet aspect énergétique, qui est au centre des préoccupations des entreprises productrices d'aluminium, fait l'objet du paragraphe suivant.

Terminons ce bref aperçu des aspects quantitatifs en donnant quatre paramètres importants au sujet des deux fluides intervenant dans le procédé Hall-Héroult :

	Aluminium	Électrolyte
Densité	$\rho_1 = 2300\, kg.m^{-3}$	$\rho_2 = 2150\, kg.m^{-3}$
Viscosité	$\eta_1 = 1.196.10^{-3}\, kg.m^{-1}.s^{-1}$	$\eta_2 = 2.558.10^{-3}\, kg.m^{-1}.s^{-1}$
Conductivité	$\sigma_1 = 3.5.10^{6}\, \Omega^{-1}.m^{-1}$	$\sigma_2 = 2.5.10^{2}\, \Omega^{-1}.m^{-1}$
Hauteur de fluide	$h_1 = 0.2\, m$	$h_2 = 0.05\, m$

TAB. 1.1 – Quelques paramètres importants

1.3.2 Questions de rendement

Un premier critère d'efficacité du procédé est le *rendement Faraday* : c'est le rapport entre la quantité d'aluminium effectivement produite et la quantité théorique prévue par la loi de Faraday, compte tenu du courant électrique fourni. La principale cause de dégradation de ce rendement est la réaction parasite de recombinaison de l'alumine (1.2). On mesure d'ailleurs le rendement Faraday par la quantité de monoxyde de carbone indésirablement produit. L'apparition de bulles de gaz sous les anodes (*effet d'anode*), provoquée essentiellement par une trop faible concentration en alumine dans l'électrolyte, peut également être à l'origine d'une chute du rendement Faraday. Ce dernier phénomène est cependant bien maîtrisé à présent, si bien que de nos jours le rendement Faraday atteint en moyenne 90% (et même 95% dans certaines configurations).

Pour améliorer le fonctionnement des cuves, il est utile de considérer un critère d'efficacité plus global appelé *rendement énergétique*, celui-là avoisinant seulement les 50% (cf. [37]). Il exprime, pour une quantité donnée d'aluminium produit, le rapport entre la variation d'enthalpie ΔH (énergie thermochimique) nécessaire à la réaction d'oxydoréduction, et l'énergie électrique totale fournie W_{el}, qui est en partie perdue sous forme d'énergie thermique (qu'on note W_{th}) :

$$r_e = \frac{\Delta H}{W_{\text{el}}}, \quad \text{avec} \quad W_{\text{el}} = \Delta H + W_{\text{th}}$$

Actuellement, les deux quantités ΔH et W_{th} sont donc à peu près les mêmes. La première étant une constante de la réaction (1.1), on n'a que l'option de diminuer W_{el} si l'on veut augmenter r_e. Sachant (cf. [37]) que W_{el} est directement proportionnel à la différence de potentiel $E = 4V$ aux bornes d'une cuve (en considérant le rendement Faraday comme approximativement égal à 1), la seule solution pour améliorer le procédé est donc de réduire E, qui comprend justement pour moitié la surtension liée au passage du courant dans l'électrolyte ($2V$) ! *C'est donc essentiellement l'effet Joule provoqué par la faible conductivité de l'électrolyte qui est responsable des pertes thermiques*, et c'est pourquoi la distance entre les anodes et l'aluminium fondu, qu'on appelle encore DAM pour *distance anode-métal* (représenté par h_2 dans le tableau TAB. 1.1) est si faible.

Même si des perfectionnements sont toujours possibles, l'optimisation des coûts énergétiques passe donc en grande partie par la diminution de la DAM, et non par l'augmentation du rendement Faraday. Pour se fixer les idées, on notera qu'un demi-centimètre de hauteur d'électrolyte coûte environ 100 millions de dollars par an.

Parallèlement à cela, on observe sur le terrain qu'augmenter (resp. diminuer) la distance anode-métal a un effet stabilisant (resp. déstabilisant) sur le déroulement du procédé. Tout l'enjeu est ainsi de pouvoir augmenter le rendement des cuves sans les déstabiliser, ce qui constitue un cas typique de problème d'optimisation sous contraintes.

Chapitre 2

De la modélisation à l'optimisation

Dans la nature, l'industrie et l'économie, de nombreux phénomènes suivent un comportement complexe qu'il est difficile de prévoir avec précision, en raison des contraintes matérielles faisant obstacle à la multiplication des expériences, et même à la simple mesure. L'électrolyse de l'alumine en fait partie, du fait de la variété des phénomènes physiques qu'elle met en jeu et des conditions extrêmes du procédé Hall-Héroult (température, corrosion, courants intenses).

Ainsi, on préfère souvent *simuler* les expériences, en construisant un *modèle* mathématique, par lequel on espère relier de façon formelle le comportement d'un système réel (*variables d'état*) aux conditions censées déterminer ce comportement (*paramètres*). Une fois cette loi établie, si on a réussi à démontrer qu'à un jeu de paramètres réalistes correspond un et un seul jeu de variables d'état (problème *bien posé*), on peut considérer avoir terminé la phase de *modélisation*. Il reste alors à *résoudre* le modèle par une méthode adaptée pour obtenir la prévision escomptée. Évidemment, la complétion de cette seconde étape dépend fortement de la complexité du modèle : en général, celui-là prend la forme d'un système d'équations différentielles, qu'on peut parfois résoudre "à la main" (ce qui mène à des solutions *explicites*), mais qui se révèle bien souvent inextricable par des méthodes analytiques. La plupart des modèles physiques entrent dans la deuxième catégorie, car ils sont basés sur des *équations aux dérivées partielles*, comme les modèles usuels de la thermique (équation de la chaleur), de l'électromagnétisme (équations de Maxwell) et de la mécanique des fluides (équations de Navier-Stokes), qu'on peut utiliser pour décrire le procédé Hall-Héroult. La seule issue est alors le recours à la simulation *numérique*, domaine scientifique qui a vu le jour dans les années 1950, suite à l'apparition des ordinateurs.

La simulation, qui se substitue donc - au moins en partie - à l'expérimentation matérielle, est aujourd'hui fondamentale dans de nombreux secteurs scientifiques. En plus de permettre des économies conséquentes (en temps et en argent) par cette substitution, elle joue aussi un rôle de validation des modèles et des expériences, et offre des possibilités bien plus grandes en matière d'exploration, la variation des paramètres et variables d'état étant virtuelle donc sans limites. Cependant, cette technique est dans une large mesure tributaire de la puissance de calcul, car même si celle-là évolue constamment, tous les problèmes posés sont loin d'être résolus, et certains ne le seront sûrement jamais. Pour cette raison, la recherche d'algorithmes efficaces, qui constitue une branche des mathématiques à part entière appelée *analyse numérique*, est déterminante dans la portée de toute démarche de simulation. Dans le cas où les performances des méthodes de résolution sont encore insuffisantes au regard de la précision escomptée, la seule alternative est alors de simplifier le modèle en limitant au maximum la dégradation de sa fiabilité. Dans le problème physique qui nous intéresse, nous retenons deux modèles, l'un (non linéaire) se voulant plus précis, et l'autre (linéaire) plus rapide à résoudre.

Ainsi, nous disposons de deux moyens d'estimer le comportement des cuves en fonction de leurs paramètres d'entrée. Ceux-là, qui décrivent l'environnement dans lequel se déroule le procédé, sont constitués de caractéristiques imposées par le contexte physique (densité, viscosité, conductivité, attraction, etc.), et de variables "réglables" (forces électromagnétiques, conditions aux limites, distance anode-métal, etc.), ce terme restant à définir au regard des moyens matériels à disposition et du coût de leur utilisation. On désigne par *données* les paramètres de la première catégorie, et par *commandes* ou *actionneurs* ceux de la deuxième. Si l'on suppose cette dernière non vide, on peut envisager de modifier le phénomène physique, en vue d'optimiser un certain aspect de son comportement. Alors, l'existence d'un modèle déduisant les variables d'état des actionneurs est d'une grande utilité, car il permet de recourir à la théorie mathématique du *contrôle optimal* pour déterminer automatiquement, à partir d'un critère fixé portant sur les variables d'état (*fonction coût*), des valeurs d'actionneurs qui engendrent le comportement escompté.

Remarques

1) Du point de vue du modèle, les configurations optimales évoquées ci-dessus peuvent être remplacées par d'autres types de comportements prescrits. Dans ce cadre, le contrôle optimal permet de cerner les conditions dans lesquelles se produisent certains phénomènes.

2) Dans le cas où les données ne sont pas connues précisément, mais où ce manque d'information est sans conséquences sur la validité mathématique du modèle, il est possible de les déterminer si l'on dispose d'observations physiques sur le phénomène. Mathématiquement, il s'agit d'un problème de contrôle optimal dans lequel ces données prennent la forme de commandes servant à minimiser l'écart entre le résultat de la simulation et l'observation. Cette méthode, qui permet de *caler* ou *calibrer* les modèles, porte le nom de *problème inverse*.

On voit que d'une part, avoir un modèle à disposition permet d'exploiter la théorie du contrôle optimal, et que d'autre part, l'utilisation d'un programme de contrôle permet d'obtenir des informations qualitatives et quantitatives sur la validité du modèle. En plus de permettre la modification d'un comportement en fonction d'objectifs préétablis, le contrôle optimal est donc utile à la modélisation des phénomènes, au même titre que la simulation en elle-même (et la remplace même avantageusement dans des procédures du type essai-erreur). Pour les mettre en œuvre, on cherche souvent les solutions de problèmes de minimisation du type (où $\mathcal{E}$, $\mathcal{P}$ et $\mathcal{F}$ sont à définir) :

$$\inf_{u \in U} \mathcal{J}(u)\,, \quad \text{avec} \quad \mathcal{J}(u) = \mathcal{E}(\varphi) + \mathcal{P}(u)\,, \quad \text{et } \varphi \text{ tel que } \mathcal{F}(\varphi, u) = 0\,.$$

Les variables d'état (vitesse, pression, etc.) sont ici représentées par φ, les commandes par u, la fonction coût par $\mathcal{J}$ et les équations d'état (qui comprennent les données) par $\mathcal{F} = 0$. Ainsi, l'étude mathématique et la résolution numérique de ce problème reposent en partie sur celles des équations d'état, qui prennent dans notre cas la forme d'équations aux dérivées partielles (EDP).

Contexte mathématique

Même s'il n'existe pas de théorie générale pour le problème ci-dessus, les deux modèles utilisés dans les présents travaux possèdent des points communs à la fois au niveau :

- des outils mathématiques employés pour l'étude de leurs solutions,
- des méthodes par lesquelles on les résout numériquement,
- des algorithmes de contrôle optimal auxquels ils peuvent se prêter.

Leur présentation détaillée fera l'objet du chapitre **3**, mais on peut d'ores et déjà les classer pour l'un dans la catégorie des problèmes *paraboliques-elliptiques* non linéaires, et pour l'autre dans celle des problèmes *hyperboliques du deuxième ordre* linéaires. Ces deux classes sont couramment

représentées par des "problèmes types" linéaires, respectivement :

$$\text{- l'}\textit{équation de la chaleur}\text{ :} \quad \left\{ \begin{array}{rcll} \partial_t \varphi(t,x) - \Delta\varphi(t,x) & = & f(t,x)\,, & \forall\,(t,x) \in [0,T]\times\Omega\,, \\ \varphi(0,x) & = & \varphi^0(x)\,, & \forall\, x \in \Omega\,, \end{array} \right. \tag{2.1}$$

$$\text{- l'}\textit{équation des ondes}\text{ :} \quad \left\{ \begin{array}{rcll} \partial_t^2 \varphi(t,x) - \Delta\varphi(t,x) & = & f(t,x)\,, & \forall\,(t,x) \in [0,T]\times\Omega\,, \\ \left\{ \begin{array}{rcl} \varphi(0,x) & = & \varphi^0(x)\,, \\ \partial_t\varphi(0,x) & = & \varphi^1(x)\,, \end{array} \right. & & & \forall\, x \in \Omega\,, \end{array} \right. \tag{2.2}$$

où $T \in \mathbb{R}^+$, Ω est un ouvert dans $\mathbb{R}^d$, et f et $(\varphi^k)_{0\le k\le m-1}$ sont des données. Lorsque $\Omega = \mathbb{R}^d$, ces deux modèles rentrent dans le formalisme du *problème de Cauchy*, qui constitua le premier véritable sujet d'étude des EDP en tant que telles, et pour lequel le *théorème de Cauchy-Kovalevsky* (1842) fournit un résultat d'existence et d'unicité des solutions analytiques au voisinage de $x = 0$, pour des données analytiques. Ce type de solutions a pour inconvénient de ne pas satisfaire la *condition de Hadamard* (1923), qui veut que la solution dépende continûment des données, en conformité avec les systèmes physiques modélisés. Par ailleurs, un phénomène physique est souvent restreint à un ouvert Ω *borné* dans $\mathbb{R}^d$, aux frontières duquel il est nécessaire d'imposer des *conditions aux limites*. La nature locale du théorème de Cauchy-Kovalevsky se prête mal à de telles contraintes globales. C'est ainsi qu'on est amené à s'affranchir du cadre des fonctions classiques en considérant les solutions comme des *distributions*, qui sont mathématiquement définies dans le cadre d'une formulation *variationnelle* des EDP. Souvent, on peut physiquement justifier ce formalisme par le *principe de moindre action*, comme par exemple dans le cas de l'équation des ondes définie sur un intervalle $[0,\ell]$, représentant les vibrations d'une corde fixée en ses extrémités :

$$\left\{ \begin{array}{rcll} \mu\,\partial_t^2\varphi(t,x) - \kappa\,\partial_x^2\varphi(t,x) & = & 0\,, & \forall\,(t,x) \in [0,T]\times[0,\ell]\,, \\ \varphi(t,0) = \varphi(t,\ell) & = & 0\,, & \forall\, t \in [0,T]\,, \\ \varphi(0,x) & = & \varphi^0(x)\,, & \forall\, x \in [0,\ell]\,, \\ \partial_t\varphi(0,x) & = & \varphi^1(x)\,, & \forall\, x \in [0,\ell]\,. \end{array} \right. \tag{2.3}$$

En effet, si on modélise la corde par un système de $N_h + 2$ points $(x_j, y_j = \varphi(x_j))_{0\le j\le N_h+1}$ de masse m reliés par des fils élastiques sans masse, tels que $x_{j+1} - x_j = h$, $m = \mu\,h$, $x_0 = 0$ et $x_{N_h+1} = 0$:

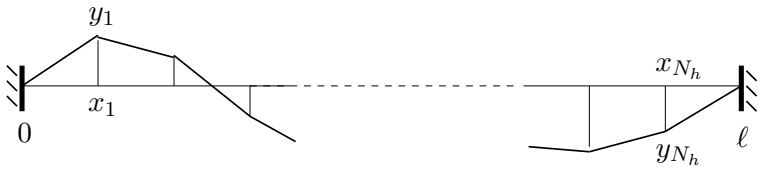


FIG. 2.1 – Chaîne élastique

l'équation en $\varphi(t,x)$ devient, en remplaçant $\partial_x\varphi(x)$ par $\dfrac{\varphi(x+h/2) - \varphi(x-h/2)}{h}$:

$$m\,\ddot{y}_j(t) - k\,[y_{j-1}(t) - 2y_j(t) + y_{j+1}(t)] = 0\,, \quad \forall\, j \in \{1,\dots,N_h\}\,, \tag{2.4}$$

où $k = \kappa/h$ est la constante de raideur des fils. Les équations de Newton (2.4) peuvent être vues comme une condition nécessaire et suffisante de stationnarité de l'*action* $\mathcal{A}(y = (y_1,..,y_{N_h}))$ pour tout *déplacement virtuel admissible* purement vertical $\tilde{y} = (\tilde{y}_1,..,\tilde{y}_{N_h})$:

$$(2.4) \;\Leftrightarrow\; \lim_{\varepsilon\to 0} \frac{\mathcal{A}(y+\varepsilon\tilde{y}) - \mathcal{A}(y)}{\varepsilon} = 0\,, \quad \forall\,\tilde{y}\,, \quad \text{avec} \quad \mathcal{A}(y) = \int_0^T \frac{1}{2}\sum_{j=1}^{N_h} \left[m\dot{y}_j^2 - k\,(y_{j+1}-y_j)^2 \right]. \tag{2.5}$$

Alors, en faisant tendre N_h vers l'infini (soit h vers dx et $\sum$ vers $\int$) dans (2.5), on obtient que l'EDP (2.3) est une condition de stationnarité de

$$\mathcal{A}(\varphi) = \int_0^T \int_0^\ell \frac{1}{2} \left[\mu \, (\partial_t \varphi)^2 - \kappa \, (\partial_x \varphi)^2 \right],$$

qui s'exprime sous forme variationnelle

$$\lim_{\varepsilon \to 0} \frac{\mathcal{A}(\varphi + \varepsilon \tilde{\varphi}) - \mathcal{A}(\varphi)}{\varepsilon} = 0\,, \quad \forall\, \tilde{\varphi} \quad \Leftrightarrow \quad \mu \int_0^T \int_0^\ell \partial_t \varphi \, \partial_t \tilde{\varphi} - \kappa \int_0^T \int_0^\ell \partial_x \varphi \, \partial_x \tilde{\varphi} = 0\,, \quad \forall\, \tilde{\varphi}\,.$$

Pour un exposé approfondi des principes variationnels en dynamique, on pourra consulter B.A. Kupershmidt [57], qui couvre des modèles en magnétohydrodynamique. Plus généralement, on obtient la formulation variationnelle d'une EDP en la multipliant par une *fonction test* adéquate $\tilde{\varphi}$, puis en l'intégrant par parties en tenant compte des conditions aux limites. On est alors en mesure de traiter des problèmes tels que (2.1) ou (2.2) posés sur des domaines bornés, qu'on qualifie de *problèmes aux limites*. Cette manière de procéder implique l'utilisation d'espaces fonctionnels particuliers pour manipuler les distributions, les *espaces de Sobolev*. Ainsi, nous présentons en section **2.1** les rudiments de la théorie variationnelle des équations des ondes et de la chaleur, avant de nous pencher sur leur approximation numérique en section **2.2**. Sans le vouloir, nous l'avons déjà utilisée ci-dessus en assimilant une corde à un nombre fini de points reliés par des élastiques. En effet, nous avons été en mesure de réduire l'EDP (de dimension infinie) à un système de dimension finie, mais aussi d'approcher l'opérateur de dérivation $-\Delta$ par une somme d'opérateurs de multiplication, et d'obtenir par là un système d'*équations différentielles ordinaires* (EDO). Il s'agit d'un schéma aux *différences finies*, qui constituent (au moins dans certaines configurations) un cas particulier de la méthode des *éléments finis*, de nature fondamentalement variationnelle.

Enfin, de la même manière que la formulation variationnelle des modèles physiques permet de caractériser la stationnarité de l'action, les solutions des problèmes de contrôle étudiés remplissent une condition nécessaire du premier ordre variationnelle portant sur la stationnarité du critère $\mathcal{J}$. Celle-là est définie non pas par rapport à la variable d'état, mais par rapport à la commande. À titre d'exemple, considérons le problème de contrôle de la température d'une barre de longueur ℓ par un terme source défini sur $[0, \ell]$. Le problème est d'obtenir une répartition des températures la plus proche possible de la distribution φ_{opt}, en limitant la norme de u pour modéliser le coût de l'activation de la source :

$$\inf_{u \in U} \mathcal{J}(u)\,, \quad \text{où} \quad \mathcal{J}(u) = \frac{1}{2} \int_0^T \int_0^\ell \|\varphi(u) - \varphi_{\text{opt}}\|^2 + \frac{Q}{2} \int_0^T \int_0^\ell \|u\|^2\,, \quad Q \in \mathbb{R}_+^*\,, \tag{2.6}$$

$$\text{avec} \quad \left\{ \begin{array}{rcll} \partial_t \varphi(t,x) - \kappa \, \partial_x^2 \varphi(t,x) & = & u(t,x)\,, & \forall\, (t,x) \in [0,T] \times [0,\ell]\,, \\ \varphi(t,0) = \varphi(t,\ell) & = & 0\,, & \forall\, t \in [0,T]\,, \\ \varphi(0,x) & = & \varphi^0(x)\,, & \forall\, x \in [0,\ell]\,. \end{array} \right. \tag{2.7}$$

Le gradient de la fonction coût (2.6) s'écrit

$$\nabla_u \mathcal{J}(u) = \lambda + Qu\,,$$

où λ est solution du *problème adjoint* :

$$\left\{ \begin{array}{rcll} -\partial_t \lambda(t,x) - \kappa \, \partial_x^2 \lambda(t,x) & = & \varphi(u)(t,x) - \varphi_{\text{opt}}(t,x)\,, & \forall\, (t,x) \in [0,T] \times [0,\ell]\,, \\ \lambda(t,0) = \lambda(t,\ell) & = & 0\,, & \forall\, t \in [0,T]\,, \\ \lambda(T,x) & = & 0\,, & \forall\, x \in [0,\ell]\,, \end{array} \right. \tag{2.8}$$

$\varphi(u)$ étant la solution de l'équation d'état (2.7) pour un certaine valeur de u. C'est ainsi que la condition de stationnarité $\lambda + Qu = 0$ avec les équations (2.7) et (2.8) vérifiées caractérise la (les) solution(s) (u, φ) du problème (2.6)-(2.7). Nous présentons en section **2.3** la généralisation de ce type de propriété à plusieurs types de problèmes de contrôle et son rôle dans leur résolution.

2.1 Outils d'analyse fonctionnelle pour l'étude des modèles

Contrairement aux problèmes hyperboliques du premier ordre dont l'étude mathématique est assez délicate (chocs...), les équations (2.1) et (2.2) comportent l'opérateur Δ, qui génère un *semi-groupe continu* de *contractions* par la propriété d'*ellipticité* de $-\Delta$. Avant de préciser ces aspects, on peut déjà exhiber par la transformée de Fourier la structure de semi-groupe des opérateurs qui, à une condition initiale $\Phi^0 = (\varphi^k)_{0\le k\le m-1}$ associent la solution $\Phi(t) = (\varphi^{(k)}(t))_{0\le k\le m-1}$, pour $m = 1$ et $m = 2$. On ramène les équations (2.1) ($m = 1$) et (2.2) ($m = 2$) à des systèmes d'EDO.

2.1.1 Transformée de Fourier

Aux XVIII^e et XIX^e siècles, lorsque les premières EDP apparaissent dans les modèles physiques, on commence par rechercher des solutions *explicites* ("transcendantes") à leur formulation (forte) newtonienne au moyen de divers outils, dont la transformée de Fourier. Pour pouvoir définir celles-là dans le cadre moderne des distributions, on considère les inconnues $\varphi^{(k)}(t,\cdot)$ dans l'espace $\mathcal{S}'(\mathbb{R}^d)$ des *distributions tempérées*. Ainsi, leurs transformées de Fourier en espace $\widehat{\varphi}^{(k)}(t,\cdot)$ satisfont :

$$\left\{\begin{array}{rcl} \widehat{\varphi}^{(m)}(t) + 4\pi^2\|\xi\|^2\,\widehat{\varphi}(t) &=& \widehat{f}(t)\,, \quad \forall\, t\in\mathbb{R}^+\,, \\ \widehat{\varphi}^{(k)}(0) &=& \widehat{\varphi}^k\,, \quad \forall\, k\in\{0,..,m-1\}\,, \end{array}\right. \tag{2.9}$$

où ξ joue le rôle d'un paramètre. Par une extension aux distributions des résultats concernant les EDO classiques, on montre que ce problème est bien posé dans $\mathcal{C}^m(\mathbb{R}^+,\mathcal{S}'(\mathbb{R}^d))$, et que sa solution prend la forme (voir R. Dautray et J.-L. Lions [22] V) :

$$\widehat{\varphi}(t) = \sum_{k=0}^{m-1} \widehat{E}_{m-k}(t)\,\widehat{\varphi}^k + \int_0^{+\infty} \widehat{E}_0(t-s)\,\widehat{f}(s)\,ds\,,$$

où les $\widehat{E}_k$ ($k = 0,..,m$) sont solutions des problèmes (δ désignant le symbole de Kronecker) :

$$\left\{\begin{array}{rcl} \widehat{E}_k^{(m)}(t) + 4\pi^2\|\xi\|^2\,\widehat{E}_k(t) &=& 0\,, \quad \forall\, t\in\mathbb{R}^+\,, \\ \widehat{E}_k^{(l)}(0) &=& \delta_{m-k,l}\,, \quad \forall\, l\in\{0,..,m-1\}\,. \end{array}\right.$$

Pour (2.1) comme pour (2.2), on peut calculer explicitement $(\widehat{E}_k(t))_{0\le k\le m}$, et il existe une seule solution $\varphi(t,x)$ dans $\mathcal{C}^\infty(\mathbb{R}^+,\mathcal{S}'(\mathbb{R}^d))$, donnée par la transformée de Fourier inverse en espace de $\widehat{\varphi}(t,\xi)$. Alors, la solution s'écrit, en utilisant une matrice de convolution en espace (cf. [22] XIV) :

$$\Phi(t) = G(t)\Phi^0 + (G \underset{t}{*} F)(t)\,, \quad \forall\, t\ge 0\,,$$

où la famille $(G(t))_{t\in\mathbb{R}^+}$ d'opérateurs de $\mathcal{L}(\mathcal{S}'(\mathbb{R}^d))$ forme un *semi-groupe* : $G(s+t) = G(s)G(t)$.

Étudions à présent le cas Ω borné, soit $[0,\ell]$ en dimension 1 pour simplifier : on peut utiliser la transformée de Fourier à condition de prolonger φ à $\mathbb{R}$ tout entier, par sa "périodisée en espace"

$$\varphi_P(t,x) = \varphi(t,x) \underset{x}{*} W_\ell(x)\,, \quad \text{avec} \quad W_\ell(x) = \sum_{j\in\mathbb{Z}} \delta(x-j\ell) \quad (\delta : \text{mesure de Dirac})\,.$$

$$\text{Alors} \quad \widehat{\varphi_P}(t,\xi) = \frac{1}{\ell}\,\widehat{\varphi}(t,\xi)\,W_{\frac{1}{\ell}}(\xi)\,, \quad \text{et donc} \quad \varphi_P(t,x) = \frac{1}{\ell}\sum_{j\in\mathbb{Z}} \widehat{\varphi}(t,\frac{j}{\ell})\,\mathrm{e}^{2i\pi j\frac{x}{\ell}}\,.$$

On obtient ainsi une représentation de l'inconnue sous forme de *série de Fourier*, qu'on peut récrire compte tenu des conditions aux limites $\varphi(t,0) = \varphi(t,\ell) = 0$:

$$\varphi(t,x) = \sum_{j\in\mathbb{N}} y_j(t)\sin(\omega_j x)\,, \quad \text{avec} \quad \omega_j = \frac{2\pi j}{\ell}\,, \tag{2.10}$$

de sorte que (2.3) et (2.7), par exemple, prennent la forme de systèmes d'EDO. Nous allons voir, au moyen des espaces de Sobolev, que cette propriété s'inscrit dans la théorie *spectrale* du laplacien.

2.1.2 Formulation variationnelle et espaces de Sobolev

Dans la théorie classique des fonctions, l'opérateur de dérivation n'est pas continu, ce qui rend d'un emploi malaisé la notion de convergence des solutions d'EDP. C'est pourquoi, dans la première moitié du XX^e siècle, les mathématiciens commencèrent à s'intéresser à la notion de solution *faible*, au sens des distributions. Ce point de vue fournit un cadre théorique dans lequel la dérivation est une opération régulière au sens suivant : tout élément de l'espace des distributions $\mathcal{D}'(\Omega)$ est indéfiniment dérivable, et si une suite de distributions (ψ_n) converge vers ψ, alors $(\partial\psi_n)$ converge vers $\partial\psi$. En plus de cette propriété, il est utile de définir une *norme* sur l'espace considéré, et que la propriété de *complétude*[1] soit remplie au sens de cette norme. C'est ainsi que les opérateurs linéaires dans les *espaces de Banach*, qui sont des espaces vectoriels normés complets, vérifient, en vertu du théorème de l'application ouverte (cf. [15]), la condition de Hadamard. Par ailleurs, dans les problèmes aux limites, on a besoin de définir les distributions aux frontières d'un domaine (notion de *trace*), ce qui requiert certaines propriétés de *régularité*. Enfin, la définition d'un *produit scalaire* permet d'utiliser la notion d'orthogonalité, qui procure de nombreux avantages pratiques. Toutes ces considérations amènent Sobolev à définir en 1934 un nouveau type d'espaces fonctionnels, qui donneront naissance à la théorie des distributions, formalisée par Schwartz en 1950. Les distributions s'imposeront dès lors dans le traitement des EDP linéaires, mais aussi non linéaires car celles-là génèrent, même à partir de données régulières, des solutions singulières.

Si dx représente la mesure de Lebesgue, les formes bilinéaires symétriques positives

$$(\varphi,\psi)\mapsto(\varphi,\psi)_{H^p(\Omega)}=\int_\Omega \partial^\alpha\varphi(x)\,\partial^\alpha\psi(x)\,dx\,,\quad \alpha=(\alpha_i)_{1\le i\le d}\,,$$

lorsqu'elles existent, permettent de définir les produits scalaires $(\varphi,\psi)_{H^p(\Omega)}$ et les normes $\varphi\mapsto|\varphi|_{H^p(\Omega)}=\sqrt{(\varphi,\varphi)_{H^p(\Omega)}}$. Ainsi, les *espaces de Sobolev* d'ordre p

$$H^p(\Omega)=\Big\{\psi\in\mathcal{D}'(\Omega)\;\Big|\;\int_\Omega(\partial^\alpha\psi(x))^2\,dx<\infty\,,\;\forall|\alpha|\in\{0,..,p\}\Big\}$$

sont des espaces de Banach, contrairement aux espaces $\mathcal{C}^p(\Omega)$ munis de la même norme. On les qualifie parfois d'espaces à *énergie finie* car $(\partial^\alpha\varphi)^2$ représente souvent, pour une certaine valeur de α, une densité d'énergie dans les problèmes traités. De plus, les produits scalaires ci-dessus en font des *espaces de Hilbert*, qui admettent une base orthonormale dénombrable (ils sont isomorphes à l^2). En généralisant ainsi la distance euclidienne aux espaces de dimension infinie, ils sont bien adaptés aux approximations numériques (*méthodes de Galerkin*). Une de leurs propriétés fondamentales est le *théorème de Riesz-Fréchet*, selon lequel toute forme linéaire continue sur un espace de Hilbert V s'exprime sous la forme du produit scalaire par un élément de V, et réciproquement :

$$\left\{\begin{array}{l}\forall\,L\in V'\,,\quad\exists!\,\varphi_L\in V\;\;/\;\;L(\psi)=(\varphi_L,\psi)_V\,,\quad\forall\,\psi\in V\,,\\ \forall\,\varphi\in V\,,\quad\exists!\,L_\varphi\in V'\;\;/\;\;(\varphi,\psi)_V=L_\varphi(\psi)\,,\quad\forall\,\psi\in V\,.\end{array}\right.$$

Il en résulte qu'un espace de Hilbert peut être identifié à son dual, ce qui en fait un espace de Banach *réflexif*. On dispose alors de la propriété fondamentale suivante (*compacité faible*) : pour toute suite bornée dans V, il existe une sous-suite faiblement convergente[2].

[1] Toute suite de Cauchy (φ_n) est *fortement* convergente : $\exists\,\varphi\,,\ \lim_{n\to\infty}|\varphi-\varphi_n|=0$.

[2] On dit que (φ_n) converge *faiblement* vers φ si $\forall\,\psi\in V,\ \lim_{n\to\infty}(\varphi_n,\psi)_V=(\varphi,\psi)_V$.

Une conséquence directe en est la propriété de *minimisation des fonctionnelles convexes*, dont la démonstration sera donnée en section **2.3.2.b** dans le cadre des problèmes de contrôle optimal. En particulier, on a le théorème suivant :

Si V est un espace de Hilbert, $f \in V'$ et $a(\cdot,\cdot)$ est une forme bilinéaire sur V

$$\begin{array}{lll} \text{(i)} & \text{continue :} & \exists\, c \in \mathbb{R}_+ \,\forall\,(\varphi,\psi) \in V^2\,, \quad a(\varphi,\psi) \le c\,|\varphi|_V\,|\psi|_V\,, \\ \text{(ii)} & \text{coercive :} & \exists\, k \in \mathbb{R}_+^* \,\forall\,\varphi \in V\,, \quad a(\varphi,\varphi) \ge k\,|\varphi|_V^2\,, \\ \text{(iii)} & \text{symétrique :} & \forall\,(\varphi,\psi) \in V^2\,, \quad a(\varphi,\psi) = a(\psi,\varphi)\,, \end{array} \tag{2.11}$$

alors le problème $a(\varphi,\psi) = \langle f,\psi\rangle_{V',V}\ \forall \psi \in V$ admet une unique solution dans V.

Ce résultat s'étend aux formes bilinéaires continues coercifs non nécessairement symétriques (*théorème de Lax-Milgram*, [60]). On voit donc qu'il peut être avantageux de poser les problèmes sous forme variationnelle dans les espaces de Hilbert. Les espaces de Sobolev le permettent, dès lors qu'ils sont choisis d'ordre suffisamment élevé (fonctions "régulières") pour pouvoir intégrer par parties des distributions sur des domaines bornés. À titre d'exemple, considérons le problème de Dirichlet homogène

$$\left\{ \begin{array}{rcll} -\Delta\varphi & = & f & \text{dans } \Omega\,, \\ \varphi & = & 0 & \text{sur } \partial\Omega\,. \end{array} \right.$$

Ce problème est bien posé si l'on choisit φ dans $H^1(\Omega)$, car, d'abord, $\mathcal{D}(\overline{\Omega})$ est *dense* dans $H^1(\Omega)$, ce qui permet de définir la trace de φ dans $L^2(\partial\Omega)$, et par conséquent l'espace $H_0^1(\Omega)$. En multipliant alors la première équation par une fonction test $\tilde{\varphi} \in H_0^1(\Omega)$ et en intégrant par parties (*formule de Green*) sur Ω, on se retrouve dans les conditions du théorème ci-dessus, l'ellipticité de a résultant de l'inégalité de Poincaré (qui n'est pas valable dans $H^1(\Omega)$)

$$\forall\,\psi \in H_0^1(\Omega)\,, \quad \exists\, C_\Omega \ge 0 \quad / \quad \sum_{i=1}^{d} |\partial_{x_i}\psi|^2_{L^2(\Omega)} \ge C_\Omega |\psi|^2_{L^2(\Omega)}\,.$$

Les espaces H^p possèdent de nombreuses autres propriétés fondamentales : lorsque $\Omega = \mathbb{R}^n$, ce sont des sous-espaces des distributions *tempérées* $\mathcal{S}'$ (stables par transformation de Fourier). Par ailleurs, les espaces H^p ont des propriétés de *compacité* regroupées sous le nom d'*injections de Sobolev*, qui interviennent dans la justification des propriétés *spectrales* de certains opérateurs, ou encore dans l'obtention de résultats de convergence dans les problèmes non linéaires. C'est ainsi que nous allons pouvoir expliquer la formule (2.10) obtenue en section précédente : ce résultat s'inscrit en fait dans la théorie spectrale des opérateurs coercifs (ou *elliptiques*) symétriques :

Soient V et H deux espaces de Hilbert réels tels que l'injection de V dans H soit compacte, $a(\cdot,\cdot)$ une forme bilinéaire sur V répondant aux conditions (2.11). Les solutions du problème :

$$a(\varphi,\psi) = \lambda\,(\varphi,\psi)_H\,, \quad \forall\,\psi \in V \tag{2.12}$$

sont les couples $(\lambda_j, w_j)_{j\in\mathbb{N}}$, où les valeurs propres $(\lambda_j)_j$ forment une suite croissante dans $\mathbb{R}_+^*$ tendant vers $+\infty$, et les vecteurs propres $(w_j)_j$ forment une base orthonormale de l'espace H.

Ainsi, dans le cas des séries de Fourier, la famille $\left(\sqrt{\frac{2}{\ell}}\sin(\omega_j x)\right)_{j\in\mathbb{N}}$ forme une base orthonormale de l'espace $H = \{\varphi \in L^2(]0,\ell[) \mid \varphi(0) = \varphi(\ell) = 0\}$. On dispose en effet du *théorème de Rellich* (cf. [15]), qui prévoit que lorsque Ω est borné, alors l'injection de V dans $L^2(\Omega)$ est compacte si $H_0^1(\Omega) \subseteq V \subseteq H^1(\Omega)$. On pourra consulter [84] pour l'application de ce type de raisonnement à de nombreux problèmes aux limites linéaires (dont (2.1) et (2.2) posés sur des domaines bornés).

2.1.3 Estimations *a priori*

Les théorèmes ci-dessus sont très puissants pour l'étude des problèmes linéaires, *une fois mis en place les espaces fonctionnels permettant leur application*. Ils permettent en général de séparer les variables temps et espace en décomposant pour tout t l'inconnue $\varphi(t,\cdot)$ sur une base (infinie) des espaces en question, de sorte que l'EDP considérée se ramène à un système d'EDO de dimension infinie, pour l'analyse duquel on dispose alors du *théorème de Cauchy-Lipschitz* :

Soient V un espace de Banach et $F : V \to V$ une application lipschitzienne :

$$\exists\, M \in \mathbb{R}^+ \quad / \quad \forall\, (\varphi,\psi) \in V^2\,, \quad |F(\varphi) - F(\psi)|_V \leq M|\varphi - \psi|_V\,. \tag{2.13}$$

Alors, pour tout $\varphi^0 \in E$, il existe $\varphi \in \mathcal{C}^1(\mathbb{R}^+, E)$ unique telle que $\left\{ \begin{array}{rcll} d_t\varphi(t) & = & F(\varphi(t)) & \text{sur } \mathbb{R}^+\,, \\ \varphi(0) & = & \varphi^0 & \text{en } t = 0\,. \end{array} \right.$

Une grande partie de la question repose ainsi sur l'obtention de la propriété (2.13), qui sert en fait de guide dans le choix de l'espace V. Par exemple, si l'on multplie les problèmes de Cauchy-Dirichlet que constituent les équations de la chaleur et des ondes respectivement par φ et $\partial_t\varphi$, on obtient après intégration sur $[0,t]\times\Omega$ les *égalités d'énergie*

$$\frac{1}{2}\,|\varphi(t)|_H^2 + \int_0^t \sum_{i=1}^d |\partial_{x_i}\varphi(s)|_H^2\, ds = \frac{1}{2}|\varphi(0)|_H^2 + \int_0^t (f(s),\varphi(s))_H\, ds \quad \text{et} \tag{2.14}$$

$$\frac{1}{2}\left(|\partial_t\varphi(t)|_H^2 + \sum_{i=1}^d |\partial_{x_i}\varphi|_H^2\right) = \frac{1}{2}\left(|\partial_t\varphi(0)|_H^2 + \sum_{i=1}^d |\partial_{x_i}(0)\varphi|_H^2\right) + \int_0^t (f(s),\partial_t\varphi(s))_H\, ds\,. \tag{2.15}$$

où $H = L^2(\Omega)$. Alors, l'ellipticité de $-\Delta$ dans $V = H_0^1(\Omega)$, et l'application au second membre des inégalités d'Young pour la première, et de Cauchy-Schwartz et de Gronwall pour la deuxième, conduisent aux *estimations a priori* (avec $T \geq t$) :

$$|\varphi(t)|_H^2 + \int_0^t |\varphi(s)|_V^2\, ds \leq |\varphi^0|_H^2 + C\int_0^T |f(t)|_{V'}^2\, dt \quad \text{et}$$

$$|\varphi(t)|_V^2 + |\partial_t\varphi(t)|_H^2 \leq C\left(|\varphi^0|_V^2 + |\varphi^1|_H^2 + \int_0^T |f(t)|_H^2\, dt\right),$$

où les $C \in \mathbb{R}^+$ désignent des constantes diverses. C'est ainsi qu'on arrive à déterminer des espaces de solutions en fonction de ceux auxquels appartiennent les données, et à valider la condition de Hadamard. Pour le moment, l'existence de ces solutions n'est pas encore prouvée, donc le raisonnement demeure formel ("*a priori*"). Il peut cependant être appliqué à des solutions approchées à variables séparées de la forme :

$$\varphi_h(t,x) = \sum_{j=1}^{N_h} y_j(t)\, w_j(x)\,, \qquad (w_j)_{j\in\mathbb{N}_h} \text{ base de } V\,,$$

de sorte qu'on obtient une estimation sur les solutions $(\varphi_h)_{N_h\in\mathbb{N}}$ qui, cette fois, *existent*, en vertu du théorème de Cauchy-Lipschitz. Il en résulte que la suite (φ_h) est bornée dans $L^\infty([0,T],H)\cap L^2([0,T],V)$ pour l'équation de la chaleur, et dans $L^\infty([0,T],V)$ pour l'équation des ondes. De plus, pour cette dernière, la suite $(\partial_t\varphi_h)$ est bornée dans $L^\infty([0,T],H)$. Alors, la propriété de compacité faible (cf. **2.1.2**) permet d'extraire de ces suites des sous-suites qui convergent faiblement. On obtient ainsi l'existence des solutions faibles, en démontrant que ces limites sont bien solutions des équations (2.1) et (2.2), et qu'elles sont uniques. Nous ne détaillerons pas ces derniers points et renvoyons à J.-L. Lions et E. Magenes ([66] vol. 1) pour plus de précisions.

Le point de vue des semi-groupes

Sous l'hypothèse de régularité supplémentaire sur la solution $\varphi(t) \in H^2(\Omega)$, le *théorème de Hille-Yosida* ([49], [99]) permet d'expliciter le semi-groupe $(G(t))_{t\geq 0}$, et de démontrer sa continuité dans l'espace avec $W = H$ (resp. $W = V \times H$) pour $m = 1$ (resp. $m = 2$). Pour cela, on considère l'opérateur A défini sur le sous-espace $H^2(\Omega) \cap V$ (resp. $H^2(\Omega) \cap V \times V$) noté $D(A)$, de sorte que

$$\partial_t \Phi = A\Phi \quad \text{soit} \quad A = \Delta \quad \left(\quad \text{resp.} \quad A = \begin{bmatrix} 0 & I \\ \Delta & 0 \end{bmatrix} \quad \right).$$

Dans les deux cas, $D(A)$ est *dense* dans W, et l'opérateur A possède de plus la propriété d'être fermé. On est alors en mesure de définir son spectre, et son ensemble résolvant $\rho(A)$, qui est le complémentaire de son spectre dans $\mathbb{C}$. Alors, le théorème de Hille-Yosida s'énonce :

S'il existe $(\omega, \mu) \in \mathbb{R} \times \rho(A)$ *avec* $\mathrm{Re}(\mu) > \omega$, *et* $M \geq 1$ *tels que*

$$|(\mu I - A)^{-k}|_{\mathcal{L}(W)} \leq \frac{M}{(\mathrm{Re}(\mu) - \omega)^k}, \quad \forall\, k \geq 0,$$

alors A *est le générateur d'un semi-groupe* G *(fortement) continu sur* W, *avec*

$$G(t) = \mathrm{e}^{At} = \sum_{k\geq 0} \frac{A^k t^k}{k!}, \quad \text{et} \quad |G(t)|_{\mathcal{L}(W)} \leq M\, \mathrm{e}^{\omega t}.$$

Ainsi, en prenant $\{\mu = 1, \omega = 0, M = 1\}$ si $m = 1$ et $\{\mu = 2, \omega = 1, M = 1\}$ si $m = 2$, on obtient l'existence de $(\mu I - A)^{-1}$ par le théorème de Lax-Milgram, et les estimations *a priori* permettent d'exhiber l'inégalité requise. On remarquera que $|G(t)|_{\mathcal{L}(W)} \leq 1$ quel que soit t dans le cas de la chaleur, ce qui signifie que G est un semi-groupe de *contractions*. On peut démontrer que le semi-groupe assosié à l'équation des ondes possède également cette propriété (cf. [22] Chap. XVII).

2.2 Calcul scientifique et analyse numérique

Plus peut-être que tout autre domaine mathématique, les équations aux dérivées partielles étaient prédisposées à bénéficier de l'utilisation des ordinateurs. Un des premiers travaux en ce sens - le mémoire de Daniel Bernoulli publié en 1753 - contenait deux procédés d'approximation de la solution du problème de corde vibrante que nous avons déjà rencontrés : l'un consistait à remplacer la corde par un nombre fini de masses ponctuelles reliées par un fil élastique sans masse, l'autre à développer la solution en série trigonométrique. Ces idées sont le reflet de deux méthodes encore dominantes aujourd'hui en analyse numérique des EDP, respectivement :

- la discrétisation par interpolation (méthodes de différences finies, d'éléments finis, etc.).
- la représentation des solutions par des intégrales (transformées de Fourier, de Laplace, etc.), suivie d'une approximation portant sur ces représentations,

Dans les travaux présentés ici, trois procédés d'approximation seront utilisés : les méthodes des différences et des éléments finis, et les méthodes spectrales (de la deuxième catégorie). Nous en exposons ci-dessous les grandes lignes, en guise d'introduction au paragraphe **3.3**.

Le calcul scientifique consiste à trouver des techniques de résolution informatique de problèmes complexes pour lesquels aucune solution explicite n'est connue. Il est important de noter que cette approche fournit souvent des preuves - qu'on appelle alors *constructives* - d'existence et d'unicité de solutions abstraites au niveau continu (par passage à la limite). On entre alors dans le domaine de l'analyse numérique, dont l'objet est d'établir des preuves de convergence pour les méthodes de discrétisation employées.

2.2.1 Méthodes d'interpolation

2.2.1.a Différences finies

L'outil universel de résolution des problèmes aux dérivées partielles a longtemps été la méthode des *différences finies*, dont le principe est simple : presque par définition, les expressions

$$\frac{\psi(x)-\psi(x-h)}{h} \quad \text{(dérivée } \textit{retardée}\text{)} \quad \text{et} \quad \frac{\psi(x+h)-\psi(x)}{h} \quad \text{(dérivée } \textit{avancée}\text{)}$$

sont, pour h petit, des approximations de $\partial_x\psi$. En soustrayant la première à la deuxième et en divisant le tout par h, on obtient une dérivée du deuxième ordre approchée *centrée*, qu'on peut utiliser pour construire une approximation en espace de l'équation de corde vibrante (2.3). Si l'on procède de la sorte en chaque point d'abscisse $x_j = jh$, $j = 1,..,N_h$ avec $h = d$ et qu'on cherche $y_j = \varphi(x_j)$, on retrouve simplement le système (2.4). Pour que la discrétisation soit totale, il faut également échantillonner le temps, et il reste alors à s'assurer que le *schéma* ainsi construit permet d'obtenir une solution approchée qui *converge* vers la solution exacte lorsque les pas d'échantillonnage, en temps et en espace, tendent vers zéro. Pour cela, le schéma doit posséder les propriétés de *consistance* et de *stabilité*. On considère ainsi le problème modèle

$$\partial_t^m y(t) + Ry(t) = 0\,, \quad \text{avec} \quad y = [y_1 \dots y_{N_h}]^T\,, \quad \text{et} \quad R = \frac{1}{h^2}\left(-\delta_{j-1,j'} + 2\delta_{j,j'} - \delta_{j,j'-1}\right)_{1\le j,j'\le N_h}\,,$$

qu'on échantillonne en temps en définissant le pas $\Delta t = T/N_f$, $N_f \in \mathbb{N}^*$. On peut alors écrire de manière générale le schéma en fonction d'une matrice $G_h(\Delta t) = (G_h^{k,i}(\Delta t))_{0\le k,i\le m-1}$:

$$y^{k,n+1} = \sum_{i=0}^{m-1} G_h^{k,i}(\Delta t)\, y^{i,n}\,, \quad \forall\, k \in \{0,\dots,m-1\}\,,$$

où $y^{k,n}$ représente la version totalement discrétisée de la dérivée $k^{\text{ème}}$ de la solution semi-discrétisée $\partial_t^k y$ à l'instant $n\Delta t$. Le schéma est dit *consistant* d'ordre p en temps et q en espace si

$$\exists\,(p,q) \in \mathbb{N}^{*2} \quad / \quad \sup_{n,k} \ \frac{1}{\Delta t}\left\| \varphi^{k,n+1} - \sum_{i=0}^{m-1} G_h^{k,i}(\Delta t)\,\varphi^{i,n} \right\| = \mathcal{O}(\Delta t^p + h^q)\,,$$

$\varphi^{k,n}$ étant la projection sur le maillage de la solution exacte $\partial_t^k\varphi$ à l'instant $n\Delta t$. Enfin, le schéma est *stable* s'il existe $K \ge 0$ indépendant de Δt et h tel que $\|G_h(\Delta t)^n\| \le K\,,\ \forall\, n \in \mathbb{N}$.

1) L'équation de la chaleur ($m = 1$) est très souvent discrétisée par le *θ-schéma* ($\theta \in [0,1]$) :

$$\frac{y^{n+1}-y^n}{\Delta t} = -R\left[\theta y^{n+1} + (1-\theta)y^n\right].$$

Par exemple, on montre par des développements de Taylor sur φ que le schéma d'*Euler explicite* ($\theta = 0$) est consistant d'ordre 1 en temps et 2 en espace. Pour l'étude de sa stabilité, il résulte de la transformée de Fourier de l'égalité $\overline{y^n} = G_h(\Delta t)^n\overline{y^0}$ (où l'on a interpolé y^n par $\overline{y^n}$) :

$$\widehat{\overline{y^n}}(\xi) = \Big[1 - \frac{4\,\Delta t}{h^2}\sin^2\Big(\frac{\xi\, h}{2}\Big)\Big]^n\,\widehat{\overline{y^0}}(\xi)$$

et de la relation de Plancherel que $\|y^n\| \le \|y^0\|$ (et donc $\|G_h(\Delta t)^n\| \le 1$) si et seulement si la *condition CFL* (pour Courant-Friedrichs-Lewy, cf. [19]) est vérifiée :

$$\Delta t \le \frac{h^2}{2}\,. \tag{2.16}$$

Les même démarche permet de montrer que le schéma d'*Euler implicite* ($\theta = 1$) est consistant d'ordre 1 en temps, 2 en espace, et inconditionnellement stable.

2) Dans le cas de l'équation des ondes ($m = 2$), on utilise souvent la *méthode de Newmark* :

$$\left\{ \begin{array}{rcl} \dfrac{y^{0,n+1} - y^{0,n}}{\Delta t} & = & y^{1,n} - \Delta t\, R\,[\theta_0\, y^{0,n+1} + (\frac{1}{2} - \theta_0)\, y^{0,n}]\,, \\ \dfrac{y^{1,n+1} - y^{1,n}}{\Delta t} & = & -R\,[\theta_1 y^{0,n+1} + (1-\theta_1) y^{0,n}]\,. \end{array} \right.$$

Lorsque $\theta_1 = 1/2$, on montre que celle-là est consistante d'ordre 2 par des développements de Taylor à l'ordre 3 sur φ et à l'ordre 2 sur $\partial_t \varphi$. De plus, si on choisit $\theta_0 = 0$, on obtient alors une méthode explicite qui est stable seulement si

$$\Delta t \leq h\,. \tag{2.17}$$

En effet, par le même type de raisonnement (transformée de Fourier) que pour l'équation de la chaleur, on obtient en posant $\omega(\xi) = \frac{2}{h} \sin\left(\frac{\xi h}{2}\right)$:

$$\left\{ \begin{array}{rcl} \widehat{y^{0,n+1}} & = & \left(1 - \dfrac{\Delta t^2}{2}\,\omega^2\right) \widehat{y^{0,n}} + \Delta t\, \widehat{y^{1,n}}\,, \\ \widehat{y^{1,n+1}} & = & \left(\dfrac{\Delta t^3}{4}\,\omega^4 - \Delta t\,\omega^2\right) \widehat{y^{0,n}} + \left(1 - \dfrac{\Delta t^2}{2}\,\omega^2\right) \widehat{y^{1,n}}\,, \end{array} \right.$$

Ainsi, on montre que la norme spectrale (subordonnée à la norme euclidienne dans $\mathbb{R}^2$) de cette matrice est inférieure à 1 seulement si (cf. [84])

$$\frac{\Delta t^2}{h^2} \sin^2\left(\frac{\xi h}{2}\right) \leq 1\,,$$

d'où le résultat par la relation de Plancherel.

2.2.1.b Éléments finis

Penchons-nous encore une fois sur l'exemple de la "chaîne élastique" (FIG. 2.1 p.9). Soit $l \in \{1, .., N_h + 1\}$ le numéro d'un ressort compris entre les points (x_{l-1}, y_{l-1}) et (x_l, y_l), et (x, y) un point de ce ressort. Ce dernier satisfait l'équation affine $y = ax + b$, avec

$$a = \frac{y_l - y_{l-1}}{x_l - x_{l-1}} \quad \text{et} \quad b = \frac{y_{l-1}\, x_l - y_l\, x_{l-1}}{x_l - x_{l-1}}\,,$$

que l'on peut récrire sous la forme $y = y_{l-1}\, \tau_1(x) + y_l\, \tau_2(x)$, avec

$$\tau_1(x) = \frac{x_l - x}{x_l - x_{l-1}} \quad \text{et} \quad \tau_2(x) = \frac{x - x_{l-1}}{x_l - x_{l-1}}\,.$$

On a ainsi décomposé la fonction $y(x)$ de l'espace vectoriel $\mathbb{P}^1([x_{l-1}, x_l]) = \text{Vect}\,\{1, x\}$ sur une base formant une *partition de l'unité* :

$$\tau_1(x) + \tau_2(x) = 1\,, \quad \forall\, x \in\, [x_{l-1}, x_l]\,,$$

et si $K = [x_{l-1}, x_l]$, $P = \mathbb{P}^1(K)$, $M_1 = x_{l-1}$ et $M_2 = x_l$ les points de K tels que $\tau_i(M_j) = \delta_{ij}$, et $\Sigma = \{M_1, M_2\}$, alors l'application

$$\begin{array}{rcl} P & \to & \mathbb{R}^{\text{Card}\,\Sigma} \\ p & \mapsto & (p(M_1), \ldots, p(M_{\text{Card}\,\Sigma})) \end{array}$$

est bijective. On dit que P est Σ-*unisolvant*, et cela définit un *élément fini de Lagrange d'ordre 1.*

En construisant ainsi pour chaque intervalle K^l les fonctions τ_1^l et τ_2^l, si on considère l'application

$$\ell_g \; : \; \begin{array}{ccc} \{1,..,N_h+1\}\times\{1,2\} & \to & \{0,..,N_h+1\} \\ (l,i) & \mapsto & j \end{array} \; ,$$

et si on définit les *fonctions de formes* par

$$w_j = \sum_{\ell_g(l,i)=j} \tau_i^l \, , \tag{2.18}$$

alors

$$\{\psi \in H_0^1([0,\ell]) \,|\, \psi_{|K^l} \in \mathbb{P}^1(K^l) \, , \, l=1,..,N_h+1\} = \underset{1\le j\le N_h}{\text{Vect}} \, (w_j) \, .$$

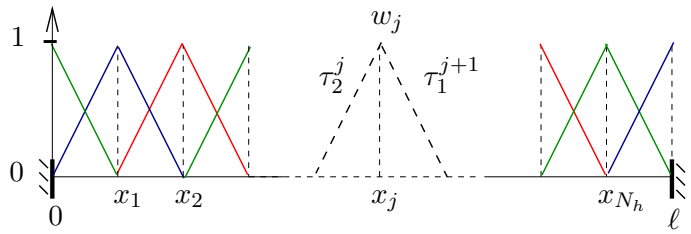


FIG. 2.2 – Fonctions de forme P^1-*Lagrange* dans $H^1([0,\ell])$

À l'aide de cette base, on peut chercher une approximation en espace de la solution φ du problème continu sous la forme

$$\varphi_h(t,x) = \sum_{j=1}^{N_h} y_j(t)\, w_j(x) \, , \quad \text{avec} \quad y_j(t) = \varphi_h(x_j,t) \, .$$

Ainsi, la discrétisation de la formulation variationnelle en espace du problème (2.3) :

$$\mu \frac{\partial^2}{\partial t^2} \int_0^\ell \varphi\, \tilde\varphi + \kappa \int_0^\ell \frac{\partial \varphi}{\partial x} \frac{\partial \tilde\varphi}{\partial x} = 0 \, , \quad \forall \, \tilde\varphi \in H_0^1([0,\ell]) \, , \quad \forall \, t \in [0,T]$$

dans $V_h = \text{Vect}_{1\le I\le N_h} \, (w_I)$ s'écrit

$$\mu \frac{\partial^2}{\partial t^2} \int_0^\ell \varphi_h\, \tilde\varphi_h + \kappa \int_0^\ell \frac{\partial \varphi_h}{\partial x} \frac{\partial \tilde\varphi_h}{\partial x} = 0 \, , \quad \forall \, \tilde\varphi_h \in V_h \, , \quad \forall \, t \in [0,T] \, .$$

Matriciellement, on obtient - avec une dérivée temporelle d'ordre quelconque et un second membre interpolé dans V_h de composantes $[f_h]$ dans la base (w_j) :

$$My^{(m)}(t) + Ry(t) = M[f_h](t) \, , \quad \text{avec} \quad M = \left(\mu \int_0^\ell w_I \, w_J \right)_{I,J} \quad \text{et} \quad R = \left(\kappa \int_0^\ell \frac{\partial w_I}{\partial x} \frac{\partial w_J}{\partial x} \right)_{I,J} .$$

Comme dans le paragraphe précédent, on peut achever de discrétiser ce système d'EDO par différences finies en temps. Cette méthode se généralise aisément, d'une part, en dimension supérieure, et d'autre part à des supports K_l de tailles différentes, c'est dire à des maillages spatiaux *non structurés*. Elle est qualifiée d'approximation *interne* ou de *Galerkin* parce qu'elle approche la solution exacte dans un sous-espace vectoriel de son espace d'existence.

En dehors des éléments finis lagrangiens d'ordre 1, qui seront au centre des méthodes de simulation présentées ici, il existe de nombreux autres éléments finis, adaptés à divers types de problèmes. L'avantage des éléments finis sur les différences finies apparaît dans les problèmes dont la géométrie du domaine de définition requiert d'utiliser un maillage non structuré. On peut d'ailleurs interpréter les différences finies comme un cas particulier des éléments finis dans lequel, d'une part, le maillage est structuré, et, d'autre part, la matrice M est approchée par une quadrature particulière, qui donne lieu à une version diagonale de M (*mass lumping*, cf. [29]).

Ainsi, dans la résolution du modèle non linéaire, nous utiliserons la méthode très répandue qui consiste à discrétiser le temps par différences finies et l'espace par éléments finis.

2.2.1.c Résolution des systèmes linéaires

Après discrétisation totale, les deux méthodes présentées ici conduisent souvent à des schémas du type (pour les problèmes du premier ordre en temps) :

$$\left(\frac{1}{\Delta t}\,M + \theta\,R\right) y^{n+1} = M z^n + \left(\frac{1}{\Delta t}\,M + (\theta - 1)\,R\right) y^n\,,$$

qui requièrent ainsi de résoudre en chaque temps $i\Delta t$ un système linéaire de la forme $Ax = b$. La taille N de l'inconnue numérique x étant de l'ordre du produit du nombre d'inconnues scalaires par le nombre de points d'interpolation, il se peut que x ait plusieurs milliers (voire millions) de composantes. Nous aurons typiquement $N \simeq 10^5$ dans les applications. Il est donc primordial de choisir une méthode d'inversion du système linéaire performante. Les méthodes *directes* (du type élimination de Gauss) ne sont pas efficaces compte tenu de la structure des matrices, en général creuses et en bandes. En effet, elles ont pour effet de transformer la matrice creuse en matrice pleine, ce qui est embarrassant du point de vue de l'occupation mémoire. C'est pourquoi on emploie plutôt des méthodes *itératives*, reposant sur une modification progressive de l'inconnue à partir du *résidu* du problème $r_k = b - A\,x_k$, si on note x_k l'inconnue après k itérations. La plus simple est l'algorithme de Richardson stationnaire : en définissant par P et N deux matrices positives telles que P soit inversible et $A = P - N$, on effectue

$$x_{k+1} \leftarrow x_k + \alpha\,P^{-1} r_k\,, \tag{2.19}$$

où $\alpha \geq 0$ est fixé. L'intérêt de cette technique est de ne pas traiter directement la matrice A, mais plutôt un problème possédant des propriétés de vitesse de convergence accrues par le *préconditionneur* P. Les méthodes de différences et d'éléments finis, en effet, conduisent à des matrices d'autant plus *mal conditionnées* que la disctrétisation est fine. Cet aspect est en relation avec le spectre de la matrice, qui fait intervenir le pas du maillage (cf. **2.2.1.a**, **2.2.3**). On renvoie à [83] et [85] pour une description détaillée, ainsi qu'en section **6.3.2** pour un exemple.

2.2.2 Décomposition spectrale

Comme nous l'avons vu en **2.1.2**, les vecteurs propres de certains opérateurs constituent une base orthogonale de certains espaces fonctionnels utiles pour l'analyse (continue) de certaines EDP. C'est notamment le cas des équations du type des ondes, de la chaleur, de l'élasticité linéaire ou encore de l'hydrodynamique. Dans le cas où l'on peut calculer ces vecteurs propres, d'une part il est possible de déterminer par projection les composantes du membre de droite sur la base qu'ils forment, et d'autre part l'équation se trouve grandement simplifiée après substitution de l'opérateur par une "matrice diagonale infinie". Par troncature de la base des vecteurs propres, le problème se ramène à un système d'EDO qu'on peut résoudre par différences finies (cf. [84]).

Les méthodes spectrales, comme les méthodes d'éléments finis, rentrent ainsi dans la classe des approximations *internes*. Elles ont longtemps été l'un des principaux outils de la physique mathématique (études de stabilité à l'aide des *modes propres*), et s'étendent aujourd'hui à des situations où la diagonalisation mise en œuvre ne porte pas sur un opérateur en particulier (polynômes de Legendre etc., cf. [9]). Par ailleurs, elles sont très puissantes pour l'étude du caractère bien posé des problèmes, celle-là ne nécessitant pas la connaissance explicite du spectre mais seulement son existence. C'est le cas par exemple dans l'analyse continue des équations de la MHD linéaires et non linéaires (cf. **3**). Enfin, elles sont utiles dans l'analyse numérique des méthodes (voir le paragraphe suivant). Pour notre part, une méthode spectrale sera mise en œuvre dans la discrétisation spatiale du problème linéaire (chapitre **5**).

2.2.3 Résultats de convergence

Les méthodes variationnelles de discrétisation spatiale, de type éléments finis ou décomposition spectrale, procurent en fonction de h (ou N_h) des ordres de convergence sur l'écart entre la solution de l'équation et sa version numérique à l'instant t. Si l'on utilise alors des différences finies pour discrétiser le temps, des arguments de consistance et stabilité (cf. **2.2.1.a**) permettent d'établir des ordres de convergence pour la discrétisation totale (c'est à dire également en fontion de Δt).

On considère pour cela un ouvert borné $\Omega \in \mathbb{R}^d$, et on note $(\lambda_j, v_j)_{j\geq 1}$ une solution du problème spectral (cf. **2.1.2**) :

$$\int_\Omega \nabla\varphi \cdot \nabla\tilde{\varphi} = \lambda \int_\Omega \varphi\,\tilde{\varphi}\,, \quad \forall\,\tilde{\varphi} \in H_0^1(\Omega)\,,$$

telles que (v_j) forme une base de $H_0^1(\Omega)$ orthonormale dans $L^2(\Omega)$ et orthogonale dans $H^1(\Omega)$. Alors, la solution du problème (cf. P.-A. Raviart et J.-M. Thomas [84])

$$\partial_t^m \int_\Omega \varphi\,\tilde{\varphi} + \int_\Omega \nabla\varphi\cdot\nabla\tilde{\varphi} = \int_\Omega f\,\tilde{\varphi}\,, \quad \forall\,\tilde{\varphi} \in H_0^1(\Omega)\,,$$

est donnée par (on note $(\cdot,\cdot)$ le produit scalaire $L^2(\Omega)$)

$$\varphi(t) = \sum_{j\geq 1} (\varphi^0, v_j)\, e^{-\lambda_j t} + \int_0^t (f(s), v_j)\, e^{-\lambda_j (t-s)} \quad \text{si } m = 1\,,$$

et (en posant $\omega_j = \sqrt{\lambda_j}$)

$$\varphi_j(t) = \sum_{j\geq 1} (\varphi^0, v_j)\cos(\omega_j t) + \frac{1}{\omega_j}(\varphi^1, v_j)\sin(\omega_j t) + \frac{1}{\omega_j}\int_0^t (f(s), v_j)\sin(\omega_j(t-s)) \quad \text{si } m = 2\,.$$

Les mêmes résultats sont valables dans des sous-espaces de dimension finie V_h de $H_0^1(\Omega)$, comme celui engendré par la décomposition spectrale ci-dessus tronquée pour $j > N_h$, ou encore par une méthode d'éléments finis lagrangiens d'ordre 1. Dans ce dernier cas, si l'on suppose Ω polyédrique et partitionné en n-simplexes de diamètre inférieur à h, alors, pour toutes fonctions $\varphi \in H^1(\Omega)$ et leur version discrète $\varphi_h \in V_h$, il existe une constante $c \in \mathbb{R}^+$ telle que

$$\inf_{\varphi_h \in V_h} \Big(|\varphi - \varphi_h|_{L^2(\Omega)} + h|\varphi - \varphi_h|_{H^1(\Omega)} \Big) \leq c\, h^2 |\varphi|_{H^2(\Omega)}\,.$$

Ainsi, on est en mesure d'établir la convergence des schémas en temps explicites (cf. **2.2.1.a**) d'Euler pour l'équation de la chaleur et de Newmark pour l'équation des ondes. Si l'on note $\mu_h = \max\limits_{1\leq j\leq N_h} \lambda_j$, alors sous les hypothèses de stabilité

$$\Delta t\,\mu_h \leq 2 \quad \text{si } m = 1 \quad \text{et} \quad \Delta t^2\,\mu_h \leq 4 \quad \text{si } m = 2\,, \tag{2.20}$$

il existe une unique solution au problème (2.1) (resp. (2.2)) discrétisé dans ce cadre, qui, pour des données suffisamment régulières, satisfait l'estimation d'erreur :

$$|\varphi_h^n - \varphi(n\Delta t)|_{L^2(\Omega)} \leq O\Big(|\varphi_h^0 - \varphi^0|_{L^2(\Omega)} + h^2 + \Delta t \Big)\,,$$

resp.

$$|\varphi_h^n - \varphi(n\Delta t)|_{L^2(\Omega)} \leq O\Big(|\varphi_h^0 - \varphi^0|_{L^2(\Omega)} + |\varphi_h^1 - \varphi^1|_{L^2(\Omega)} + h^2 + \Delta t^2 \Big)\,.$$

On notera qu'en dimension 1 et sur maillage structuré (cf. **2.2.1.a**), on retrouve les conditions (2.16) et (2.17) à partir de (2.20), étant donné que les valeurs propres de la matrice R sont dans ce cas majorées par $4/h^2$ (cf. R. Dautray et J.-L. Lions [22] III).

2.3 Contrôle optimal

2.3.1 Notions d'automatique

Modifier les paramètres d'un système dans le but de lui faire adopter un comportement prescrit constitue, pourrait-on dire, le quotidien de n'importe quel être humain ou société. Pour limiter le plus possible ces interventions, il est logique de chercher à maximiser l'autonomie du système en le rendant plus "intelligent". L'exemple le plus ancien remonte à Héron d'Alexandrie, qui mit au point la fontaine à vin, destinée à maintenir une hauteur de boisson constante dans un récipient au fil des prélèvements qui s'y font, en reliant un flotteur à une soupape par un mécanisme de levier. Ce dispositif comporte toutes les caractéristiques usuelles d'un *système commandé* :

- une *donnée d'entrée* (consommation de vin), qu'on ne maîtrise pas,
- une *variable d'état* (hauteur de vin), dont on veut qu'elle prenne une certaine valeur,
- un *actionneur* (soupape), qui permet de modifier la variable d'état
- un *observateur* (flotteur), qui mesure l'état du système
- une *rétroaction* (levier), qui associe à la mesure une certaine valeur de l'actionneur.

C'est ainsi qu'on appelle *système* la transformation (naturelle) qui *détermine* les variables d'état à partir des données et des actionneurs, et *correcteur* le dispositif (humain) constitué de l'observateur et de la rétroaction :

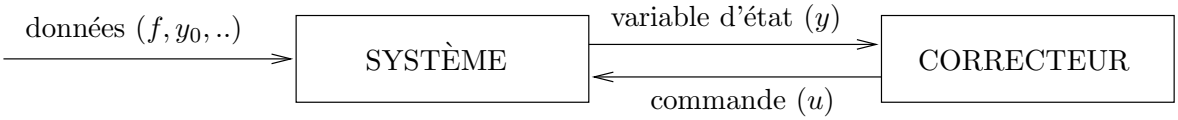


FIG. 2.3 – Système commandé

Nous nous plaçons ici dans un cas idéal, où les imperfections de l'observateur, de l'actionneur et de la rétroaction (qui sont des éléments physiques) ne sont pas prises en compte dans le modèle.

Dans le cas où $f(t)$, $y(t)$, $z(t)$ et $u(t)$ sont des vecteurs de dimension **finie**, et où le comportement du système peut être modélisé par des équations linéaires, ce type de problème rentre dans la classe des *systèmes dynamiques linéaires autonomes* (A et B sont des matrices réelles) :

$$\left\{ \begin{array}{rcl} \dot{y}(t) & = & Ay(t) + Bu(t) + f(t) \quad \text{sur } [0,T]\,, \\ y(0) & = & y_0\,. \end{array} \right. \tag{2.21}$$

Dans ce cadre, la théorie des EDO montre que l'état est relié à la commande par la loi

$$y(t) = G(t)\, y_0 + \int_0^t G(t-s)\,[Bu(s) + f(s)]\,, \quad \text{avec} \quad G(t) = \mathrm{e}^{At}\,.$$

L'objet de l'*automatique* est alors de trouver une *loi de commande* (en boucle fermée)

$$u(t) = F[t, y(t)]$$

pour construire une rétroaction (ou *feedback*) permettant de minimiser un certain critère $\mathcal{J}(u) = \mathcal{J}(y(u), u)$, comme par exemple le problème de *poursuite* (*tracking*) visant à minimiser, sur l'intervalle $[0,T]$, l'écart $z(t) = y(t) - y_{\text{opt}}(t)$ entre y et une trajectoire donnée y_{opt} :

$$\mathcal{J}(u) = \frac{\sigma}{2}\,\|z(T)\|^2 + \int_0^T \frac{1}{2}\Big[\|z(t)\|^2 + u(t)^T Q u(t)\,dt\Big]\,, \tag{2.22}$$

où $\sigma \in \mathbb{R}^+$, et Q est une matrice symétrique définie positive qui quantifie le coût de la commande. Il s'agit là d'un problème *linéaire-quadratique*, dont la théorie a été développée dans les années 1940-50, notamment par R.E. Kalman [54], L.S. Pontryagin [82] et R. Bellman [7].

Ainsi, la première question qui se pose est celle de la *contrôlabilité*, à savoir, étant donné deux états arbitraires y_0 et y_T, existe-t-il une commande $u(t)$ telle que $y(T) = y_T$? Le *critère de Kalman* permet d'y répondre en donnant la condition nécessaire et suffisante de contrôlabilité

$$\mathrm{Rg}\,[B, AB, \ldots, A^{N-1}B] = N\,, \quad \text{où } N \text{ est la taille de } y\,.$$

En supposant cette condition remplie, existe-t-il une commande $u(t)$ joignant y_0 à y_T qui, de plus, minimise le critère $\mathcal{J}(u)$? C'est le problème du *contrôle optimal*. Ici, cette propriété a lieu (et la commande optimale est même unique) car la fonction $\mathcal{J}$ est α-*convexe*, la matrice Q étant symétrique définie positive. Il reste alors à obtenir la commande, en résolvant un système qui caractérise la solution u, et si possible la détermine. Le *principe du minimum de Pontryagin* conduit à des *conditions nécessaires*[3] *d'optimalité* : pour que la commande u d'état associé $y(u)$ soit optimale, il faut qu'il existe un vecteur *adjoint* p tel que (voir par ex. [31])

$$\left\{ \begin{array}{rcl} \dot{p}(t) & = & -A^T p(t) - z(t)\,, \\ p(T) & = & \sigma\, z(T)\,, \end{array} \right. \tag{2.23}$$

et que la *fonction de Pontryagin* du système atteigne son minimum en (y, u, p) :

$$H_P(y,p,u) = \inf_v H_P(y,p,v)\,, \quad \text{avec} \quad H_P(y,p,v) = \frac{1}{2}\left(\|z\|^2 + v^T Q v\right) + p^T\left(Ay + Bv + f\right).$$

Les conditions nécessaires d'optimalité se résument ainsi par

$$\left\{ \begin{array}{rcl} \dot{y} & = & \partial_p H_P(y,p,u)\,, \\ \dot{p} & = & -\partial_y H_P(y,p,u)\,, \\ 0 & = & \partial_v H_P(y,p,u)\,, \end{array} \right. \tag{2.24}$$

qu'on peut encore exprimer par la stationnarité de la *fonction de Lagrange* $\mathcal{L}$:

$$\partial_u \mathcal{L} = \partial_y \mathcal{L} = \partial_p \mathcal{L} = 0\,, \quad \text{avec} \quad \mathcal{L}(u,y,p) = J(u,y) - \int_0^T p^T(\dot{y} - Ay - Bu - f)\,.$$

Il est par ailleurs possible d'expliciter la loi de commande par le biais de la résolution d'une équation matricielle (*de Ricatti*), mais dans notre contexte (voir ci-dessous), il est inutile de détailler cet aspect. De même, l'approche de la *programmation dynamique* de Bellman, basée sur la théorie de Hamilton-Jacobi, ne sera pas abordée en raison de son coût numérique trop élevé dans les problèmes de contrôle qui nous intéressent.

Remarque : le principe du minimum est souvent appelé le principe du *maximum* : si l'on change le signe des variables adjointes p, il faut maximiser et non minimiser le *hamiltonien* (cf. [96])

$$H(y,p,v) = p^T\left(Ay + Bv + f\right) - \frac{1}{2}\left(\|z\|^2 + v^T Q v\right).$$

Cet autre formalisme correspond à celui du *problème de Lagrange* (cf. V. Alexéev *et al.* [2]) :

$$\left\{ \begin{array}{l} \mathcal{A}(y) = \displaystyle\int_0^T L(t, y(t), \dot{y}(t)) \quad \longrightarrow \quad \mathrm{extr}\,, \\ \mathcal{G}(t, y(t), \dot{y}(t)) = 0\,, \quad y(0) = y_0\,, \quad y(T) = y_T\,. \end{array} \right.$$

En mécanique, lorsque L représente le *lagrangien* d'un système (différent de la fonction de Lagrange), l'objet de ce problème est de caractériser les trajectoires qui rendent stationnaires l'action $\mathcal{A}$ sous les contraintes $\mathcal{G} = 0$. Le principe du maximum de Pontryagin est alors utile pour traiter des contraintes du type $\dot{y} = g(t, y(t), \dot{y}(t))$ (*non holonômes*). Les variables adjointes sont encore appelées des *multiplicateurs de Lagrange*, et les conditions d'optimalité (2.24) prises avec $H(y,p,\dot{y}) = p^T g(y,\dot{y}) - L(y,\dot{y})$ à la place de H_P les équations d'*Euler-Lagrange*.

[3] et suffisantes dans ce cas précis

2.3.2 Position du problème

Nous venons de décrire un moyen permettant de commander le comportement d'un système de dimension finie au moyen d'un actionneur u apparaissant dans le second membre de l'équation d'état. La démarche a consisté, après s'être assuré de l'existence et de l'unicité de la commande pour le critère donné, à résoudre trois systèmes différentiels ordinaires pour obtenir le résultat. Nous avons par ailleurs mentionné la possibilité d'obtenir la loi de commande par l'intermédiaire de l'équation de Ricatti, ou encore de traiter le problème par la méthode de la programmation dynamique. Dans certaines situations, l'automatique peut s'étendre à des problèmes de dimension **infinie**, dans lesquels l'équation d'état est une EDP et représente ainsi un *système à paramètres distribués*. Dans ce cadre, la théorie du contrôle optimal se complexifie notablement, sur la question de la contrôlabilité comme sur celles de l'existence et de l'obtention de solutions.

2.3.2.a Contrôlabilité approchée

La régularité minimale $\varphi \in H^1$ des solutions des EDP que nous étudions (cf. **2.1.3**) empêche, par exemple, d'atteindre un état cible moins régulier $\varphi_{\rm opt} \in L^2$ dans le problème (2.6)-(2.7) :

$$\inf_{u\in U} \mathcal{J}(u)\,, \quad \text{où} \quad \mathcal{J}(u) = \frac{1}{2}\int_0^T\int_0^\ell \|\varphi(u)-\varphi_{\rm opt}\|^2 + \frac{Q}{2}\int_0^T\int_0^\ell \|u\|^2\,, \quad Q\in\mathbb{R}_+^*\,,$$

$$\text{et } \varphi(u) \text{ est tel que } \left\{\begin{array}{rcll} \partial_t\varphi(t,x) - \kappa\,\partial_x^2\varphi(t,x) &=& u(t,x)\,, & \forall\,(t,x)\,\in\,[0,T]\times[0,\ell]\,,\\ \varphi(t,0)=\varphi(t,\ell) &=& 0\,, & \forall\, t\in[0,T]\,,\\ \varphi(0,x) &=& \varphi^0(x)\,, & \forall\, x\,\in\,[0,\ell]\,. \end{array}\right.$$

qui est une version distribuée du problème d'automatique (2.22)-(2.21) avec $A\equiv\kappa\,\partial_x^2$, $B\equiv I$, $f\equiv 0$ et $\theta = 0$. En effet, pour tout T, il est impossible que $\varphi = \varphi_{\rm opt}$, ce qui signifie que le système commandé ne satisfait la condition de *contrôlabilité exacte*. En revanche, on peut déceler une propriété de *contrôlabilité approchée* si l'on impose à u d'évoluer dans $U = L^2([0,T],L^2(\Omega))$. En effet, en ayant recours à l'état adjoint défini par (2.8), on démontre alors (cf. J.-L. Lions [64]) que $\varphi(u)$ parcourt un espace Z dense dans l'espace cible $L^2([0,T],L^2(\Omega))$. Ce résultat s'appuie sur un corollaire du théorème de Hahn-Banach (cf. [22] Chap. VI), selon lequel un sous-espace vectoriel $Z\subset Y$ est dense dans Y si et seulement si toute forme linéaire continue sur Z qui s'annule sur Y s'annule sur Z tout entier. Ainsi, on choisit pour la forme linéaire en question un élément ψ de Z^T (théroème de Riesz), et on considère la solution λ du problème

$$\left\{\begin{array}{rcll} -\partial_t\lambda(t,x) - \kappa\,\partial_x^2\lambda(t,x) &=& \psi(t,x)\,, & \forall\,(t,x)\,\in\,[0,T]\times[0,\ell]\,,\\ \lambda(t,0)=\lambda(t,\ell) &=& 0\,, & \forall\, t\in[0,T]\,,\\ \lambda(T,x) &=& 0\,, & \forall\, x\,\in\,[0,\ell]\,. \end{array}\right.$$

Alors

$$\forall\,\varphi\,\in\,Z\,,\quad 0 = \int_0^T\int_0^\ell (-\partial_t\lambda - \partial_x^2\lambda)\,\varphi = \int_0^T\int_0^\ell \lambda\,(\partial_t\varphi - \partial_x^2\varphi) = \int_0^T\int_0^\ell \lambda\,u\,,$$

donc nécessairement $\lambda = 0$ puisque u est arbitraire. Il en résulte que $\psi = 0$ sur $L^2([0,T],L^2(\Omega))$. On dit alors que U est l'ensemble des *commandes admissibles* et Y l'ensemble *atteignable*. Par densité, l'état cible $\varphi_{\rm opt}$ peut alors être approché sous la forme

$$\varphi(u)(t)\;\in\;\varphi_{\rm opt}(t) + [-\varepsilon,\varepsilon]\,,\quad \forall\,\varepsilon>0\,.$$

La question de la contrôlabilité exacte se pose dans de nombreux problèmes (cf. R. Glowinski et J.-L. Lions [43] [44], J.-M. Coron [18], A.V Fursikov [36]), comme par exemple lorsque la commande n'est définie que sur le bord d'un domaine, où lorsque l'état n'est disponible qu'à travers un observateur (*observabilité*). Pour un choix convenable de U, des arguments de densité permettent alors de se ramener éventuellement à un problème de contrôlabilité approchée.

2.3.2.b Minimisation de fonctionnelles convexes

En dimension finie, la compacité du domaine de définition et la continuité du critère assurent qu'il atteint son minimum. Dans le cas des EDP, on perd la compacité et seule la convexité permet d'obtenir des gages d'existence et d'unicité sur la solution optimale. On considère pour cela un ensemble $U_{\rm ad}$, sous-espace convexe fermé d'un espace de Hilbert U. Soient b une forme bilinéaire symétrique, continue et coercive sur U (ou α-convexe : $\exists\alpha>0\,,\ \forall v\in U\,,\ b(v,v)\geq\alpha\|v\|_U^2$), M une forme linéaire continue sur U. Alors il existe une unique solution $u\in U_{\rm ad}$ au problème :

$$\inf_{v\in U_{\rm ad}} J(v)\,,\quad J(v)=\frac{1}{2}\,b(v,v)-M(v) \tag{2.25}$$

L'unicité résulte immédiatement de la coercivité de b, qui implique la stricte convexité de J. Au sujet de l'existence, la coercivité de b implique que toute suite minimisante pour J est bornée. Soit (v_n) une telle suite. Comme U est un espace de Banach réflexif, on peut extraire (cf. **2.1.2**) de (v_n) une sous-suite (v_m) qui converge faiblement vers u. De plus $u\in U_{\rm ad}$, du fait que $U_{\rm ad}$ est faiblement fermé, car fermé et convexe. Ainsi :

$$\lim_{m\to\infty} J(v_m)=\inf_{v\in U_{\rm ad}} J(v)\,,\quad\text{avec}\quad v_m\underset{m\to\infty}{\rightharpoonup} u\in U_{\rm ad}\,.$$

Enfin, les fonctions $v\mapsto b(v,v)$ et $v\mapsto M(v)$ sont, au sens faible, respectivement semi-continue inférieurement (car elliptique) et continue ; il en résulte que J est elle-même faiblement semi-continue inférieurement, donc (voir [15]) :

$$\liminf_{m\to\infty} J(v_m)\geq J(u)\,,\quad\text{ce qui implique}\quad J(u)=\inf_{v\in U_{\rm ad}} J(v)\,.$$

2.3.2.c Condition nécessaire et suffisante d'optimalité

La fonctionnelle J admet une affine minorante $(-M)$, donc une sous-différentielle en tout point de $U_{\rm ad}$, étant donné qu'elle est convexe. Elle est de plus continue donc admet une dérivée au sens de Gâteaux J' donnée par

$$(J'(u),\tilde u)_U=\lim_{\varepsilon\to0}\ \frac{J(u+\varepsilon\tilde u)-J(u)}{\varepsilon}\,.$$

On peut ainsi écrire la condition d'optimalité

$$\forall\, v\in U_{\rm ad}\,,\quad\forall\theta\in]0,1[\,,\quad J(u)\leq J((1-\theta)u+\theta v)$$

sous la forme $J'(u)\cdot(v-u)\geq0\,,\ \forall\, v\in U_{\rm ad}$, soit

$$b(u,v-u)\geq M(v-u)\,,\quad\forall\, v\in U_{\rm ad}\,. \tag{2.26}$$

L'implication réciproque :

$$b(u,v-u)\geq M(v-u)\quad\Rightarrow\quad J(u)=\inf_{v\in U_{\rm ad}} J(v)$$

se démontre sans peine. La caractérisation (2.26) de la solution optimale u porte le nom d'*inéquation d'Euler*. Nous allons voir qu'elle s'étend au contrôle optimal des systèmes gouvernés par des équations elliptiques.

Remarque : lorsque $U_{\rm ad}=U$ ou $u\in\overset{\circ}{U}$, la condition (2.26) prend la forme :

$$b(u,v)=M(v)\,,\quad\forall\, v\in U_{\rm ad}\,.$$

C'est l'*équation d'Euler* du problème (2.25). Sa solvabilité avait déjà été mentionnée en section **2.1.2** dans la présentation de l'étude des modèles.

2.3.3 Cas elliptique

Soient $\Omega \in \mathbb{R}^d$ un ouvert borné, H et V deux espaces de Hilbert tels que :

$$V \subset H = H' \subset V' , \quad \text{avec injections denses et continues.}$$

(on identifie H à son dual). Considérons par ailleurs les formes continues a et L respectivement bilinéaire coercive et linéaire sur V. D'après le théorème de Riesz (cf. **2.1.2**),

$$\exists ! \, f \in V' \quad / \quad L(\tilde{\varphi}) = (f, \tilde{\varphi}) , \quad \forall \, \tilde{\varphi} \in V .$$

On pose en fait une hypothèse plus forte sur L, à savoir $f \in H$ (c'est pourquoi on a pris le produit scalaire L^2 dans le membre de droite). Le lemme de Lax-Milgram assure l'existence d'une unique solution $\varphi \in V$ au problème elliptique :

$$a(\varphi, \tilde{\varphi}) = (f, \tilde{\varphi}) , \quad \forall \, \tilde{\varphi} \in V .$$

La forme a est continue par rapport à chacune de ses variables, donc en utilisant une nouvelle fois le théorème de Riesz, on lui associe un opérateur $L_a \in \mathcal{L}(V, V')$, et on pose $A = -L_a$, de sorte que le problème précédent est équivalent à

$$-A\varphi = f .$$

(par exemple, dans le cas du problème de Dirichlet homogène : $H = L^2(\Omega)$, $V = H_0^1(\Omega)$ et $A = \Delta$). Enfin, on définit un opérateur coercif $Q \in \mathcal{L}(U)$ ($\forall \, v \in U$, $(Qv, v)_U \geq \beta \|v\|_U^2$), ainsi qu'un état-cible $\varphi_{\text{opt}} \in H$ pour définir le problème de *contrôle optimal* :

$$\inf_{v \in U_{\text{ad}}} \quad \mathcal{J}(v) , \quad \text{où} \quad \mathcal{J}(v) = \frac{1}{2} |\varphi(v) - \varphi_{\text{opt}}|_H^2 + \frac{1}{2} (Qv, v)_U , \tag{2.27}$$

$$\text{et} \quad \varphi(v) \quad \text{est tel que} \quad -A\varphi = f + v . \tag{2.28}$$

2.3.3.a Première caractérisation de l'optimum

Grâce à l'existence d'une solution à l'équation d'état (2.28) posée avec $v = 0$, nous allons montrer qu'un tel problème peut se ramener à la minimisation d'une fonctionnelle convexe comme exposé précédemment. On a :

$$\mathcal{J}(v) = \frac{1}{2} \, |\varphi(v) - \varphi(0) + \varphi(0) - \varphi_{\text{opt}}|_H^2 + \frac{1}{2} \, (Qv, v)_U ,$$

et si l'on pose à présent

$$\begin{aligned} b(u, v) &= \Big(\varphi(u) - \varphi(0) \, , \, \varphi(v) - \varphi(0)\Big)_H + (Qu, v)_U , \\ M(v) &= \Big(\varphi_{\text{opt}} - \varphi(0) \, , \, \varphi(v) - \varphi(0)\Big)_H , \end{aligned} \tag{2.29}$$

du fait de l'ellipticité de Q, qui implique celle de b, on peut traiter le problème en se servant des résultats des sections **2.3.2.b** et **2.3.2.c**. En effet, le critère s'écrit :

$$\mathcal{J}(v) = \frac{1}{2} \, b(v, v) - M(v) + \frac{1}{2} \, |\varphi_{\text{opt}} - \varphi(0)|_H^2 ,$$

dont la minimisation est équivalente à celle de $J(v)$.

2.3.3.b Caractérisation par l'état adjoint

Lorsque $U_{\rm ad} \neq U$, nous n'avons qu'à disposition l'inégalité (2.26). C'est une condition nécessaire et suffisante de solvabilité du problème de contrôle, qui mérite d'être développée dans le cadre des hypothèses introduites ci-dessus. On réécrit la différence entre $b(u, v-u)$ et $L(v-u)$ en faisant à nouveau appel au théorème de Riesz, qui permet de définir l'*isomorphisme canonique* Λ de U sur U'. Celui-là est nécessaire pour exprimer le produit scalaire en fonction du produit de dualité (et vice-versa), du fait qu'on n'a pas identifié U à son dual. Ainsi, on a la propriété que pour tout $u \in U$, $(u,v)_U = \langle \Lambda\, u, v\rangle_{U',U}$, $\forall\, v \in U$. On commence par substituer v par $v-u$ dans (2.29) :

$$b(u, v-u) = \langle (\varphi(u) - (\varphi(0))\,, \varphi(v-u) - \varphi(0)\rangle_{U',U} + (Qu, v-u)_U\,,$$

$$M(v-u) = \langle (\varphi_{\rm opt} - \varphi(0)), \varphi(v-u) - \varphi(0)\rangle_{V',V}\,.$$

Il vient alors :

$$b(u, v-u) - M(v-u) = \langle (\varphi(u) - \varphi_{\rm opt}), \varphi(v) - \varphi(u)\rangle_{V',V} + (Qu, v-u)_U\,.$$

Nous introduisons à présent l'opérateur A^* *adjoint* de A (extension de la transposée d'une matrice à la dimension infinie), et appellons état adjoint la solution $\lambda(v)$ du *problème adjoint* :

$$-\,A^*\lambda = \varphi(v) - \varphi_{\rm opt}\,. \tag{2.30}$$

À l'aide de cette nouvelle inconnue, il ressort une condition nécessaire sur la dérivée du critère ne faisant plus intervenir $\varphi(v)$:

$$b(u, v-u) \geq M(v-u) \quad \Leftrightarrow \quad (\Lambda^{-1}\lambda(u) + Qu, v-u)_U \geq 0\,, \quad \forall\, v \in U_{\rm ad}\,, \tag{2.31}$$

qu'on appelera désormais *condition d'optimalité*. C'est ainsi qu'on arrive à établir, comme dans l'exemple en dimension finie (2.24), un système de trois conditions d'optimalité en regroupant (2.28), (2.30) et (2.31). Dans le cadre strictement elliptique que nous venons de présenter, la commande optimale est déterminée par la solution de ce système.

Enfin, on notera que le problème peut se voir sous l'angle de la recherche d'un point-selle en *optimisation sous contraintes* : si on définit la fonction de Lagrange par

$$L(u,\lambda) = \mathcal{J}(u) - \langle \lambda\,,\, F(u)\rangle_{V,V'}\,, \quad \text{où} \quad F(u) = -A\varphi(u) - f - u\,,$$

$\mathcal{F} : U \to V'$ (sous la condition de *qualification* que $\mathcal{F}'(u)^* : V \to U'$ soit un opérateur linéaire borné surjectif, cf. M. Burger [17]) alors (2.27)-(2.28) est équivalent au problème

$$\inf_{u\in U}\ \sup_{\lambda\in V}\ L(u,\lambda) \quad \text{qui implique} \quad \partial_v L(u,\lambda) = \partial_\mu L(u,\lambda) = 0\,, \tag{2.32}$$

ou encore

$$\partial_\varphi \mathcal{L}(\varphi, u, \lambda) = \partial_u \mathcal{L}(\varphi, u, \lambda) = \partial_\lambda \mathcal{L}(\varphi, u, \lambda) = 0\,,$$

$$\text{avec} \quad \mathcal{L}(\varphi, u, \lambda) = \mathcal{J}(\varphi, u) - \langle \lambda\,,\, \mathcal{F}(\varphi, u)\rangle\,, \quad \mathcal{F}(\varphi, u) = -A\varphi - f - u\,.$$

La propriété (2.32) est l'extension en dimension infinie du *théorème de Kuhn-Tucker* dans le cas des contraintes d'égalité[4] (cf. [56]). Cela donne un moyen pratique d'établir la forme du problème adjoint ($\partial_\varphi \mathcal{L} = 0$), auquel nous aurons constamment recours dans les applications (cf. **5** et **8**).

[4] $F = (f_i)_{1\leq i\leq N}$, avec $f_i : U \to \mathbb{R}$: la condition de qualification est alors l'indépendance linéaire des f_i'.

2.3.4 Problèmes d'évolution

De même que dans le cas elliptique, nous exposons à présent brièvement une extension du principe de Pontryagin aux systèmes gouvernés par des EDP paraboliques ou hyperboliques, donc avec commande, état et multiplicateurs de Lagrange de dimension infinie. Dans les sections précédentes, nous avons mis en évidence la structure de *semi-groupe continu de contractions*[5] de la famille d'opérateurs $(G(t))_{t\in\mathbb{R}^+}$ qui à une condition initiale associent la variable d'état à l'instant t, sur le plan continu (**2.1.3**) comme sur les plans discret (**2.2.1.a**) et semi-discrétisé (**2.2.3**). En résumé, considérant un espace de Hilbert W, nous avons pu expliciter $G(t)$ par son *générateur infinitésimal* $A \in \mathcal{L}(W)$ sous la forme

$$\Phi(t) = G(t)\Phi^0 + (G \underset{t}{*} F)(t)\,, \quad \forall\, t \geq 0\,, \quad \text{avec} \quad G(t) = e^{At}\,.$$

2.3.4.a Existence de contrôles optimaux

Dans le cas de l'équation de la chaleur ($W = H = L^2(\Omega)$, $A = \Delta$), ce semi-groupe est par ailleurs *compact*, du fait de la compacité de sa *résolvante* $(\mu I - A)^{-1}$ (cf. **2.1.3**), elle-même permise par le choix de l'espace $V = H_0^1(\Omega)$ et l'ellipticité de $-\Delta$. Ainsi, et sous d'autres hypothèses techniques que nous ne détaillerons pas ici, un problème de contrôle optimal de la forme (où $M(\cdot,\varphi(\cdot),u(\cdot)) \in L^1(]0,T[)$, $g \in \mathcal{C}([0,T], H\times U)$ et $\mathcal{S}$ est un sous-espace fermé convexe de $H\times H$) :

$$\inf_{u\in U} \mathcal{J}(u)\,, \quad \text{où} \quad \mathcal{J}(u) = \int_0^T M(t,\varphi(u(t)),u(t))\,dt\,, \tag{2.33}$$

$$\text{et} \quad \varphi(u) \quad \text{est tel que} \quad \partial_t\varphi(t) = A\varphi(t) + g(t,\varphi(t),u(t))\,, \quad (\varphi(0),\varphi(T)) \in \mathcal{S} \tag{2.34}$$

admet au moins une solution (cf. X. Li et J. Yong [62]) si G est compact. L'unicité, en revanche, n'est pas établie. On notera que ce modèle permet de traiter des problèmes non linéaires dominés par un opérateur ellitptique, comme par exemple les équations de Navier-Stokes (cf. **3.1**), et que la contribution de la commande à l'équation d'état n'est plus régie par une relation linéaire.

L'équation des ondes, en revanche, n'est pas régie par un semi-groupe *compact*[6], pour la simple raison que c'est en fait un groupe (cf. [22] Chap. XIV). On peut cependant démontrer l'existence de commandes optimales en ayant recours à la formulation variationnelle de l'équation et aux estimations *a priori* appliquées au système commandé.

2.3.4.b Principe de Pontryagin

Si l'on suppose, dans le problème (2.33)-(2.34), que A est le générateur d'un semi-groupe continu, que M et g sont continûment différentiable au sens de Fréchet par rapport à la variable d'état, et, de plus, qu'il existe une constante $C > 0$ telle que pour tout $(t,\varphi,u) \in [0,T]\times H\times U$:

$$|\nabla_\varphi M(t,\varphi,u)|\,,\ |\nabla_\varphi g(t,\varphi,u)|\,,\ |M(t,0,u)|\,,\ |g(t,0,u)| \leq C\,,$$

alors, le principe de Pontryagin s'applique au problème (2.33)-(2.34), sous réserve que l'ensemble :

$$\mathcal{Q} = \left\{ \varphi_1 - \varphi(T) \mid \varphi(t) = \mathrm{e}^{At}\varphi_0 + \int_0^t \mathrm{e}^{A(t-s)}\nabla_\varphi g(s,\bar\varphi(s),\bar u(s))\,ds\,,\ t\in[0,T]\,,\ (\varphi_0,\varphi_1)\in S \right\}$$

soit de codimension finie dans H (cf. X. Li et J. Yong [62]). On notera que la propriété de compacité du semi-groupe n'est pas requise ; ainsi, le principe est également valable pour les problèmes hyperboliques en remplaçant φ par Φ dans (2.33)-(2.34).

[5] Pour son extension aux problèmes non linéaires, on pourra consulter par exemple R. Temam [95] où le raisonnement est appliqué aux équations de Navier-Stokes.

[6] mais possède des propriétés de contrôlabilité exacte (cf. [65]), contrairement à l'équation de la chaleur...

Ainsi, les conditions nécessaires d'optimalité prennent la forme, si $u(t)$ est la commande optimale :

$$\lambda(t) = \mathrm{e}^{A^*(t-T)}\lambda(T) + \int_t^T \mathrm{e}^{A^*(s-t)}\Big(\nabla_\varphi g(s,\varphi(s),u(s))\,,\ \lambda(s)\Big)_H\,ds - \int_t^T \mathrm{e}^{A^*(s-t)}\,M(s,\varphi(s),u(s))\,ds$$

avec φ solution de (2.34), et

$$H_P(\varphi,u,\lambda) = \inf_{v\in U} H_P(\varphi,v,\lambda)\,,\quad \text{avec}\quad H_P(\varphi,u,\lambda) = M(\varphi,u) + \Big(\lambda\,,\ A\varphi + g(\varphi,u)\Big)_H\,,$$

qu'on exprime encore par la stationnarité de la fonction de Lagrange par rapport φ, u et λ :

$$\partial_\varphi\mathcal{L} = \partial_u\mathcal{L} = \partial_\lambda\mathcal{L}\,,\quad \text{avec}\ \ \mathcal{L}(\varphi,u,\lambda) = \mathcal{J}(\varphi,u) - \int_0^T \Big(\lambda\,,\ \mathcal{F}(\varphi,u)\Big)_H\,,\ \ \mathcal{F}(\varphi,u) = \partial_t\varphi - A\varphi - g(\varphi,u)\,. \tag{2.35}$$

2.3.5 Mise en œuvre

Il existe essentiellement deux grandes familles de méthodes pour déterminer la solution d'un problème de contrôle optimal : les approches déterministes et les approches stochastiques (algorithmes génétiques etc.). Ces dernières sont bien adaptées à la recherche d'un minimum global, mais souvent trop coûteuses dans le domaine des systèmes gouvernés par des EDP, qui requiert une discrétisation de l'équation d'état conduisant à des problèmes de grande taille. Dans la première catégorie, on distingue principalement trois types de méthodes, à savoir :

- le calcul direct de la commande optimale en fonction de l'état par une loi de commande (cf. **2.3.1**) qui requiert de résoudre l'équation de Ricatti,
- la recherche d'un minimum local en résolvant le système de Pontryagin (2.35),
- la méthode de la programmation dynamique, basée sur la recherche de *solutions de viscosité* (cf. M.G. Crandall et P.-L. Lions [20]) à l'équation de *Hamilton-Jacobi-Bellman*.

Dans le premier cas, la version en dimension infinie de l'équation (intégro-différentielle) de Ricatti (cf. J.-L. Lions [64]) se prête difficilement au calcul numérique, tandis que dans le troisième, c'est encore la taille des problèmes qui compromet son utilisation. On rencontre plutôt l'approche de la programmation dynamique en contrôle stochastique, ou encore dans les *problèmes en temps discret*, comme en logistique par exemple.

C'est ainsi qu'on adopte la méthode de Pontryagin pour déterminer un minimum local (qu'on espère global...) au moyen du système des conditions nécessaires d'optimalité. Celui-là peut paraître ardu car triplement couplé, mais il est possible de contourner la difficulté en utilisant une méthode **itérative** pour minimiser $\mathcal{J}$. Une première idée peut consister à calculer son gradient en l'approchant par différences finies connaissant sa valeur pour deux commandes données, mais on s'aperçoit que cela requiert d'utiliser un pas d'échantillonage trop fin et conduit à un coût prohibitif en temps de calcul. En revanche, la résolution du problème adjoint pour une commande (et donc une variable d'état) donnée s'avère très intéressante, car elle permet de calculer le gradient du critère de manière beaucoup plus efficace que par différences finies. On dispose alors de l'ensemble des méthodes de minimisation (méthode de gradient, gradient conjugué, Newton, quasi-Newton) pour un exposé approfondi desquelles on renvoie à D.G. Luenberger [68] ou F. Bonnans *et al.* [14].

À ce stade, on se retrouve encore face à deux possibilités : calculer le gradient du critère discrétisé, ou discrétiser une formule continue du gradient, dont nous allons à présent expliquer l'obtention dans un cadre formel.

2.3.5.a Calcul du gradient de la fonction coût par la méthode de l'adjoint

Exprimons la *dérivée directionnelle* du critère $\mathcal{J}$ donné par (2.27) :

$$\Big(\nabla\mathcal{J}(u),\tilde{u}\Big)_U = \lim_{\varepsilon\to 0}\frac{\mathcal{J}(\varphi+\varepsilon\tilde{\varphi},u+\varepsilon\tilde{u})-\mathcal{J}(\varphi,u)}{\varepsilon}\,, \quad \text{soit}$$

$$\Big(\nabla\mathcal{J}(u),\tilde{u}\Big)_U = (\varphi(u)-\varphi_{\rm opt}\,,\,\tilde{\varphi})_H \,+\, (Qu\,,\tilde{u})_U\,, \tag{2.36}$$

où $\tilde{\varphi}$ est solution du système des équations de *sensitivité* :

$$\nabla_u\mathcal{F}(\varphi,u)=0 \quad\Leftrightarrow\quad \lim_{\varepsilon\to 0}\frac{\mathcal{F}(\varphi+\varepsilon\tilde{\varphi},u+\varepsilon\tilde{u})-\mathcal{F}(\varphi,u)}{\varepsilon}=0\,, \tag{2.37}$$

le système étant gouverné par l'équation d'état $\mathcal{F}(\varphi,u)=0$. Cette approche ne fournit pas d'expression explicite du gradient, du fait de la présence de $\tilde{\varphi}$: pour obtenir $\nabla\mathcal{J}(u)$, il faudrait résoudre (2.37) dans toutes les directions $\tilde{u}$ ... mais on démontre la propriété

$$\nabla\mathcal{J}(u)=Q\,u\;-\;\Big\langle\lambda,\partial_u\mathcal{F}(y,u)\Big\rangle \quad\text{lorsque}\quad \partial_\varphi\mathcal{L}(\varphi,u,\lambda)=0\,. \tag{2.38}$$

En effet, compte-tenu de (2.37) :

$$\Big(\nabla\mathcal{J}(u),\tilde{u}\Big)_U=\lim_{\varepsilon\to 0}\frac{\mathcal{L}(\varphi+\varepsilon\tilde{\varphi},u+\varepsilon\tilde{u},\lambda)-\mathcal{L}(\varphi,u,\lambda)}{\varepsilon}\,,$$

donc, en développant au premier ordre :

$$\Big\langle\nabla\mathcal{J}(u),\tilde{u}\Big\rangle=\Big\langle\partial_u\mathcal{L}(\varphi,u,\lambda),\tilde{u}\Big\rangle+\Big\langle\partial_\varphi\mathcal{L}(\varphi,u,\lambda),\tilde{\varphi}\Big\rangle\,,$$

et comme $\partial_\varphi\mathcal{L}(\varphi,u,\lambda)=0$:

$$\nabla\mathcal{J}(u)=\partial_u\mathcal{L}(\varphi,u,\Lambda)=Q\,u\;-\;\Big\langle\lambda,\partial_u\mathcal{F}(\varphi,u)\Big\rangle\,.$$

Nous disposons ainsi d'une expression de $\nabla\mathcal{J}(u)$ exploitable : une fois λ obtenu par la résolution de l'équation d'état et du problème adjoint (qui est en général un peu plus simple), le calcul de $\Big\langle\lambda,\partial_u\mathcal{F}(\varphi,u)\Big\rangle$ se fait sans difficultés.

2.3.5.b Discrétisation et optimisation : dans quel ordre ?

Pour calculer le gradient du critère en vue de lancer une procédure d'optimisation, un première possibilité est d'appliquer cette méthode telle quelle en construisant le problème adjoint au niveau continu, puis en le discrétisant de la même manière que l'équation d'état, et enfin en calculant le critère à partir des données discrètes et de la formule continue (2.38). Il reste alors à lancer un programme de minimisation en dimension finie à partir des données approchées de u et $\mathcal{J}(u)$.

L'autre option est d'établir des conditions d'optimalité en dimension finie à partir de l'équation d'état et du critère discrétisés. Il existe pour cela deux méthodes, à savoir par le même type de raisonnement qu'en **2.3.1**, ou en générant directement le problème à partir du code source du critère et de l'équation d'état par *différentiation automatique* (cf. I. Danaila *et al.* [21]).

L'avantage de la deuxième approche est qu'elle traite un problème dans lequel les versions discrétisées du gradient et du critère se correspondent exactement, contrairement au cas précédent où elles étaient construites à partir des conditions d'optimalité continues. L'inconvénient est une implémentation qui n'est pas toujours pratique (cf. M.D. Gunzburger [45]). Un bon compromis semble être d'opter pour la première possibilité, ce que nous ferons dans les travaux présentés ici. Des exemples de cette démarche peuvent être trouvés pour le contrôle frontière d'un problème de MHD stationnaire non linéaire (L.S. Hou et A.J. Meir [52]), ou encore dans un domaine aussi éloigné que la calibration de la volatilité locale des options européennes et américaines en finance (Y. Achdou et O. Pironneau [1]).

Chapitre 3

Présentation des modèles utilisés

Les contraintes pratiques et la complexité du procédé Hall-Héroult ont toujours été un frein à son optimisation. C'est la raison pour laquelle d'actives recherches sur la modélisation des cuves d'électrolyse, et en particulier sur le comportement de l'interface électrolyte/aluminium, ont été menées depuis une quarantaine d'années maintenant, avec les premiers travaux de J.-P. Givry [42]. Dix ans plus tard, T. Sele [91] proposait un modèle d'instabilités qui eut un certain succès dans la communauté de l'aluminium et, sur le terrain, fut validé notamment par N. Urata [98]. Cette approche *linéaire*, utile pour appréhender des mécanismes qualitatifs menant aux instabilités, se révèle insuffisante pour cerner des aspects plus quantitatifs. C'est pourquoi une méthode *non linéaire*, inspirée des travaux de M. Sermange et R. Temam [92] sur les modèles magnétohydrodynamiques, a été développée par J.-F. Gerbeau *et al.* [37]. L'objectif ici est de résumer les principaux résultats obtenus dans chacune des deux démarches, en commençant par la deuxième, la première pouvant être regardée comme une de ses simplifications possibles.

3.1 Modélisation non linéaire

Comme nous l'avons déjà signalé au premier chapitre, les phénomènes qui se produisent dans les cuves de réduction de l'alumine sont de nature fondamentalement multiphysique, en effet ils mettent en jeu des aspects chimiques, thermiques, électromagnétiques et hydrodynamiques. L'objectif principal de la modélisation de ce genre de processus étant la simulation numérique, il est inévitable, encore aujourd'hui, de devoir simplifier les modèles trop compliqués car la puissance de calcul n'est pas inépuisable. De nombreux auteurs (par exemple [23], [25], [74], [90], [94]) s'accordent à attribuer l'origine des instabilités aux phénomènes hydrodynamiques et magnétiques, entre autres parce que le temps caractéristique des effets thermiques est bien supérieur à celui des effets magnétohydrodynamiques, aussi semble-t-il raisonnable d'adopter le même point de vue.

Ainsi, on se propose de décrire, dans chacun des deux fluides, les phénomènes hydrodynamiques par les équations de Navier-Stokes incompressibles, et les phénomènes électromagnétiques par les équations de Maxwell sans courants de déplacement. Ces deux modèles bien connus sont couplés à deux niveaux, par la loi d'Ohm et la force de Lorentz. Enfin, la prise en compte de deux fluides s'appuie une équation globale de conservation de la masse, mais avec une densité non homogène dépendant du côté de l'interface duquel on se situe.

3.1.1 Équations de Navier-Stokes

Traditionnellement, on distingue les fluides *newtoniens*, dans lesquels il existe une relation linéaire[1] entre le *tenseur des contraintes* le *tenseur des déformations* (cf. *infra*) et les fluides *non-newtoniens* ou *complexes* pour lesquels cette relation n'est pas valable. Dans les cuves d'électrolyse, on suppose que les fluides sont newtoniens, mais avant de préciser cette propriété, il nous faut aborder la notion de *cinématique* du fluide, qui consiste à étudier ses déformations indépendamment des forces qui lui sont appliquées. Pour cela, on appelle *particule* un élément de fluide de taille très petite devant les échelles de longueur caractéristiques (dimensions d'une cuve...), mais suffisamment grande pour modéliser le phénomène d'un point de vue macroscopique. Alors, on distingue deux manières d'exprimer la vitesse d'un fluide en fonction de l'espace et du temps :

- dans la description *eulérienne*, on s'intéresse à la vitesse $v(x,t)$ d'une particule de fluide qui coïndice à l'instant t avec le point fixe de position x ; à chaque instant, on regarde donc les vitesses de particules différentes ;
- dans la description *lagrangienne*, on suit le mouvement d'une particule fixée, en spécifiant sa position x_0 à l'instant 0 ; on note ainsi $V(x_0,t)$ la vitesse du fluide.

La première est la plus couramment utilisée car la plus commode pour calculer la variation spatiale d'une propriété au temps t. Ainsi, dans les simulations, la description eulérienne est plus robuste car elle ne nécessite pas de déformation du maillage pour suivre une particule donnée. Nous serons cependant amenés à nuancer quelque peu cette position lorsqu'il s'agira de prendre en compte les mouvements de *deux* fluides (cf. *infra*). Pour le moment, on cherche à appliquer les lois de Newton en coordonnées eulériennes, et comme la dérivée de la vitesse d'une particule $v(x,t)$ est due à la fois à la variation explicite du champ de vitesse en fonction du temps et à l'exploration du champ de vitesse par la particule (*convection*), l'accélération s'obtient par la règle de dérivation en chaîne :

$$\frac{dv}{dt} = \frac{\partial v}{\partial t} + \sum_{i=1}^{d} \frac{\partial v}{\partial x_i}\frac{dx_i}{dt} = \frac{\partial v}{\partial t} + (v\cdot\nabla)\,v \qquad (d : \text{dimension d'espace})\,.$$

Nous verrons que d'autres grandeurs exprimées en coordonnées eulériennes (x,t) se dérivent suivant la même règle : $d/dt = \partial/\partial t + v\cdot\nabla$, que nous aurons l'occasion d'appliquer à la densité par exemple. Venons-en à présent à la définition du tenseur des déformations ; celui-là exprime la variation de la vitesse en fonction du changement de position de la particule à un instant fixé :

$$\nabla v = \left(\frac{\partial v_i}{\partial x_j}\right)_{1\le i,j\le d}.$$

Le fluide est newtonien lorsque le tenseur des contraintes, qui regroupe les forces surfaciques que subit une particule de fluide occupant un domaine ω, à savoir :

- les forces en pression $-\int_{\partial\omega} p\,n$
- les contraintes visqueuses dues à la déformation du fluide $\int_{\partial\omega}\sigma'\,n$ (σ' est un tenseur)

prend la forme

$$\sigma = \sigma' - p\,I_d\,,\quad \text{avec } \sigma' = \eta(\nabla v + \nabla v^T) + \left[\left(\xi - \frac{2\eta}{3}\right)\operatorname{div} v\right] I_d$$

où η représente la *viscosité* du fluide et ξ une autre grandeur caractéristique qu'on appelle souvent *coefficient d'atténuation du son*. On est bien en présence d'une relation linéaire entre le tenseur des contraintes et le tenseur des déformations. Cette *loi de comportement* est encore qualifiée de *relation de fermeture* car elle permet d'aboutir à un problème bien posé.

[1] Cela constitue en quelque sorte une extension aux fluides de la loi de Hooke (1678), qui prévoit la proportionnalité des déformations aux efforts en élasticité linéaire.

Par ailleurs, si l'on tient compte d'éventuelles forces volumiques (gravité, force de Lorentz, etc.) notées f, la loi de Newton appliquée à un élément de fluide de densité ρ compris dans un volume élémentaire ω s'écrit :

$$\frac{d}{dt}\int_\omega \rho\, v\, dx = \int_{\partial\omega} \left[\eta\, D(v) + \left[\left(\xi - \frac{2\eta}{3} \right) \operatorname{div} v - p \right] I_d \right] n_{|\partial\omega} \cdot d(\partial\omega) + \int_\omega f\, dx\,, \tag{3.1}$$

où $D(v) = \nabla v + \nabla v^T$. De plus, on doit tenir compte de la *conservation de la masse*, qui exprime que la variation de la masse de fluide contenue dans un certain volume $\mathcal{V}$ ne peut être due qu'aux flux entrant dans (ou sortant de) ce volume :

$$\frac{d}{dt}\int_{\mathcal{V}} \rho\, dx = -\int_{\partial\mathcal{V}} \rho\, v \cdot d(\partial\mathcal{V})\,.$$

Il résulte de la formule d'Ostrogradski, et de la validité de ce bilan quel que soit $\mathcal{V}$:

$$\frac{\partial\rho}{\partial t} + \operatorname{div}(\rho\, v) = 0\,, \quad \text{ou encore} \quad \frac{d\rho}{dt} + \rho \operatorname{div} v = 0\,.$$

Ainsi, l'*incompressibilité* d'un fluide ($d\rho/dt = 0$) se traduit par la relation $\operatorname{div} v = 0$, qui a pour conséquence de simplifier l'équation (3.1). Alors, à nouveau par la formule d'Ostrogradski, l'écoulement d'un fluide newtonien incompressible répond aux *équations de Navier-Stokes* :

$$\left\{ \begin{array}{rcl} \rho \dfrac{\partial v}{\partial t} + \rho\, v \cdot \nabla v - \eta\, \Delta v + \nabla p &=& f\,, \\ \operatorname{div} v &=& 0\,. \end{array} \right. \tag{3.2}$$

auxquelles on ajoute une condition aux limites exprimant, dans le cas d'un écoulement confiné dans une cuve, que le fluide ne peut sortir de cette cuve ni glisser sur ses parois :

$$v = 0 \quad \text{sur le bord du domaine.}$$

Le lecteur intéressé pourra approfondir le raisonnement ci-dessus en consultant les ouvrages spécialisés, comme par exemple [47] ou encore [88] pour une approche plus mathématique.

Après l'établissement du modèle, il convient de vérifier s'il détermine une solution et une seule ; on entre là dans le domaine de l'analyse mathématique des EDP, pour laquelle il est en général utile d'avoir recours à une formulation variationnelle du problème. C'est en particulier le cas des équations de Navier-Stokes, comme en témoigne l'intérêt que leur porte le *Clay Mathematics Institute* en accordant un million de dollars à quiconque prouvera, par exemple, l'existence de solutions en temps long dans $\mathbb{R}^3$ (cf. [32]). Quand bien même, on dispose à l'heure actuelle d'informations suffisantes pour entreprendre des expérimentations numériques. En effet, au sujet de l'existence de solutions faibles, de nombreux progrès ont été réalisés, depuis les premiers travaux de Leray [61]. Nous retranscrirons pour notre part (de manière très synthétique encore une fois) le raisonnement de J.-L. Lions [63] basé sur des estimations *a priori* (cf. **2.1**). Pour cela, il est nécessaire d'introduire l'espace des vitesses définies pour tout $t \in [0,T]$ sur un ouvert borné $\Omega \subset \mathbb{R}^d$:

$$V = \{\tilde{v} \in H_0^1(\Omega)^d \,|\, \operatorname{div}\tilde{v} = 0\} \quad \text{et sa norme} \quad |\tilde{v}|_V = \left(\int_\Omega \|\nabla\tilde{v}\|^2 \right)^{1/2}$$

(on omettra souvent la mesure dx par concision), qui permet d'éliminer la pression des inconnues, celle-là jouant en fait le rôle de multiplicateur de Lagrange pour la contrainte $\operatorname{div} v = 0$.

Ainsi, par application du *principe des puissances virtuelles* (variante du principe de moindre action mentionné page 9) dans l'espace V des *vitesses cinématiquement admissibles* (cf. [10]), ou plus simplement en multipliant la première équation dans (3.2) par une *fonction test* $\tilde{v} \in V$ et en intégrant par parties, on obtient la formulation variationnelle pour tout $t \in [0,T]$:

$$\rho \int_\Omega \partial_t v \cdot \tilde{v} + \rho \int_\Omega v \cdot \nabla v \cdot \tilde{v} + \eta \int_\Omega \nabla v : \nabla \tilde{v} = \int_\Omega f \cdot \tilde{v}\,, \quad \forall\, \tilde{v} \in V\,. \tag{3.3}$$

Enfin, on ajoute au modèle la condition initiale :

$$v(0,x) = v^0(x) \in V\,. \tag{3.4}$$

Alors, il reste à appliquer le même type de raisonnement qu'en section **2.1.3** avec $H = H_n^{\mathrm{div}}(\Omega)$. Pour cela, on remplace $\tilde{v}$ par v dans (3.3) puis on intègre en temps, ce qui conduit à la majoration (par intégrations par parties en espace et par une inégalité d'Young au second membre) :

$$\rho\,|v(t)|_H^2 + \eta \int_0^t |\nabla v|_H^2 \;\leq\; \rho\,|v(0)|_H^2 + c\,|f|^2_{L^2(0,T;H)}\,, \quad c \in \mathbb{R}^+\,,$$

en supposant pour simplifier que $f(t) \in H$. Ensuite, on remplace cette inéquation formelle, puisqu'on n'est pas encore assuré de l'existence de v, par son équivalent discrétisé sur la base des vecteurs propres $(w_n)_{n\in\mathbb{N}}$ - de valeurs propres associées $(\lambda_n)_{n\in\mathbb{N}}$ - du problème spectral

$$\int_\Omega \nabla w_n : \nabla \tilde{v} = \lambda_n \int_\Omega w_n \cdot \tilde{v}\,, \quad \forall\, \tilde{v} \in V\,,$$

qui constituent une base orthonormale de H du fait de l'ellipticité dans V de l'opérateur $(v,\tilde{v}) \mapsto \int_\Omega \nabla v : \nabla \tilde{v}$, et de la compacité de l'injection de V dans H (cf. p.13). Alors, pour $N_h \in \mathbb{N}$, en posant $v_h(t,x) = \sum_{n=1}^{N_h} (v(t,x), w_n(x))_H\, w_n(x)$, on obtient - rigoureusement cette fois :

$$|v_h(t)|_H^2 + \frac{\eta}{\rho} \int_0^t |\nabla v_h|_H^2 \;\leq\; K\,, \quad K \in \mathbb{R}^+$$

ce qui permet d'affirmer que la suite $(v_h)_{N_h\in\mathbb{N}}$ est bornée dans $L^2(0,T;V) \cap L^\infty([0,T],H)$, et donc d'en extraire une sous-suite convergeant faiblement (resp. *faiblement-*$*$) dans $L^2(0,T;V)$ (resp. $L^\infty([0,T],H)$) vers une certaine fonction v. Par ailleurs, on a besoin de démontrer que la suite $(\partial_t v_h)_{N_h}$ également demeure dans un borné (de $L^2(0,T;V')$) pour pouvoir en déduire l'existence d'une limite $\partial_t v$. C'est une difficulté nouvelle - liée au terme non linéaire $v\cdot\nabla v$ - qui ne se présentait pas dans le cas de l'équation des ondes par exemple. En ayant recours aux propriétés de la base (w_n), et, en dimension trois, à une hypothèse plus forte sur la régularité de la vitesse (qu'on considère dans $H^2(\Omega)$), on obtient une estimation *a priori* sur le terme de convection, qui permet d'en déduire le résultat escompté. Enfin, la non-linéarité génère une difficulté dans la procédure de passage à la limite, car celle-là ne peut se faire que si la suite v_h converge *fortement* (au sens de la norme) vers v. Il faut pour cela disposer d'une propriété d'injection compacte (du type Rellich), qui a bien lieu dans notre cas. Là encore, nous n'entrerons pas plus en détail dans l'analyse du problème de l'existence de solutions au problème de Navier-Stokes, et n'aborderons pas du tout celui de l'unicité, qui n'est d'ailleurs pas encore prouvé en dimension trois avec des données quelconques (cf. [32]). Nous renvoyons une nouvelle fois à J.-L. Lions [63] pour la démonstration précise, ou encore à O. Pironneau [81] pour une version plus synthétique et J.-F. Gerbeau *et al.* [39] pour un exposé plus pédagogique.

3.1.2 Équations de la MHD

Comme nous l'avons déjà mentionné à plusieurs reprises, le modèle adopté pour la modélisation des cuves repose sur un couplage entre les équations de Navier-Stokes et celles de Maxwell. Dans les milieux continus, ces dernières prennent la forme de quatre équations comprenant quatre inconnues vectorielles, à savoir le *champ électrique* E, l'*induction électrique* D, le *champ magnétique* H et l'*induction magnétique* B :

$$\left\{\begin{array}{rcll} -\partial_t D + \operatorname{rot} H &=& J & \text{(Maxwell-Ampère)}\,, \\ \partial_t B + \operatorname{rot} E &=& 0 & \text{(Maxwell-Faraday)}\,, \\ \operatorname{div} D &=& P & \text{(Maxwell-Coulomb)}\,, \\ \operatorname{div} B &=& 0 & \text{(Maxwell-Gauss)}\,, \end{array}\right. \tag{3.5}$$

où J et P sont les densités de courant et de charge électriques. Nous nous restreignons par ailleurs à l'hypothèse des milieux *parfaits*, qui sont *linéaires*, *isotropes* et *homogènes*, de sorte qu'il existe des relations entre D et E d'une part, et entre H et B d'autre part, qui s'écrivent :

$$D(t,x) = \epsilon\, E(t,x) \quad \text{et} \quad H(t,x) = \frac{1}{\mu}\, B(t,x)\,,$$

où ϵ et μ sont respectivement la *permittivité électrique* et la *perméabilité magnétique* du milieu considéré. Ces lois se justifient lorsqu'on considère des champs suffisamment faibles (*dixit* [58]), ce qui est le cas ici d'après R. Moreau [73]. Le même auteur fait de plus l'hypothèse que les *courants de déplacement* $\partial_t D$ sont négligeables. Enfin, la loi d'Ohm pour les milieux en mouvement à la vitesse v permet de substituer E par v et B (on note σ la conductivité du milieu) :

$$\left.\begin{array}{rcll} J &=& \sigma\,(E + v \times B) & \text{(Ohm)} \\ \operatorname{rot} B &=& \mu\, J & \text{(Maxwell-Ampère)} \end{array}\right\} \quad \Rightarrow \quad E = \frac{\operatorname{rot} B}{\mu\sigma} - v \times B\,,$$

d'où une simplification considérable des équations de Maxwell, qui se réduisent à :

$$\left\{\begin{array}{rcl} \partial_t B + \dfrac{1}{\mu\sigma}\operatorname{rot}\operatorname{rot} B - \operatorname{rot}(v \times B) &=& 0\,, \\ \operatorname{div} B &=& 0\,. \end{array}\right.$$

On voit ainsi survenir un premier couplage entre les variables hydrodynamiques (la vitesse en l'occurence) et (électro-)magnétiques. Un deuxième, dû à la force de Lorentz (1.3), apparaît dans les équations de Navier-Stokes par l'intermédiaire du champ magnétique, si bien que le problème magnétohydrodynamique soumis à la seule pesanteur (g désigne la gravité) répond au modèle :

$$\left\{\begin{array}{rcl} \rho\,\partial_t v + \rho\, v\cdot\nabla v - \eta\,\Delta v + \nabla p - \dfrac{1}{\mu}\operatorname{rot} B \times B &=& \rho\, g\,, \\ \operatorname{div} v &=& 0\,, \\ \partial_t B + \dfrac{1}{\mu\sigma}\operatorname{rot}\operatorname{rot} B - \operatorname{rot}(v \times B) &=& 0\,, \\ \operatorname{div} B &=& 0\,, \end{array}\right. \tag{3.6}$$

que l'on complète par les conditions aux limites (en temps et en espace) suivantes :

$$\left\{\begin{array}{rcl} v(0,x) &=& v^0(x) \\ B(0,x) &=& B^0(x) \end{array}\right. , \quad \text{et} \quad \left\{\begin{array}{rcl} v &=& 0 \\ B\cdot n &=& 0 \\ \operatorname{rot} B \times n &=& 0 \end{array}\right. \quad \text{sur } \partial\Omega\,. \tag{3.7}$$

Donnons à présent quelques informations sur l'analyse mathématique du système (3.6)-(3.7). L'étude de ce genre de modèle a été menée pour la première fois par G. Duvaut et J.-L. Lions [28] dans le contexte des inéquations variationnelles, puis, comme nous l'avons mentionné en introduction, étendue par M. Sermange et R. Temam [92]. La procédure suivie pour la démonstration de l'existence de solutions (faibles) est très similaire à celle résumée p.34 pour les équations de Navier-Stokes. En effet, elle consiste dans un premier temps à écrire le problème sous la forme variationnelle

$$\left\{\begin{array}{rcl} \rho\displaystyle\int_\Omega \partial_t v\cdot\tilde v+\rho\int_\Omega (v\cdot\nabla)\tilde v+\eta\int_\Omega \nabla v:\nabla\tilde v-\frac{1}{\mu}\int_\Omega(\operatorname{rot} B\times B)\cdot\tilde v &=& \displaystyle\int_\Omega f\cdot\tilde v\,,\quad \forall\,\tilde v\in V\,,\\[2ex] \rho\displaystyle\int_\Omega \partial_t B\cdot\tilde B+\frac{1}{\mu\sigma}\int_\Omega \operatorname{rot} B\cdot\operatorname{rot}\tilde B-\int_\Omega \operatorname{rot}(v\times B)\cdot\tilde B &=& 0\,,\quad \forall\,\tilde B\in W\,, \end{array}\right.$$

avec $W=\{\tilde B\in H^1(\Omega)^d\,|\,\operatorname{div}\tilde B=0,\ \tilde B\cdot n=0\}$ de norme $\int_\Omega\|\operatorname{rot}\tilde B\|^2$, puis à établir l'estimation

$$\rho\,|v(t)|_H^2+|B(t)|_H^2+\int_0^t\left(\eta\,|\nabla v|_H^2+\frac{2}{\mu}\,|\nabla B|_H^2\right)\ \le\ \rho\,|v(0)|_H^2+|B(0)|_H^2+c\,|f|^2_{L^2(0,T;H)}\,,\quad c\in\mathbb{R}^+.$$

Dans un deuxième temps, on considère la décomposition B_h de B sur la base orthonormale de H formée des vecteurs propres du problème spectral en (μ_n, C_n)

$$\int_\Omega \operatorname{rot} C_n\cdot\operatorname{rot}\tilde B=\mu_n\int_\Omega C_n\cdot\tilde B\,,\quad\forall\,\tilde B\in W\,,$$

assortie de celle déjà décrite pour v p.34, ce qui permet d'établir l'analogue de l'estimation ci-dessus pour ces solutions de dimension finie, dont on connaît l'existence. De là on déduit que (v_h) et $(\partial_t v_h)$ sont bornées dans les mêmes espaces que précédemment, et (B_h) et $(\partial_t B_h)$ sont bornées respectivement dans $L^2(0,T;W)\cap L^\infty(0,T;H)$ et $L^2(0,T;W')$. De tous ces résultats, et d'arguments de compacité du même type que pour le terme de convection appliqués aux termes non linéaires que sont la loi d'Ohm et la force de Lorentz, on déduit que les suites v_h et B_h convergent vers des limites v et B, solutions des équations de la MHD. Signalons pour finir que M.D. Gunzburger *et al.* [46] se sont penchés sur le modèle stationnaire correspondant, avec des conditions aux limites magnétiques différentes du type $B\times n=k$, qui nous seront utiles dans les simulations (voir le chapitre **7**).

3.1.3 Interface libre

Pour achever la modélisation, il reste à traiter le cas où deux fluides non miscibles coexistent. Mathématiquement, cela revient à considérer un fluide dont les caractéritiques (densité, viscosité, conductivité et perméabilité) prennent des valeurs dépendant du sous-domaine *mouvant* dans lequel on se trouve. Il en résulte au niveau du modèle qu'on ne peut plus extraire ces caractéristiques des opérateurs de dérivation (en temps comme en espace). Ainsi, le système (3.6) devient :

$$\left\{\begin{array}{rcl} \partial_t(\rho v)+\operatorname{div}(\rho\, v\otimes v)-\operatorname{div}[\eta\, D(v)]+\nabla p-\dfrac{1}{\mu}\operatorname{rot} B\times B &=& \rho g\\[2ex] \operatorname{div} v &=& 0\\[2ex] \partial_t\rho+\operatorname{div}(\rho v) &=& 0\\[2ex] \partial_t B+\operatorname{rot}\left(\dfrac{1}{\mu\sigma}\operatorname{rot} B\right)-\operatorname{rot}(v\times B) &=& 0\\[2ex] \operatorname{div} B &=& 0 \end{array}\right. \tag{3.8}$$

On notera, en plus de la contrainte d'incompressibilité, la présence de l'équation complète de conservation de la masse, nécessaire pour pouvoir tenir compte des mouvements de l'interface sans transfert de masse entre les deux fluides. Ainsi, comme on peut déduire les deux autres paramètres hétérogènes η et σ - μ étant en fait supposé homogène - de la valeur de ρ, on se retrouve avec une équation (non linéaire) et une inconnue (ρ) supplémentaires. Cette équation, de par sa nature hyperbolique, ne procure pas *a priori* de propriété de compacité sur ρ permettant d'assurer la convergence des termes non linéaires y figurant, suivant le type de raisonnement déjà évoqué au sujet du terme de convection par exemple (p. 34). Il est cependant possible de traiter cette difficulté par la notion de *solution renormalisée*, dont la théorie a été développée par R.J. DiPerna et P.-L. Lions [26]. Par le même type d'approche, J.-F. Gerbeau et C. Le Bris [38] ont obtenu un résultat d'existence de solutions faibles aux équations (3.8)-(3.7) :

Sous les hypothèses $\left\{ \begin{array}{c} \rho^0 \in L^\infty(\Omega)\,;\ u^0 \in L^2(\Omega)^d\,;\ B^0 \in H^{\mathrm{div}}_n(\Omega) \\ f \in L^2(0,T;L^2(\Omega)^d) \\ \eta(\cdot)\,,\ \rho(\cdot) \in \mathcal{C}^0_b(\mathbb{R}^+,\mathbb{R}^+_*)\,; \end{array} \right.$, *il existe*

$$\left\{ \begin{array}{l} \rho \in L^\infty(\Omega\times(0,T)) \cap \mathcal{C}^0(0,T;L^p(\Omega))\,, \quad \forall\, p \geq 1 \\ v \in L^2(0,T;V) \cap L^\infty(0,T;H^{\mathrm{div}}_n(\Omega)) \cap \mathcal{C}^0(0,T;H_w) \\ B \in L^2(0,T;W) \cap L^\infty(0,T;H^{\mathrm{div}}_n(\Omega)) \cap \mathcal{C}^0(0,T;H_w) \end{array} \right. \quad \textit{solution de l'équation}$$

de conservation de la masse au sens des distributions, et du problème

$$\left\{ \begin{array}{l} \displaystyle\iint_{\Omega\times(0,\infty)} \Big(-\rho v\cdot\frac{\partial\phi}{\partial t} - \rho v\otimes v : \nabla\phi + 2\eta D(v) : D(\phi) - \frac{1}{\mu}\mathrm{rot}\, B\times B)\Big)\, dx\, dt = \\ \qquad\qquad \displaystyle\iint_{\Omega\times(0,\infty)} \rho\, g\cdot\phi\, dx\, dt + \int_\Omega \rho^0 v^0\cdot\phi(0,x)\, dx \\ \\ \displaystyle\iint_{\Omega\times(0,\infty)} \Big(-B\cdot\frac{\partial\phi}{\partial t} + \frac{1}{\mu\sigma}\mathrm{rot}\, B\cdot\mathrm{rot}\,\phi - \mathrm{rot}\,(v\times B)\cdot\phi\Big)\, dx\, dt = \int_\Omega B^0\cdot\phi(0,x)\, dx \end{array} \right.$$

pour tout $\phi \in \mathcal{D}(\Omega\times[0,\infty))^d$ *. De plus,* $\mathrm{Mes}\,\{x\in\Omega\,|\,\alpha\leq\rho(t,x)\leq\beta\}$ *reste constant au cours du temps pour tout* $0\leq\alpha\leq\beta<\infty$ *.*

De même que dans le cas des équations monofluides de Navier-Stokes et de la MHD, l'existence d'une solution forte en temps long et l'unicité de la solution faible en général restent à démontrer. La preuve du théorème ci-dessus est basée sur l'étude d'un problème *régularisé* intermédiaire, dans lequel la vitesse, la viscosité, la conductivité et les termes non linéaires sont approchés par des expressions "lissées" permettant la prise en compte de la discontinuité des grandeurs physiques à l'interface. Les travaux de S.N. Antontsev *et al.* ([4]) contiennent le même type d'idées pour traiter les équations de Navier-Stokes bidimensionnelles de densité variable. On pourra se référer à P.-L. Lions [67] pour une étude mathématique générale des fluides à densité variable.

Remarque : dans les simulations numériques (cf. **3.3.1**), nous imposerons d'autres conditions aux frontières que (3.7) pour permettre le glissement de l'interface sur les parois, typiquement $v\cdot n = 0$ sur les bords en question. Les deux alternatives entrent dans le cadre général des *conditions de glissement de Navier*, du type $\left\{ \begin{array}{r} v\cdot n = 0 \\ \theta\, v\cdot n + (1-\theta)\,\sigma\, n\cdot t = 0 \end{array} \right.$, $\theta\in[0,1]$, pour lesquelles le théorème ci-dessus reste valable.

3.2 Modélisation linéaire

Nous avons mentionné dans le premier chapitre un type d'instabilités reposant sur une interaction entre les courants électriques horizontaux et le champ magnétique vertical ambiant. Dans le cas où ce dernier est supposé constant, ce phénomène est désigné par *mécanisme de Sele*, du nom de celui qui l'a pour la première fois mis en évidence (cf. [91]). D'autres auteurs ont par la suite proposé des modèles linéaires de cuves reposant sur un certain nombre d'hypothèses physiques conduisant à des critères d'instabilité du même type, qu'on regroupe génériquement sous le nom de *critère de Sele*. On citera notamment P.A. Davidson et R.I. Lindsay [23], ainsi que V. Bojarevics et M.V. Romerio [13]. Dans l'exposé ci-dessous, nous focalisons notre attention sur ces derniers, et renvoyons à J.-F. Gerbeau *et al.* [39] Chap. 6 pour une revue détaillée de la littérature à ce sujet.

3.2.1 Équations de Saint-Venant pour la MHD

Les premières hypothèses physiques consistent, à partir du modèle (3.8)-(3.7), à négliger :

H_1. les courants induits dans la loi d'Ohm,

H_2. la dérivée temporelle $\partial_t B$ dans l'équation de Maxwell,

H_3. la dissipation mécanique dans les équations fluides,

de sorte qu'on obtient un modèle MHD simplifié basé sur un couplage entre les équations d'Euler et de Maxwell :

$$\left\{\begin{array}{rcl} \partial_t(\rho v) + \operatorname{div}(\rho\, v\otimes v) + \nabla p - \dfrac{1}{\mu}\operatorname{rot} B\times B &=& \rho g \\ \operatorname{div} v &=& 0 \\ \partial_t\rho + \operatorname{div}(\rho v) &=& 0 \\ \dfrac{1}{\mu}\operatorname{rot}\Big(\dfrac{1}{\sigma}\operatorname{rot} B\Big) &=& 0 \\ \operatorname{div} B &=& 0 \end{array}\right. \tag{3.9}$$

On pose par ailleurs des conditions aux bords de non-pénétration ($u\cdot n = 0$). Dans un deuxième temps, on réintroduit la densité de courant électrique

$$J = \frac{1}{\mu}\operatorname{rot} B\,,\quad \text{ainsi}\quad \operatorname{rot}\Big(\frac{1}{\sigma}J\Big) = 0\,,\quad \text{et donc}\quad \operatorname{div} J = 0\,,\quad \text{et}\quad \exists\,\Phi\quad /\quad J = -\sigma\,\nabla\Phi\,.$$

De cette manière, on peut réécrire le problème (3.9) sous la forme :

$$\left\{\begin{array}{rcl} \partial_t(\rho v) + \operatorname{div}(\rho\, v\otimes v) + \nabla p - J\times B &=& \rho g \\ \operatorname{div} v &=& 0 \\ \partial_t\rho + \operatorname{div}(\rho v) &=& 0 \\ J + \sigma\,\nabla\Phi &=& 0 \\ \operatorname{div} J &=& 0 \end{array}\right.$$

Alors, on procède à l'approximation des eaux peu profondes, en supposant que :

H_4. les dimensions verticales sont négligeables devant les dimensions horizontales,

H_5. les perturbations de l'interface sont négligeables devant les dimensions verticales,

si bien que qu'il est possible de négliger la variation et la composante des inconnues le long de la verticale. Ainsi, on peut approcher la pression par une pression hydrostatique :

$$p = p_{\text{int}} + \rho g\,(h - z)$$

où $p_{\rm int}(t,x,y)$ est la pression sur l'interface, et $h(t,x,y)$ sa hauteur. On pourra consulter O. Pironneau [81] pour plus de précisions sur les approximations de type *shallow water* (ou de Saint-Venant ou des eaux peu profondes). On note alors le nouveau système obtenu :

$$\left\{\begin{array}{rcl} \rho_1\,[\partial_t\,v_{H,1}+v_{H,1}\cdot\nabla_H\,v_{H,1}+g\,\nabla_H h]-(J\times B)_{H,1} &=& -\nabla_H\,p_{\rm int}\,,\\ \partial_t h+{\rm div}\,(h\,u_{H,1}) &=& 0\,,\\ \rho_2\,[\partial_t\,v_{H,2}+v_{H,2}\cdot\nabla_H\,v_{H,2}+g\,\nabla_H h]-(J\times B)_{H,2} &=& -\nabla_H\,p_{\rm int}\,,\\ \partial_t h-{\rm div}\,(h\,u_{H,2}) &=& 0\,, \end{array}\right. \tag{3.10}$$

où l'indice H souligne que seules les composantes horizontales sont prises en compte (à l'exception du terme $(J\times B)$ où il s'agit d'une moyenne verticale) et l'indice 1 (resp. 2) que la quantité est prise l'aluminium (resp. l'électrolyte). De cette manière, les inconnues sont définies sur le rectangle $\Omega_H=[0,L_x]\times[0,L_y]$.

3.2.2 Linéarisation

On considère les perturbations autour d'un état stationnaire où l'interface est horizontale, et $(u_{H,\{1,2\}})_0=0$, ce qui permet de supprimer le terme de convection. Par ailleurs, en notant h_1 (resp. h_2) la hauteur moyenne d'aluminium (resp. d'électrolyte), on note $\eta(t,x,y)=h(t,x,y)-h_1$ la perturbation sur la hauteur d'interface à l'équilibre, si bien qu'après des manipulation algébriques élémentaires, le système (3.10) prend la forme :

$$\Big(\frac{\rho_1}{h_1}+\frac{\rho_2}{h_2}\Big)\,\partial_t^2\eta-(\rho_1-\rho_2)\,g\,\Delta_H\eta={\rm div}\,((J\times B)_{H,2}-(J\times B)_{H,1})\,. \tag{3.11}$$

La condition de non-pénétration, quant à elle, s'exprime (n désignant la normale à $\partial\Omega_H$) :

$$[(\rho_1-\rho_2)\,g\,\nabla_H\eta+(J\times B)_{H,2}-(J\times B)_{H,1}]\cdot n=0\,. \tag{3.12}$$

On suppose de plus qu'à l'état stationnaire, $J_0=-\|J_0\|e_z$, ${\rm div}\,B_0=0$ et ${\rm rot}\,(J_0\times B_0)=0$ (équivalent à $\partial_z B_0=0$, cf. A.D. Sneyd [93]), et on émet l'hypothèse supplémentaire :

H_6. le problème électrique est simplifié en passant à la limite

$$\sigma_2\ll\sigma_{\rm anode}\ll\sigma_1\,.$$

Alors, on note $j=J-J_0$ et $b=B-B_0$ les perturbations en courant électrique et champ magnétique, de sorte que :

$$j_2=-\frac{J_0\eta}{h_2}\,e_z\quad\text{et}\quad j_1=j_{H,1}-\frac{J_0\eta}{h_2}\frac{z}{h_1}\,e_z\,,\quad\text{où}\quad j_{H,1}=-\sigma_1\,\nabla_H\,\varphi\,, \tag{3.13}$$

et où φ est la perturbation sur le potentiel électrique, solution de :

$$\left\{\begin{array}{rcll} -\Delta_H\,\varphi &=& \dfrac{J_0\,\eta}{h_1\,h_2\,\sigma_1} & \text{dans}\quad \Omega_H\,,\\ \partial_n\varphi &=& 0 & \text{sur}\quad \partial\Omega_H\,. \end{array}\right. \tag{3.14}$$

Enfin, les courants électriques horizontaux étant faibles devant les courants verticaux (d'après l'hypothèse H_4), et par d'autres arguments portant sur les ordres de grandeur (cf. [23], [94]), on peut approcher la perturbation sur le membre de droite dans (3.11) par :

$$((J\times B)_{H,2}-(J\times B)_{H,1})\approx -j_{H,1}\times(B_{0,z}\,e_z)\,. \tag{3.15}$$

Alors, en combinant (3.11), (3.12), (3.13), (3.14) et (3.15), on obtient le système final

$$\left\{\begin{array}{l} \left\{\begin{array}{rcll} \partial_t^2\,\eta - c^2\,\Delta_H\,\eta &=& c^2\,(\partial_y\,\varphi\,\partial_x\,B_{0,z} - \partial_x\,\varphi\,\partial_y\,B_{0,z}) & \text{dans} \quad \Omega_H\times[0,T] \\ -\Delta_H\,\phi &=& S\,\eta & \end{array}\right. \\ \\ \left\{\begin{array}{rcll} \partial_n\,\eta &=& B_{0,z}\,(\partial_x\,\phi\,n_y - \partial_y\,\phi\,n_x) & \text{sur} \quad \partial\Omega_H\times[0,T] \\ \partial_n\,\varphi &=& 0 & \end{array}\right. \\ \\ \left\{\begin{array}{rcll} \eta(t=0) &=& \eta^0 & \text{dans} \quad \Omega_H \\ \partial_t\eta(t=0) &=& \eta^1 & \end{array}\right. \end{array}\right. \tag{3.16}$$

$$\text{avec} \quad c^2 = \frac{(\rho_1-\rho_2)\,g}{\frac{\rho_1}{h_1}+\frac{\rho_2}{h_2}}\,\frac{\mathcal{T}^2}{L^2} \quad \text{et} \quad S = \frac{J_0\,B_0\,L^2}{h_1\,h_2\,(\rho_1-\rho_2)\,g} \quad \text{(constante de Sele)}\,,$$

où L et $\mathcal{T}$ sont la longueur et le temps caractéristiques.

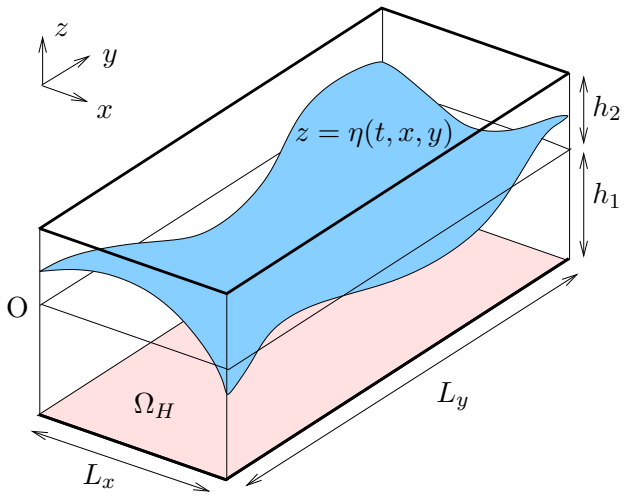


FIG. 3.1 – Modèle de type shallow-water linéarisé

En s'appuyant sur les propriétés spectrales de l'opérateur de Neumann (cf. **3.3.2**), il est possible d'écrire le système ci-dessus sous la forme

$$d_t^2\eta_i + c^2\,k_i^2\eta_i = c^2\,S\sum_{j\in\mathbb{N}} G_{i,j}\eta_j\,, \quad \forall\, i\in\mathbb{N}\,,$$

avec k_i^2 de l'ordre de i^2, et $|G_{i,j}|$ de l'ordre de $1/(k_ik_j)$. Alors, en multipliant par $d_t\eta_i$ et en sommant sur i, on obtient (C désignant des constantes diverses)

$$\frac{1}{2}\sum_{i\in\mathbb{N}} d_t[(d_t\eta_i)^2 + c^2\,k_i^2\eta_i^2] \le C\sum_{(i,j)\in\mathbb{N}^2}\eta_j\frac{d_t\eta_i}{k_i\,k_j} \le C\Big(\sum_{i\in\mathbb{N}}\frac{d_t\eta_i}{k_i}\Big)\Big(\sum_{j\in\mathbb{N}}\frac{\eta_j}{k_j}\Big) \le C\sqrt{\sum_{i\in\mathbb{N}}(d_t\eta_i)^2}\sqrt{\sum_{j\in\mathbb{N}}\eta_j^2}$$

(où l'on a utilisé le fait que $\sum_{i\in\mathbb{N}}(1/k_i^2)<\infty$) ; ainsi

$$\frac{1}{2}\sum_{i\in\mathbb{N}} d_t[(d_t\eta_i)^2 + k_i^2\eta_i^2] \le C\sum_{i\in\mathbb{N}}\Big((d_t\eta_i)^2 + k_i^2\eta_i^2\Big)\,,$$

d'où une estimation a priori, qu'on peut utiliser dans une preuve de type point fixe (comme pour le thérème de Cauchy-Lipschitz), en travaillant dans l'espace des suites $(\eta_i(t))$ qui sont continues en temps à valeur dans l^∞.

3.3 Simulation

3.3.1 Modèle non linéaire

Dans les démonstrations d'existence de solutions aux équations de Navier-Stokes et de la MHD, la contrainte d'incompressibilité a été incorporée dans l'espace des vitesses, ce qui a permis d'éliminer la pression des inconnues. Du point de vue pratique, cette approche a pour inconvénient d'imposer aux vitesses d'être discrétisées sur un espace à divergence nulle, pour lequel il est possible de construire une base (cf. [48]) qu'il est cependant compliqué d'implémenter. Ainsi, la méthode des éléments P^1 fournit une discrétisation commode de l'espace $V = H_0^1(\Omega)^d$, qu'il est possible de choisir comme espace des vitesses. En effet, si l'on pose $M = \{\tilde{p} \in L^2(\Omega) \,|\, \int_\Omega \tilde{p} = 0\}$, on peut établir (cf. V. Girault et P.-A. Raviart [41]) l'existence d'une solution (v, p) à la formulation variationnelle alternative à (3.3) - V étant différent :

$$\left\{ \begin{array}{rcll} \rho\displaystyle\int_\Omega \partial_t v\cdot\tilde{v} + \rho\int_\Omega v\cdot\nabla v\cdot\tilde{v} + \eta\int_\Omega \nabla v : \nabla\tilde{v} - \int_\Omega p\,\mathrm{div}\,\tilde{v} & = & \displaystyle\int_\Omega f\cdot\tilde{v}\,, & \forall\,\tilde{v}\in V\,, \\ \displaystyle\int_\Omega \mathrm{div}\, v\,\tilde{p} & = & 0\,, & \forall\,\tilde{p}\in M\,. \end{array} \right. \tag{3.17}$$

Cette approche n'est pas sans conséquence sur la discrétisation à mettre en œuvre : le choix des espaces V et M est contraint par la condition *inf-sup* (ou condition de *Babuška-Brezzi*, cf. [6],[16]). Il en résulte que l'interpolation par éléments finis P^1-Lagrange en vitesse et en pression n'est pas stable pour la formulation ci-dessus, mais qu'il faut soit :

- monter en précision sur la vitesse en augmentant $\mathrm{Card}\,\Sigma$ (cf. **2.2.1.b**), ce qui mène par exemple aux éléments finis P^2/P^1 [51], P^1-*iso*-P^2/P^1 [8] ou encore P^1-*bulle*$/P^1$ [5].
- *stabiliser* la discrétisation de la formulation (3.17) (cf. T.J.R. Hughes *et al.* [53]), ce qui est moins coûteux en temps de calcul mais ajoute un paramètre numérique au modèle.

Les mêmes considérations s'appliquent aux équations de la MHD, de sorte qu'une formulation variationnelle en éléments finis P^1 *stablisés* comprenant la pression peut être utilisée pour la simulation du modèle (3.6)-(3.7). Précisons quand même que des hypothèses supplémentaires de régularité et de convexité du domaine de définition Ω sont nécessaires pour l'interpolation P^1 du champ magnétique. Dans le cas non convexe, on doit utiliser l'élément fini de Nédelec (cf. [77]).

La principale difficulté du passage des équations de Navier-Stokes au système qui nous intéresse est, encore une fois, la prise en compte de l'interface libre. Pour des raisons de clarté, nous exprimons dans un premier temps les équations bifluides de la MHD sous leur forme *adimensionnée* : en définissant par $\mathcal{L}$, $\mathcal{U}$ et $\mathcal{B}$ les longueur, vitesse et champ magnétique caractéristiques, on pose

$$\begin{array}{rcll} Re_i & = & \dfrac{\rho_i\,\mathcal{U}\,\mathcal{L}}{\eta(\rho_i)} & \text{(Reynolds)}\,, \\ Rm_i & = & \mu\,\sigma(\rho_i)\,\mathcal{U}\,\mathcal{L} & \text{(Reynolds magnétique)}\,, \\ \zeta_i & = & \rho_i/\rho_1 & \text{(densité adimensionnée)}\,, \end{array} \qquad \begin{array}{rcll} Fr & = & \dfrac{\mathcal{U}^2}{g\,L} & \text{(Froude)}\,, \\ S & = & \dfrac{\mathcal{B}^2}{\mu\rho_1\mathcal{U}^2} & \text{(couplage)}\,, \end{array}$$

de sorte qu'on écrit - où Ω_1 (resp. Ω_2) est le domaine d'évolution de l'aluminium (resp. électrolyte)

$$\left\{ \begin{array}{rcl} \left.\begin{array}{rcl} \mathrm{div}\, v \;=\; \mathrm{div}\, B & = & 0 \\ \partial_t\,\rho + \mathrm{div}\,(\rho v) & = & 0 \end{array}\right\} & & \text{dans } \Omega\,, \\ \left.\begin{array}{rcl} \zeta_i\dfrac{\partial v}{\partial t} + \zeta_i(v\cdot\nabla)\,v - \mathrm{div}\left(\dfrac{\zeta_i}{Re_i}\nabla v\right) + \nabla p - S\,\mathrm{rot}\,B\times B & = & -\zeta_i\dfrac{e_z}{Fr} \\ \dfrac{\partial B}{\partial t} + \mathrm{rot}\Big(\dfrac{1}{Rm_i}\,\mathrm{rot}\,B\Big) - \mathrm{rot}\,(v\times B) & = & 0 \end{array}\right\} & & \text{dans } (\Omega_i)_{i=1,2}\,. \end{array} \right. \tag{3.18}$$

On complète ce système par les conditions aux limites

$$\left\{\begin{array}{rcll} \left.\begin{array}{rcl} v &=& v^0 \\ B &=& B^0 \end{array}\right\} & & \text{en } t=0 & \forall\, x\,, \\ \left.\begin{array}{rcl} v\cdot n &=& 0 \\ B\times n &=& B_0\times n \end{array}\right\} & & \text{sur } \Gamma_1\,, & \forall t\,, \\ \left.\begin{array}{rcl} v &=& 0 \\ B\cdot n &=& 0 \\ \mathrm{rot}\, B\times n &=& 0 \end{array}\right\} & & \text{sur } \Gamma_2\,, & \forall t\,, \end{array}\right. \tag{3.19}$$

où $Gamma_1$ est la partie de $\partial\Omega$ en contact avec l'interface, $\Gamma_2 = \partial\Omega \setminus \Gamma_1$; et les conditions interfaciales (continuité du tenseur des contraintes et du champ électrique)

$$\left.\begin{array}{rcl} \left(\dfrac{1}{Re_1}\, D(v) - p\, I_d\right)\Big|_{\Omega_1} n &=& \left(\dfrac{1}{Re_2}\, D(v) - p\, I_d\right)\Big|_{\Omega_2} n \\ \left(\dfrac{1}{Rm_1}\mathrm{rot}\, B - v\times B\right)\Big|_{\Omega_1} \times n &=& \left(\dfrac{1}{Rm_2}\mathrm{rot}\, B - v\times B\right)\Big|_{\Omega_2} \times n \end{array}\right\} \quad \text{sur l'interface}\,, \quad \forall t\,. \tag{3.20}$$

La forme des cuves est modélisée par un parallélépipède ou un cylindre vertical en trois dimensions, et par un rectangle en deux dimensions ; et les paramètres physiques suivants sont fixés :

$$\left\{\begin{array}{rcl} \mathcal{L} &=& 1\, m \\ \mathcal{U} &=& 0.1\, m.s^{-1} \\ \mathcal{B} &=& 5\, mT \end{array}\right. , \quad \left\{\begin{array}{rcl} \rho_1 &=& 2300\, kg.m^{-3} \\ Re_1 &=& 1923 \\ Rm_1 &=& 1 \end{array}\right. , \quad \left\{\begin{array}{rcl} \rho_2 &=& 2150\, kg.m^{-3} \\ Re_2 &=& 840 \\ Rm_2 &=& 10^{-4} \end{array}\right. .$$

Tous ces paramètres sont réalistes, à l'exception des nombres de Reynolds qui sont divisés par un facteur 100 (cf. Tab. 1.1 p.6). En choisissant ainsi la manière la plus simple de modéliser la dissipation d'énergie occasionnée les phénomènes turbulents, on s'affranchit d'un modèle de type $k-\varepsilon$ (cf. [59]) ou encore *LES* (*Large-Eddy Simulation*). Les simulations effectuées ici rentrent plutôt dans la classe *DNS* (*Direct Numerical Simulation*), à ceci près que l'écoulement est laminarisé. Cette approche est retenue pour privilégier une étude qualitative des effets magnétohydrodynamiques sur le comportement de l'interface, à travers son suivi explicite (*interface tracking*). Pour cela, nous représentons son mouvement par la dérivée temporelle de la transformation régulière $\hat{\mathcal{A}}_t(\hat{x})$ qui à un domaine de référence $\hat{\bar{\Omega}} = \hat{\bar{\Omega}}_1 \cup \hat{\bar{\Omega}}_2$, fait correspondre $\bar{\Omega}(t) = \bar{\Omega}_1(t) \cup \bar{\Omega}_2(t)$, et telle que $\hat{\mathcal{A}}_0 = I_d$:

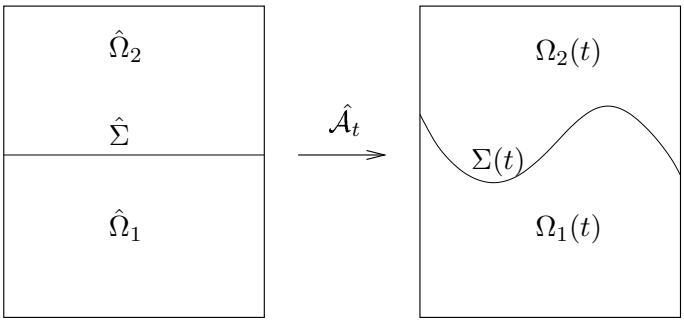


Fig. 3.2 – Une bijection qui associe $\Omega(t)$ à $\hat{\Omega}$.

Ainsi, la nouvelle inconnue est la *vitesse du domaine*

$$w(t,x) = \partial_t \hat{\mathcal{A}}_t(\hat{\mathcal{A}}_t^{-1}(x))\,.$$

Si l'on définit à présent, pour toute fonction $\varphi(t,x)$, son analogue "lagrangienne" $\Phi(t,\hat{x})$ par

$$\Phi(t,\hat{x}) = \varphi(t,\hat{\mathcal{A}}_t(\hat{x}))\,,$$

sa dérivée temporelle est égale à la dérivée de φ sur le domaine mouvant, qu'on note $D_t\varphi$:

$$\partial_t\Phi(t,\hat{x}) = D_t\varphi(t,x) = \partial_t\varphi(t,x) + \partial_x\varphi(t,x)\cdot\partial_t\hat{\mathcal{A}}_t(\hat{x})\,,\quad \text{avec}\quad x = \hat{\mathcal{A}}_t(\hat{x})\,,$$

$$\text{soit}\quad D_t\varphi = \partial_t\varphi + w\cdot\nabla\varphi\,.$$

La "dérivée lagrangienne" de la densité, en particulier, est nulle, car elle prend une valeur constante sur chaque partie $\Omega_1(t)$ et $\Omega_2(t)$:

$$D_t\rho = 0\,,\quad \text{et donc}\quad \partial_t\rho = -w\cdot\nabla\rho\,.$$

Compte tenu de la conservation de la masse et de l'incompressibilité de chaque fluide, il en résulte

$$(w-v)\cdot\nabla\rho = 0\,.$$

Par ailleurs, si on note n_Σ la normale à l'interface extérieure à Ω_1, et δ_Σ la mesure de Dirac sur Σ (voir aussi **8.2**), alors $\nabla\rho = (\rho_2-\rho_1)\,\delta_\Sigma\,n_\Sigma$, d'où l'équation

$$w\cdot n_\Sigma = v\cdot n_\Sigma\,,$$

complétée par la condition aux frontières

$$w\cdot n = v\cdot n\quad \text{sur } \partial\Omega\,.$$

On a ainsi obtenu une caractérisation de l'hypothèse implicite que $\hat{\mathcal{A}}_t$ transforme $\hat{\Omega}_1$ en $\Omega_1(t)$ et $\hat{\Omega}_2$ en $\Omega_2(t)$. Cela suffit à déterminer le mouvement de l'interface, et autorise même plusieurs choix possibles pour w. En particulier, $w = v$ sur l'ensemble du domaine correspond à une description lagrangienne du mouvement, mais n'est pas une solution adaptée aux méthodes d'éléments finis, car il faudrait dans ce cas que les nœuds du maillage suivent le mouvement du fluide, ce qui n'est pas possible compte tenu de la condition de convexité des éléments (cf. **6.3.1.b**). On adopte ainsi un point de vue *arbitrairement* lagrangien ou eulérien en fonction des facilités numériques que celui-là procure. Cette méthode, qui porte le nom de formulation *ALE* (pour *Arbitrary Lagrangian-Eulerian*) a été introduite pour la première fois par C.W. Hirt *et al.* [50], et largement utilisée par la suite. Dans notre situation, le meilleur compromis est de considérer une vitesse du domaine purement verticale, qui n'existe que pour assurer le mouvement de l'interface et éventuellement équilibrer la taille des éléments sur l'ensemble du maillage. Ainsi, on résout à chaque pas de temps le problème

$$\left\{\begin{array}{rcll} -\Delta w &=& 0 & \text{dans } (\Omega_i)_{i=1,2}\,,\\ w\cdot n &=& 0 & \text{sur } \partial\Omega\,,\\ w\cdot n_\Sigma &=& \dfrac{v\cdot n_\Sigma}{n_\Sigma\cdot e_z} & \text{sur } \Sigma\,. \end{array}\right. \tag{3.21}$$

Nous sommes à présent en mesure d'écrire la formulation faible du problème, en vue de sa discrétisation. On utilise à cette fin la *formule de Reynolds*

$$\frac{d}{dt}\int_{\Omega(t)}\varphi = \int_{\Omega(t)}\left(\frac{\partial\varphi}{\partial t} + \operatorname{div}(w\varphi)\right),$$

qui permet de dériver une intégrale définie sur un domaine transporté à la vitesse w.

Cette propriété s'étend aux formulations variationnelles définies à l'aide de fonctions test $\tilde{\varphi}$ telles que $D_t\tilde{\varphi} = 0$, en effet dans ce cas

$$\frac{d}{dt}\int_{\Omega(t)} \varphi\,\tilde{\varphi} = \int_{\Omega(t)} \left(\frac{\partial \varphi}{\partial t} + \operatorname{div}(w\varphi)\right)\tilde{\varphi}\,.$$

Alors, la formulation faible ALE du problème (3.18) s'écrit :

$$\left\{\begin{array}{l} \dfrac{d}{dt}\displaystyle\int_{\Omega(t)} \zeta\, v\cdot\tilde{v} + \int_{\Omega(t)} \zeta\,(v-w)\cdot\nabla v\cdot\tilde{v} - \int_{\Omega(t)} \zeta\, v\cdot\tilde{v}\operatorname{div} w + \int_{\Omega(t)} \dfrac{\zeta}{Re}\nabla v : \nabla\tilde{v} - \int_{\Omega(t)} p\operatorname{div}\tilde{v} \\ \qquad\qquad -S\displaystyle\int_{\Omega(t)} \operatorname{rot} B\times B\cdot\tilde{v} = \int_{\Omega(t)} \dfrac{\zeta}{Fr}\, e_z\cdot\tilde{v}\,, \\ \dfrac{d}{dt}\displaystyle\int_{\Omega(t)} B\cdot\tilde{B} - \int_{\Omega(t)} w\cdot\nabla B\cdot\tilde{B} - \int_{\Omega(t)} B\cdot\tilde{B}\operatorname{div} w + \int_{\Omega(t)} \dfrac{1}{Rm}\left(\operatorname{rot} B\cdot\operatorname{rot}\tilde{B} + \operatorname{div} B\operatorname{div}\tilde{B}\right) \\ \qquad\qquad -\displaystyle\int_{\Omega(t)} v\times B\cdot\operatorname{rot}\tilde{B} = 0\,, \\ \qquad\qquad -\displaystyle\int_{\Omega(t)} \operatorname{div} v\,\tilde{p} = 0\,, \end{array}\right.$$

$$\forall\,(\tilde{v},\tilde{B},\tilde{p})\in V_T\times W_T\times M_T\,,\ \text{avec}\ \left\{\begin{array}{lcl} V_T &=& \{\tilde{v}\in H_0^1(\Omega)^d \,|\, D_t\tilde{v}=0\} \\ W_T &=& \{\tilde{B}\in H_n^1(\Omega) \,|\, D_t\tilde{B}=0\} \\ M_T &=& \{\tilde{p}\in L^2(\Omega) \,|\, \textstyle\int_{\Omega(t)}\tilde{p}=0\} \end{array}\right. . \tag{3.22}$$

Pour discrétiser ces espaces par éléments finis, il suffit alors d'appliquer la transformation $\hat{\mathcal{A}}_t$ en chaque point du maillage et de calculer leurs fonctions de forme sur le maillage transporté $\Omega(t)$. En définissant la distribution de densité non pas en fonction des coordonnées spatiales, mais en fonction des éléments géométriques parcourus lors de l'*assemblage* (cf. **6.3.1.b**), il suffit d'attribuer la densité de l'aluminium au sous-domaine discret constitué de la réunion des éléments correspondants, et de même pour l'électrolyte. C'est ainsi qu'on remplace l'inconnue ρ par l'inconnue w.

Sur des quadrangles, l'interpolation la moins coûteuse possible est la méthode Q^1 stabilisée, dont on fixe son paramètre à 0.1 (cf. L.P. Franca et S.L. Frey [35]) pour obtenir des résultats intéressants sur un maillage de quelques milliers d'éléments (cf. *infra* et chapitre **7**). La discrétisation temporelle s'effectue au moyen d'un *schéma d'Euler semi-implicite*, qui nécessite la résolution d'un problème du type (2.19) à chaque itération en temps, dont le détail sera abordé au chapitre **6.3**. Si l'on note (v_h^n, B_h^n) la vitesse et le champ magnétiques discrétisés au temps $n\Delta t$, cet algorithme possède la propriété de stabilité dans la norme de l'énergie

$$\frac{1}{2\Delta t}\int_\Omega (\zeta\|v^{n+1}\|^2 + S\|B^{n+1}\|^2) + \int_\Omega \left(\frac{\zeta}{Re}\|\nabla v^{n+1}\|^2 + \frac{S}{Rm}\|\nabla B^{n+1}\|^2\right) \le \frac{1}{2\Delta t}\int_\Omega (\zeta\|v^n\|^2 + S\|B^n\|^2)\,. \tag{3.23}$$

Par ailleurs, la masse n'est conservée dans chaque fluide que si la *loi de conservation géométrique* (où $\mathcal{A}_{n,n+1}$ est la transformation géométrique associée à w^n)

$$\int_{\Omega_i^{n+1}} \varphi(x)\,dx - \int_{\Omega_i^n} \varphi\circ\mathcal{A}_{n,n+1}(x)\,dx = \Delta t\int_{\Omega_i^n} \varphi\circ\mathcal{A}_{n,n+1}(x)\operatorname{div} w^n(x)\,dx$$

et la contrainte discrète

$$\int_{\Omega_1^{i+1}} \operatorname{div} v_h^i = 0 \tag{3.24}$$

sont satisfaites (cf. [39]). La première est simplement assurée par le choix purement vertical de w (et donc w^n). La prise en compte de la seconde, en revanche, a des répercussions sérieuses au niveau de l'implémentation comme nous pourrons le voir au chapitre **6**.

Pour l'heure, terminons ce bref aperçu par un résultat numérique important de l'approche non linéaire : on résout le modèle (3.22) dans un domaine cylindrique avec les conditions aux limites

$$v^0 = 0\,, \quad B^0 = 0\,, \quad \text{et} \quad B_0 = \begin{cases} -\alpha\, r \\ 0 \\ B_z \end{cases}, \quad \alpha > 0\,, \quad \text{en coordonnées cylindriques}\,,$$

qui correpondent à une distribution anodique uniforme de l'arrivée de courant. Alors, si on perturbe la gravité pendant la première unité de temps, on obtient un phénomène de roulement de la nappe de métal (*metal pad rolling*), à savoir une rotation inclinée de l'interface dans le sens positif d'axe e_z. Lorsque cette inclinaison s'amplifie au cours du temps jusqu'à toucher le bord supérieur du cylindre, on dit alors que le phénomène est *instable*. Dans le cas contraire, c'est à dire celui où l'amplitude des oscillations diminue progressivement jusqu'à un état stationnaire, on parle de configuration *stable*. En pratique, les cas instables engendrent au bout d'un certain temps des déformations trop importantes du maillage, qui finissent par mettre un terme à la simulation. Le modèle non linéaire permet l'obtention de cas stables sous un certain seuil (typiquement 0.1) portant sur B_z. Ce résultat est à opposer à ceux fournis par les modèles linéaires (cf. Chap. **4**). Donnons pour finir une interprétation physique du phénomène de rolling : la visualisation des champs de vitesses et des courants électriques horizontaux laisse à penser que ce phénomène peut

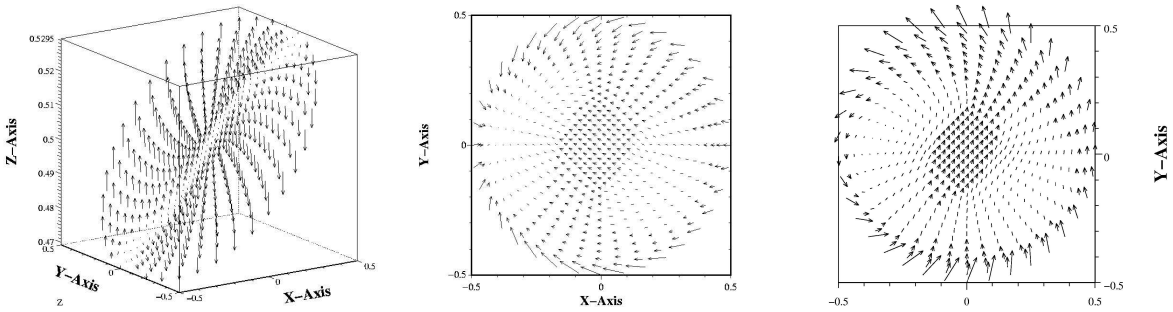


FIG. 3.3 – Vitesse, courants horizontaux et force de Lorentz qui en résulte sur l'interface

s'expliquer simplement par la force de Lorentz, qui trouve son origine dans le mécanisme exposé en **1.2.2**. En effet, d'après la figure 3.3, les courants électriques horizontaux, qui se trouvent majoritairement au voisinage de la paroi, engendrent par interaction avec le champ magnétique vertical une force de Lorentz, qui d'un côté est dirigée vers la paroi, et de l'autre, vers l'intérieur de la cuve. Il en résulte le sens de rotation observé :

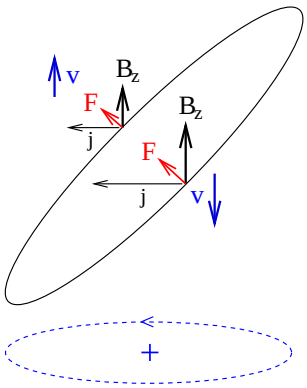


FIG. 3.4 – Sens de rotation induit par une arrivée uniforme du courant

3.3.2 Modèle linéaire

En multipliant la première (resp. deuxième) équation du système (3.16) par la fonction test ζ (resp. ψ) de $H^1(\Omega_H)$, en intégrant par parties et en utilisant les conditions au bord, le problème s'écrit $\forall\, t \in [0,T]$ sous forme variationnelle en espace :

$$\left\{ \begin{array}{l} \text{Trouver} \quad (\eta(t),\varphi(t)) \,\in\, (H^1(\Omega_H))^2 \quad \text{tels que} \quad \forall\,(\zeta,\psi) \,\in\, (H^1(\Omega_H))^2 \quad : \\ \displaystyle\int_{\Omega_H} \Big[\partial_t^2\eta(t)\,\zeta + c^2\,[\nabla_H\eta(t)\cdot\nabla_H\zeta + B_{0,z}\,(\partial_y\varphi(t)\,\partial_x\zeta - \partial_x\varphi(t)\,\partial_y\zeta)]\Big] \;=\; 0\,, \\ \displaystyle\int_{\Omega_H} \Big[\nabla_H\varphi(t)\cdot\nabla_H\psi - S\,\eta(t)\,\psi\Big] \;=\; 0\,. \end{array} \right. \tag{3.25}$$

plus les conditions initiales $\eta_{|t=0} = \eta_0$ et $(\partial_t\eta)_{|t=0} = \eta_1$.

On peut décomposer les solutions η et φ sur la base des "modes gravitationnels" (cf. **4**) :

$$\left\{ \begin{array}{l} \eta(t,x,y) = \displaystyle\sum_{(m,n)\in\mathbb{N}^2} \eta_{m,n}(t)\, f_{m,n}(x,y) \\ \varphi(t,x,y) = \displaystyle\sum_{(m,n)\in\mathbb{N}^2} \varphi_{m,n}(t)\, f_{m,n}(x,y) \end{array} \right. ,$$

avec

$$f_{m,n}(x,y) = \frac{2\,\epsilon_{m,n}}{\sqrt{L_x L_y}} \cos\left(\frac{m\pi}{L_x}\,x\right) \cos\left(\frac{n\pi}{L_y}\,y\right),$$

où la normalisation

$$\epsilon_{m,n} = \left\{ \begin{array}{ccc} 1/2 & \text{si} & (m,n) = (0,0) \\ \sqrt{1/2} & \text{si} & m = 0 \text{ ou } n = 0, \text{ et } (m,n) \neq (0,0) \\ 1 & \text{si} & m \neq 0 \text{ et } n \neq 0 \end{array} \right.$$

fait de $\Big(f_{m,n}(x,y)\Big)_{(m,n)}$ une base orthonormale dans $L^2(\Omega_H)$ et orthogonale dans $H^1(\Omega_H)$. Les fonctions $f_{m,n}$ sont les vecteurs propres de l'opérateur de Neumann $-\Delta_H$ sur le domaine Ω_H, associés aux valeurs propres

$$k_{m,n}^2 = \left(\frac{m\,\pi}{L_x}\right)^2 + \left(\frac{n\,\pi}{L_y}\right)^2.$$

Ainsi, en prenant les éléments de cette base pour fonctions test, le système d'équations (3.25) peut se reformuler

$$\frac{d^2\xi_{m,n}}{dt^2} + c^2\left(k_{m,n}^2\xi_{m,n} \,-\, S\sum_{(m',n')\in\mathbb{N}^2} G_{(m,n),(m',n')}\,\xi_{m',n'}\right) = 0\,, \qquad \text{avec} \qquad \xi_{m,n} = \frac{\eta_{m,n}}{k_{m,n}}\,, \tag{3.26}$$

où la matrice G est facilement calculable et de l'ordre de $\frac{1}{k_{m,n}\,k_{m',n'}}$ (on renvoie en section **4.3.1** pour la forme précise de G). Ainsi, la discrétisation en espace du modèle linéaire s'opère par une méthode spectrale (cf. **2.2.2**), qui permet de simplifier grandement les équations en "diagonalisant" les laplaciens. Pour parvenir à une discrétisation totale du modèle, on peut utiliser la *méthode de Newmark* (cf. [78]), qui est un schéma au différence finies du deuxième ordre devant satisfaire des conditions du même type que dans le cas du schéma d'Euler (stabilité, consistance, cf. **2.2.1.a**). L'implémentation de cette procédure pourra être trouvée au chapitre **5**.

Chapitre 4

Linéaire *versus* non linéaire

4.1 Introduction

La modélisation des phénomènes magnétohydrodynamiques dans les cuves de production d'aluminium par électrolyse est un problème encore très ouvert. D'un côté, le modèle de base que constituent les équations paraboliques de la MHD sans courants de déplacement, pour deux fluides non miscibles, contient diverses non-linéarités (cf. **3.1**), dont la présence d'une interface libre entre les deux fluides. La simulation numérique de ce modèle non linéaire permet d'appréhender les problèmes de stabilité dans les cuves. D'un autre côté, de nombreuses approches reposent sur une version linéarisée de ces équations (cf. **3.2**), qui permet de mener une analyse de stabilité basée sur une étude des modes propres de la cuve. Nous nous proposons dans le présent chapitre de comparer et discuter les résultats issus de ces deux points de vue.

Ces travaux ont fait l'objet d'un proceeding de la conférence ECCOMAS 2006 [40].

4.2 Étude fréquentielle purement hydrodynamique

L'étude de la stabilité linéaire des cuves repose sur la décomposition de la solution du système (3.16) sur la base des "modes gravitationnels". Nous étudions ici, sur un cas bidimensionnel, la cohérence entre cette approche analytique et les résultats fournis par la simulation numérique du problème de Navier-Stokes bifluide.

4.2.1 Calcul analytique des modes gravitationnels sur deux modèles linéarisés

4.2.1.a Un modèle de fluide potentiel à surface libre

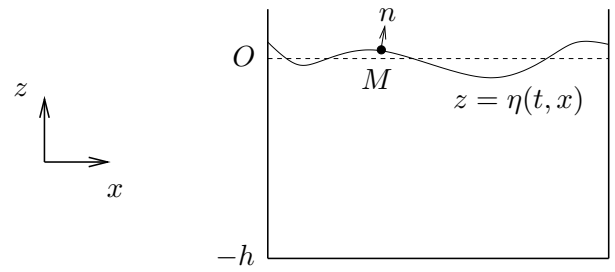
Lorsqu'un fluide parfait incompressible ($\operatorname{div} v = 0$) n'est soumis qu'à des forces dérivant d'un potentiel (comme la gravité), les vitesses telles que

$$\operatorname{rot} v = 0\,,$$

sont des solutions dites *irrotationnelles* aux équations d'Euler (équations de Navier-Stokes sans le terme de viscosité). Pour des conditions aux limites du type $v \cdot n = g$ sur $\partial\Omega$, cette propriété se propage dans le temps et assure ainsi l'unicité de la solution. Il est alors possible de faire dériver la vitesse également d'un potentiel scalaire : $v = \nabla\Phi$, ce qui aboutit au problème de Neumann :

$$\left\{ \begin{array}{rcl} \Delta\Phi & = & 0 \quad \text{dans} \quad \Omega\,, \\ \partial_n\Phi & = & 0 \quad \text{sur} \quad \partial\Omega\,. \end{array} \right. \tag{4.1}$$

Intéressons-nous alors pour commencer à un problème de surface libre dans le cadre de ces hypothèses : considérons un fluide contenu dans un récipient rectangulaire (le cas tridimensionnel ne présentant pas de difficultés supplémentaires), dont on repère le fond par $z = -h$ et la hauteur moyenne de fluide par $z = 0$. Enfin, on suppose la surface paramétrée par la fonction $\eta(t, x)$:



$$OM = \begin{pmatrix} x \\ \eta(t,x) \end{pmatrix}$$

FIG. 4.1 – Paramétrisation de la surface

L'évolution d'un point $M(t)$ de la surface est régie par la condition aux limites caractéristique des écoulements à surface libre :

$$\partial_t (OM) \cdot n = u \cdot n \,,$$

que l'on peut récrire sous la forme

$$- \partial_t \eta = u_x \, \partial_x \eta - u_z \,, \tag{4.2}$$

étant donné que

$$n = \frac{1}{\sqrt{1 + (\partial_x \eta)^2}} \begin{pmatrix} -\partial_x \eta \\ 1 \end{pmatrix} .$$

Nous aurons l'occasion d'utiliser une technique de paramétrisation analogue lorsqu'il s'agira de contrôler l'interface au chapitre **8**. Par ailleurs, l'évolution de la pression suit l'*équation de Bernoulli* (cf. [47]) :

$$\rho \, \partial_t \Phi + \rho \, \frac{\|u\|^2}{2} + p + \rho \, g \, z = \text{constante} \,.$$

Ainsi, si l'on prend pour origine des pressions la pression atmosphérique, le potentiel Φ suit la loi suivante sur la surface libre :

$$\partial_t \Phi + \frac{\|u\|^2}{2} + g \, z = 0 \,. \tag{4.3}$$

Linéarisation

La linéarisation des équations (4.2)-(4.3) autour de l'état d'équilibre $\{u = 0, \, \eta = 0, \, \Phi = 0\}$ où l'interface est plate ($z = 0$) conduit au système (en notant encore η et Φ les perturbations) :

$$\begin{cases} \partial_t \, \eta &= \partial_z \Phi \,, \\ \partial_t \Phi &= -g \, \eta \,, \end{cases}$$

que l'on peut encore écrire :

$$g \, \partial_z \, \Phi + \partial_t^2 \Phi = 0 \,. \tag{4.4}$$

Enfin, rappelons qu'on dispose de la condition de Neumann :

$$\partial_z \, \Phi = 0 \quad \text{en} \quad z = -h.$$

On peut alors définir pour le système (4.1)-(4.4) des solutions du type :

$$\Phi(t, x) = A \cos \left(kx - \omega_k t\right) \text{ch} \left(k(z + h)\right) , \quad \omega_k^2 = g \, k \tanh \left(kh\right) .$$

On admettra que ces résultats se généralisent en trois dimensions à deux fluides de hauteurs moyennes h_1 et h_2 :

$$\omega_k = \sqrt{\frac{g\,(\rho_1-\rho_2)\,k}{\rho_1\coth(kh_1)+\rho_2\coth(kh_2)}}\,,\quad \text{avec}\quad k=\pi\sqrt{\frac{m^2}{L_x^2}+\frac{n^2}{L_y^2}}\,. \tag{4.5}$$

Dans notre cas, rappelons que les longueur et vitesse caractéristiques du phénomène sont $\mathcal{L}=1$ m et $\mathcal{U}=0.1$ m/s. Il en résulte une fréquence caractéristique $\mathcal{F}=\mathcal{U}/\mathcal{L}=0.1$ Hz, par laquelle il faut diviser les fréquences physiques pour les comparer aux valeurs que nous obtenons. Par ailleurs, nous effectuons nos simulations avec un nombre de Froude multiplié par 10, donc il faut prendre $g=1$ dans la formule (4.5). Enfin, si $h=h_1=h_2$, on obtient comme formule analytique pour les fréquences adimensionnées (L_x est la largeur et L_y la longueur de la cuve) :

$$\tilde{f}_{m,n} = \sqrt{\frac{75}{89}\frac{\sqrt{\left(\frac{m}{L_x}\right)^2+\left(\frac{n}{L_y}\right)^2}}{\coth\left[\sqrt{\left(\frac{m}{L_x}\right)^2+\left(\frac{n}{L_y}\right)^2}\,h\right]}}\,,\quad (m,n)\in\mathbb{N}^2\,. \tag{4.6}$$

4.2.1.b Équations de Saint-Venant

Le même type de raisonnement que celui exposé dans le cas magnétohydrodynamique (cf. **3.2**) permet d'arriver à l'équation du type des ondes :

$$\left(\frac{\rho_1}{h_1}+\frac{\rho_2}{h_2}\right)\partial_t^2\eta-(\rho_1-\rho_2)\,g\,\Delta_H\eta=0\,, \tag{4.7}$$

à laquelle on adjoint la condition aux limites de non-pénétration :

$$(\rho_1-\rho_2)\,g\,\nabla_H\eta\cdot n=0\,. \tag{4.8}$$

On peut alors obtenir pour le système (4.7)-(4.8) des solutions du type "ondes gravitationnelles" s'exprimant comme suit :

$$\eta(t,x,y)=\cos(kx+ly-\omega_{k,l}t)\,,$$
$$\text{avec}\quad \omega_{k,l}=\sqrt{(k^2+l^2)\,g\,\frac{\rho_1-\rho_2}{\frac{\rho_1}{h_1}+\frac{\rho_2}{h_2}}}\,,\quad k=\frac{m\,\pi}{L_x}\,,\quad l=\frac{n\,\pi}{L_y}\,,\quad (m,n)\in\mathbb{N}^2\,. \tag{4.9}$$

On applique les mêmes paramètres que dans le cas du modèle de fluide potentiel, si bien que (4.9) conduit aux fréquences adimensionnées :

$$\tilde{f}_{k,l}=\sqrt{\frac{75}{89}\left[\left(\frac{m}{L_x}\right)^2+\left(\frac{n}{L_y}\right)^2\right]h}\,,\quad (m,n)\in\mathbb{N}^2\,. \tag{4.10}$$

Remarquer que l'expression des pulsations $\omega_{k,l}$ peut être obtenue à partir de celle issue du modèle de fluides potentiels lorsque h est "petit". En effet, il suffit dans ce cas de se servir de l'équivalence $\coth x\sim x$ pour la retrouver.

4.2.2 Résultats numériques du modèle non linéaire

Le cas étudié est avec celui d'une cuve bidimensionnelle (rectangle) de largeur $L_x = 4$; et pour différentes valeurs de $h = h_1 = h_2$: 1, 0.5 et 0.3. Par ailleurs, nous perturbons la gravité en modifiant en permanence sa direction de manière périodique (*sloshing*, cf. FIG. 4.3) :

$$\frac{\overline{g}}{\|g\|} = \sin\left(\frac{\pi}{6}\sin\left(4\pi t\right)\right) e_x - \cos\left(\frac{\pi}{6}\sin\left(4\pi t\right)\right) e_y \,, \tag{4.11}$$

de sorte que le système soit excité à la fréquence 2. À celle-là se superposent d'autres fréquences issues des propriétés physiques du problème, les *modes gravitationnels* :

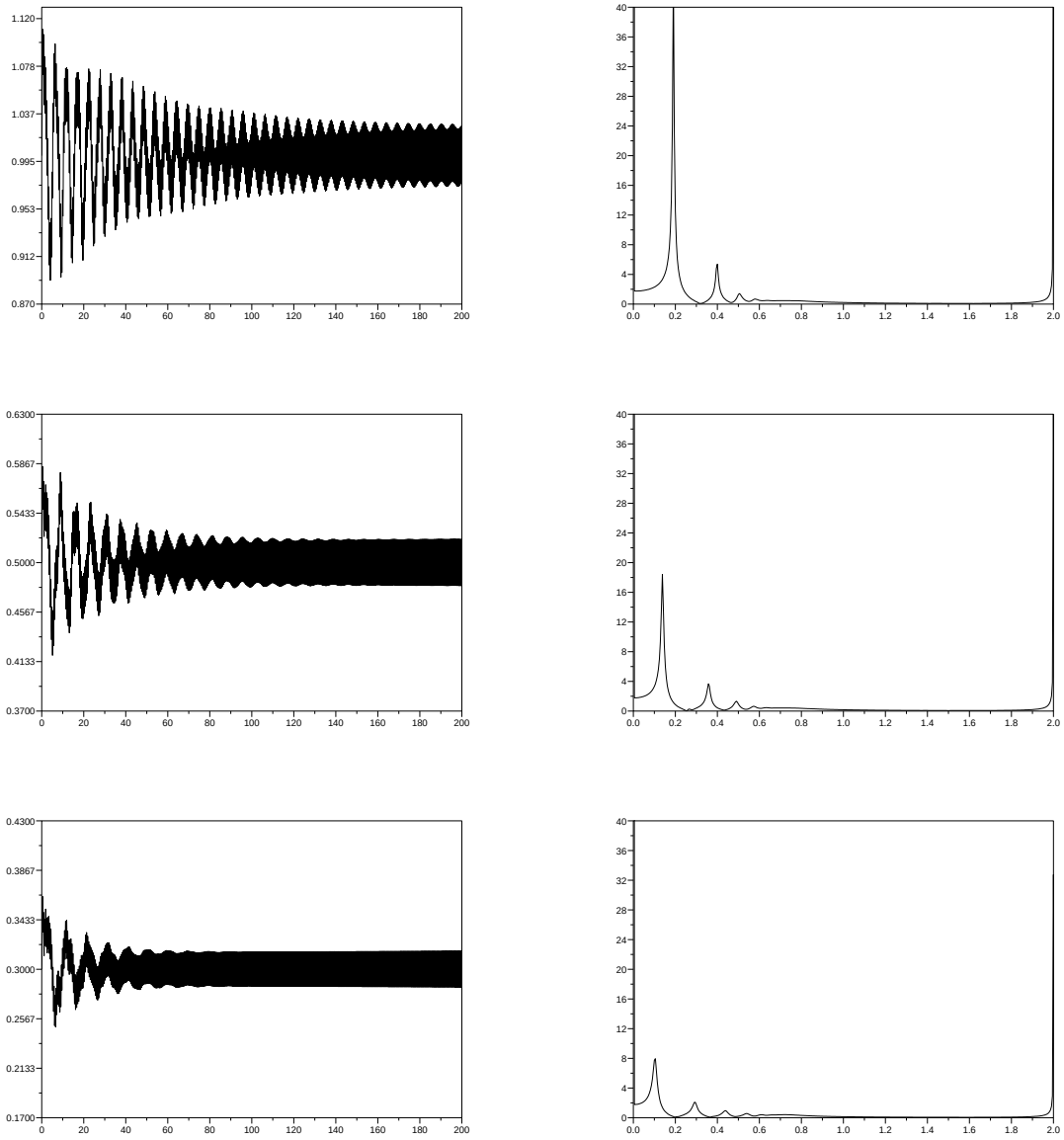


FIG. 4.2 – Altitude du bord de l'interface ($x = 2$) en fonction du temps (à gauche) et transformée de Fourier du signal pour $h = 1$, 0.5 et 0.3. Ces tests sont issus du code parallélisé Mistral (cf. **6**).

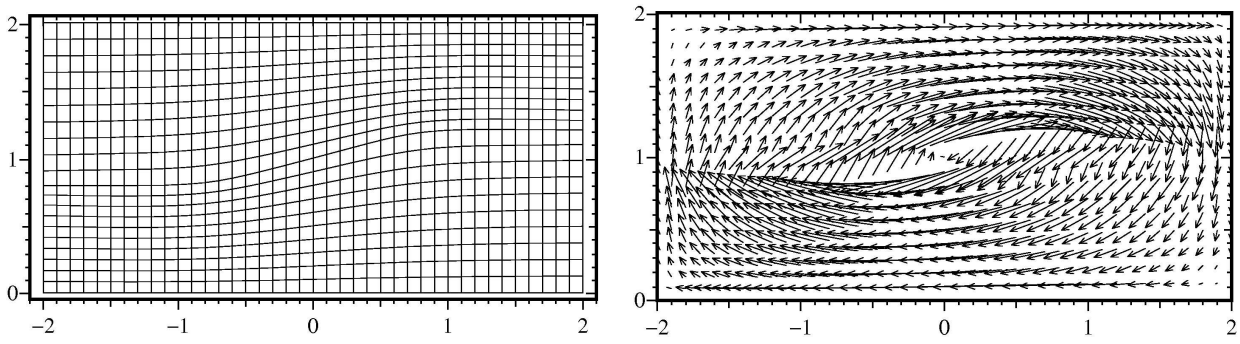


FIG. 4.3 – Le cas test du sloshing ($h = 1$, maillage et vitesse)

On regroupe ici les fréquences adimensionnées obtenues analytiquement et numériquement pour :

· $h = 1.0$:

Mode	Formule (4.6)	Formule (4.10)	*Mistral*
1	0.227	0.229	0.191
2	0.441	0.459	0.400
3	0.634	0.688	0.505
4	0.801	0.918	0.581

· $h = 0.5$:

Mode	Formule (4.6)	Formule (4.10)	*Mistral*
1	0.162	0.162	0.138
2	0.321	0.325	0.357
3	0.476	0.487	0.490
4	0.624	0.649	0.571

· $h = 0.3$:

Mode	Formule (4.6)	Formule (4.10)	*Mistral*
1	0.126	0.126	0.105
2	0.250	0.251	0.293
3	0.374	0.377	0.438
4	0.495	0.503	0.538

On observe que les trois colonnes sont quantitativement proches, ce qui permet de déceler une certaine cohérence du modèle non linéaire avec les modèles simplifiés. D'autre part, il est intéressant de remarquer que Mistral fournit des résultats plus proches du modèle de fluide potentiel que de celui des eaux peu profondes dans les hauteurs de fluide élevées, et, à l'inverse, est en meilleur accord avec le modèle des eaux peu profondes lorsque h est faible. Enfin, la cohérence de Mistral avec les deux modèles linéarisés se dégrade pour $h = 1$, ce qui était prévisible étant donné que ceux-là sont plutôt adaptés aux faibles hauteurs, car construits sur la base de petites perturbations autour d'un état stationnaire.

Par ailleurs, ces résultats montrent qu'une **diminution des hauteurs d'aluminium et d'électrolyte conduit à une diminution de l'amplitude des signaux**. Nous verrons que ce comportement subsiste dans le cas du phénomène de rolling, dont l'amplitude des oscillations augmente par hausse de la hauteur d'électrolyte (cf. **7.1**).

4.3 Étude comparative sur la stabilité du phénomène de rolling

Nous nous proposons à présent de comparer les deux modèles présentés au chapitre **3**. L'approche non linéaire est connue pour être plus compliquée, mais plus solide que l'analyse linéaire sur les questions de stabilité. L'analyse de stabilité linéaire peut en effet amener à la conclusion fausse qu'un état stationnaire est stable. Cela est à relier au fait que l'analyse spectrale ne garantit pas la stabilité des systèmes de dimension infinie (contrairement aux systèmes de dimension finie). De plus, les équations linéarisées autour d'un état stationnaire ne sont valides que pour de petites perturbations. Dans le cas de déviations importantes par rapport à cet état, cette supposition n'est donc plus valable. Cependant, il est intéressant de comprendre les limites de l'approche linéarisée, dans la mesure où celle-là est moins coûteuse du point de vue du temps de calcul.

4.3.1 Modèle linéaire

Rappelons la formulation discrétisée en espace (3.26) du système (3.16) :

$$\frac{d^2\xi_{m,n}}{dt^2} + c^2 k_{m,n}^2 \xi_{m,n} = c^2\, S \sum_{(m',n')\in\mathbb{N}^2} G_{(m,n),(m',n')}\xi_{m',n'}$$

où pour $(m,n),(m',n') \in (\mathbb{N}^2 \setminus (0,0)) \times (\mathbb{N}^2 \setminus (0,0))$,

$$\begin{aligned} G_{(m,n),(m',n')} &= \frac{\epsilon_{m,n}\epsilon_{m',n'}}{L_x\, L_y\, k_{m,n}\, k_{m',n'}} && (4.12)\\ &\Big(m'n(q_{m'+m,n'+n} - q_{m'-m,n'-n} + q_{m'-m,n'+n} - q_{m+m',n'-n}) && (4.13)\\ &+n'm(-q_{m'+m,n'+n} + q_{m'-m,n'-n} + q_{m'-m,n'+n} - q_{m'+m,n'-n})\Big), && (4.14) \end{aligned}$$

et, pour $(m,n) \in \mathbb{Z}^2$,

$$q_{m,n} = \frac{\pi^2}{L_x L_y}\int_{\Omega_H} B_{0,z} \sin\left(\frac{m\pi}{L_x}x\right)\sin\left(\frac{n\pi}{L_y}y\right).$$

Noter que si pour tout ϵ_m, $\epsilon_n \in \{-1,1\}$, $b_{\epsilon_m m,\epsilon_n n} = \epsilon_m\epsilon_n b_{m,n}$, G est une matrice antisymétrique : $G_{(m,n),(m',n')} = -G_{(m',n'),(m,n)}$. On pourra trouver plus de détails dans le chapitre **5**, où nous utilisons le même type de décomposition pour résoudre un problème de stabilisation au moyen des actionneurs h_2 et $B_{0,z}$.

Ainsi, la stabilité de (3.16) repose sur l'analyse spectrale de la matrice $c^2\,(K - SG)$, où les matrices sont indexées par les double indices $(m,n) \in \mathbb{N}^2 \setminus (0,0)$, et K est la matrice diagonale définie par : $K_{(m,n),(m',n')} = k_{m,n}^2\delta_{(m,n),(m',n')}$, où $\delta_{(m,n),(m',n')}$ est la matrice identité. En effet, si (λ,η) est un mode propre de $c^2\,(K - SG)$, une solution de (3.16) est $\exp(-i\omega t)\eta$ avec $\omega^2 = \lambda$. Il est facile de vérifier que la partie réelle de λ est positive (en utilisant le fait que K est une matrice symétrique définie positive et G est une matrice réelle antisymétrique). Par conséquent, ou bien toutes les valeurs propres de $c^2\,(K - SG)$ sont réelles et le système est stable, ou bien il existe une valeur propre avec une partie imaginaire non nulle et le système est instable.

Lorsqu'on augmente la valeur de la composante verticale du champ magnétique, c'est à dire lorsqu'on augmente S, des modes instables apparaissent car certaines valeurs propres commencent à prendre une partie imaginaire non nulle. Pour comprendre cela, on peut simplememnt considérer le problème modèle $K = \begin{bmatrix} k_1^2 & 0 \\ 0 & k_2^2 \end{bmatrix}$ et $G = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Alors les valeurs propres de

$K - SG = \begin{bmatrix} k_1^2 & -S \\ S & k_2^2 \end{bmatrix}$ sont solutions de $\lambda^2 - (k_1^2 + k_2^2)\lambda + (k_1^2 k_2^2 + S^2) = 0$ ce qui est équivalent à $\left(\lambda - \frac{k_1^2+k_2^2}{2}\right)^2 = \left(\frac{k_1^2-k_2^2}{2}\right)^2 - S^2$. Ainsi, quand S augmente, les deux valeurs propres, partant de k_1^2 et k_2^2 pour $\beta = 0$, se rapprochent de plus en plus jusqu'à se rencontrer lorsque $S = \left|\frac{k_1^2-k_2^2}{2}\right|$. Au-delà de ce seuil $\left(S > \left|\frac{k_1^2-k_2^2}{2}\right|\right)$, les deux valeurs propres sont complexes conjuguées : la valeur absolue de leur partie imaginaire est $\left(S^2 - \left(\frac{k_1^2-k_2^2}{2}\right)^2\right)^{1/2}$. Cela illustre le fait qu'une instabilité peut survenir par "collision de deux valeurs propres". Remarquer que la valeur critique $S_{\text{critique}} = \left|\frac{k_1^2-k_2^2}{2}\right|$ est nulle si les deux valeurs propres sont initialement les mêmes : $k_1^2 = k_2^2$. Cette situation se présente en pratique en géométries carrées ($L_x = L_y$) par example, puisque $k_{1,0}^2 = k_{0,1}^2$. Dans ce cas, comme pour les cuves cylindriques, le modèle prévoit que la cuve est instable dès que $S > 0$.

Essayons à présent de transposer ces résultats en géométrie "réaliste". Le temps caractéristique est fixé à $\tau \simeq 6\,s$ de sorte que $c = 1$. Les dimensions de la cuve sont $L_x = 3\,m$, $L_y = 10\,m$ et le champ magnétique vertical est uniforme : $B_{0,z} \equiv 1$. Dans ce cas, $q_{m,n} = b_m\, q_n$, where $q_m = \int_0^\pi \sin(mx)\,dx = \begin{cases} 0 & \text{si } m \text{ est pair} \\ \frac{2}{m} & \text{si } m \text{ est impair} \end{cases}$. Les valeurs propres de $K - SG$ peuvent être calculées numériquement, comme une fonction de S. Nous nous retreignons aux modes (m, n) tels que $0 \leq m, n \leq 3$ (et $(m, n) \neq (0, 0)$). Les résultats sont exposés FIG. 4.4. Quand S augmente, la première valeur propre avec une partie imaginaire apparaît pour $S \in (0.244, 0.245)$, de la collision des modes $(0, 3)$ et $(1, 0)$. Alors, ce mode instable se "restabilise" pour $S \in (0.502, 0.503)$. Le système est alors stable pour $S \in (0.503, 0.514)$. Finalement, le système devient définitivement instable pour S plus grand qu'une valeur critique dans l'intervalle $(0.514, 0.515)$.

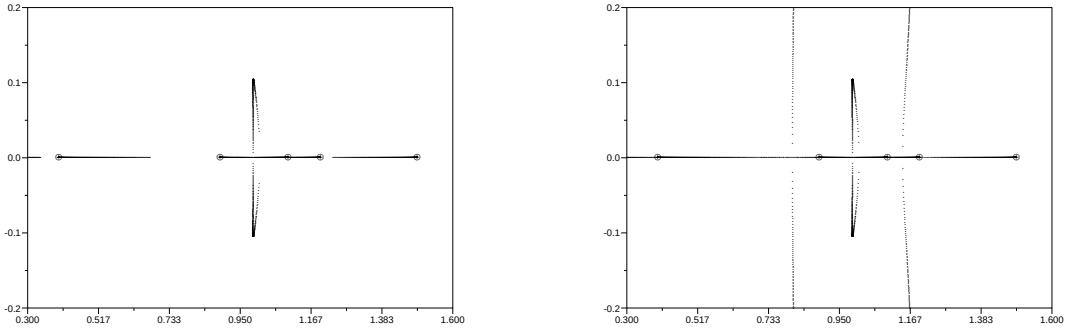


FIG. 4.4 – Trajectoires des valeurs propres de $K - SG$ évoluant dans le plan complexe quand S augmente (entre deux points, l'incrément sur S est de 0.001) : $S \in [0, 0.5]$ en haut, et $S \in [0, 0.59]$ en bas. Les positions initiales des valeurs propres quand $S = 0$ sont indiquées par des cercles.

Cette expérience illustre le fait surprenant que le système peut être instable pour certaines valeurs de S, et stable pour de valeurs plus grandes de S : des modes instables peuvent se "restabiliser". On voit aussi que le S critique au-delà duquel de véritables modes instables apparaissent est plutôt faible comparé aux valeurs réalistes (S varie typiquement entre 6 et 340 pour les cuves industrielles). Cela illustre le fait qu'il est nécessaire de déterminer un seuil de stabilité pour les parties imaginaires des valeurs propres, afin d'obtenir une valeur réaliste pour S_{critique}.

Nous voudrions également mentionner qu'il est possible d'obtenir une valeur plus grande de S_{critique} en introduisant un terme de friction dans le modèle linéaire (3.16), pour modéliser la dissipation mécanique (pour relaxer l'hypothèse H_3 p.38). Dans [12], il est démontré qu'introduire une loi de friction comme dans [74] revient à remplacer l'équation sur η dans (3.16) par :

$$\frac{\partial^2 \eta}{\partial t^2} + \gamma \frac{\partial \eta}{\partial t} - c^2 \Delta_H \eta = c^2 \left(\frac{\partial \varphi}{\partial y} \frac{\partial B_{0,z}}{\partial x} - \frac{\partial \varphi}{\partial x} \frac{\partial B_{0,z}}{\partial y} \right), \tag{4.15}$$

où

$$\gamma = \tau \frac{\frac{\rho_1 \gamma_1}{h_1} + \frac{\rho_2 \gamma_2}{h_2}}{\frac{\rho_1}{h_1} + \frac{\rho_2}{h_2}}$$

est un coefficient de friction adimensionné, et γ_1 et γ_2 sont les coefficients de friction respectivement dans l'aluminium et dans l'électrolyte. Des valeurs typiques pour les coefficients de friction sont $0.01\, s^{-1} \leq \gamma_1 = \gamma_2 \leq 0.1\, s^{-1}$ (voir le tableau 1 dans [100] où ces coefficients sont étalonnés en comparant des vitesses calculées avec des expériences, ou [75]), ainsi $\gamma \in (0.06, 0.6)$. Si (λ, η) est un mode propre de $K - SG$, une solution du modèle linéaire (3.16) avec un terme d'amortissement tel que dans (4.15) est donnée par $\exp(-i\omega t)\eta$, avec $-\omega^2 - i\gamma\omega + \lambda = 0$. Comme[1] $\mathrm{Re}(\lambda) > 0$, on peut vérifier que pour $\gamma \geq |\mathrm{Im}(\lambda)|/\sqrt{\mathrm{Re}(\lambda)}$, $\mathrm{Im}(\omega) \leq 0$ et donc le système est stable. Le coefficient d'amortissement γ est ainsi relié au seuil de stabilité (que nous avons mentionné ci-dessus) pour les parties imaginaires des valeurs propres du modèle linéaire (3.16).

Dans notre exemple numérique, on observe que pour $\gamma = 0.06$, quand β augmente, le premier mode non linéaire apparaît pour $S \in (0.289, 0.290)$, puis le système se "restabilise" pour $S \in (0.492, 0.493)$, et finalement devient définitivement instable pour $S \in (0.521, 0.522)$. Pour $\gamma = 0.6$, le système devient definitivement instable pour S plus grand qu'une valeur critique dans l'intervalle $(0.800, 0.801)$. Ce sont encore des valeurs très pessimistes de S_{critique}, comparé aux valeurs typiques de S dans les cuves réelles : $6 \leq S \leq 340$. Inversement, pour $S = 6$, le système est stable pour γ plus grand que 6.1. Pour $S = 340$, le système est stable pour γ plus grand que 295. Ces valeurs de γ correspondent à des valeurs non réalistes de coefficients de friction γ_1 et γ_2. En conclusion, le modèle linéaire semble trop pessimiste par rapport à la stabilité des cuves industrielles.

4.3.2 Modèle non linéaire

On considère à présent le même problème modélisé par l'approche non linéaire. On utilise le protocole classique permettant d'obtenir le phénomène de rolling (voir **3.3**) : un courant vertical traverse la cellule et on ajoute une composante verticale au champ magnétique. Dans le modèle non linéaire, cela est imposé au travers du champ magnétique B_0 utilisé[2] pour définir les conditions aux limites sur le champ magnétique (voir [93] pour des conditions aux limites similaires sur une cuve parallélépipédique) :

$$B_{0,x} = -\frac{\mu_0 j_0 y}{2}, \qquad B_{0,y} = \frac{\mu_0 j_0 x}{2}, \qquad B_{0,z} \text{ uniforme}, \tag{4.16}$$

avec $j_0 \approx 10\, kA/m^2$. Nous décomposons l'évolution temporelle de l'altitude d'un point de l'interface sous la forme $z(t) = z_0 + \sum_{i=1}^{N} \alpha_i \exp(\tau_i t) \cos(2\pi t/T_i - \phi_i)$, où N représente le nombre de modes, τ_i le facteur de croissance de chaque mode, et T_i la période de chaque mode.

[1] On désigne par $\mathrm{Re}(\lambda)$ (resp. $\mathrm{Im}(\lambda)$) la partie réelle (resp. imaginaire) d'un nombre complexe λ.

[2] Dans l'approche linéaire, la linéarisation a été effectuée autour de ce champ magnétique stationnaire B_0.

Un signal typique obtenu pour $B_{0,z} = 280\,G$ (où G désigne $Gauss$: $1G = 10^{-4}\,T$) est le suivant :

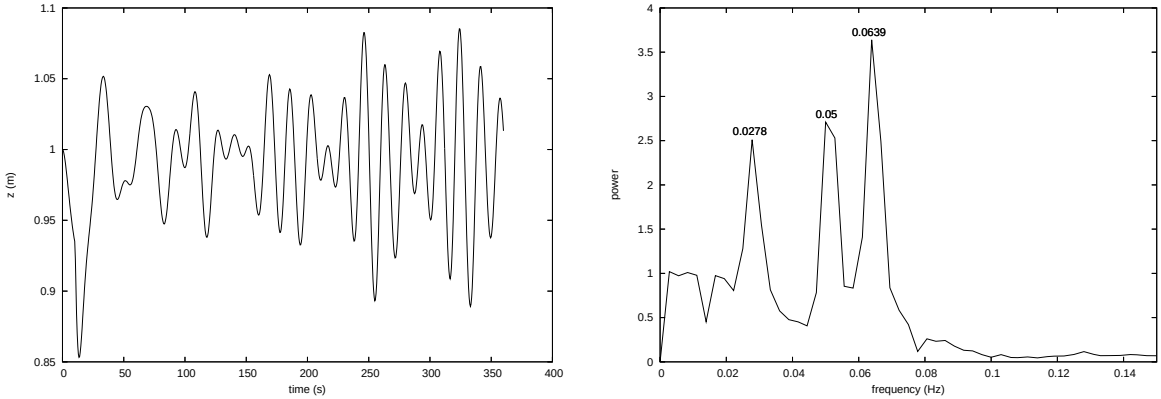


FIG. 4.5 – Altitude d'un point de l'interface en fonction du temps et transformée de Fourier associée, pour $B_{0,z} = 280\,G$. Cette expérience numérique est effectuée en utilisant le modèle non linéaire, sur une cuve parallélépipédique avec $L_x = L_y = 2m$ et $h_1 = h_2 = 1m$.

Pour de faibles valeurs de $B_{0,z}$ ($B_{0,z} < 200\,G$), un unique mode de Fourier ($N = 1$) est suffisant pour décrire le signal. Le tableau ci-dessous montre le taux d'accroissement τ_1 et la période T_1 de ce mode, pour différentes valeurs de la composante verticale du champ magnétique :

$B_{0,z}\,(G)$	0	30	50	65	80	95	130	160	180
$\tau_1(s^{-1})$	-0.0335	-0.0191	-0.0137	-0.0080	-0.0032	0.0000	0.0049	0.0092	0.0097
$T_1(s)$	29.4	29.3	29.2	29.1	29.3	29.3	30.0	29.8	30.0

TAB. 5.1 - Taux d'accroissement (τ_1) et période (T_1) du principal mode propre pour divers champs magnétiques verticaux $B_{0,z}$. Les expériences numériques sont effectuées sur le modèle non linéaire, sur une cuve parallélépipédique telle que $L_x = L_y = 2m$ et $h_1 = h_2 = 1m$.

Nous observons que la période du mode propre dépend peu de $B_{0,z}$, mais l'observation la plus importante est qu'il existe un seuil critique de stabilité sur $B_{0,z}$. Pour $B_{0,z} < 95\,G$, τ_1 est négatif et la cuve est alors stable. Pour $B_{0,z} > 95\,G$, τ_1 est positif et la cuve est instable. De même, sur cuve cylindrique, le phénomène est stable pour de faibles valeurs de B_z :

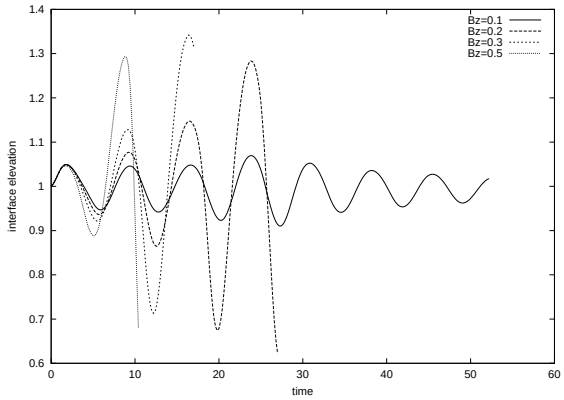


FIG. 4.6 – Influence du champ magnétique vertical sur la stabilité du phénomène de rolling sur cuve cylindrique (cf. [39])

Cela contredit l'analyse spectrale des systèmes linéaires simplifiés qui prévoit qu'une cuve est instable si $L_x = L_y$ quelle que soit $B_{0,z}$ (voir la section précédente). Par ailleurs, des expériences numériques (non reproduites ici) ont montré que le seuil ci-dessus n'est pas significativement sensible aux variations du facteur d'amplification artificiel appliqué à la viscosité (cf. **3.3**).

D'un autre côté, pour de grandes valeurs de $B_{0,z}$, trois modes de Fourier ($N = 3$) sont nécessaires pour décrire le signal (voir FIG. 4.5 pour le cas $B_{0,z} = 280\,G$). Nous observons que quand $B_{0,z}$ augmente, le premier mode (celui qui a la plus grande période) redevient stable (τ_1 redevient négatif) et l'instabilité de la cuve est due aux premier et troisième modes :

$B_{0,z}$	τ_1	T_1	τ_2	T_2	τ_3	T_3
$250G$	-0.004	35	0.0051	20	0.0085	15
$280G$	-0.01	35	0.0045	20	0.0092	15

TAB. 5.2 - Taux d'accroissement (τ_i) et période (T_i) des trois principaux modes propres pour deux valeurs du champ magnétique vertical. Les expériences numériques sont effectuées avec le modèle non linéaire, sur une cuve parallélépipédique telle que $L_x = L_y = 2m$ et $h_1 = h_2 = 1m$.

De façon intéressante, la "restabilisation" d'un mode instable lorsqu'on augmente le champ magnétique vertical est aussi observée sur le système linéaire simplifié (voir FIG. 4.4).

4.4 Discussion

Ces expériences numériques montrent que l'analyse de stabilité linéaire sur des systèmes simplifiés et des expériences numériques sur un modèle non linéaire plus complet se complètent avantageusement. Nous avons observé une bonne cohérence entre les deux modèles à plusieurs points de vue : fréquences gravitationnelles, comportement des fréquences quand la composante verticale du champ magnétique augmente.

La différence qualitative la plus importante que nous avons mentionnée est que le modèle linéaire prévoit que les cuves cylindriques et carrées sont instables dès lors que la composante veticale du champ magnétique est non nulle. Cela contredit les résultats obtenus par le modèle non linéaire. Cette différence peut être corrigée en ajoutant un terme d'amortissement dans le modèle linéaire, pour modéliser les effets de dissipation qui ont été négligés, comme expliqué à la fin de la section **4.3.1**. Cependant, ce n'est pas satisfaisant dans la mesure où des valeurs non réalistes du terme d'amortissement sont requises pour ajuster le modèle aux observations, et où le seuil de stabilité est très sensible à la valeur de ce paramètre artificiel.

D'autres comparaisons des résultats fournis par les deux modèles sont en cours. Nous voudrions en particulier comprendre si les différences qualitatives obervées proviennent de la procédure de linéarisation ou des approximations H_1-H_6 faites sur les équations (cf. **3.2**).

Les conclusions sur le modèle linéarisé laissent penser que certains comportements qualitatifs peuvent être appréhendés dans ce contexte. Aussi nous proposons à présent de contrôler l'évolution de l'interface au moyen de deux commandes à l'aide de ce modèle, à savoir la distance anode-métal h_2 et le champ magnétique vertical B_z.

Chapitre 5

Contrôle de l'évolution de l'interface en eaux peu profondes

5.1 Présentation du problème

Dans le cadre de l'approximation des eaux peu profondes (voir **3** et **4**), le modèle des cuves parallélépipédiques est défini sur un rectangle $\Omega_H = [0, L_x] \times [0, L_y]$. Nous cherchons à stabiliser l'interface électrolyte-aluminium en minimisant, pendant une durée fixée T, une certaine énergie $\mathcal{E}(\eta)$, où $\eta(t, x, y)$ est une perturbation sur la hauteur d'interface moyenne. Pour cela, nous utilisons un actionneur $u(t, x, y)$, ce qui représente un certain coût qu'on quantifie par la fonction $\mathcal{P}(u)$. Ainsi, le problème est de trouver le minimum de la fonctionnelle

$$\mathcal{J}(u) = \mathcal{E}(\eta(u)) + \mathcal{P}(u)\,,$$

où η est lié à u par les équations d'état. Plus précisément, on écrit les fonctions $\mathcal{E}$ et $\mathcal{P}$:

$$\mathcal{E}(\eta) = \frac{\sigma_0}{2}|\eta|^2_{L^2([0,T],\Omega_H)} + \frac{\sigma_1}{2}|\partial_t\eta|^2_{L^2([0,T],\Omega_H)}\,, \qquad \mathcal{P}(u) = \frac{Q}{2}\|u\|^2\,, \quad \sigma_i > 0 \quad \forall\, i\,, \quad Q > 0\,,$$

où U est un espace de Hilbert contenant l'ensemble des commandes admissibles (cf. **2.3.2.a**). Ces travaux ont fait l'objet d'une communication orale [79] au Congrès National d'Analyse numérique (CANUM) 2006.

5.1.1 Équations d'état

On rappelle (cf. **3.1**) le modèle régissant l'évolution des inconnues η et φ en présence d'un champ magnétique vertical $B_z(t, x, y)$, et pour une distance anode-métal $h_2(t)$:

$$\left\{\begin{array}{rcccc} \partial_t^2\,\eta - c^2\,(\Delta_H\,\eta + \partial_y\varphi\,\partial_x B_z - \partial_x\varphi\,\partial_y B_z) & = & 0 & \text{dans} & \Omega_H \times [0,T] \\ -\Delta_H\,\varphi - S\,\eta & = & 0 & \text{dans} & \Omega_H \times [0,T] \\ \partial_n\,\eta + B_z\,(\partial_y\,\varphi\,n_x - \partial_x\,\varphi\,n_y) & = & 0 & \text{sur} & \partial\Omega_H \times [0,T] \\ \partial_n\,\varphi & = & 0 & \text{sur} & \partial\Omega_H \times [0,T] \\ \eta_{|t=0} & = & \eta_0 & \text{dans} & \Omega_H \\ (\partial_t\,\eta)_{|t=0} & = & \eta_1 & \text{dans} & \Omega_H \end{array}\right. \tag{5.1}$$

$$\text{avec} \quad c^2 = \frac{(\rho_1 - \rho_2)\,g}{\dfrac{\rho_1}{h_1} + \dfrac{\rho_2}{h_2}}\,\frac{\tau^2}{\mathcal{L}^2} \quad \text{et} \quad S = \frac{J_0\,B_0\,L^2}{h_1\,h_2\,(\rho_1 - \rho_2)\,g} \quad \text{(constante de Sele)}\,,$$

où $\mathcal{L}$ et τ sont la longueur et le temps caractéristiques, et $\Omega_H = [0, L_x] \times [0, L_y]$.

Ainsi, nous sommes en présence d'un système hyperbolique du deuxième ordre homogène triplement couplé, à savoir dans chacune des deux équations volumiques et dans la condition aux limites portant sur $\partial_n \eta$. Des données usuelles pour ce problème sont une distance anode-métal moyenne constante ($h_2(t) = h_2^0$) et un champ magnétique vertical constant et uniforme ($B_z(t,x,y) = \overline{B}_z$). Le comportement observé dans cette configuration est alors une déstabilisation assez rapide de la cuve - typiquement des fluctuations importantes de l'interface au bout d'une unité de temps - pour un choix des paramètres conforme à la réalité physique (voir les tests en **5.4**).

5.1.2 Commandes et fonctions coût

On peut envisager deux manières d'atténuer le phénomène de déstabilisation décrit ci-dessus, à savoir une approche de type "régulation", où l'on donne la possibilité à un technicien (ou un automate) de modifier en temps réel un paramètre du problème (typiquement la distance anode-métal moyenne) ; et une approche de type "design de cuve", où l'on cherche cette fois à modifier une caractéristique intrinsèque de la cuve, comme le champ magnétique vertical (stationnaire) généré par son environnement en usine. Enfin, bien que ce soit un peu moins naturel, on peut considérer que le champ magnétique vertical soit lui aussi modifiable en temps réel. Dans les deux cas, on remarquera que la commande apparaît, d'une part, dans le coefficient des inconnues, et, d'autre part, dans plusieurs du système.

5.1.2.a Contrôle par la distance anode-métal moyenne : $h_2(t)$

On cherche la hauteur optimale $u(t) = h_2(t)$ à donner au plan anodique au cours du temps, de manière à obtenir le système le plus stable possible sur une certaine durée T. La minimisation au sens des moindres carrés de la perturbation η semble être un critère adapté à cet objectif :

$$\mathcal{J}_1(h_2) = \mathcal{E}(\eta(h_2)) \; + \; \frac{Q_1}{2} |h_2|^2_{L^2([0,T],\mathbb{R}^+)} \, . \tag{5.2}$$

Le coût du contrôle est directement représenté par la DAM moyenne, sachant qu'un demi-centimètre de cette dernière se chiffre à environ $\$10^8$ annuels... Le cas particulier h_2 constant

$$\mathcal{J}_1(h_2) = \mathcal{E}(\eta(h_2)) \; + \; \frac{Q_1}{2} \, T \, h_2^2 \, ,$$

qui permet de travailler avec une commande unidimensionnelle, est un bon moyen de tester la validité du programme, dont des parties importantes (équation d'état et problème adjoint) varient peu d'un type de commande à l'autre.

5.1.2.b Contrôle par le champ magnétique vertical : $B_z(t,x,y)$

Seule origine des instabilités observées sur ce type de modèle, le champ magnétique vertical est imposé par l'environnement du système. Pour une DAM fixée à h_2^0, on prend pour commande une perturbation $u(t,x,y) = b_z(t,x,y) = B_z(t,x,y) - \overline{B}_z$ à moyenne spatiale nulle pour tout $t \in [0,T]$, minimisant la fonctionnelle :

$$\mathcal{J}_2(b_z) = \mathcal{E}(\eta(b_z)) \; + \; \frac{Q_2}{2} |b_z|^2_{L^2([0,T],\Omega_H)} \, . \tag{5.3}$$

De même que pour la commande h_2, nous étudierons le cas indépendant du temps

$$\mathcal{J}_2(b_z) = \mathcal{E}(\eta(b_z)) \; + \; \frac{Q_2}{2} \, T \, |b_z|^2_{L^2(\Omega_H)} \, .$$

Notons que contrairement à h_2 qui est un scalaire, la commande b_z est distribuée sur l'ensemble du domaine.

5.2 Expression des gradients des fonctions coût

Nous adoptons la démarche exposée en **2.3** pour le contrôle des systèmes gouvernés par des EDP, à savoir la recherche numérique d'un minimum local de $\mathcal{J}(u)$ à l'aide de $\nabla\mathcal{J}(u)$, en supposant ces deux données numériquement calculables pour toute commande admissible u. Alors, il suffit d'utiliser une des nombreuses librairies d'optimisation en dimension finie comprenant une routine de type gradient ou mieux (cf. **2.3.5.a**). La fonction `optim` de Scilab [89] fournit ce genre d'outils.

5.2.1 Problème adjoint

5.2.1.a Formulation faible

Soient trois espaces de Hilbert U, V et W, de sorte que $(\eta,\varphi) \in V^2$, et que les équations d'état s'écrivent sous la forme $\mathcal{F}(\eta,\varphi,u)=0$, où $\mathcal{F}$ est défini de $V^2\times U$ dans W. Ainsi la dérivée directionnelle (cf. (2.36)) de $\mathcal{J}(u)=\mathcal{J}(\eta(u),u)$ dans toute direction $\tilde{u}\in U$ s'écrit :

$$\Big(\nabla\mathcal{J}(u),\tilde{u}\Big)_U \;=\; \int_0^T\int_{\Omega_H}(\sigma_0\,\eta\,\tilde{\eta}+\sigma_1\,\partial_t\eta\,\partial_t\tilde{\eta}) \;+\; Q\,(u\,,\tilde{u})_U\,, \tag{5.4}$$

où $\tilde{\eta}$ est lié à $\tilde{u}$ par les équations de sensitivité (cf. **2.3.5.a**) :

$$\lim_{\varepsilon\to 0}\frac{\mathcal{F}(\eta+\varepsilon\tilde{\eta},\varphi+\varepsilon\tilde{\varphi},u+\varepsilon\tilde{u})-\mathcal{F}(\eta,\varphi,u)}{\varepsilon}=0\,. \tag{5.5}$$

Alors, l'utilisation de la solution du problème adjoint :

$$\left\{\begin{array}{l} \text{Trouver}\quad \Lambda=(\alpha,\beta,\gamma,\delta,\lambda,\mu)\;\in W'\quad\text{tel que pour tout}\quad(\tilde{\eta},\tilde{\varphi})\;\in V^2\quad: \\[2ex] \displaystyle\int_0^T\int_{\Omega_H}[\alpha\,(\partial_t^2\tilde{\eta}-c^2\,\Delta_H\tilde{\eta})-\beta\,S\,\tilde{\eta}]+\int_0^T\int_{\partial\Omega_H}\gamma\,\partial_n\tilde{\eta}+\int_{\Omega_H}(\lambda\,\tilde{\eta}_{|t=0}+\mu\,\partial_t\tilde{\eta}_{|t=0}) \\ \displaystyle=\int_0^T\int_{\Omega_H}(\sigma_0\,\eta\,\tilde{\eta}+\sigma_1\,\partial_t\eta\,\partial_t\tilde{\eta})\,, \\[2ex] \displaystyle\int_0^T\int_{\Omega_H}[\alpha\,c^2(\partial_y\tilde{\varphi}\,\partial_x B_z-\partial_x\tilde{\varphi}\,\partial_y B_z)+\beta\,\Delta_H\tilde{\varphi}]-\int_0^T\int_{\partial\Omega_H}[\gamma\,B_z\,(\partial_y\tilde{\varphi}\,n_x-\partial_x\tilde{\varphi}\,n_y)+\delta\,\partial_n\tilde{\varphi}] \\ =0\,, \end{array}\right. \tag{5.6}$$

permet de remplacer dans (5.4) la variable difficilement calculable $\tilde{\eta}$ par une expression ne dépendant que de Λ et u. Cette opération est spécifique à chaque problème (voir **5.2**).

5.2.1.b Formulation forte

En intégrant par parties les équations (5.6), on obtient la formulation forte en (α,β) :

$$\left\{\begin{array}{ccccc} \partial_t^2\,\alpha-c^2\Delta_H\alpha-S\beta &=& \sigma_0\,\eta-\sigma_1\,\partial_t^2\eta & \text{dans} & \Omega_H\times[0,T]\,, \\ -\Delta_H\beta+c^2\,(\partial_y\alpha\,\partial_x B_z-\partial_x\alpha\,\partial_y B_z) &=& 0 & \text{dans} & \Omega_H\times[0,T]\,, \\ \partial_n\,\alpha &=& 0 & \text{sur} & \partial\Omega_H\times[0,T]\,, \\ \partial_n\beta-c^2\,B_z\,(\partial_y\,\alpha\,n_x-\partial_x\,\alpha\,n_y) &=& 0 & \text{sur} & \partial\Omega_H\times[0,T]\,, \\ \alpha_{|t=T} &=& 0 & \text{dans} & \Omega_H\,, \\ (\partial_t\,\alpha)_{|t=T} &=& -\sigma_1\,\partial_t\eta_{|t=T} & \text{dans} & \Omega_H\,. \end{array}\right. \tag{5.7}$$

(les autres solutions s'obtiennent facilement à partir de α et β : $\gamma=c^2\alpha_{|\partial\Omega_H}$, $\delta=\beta_{|\partial\Omega_H}$, $\lambda=-\eta_1-(\partial_t\alpha)_{|t=0}$, $\mu=\alpha_{|t=0}$).

5.2.2 Commande $h_2(t)$

Exprimons le gradient du critère sous la forme (5.4) :

$$\int_0^T \nabla \mathcal{J}_1(h_2)\,\tilde{h}_2 = \int_0^T \int_{\Omega_H} (\sigma_0\,\eta\,\tilde{\eta} + \sigma_1\,\partial_t\eta\,\partial_t\tilde{\eta}) \;+\; Q_1 \int_0^T h_2\,\tilde{h}_2\,, \tag{5.8}$$

et explicitons les équations de sensitivité (5.5) relatives à la commande h_2 :

$$\left\{\begin{array}{rcccc} \partial_t^2\tilde{\eta} - c^2(h_2)\,(\Delta_H\tilde{\eta} + \partial_y\tilde{\varphi}\,\partial_x B_z - \partial_x\tilde{\varphi}\,\partial_y B_z) & = & \dfrac{k(h_2)}{h_2}\,\partial_t^2\eta\,\tilde{h}_2 & \text{dans} & \Omega_H\times[0,T] \\ -\Delta_H\tilde{\varphi} - S(h_2)\,\tilde{\eta} & = & -\dfrac{S(h_2)}{h_2}\,\eta\,\tilde{h}_2 & \text{dans} & \Omega_H\times[0,T] \\ \partial_n\tilde{\eta} + B_z(\partial_y\tilde{\varphi}\,n_x - \partial_x\tilde{\varphi}\,n_y) & = & 0 & \text{sur} & \partial\Omega_H\times[0,T] \\ \partial_n\tilde{\varphi} & = & 0 & \text{sur} & \partial\Omega_H\times[0,T] \\ \tilde{\eta}_{|t=0} & = & 0 & \text{dans} & \Omega_H \\ (\partial_t\tilde{\eta})_{|t=0} & = & 0 & \text{dans} & \Omega_H \end{array}\right.,$$

où l'on a utilisé les expressions suivantes des dérivées c^2 et S :

$$[c^2]'(h_2) = \frac{k(h_2)}{h_2}\,c^2(h_2) \quad \text{avec} \quad k(h_2) = \frac{\rho_2 h_1}{\rho_2 h_1 + \rho_1 h_2}\,, \quad \text{et} \quad S'(h_2) = -\frac{1}{h_2}\,S(h_2)\,.$$

En multipliant à présent chacune de ces équations respectivement par les solutions α, β, γ, δ, λ, μ du problème adjoint, et en les intégrant sur leur domaine de définition, on obtient par l'intermédiaire de (5.6) :

$$\int_0^T \int_{\Omega_H} (\sigma_0\,\eta\,\tilde{\eta} + \sigma_1\,\partial_t\eta\,\partial_t\tilde{\eta}) = \int_0^T \int_{\Omega_H} \left[k\,\partial_t^2\eta\,\alpha - S\,\eta\,\beta\right]\frac{\tilde{h}_2}{h_2}\,,$$

et donc finalement

$$\nabla \mathcal{J}_1(h_2) = \frac{1}{h_2}\int_{\Omega_H}\left[\frac{\rho_2 h_1}{\rho_2 h_1 + \rho_1 h_2}\,\alpha\,\partial_t^2\eta - S\,\beta\,\eta\right] \;+\; Q_1\,h_2\,. \tag{5.9}$$

Dans le cas h_2 indépendant du temps, le gradient du critère prend la forme

$$\nabla \mathcal{J}_1(h_2) = \frac{1}{h_2}\int_0^T\int_{\Omega_H}\left[\frac{\rho_2 h_1}{\rho_2 h_1 + \rho_1 h_2}\,\alpha\,\partial_t^2\eta - S\,\beta\,\eta\right] \;+\; Q_1\,T\,h_2\,.$$

5.2.3 Commande $B_z(t,x,y)$

Par la même démarche, on établit dans un premier temps :

$$\int_0^T\int_{\Omega_H} \nabla \mathcal{J}_2(b_z)\,\tilde{b}_z = \int_0^T\int_{\Omega_H} (\sigma_0\,\eta\,\tilde{\eta} + \sigma_1\,\partial_t\eta\,\partial_t\tilde{\eta}) \;+\; Q_2\int_0^T\int_{\Omega_H} b_z\,\tilde{b}_z\,,$$

puis on exprime la dépendance de $\tilde{\eta}$ en $\tilde{b}_z$ par les équations de sensitivité :

$$\left\{\begin{array}{rcccc} \partial_t^2\tilde{\eta} - c^2\,(\Delta_H\tilde{\eta} + \partial_y\tilde{\varphi}\,\partial_x B_z - \partial_x\tilde{\varphi}\,\partial_y B_z) & = & c^2\,(\partial_y\varphi\,\partial_x\tilde{b}_z - \partial_x\varphi\,\partial_y\tilde{b}_z) & \text{dans} & \Omega_H\times[0,T] \\ -\Delta_H\tilde{\varphi} - S\,\tilde{\eta} & = & 0 & \text{dans} & \Omega_H\times[0,T] \\ \partial_n\tilde{\eta} + B_z(\partial_y\tilde{\varphi}\,n_x - \partial_x\tilde{\varphi}\,n_y) & = & -\tilde{b}_z\,(\partial_y\varphi\,n_x - \partial_x\varphi\,n_y) & \text{sur} & \partial\Omega_H\times[0,T] \\ \partial_n\tilde{\varphi} & = & 0 & \text{sur} & \partial\Omega_H\times[0,T] \\ \tilde{\eta}_{|t=0} & = & 0 & \text{dans} & \Omega_H \\ (\partial_t\tilde{\eta})_{|t=0} & = & 0 & \text{dans} & \Omega_H \end{array}\right.,$$

Comme ci-dessus, on multiplie chaque équation par la solution du problème adjoint associée, et on intègre. Alors, une intégration par parties des dérivées de $\tilde{b}_z$ dans la première équation permet d'aboutir à l'égalité (toujours compte tenu de (5.6)) :

$$\int_0^T \int_{\Omega_H} (\sigma_0\, \eta\, \tilde{\eta} + \sigma_1\, \partial_t\eta\, \partial_t\tilde{\eta}) = \int_0^T \int_{\Omega_H} c^2\, (\partial_y\alpha\, \partial_x\varphi - \partial_x\alpha\partial_y\varphi)\, \tilde{b}_z$$

(sachant que $\gamma = c^2\alpha_{|\partial\Omega_H}$). Ainsi

$$\nabla\mathcal{J}_2(b_z) = c^2\, (\partial_y\alpha\, \partial_x\varphi - \partial_x\alpha\partial_y\varphi)\ +\ Q_2\, b_z\,. \tag{5.10}$$

Dans le cas B_z indépendant du temps, on trouve le gradient :

$$\nabla\mathcal{J}_2(b_z) = \int_0^T c^2\, (\partial_y\alpha\, \partial_x\varphi - \partial_x\alpha\partial_y\varphi)\ +\ Q_2\, T\, b_z\,.$$

5.3 Discrétisation

Rappelons la formulation variationnelle en espace (5.11) de l'équation d'état, qui va nous permettre d'approcher la solution par une méthode de Galerkin de type spectral : $\forall\, t \in [0,T]$,

$$\left\{\begin{array}{l} \text{Trouver}\quad (\eta(t),\varphi(t))\ \in\ (H^1(\Omega_H))^2\quad \text{tels que}\quad \forall\,(\zeta,\psi)\ \in\ (H^1(\Omega_H))^2\quad : \\ \displaystyle\int_{\Omega_H}\Big[\partial_t^2\eta(t)\,\zeta + c^2\,[\nabla_H\eta(t)\cdot\nabla_H\zeta + B_z\,(\partial_y\varphi(t)\,\partial_x\zeta - \partial_x\varphi(t)\,\partial_y\zeta)]\Big] \quad = \quad 0\,, \\ \displaystyle\int_{\Omega_H}\Big[\nabla_H\varphi(t)\cdot\nabla_H\psi - S\,\eta(t)\,\psi\Big] \quad = \quad 0\,. \end{array}\right. \tag{5.11}$$

plus les conditions initiales $\eta_{|t=0} = \eta_0$ et $(\partial_t\eta)_{|t=0} = \eta_1$. Le problème adjoint (5.7) s'écrit sous forme variationnelle en espace $\forall\, t \in [0,T]$:

$$\left\{\begin{array}{l} \text{Trouver}\quad (\alpha(t),\beta(t))\ \in\ (H^1(\Omega_H))^2\quad \text{tels que}\quad \forall\,(\zeta,\psi)\in\quad (H^1(\Omega_H))^2\quad : \\ \displaystyle\int_{\Omega_H}\Big[\partial_t^2\alpha(t)\,\zeta + c^2\,\nabla_H\alpha(t)\cdot\nabla_H\zeta - S\,\beta(t)\,\zeta\Big] \quad = \quad \int_{\Omega_H}\Big[\sigma_0\,\eta(t) - \sigma_1\,\partial_t^2\eta(t)\Big]\,\zeta\,, \\ \displaystyle\int_{\Omega_H}\Big[\nabla_H\beta(t)\cdot\nabla_H\psi - c^2\,B_z\,[\partial_y\alpha(t)\,\partial_x\psi - \partial_x\alpha(t)\,\partial_y\psi]\Big] \quad = \quad 0\,. \end{array}\right. \tag{5.12}$$

plus les conditions initiales $\alpha_{|t=T} = 0$ et $(\partial_t\alpha)_{|t=T} = -\sigma_1(\partial_t\eta)_{|t=T}$.

5.3.1 Approximation de Galerkin sur la base des modes gravitationnels

Les propriétés spectrales du laplacien permettent de chercher toute solution du système (5.11)-(5.12) sur la base des "modes gravitationnels" :

$$\left\{\begin{array}{l} \eta(t,x,y) = \displaystyle\sum_{(m,n)\in\mathbb{N}^2} \eta_{m,n}(t)\, f_{m,n}(x,y) \\ \varphi(t,x,y) = \displaystyle\sum_{(m,n)\in\mathbb{N}^2} \varphi_{m,n}(t)\, f_{m,n}(x,y) \end{array}\right. , \quad \left\{\begin{array}{l} \alpha(t,x,y) = \displaystyle\sum_{(m,n)\in\mathbb{N}^2} \alpha_{m,n}(t)\, f_{m,n}(x,y) \\ \beta(t,x,y) = \displaystyle\sum_{(m,n)\in\mathbb{N}^2} \beta_{m,n}(t)\, f_{m,n}(x,y) \end{array}\right. ,$$

où $(f_{m,n}(x,y))$ est une base orthonormale dans $L^2(\Omega_H)$ et orthogonale dans $H^1(\Omega_H)$ (cf. **3.3.2**). De la même manière que pour le problème direct, la projection sur cette base de la deuxième

équation du système (5.12) permet de trouver

$$\beta_{m,n} = -\frac{c^2}{k_{m,n}^2} \sum_{(m',n')\in\mathbb{N}^2\setminus(0,0)} \alpha_{m',n'}\, B_{(m,n),(m',n')}\,, \tag{5.13}$$

où

$$B_{(m,n),(m',n')} = \int_{\Omega_H} B_z(\partial_x f_{m',n'}\partial_y f_{m,n} - \partial_y f_{m',n'}\partial_x f_{m,n})\,.$$

Alors on substitue β dans la première, pour arriver à :

$$d_t^2\alpha_{m,n} + c^2\left[k_{m,n}^2\,\alpha_{m,n} + \frac{S}{k_{m,n}^2}\sum_{(m',n')\in\mathbb{N}^2\setminus(0,0)} \alpha_{m',n'}\, B_{(m,n),(m',n')}\right] = \sigma_0\,\eta_{m,n} - \sigma_1\,\partial_t^2\eta_{m,n}$$

$\forall\,(m,n)\in\mathbb{N}^2\setminus(0,0)$. En effectuant le changement de variable $\chi_{m,n} = k_{m,n}\,\alpha_{m,n}$, on établit finalement la version semi-discrétisée du système (5.11)-(5.12) dans le sous-espace :

$$V_h = \text{Vect}\,\{f_{m,n}(x,y)\mid 0\le m\le N_X\,,\ 0\le n\le N_Y\,,\ \text{et}\ (m,n)\neq(0,0)\}\ \subset\ H_0^1(\Omega)\ :$$

$$\left\{\begin{array}{l}\text{Trouver}\quad \xi(t) = [\xi_{0,1}(t),...,\xi_{1,0}(t),...,\xi_{N_X,N_Y}(t)]^T\quad\text{et}\quad \chi(t) = [\chi_{0,1}(t),...,\chi_{1,0}(t),...,\chi_{N_X,N_Y}(t)]^T\\[2ex] \text{tels que}\quad \left\{\begin{array}{c}\xi''(t) + R(t)\,\xi(t) = 0\\ \xi(0) = \xi^0\,,\ \xi'(0) = \xi^1\end{array}\right.\quad\text{et}\quad \left\{\begin{array}{c}\chi\,''(t) + R^T(t)\,\chi(t) = K\,[\sigma_0\,\xi(t) + \sigma_1\,R(t)\,\xi(t)]\\ \chi(T) = 0\,,\ \chi\,'(T) = -\sigma_1\,K\,\xi'(T)\end{array}\right.\end{array}\right.\,, \tag{5.14}$$

où

$$K_{(m,n),(m',n')} = k_{m,n}^2\delta_{(m,n),(m',n')}\quad\text{et}\quad R(t) = c^2(t)\,[K - S(t)\,G(t)]\,,$$

G étant donnée par (4.12). Noter qu'en toute généralité, les matrices R et G dépendent du temps du fait de la présence dans leur définition de c^2 et S (fonctions de h_2 et donc de t) pour la première, et de $B_z(t,x,y)$ pour la deuxième. On renvoie en **5.3.4** pour le calcul de G.

5.3.2 Discrétisation en temps : méthode de Newmark explicite

Le problème (5.14) est composé de deux systèmes différentiels ordinaires du second ordre que l'on peut résoudre chacun par le schéma de Newmark [78] *explicite* ($\{\theta_0 = 0\,,\ \theta_1 = \frac{1}{2}\}$, cf. **2.2.1.a**), étant donné que la matrice multipliant $\xi''(t)$ et $\chi''(t)$ est diagonale[1] (cf. P.-A. Raviart et J.-M. Thomas [84]). Ainsi, en définissant un pas de temps $\Delta t = T/N$ et les échantillonnages associés $\xi_p \simeq \xi(p\,\Delta t)$ et $\chi_p \simeq \chi(p\,\Delta t)$, des développements limités au premier ordre (resp. deuxième) en ξ et χ (resp. ξ' et χ') permettent d'écrire les schémas aux différences finies

$$\left\{\begin{array}{rcl}\xi_{p+1} &=& \xi_p + \Delta t\,\xi_p' + \dfrac{\Delta t^2}{2}\,\xi_p''\\[2ex] \xi_{p+1}' &=& \xi_p' + \Delta t\,\dfrac{\xi_p'' + \xi_{p+1}''}{2}\end{array}\right.\quad\text{et}\quad \left\{\begin{array}{rcl}\chi_{p-1} &=& \chi_p - \Delta t\,\chi_p' + \dfrac{\Delta t^2}{2}\,\chi_p''\\[2ex] \chi_{p-1}' &=& \chi_p' - \Delta t\,\dfrac{\chi_{p-1}'' + \chi_p''}{2}\end{array}\right.\quad,\quad\forall\,p\ge 0\,.$$

Alors, l'expression du système précédent en deux instants p et $p+1$ conduit à

$$\left\{\begin{array}{rcl}\xi_{p+2} - \xi_{p+1} &=& \xi_{p+1} - \xi_p + \Delta t^2\,\xi_{p+1}''\,,\\[2ex] \chi_{p-2} - \chi_{p-1} &=& \chi_{p-1} - \chi_p + \Delta t^2\,\chi_{p-1}''\,.\end{array}\right.$$

[1] Si l'on avait utilisé une méthode d'éléments finis, qui fait apparaître une matrice non diagonale devant ξ'', il aurait suffit d'appliquer la technique de condensation statique pour se ramener au cas diagonal (cf. **2.2.1.b**).

Ainsi, en utilisant (5.14), on écrit les schémas sous la forme :

$$\left\{ \begin{array}{l} \left\{ \begin{array}{l} \xi_{p+2} = 2\xi_{p+1} - \xi_p - \Delta t^2\, R_{p+1}\, \xi_{p+1}\,, \\ \\ \xi_0 = \xi^0\,, \qquad \xi_1 = \xi^0 + \Delta t\, \xi^1 - \dfrac{\Delta t^2}{2}\, R_0\, \xi^0\,, \end{array} \right. \\ \\ \left\{ \begin{array}{l} \chi_{p-2} = 2\chi_{p-1} - \chi_p - \Delta t^2\, [R_{p-1}^T\, \chi_{p-1} - K\, (\sigma_0\, \xi_{p-1} + \sigma_1\, R_{p-1}\xi_{p-1})]\,, \\ \\ \chi_N = 0\,, \qquad \chi_{N-1} = \Delta t\, \sigma_1\, K\, \xi'_N + \dfrac{\Delta t^2}{2}\, K\, (\sigma_0\, \xi_N + \sigma_1\, R_N\, \xi_N)\,, \end{array} \right. \end{array} \right. \tag{5.15}$$

où $R_p \simeq R(p\,\Delta t)$.

5.3.3 Discrétisation des critères et de leurs gradients

Pour calculer les intégrales en temps, on considère que les variables sont interpolées par éléments finis P^0 sur le maillage :

$$\bigcup_{p=0}^{N} \omega_p = \Big[0, \frac{\Delta t}{2}\Big] \cup \left(\bigcup_{p=1}^{N-1} \Big[\Big(p - \frac{1}{2}\Big)\Delta t, \Big(p + \frac{1}{2}\Big)\Delta t\Big] \right) \cup \Big[T - \frac{\Delta t}{2}, T\Big]\,,$$

On obtient ainsi des fonctions en escalier qu'on intègre exactement.

Commande $h_2(t)$

En remarquant, comme $(f_{m,n}(x,y))_{(m,n)}$ forme une base orthonormale de $L^2(\Omega_H)$, que

$$\int_0^T \int_{\Omega_H} \eta^2 = \int_0^T \sum_{(m,n)\in\mathbb{N}^2} \eta_{m,n}^2 \quad \simeq \quad \int_0^T \sum_{(m,n)\in(0,..,N_X)\times(0,..,N_Y)\setminus(0,0)} k_{m,n}^2\, \xi_{m,n}^2 = \int_0^T \xi^T\, K\, \xi\,,$$

on a comme version discrète de l'expression (5.2) :

$$\mathcal{J}_1(h_2) \simeq \frac{1}{2} \sum_{p=0}^{N} |\omega_p|\, (\sigma_0\, \xi_p^T\, K\, \xi_p + \sigma_1\, \xi_p'^T\, K\, \xi'_p) + \frac{Q_1}{2} \sum_{p=0}^{N} |\omega_p|\, (h_2^2)_p\,.$$

$$\text{(noter que pour tout } p \geq 1 \ :\ \xi'_p = \frac{\xi_p - \xi_{p-1}}{\Delta t} - \frac{\Delta t}{2} R_{p-1}\, \xi_{p-1} \quad)\,.$$

Au sujet du gradient (5.9), l'utilisation des formules (3.26) et (5.13) permet d'obtenir :

$$\int_{\Omega_H} \alpha\, \partial_t^2 \eta = \sum_{(m,n)\in\mathbb{N}^2} \alpha_{m,n}\, d_t^2\, \eta_{m,n} \simeq -\chi^T\, R\, \xi\,, \quad \text{et}\,, \quad \int_{\Omega_H} \beta\, \eta \simeq c^2\, \chi^T\, G\, \xi$$

Par ailleurs, on décompose sur la même base que h_2 pour des raisons de consistance :

$$\nabla \mathcal{J}_1(h_2) \simeq \sum_{p=0}^{N} [\nabla \mathcal{J}_1(h_2)]_p\, \mathbb{1}_{\omega_p}\,,$$

où $[\nabla \mathcal{J}_1(h_2)]_p$ est la projection de $\nabla \mathcal{J}_1(h_2)$ sur le vecteur de base $\mathbb{1}_{\omega_p}$:

$$[\nabla \mathcal{J}_1(h_2)]_p = |\omega_p| \left\{ \frac{1}{(h_2)_p} \left[- \frac{\rho_2 h_1}{\rho_2 h_1 + \rho_1 (h_2)_p} \chi_p^T R_p \xi_p - S_p c_p^2 \chi_p^T G \xi_p \right] + Q_1 (h_2)_p \right\}.$$

Dans le cas h_2 indépendant du temps, le critère discrétisé s'écrit :

$$\mathcal{J}_1(h_2) \simeq \frac{1}{2} \sum_{p=0}^{N} |\omega_p| \, (\sigma_0 \, \xi_p^T \, K \, \xi_p + \sigma_1 \, \xi_p'^T \, K \, \xi_p') + \frac{Q_1}{2} T \, h_2^2 \,,$$

et son gradient :

$$\nabla \mathcal{J}_1(h_2) \simeq \frac{1}{h_2} \sum_{p=1}^{N-1} |\omega_p| \left(- \frac{\rho_2 h_1}{\rho_2 h_1 + \rho_1 h_2} \chi_p^T \, R \, \xi_p - S \, c^2 \chi_p^T \, G \, \xi_p \right) + Q_1 \, T \, h_2 \,.$$

Commande $B_z(t,x,y)$

En plus de la discrétisation en temps, on doit ici décomposer l'expression (5.10) également en espace. On utilise la base des modes gravitationnels :

$$b_z(t,x,y) = \sum_{(i,j) \in \mathbb{N}^2 \setminus (0,0)} b_{i,j}(t) \, f_{i,j}(x,y) \,, \quad \text{avec} \quad b_{i,j} = \Big(b_z, f_{i,j} \Big)_{L^2(\Omega_H)} \,,$$

et on restreint l'espace à un sous-espace de dimension $(N_X^{b_z} + 1) \times (N_Y^{b_z} + 1) - 1$. Par ailleurs, on note $b_{i,j,p} = \langle b_{i,j}, \mathbb{1}_{\omega_p} \rangle_{L^2[0,T]}$. Ainsi :

$$\mathcal{J}_2(b_z) \simeq \frac{1}{2} \sum_{p=0}^{N} |\omega_p| \, (\sigma_0 \, \xi_p^T \, K \, \xi_p + \sigma_1 \, \xi_p'^T \, K \, \xi_p') + \frac{Q_2}{2} \sum_{(i,j) \in \{0,..,N_X^{b_z}\} \times \{0,..,N_Y^{b_z}\} \setminus (0,0)} \sum_{p=0}^{N} b_{i,j,p}^2 \,,$$

et :

$$\nabla \mathcal{J}_2(b_z) \simeq \sum_{(i,j) \in \{0,..,N_X^{b_z}\} \times \{0,..,N_Y^{b_z}\} \setminus (0,0)} \left[\sum_{p=0}^{N} [\nabla \mathcal{J}_2(b_z)]_{i,j,p} \, \mathbb{1}_{\omega_p} \right] f_{i,j}$$

où

$$\begin{aligned} [\nabla \mathcal{J}_2(b_z)]_{i,j,p} &= \Big((c^2 \, (\partial_y \alpha \, \partial_x \varphi - \partial_x \alpha \partial_y \varphi) \; + \; Q_2 \, b_z \, , \, f_{i,j})_{L^2(\Omega_H)} \, , \, \mathbb{1}_{\omega_p} \Big)_{L^2[0,T]} \\ &\simeq \Big(S \, c^2 \, \chi^T F^{i,j} \xi + Q_2 \, b_{i,j} \, , \, \mathbb{1}_{\omega_p} \Big)_{L^2[0,T]} \\ &\simeq |\omega_p| \, (S \, c^2 \, \chi_p^T F^{i,j} \xi_p + Q_2 \, b_{i,j,p}) \,, \end{aligned}$$

avec

$$F^{i,j}_{(m,n),(m',n')} = \frac{1}{k_{m,n} \, k_{m',n'}} \int_{\Omega_H} (\partial_x f_{m',n'} \partial_y f_{m,n} - \partial_y f_{m',n'} \partial_x f_{m,n}) \, f_{i,j} \,.$$

Dans le cas B_z indépendant du temps,

$$\mathcal{J}_2(b_z) \simeq \frac{1}{2} \sum_{p=0}^{N} |\omega_p| \, (\sigma_0 \, \xi_p^T \, K \, \xi_p + \sigma_1 \, \xi_p'^T \, K \, \xi_p') + \frac{Q_2}{2} \, T \sum_{(i,j) \in \{0,..,N_X^{b_z}\} \times \{0,..,N_Y^{b_z}\} \setminus (0,0)} b_{i,j}^2 \,,$$

et

$$\nabla \mathcal{J}_2(b_z) \simeq \sum_{(i,j)\in\{0,..,N_X^{bz}\}\times\{0,..,N_Y^{bz}\}\setminus(0,0)} [\nabla \mathcal{J}_2(b_z)]_{i,j}\, f_{i,j}\,,$$

avec

$$\begin{aligned}[\nabla \mathcal{J}_2(b_z)]_{i,j} &= \Big(\int_0^T c^2\,(\partial_y\alpha\,\partial_x\varphi - \partial_x\alpha\partial_y\varphi) \;+\; Q_2\,T\,b_z\,,\; f_{i,j}\Big)_{L^2(\Omega_H)} \\ &\simeq \sum_{p=0}^{N} |\omega_p|\,S\,c^2\,\chi_p^T F^{i,j}\xi_p + Q_2\,T\,b_{i,j}\,.\end{aligned}$$

5.3.4 Calcul des matrices antisymétriques

Les matrices G et $F^{i,j}$ s'écrivent sous la forme $G = H(B_z)$ et $F^{i,j} = H(f_{i,j})$, où

$$H \;:\; g \;\mapsto\; \left(\frac{1}{k_{m,n}\,k_{m',n'}}\int_{\Omega_H}(\partial_x f_{m',n'}\partial_y f_{m,n} - \partial_y f_{m',n'}\partial_x f_{m,n})\,g\right)_{(m,n),(m',n')}$$

se développe comme suit :

$$\begin{aligned}H_{(m,n),(m',n')}(g) = \\ \frac{4\pi^2\,\epsilon_{m,n}\,\epsilon_{m',n'}}{(L_x\,L_y)^2\,k_{m,n}\,k_{m',n'}}\int_{\Omega_H} g\Big[&m'n\sin\left(\frac{m'\pi}{L_x}\,x\right)\cos\left(\frac{n'\pi}{L_y}\,y\right)\cos\left(\frac{m\pi}{L_x}\,x\right)\sin\left(\frac{n\pi}{L_y}\,y\right) \\ &-n'm\cos\left(\frac{m'\pi}{L_x}\,x\right)\sin\left(\frac{n'\pi}{L_y}\,y\right)\sin\left(\frac{m\pi}{L_x}\,x\right)\cos\left(\frac{n\pi}{L_y}\,y\right)\Big]\end{aligned}$$

ou encore $H_{(m,n),(m',n')}(g) =$

$$\begin{aligned}\frac{\epsilon_{m,n}\,\epsilon_{m',n'}}{k_{m,n}\,k_{m',n'}\,L_x\,L_y}\;[\;&m'n\,(q_{m'+m,n'+n} - q_{m'-m,n'-n} + q_{m'-m,n'+n} - q_{m'+m,n'-n}) \\ &-n'm\,(q_{m'+m,n'+n} - q_{m'-m,n'-n} - q_{m'-m,n'+n} + q_{m'+m,n'-n})]\end{aligned}$$

$$\text{avec}\quad q_{m,n} = \frac{\pi^2}{L_x\,L_y}\int_{\Omega_H} g\sin\left(\frac{m\pi}{L_x}\,x\right)\sin\left(\frac{n\pi}{L_y}\,y\right).$$

Cette dernière formule est intéresante car lorsque g est uniforme,

$$q_{m,n} = g\int_0^\pi \sin(mx)\int_0^\pi \sin(nx)\,,\quad \text{avec}\quad \int_0^\pi \sin(kx) = \begin{cases} 0 & \text{si } k \text{ est pair} \\ 2/k & \text{sinon}\end{cases}.$$

Si ce n'est pas le cas, on décompose g sur la base des $f_{i,j}$, et on obtient :

$$\begin{aligned}q_{m,n} &= \frac{2\pi^2}{L_xL_y\sqrt{L_xL_y}}\sum_{(i,j)\in\mathbb{N}^2}\epsilon_{i,j}\,g_{i,j}\int_{\Omega_H}\cos\left(\frac{i\pi}{L_x}\,x\right)\cos\left(\frac{j\pi}{L_y}\,y\right)\sin\left(\frac{m\pi}{L_x}\,x\right)\sin\left(\frac{n\pi}{L_y}\,y\right) \\ &= \frac{1}{2\sqrt{L_xL_y}}\sum_{(i,j)\in\mathbb{N}^2}\epsilon_{i,j}\,g_{i,j}\,(r_{m+i,j+n} - r_{m-i,j-n} + r_{m-i,j+n} - r_{m+i,j-n})\end{aligned}$$

$$\text{avec}\quad r_{k,l} = \frac{\pi^2}{L_x\,L_y}\int_{\Omega_H}\sin\left(\frac{k\pi}{L_x}\,x\right)\sin\left(\frac{l\pi}{L_y}\,y\right).$$

Nous disposons à présent de tous les éléments nécessaires à l'implémentation des problèmes de contrôle étudiés.

5.4 Résultats numériques

On fixe dans tous les test présentés ici les paramètres physiques et numériques suivants :

- $L_x = 3\,,\ L_y = 10\,m$
- $\rho_1 = 2300\,,\ \rho_2 = 2150\ kg/m^3$
- $h_1 = 0.2\,,\ h_2^0 = 0.05\,m$
- $\overline{B}_z = 0.5\,,\ B_0 = 10^{-3}\,T$
- $J_0 = 10^5\ A/m^2$
- $\tau = 5\,s\,,\quad L = 1\,m$
- $N_X = 3\,,\ N_Y = 3$
- $\Delta t = 0.05$
- $N_X^{b_z} = 2\,,\ N_Y^{b_z} = 2$

Choix de la donnée initiale

En faisant évoluer sur deux unités de temps la gaussienne centrée à moyenne nulle :

$$\eta(x,y) = 0.04\left(\exp[-(x-\bar{x})^2 - 0.1\,(y-\bar{y})^2] - \overline{\exp[-(x-\bar{x})^2 - 0.1\,(y-\bar{y})^2]}\right)$$

(où $(\bar{x},\bar{y})$ désigne le centre du domaine Ω_H), on obtient comme déformée d'interface :

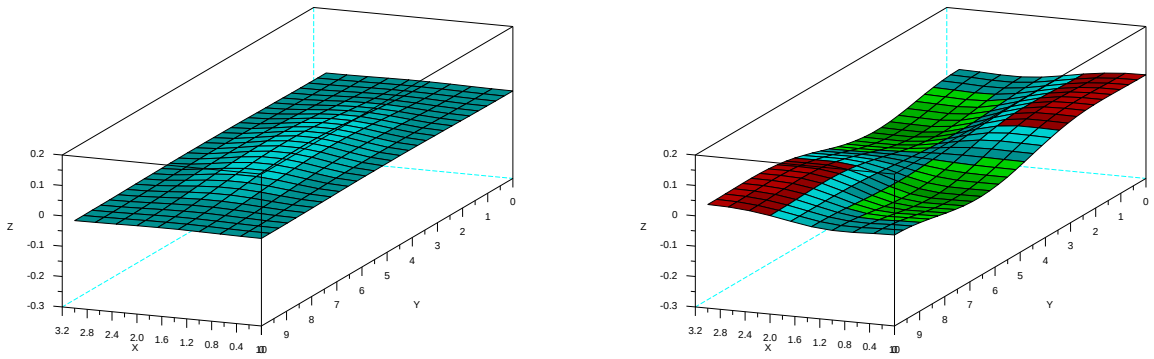


FIG. 5.1 – Évolution d'une gaussienne sur deux unités de temps

qu'on choisit, assortie de sa dérivée temporelle, comme donnée initiale (η_0, η_1).

Problèmes d'optimisation

Par une méthode de gradient conjugué, on cherche à atténuer $T = 1$ puis en $T = 2$ la déformée d'interface d'une part $((\sigma_0,\sigma_1) = (1,0))$ et sa vitesse d'autre part $((\sigma_0,\sigma_1) = (0,1))$:

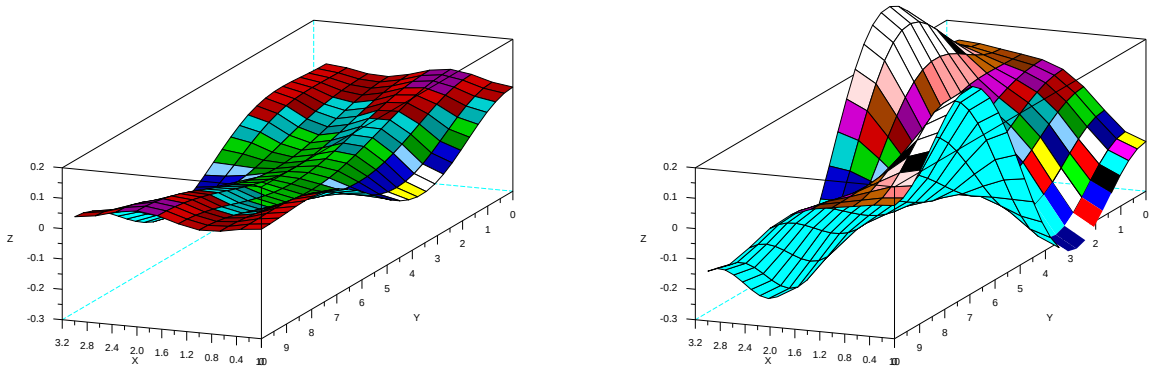


FIG. 5.2 – Déformées d'interface en $T = 1$ et $T = 2$ sans optimisation

Les gradients des fonctions coût sont en général divisés par 100 à la fin de l'algorithme.

5.4.1 Commande $h_2(t)$ sur une unité de temps ; $Q_1 = 1$

Cas indépendant du temps

Contrôle de l'interface - on parvient à réduire le coût initial $\mathcal{J}_1(h_2^0) = 0.0201$ de 4 % par la commande optimale $h_2 = 0.072$, en accord avec les profils de la fonction coût et de son gradient :

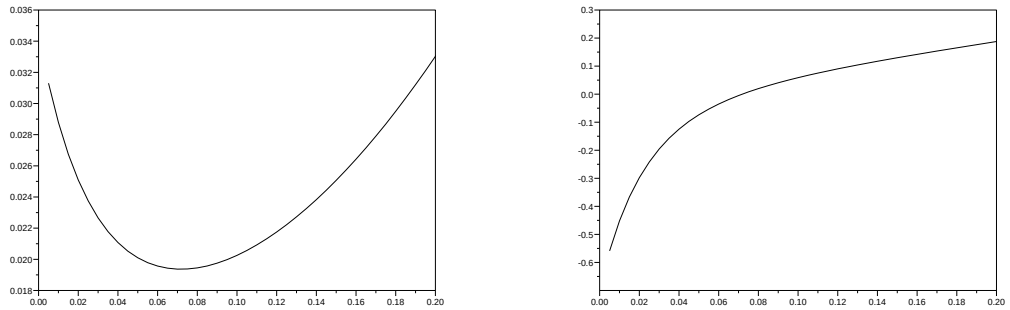


FIG. 5.3 – $\mathcal{J}_1(h_2)$ et $\nabla\mathcal{J}_1(h_2)$

Contrôle de la vitesse de l'interface - le minimum est atteint en $h_2 = 0.105$, qui donne un gain de 12 % par rapport à $\mathcal{J}_1(h_2^0) = 0.0623$:

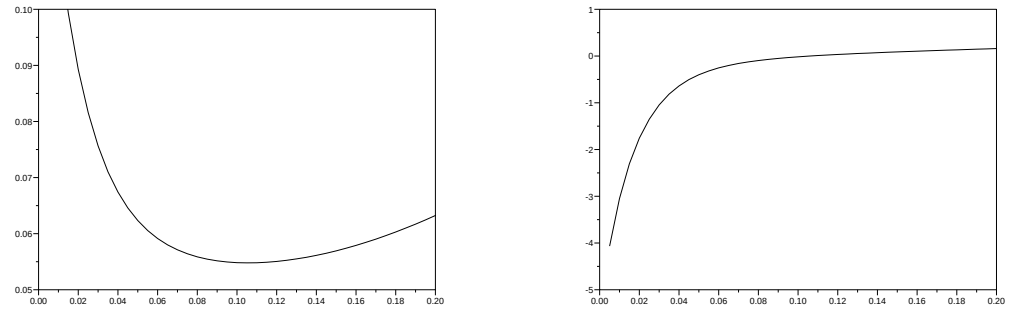


FIG. 5.4 – $\mathcal{J}_1(h_2)$ et $\nabla\mathcal{J}_1(h_2)$

Cas h_2 constant sur $[0, T/2]$ et sur $[T/2, T]$ (commande en dimension 2)

Contrôle de l'interface - on réduit le coût $\mathcal{J}_1(h_2^0)$ de 6 % pour $h_2 = (h_2^1, h_2^2) = (0.093, 0.043)$.

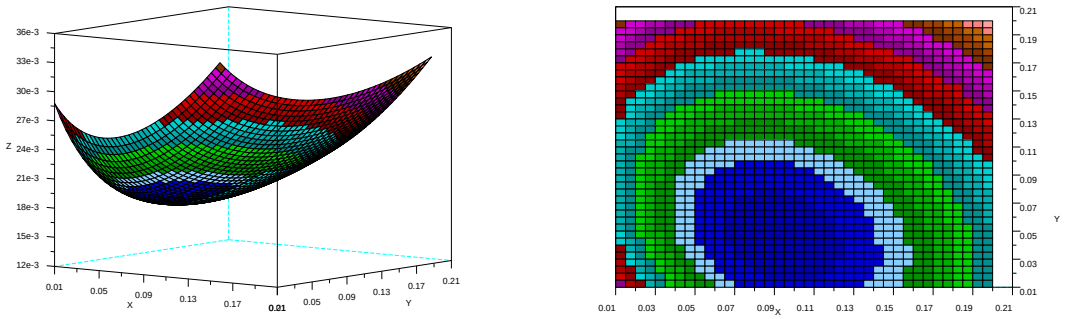


FIG. 5.5 – $\mathcal{J}_1(h_2^1, h_2^2)$ (vues de profil et de dessus)

Cas général

Contrôle de l'interface - on réduit le coût $\mathcal{J}_1(h_2^0)$ de 7 % par la commande représentée ci-dessous avec les deux précédentes et l'interface assosiée en T (qui diffère peu des deux autres cas) :

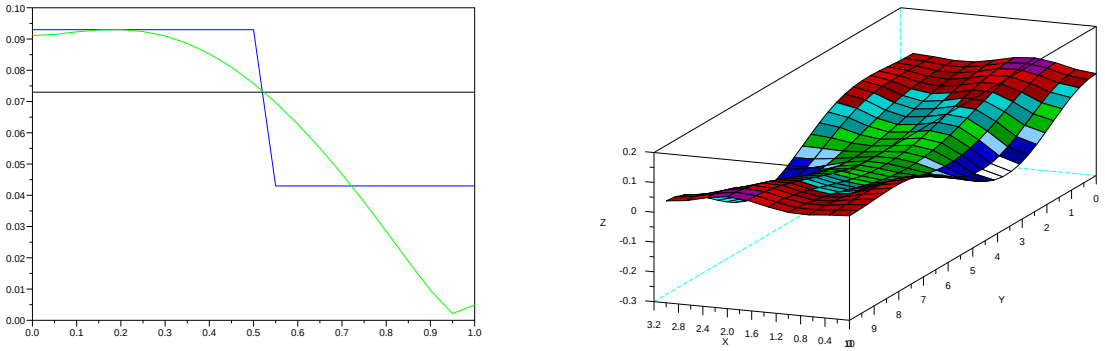


FIG. 5.6 – Commande optimale $h_2(t)$ et interface obtenue en T

Contrôle de la vitesse de l'interface - le gain obtenu est de 13 % sur le coût $\mathcal{J}_1(h_2^0)$:

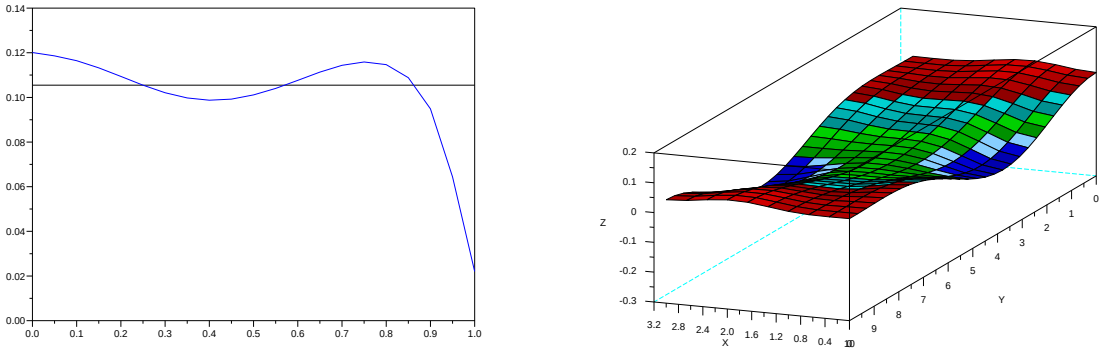


FIG. 5.7 – Commande optimale $h_2(t)$ et interface obtenue en T

5.4.2 Commande $B_z(t,x,y)$ sur une unité de temps ; $Q_3 = 10^{-4}$

Cas indépendant du temps

Contrôle de l'interface - on réduit le coût $\mathcal{J}_2(0) = 0.0189$ de 13 % par la commande :

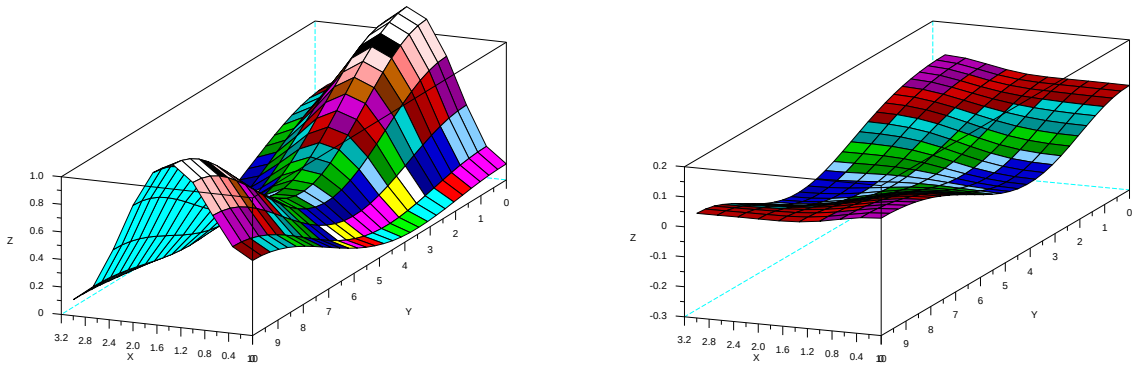


FIG. 5.8 – Commande optimale $B_z(x,y)$ et interface obtenue en T

Contrôle de la vitesse de l'interface - on obtient un gain de 6% sur le coût $\mathcal{J}_2(0) = 0.0611$:

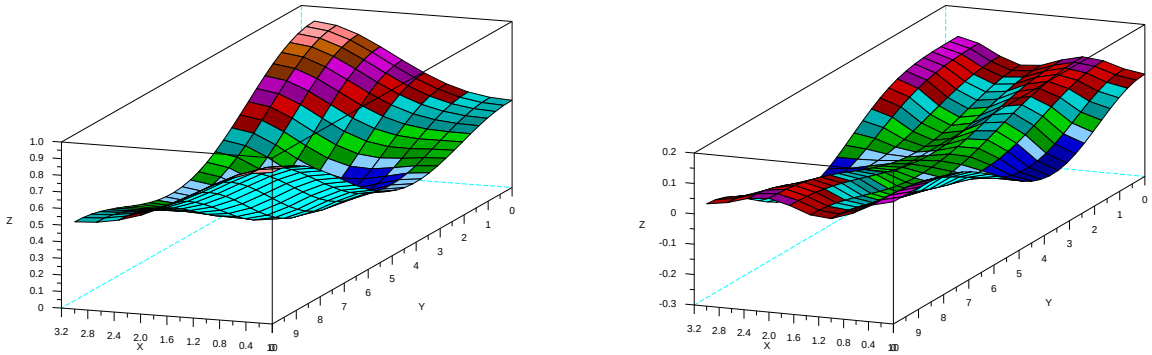


FIG. 5.9 – Commande optimale $B_z(x, y)$ et interface obtenue en T

Cas général

Contrôle de l'interface - on réduit de 13 % le coût $\mathcal{J}_2(0)$ par la commande B_z suivante :

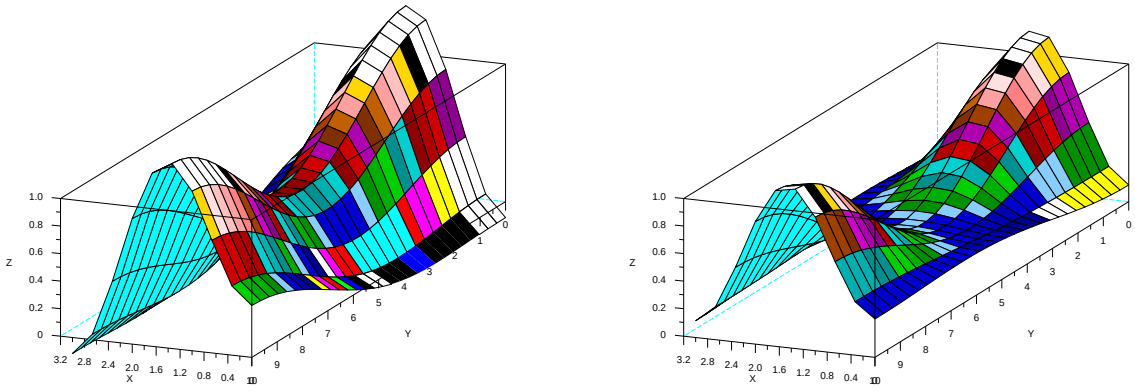


FIG. 5.10 – Commande optimale $B_z(x, y)$ en $t = 0.25$ et $t = 0.5$

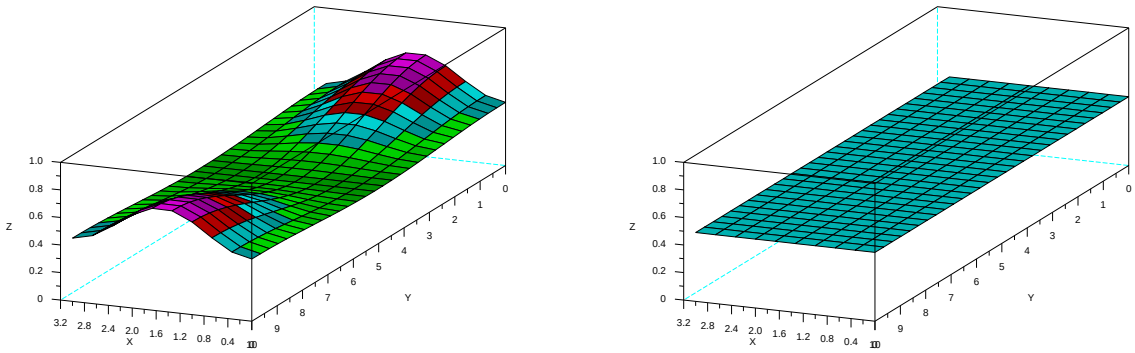


FIG. 5.11 – Commande optimale $B_z(x, y)$ en $t = 0.75$ et $t = 1$

La déformée d'interface en T, quant à elle, diffère peu du cas B_z constant.

Contrôle de la vitesse de l'interface - on réduit le coût $\mathcal{J}_2(0) = 0.0611$ de 11 % par la commande B_z suivante :

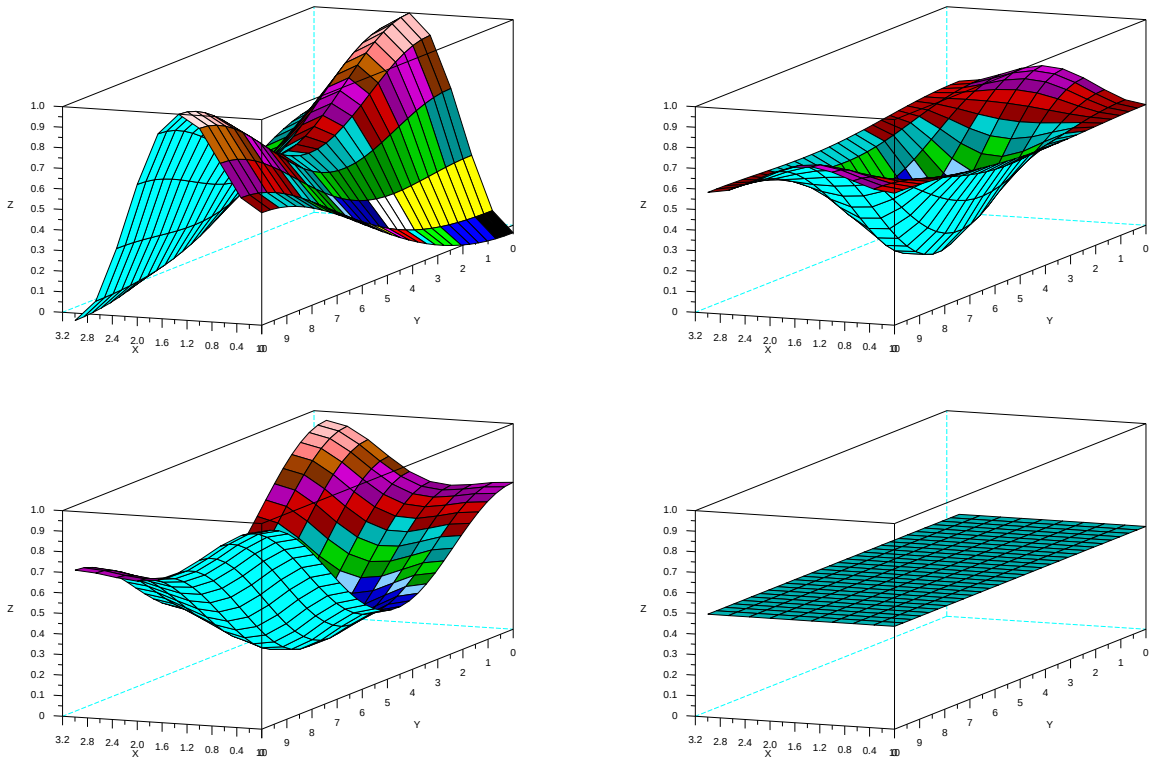


FIG. 5.12 – Commande optimale $B_z(x, y)$ en $t = 0.25$, 0.5, 0.75 et 1.

De même que pour le contrôle de l'interface, la déformée diffère peu du cas constant.

5.4.3 Commande $h_2(t)$ sur deux unités de temps ; $Q_1 = 1$

Cas indépendant du temps

Contrôle de l'interface - par les commandes h_2 et $h_2(t)$, on réduit le coût $\mathcal{J}_1(h_2^0) = 0.197$ de 43 % et 45 %.

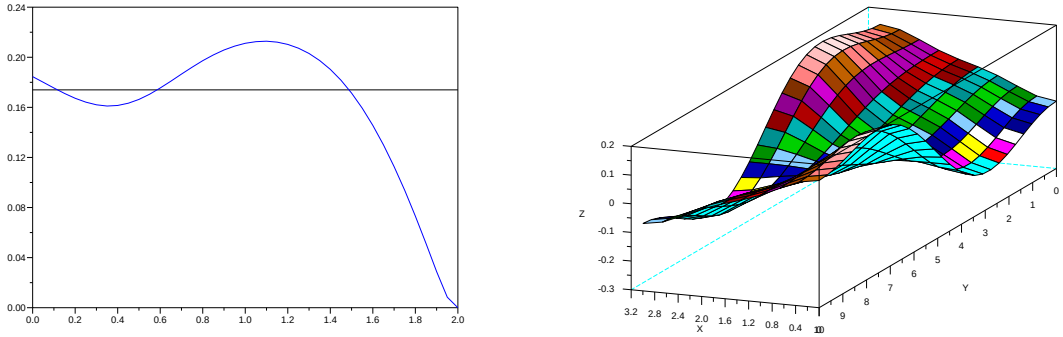


FIG. 5.13 – Commandes optimales h_2 et $h_2(t)$ et interface obtenue en T

Contrôle de la vitesse de l'interface - les gains obtenus sur $\mathcal{J}_1(h_2^0) = 0.428$ sont de 63 % et 64 % :

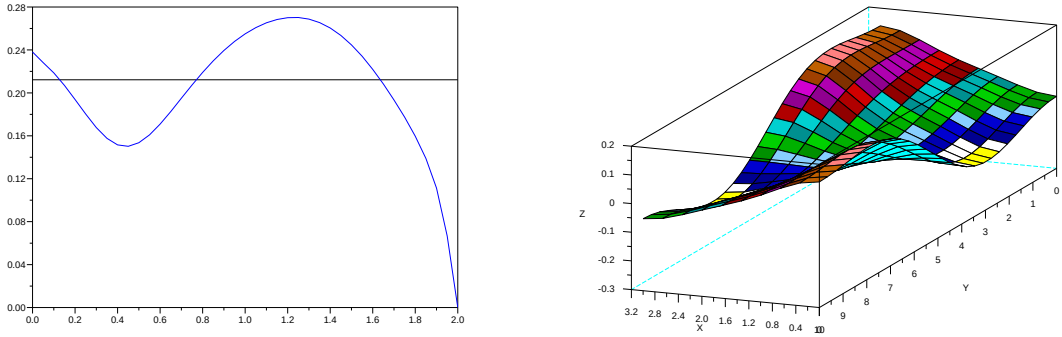


FIG. 5.14 – Commande optimale $h_2(t)$ et interface obtenue en T

5.4.4 Commande $B_z(t, x, y)$ sur deux unités de temps ($Q_2 = 5 \cdot 10^{-5}$)

Cas indépendant du temps

Contrôle de l'interface - on réduit le coût $\mathcal{J}_2(0) = 0.194$ de 10 % par la commande :

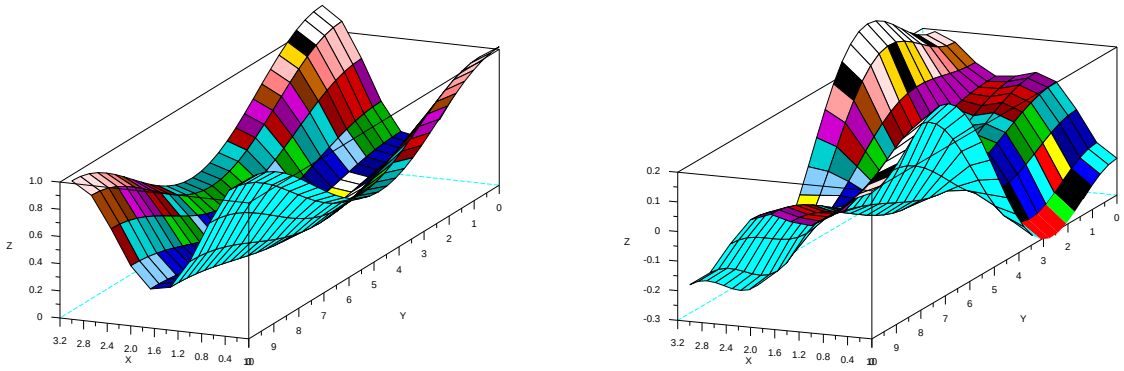


FIG. 5.15 – Commande optimale $B_z(x, y)$ et interface obtenue en T

Contrôle de la vitesse de l'interface - on réduit le coût $\mathcal{J}_2(0) = 0.426$ de 33 % par la commande B_z suivante :

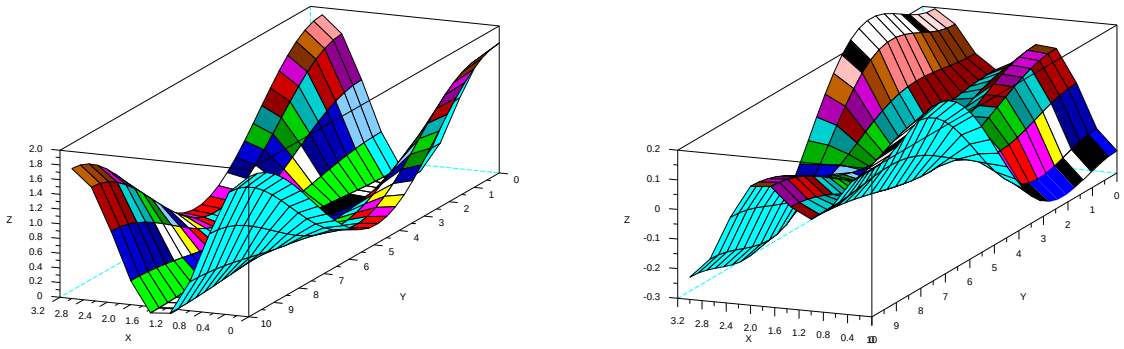


FIG. 5.16 – Commande optimale $B_z(x, y)$ et interface obtenue en T

Cas général

Contrôle de l'interface - on réduit le coût $\mathcal{J}_2(0)$ de 50 % en obtenant une déformée d'interface sensiblement meilleure que dans le cas B_z constant :

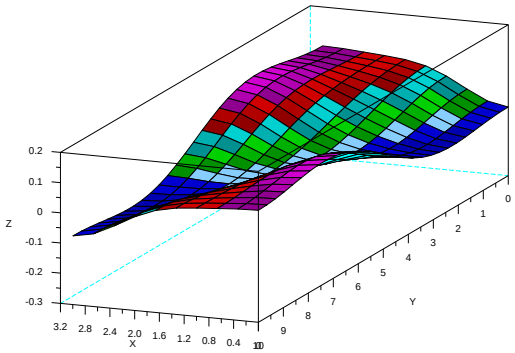


FIG. 5.17 – Déformée d'interface en T

Ci-dessous la commande permettant d'arriver à cette performance :

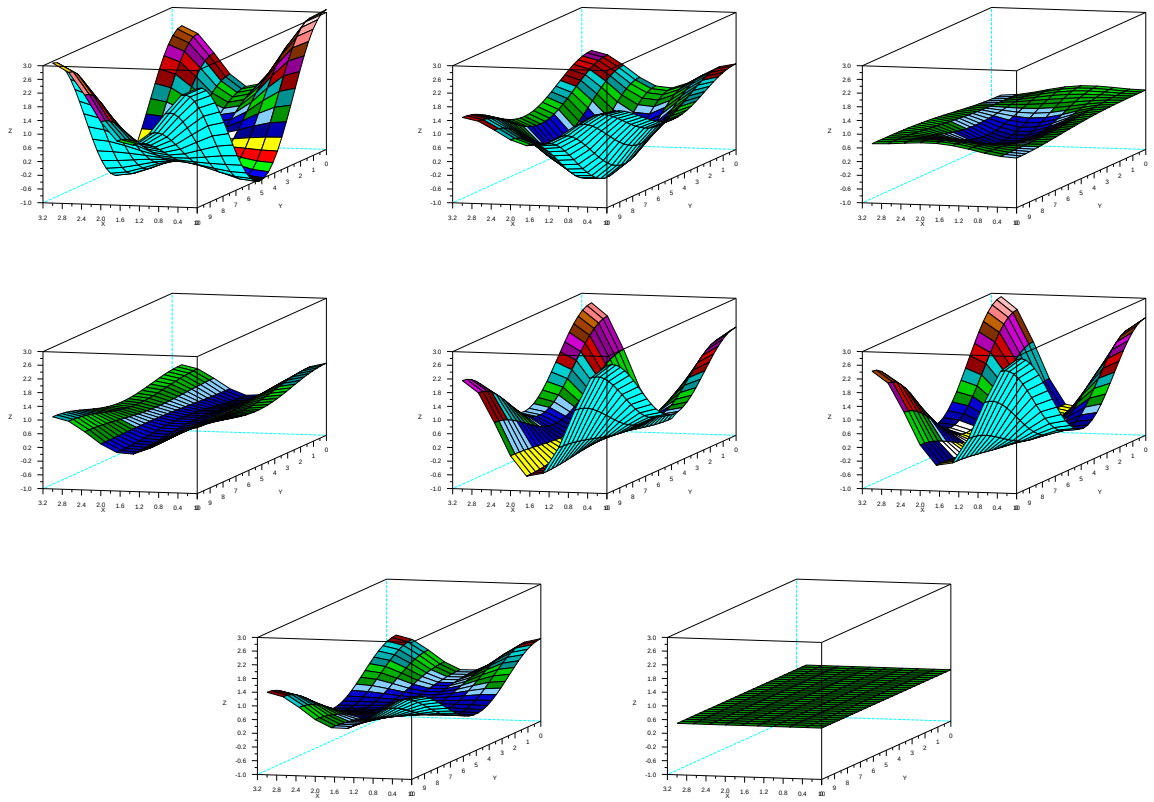


FIG. 5.18 – Commande optimale $B_z(x, y)$ en $t = 0.25,\ 0.5,\ 0.75,\ 1,\ 1.25,\ 1.5,\ 1.75$ et 2

Contrôle de la vitesse de l'interface - on obtient 61 % de gain, et l'interface :

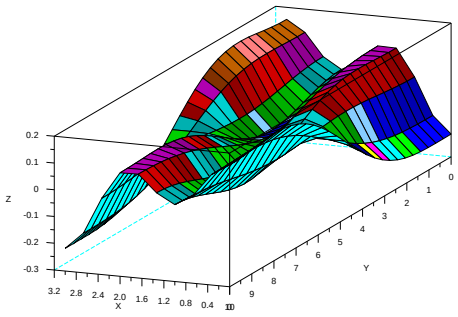


FIG. 5.19 – Déformée d'interface en T

Ce résultat n'est pas spectaculaire, donc *a priori* étonnant compte tenu du gain observé sur la fonction coût, mais en visionnant l'évolution de l'interface au cours du temps, on se rend compte que cette dernière bouge peu, en accord avec le critère choisi. Ci-dessous la commande correspondante :

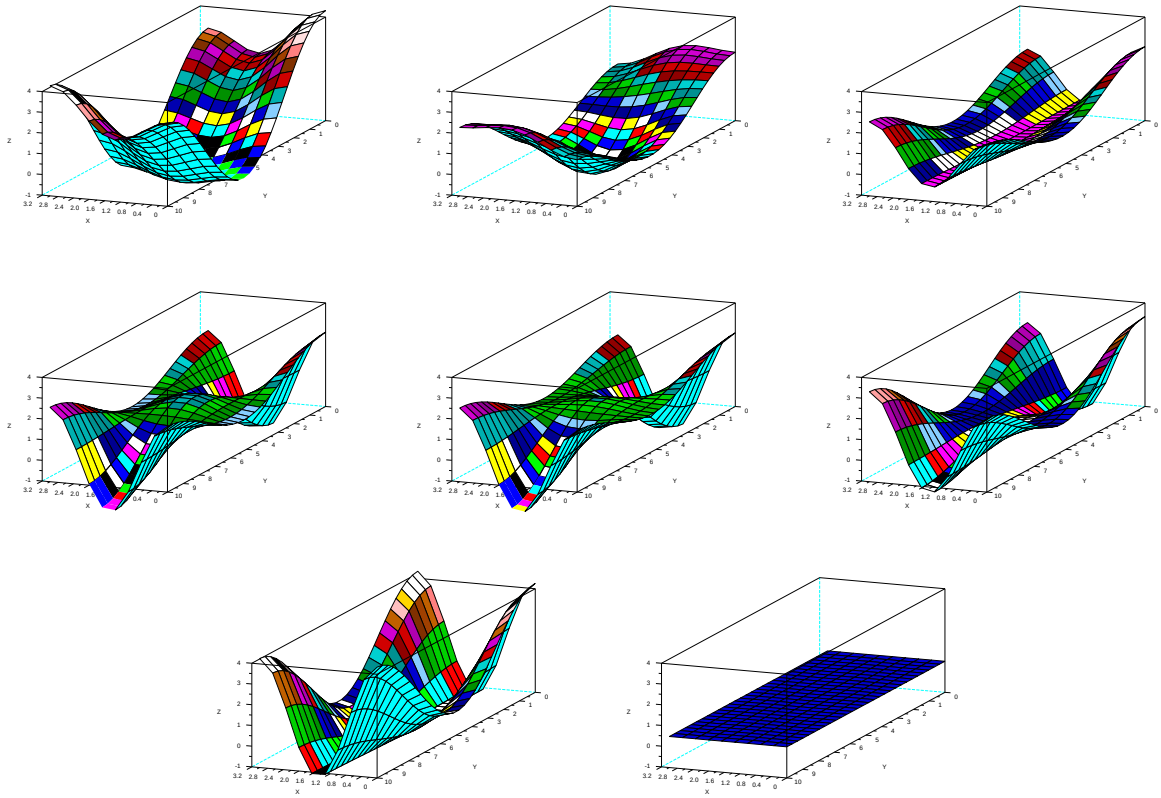


FIG. 5.20 – Commande optimale $B_z(x, y)$ en $t = 0.25$, 0.5, 0.75, 1, 1.25, 1.5, 1.75 et 2.

5.5 Bilan

Commande $h_2(t)$

Le premier cas exposé (contrôle de la déformée sur une unité de temps) montre clairement la cohérence entre les cas indépendant et dépendant du temps. En effet, on observe que la commande instationnaire h_2, d'une part, oscille autour de la commande constante optimale, et, d'autre part, est en accord avec le cas h_2 bidimensionnel. Par ailleurs, on observe que la commande instationnaire fournit des coûts optimaux certes meilleurs que dans le cas constant, mais améliore peu ce dernier. Enfin, la commande oscille plus pour le critère "vitesse" que pour le critère "déformée", ainsi que pour une durée plus grande.

$T = 1$	déformée	vitesse
stationnaire	4	12
instationnaire	7	13

$T = 2$	déformée	vitesse
stationnaire	43	63
instationnaire	45	64

TAB. 5.1 Tableaux récapitulatifs des gains pour $u = h_2$

Commande $B_z(t, x, y)$

Le contrôle par le champ magnétique vertical révèle que si l'on choisit ce dernier indépendant du temps, on obtient de meilleurs gains en temps court qu'en temps plus long. Par ailleurs, le fait de permettre à B_z d'évoluer dans le temps améliore très peu les performances, toujours pour $T = 1$. La situation change complètement dès lors qu'on cherche à contrôler en temps plus long, en effet dans ce cas les gains sont bien meilleurs pour une commande instationnaire, ce qui est une différence notable par rapport à la commande h_2. Enfin, soulignons l'intérêt de la commande B_z, qui permet de maintenir la DAM à sa valeur basse h_2^0.

$T = 1$	déformée	vitesse
stationnaire	13	6
instationnaire	13	11

$T = 2$	déformée	vitesse
stationnaire	10	33
instationnaire	50	61

TAB. 5.2 Tableaux récapitulatifs des gains pour $u = b_z$

Conclusion

La minimisation de la vitesse de l'interface, d'une manière générale - excepté pour le cas B_z stationnaire et $T = 1$ - apparaît comme un critère qui s'optimise plus facilement que la déformée elle-même. D'autre part, la commande h_2, semble-t-il, na pas besoin de dépendre du temps pour être efficace, tandis que le champ magnétique vertical, au contraire, requiert d'être modifié en permanence pour offrir de bonnes performances. D'un point de vue pratique, ce modèle linéaire préconise donc de disposer d'un appareil électromagnétique permettant de régler en temps réel le champ magnétique ambiant. Ce dernier étant principalement généré par des conducteurs fixes, on imagine difficilement la mise en œuvre d'un tel procédé. Afin de détecter un autre moyen d'améliorer la stabilité des cuves, nous utilisons dans la suite le modèle non linéaire pour explorer plus en détail le phénomène physique.

Chapitre 6

Parallélisation du code non linéaire

6.1 Présentation du problème

La simulation d'un phénomène de rolling (cf. **3.3**) sur 10 unités de temps nécessite une journée de calcul sur une machine moyenne possédant quelques GHz de fréquence d'horloge, alors que le modèle utilisé est une grossière approximation de la réalité physique (suppression des phénomènes thermiques, chimiques et turbulents, géométrie cylindrique, etc.). Ce coût élevé provient de la complexité mathématique du problème (3.8)-(3.7), qui demeure importante (nombre d'inconnues, déplacement du maillage) même une fois les hypothèses ci-dessus prises en compte. D'une manière générale, plus un modèle est précis, et plus il est cher à résoudre, au sens propre du terme. Il est alors possible de réduire ce coût en améliorant soit l'efficacité du modèle (le rapport entre sa pertinence et sa complexité mathématique), soit celle de son approximation. La première solution relève de la physique fondamentale, tandis que la deuxième se situe à l'intersection des mathématiques et de l'informatique, mais nourrit également la modélisation pure en fournissant des informations sur sa portée physique. La simulation numérique n'a donc pas qu'une signification *quantitative* de prévision de phénomènes en fonction de paramètres donnés, mais joue aussi un rôle *qualitatif* de mise à l'épreuve des modèles lorsqu'ils ne sont pas encore complètement validés. Dans cette phase exploratoire, il est parfois plus utile de connaître rapidement la réponse du modèle à un paramètre particulier, que de tester plusieurs paramètres simulanément, mais plus lentement. Cette situation se présentera typiquement au chapitre **7**, lorsqu'il s'agira d'exhiber des actionneurs en vue du contrôle de l'interface électrolyte-aluminium (cf. **8**).

Un autre exemple est le problème du contrôle en lui-même, dont la raison d'être est de décider *automatiquement* des valeurs à donner aux paramètres pour obtenir une réponse particulière du modèle, plutôt que de lancer plusieurs tests en espérant tomber sur la valeur en question. En utilisant par exemple un algorithme de gradient (cf. **2** et **5**), il est nécessaire de résoudre à chaque itération deux problèmes du type (3.8)-(3.7).

Ainsi, lorsque ce genre de cas se présente, et que l'on dispose de plusieurs ordinateurs, il est avantageux d'optimiser la puissance de calcul en utilisant *tous* les ordinateurs pour l'exécution d'un *seul* programme. Cette technique, qui porte le nom de *parallélisation*, requiert de diviser le programme en parties les plus indépendantes possibles les unes des autres, afin de limiter le nombre de *communications* (transferts de données) entre ordinateurs. La méthode des éléments finis possède justement une structure *locale* remplissant cette condition, qu'on exploite ici pour paralléliser le code non linéaire. En décomposant l'inconnue sur des sous-domaines du maillage, puis en appliquant au *problème distribué* résultant un solveur parallèle de la librairie Aztec [97], on ramène le temps d'exécution à l'échelle de l'heure plutôt que de la journée.

6.2 Notions sur le calcul parallèle

6.2.1 Latence - granularité

En informatique, la *performance* d'un programme s'évalue directement comme l'inverse de son temps d'exécution T_X. Si, de plus, on prend en compte le coût de l'exécution en rapportant cette performance à la puissance totale de calcul utilisée N_p (exprimée en nombre de processeurs), on définit l'*efficacité* par

$$E = \frac{1}{T_X\, N_p}\,.$$

D'après cette définition, si l'on considère un problème décomposé en N_p sous-problèmes de taille N_p fois plus petite, les temps de communication $\varepsilon > 0$ entraînent que le problème décomposé est moins efficace que le problème entier (programme *séquentiel*), car

$$\frac{1}{\left(\dfrac{T_X}{N_p} + \varepsilon\right) N_p} < \frac{1}{T_X}\,.$$

Il est donc préférable de *ne pas* paralléliser le programme.

En calcul scientifique, ce n'est plus vraiment l'efficacité du programme en lui-même qui compte, mais plutôt l'intérêt des informations qu'il renvoie. Lorsque celui-là est suffisant, par exemple, pour faire l'économie du lancement de N_r programmes simultanés, il est clair que la parallélisation est souhaitable si les temps de communication sont suffisamment faibles, en effet

$$\varepsilon < \frac{T_X}{N_p}\,(N_r - 1) \quad \Rightarrow \quad \frac{1}{\left(\dfrac{T_X}{N_p} + \varepsilon\right) N_p} > \frac{1}{T_X N_r}\,.$$

Ainsi, il s'agit typiquement d'un problème de *latence* et non de *débit* : un Concorde a une meilleure latence qu'un Boeing 747 car il peut transporter deux fois plus rapidement les passagers, mais comme il en contient trois fois moins, il possède au final un débit inférieur. L'enjeu de la parallélisation est d'optimiser la latence du code, de manière à pouvoir connaître rapidement la réponse du modèle à un paramètre donné, dans une procédure *essai-erreur* par exemple.

L'efficacité d'un programme parallèle est non seulement tributaire des temps de communication qu'il engendre, mais aussi de l'exploitation maximale de la puissance des ordinateurs. Celle-là se trouve freinée par les parties non parallélisables du code, qui sont exécutées identiquement par chaque processeur (ou *nœuds*). En général, ces sections trouvent leur origine dans la nécessité, à certains moments de l'exécution, de regrouper sur chacun des processeurs l'information issue de l'ensemble des nœuds du programme pour pouvoir poursuivre ce dernier. Alors, dans le cas où certains processeurs ont mis plus de temps que les autres à effectuer la tâche permettant l'obtention de ladite information, tous ceux qui ont travaillé plus rapidement se retrouvent dans une situation d'attente, et n'utilisent donc pas tout leur potentiel. Cet inconvénient est malheureusement inévitable, car certaines parties du code se prêtent mal à la parallélisation, à partir du moment où le gain obtenu en nombre d'opérations est neutralisé (voire surpassé) par la perte occasionnée due aux communications qu'il nécessite. Ainsi, on définit la *granularité* d'un code parallèle par le rapport entre la charge de travail W (nombre d'opérations) par processeur et les temps de communication :

$$G = \frac{W}{\varepsilon}\,,$$

et donc plus la granularité est élevée, meilleure est l'efficacité.

6.2.2 Classification de Flynn

Dès 1966, M.J. Flynn [33] propose une classification des architectures d'ordinateurs, basée sur la combinaison entre les lettres S et M (pour *Single* et *Multiple*) d'une part, et I et D (pour *Instruction* et *Data*) d'autre part. Ainsi trois catégories sont distinguées :

- SISD : la plupart des ordinateurs actuels (exécution séquentielle)
- SIMD : la même instruction est exécutée *simultanément* sur des données différentes, de manière transparente pour le programmeur (machines vectorielles par exemple),
- MIMD : des flux d'instructions indépendants sont envoyés de manière *asynchrone* sur des données indépendantes. Citons le cas particulier (important) SPMD : *Single Program, Multiple Data*, où les instructions sont les mêmes, à ne pas confondre avec SIMD.

L'architecture la plus répandue est de loin la dernière, car elle peut être construite à partir de machines ordinaires (réseaux Ethernet de stations de travail, clusters) communiquant par passage de messages. On peut même affirmer que c'est sa variante SPMD qui est utilisée par la majorité des codes parallèles, car elle permet d'écrire un unique code source pour l'ensemble des processeurs.

6.2.3 Mémoire distribuée - mémoire partagée

On peut distinguer deux types d'architectures MIMD, dans lesquelles la mémoire vive se trouve propre à chaque machine (mémoire distribuée) ou commune à l'ensemble des processeurs (mémoire partagée). Dans le premier cas, l'avantage est un accès rapide à la mémoire, tandis que l'inconvénient est un ralentissement éventuel de communication entre les machines. Dans le deuxième, c'est l'inverse (gestion de cohérence des caches ...).

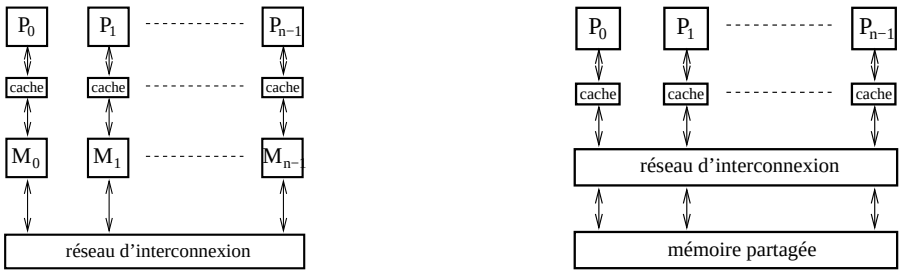


FIG. 6.1 – Mémoires distribuée et partagée (P : processeur, M : mémoire)

Nous effectuerons nos tests sur une machine d'architecture hybride : c'est un cluster de biprocesseurs qui possèdent chacun leur propre mémoire, ce qui en fait un système à mémoire distribuée, mais les deux processeurs qui constituent un élément fonctionnent, eux, en mémoire partagée :

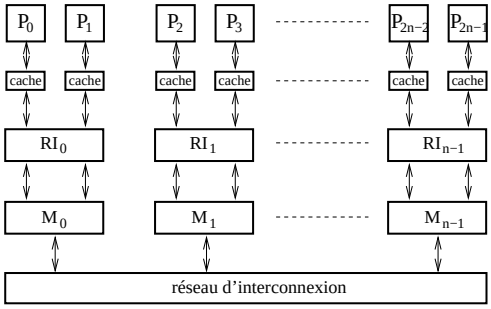


FIG. 6.2 – CLUMPS (cluster de SMP, Symmetric Multiprocessor)

6.3 Description du code initial

Le code *Mistral*, écrit en *C++* entre 1995 et 2003 par J.-F. Gerbeau, seul d'abord puis avec T. Lelièvre, est un programme de résolution par éléments finis d'une classe assez générale d'EDP, typiquement des problèmes paraboliques-elliptiques multi-inconnues comme les équations de Navier-Stokes ou de la MHD. En plus de cela, il possède des utilitaires capables de déformer le maillage au cours du temps, et donc de traiter des problèmes en formulation ALE (cf. **3.3**). Dans un premier temps, on peut distinguer deux grandes phases dans son exécution (et donc dans sa conception) :

- premièrement, la *construction* du problème initialise le maillage, l'interpolation et la structure matricielle qui en résulte ; cette opération est effectuée *une seule fois* au début du programme.
- deuxièmement, le programme consiste à remplir la structure matricielle ainsi construite en fonction des données d'entrées (*assemblage*), puis à résoudre le système linéaire qui en résulte.

Dans les problèmes itératifs (cas instationnaires et / ou non linéaires), la deuxième procédure est exécutée à *chaque itération*, et pour les phénomènes instationnaires qui nous intéressent principalement, le maillage se déforme au cours du temps (ce qui, en dehors du calcul de w, nécessite des opérations supplémentaires de coût non négligeable). Dans les problèmes de MHD bifluide, on observe que la phase d'assemblage consomme en réalité près de 80% du temps de calcul, contre 20% pour la résolution. Soulignons enfin qu'un tel problème est très consommateur de ressources informatiques, d'où l'utilisation majoritaire de formulations en éléments finis stabilisés d'ordre 1. Dans le cas magnétohydrodynamique tridimensionnel, la taille du système linéaire à inverser à chaque itération correspond alors à sept fois le nombre de sommets du maillage (du fait des deux inconnues vectorielles v et B, et de l'inconnue scalaire p).

6.3.1 Initialisation

6.3.1.a Maillage

Solveur "pur", dans le sens où il ne contient pas de mailleur, mais prend en entrée un fichier au format *Fidap Neutral File*, Mistral commence par construire l'objet *Geometrie*, chargement de ce fichier en mémoire. Un maillage Ω_h d'un domaine Ω est constitué au minimum :

- d'un ensemble de points de Ω identifés par un numéro *global* $I \in \{1, ..N_h\}$
- d'un ensemble d'arêtes reliant ces points, de manière à former une partition du domaine approché Ω_h en polyèdres $(K^l)_{1 \leq l \leq L_h}$ (*éléments géométriques volumiques*), tels que l'intersection de deux polyèdres distincts soit l'ensemble vide, un point, une arête ou une face.

Par ailleurs, il doit être possible de regrouper les éléments en sous-domaines particuliers, les *figures*, pour distinguer des endroits auxquels sont appliqués des opérateurs différents. Ils s'ensuit une renumérotation *locale* des éléments de manière à pouvoir y accéder rapidement sur une figure donnée. De même, dans un élément donné, les points doivent être localement numérotés pour identifier leurs positions relatives. Un point commun à plusieurs éléments a donc plusieurs numéros locaux. Cet aspect est crucial pour le calcul de la contribution élémentaire du problème sur un polyèdre (voir le paragraphe suivant). Parallèlement à cela, des portions de la frontière de certains éléments géométriques doivent être repertoriés comme des éléments *surfaciques*, afin de permettre la prise en compte des conditions aux limites. Comme dans le cas des éléments volumiques, ces portions sont alors regroupées en figures surfaciques et leurs points et éléments localement renumérotés (cf. Fig. 6.3). Enfin, on peut avoir besoin d'accéder à l'ensemble des points d'une figure (typiquement pour les conditions de Dirichlet), d'où un troisième type d'identifiant qui est le numéro du point dans la figure qui le contient (il peut y en avoir plusieurs).

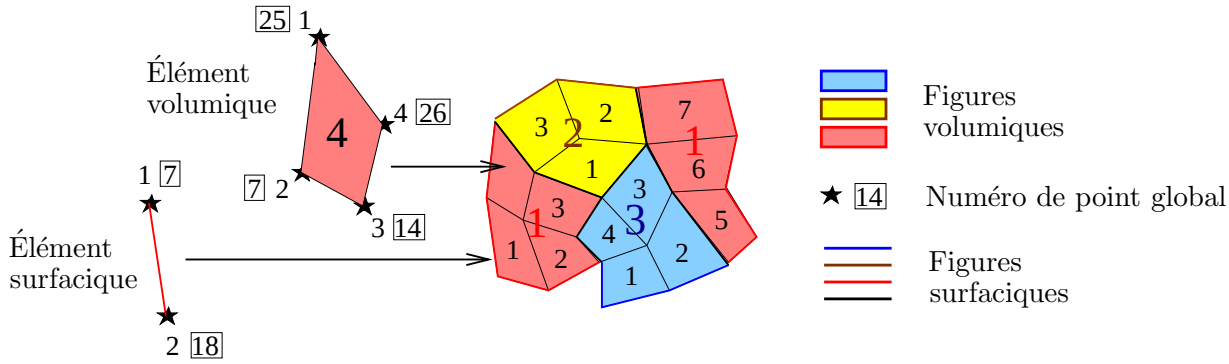


FIG. 6.3 – Maillage et identifiants locaux

En plus des sommets des éléments, d'autres points peuvent être ajoutés au maillage (sur leurs arêtes par exemple), afin de pouvoir supporter divers types d'interpolation. Au niveau informatique, cette structure de données se traduit comme suit :

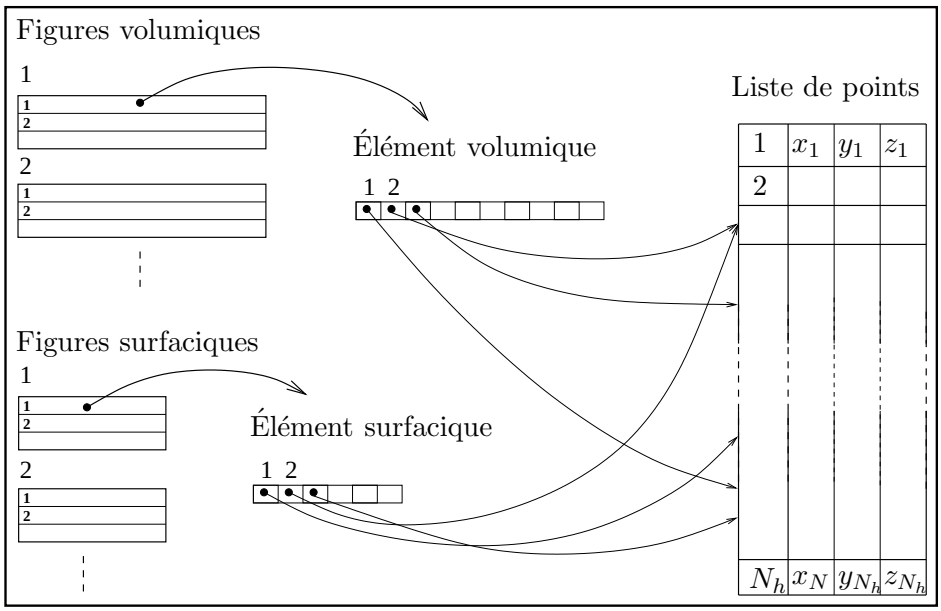


FIG. 6.4 – Structure de données *Geometrie*

En dehors de cette structure de base, deux outils importants sont présents dans l'objet *Geometrie* (sous forme de *fonctions membre*), à savoir le calcul de normales en des points d'une figure surfacique, et le déplacement du maillage (qui consiste à ne modifier que les coordonnées dans la liste de points FIG. 6.4). On dispose ainsi d'un maillage *non structuré* du domaine de résolution. Cette configuration a pour avantage de permettre une souplesse d'emploi de la discrétisation spatiale par rapport à la géométrie du domaine, mais pour inconvénient de ne pas pouvoir disposer d'une connectivité simple du maillage (cf. **6.3.1.c**).

Dans le problème qui nous intéresse, la configuration standard repose sur deux figures volumiques (aluminium, cryolithe) et cinq figures surfaciques (cathode, anode, parois latérales pour l'aluminium et la cryolithe, interface), auxquelles il faut ajouter leurs réunions respectives.

6.3.1.b Interpolation

Nous avons vu en **2.2** que la méthode des éléments finis conduisait à résoudre au moins une fois un système de la forme

$$Ax = b\,,$$

où la matrice A est consituée de contributions *inter-nodales* de l'équation, notamment du type

$$\int_\Omega \nabla w_I \cdot \nabla w_J\,, \quad 1 \le I, J \le N_h\,,$$

où les $(w_I)_{1\le I\le N_h}$ sont des fonctions de forme sur Ω, ou encore des fonctions de base *globales* (une par point), par opposition aux fonctions de base *locales* τ_i^l (plusieurs par élément). Sur un maillage non structuré, le calcul de ces termes requiert *a priori* la connaissance de chaque fonction w_I. En pratique, il existe une méthode plus simple, basée sur l'*assemblage* de A_{IJ} (voir **6.3.2.a**) à partir des contributions *élémentaires* des fonctions de base locales qui constituent w_I et w_J. Ainsi, il est possible de parcourir le maillage élément par élément plutôt que point par point, ce qui a pour avantage de ne pas considérer des réunions d'éléments comme domaines de calcul, mais aussi de ramener tous les calculs élémentaires sur un *même élément de référence* $\hat K$ par transformation géométrique. C'est ce dernier point qui détermine la structure informatique d'un élément fini : il s'agit d'une combinaison entre un élément géométrique (l'ensemble des sommets d'un polyèdre) et les fonctions de base locales $\hat\tau_i$ de l'élément de référence, choisi en fonction des commodités calculatoires qu'il procure. Concrètement, dans Mistral, les éléments géométriques volumiques en 2-d sont des quadrangles (cf. FIG. 6.3 ci-dessus), aussi l'élément de référence est-il le carré unité. Si l'on fixe l'ordre d'interpolation à 1 pour construire la base (w_I), alors $\operatorname{Card}\Sigma = 4$ et il faut donc réaliser les 16 opérations

$$\mathbb{A}_{ij}^l = \int_{K^l} \nabla\tau_i^l \cdot \nabla\tau_j^l\,dx\,, \quad 1 \le i, j \le 4$$

sur un élément pour balayer l'ensemble des interactions entre les différentes fonctions de base globales non nulles à cet endroit. Or il s'avère que les applications géométriques définies et à valeurs dans les quadrangles convexes sont d'ordre 4, ainsi l'élément fini considéré (de type Q^1) possède une propriété d'*isoparamétrie* qui se traduit par

$$\forall\,\hat x \in \hat K\,, \quad \hat{\mathcal{F}}^l(\hat x) = \sum_{i=1}^4 M_i^l\,\tau_i^l[\hat{\mathcal{F}}^l(\hat x)] = \sum_{i=1}^4 M_i^l\,\hat\tau_i(\hat x)\,,$$

pour toute transformation $\hat{\mathcal{F}}^l$ faisant passer de l'élément de référence $\hat K$ à l'élément K^l du domaine physique, où les $(M_i^l)_i$ sont les (coordonnées des) sommets de K^l. Il en résulte, en désignant par J^l la matrice jacobienne de cette transformation :

$$\mathbb{A}_{ij}^l = \int_{\hat K} [\hat\nabla\hat\tau_i(\hat x)]^T\,[J^l(\hat x)]^{-1}\,[J^l(\hat x)]^{-T}\,[\hat\nabla\hat\tau_j(\hat x)]\,|\operatorname{Det} J^l(\hat x)|\,d\hat x\,.$$

Sur maillage fixe, ce genre de formule a pour avantage de ne devoir être calculée qu'une fois dans le cas des opérateurs linéaires, tandis pour les opérateurs non linéaires on se contente des données $J^{-T}\ \hat\nabla\hat\tau_i$ et $\operatorname{Det} J$. Enfin, sur maillage mobile, J change en permanence, si bien que les seuls invariants restants sur un élément fini donné sont les *numéros* $I = \ell_g(l,i)$ des nœuds qu'il contient ; la matrice $\hat\nabla\hat\tau_i$, quant à elle, ne varie pas non plus mais n'est pas spécifique à l'élément.

Ainsi, la structure minimale d'un élément fini peut être établie à partir d'une simple liste de numéros globaux de points, rangés dans l'ordre $(I_1, .., I_4) = (\ell_g(l,1), .., \ell_g(l,4))$ pour pouvoir relier chacun à la fonction de base locale $\hat{\tau}_i$ qui lui est associée. Pour cela, le plus simple est de construire un objet *Fctform* contenant les fonctions $\hat{\tau}_i$ calculées une fois pour toutes. D'un point de vue informatique, on englobe l'objet *Fctform* et les objets et *Elemfi* (éléments finis volumiques) *Elembd* (éléments finis surfaciques) dans l'objet *Interpol* :

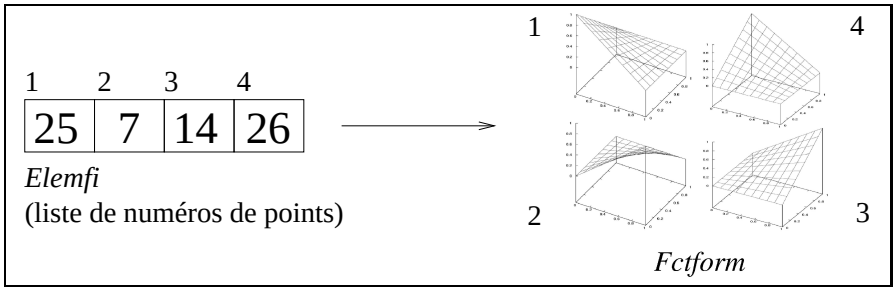


FIG. 6.5 – Structure de données *Interpol*

Enfin, rappelons que dans le problème étudié, plusieurs inconnues entrent en compte : il peut donc être nécessaire de définir plusieurs types d'interpolation par élément géométrique (typiquement pour respecter la condition inf-sup dans le cadre d'une formulation non stabilisée, cf. **3.3**), à savoir plusieurs listes de numéros globaux de points (qui peuvent être fictifs dans le cas des éléments finis non lagrangiens comme P^1 sur quadrangle par exemple). En réalité, on peut montrer que dans le cas de la MHD, l'interpolation optimale du champ magnétique n'est autre que celle appliquée à la vitesse (cf. Gerbeau *et al.* [39]), si bien qu'un élément fini dans Mistral est au plus constitué de *deux* listes de points (une seule dans le cas Q^1-stablisé). L'objet *Fctform* doit donc contenir deux types d'interpolation sur l'élément de référence.

6.3.1.c Structure matricielle

En décomposant les inconnues *et* l'action des opérateurs sur une base de fonctions de support limité (aux éléments contenant le point rattaché à la fonction de forme) , la méthode des éléments finis conduit à une *localisation en espace* des EDP, dont une conséquence directe est le faible taux de remplissage des matrices M et K (cf. **2.2.1.b**). Ainsi, la matrice A possède une structure *creuse*, qu'on exploite pour optimiser son occupation mémoire. On utilise ainsi le *format creux* de matrice le plus répandu, à savoir le format CSR (pour *Compressed Sparse Row*), qui consiste à stocker ligne par ligne les éléments non nuls de la matrice. Ainsi, à partir de deux tableaux d'indices I_A et J_A et d'un tableau de valeurs (non nulles) V_A, on a

$$A_{IJ} = V_A[K]\,, \quad \text{avec} \quad J = J_A[K]\,, \quad I_A[I] \leq K < I_A[I+1]\,.$$

Exemple : $A = \begin{bmatrix} 0. & 1. & 0. & 2. \\ 3. & 0. & 0. & 4. \\ 0. & 0. & 5. & 0. \\ 6. & 7. & 0. & 8. \end{bmatrix}$ donne $\left\{ \begin{array}{rcl} I_A &=& \boxed{1}\boxed{3}\boxed{5}\boxed{6}\boxed{9} \\ J_A &=& \boxed{2}\boxed{4}\boxed{1}\boxed{4}\boxed{3}\boxed{1}\boxed{2}\boxed{4} \\ V_A &=& \boxed{1.}\boxed{2.}\boxed{3.}\boxed{4.}\boxed{5.}\boxed{6.}\boxed{7.}\boxed{8.} \end{array} \right.$

I_A est donc de taille $N_h + 1$ tandis que J_A et V_A sont de taille N_Z, où N_Z est le nombre de valeurs non nulles dans la matrice. Ces tableaux sont regroupés dans la structure *Matrice*.

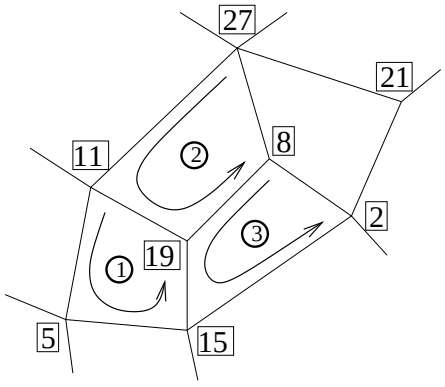
Il existe de nombreuses librairies de solveurs itératifs prenant en entrée ce format (nous utilisons pour notre part *IML++* [27]). L'initialisation de la structure matricielle consiste alors à construire les tableaux I_A et J_A (objet *DegLib*) à partir de la *connectivité* du maillage - à savoir la connaissance, pour un point donné, de l'ensemble des points avec lesquels il interagit - et du type d'interpolation choisi, qui donne la liste des points par élément (sommets ou autres). Elle se base alors sur la recherche des interactions non nulles entre les *degrés de liberté*, qui sont les points du maillage en lesquels aucune condition de Dirichlet n'est imposée (fonction *dirichlet()*). Le problème variationnel discrétisé, en effet, peut prendre deux formes possibles lorsqu'une partie D des composantes de l'inconnue x se trouve contrainte par des conditions aux limites de type Dirichlet :

$$\left[\begin{array}{c|c} I_D & 0 \\ \hline & \\ A_D & A_{\overline{D}} \end{array}\right] \left[\begin{array}{c} x_D \\ \hline \\ x_{\overline{D}} \end{array}\right] = \left[\begin{array}{c} x_D \\ \hline \\ b_{\overline{D}} \end{array}\right] \quad \text{ou} \quad \left[\begin{array}{c} \\ A_{\overline{D}} \\ \\ \end{array}\right] \left[x_{\overline{D}}\right] = \left[\begin{array}{c} \\ b_{\overline{D}} - A_D\, x_D \\ \\ \end{array}\right] .$$

Dans Mistral, on retient la deuxième approche, d'où la définition. On l'étend à des problèmes multi-inconnues en attribuant un numéro non nul I à toute combinaison $(k, \ell_g(l,i))$ où la $k^{\text{ème}}$ inconnue n'est pas éliminée par une condition de Dirichlet au point $\ell_g(l,i)$ (attention : l'application ℓ_g n'est pas injective, donc ce point peut être déjà traité). On définit à cette fin l'application :

$$\ell_d \quad : \quad \begin{array}{rcl} \{1,..,n_{\text{inc}}\} \times \{1,..,L_h\} \times \{1,..,n_k\} & \rightarrow & \{0,..,N\} \\ (k,l,i) & \mapsto & \left\{ \begin{array}{ll} I & \text{si } (k,\ell_g(l,i)) \text{ est un degré de liberté} \\ 0 & \text{sinon} \end{array} \right. \end{array}$$

où n_{inc} est le nombre d'inconnues scalaires, n_k le nombre de points d'interpolation par élément pour l'inconnue k, et N non plus le nombre de points du maillage mais le nombre de degrés de liberté. Parallèlement à la définition de ces derniers, il est également nécessaire d'établir la connectivité du maillage, qui consiste à parcourir les éléments volumiques du maillage, en enregistrant les numéros globaux de tous les points de l'élément courant dans des listes associées à ces points, sauf lorsqu'ils ont déjà été répertoriés par un passage antérieur sur un autre élément :



Point	Liste des voisins
2	8 19 15 2
5	11 5 15 19
8	27 11 19 8 15 2
11	11 5 15 19 27 8
15	11 5 15 19 8 2
19	11 5 15 19 27 8 2
21	
27	27 11 19 8

FIG. 6.6 – Établissement de la connectivité du maillage

Si plusieurs interpolations sont utilisées, il faut établir la connectivité de toutes les couples d'éléments finis possibles (quatre au maximum dans Mistral), de manière à pouvoir prendre en compte les opérateurs variationnels mixtes comme l'incompressibilité ou le gradient de pression.

Enfin, on est en mesure de construire les vecteurs I_A et J_A à partir de deux tableaux temporaires $K[I]$ et $C[I,K]$, où $K[I]$ est la position de J dans la liste des indices de colonne $C[I,\cdot]$ rattachés à la ligne I. Sur un maillage de N_h points, et dans le cas où la même interpolation est appliquée à toutes les inconnues, on effectue (où $(\nu_p(m))_{1\leq m\leq n_p^\nu}$ est la liste des voisins du point p) :

```
1   N_Z ← 0
2   ∀ I ∈ {1,..,N} , K[I] ← 1
3   Pour p = 1 ... N_h :
4   |   Pour k_i = 1 ... n_inc :
5   |   |   I ← ℓ_d(k_i, p)
6   |   |   Si I ≠ 0 :
7   |   |   |   Pour m = 1 ... n_p^ν :
8   |   |   |   |   Pour k_j = 1 ... n_inc :
9   |   |   |   |   |   J ← ℓ_d(k_j, ν_p(m))
10  |   |   |   |   |   Si J ≠ 0 :
11  |   |   |   |   |   |   C[I, K[I]] ← J
12  |   |   |   |   |   |   K[I] ← K[I] + 1
13  |   |   |   |   |   |   N_Z ← N_Z + 1
14  |   |   |   |   |   Fin si
15  |   |   |   |   Fin pour
16  |   |   |   Fin pour
17  |   |   Fin si
18  |   Fin pour
19  Fin pour
```

6.3.2 Assemblage et résolution

Nous avons vu en **6.3.1.b** que la prise en compte des opérateurs se faisait élément par élément, par l'intermédiaire des *matrices élémentaires* $\mathbb{A}^l = (\mathbb{A}^l_{ij})_{i,j}$. L'objet de l'assemblage est d'incorporer ces dernières dans la matrice globale A à l'endroit qui leur correspond. On utilise pour cela une généralisation de la propriété (2.18) (cf. **2.2.1.b**) :

$$w_I^k = \sum_{\ell_d(k,(l,i))=I} \tau_i^l$$

au cas d'un problème à n_{inc} inconnues scalaires, dont chacune est interpolée par un élément fini à n_k fonctions de forme, $1 \leq k \leq n_{\text{inc}}$. Ainsi

$$\mathbb{A}^l = \left(\left(\left(\mathbb{A}^l_{i,j} \right)_{1\leq i\leq n_{k_i},\, 1\leq j\leq n_{k_j}} \right)_{k_i,k_j} \right)_{1\leq k_i,k_j\leq n_{\text{inc}}}$$

désigne la contribution au problème de l'élément K^l, où $\mathbb{A}^l_{i,j,k_i,k_j}$ représente l'action de l'équation k_i au point (l,i) sur l'inconnue k_j au point (l,j). On suppose de plus que $n_{\text{op}}^{k_i,k_j}$ opérateurs sont appliqués dans l'équation k_i à l'inconnue k_j :

$$\mathbb{A}^l_{i,j,k_i,k_j} = \sum_{m=1}^{n_{\text{op}}^{k_i,k_j}} \mathbb{O}^{l,m}_{i,j,k_i,k_j} \,.$$

Il reste à définir les vecteurs élémentaires $\mathbb{B}^l_{i,k_i}$, constitués de la somme de $n_{\text{f}}^{k_i}$ termes de données (ou inconnues à l'itération précédente) notés $\mathbb{F}^{l,q}_{i,k_i}$, pour écrire l'algorithme d'assemblage.

6.3.2.a Assemblage

À chaque pas de temps ou itération non linéaire, le programme parcourt les éléments. Sur chacun d'entre eux, il calcule :

1. les outils nécessaires à la transformation géométrique,
2. la matrice élémentaire en tenant compte de l'ensemble des opérateurs,
3. le second membre élémentaire en tenant compte de toutes les données,

puis assemble la matrice et le vecteur ainsi obtenus dans la matrice et le second membre globaux $A_{\overline{D}}$ et $b_{\overline{D}}$ (qu'on notera encore A et b dans toute la suite) par l'intermédiaire de l'application ℓ_d :

$A \leftarrow 0,\ b \leftarrow 0$
Pour $l = 1...L_h$:
| *1) Calcul de la matrice et du second membre élémentaires :*
| Calculer $[J^l]^{-T}$ et $|\mathrm{Det}\, J^l|$
| Pour $k_i = 1...n_{\mathrm{inc}}$:
| | Pour $k_j = 1...n_{\mathrm{inc}}$:
| | | Pour $p = 1...n_{\mathrm{op}}^{k_i,k_j}$:
| | | | Pour $i = 1...n_{k_i}$ et $j = 1...n_{k_j}$:
| | | | | $\mathbb{A}^l_{i,j,k_i,k_j} \leftarrow \mathbb{A}^l_{i,j,k_i,k_j} + \mathbb{O}^{l,p}_{i,j,k_i,k_j}$
| | | | Fin pour
| | | Fin pour
| | Fin pour
| | Pour $q = 1...n_{\mathrm{f}}^{k_i}$:
| | | Pour $i = 1...n_{k_i}$:
| | | | $\mathbb{B}^l_{i,k_i} \leftarrow \mathbb{B}^l_{i,k_i} + \mathbb{F}^{l,q}_{i,k_i}$
| | | Fin pour
| | Fin pour
| Fin pour
| *2) Assemblage :*
| Pour $k_i = 1...n_{\mathrm{inc}}$ et $i = 1...n_{k_i}$:
| | $I \leftarrow \ell_d(k_i,(l,i))$
| | Si I est un degré de liberté :
| | | Pour $k_j = 1...n_{\mathrm{inc}}$ et $j = 1...n_{k_j}$:
| | | | $J \leftarrow \ell_d(k_j,(l,j))$
| | | | Si J est un degré de liberté :
| | | | | $A_{IJ} \leftarrow A_{IJ} + \mathbb{A}^l_{i,j,k_i,k_j}$
| | | | sinon :
| | | | | $b_I \leftarrow b_I - \mathbb{A}^l_{i,j,k_i,k_j}\, X_D(J)$
| | | | Fin si
| | | Fin pour
| | | $b_I \leftarrow b_I + \mathbb{B}^l_{i,k_i}$
| | Fin si
| Fin pour
Fin pour

Matrice et vecteur élémentaires sont contenus dans un même objet *Tabelem*, réinitialisé à zéro sur chaque élément. La fonction *assemble()* assure le transfert dans la structure {*Matrice, Vecteur*}.

6.3.2.b Résolution du système $Ax = b$

On note $x = x_{\overline{D}}$ dans toute la suite.

Nous avons mentionné en section **2.2** que les méthodes directes classiques étaient mal adaptées à la structure creuse de la matrice A, d'où l'utilisation de méthodes itératives du type (2.19) :

$$x_{k+1} \leftarrow x_k + \alpha\, r_k\,, \quad \alpha \text{ fixé } \geq 0 \tag{6.1}$$

où l'on a pris $P = I_N$ (problème non préconditionné) pour nous concentrer sur la méthode itérative en elle-même. Lorsque A est symétrique définie positive, l'algorithme (6.1) n'est autre que la méthode du *gradient à pas fixe*, en effet

$$r_k = -\nabla J(x_k) \quad \text{avec} \quad J(x) = \frac{1}{2}(Ax, x) - (b, x) \quad (\text{fonction convexe dans } \mathbb{R}^N\,).$$

Il est possible d'améliorer cette méthode par deux moyens. Le premier est d'adapter le paramètre α à chaque itération en résolvant le sous-problème d'optimisation (problème de *recherche linéaire*) :

$$\inf_{\alpha} J(x_k + \alpha\, r_k)\,,$$

ce qui conduit la méthode du *gradient à pas optimal* :

$$\begin{array}{|l} \\ \\ \\ \\ \end{array} \begin{array}{l} \text{Calculer } Ar_k \\ \alpha_k \leftarrow (r_k, r_k)/(Ar_k, r_k) \\ x_{k+1} \leftarrow x_k + \alpha_k\, r_k \\ r_{k+1} \leftarrow r_k - \alpha_k Ar_k \end{array}$$

On a minimisé J sur la droite $x(t) = x_k + tr_k$ à chaque itération, et obtenu ainsi une solution à l'itération k dans l'*espace de Krylov*

$$V_k = \underset{0\leq j\leq k-1}{\text{Vect}}\,(A^j r_0)\,,$$

mais on n'a pas optimisé sur *tout* l'espace V_k. C'est pourquoi on a recours à la deuxième technique qui est une modification de la direction de descente. Pour cela, on s'arrange pour que le résidu r_k se trouve dans un espace orthogonal à V_k en actualisant l'inconnue par

$$x_{k+1} \leftarrow x_k + \alpha_k\, p_k\,, \quad \text{avec} \quad (p_k, Ap_{k-1}) = 0\,, \quad \forall\, k \geq 1\,,$$

avec $p_0 = r_0$. C'est ainsi qu'on obtient la méthode du *gradient conjugué*, qui nous sera utile pour le problème de déplacement du maillage (3.21) par exemple.

Concernant les systèmes non symétriques, comme $Ax = b$ dès lors que des termes non linéaires sont pris en compte dans le modèle, on a recours à des méthodes un peu plus compliquées, mais du même type. Parmi celles-là, la méthode GMRES (Y. Saad et M.H. Schultz [86]) consiste à rechercher la solution dans l'espace V_k, et son résidu orthogonal à AV_k. La solution x_k jouit ainsi d'une propriété d'optimalité sur le sous-espace affine $r_0 + V_k$, et donc converge en N itérations exactement. En pratique, très peu d'itérations suffisent pour obtenir une approximation satisfaisante de x pour peu que le problème soit bien préconditionné (cf. **2.2.1.c**). On observe que le préconditionneur ILU(0) (*Incomplete LU*, J.A. Meijerink and H.A. van der Vorst [70]) remplit cette condition. Il consiste à utiliser $P = \tilde{L}\tilde{U}$, où $\tilde{L}$ (resp. $\tilde{U}$) est la matrice triangulaire inférieure (resp. supérieure) issue de l'algorithme du pivot de Gauss appliqué à A, dans laquelle les éléments ne se situant pas dans la structure de A sont annulés.

6.3.2.c Discrétisation en temps et traitement des termes non linéaires

Nous avons vu en section **2.2.1.a** que le schéma d'Euler implicite était inconditionnellement stable pour l'équation de la chaleur. On peut en dire la même chose au sujet du problème parabolique que constituent les équations de la MHD, mais son emploi dans ce cas peut s'avérer onéreux du fait de la présence de non-linéarités telles que la convection, la force de Lorentz et la loi d'Ohm. L'algorithme explicite, quant à lui, contraint les paramètres h et Δt à satisfaire une forme de condition CFL qu'on ne connaît pas. Il existe un compromis intéressant, appelé schéma semi-implicite, basé sur le "gel" de l'une des deux inconnues intervenant dans chaque opérateur non linéaire : on montre que l'algorithme monolithique (couplé) qu'on écrit sous forme forte

$$\left\{\begin{array}{rcl} \dfrac{v^{n+1}-v^n}{\Delta t} + v^n\cdot\nabla v^{n+1} - \dfrac{1}{Re}\Delta v^{n+1} + \nabla p^{n+1} - S\,\mathrm{rot}\,B^{n+1}\times B^n &=& \dfrac{1}{Fr}\,e_z \\ \mathrm{div}\,v^{n+1} &=& 0 \\ \dfrac{B^{n+1}-B^n}{\Delta t} + \dfrac{1}{Rm}\,\mathrm{rot}\,\mathrm{rot}\,B^{n+1} - \mathrm{rot}\,(v^{n+1}\times B^n) &=& 0 \\ \mathrm{div}\,B^{n+1} &=& 0 \end{array}\right.$$

est inconditionnellement stable et satisfait la propriété (3.23) pour le problème monofluide.

Informatiquement parlant, c'est la fonction *avance()* qui se charge, après chaque itération, d'affecter $(v^{n+1}, B^{n+1}, p^{n+1})$ à (v^n, B^n, p^n), puis de réinitialiser A, A_D et b.

6.3.2.d Formulation ALE

L'implémentation de la formulation ALE constiste à diviser l'assemblage en deux phases séparées par une itération de déplacement du maillage (on résout le problème (3.21) dans une structure matricielle indépendante par la fonction *calc_vit_maillage()*), de manière à pouvoir prendre en compte les termes $u^n/\Delta t$ et $B^n/\Delta t$ sur les bons domaines. Pour que la résolution du problème génère une solution respectant la conservation de la masse de chaque fluide, il reste à tenir compte de la contrainte (3.24). Si l'on veut pouvoir utiliser les éléments finis Q^1-stabilisés, on n'a d'autre choix que d'introduire un multiplicateur de Lagrange dans l'équation, car l'interpolation Q^1 en pression ne comporte pas la fonction indicatrice de Ω_1. Pour cela, on augmente la taille de l'inconnue (du type *Vecteur*) d'un degré de liberté, et on ajoute une ligne (Φ) et une colonne (Φ^T) à la matrice A, qu'on assemble séparément du reste à partir des contributions élémentaires :

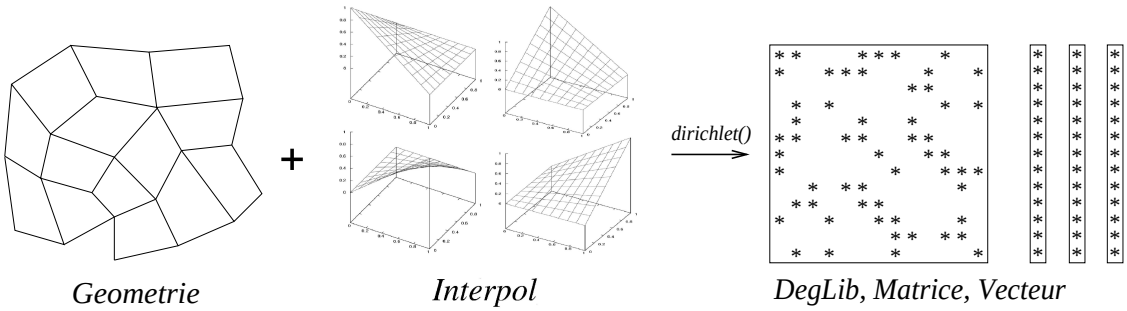
$$\Big(\int_{K^l}\partial\tau^l_{i,1}\Big)_{1\le i\le n_1} \quad \text{pour} \quad \text{pour tout } l \text{ tel que } K^l\in\Omega_1\,.$$

D'un point de vue informatique, on définit un nouvel objet pour pouvoir construire le système linéaire $A'x' = b'$ tel que

$$A' = \left[\begin{array}{c|c} & \\ A & \Phi^T \\ & \\ \hline & - \\ \Phi & 0 \end{array}\right], \quad x' = \left[\begin{array}{c} x \\ \\ \hline - \\ \lambda \end{array}\right], \quad \text{et} \quad b' = \left[\begin{array}{c} b - A_D x_D \\ \\ \hline - \\ 0 \end{array}\right]. \tag{6.2}$$

On applique alors GMRES/ILU(0) (librairies *IML++* [27] et *SparseLib++* [69]) à ce nouveau système par l'intermédiaire de la fonction *resol()*.

1. Initialisation



2. Assemblage et résolution de $A'x' = b'$

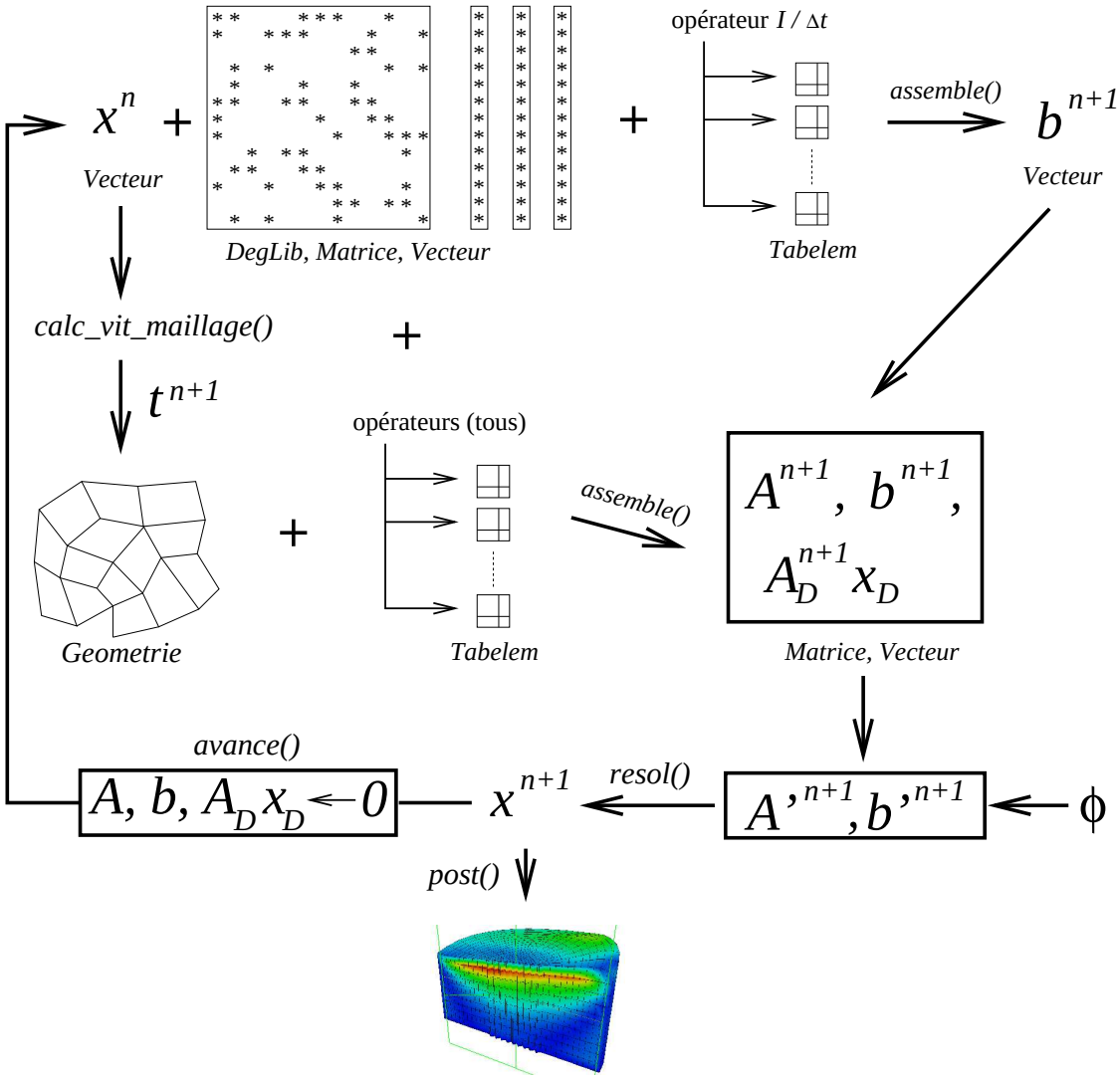


FIG. 6.7 – Résumé de la méthode

6.4 Fonctionnement de la librairie parallèle

La librairie Aztec, qui utilise *MPI* [71] version langage C, est un outil spécialement conçu pour la résolution parallèle des systèmes linéaires rencontrés en simulation numérique. Elle propose des solveurs de type Krylov (gradient conjugué, GMRES, etc.) adaptés à des matrices partitionnées sur différents processeurs, ainsi que des préconditionneurs parallèles efficaces pour les problèmes à granularité importante (cf. **6.2.1**), tout en gérant en interne les communications dans l'exécution des méthodes. C'est une suite d'utilitaires purement algébriques ou informatiques, qui ne requiert en entrée que la matrice et le second membre partitionnés dans un format spécifique. Nous résumons ici son fonctionnement, dont le détail pourra être trouvé dans son manuel de référence ([97]).

6.4.1 Données de base

Dans une configuration SPMD (cf. **6.2.2**) telle que MPI, on exécute le même programme sur N_P processeurs par la commande

`mpirun -np` N_P `-machinefile` <*liste des ordinateurs*> <*exécutable*> ,

qui génère un communicateur *MPI_COMM_WORLD* sur l'ensemble des programmes faisant appel à la fonction *MPI_Init()*. Un programme qui appelle ensuite *MPI_Finalize()* désactive le communicateur, mais peut le réactiver à tout moment par un nouvel appel à *MPI_Init()*. Alors, les fonctions minimales pour faire communiquer les processeurs sont au nombre de quatre. Il s'agit de :

- *MPI_Comm_Size()* qui donne le nombre de processeurs,
- *MPI_Comm_Rank()* qui donne l'identifiant (rang du processeur dans la liste $\{0, .., N_P - 1\}$),
- *MPI_Send()* pour envoyer des messages,
- *MPI_Receive()* pour recevoir des messages.

Un programme utilisant Aztec commence ainsi par appeler *MPI_Init()*, puit s'appuie sur diverses fonctions prédéfinies appelant notamment les quatre ci-dessus.

6.4.2 Format des sous-matrices

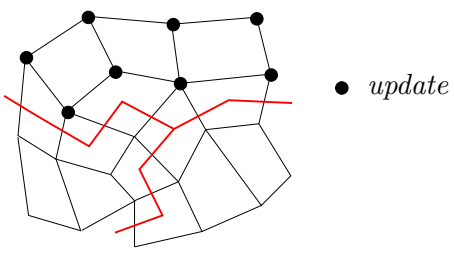
6.4.2.a Tableau *update*

Le principe de partitionnement des matrices dans *Aztec* est d'affecter un certain nombre de lignes de la matrice globale (sans qu'elle existe physiquement) à chaque processeur. La fonction *AZ_read_update()* permet à l'utilisateur de spécifier les numéros des lignes qu'il attribue à chaque processeur, par l'intermédiaire d'un fichier d'entrée contenant N_P doubles lignes du type

N_p (*Nombre de lignes affectées au processeur p*)

I_p^1, I_p^2, .. (*Indices correspondants rangés par ordre croissant*)

Ainsi, chaque processeur est doté d'un tableau d'entiers *update* contenant les numéros des lignes dont il a la charge, et le sous-système linéaire (matrice, inconnue et second membre) correspondant à *update* (voir ci-dessous). Par exemple, si l'on s'arrange, dans un problème à une seule inconnue scalaire, pour que les points du maillage soient numérotés de sorte qu'ils identifient des degrés de liberté, *update* sera constitué de numéros de ces points. On représente ci-dessous le contenu de *update* sur le sous-domaine supérieur d'un maillage partitionné en trois :

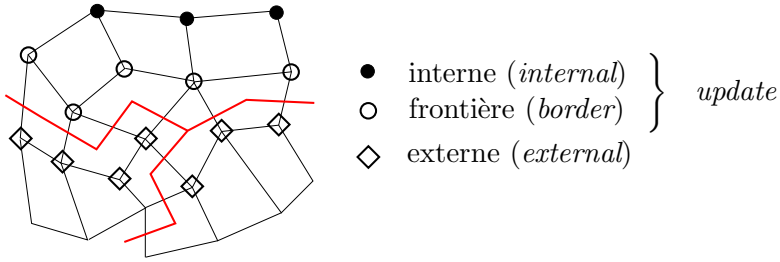


6.4.2.b Fonction *AZ_transform()*

Nous avons vu en **6.3.2.b** que les méthodes de Krylov s'appuyaient sur des produits matrice-vecteur globaux. C'est la raison pour laquelle chaque sous-matrice est constituée de lignes entières. On est alors amené à distinguer, parmi l'ensembles des indices de degrés de liberté qu'elle met en jeu (lignes et colonnes) :

- les indices *externes* : indices de colonne qui ne sont pas des indices de ligne,
- les indices *frontière* : indices des lignes où figurent des éléments d'indice externe
- les indices *internes* : indices des lignes ne contenant aucun élément d'indice externe.

Dans l'exemple ci-dessus, cela se traduit comme suit au niveau du maillage :



On peut ainsi savoir quelles sont les lignes qui nécessitent une communication avec un autre processeur dans le produit matrice-vecteur. La librairie Aztec, pour optimiser ce dernier, réordonne les lignes de la sous-matrice, de manière à ce que les lignes d'indice interne figurent au-dessus des lignes d'indice frontière. Ce réarrangement génère un certain nombre de données nouvelles, dont les paramètres sont rangés dans le tableau *data_org*. En résumé, une renumérotation locale de 0 à $M-1$ (où M est le nombre d'indices différents, y compris externes), est mise en place, de sorte que les indices de *border* succèdent aux indices *internal* dans *update*, et que les indices de *external* succèdent eux-mêmes aux indices de *update*. Pour ce faire, l'utilisateur doit fournir en entrée la sous-matrice au format MSR (*Modified Sparse Row*), basé sur un seul vecteur d'indices (globaux) *bindx* et un tableau de valeurs *val* tels que :

- les éléments diagonaux a_{II} soient rangés dans les N_P premières cases de *val*,
- les éléments $a_{I \neq J}$ non nuls soient rangés à partir de la position (N_P+1) dans *val* (la première position étant 0), en parcourant a ligne par ligne,
- $bindx[0] = N_P$, $bindx[k+1]-bindx[k]$ soit égal au nombre d'éléments non nuls *hors diagonale* dans la ligne $k \in \{0,..,N_P-1\}$, et $bindx[k]$ soit l'indice de colonne de l'élément $val[k]$.

À partir de là, Aztec est capable de générer son format interne DMSR (*Distributed* MSR) par le biais de la fonction *AZ_transform()*. Celle-là renvoie les tableaux *internal*, *border* et *external*, et les redirections *update_index* et *extern_index* de sorte que l'indice *update[i]* (resp. *external[i]*) soit représenté à la position *update_index[i]* (resp. *extern_index[i]*) dans la numérotation locale.

Exemple : soit une matrice creuse 15x15 répartie sur 3 processeurs, de sorte qu'au processeur i soient affectées les lignes $5*i$ à $5*(i+1)-1$. On s'intéresse au processeur 1, dont la sous-matrice présente le profil suivant :

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
5			*			*		*						*	
6							*	*							
7	*		*		*								*		
8								*		*					
9					*		*			*					*

Le tableau *bindx* de son format MSR prend la forme :

6	9	10	14	16	19	2	7	13	7	0	2	4	12	7	9	4	6	14

.

De *bindx* et de *update* = | 5 | 6 | 7 | 8 | 9 |, la fonction *AZ_transform()* déduit les tableaux

internal = | 6 | 8 |, *border* = | 5 | 7 | 9 |, et *external* = | 0 | 2 | 4 | 12 | 13 | 14 |,

qui engendrent la numérotation locale | 6 | 8 | 5 | 7 | 9 | 0 | 2 | 4 | 12 | 13 | 14 |, soit

update_index = | 2 | 0 | 3 | 1 | 4 | et *extern_index* = | 5 | 6 | 7 | 8 | 9 | 10 |.

La sous-matrice devient donc (on notera que les premières lignes sont bien d'indices internes) :

	0	1	2	3	4	5	6	7	8	9	10
0	*			*							
1				*	*						
2			*	*			*			*	
3						*	*	*	*		
4	*				*			*			*

dont le tableau *bindx* (format DMSR) s'écrit :

6	7	9	12	16	19	3	3	4	3	6	9	5	6	7	8	0	7	10

.

6.4.3 Résolution parallèle

Une fois la matrice locale initialisée et transformée, il reste à lancer la fonction *AZ_solve()*, en ayant au préalable distribué (*MPI_Scatterv()*) puis renuméroté (*AZ_reorder_vec()*) le membre de droite et la donnée initiale. La solution globale est alors obtenue en appliquant l'opération inverse aux solutions locales (*AZ_invorder_vec()* puis *MPI_Gatherv()*).

Dans cet esprit, c'est un système linéaire global qui est résolu simultanément par tous les processeurs. Du point de vue de la méthode itérative en elle-même, sa version parallélisée ne présente pas de différences avec sa version séquentielle, si ce n'est bien sûr au niveau informatique, où les produits matrices-vecteurs globaux sont localement effectués. Pour cela, chaque processeur reçoit des autres les composantes externes du vecteur qu'il doit multiplier, et donc envoie également les composantes frontière de son vecteur local aux processeurs qui en ont besoin. Toute la difficulté repose en fait sur le choix du préconditionneur, dont la nature globale (élimination de Gauss par exemple), en revanche, peut détériorer le gain procuré par la parallélisation. En effet, le préconditionneur doit, au même titre que la matrice A, être localisé sur chaque processeur pour pouvoir être utilisé. Deux possibilités se présentent alors :

- la méthode *globale* : un seul processeur construit le préconditionneur, puis distribue ses restrictions en correspondance avec les sous-matrices ,
- le *préconditionnement par décomposition de domaine* : chaque processeur calcule un préconditionneur local à partir de sa sous-matrice, et éventuellement par *recouvrement* .

A priori, l'option globale est la meilleure, car le problème à résoudre est de nature globale. En fait, la perte de temps occasionnée par le calcul séquentiel du préconditionneur global et sa distribution peut être bien plus grande que celle issue de sa localisation, qui permet un plus haut niveau de parallélisme. Ainsi, après un bref descriptif des options proposées par *AZ_solve()*, nous détaillons un peu plus la question du préconditionnement par décomposition de domaine.

6.4.3.a Fonction *AZ_solve()*

Aztec propose de nombreuses méthodes itératives de type Krylov comme CG, GMRES, CGS, TFQMR et BiCGSstab notamment (pour le détail desquelles on renvoie à Y. Saad [85]), ainsi qu'un large choix de techniques de préconditionnement, des plus classiques (Jacobi, Jacobi par blocs) jusqu'aux plus abouties comme les méthodes de décomposition de type Schwarz-additif, que nous aborderons à la section suivante. Il fournit également plusieurs normes possibles pour l'évalution du résidu, à savoir la norme absolue et diverses normes relatives. On peut ainsi résumer les arguments de la fonction *AZ_solve()* par les données relatives à la configuration locale *proc_config*, *update*, *data_org*, *bindx*, *val*, et les trois tableaux supplémentaires *options*, *params* et *status* décrivant respectivement le choix du solveur et du préconditionneur, les critères de convergence, et (en sortie) le bilan de la résolution. On dispose par ailleurs d'une option d'échelonnement (*scaling*) de la matrice globale, qui consiste à transformer le système $Ax = b$ en $SAx = Sb$, de manière à atténuer des différences d'échelle importantes d'un élément à l'autre. Ce cas peut se présenter lorsqu'on traite certaines conditions aux limites par pénalisation. Un autre outil intéressant est l'utilisation d'un même préconditionnement sur plusieurs résolutions successives, dont le nombre est à régler avec précaution. Cela peut s'avérer utile pour les problèmes transitoires, moyennant un pas de temps suffisamment petit pour que la matrice ne varie pas de façon disproportionnée.

6.4.3.b Préconditionneur par décomposition de domaine (Schwarz-additif)

Nous avons vu en **2.2** que pour offrir un taux de convergence acceptable, les méthodes itératives devaient être préconditionnées. Cela signifie, en un sens assez général, que l'algorithme effectue des opérations nécessitant la connaissance de la solution y du système

$$Py = z\,,$$

où z est par exemple le membre de droite, le résidu ou encore un produit matrice-vecteur impliquant A. Il faut donc que cette opération soit assez simple, étant donné qu'on l'utilise au moins une fois à chaque itération. En clair, on ne calcule pas explicitement P^{-1}, mais on se sert de son contenu pour transformer z jusqu'à obtenir y. La notation (2.19) est donc un peu ambigüe, mais fort commode pour décrire synthétiquement les algorithmes. Par exemple, pour le préconditionneur ILU(0) (cf. **6.3.2.b**), y est obtenu à partir de z par une descente-remontée du type :

$$\left.\begin{array}{rcl}\tilde{L}\bar{y} &=& z\\ \tilde{U}y &=& \bar{y}\end{array}\right\}\,,\quad \text{soit}\quad \left\{\begin{array}{rcll}\bar{y}_I &\leftarrow& z_I - \displaystyle\sum_{J=1}^{I-1}\tilde{L}_{IJ}\,\bar{y}_J & \text{de } I=1 \text{ à } N\,,\\ y_I &\leftarrow& \dfrac{1}{\tilde{U}_{II}}\left(\bar{y}_I - \displaystyle\sum_{J=1}^{I-1}\tilde{U}_{IJ}\,\bar{y}_J\right) & \text{de } I=N \text{ à } 1\,.\end{array}\right.$$

(le calcul préalable de $\tilde{L}$ et $\tilde{U}$ à partir de A étant raisonnable). On peut tranposer cette propriété au niveau local, en construisant un préconditionneur global constitué de la somme (virtuelle) des extensions à $\mathcal{M}_N(\mathbb{R})$ des préconditionneurs ILU(0) locaux des sous-matrices restreintes à *update*. Cette technique porte le nom de *procédure de Schwarz additive* sans *recouvrement*, parce que les sous-préconditionneurs locaux sont calculés indépendamment les uns et des autres, et sur des sous-matrices carrées purement locales (de taille N_p). Le revers de cette parallélisation est une détérioration du préconditionneur global, puisqu'il n'est qu'un assemblage d'informations locales. Pour pallier cet inconvénient, on augmente légèrement la taille des sous-matrices carrées de manière à transmettre une partie de l'information globale dans le préconditionnement. Deux méthodes se présentent alors : faire de même que précédemment, ou calculer les préconditionneurs locaux de

manière dépendante donc séquentielle. La deuxième est la procédure de Schwarz *multiplicative*, et n'est pas adaptée au calcul parallèle : on retombe sur le paradigme entre haut niveau de parallélisme et utilisation d'informations globales. Ainsi, les méthodes de Schwarz additives avec recouvrement offrent un bon compromis à ce niveau. Le préconditionneur obtenu s'écrit

$$P^{-1} = \sum_{p=1}^{N_P} R_p^T P_p^{-1} R_p$$

où $P_p = (\tilde{L}_p \tilde{U}_p)$ est le préconditionneur ILU(0) de la matrice

$$A_p = R_p A R_p^T ,$$

et où R_p est l'opérateur de restriction à la sous-matrice "augmentée" : si on note

$R_p^K = [0 \dots 0\ 1\ 0 \dots 0]$ la ligne de taille N où le 1 est à la position K (la première position étant 0),

on a

$$R_p = \begin{bmatrix} R_p^{K_1} \\ R_p^{K_2} \\ \vdots \\ R_p^{K_{N_p+N_p^0}} \end{bmatrix} \quad \text{tel que} \quad update \subset S_p = \{K_1, .., K_{N_p+N_p^0}\}$$

$(N_p^0 = \text{nombre de lignes/colonnes supplémentaires})$.

Exemple : $N_p^0 = 3$ et

$S_p = \{update[0] - 1, update[0], \dots, update[N_p - 1], update[N_p - 1] + 1, update[N_p - 1] + 2\}$

avec $update = \{I, \dots, I + N_p - 1\}$ (numéros globaux de N_p lignes contigües).

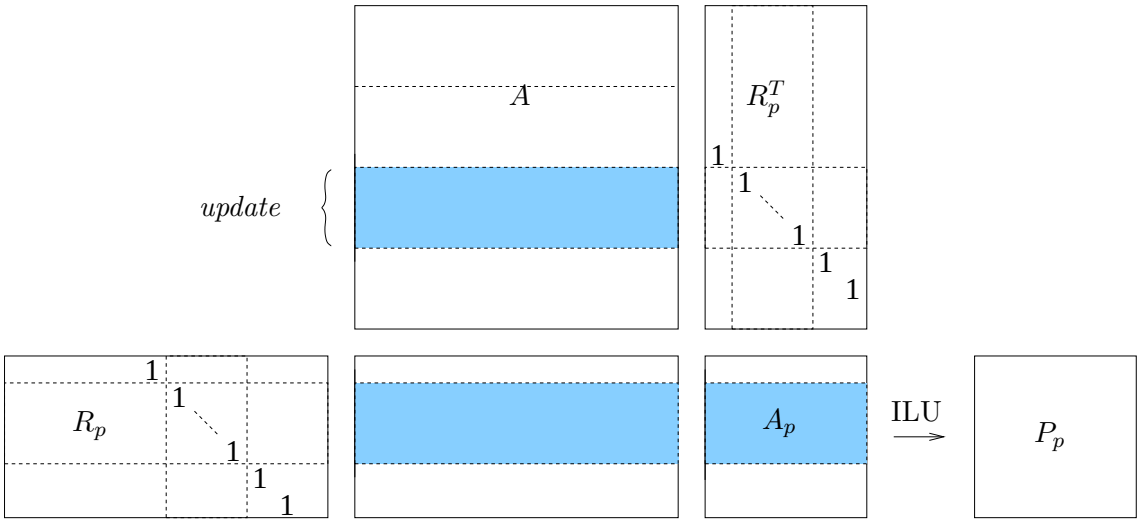


FIG. 6.8 – Restriction et préconditionnement de la matrice locale

Noter qu'une structure diagonale par blocs apparaît dans P^{-1}, parce que nous avons choisi par simplicité de considérer des sous-matrices constituées de lignes adjacentes ; ainsi, il est clair que plus les éléments de A sont concentrés autour de la diagonale, meilleur est le préconditionneur :

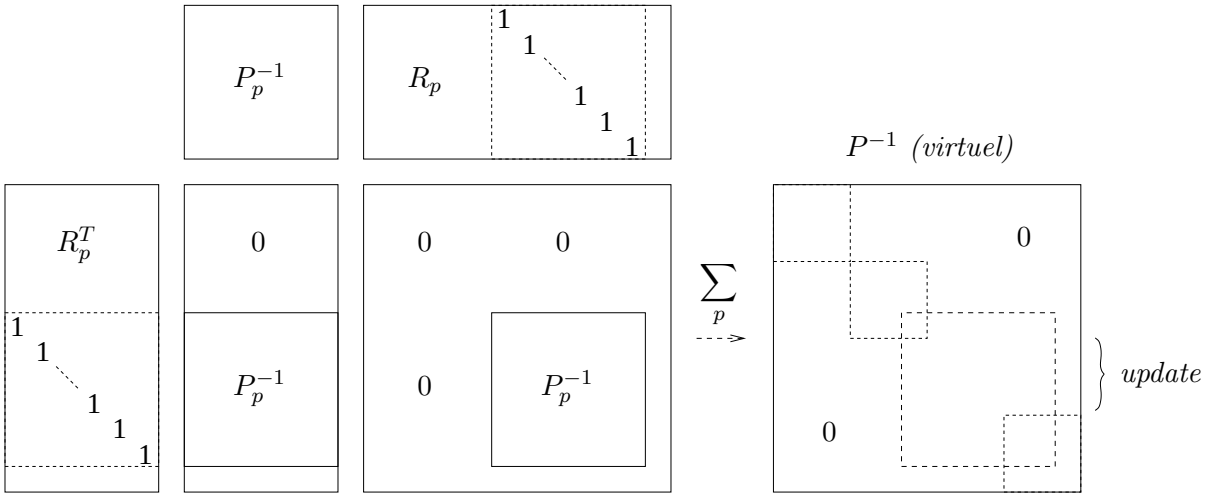


FIG. 6.9 – Prolongement et somme (virtuels) des préconditionneurs locaux

6.5 Parallélisation

Du fait du mouvement du maillage, la procédure d'assemblage coûte cher car il faut recalculer toutes les fonctions de base à chaque pas de temps, et par conséquent les matrices élémentaires de tous les opérateurs, en particulier des opérateurs linéaires (qui ne dépendent pas de la valeur de l'inconnue normalement). C'est la conséquence numérique de la non linéarité *forte* que constitue l'interface libre (cf. **3.1**). On remarque cependant que la procédure d'assemblage est de nature locale (élément par élément), et permet ainsi de restreindre sur chaque processeur les calculs aux seuls éléments contribuant à une sous-matrice donnée. Ayant à disposition une librairie (cf. *supra*) de résolution de systèmes linéaires décomposés ligne par ligne, ainsi qu'une dizaine d'ordinateurs, il paraît judicieux de diviser la taille du système linéaire d'autant. Soulignons que la librairie en question ne résout pas en parallèle plusieurs problèmes indépendants, mais bien un seul problème décomposé, ce qui requiert de disposer de la connectivité globale du maillage. L'arsenal {*Geometrie*, *Interpol*, *DegLib*} construit durant l'initialisation doit donc être conservé, pour servir en quelque sorte de repère aux structures décomposées des objets *Matrice* et *Vecteur*. À cette fin, la meilleure solution semble être de conserver telle quelle la phase d'initialisation, c'est à dire de l'exécuter de manière redondante sur l'ensemble des processeurs. Sa parallélisation engendrerait des surcoûts de communications trop importants, et, de toutes façons, procurerait un gain négligeable dans les problèmes itératifs (puisque l'assemblage et la résolution y sont effectués plusieurs fois).

A priori, on doit donc opérer des modifications essentiellement dans le format de la matrice creuse pour lui appliquer le solveur parallèle, et dans la procédure d'assemblage, qui doit être adaptée à ce format. Il reste alors à optimiser le partitionnement de la matrice pour obtenir les meilleures performances possibles. En fait, une difficulté apparaît dans la prise en compte de deux fluides, du fait de la contrainte de masse (3.24) : rappelons que pour pouvoir utiliser les éléments finis Q^1-stabilisés, il est nécessaire d'ajouter un multiplicateur de Lagrange parmi les inconnues, ce qui a pour conséquence de compliquer la structure matricielle. La solution retenue est de supprimer ce multiplicateur de Lagrange en dupliquant les degrés de liberté en pression sur l'interface.

Nous présentons ici, dans l'ordre chronologique, les trois étapes majeures qui nous ont permis de diviser le temps de calcul quasiment par le nombre de processeurs utilisés, du moins jusqu'à une dizaine d'ordinateurs. Il s'agit de la parallélisation de la méthode itérative, de l'assemblage, et de la prise en compte de degrés de libertés supplémentaires en pression sur l'interface.

6.5.1 Résolution

6.5.1.a Conversions de format

Nous avons vu que la fonction *AZ_solve()* d'Aztec, pour pouvoir résoudre un système linéaire en parallèle, requiert de diviser la matrice globale en sous-matrices au format spécial DMSR (cf. **6.4.2**). Dans un problème sur maillage structuré, on dispose en général d'une règle simple pour établir la contribution d'un point à l'ensemble du maillage, aussi est-il possible de calculer directement les contributions locales au format MSR (local), puis de leur appliquer la fonction *AZ_transform()* pour obtenir le format DMSR. Dans notre cas, chaque ligne de la matrice doit être calculée en fonction de la géométrie spécifique des éléments auxquels le point qu'elle représente appartient. C'est le rôle de la structure *DegLib* que de relier un point donné avec ses voisins (cf. **6.3.1.c**) pour pouvoir calculer les interactions qui en résultent. Ainsi, le maintien de la structure de données globale est nécessaire. À cette fin, la méthode la plus simple semble être de conserver tel quel l'établissement du format CSR (vecteurs I_A, J_A et V_A) sur tous les processeurs (au moins temporairement), de le convertir au format MSR, puis d'en déduire le format MSR local sur chaque processeur en fonction de son tableau *update*, et enfin d'appliquer la fonction *AZ_transform()* :

Il ne reste plus qu'à initialiser l'inconnue, le membre de droite et le vecteur x_D, en divisant et distribuant les originaux globaux par la fonction *MPI_Scatterv()*, puis en les réordonnant par l'équivalent vectoriel de la fonction *AZ_transform()*, à savoir *AZ_reorder_vec()* (cf. **6.4.3**). On dispose alors de toutes les données locales nécessaires au lancement de la fonction *AZ_solve()*, ainsi que des tableaux *update_index*, *extern_index* (qui ne sont pas encore utiles).

Une fois la résolution effectuée, on a besoin de regrouper les solutions locales, en effectuant l'opération inverse (cf. **6.4.3**). On doit de plus réordonner la solution globale, qui ne l'est pas si la répartition n'a pas eu lieu par lignes contigües : on utilise pour cela le regroupement des tableaux *update*, qui donne les indices de l'inconnue globale :

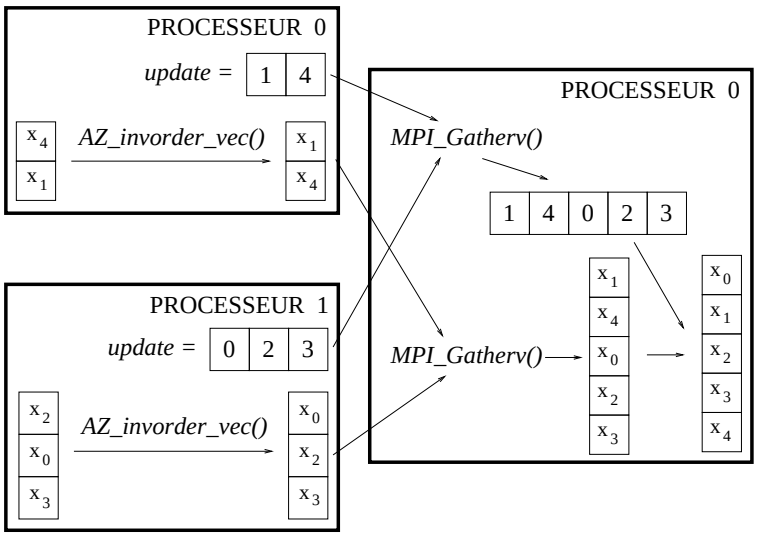


FIG. 6.10 – Exemple : regroupement de la solution sur le processeur 0

6.5.1.b Partitionnement optimal des degrés de liberté

La parallélisation d'un système linéaire génère principalement deux freins à la minimisation du temps de calcul : des communications trop importantes entre les différents processeurs et une perte de la qualité de son préconditionneur. Une bonne manière de limiter ces deux effets néfastes est de répartir les degrés de liberté de manière à ce que chaque processeur contienne le moins possible d'indices de colonnes externes (ou le plus possible dans *update*). En effet, dans ce cas, les produits matrice-vecteur nécessitent moins d'échange de données, et la technique de recouvrement dans le préconditionnement par décomposition de domaine (cf. **6.4.3.b**) devient plus efficace. Pour se fixer les idées, on pourra se représenter une partition de la matrice par lignes contigües, et remarquer que le préconditionneur (d'allure diagonale par blocs) ainsi obtenu récupèrera d'autant plus d'information globale que les éléments de la sous-matrice voisine se retrouvent proche de la diagonale, c'est à dire dans *update*. Ce raisonnement tient aussi dans le cas d'une division en lignes non contigües.

Ainsi, comme les indices de colonne de type externe représentent au niveau physique des interactions entre points voisins affectés à des processeurs différents, une solution est de partionner le maillage en limitant la taille des interfaces. On transpose alors ce partitionnement du maillage en un partitionnement des degrés de liberté en parcourant le tableau ℓ_d (*DegLib*). La première étape (maillage) est un problème *Np-complet* ; il n'existe pas, pour l'instant, d'algorithme optimal permettant de le résoudre. Toutefois, plusieurs équipes ont développé des algorithmes heuristiques ("glouton" par exemple) de très bonne qualité. Nous utiliserons pour notre part la librairie *Metis* développée par G. Karypis et V. Kumar [55], qui permet de partitionner un maillage écrit sous forme de graphe (en énumérant pour chaque point la liste des points voisins comme en **6.3.1.c**).

La procédure consiste à déduire de l'objet *Geometrie* un fichier de maillage au format d'entrée de *Metis*, puis convertir ce maillage en graphe à l'aide de l'utilitaire *mesh2nodal* (fourni par *Metis*), et enfin lancer l'outil de partitionnement *kmetis* assorti du nombre de sous-domaines qu'on désire obtenir. Le format de sortie est un fichier contenant à la ligne i le sous-domaine auquel appartient le point i. Il reste à remplir le fichier *.update* de manière à ce que tous les degrés de liberté I issus du point i se trouvent affectés au processeur qui a en charge le sous-domaine contenant i.

On arrive alors à abaisser d'environ 20% le nombre d'éléments compris dans *external* par rapport à un partitionnement en lignes contigües ; et, sur le cas du rolling par exemple, à diminuer le nombre d'itérations de l'ordre de 20% également pour la méthode GMRES préconditionnée ILU(0) par décomposition de domaine. Ainsi, c'est un gain de près de 30% (en temps) qu'on arrive à réaliser sur la méthode itérative parallélisée.

Si l'on appelle *Parallel_Solver* l'objet contenant la sous-matrice sur chaque processeur, la parallélisation de la phase d'initialisation peut alors être synthétisée par la figure suivante (où l'on n'a pas fait figurer la prise en compte de l'interpolation) :

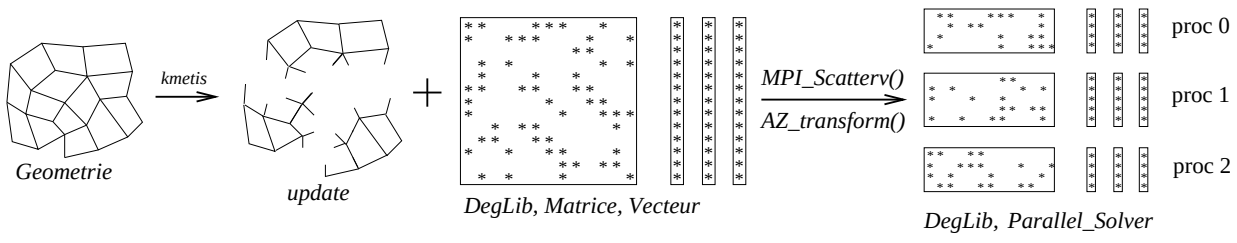


FIG. 6.11 – Phase d'initialisation sur trois processeurs

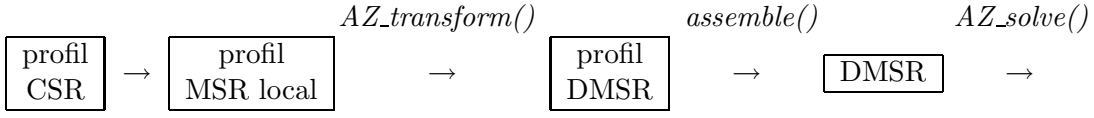
6.5.2 Assemblage

La démarche que nous avons adoptée a été de commencer par la parallélisation du problème académique de Poisson, pour lequel une seule inversion, et donc une seule phase d'assemblage de la matrice sont nécessaires. Pour cela, la méthode la plus simple consiste à convertir au format MSR local non pas la matrice CSR déjà assemblée, mais seulement sa structure. On effectue alors l'assemblage dans le profil MSR local très simplement, puisque les éléments y sont accessibles par les mêmes indices globaux que dans la procédure initiale. On obtient ainsi le schéma :

On assemble donc l'élément d'indices I et J dans la sous-matrice du processeur contenant le vecteur *update* où I figure (l'indice de colonne J en cette ligne apparaît alors forcément dans le tableau *bindx* de ce processeur). On ajoute donc le test "Si I est affecté au processeur" dans l'algorithme exposé en **6.3.2.a**, juste après "Si I est un degré de liberté".

6.5.2.a Permutation des fonctions d'assemblage et de transformation

Dans les problèmes non linéaires et/ou d'évolution, La situation n'est pas aussi simple que dans le cas précédent. En effet les sous-matrices, après résolution, se trouvent au format DMSR, ce qui empêche d'effectuer un nouvel assemblage avec la même méthode. Deux solutions se présentent alors : écrire une fonction effectuant la transformation inverse DMSR → MSR, et l'appeler après chaque résolution, ou modifier la fonction d'assemblage de manière à l'appliquer directement au format DMSR. Nous avons choisi la deuxième, parce qu'*Aztec* met à notre disposition les tableaux d'indirection *update_index* et *extern_index* permettant d'accèder aux éléments des matrices locales après transformation. Ainsi nous appliquons cette fois la fonction *AZ_transform()* au profil de la sous-matrice, et la procédure s'écrit :



En pratique, on utilise la fonction *AZ_find_index()* appliquée à l'indice I dans *update*, et à l'indice J dans *update* et *external*. Cette fonction retourne l'indice dans le tableau *update_index* (resp. *extern_index*) du numéro local correspondant au numéro global cherché dans *update* (resp. *external*) s'il y figure, -1 sinon.

Remarques

· L'assemblage des opérateurs non linéaires nécessite la donnée de l'inconnue sur tous les points du maillage, ainsi d'une part on doit effectuer un regroupement de la solution après chaque résolution, d'autre part ce regroupement doit être connu de tous les processeurs. On utilise à cet effet la procédure de la figure 6.10, mais avec la fonction *MPI_Allgatherv()* (cf. [71]).

· Pour les cas non linéaires stationnaires, qui nécessitent un calcul de résidu, on utilise le type "matrix-free" *AZ_MATRIX* et ses utilitaires (cf. [97]) pour l'effectuer en parallèle.

6.5.2.b Algorithme

L'algorithme général d'assemblage parallèle s'écrit (a et b représentent la matrice et le second membre locaux au processeur) :

$a \leftarrow 0,\ b \leftarrow 0$
Pour $l = 1...L_h$:
| S'il existe (k_i, i) tel que $\ell_d(k_i,(l,i))$ soit affecté à ce processeur ($\ell_d(k_i,(l,i)) \in$ *update*) :
| | *1) Calcul de la matrice et du second membre élémentaires (voir* **6.3.2.a***)*
| | *2) Assemblage :*
| | Pour $k_i = 1...n_{\text{inc}}$ et $i = 1...n_{k_i}$:
| | | $I \leftarrow \ell(k_i,(l,i))$
| | | Si I est un degré de liberté :
| | | | $I_P \leftarrow$ *AZ_find_index(*I, *update)*
| | | | Si $I_P \geq 0$:
| | | | | Pour $k_j = 1...n_{\text{inc}}$ et $j = 1...n_{k_j}$:
| | | | | | $J \leftarrow \ell(k_j,(l,j))$
| | | | | | Si J est un degré de liberté :
| | | | | | | $J_P \leftarrow$ *AZ_find_index(*J, *update)*
| | | | | | | Si $J_P \geq 0$:
| | | | | | | | $a_{u(I_P),u(J_P)} \leftarrow a_{u(I_P),u(J_P)} + \mathbb{A}^l_{i,j,k_i,k_j}$
| | | | | | | sinon :
| | | | | | | | $J_P \leftarrow$ *AZ_find_index(*J, *external)*
| | | | | | | | $a_{u(I_P),e(J_P)} \leftarrow a_{u(I_P),e(J_P)} + \mathbb{A}^l_{i,j,k_i,k_j}$
| | | | | | | Fin si
| | | | | | sinon :
| | | | | | | $b_{u(I_P)} \leftarrow b_{u(I_P)} - \mathbb{A}^l_{i,j,k_i,k_j}\ x_D(J)$
| | | | | | Fin si
| | | | | Fin pour
| | | | | $b_{u(I_P)} \leftarrow b_{u(I_P)} + \mathbb{B}^{i,k_i}_l$
| | | | Fin si
| | | Fin si
| | Fin pour
| Fin si
Fin pour

où nous avons noté a (resp. y) la sous-matrice (resp. le sous-second membre) porté par le processeur, et u (resp. e) le tableau d'indirection *update_index* (resp. *extern_index*).

6.5.3 Interface libre

Pour la parallélisation du problème bifluide, il reste à tenir compte de la contrainte de masse

$$\int_{\Omega_1} \operatorname{div} v_h = 0$$

dans le système linéaire, qui rappelons-le, impose l'ajout du multiplicateur de Lagrange (scalaire) λ à l'inconnue discrétisée (du moins dans le cas des éléments finis Q^1 stabilbsés, cf. **3.3**). Cette procédure se prête mal à l'assemblage distribué de la matrice, en raison de sa nature *non locale* (un élément de la ligne Φ ne représente pas une interaction physique entre deux points). C'est pour

cette raison que A et Φ sont construits séparément avant d'être regroupés dans A'. Le problème est que c'est le profil établi dans *DegLib* qui est utilisé pour définir les matrices locales, ainsi la ligne supplémentaire n'apparaît pas dans la structure parallèle, et le fonctionnement interne d'*Aztec* ne se prête bien pas à cet ajout en local.

Nous avons donc cherché une méthode alternative à la construction de l'objet $\{A'\}$ pour la prise en compte de la contrainte de masse. Une première solution est de procéder par complément de Schur, mais cela conduit à deux inversions de la matrice A. La deuxième possibilité est de "rendre discontinue" la pression au niveau de l'interface, ce qui équivaut à ajouter dans (cf. (3.22)) $M_h^n = \{\tilde{p} \in M_T \mid \tilde{p}_{|K_l} \in \mathbb{P}^1(K^l(t^n)),\ l = 1, \ldots, L\}$ les fonctions test manquantes $\mathbb{1}_{\Omega_i^n}$, $i = 1, 2$.

6.5.3.a Résolution par complément de Schur

Dans le système linéaire décrit en **6.3.2** :

$$\left[\begin{array}{c|c} & \\ A & \Phi^T \\ & \\ \hline \Phi & 0 \end{array}\right] \left[\begin{array}{c} \\ x \\ \\ \hline \lambda \end{array}\right] = \left[\begin{array}{c} \\ b - A_D x_D \\ \\ \hline 0 \end{array}\right], \tag{6.3}$$

le calcul du multiplicateur de Lagrange peut se faire par élimination de l'inconnue $x_{\overline{D}}$, ce qui entraîne deux inversions de la matrice A au lieu d'une inversion de A'. On s'inspire de la méthode de L. Formaggia et al. ([34]) où l'on cherche à imposer des flux de matière sur les bords d'un domaine, restreinte au cas où seule l'interface séparant l'aluminium de la cryolithe entre en compte, et où le flux imposé est nul. Ainsi, le premier bloc de (6.3) donne

$$x = A^{-1}(b - A_D x_D) - \lambda\, A^{-1}\Phi^T, \tag{6.4}$$

et du deuxième bloc et de (3) on déduit

$$\Phi\Big(A^{-1}(b - A_D x_D) - \lambda\, A^{-1}\Phi^T\Big) = 0,$$

soit

$$\lambda = \frac{\Phi A^{-1}(b - A_D x_D)}{\Phi A^{-1}\Phi^T}.$$

La nouvelle procédure s'écrit alors, après assemblages séparés de A et Φ :

1) Résolution parallèle de $Ax = b - A_D x_D$ *et calcul de* Φx,

2) Résolution parallèle de $Ay = \Phi^T$ *et calcul de* Φy,

3) $x \leftarrow x - \dfrac{\Phi x}{\Phi y}\, y$.

Remarquons que cela nécessite de disposer d'une version locale des vecteurs y et Φ (c'est déjà le cas pour x, b et x_D). On les ajoute dans l'objet *Parallel_Solver*, et on les initialise avec *AZ_reorder_vec()*, de la même manière que les autres données. Alors, pour multiplier x et y par Φ, il faut effectuer un regroupement de Φ, ce qui se fait sans difficultés avec *MPI_Gatherv()* (cf. Fig. 6.10) puisque ce vecteur est réparti comme le sont x et y.

Les résultats obtenus sont exactement les mêmes qu'avec la méthode utilisant A', mais bien qu'une inversion de A prenne moins de temps que celle de A' (environ 30% d'itérations en moins), c'est une méthode moins efficace étant donné qu'on fait deux inversions. On perd alors un facteur de l'ordre de 1.5 par rapport à une parallèlisation directe du système $A'x' = b'$. Cet inconvénient majeur nous a incités à chercher un autre moyen de préserver la masse de chaque fluide, sans utiliser de multiplicateur de Lagrange.

6.5.3.b Duplication des degrés de liberté en pression sur l'interface

L'ajout d'un multiplicateur de Lagrange pour prendre en compte la contrainte de masse, d'une part, est compliqué à implémenter dans le cadre de la parallélisation, et, d'autre part, n'a pas de réalité physique. En effet, considérons dans ce cas la formulation variationnelle du problème de Navier-Stokes bifluide (pour simplifier), posé sur un domaine Ω séparé en deux sous-domaines Ω_1 et Ω_2 par l'interface Σ (on note V_h l'espace de discrétisation en vitesse) :

Trouver $(v, p, \lambda) \in V_h \times M_h \times \mathbb{R}$ tel que :

$$\left\{ \begin{array}{rcl} \displaystyle\int_{\Omega_1\cup\Omega_2} \zeta\,(\partial_t\, v + v\cdot\nabla v)\cdot\tilde{v} + \int_{\Omega_1\cup\Omega_2} \frac{\zeta}{Re}\, D(v) : \nabla\tilde{v} - \int_{\Omega_1\cup\Omega_2} p\,\mathrm{div}\,\tilde{v} + \lambda \int_{\Omega_1} \mathrm{div}\,\tilde{v} & = & \displaystyle\int_{\Omega_1\cup\Omega_2} f\cdot\tilde{v}\,, \\ \displaystyle\int_{\Omega_1\cup\Omega_2} \mathrm{div}\, v\,\tilde{p} + \Big(\int_{\Omega_1} \mathrm{div}\, v\Big)\tilde{q} & = & 0\,, \end{array} \right.$$

$\forall\,(\tilde{v}, \tilde{p}, \tilde{q}) \in V_h \times M_h \times \mathbb{R}$. L'intégration par parties de la première équation donne :

$$\int_{\Omega_1\cup\Omega_2} \Big(\zeta\,(\partial_t\, v + v\cdot\nabla v) - \mathrm{div}\,\Big(\frac{\zeta}{Re}\, D(v)\Big) + \nabla p - f\Big)\cdot\tilde{v} + \int_\Sigma \Big(\Big\{\frac{\zeta}{Re} D(v) - pI_d\Big\}\, n_\Sigma + \lambda\, n_\Sigma\Big)\cdot\tilde{v} = 0\,,$$

où $\Big\{\eta D(v) - pI_d\Big\}$ est le saut du tenseur des contraintes sur l'interface :

$$\Big\{\frac{1}{Re} D(v) - pI_d\Big\} = \Big(\frac{1}{Re} D(v) - pI_d\Big)_{|\Omega_1} - \Big(\frac{1}{Re} D(v) - pI_d\Big)_{|\Omega_2}$$

et n_Σ la normale extérieure au fluide 1. On en déduit que celui-ci est non nul dès que λ est non nul, ce qui n'est pas physique d'après les conditions interfaciales (3.20).

L'approche par complément de Schur, en plus de limiter notablement l'intérêt de la parallélisation, n'est donc pas satisfaisante d'un point de vue physique. Ce constat théorique est renforcé par des tests numériques qui montrent que sur un cas de sloshing bidimensionnel, les fréquences gravitationnelles obervées (cf. **4.2.2**) dépendent de la fréquence d'excitation de la cuve. Ce qui est physique, en revanche, c'est l'existence d'une tension de surface entre les deux fluides, qui peut générer une discontinuité de la pression à l'interface. Cet opérateur est d'ailleurs déjà implémenté dans Mistral suivant la méthode expliquée dans [39]. Un moyen de rendre la pression discontinue sur l'interface *tout en conservant l'interpolation Q^1 dans chacun des fluides* est alors de lui faire prendre deux valeurs différentes à cet endroit, l'une côté aluminium et l'autre côté électrolyte. Pour implémenter cette modification, une méthode possible est de faire coexister deux degrés de liberté distincts en chaque point de l'interface, en attribuant à chaque point de l'interface deux numéros globaux, associés chacun à un des fluides de part et d'autre de l'interface :

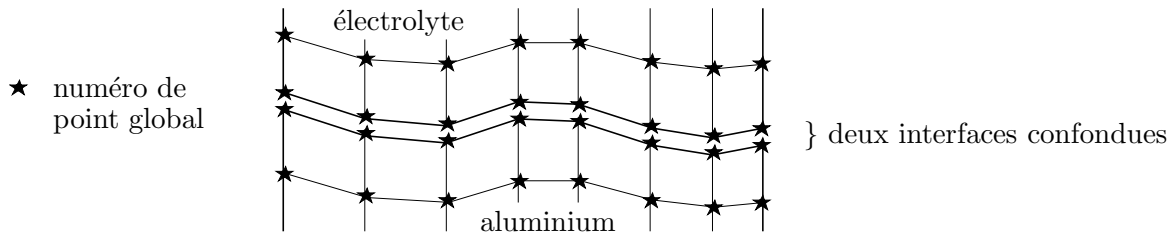


FIG. 6.12 – Dédoublement des nœuds du maillage sur l'interface

On obtient alors $(2\,d + 1)\,N_{\text{int}}$ degrés de liberté supplémentaires (où N_{int} est le nombre de points sur l'interface), parmi lesquels on ne garde que les N_{int} représentant la pression :

1) Modification de la connectivité du maillage

On génère donc un nouveau maillage dans lequel figurent deux interfaces, c'est à dire des points supplémentaires qui entraînent une modification de la connectivité du maillage. En effet, les faces des polyèdres de part et d'autres de l'interface qui initialement étaient confondues (y compris au sens de la connectivité) se retrouvent indépendantes, mais toujours géométriquement confondues. Il en résulte qu'un point de l'interface n'a plus comme voisins que ceux qui appartiennent aux éléments géométriques compris dans la figure (fluide 1 ou fluide 2) à laquelle appartient ce point.

2) Fusion des degrés de liberté relatifs à la vitesse et au champ magnétique

De ce dédoublement des points résulte un dédoublement des degrés de liberté pour l'ensemble des inconnues, étant donné que la combinaison d'un numéro global de point avec une inconnue scalaire sans conditions de Dirichlet génère *a priori* un nouveau degré de liberté. Or les inconnues v ($1 \leq k \leq d$) et B ($d+1 \leq k \leq 2d$), à l'inverse de p ($k = 2d+1$), sont impliquées dans le couplage entre les deux fluides, ce qui signifie qu'un point sur l'interface ne doit représenter qu'un degré de liberté par composante de la vitesse et du champ magnértique (et non deux). La solution retenue pour implémenter cette modification est de construire dans un premier temps la liste des couples de numéros globaux des points "jumeaux" $\{j_q^1, j_q^2\}_{1 \leq q \leq N_{\text{int}}}$, de manière à ce que $j_q^1 < j_q^2$, $\forall\, q$. La construction du tableau ℓ_d, basée sur le parcours de l'ensemble des nœuds du maillage et des inconnues scalaires, prend alors la forme :

$I \leftarrow 0$
Pour $p = 1 \dots N_h$
| $r \leftarrow 0$
| Pour $q = 1 \dots N_{\text{int}}$:
| | Si $j_q^2 = p$:
| | | $r \leftarrow j_q^1$
| | | Sortie de boucle.
| | Fin si
| Fin pour
| Pour $k = 1 \dots n_{\text{inc}}$:
| | Si $\ell_d(k,p) \neq 0$:
| | | Si $r = 0$ ou $k > 2d$:
| | | | $I \leftarrow I + 1$
| | | | $\ell_d(k,p) \leftarrow I$
| | | sinon :
| | | | $\ell_d(k,p) \leftarrow \ell_d(k,r)$
| | | Fin si
| | Fin si
| Fin pour
Fin pour

3) Modification dans la construction du format creux

Du fait de l'intervention de la liste des voisins de chaque point dans la construction de la structure matricielle (cf. **6.3.1.c**), il faut annuler dans la contruction du format CSR (voir l'algorithme page 83) les interactions redondantes (du fait que $\ell_d(k, j_q^1) = \ell_d(k, j_q^2)$) entre degrés de liberté doublement représentés pour rendre compatibles la modification *1)* et la fusion *2)*. Ces redondances,

qui se traduisent par une double apparition de l'indice de colonne $J = \ell_d(k, j_q^1) = \ell_d(k, j_q^2)$ dans les tableaux

$$C[\ell_d(k, j_q^1), \cdot]\,, \quad \forall\,(k, q)\,,$$

et par une incrémentation inutile de N_Z, peuvent être évitées en insérant le test :

$$\begin{array}{l}\text{“Si } \left(\, k_i \text{ ou } k_j = 2d+1 \,\right) \text{ ou} \\ \quad \forall\, q\,,\ j_q^1 \neq i \neq j_q^2 \ \text{ ou} \\ \quad \Big(\, \exists\, q\,,\ i = j_q^s \ \text{ et } \ \left(\, s = 1 \ \text{ ou } \ \forall\, q\,,\ j_q^1 \neq \nu_p(m) \neq j_q^2 \,\right) \Big) \ \text{“}\end{array}$$

avant la séquence d'instructions formée par les lignes 9 à 14 de l'algorithme page 83. Finalement, c'est ici dans la phase d'initialisation (construction du maillage et initialisation de la structure matricielle) que des modifications doivent être apportées au code initial, contrairement à la méthode basée sur le multiplicateur de Lagrange qui concerne les phases d'assemblage et de résolution.

Remarque

Cette manière de procéder est typique de la prise en compte de conditions périodiques, où les degrés de liberté à fusionner appartiennent non pas à une même interface mais à deux frontières sur lesquelles on souhaite que l'inconnue prenne les mêmes valeurs.

6.6 Mesures et conclusion

Nous exposons ici les performances obtenues par la parallélisation sur notre CLUMPS d'un cas de rolling sur 200 itérations en temps, et un maillage de 12000 points :

Nombre de processeurs	Temps de calcul
1	424’
2	215’
4	114’
6	81’
8	64’
10	57’

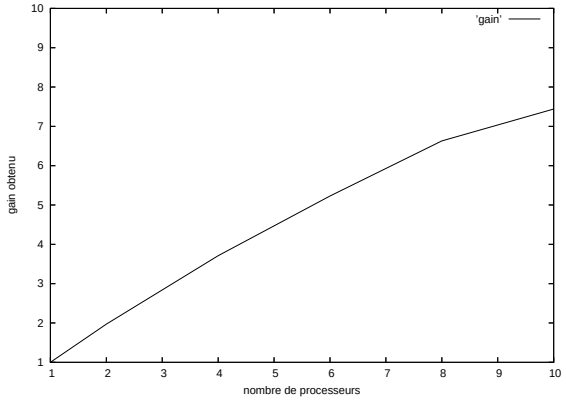


FIG. 6.13 – Mesures des temps de calcul “wall-clock” et gains correspondants

Les résultats obtenus sont satisfaisants, dans la mesure où ce type de parallélisme donne souvent des signes de fléchissement quand on approche des dix machines. Cette amélioration est suffisante pour entreprendre une campagne de tests (cf. **7**), en vue d'exhiber des commandes possibles pour la mise en œuvre d'un programme de contrôle optimal pour le modèle non linéaire (cf. **8**).

Pour faire mieux, il faudrait sûrement adopter un autre point de vue, à savoir diviser le problème en amont du système linéaire, par décomposition de domaine par exemple. On obtiendrait ainsi plusieurs systèmes linéaires indépendants, et formant chacun un problème réellement localisé... mais c'est une autre histoire, la décomposition de domaine en MHD n'étant pas encore une branche très explorée en simulation numérique. Comme on peut le voir sur la figure 6.14, la méthode globale reste la même après parallélisation, qui ne se situe qu'au niveau algébrique du problème.

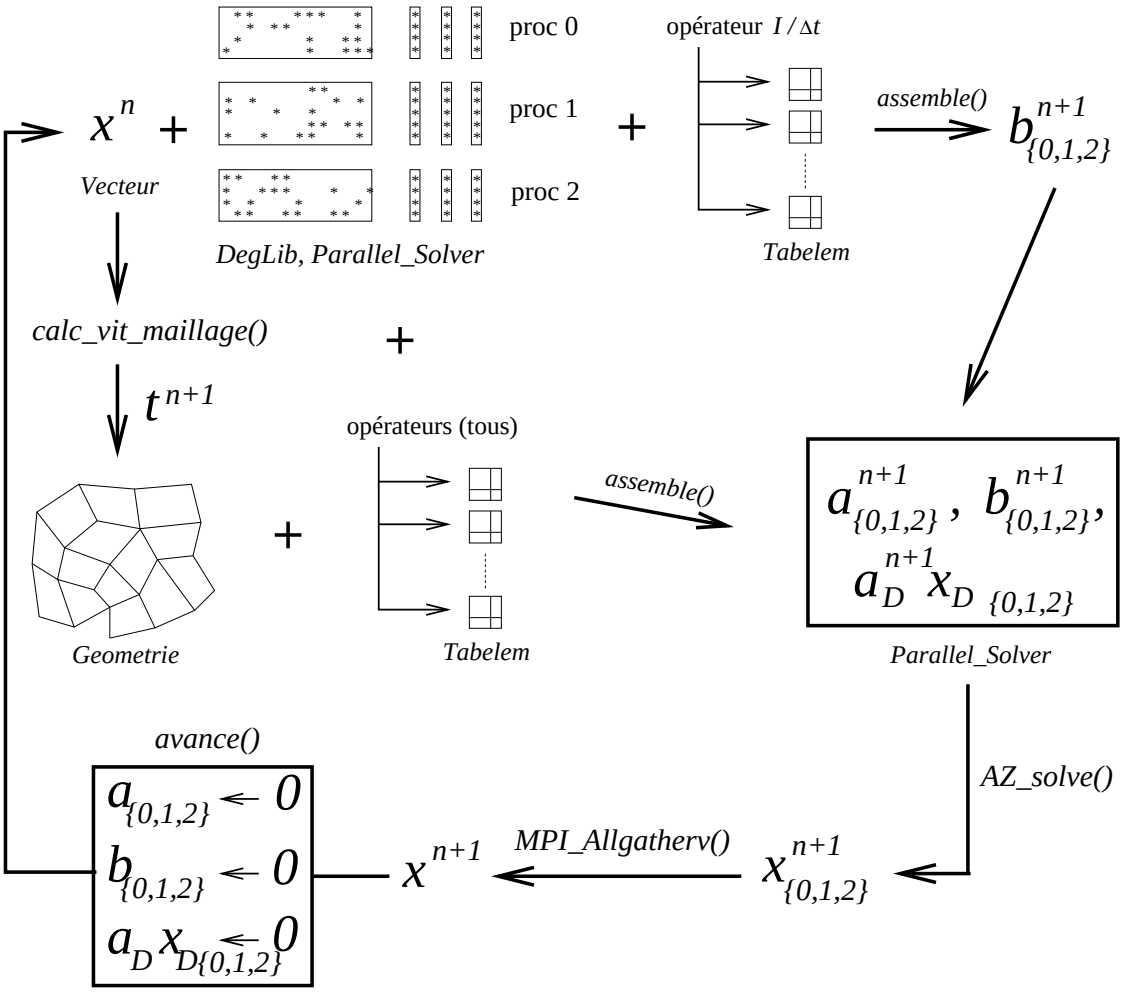


FIG. 6.14 – Parallélisation de la résolution (exemple sur 3 processeurs). Voir FIG. 6.7

Chapitre 7

Recherche d'un actionneur

Dans l'optique du contrôle optimal de l'interface bain/métal, on cherche à présent à exhiber de nouveaux mécanismes sous l'action des paramètres d'entrée du modèle. Le but ultime est, rappelons-le, non seulement de prévoir le comportement du système, mais aussi de calculer quelle modification lui apporter en fonction des signaux qu'il renvoie. Il faut disposer pour cela d'un *actionneur*, à savoir un outil physique permettant d'agir sur le système dans la direction qu'on souhaite, et d'un *critère*, le souhait en question. *A priori*, la logique voudrait que l'on commence par fixer certains objectifs avant de se donner les moyens de les atteindre. En réalité, c'est en évaluant d'abord la portée des moyens à disposition qu'on parvient à définir des critères réalistes. Cela revient en quelque sorte à chercher, par l'expérimentation numérique, des modèles contrôlables.

Dans les tests menés ici, deux paramètres sont examinés : la hauteur moyenne d'électrolyte h_2 et les conditions aux limites (stationnaires) sur le champ magnétique B. Toutes les simulations sont effectuées dans un cylindre de rayon $R = 0.5$ contenant une quantité fixée d'aluminium de hauteur moyenne $h_1 = 0.5$, et parcouru par une intensité totale constante $I = 500 \cdot 10^3$. On utilise à cet effet la version parallélisée du code, qui permet l'emploi d'un maillage assez fin (de l'ordre de $15 \cdot 10^3$ hexaèdres) pour utiliser un paramètre de stabilisation $\tau = 0.1$ (voir **3.3**).

Ainsi, quatre résultats principaux sont obtenus, qu'on peut répartir en deux catégories : l'une *dynamique*, sur la possibilité de stabiliser le phénomène en atteignant un état stationnaire ; l'autre *statique*, sur la possibilité de faire adopter à l'interface une forme particulière à l'état stationnaire :

1) Dynamique
   - 1.a) Pour toutes les conditions aux limites magnétiques testées, une diminution de la hauteur moyenne d'électrolyte conduit à une stabilisation du phénomène (**7.1**,**7.4**)
   - 1.b) Dans la situation du rolling instable (cf. **3.3**), on parvient à stabiliser le système en appliquant une distribution de courant restreinte au centre de l'anode (**7.2.2**)

2) Statique
   - 2.a) En présence de faibles champs magnétiques verticaux, on parvient à décider de la forme de l'interface en agissant sur la distribution anodique de courant (**7.2.1**,**7.4**)
   - 2.b) En présence de champs magnétiques verticaux importants, l'interface est invariablement "aspirée" en son centre lorsqu'un état stationnaire est possible (**7.2.2**,**7.4**).

On rappelle par commodité le modèle choisi pour simuler le comportement du phénomène (on ommet les conditions interfaciales par souci de concision, voir **3.3**) :

$$
\left\{
\begin{array}{rcl}
\left.\begin{array}{rcl} \operatorname{div} v \;=\; \operatorname{div} B &=& 0 \\ \partial_t \rho + \operatorname{div}(\rho v) &=& 0 \end{array}\right\} & \text{dans } \Omega\,, \forall\, t\,, \\
\left.\begin{array}{rcl} \zeta_i \dfrac{\partial v}{\partial t} + \zeta_i (v\cdot\nabla)\, v - \operatorname{div}\Big(\dfrac{\zeta_i}{Re_i}\nabla v\Big) + \nabla p - S \operatorname{rot} B \times B &=& -\zeta_i \dfrac{e_z}{Fr} \\ \dfrac{\partial B}{\partial t} + \operatorname{rot}\Big(\dfrac{1}{Rm_i}\operatorname{rot} B\Big) - \operatorname{rot}(v\times B) &=& 0 \end{array}\right\} & \text{dans } (\Omega_i)_{i=1,2}\,, \forall\, t\,, \\
\left.\begin{array}{rcl} v(0,x) &=& v^0(x) \\ B(0,x) &=& B^0(x) \end{array}\right\} & \text{dans } \Omega\,, \\
\left.\begin{array}{rcl} v\cdot n &=& 0 \\ B\times n &=& B_0 \times n \end{array}\right\} & \text{sur les parties verticales de } \partial\Omega\,, \quad \forall t\,, \\
\left.\begin{array}{rcl} v &=& 0 \\ B\cdot n &=& 0 \\ \operatorname{rot} B \times n &=& 0 \end{array}\right\} & \text{sur les parties horizontales de } \partial\Omega\,, \quad \forall t\,,
\end{array}
\right. \tag{7.1}
$$

ainsi que la perturbation initiale : on modifie l'angle de la gravité pendant une unité de temps.

7.1 Effet de la hauteur d'électrolyte sur le rolling

On reprend le cas test exposé en **3.3.1** et on modifie la hauteur de fluide supérieur pour un champ magnétique vertical fixé : le comportement observé est une **stabilisation du phénomène de rolling par diminution de la hauteur d'électrolyte**. En effet, on voit que l'amplitude des oscillations d'un point du bord de l'interface croît avec h_2 : Ce phénomène est assez gênant, car

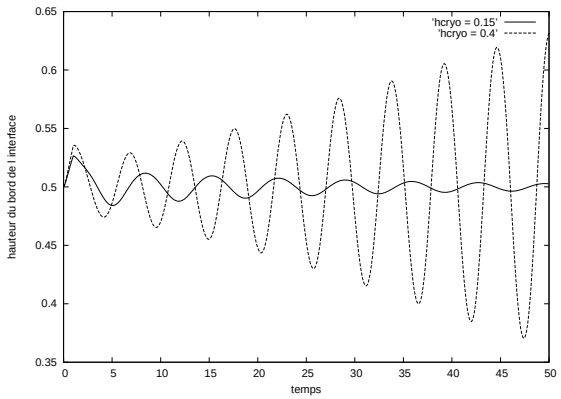


FIG. 7.1 – Comparaison de deux cas de rolling avec $h_2 = 0.15$ (stable) et $h_2 = 0.40$ (instable)

les observations faites sur le terrain montrent clairement qu'une hausse de la hauteur d'électrolyte a un effet stabilisant. C'est même une technique éprouvée, dite de "desserrement" de la cuve, utilisée sur le terrain pour réguler le fonctionnement des cuves. Le modèle est donc, sur ce point, en contradiction avec la réalité physique. Le principal objet de ce chapitre est ainsi de le modifier pour identifier un actionneur satisfaisant.

7.2 Restriction centrale de l'arrivée de courant électrique

Rappelons que dans le phénomène du rolling, les anodes sont modélisées par la surface supérieure du cylindre toute entière, ce qui occasionne une arrivée uniforme du courant électrique dans l'électrolyte (cf. **3.3.1**). Quand on connaît l'importance des courants horizontaux dans les mécanismes de stabilisation (ou déstabilisation) de la cuve (cf. **1.2.2**), une idée qui peut venir à l'esprit est d'imposer une distribution hétérogène de la densité de courant électrique à l'anode afin de générer ces courants dans l'électrolyte. En particulier, on étudie la configuration suivante de distribution *centrale* du courant :

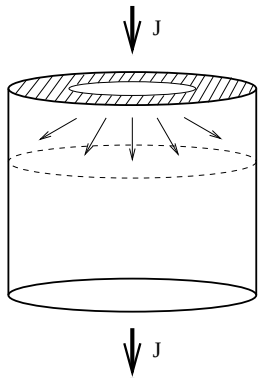


FIG. 7.2 – Restriction centrale de l'arrivée de courant

Pour implémenter cette modification, on définit le disque centré $\mathcal{S}_i$ de rayon $R_i < R$ en dehors duquel on prescrit toute arrivée de courant. On cherche alors la valeur à donner à la condition aux limites en champ magnétique sur la surface supérieure du cylindre pour satisfaire cette ditribution. Les propriétés de symétrie de cette dernière donnent lieu à un champ magnétique orthoradial ne dépendant que de la distance à l'axe central : en coordonnées cylindriques,

$$B_\theta(r) = \overline{B_\theta(r)}\, e_\theta$$

De plus, en vertu de l'équation de Maxwell-Ampère sans courants de déplacement (cf. (3.5)), l'intensité $I_{S(r)}$ du courant traversant un disque centré $S(r)$ s'exprime à l'aide de la circulation du champ magnétique sur le contour de cette section (formule de Stokes) :

$$\operatorname{rot} B_\theta(r) = \mu_0\, J(r) \quad \Rightarrow \quad \int_{\partial S(r)} B_\theta(r)\cdot d\ell = -\mu_0\, I_{S(r)} \quad \Rightarrow \quad \overline{B_\theta(r)} = -\frac{\mu_0\, I_{S(r)}}{2\pi r}$$

Deux cas se présentent ainsi :

$$\begin{aligned} &\cdot \quad r \le R_i\,, \quad \text{alors} \quad I_{S(r)} = J\pi r^2 = I\frac{r^2}{R_i^2}\,, \qquad \text{soit} \quad \overline{B_\theta(r)} = \; -\frac{\mu_0 I}{2\pi}\frac{r}{R_i^2}\,, \\ &\cdot \quad r \ge R_i\,, \quad \text{alors} \qquad I_{S(r)} = I_{\mathcal{S}_i} = I\,, \qquad\quad \text{soit} \quad \overline{B_\theta(r)} = \; -\frac{\mu_0 I}{2\pi r}\,. \end{aligned} \tag{7.2}$$

La prise en compte aisée de ce nouveau profil dans la formulation discrétisée du problème requiert d'imposer des conditions aux limites magnétiques du type

$$B \times n = B_0 \times n$$

sur $\partial\Omega$ tout entier. Il est en effet compliqué de calculer rot B sur une frontière dès lors que $B \neq 0$.

Nous nous proposons d'étudier, à hauteur moyenne d'électrolyte fixée ($h_2 = 0.25$), deux aspects : la *stabilité* des phénomènes et la *forme de l'interface* aluminium/électrolyte relativement à la modification du paramètre R_i. Pour cela, on commence par observer l'effet d'une telle redistribution de courant sur un phénomène dénué de champ magnétique vertical. Nous verrons en effet par la suite qu'on obtient des résultats opposés sur la forme de l'interface suivant la présence ou non d'un champ vertical (et en fait plus généralement qu'il existe un seuil sur B_z de part et d'autre duquel l'interface est soit "creusée", soit "bombée").

Ainsi, on trace à gauche l'évolution de la hauteur du point $(x = 0.5, y = 0)$ du bord de l'interface et à droite celle de son point central $x = y = 0$ (cf. FIG. 7.3). On dispose ainsi d'une mesure exploitable pour les deux aspects ci-dessus. Parallèlement, afin de mieux comprendre la signification physique des phénomènes, on observe des représentations en coupe des normes $\|v\|$ et $\|J_{x,y}\| = \|(\mathrm{rot}\, B)_{x,y}\|/\mu_0$ à l'état stationnaire, où la position de l'interface peut être identifiée au-dessus de la dixième maille en partant du bas (cf. *infra*). La légende consiste à suivre, du bleu sombre (valeur nulle) jusqu'au rouge (valeur maximale), les couleurs du spectre de la lumière.

7.2.1 Un problème sans champ magnétique vertical

Avec les conditions ci-dessus sur B_θ et $B_z = 0$, on observe un **creusement de l'interface en son centre, et cette déformation est d'autant plus importante que R_i est faible** (FIG. 7.3). Les représentations spatiales (FIG.7.4) sont issues d'un maillage de volume élémentaire moyen de $5 \cdot 10^{-5}$, et sont validées par des raffinements en temps et en espace, dont les courbes figurent également sur les graphes d'évolution de l'interface :

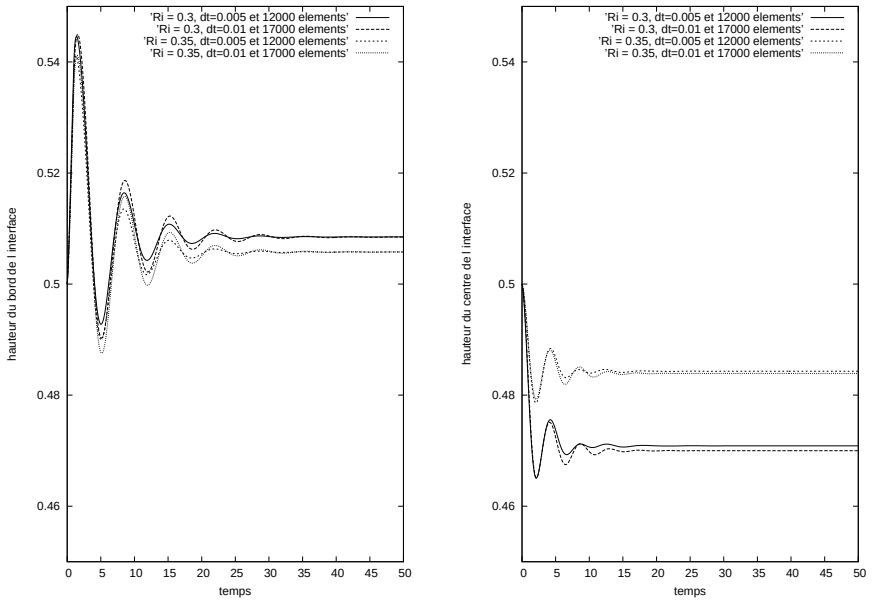


FIG. 7.3 – Hauteurs du bord et du centre de l'interface pour $R_i = 0.3$ et $R_i = 0.35$

Par ailleurs, on observe sur la figure 7.4 qu'à l'état stationnnaire, les vecteurs vitesse et courant horizontal sont centrifuges sur l'interface, et $\|v\|$ et $\|J_{x,y}\|$ sont bien plus importants dans l'électrolyte que dans l'aluminium.

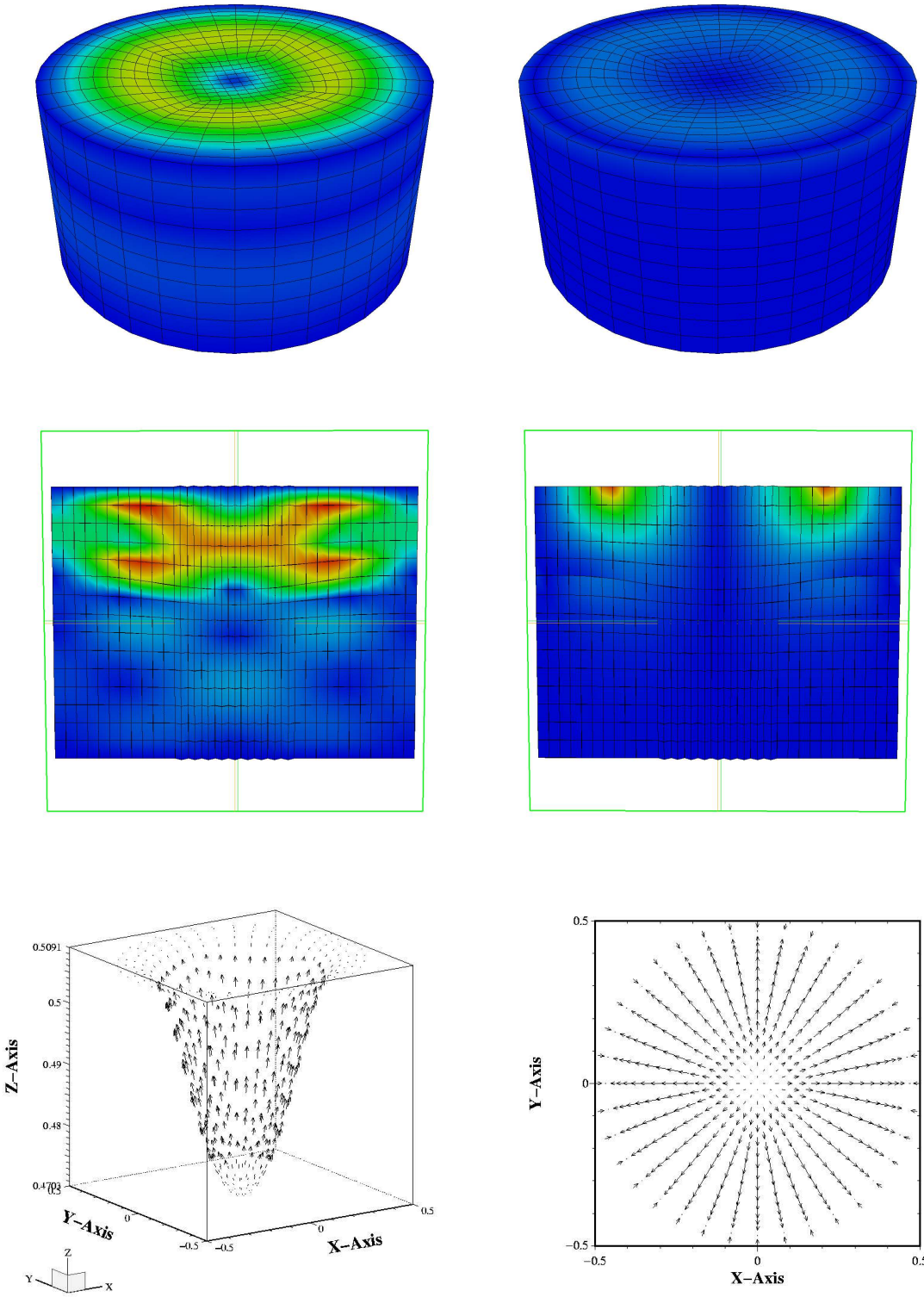


FIG. 7.4 – État stationnaire pour $R_i = 0.3$. À gauche : vitesse (norme autour de l'aluminium, norme sur une coupe verticale, vecteur sur interface). À droite : composante horizontale du courant (norme autour de l'aluminium, norme sur une coupe verticale, vecteur sur interface **aluminium**)

7.2.2 Application au phénomène de rolling

L'objectif ici est d'observer l'effet de la nouvelle distribution de courant sur les cas de rolling stable et instable présentés en **3.3.1**). La validation de ces résultats en modifiant le pas du maillage et le pas de temps pourra être trouvée dans [80].

Dans une configuration stable

On reprend le test précédent en imposant cette fois $B_z = 0.25$ sur le bord du domaine. Dans une situation de rolling classique (distribution de courant uniforme à l'anode), on observe une stabilisation progressive du phénomène vers une forme d'interface très légèrement bombée alors que là, on atteint aussi un état stationnaire mais avec une interface clairement plus haute en son centre que sur sa périphérie. On parlera d'interface *aspirée*. Par ailleurs, on observe que l'interface est d'autant plus aspirée que R_i est faible. La principale information de ce test est donc que **la restriction centrale de l'arrivée de courant en présence d'un champ magnétique vertical engendre une aspiration de l'interface**. On peut ainsi contrôler l'aspiration de l'interface en agissant uniquement sur R_i :

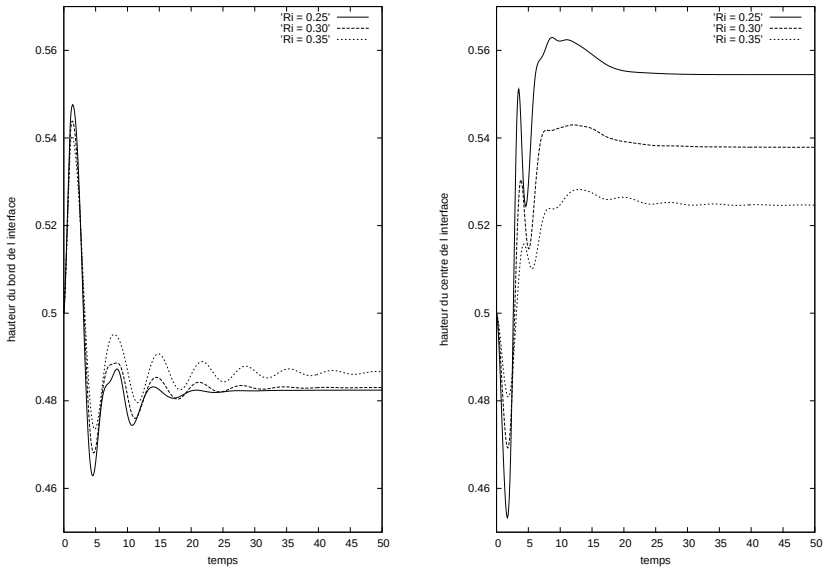


FIG. 7.5 – Contrôle de l'aspiration de l'interface ($B_z = 0.25$, $R_i = 0.25$, 0.30 et 0.35)

Si l'on fait la comparaison avec le test précédent (cf. FIG. 7.4), une autre manière de voir les choses est de constater que l'ajout d'un champ magnétique vertical dans une configuration d'arrivée du courant restreinte au centre a pour effet d'inverser la déformée d'interface. De plus, on peut voir que les courants horizontaux sous l'interface, toujours centrifuges, sont plus importants (FIG. 7.6). Par ailleurs, la vitesse des fluides, qui n'est plus centrifuge mais orthoradiale, n'est plus cantonnée à la seule électrolyte. Enfin, par rapport au rolling, cette vitesse de rotation est de *sens opposé*, et on notera que plus la surface d'émission du courant est restreinte, plus les oscillations de l'interface s'atténuent rapidement.

On représente sur la figure 7.7 l'évolution de la déformée d'interface au cours du temps, dans laquelle on voit clairement la transition d'un comportement de type rolling à une configuration d'interface aspirée, mais qui donne aussi une idée de la complexité du phénomène.

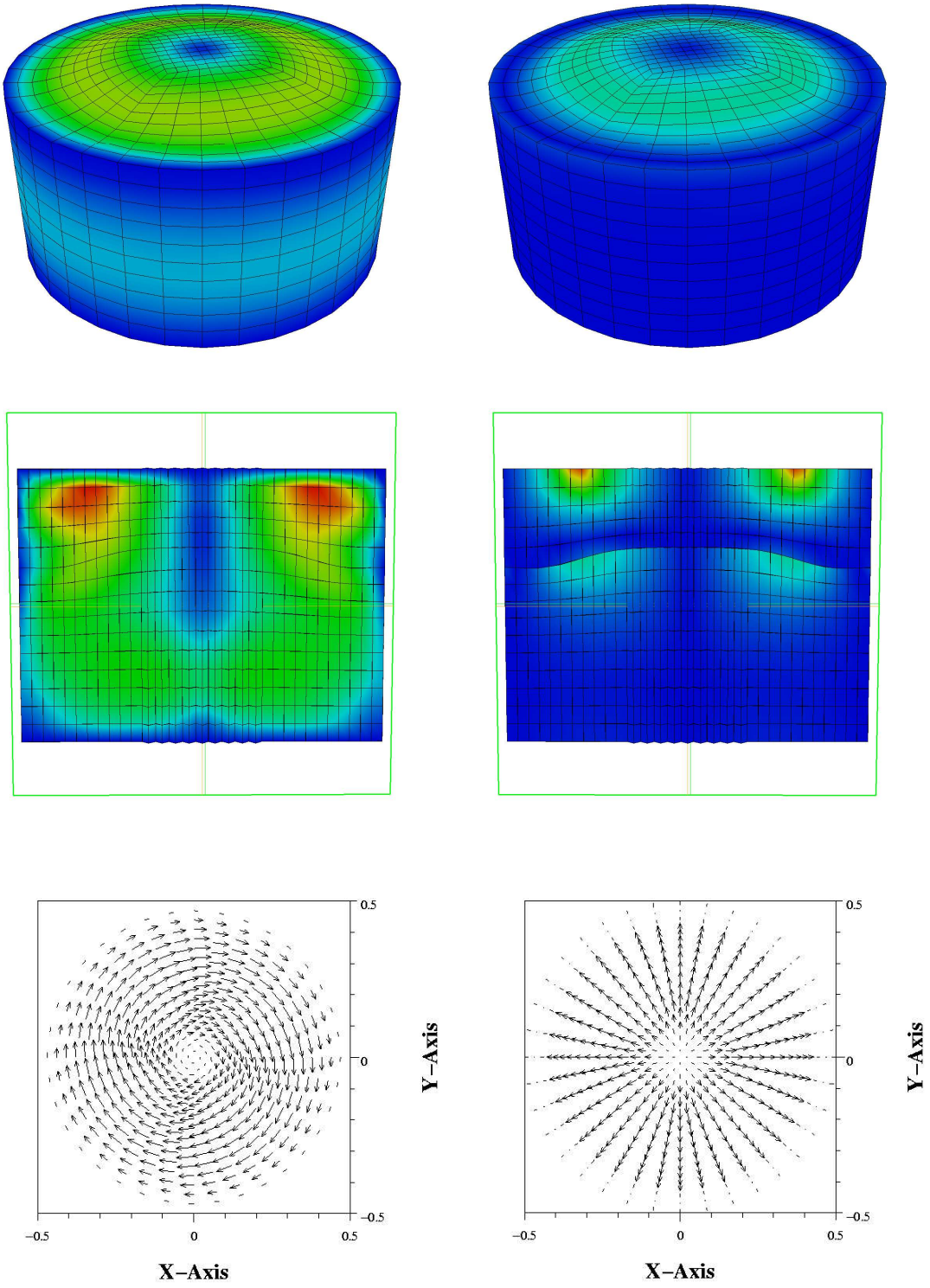


FIG. 7.6 – État stationnaire. À gauche : vitesse (norme autour de l'aluminium, norme sur une coupe verticale, composante horizontale sur interface). À droite : composante horizontale du courant (norme autour de l'aluminium, norme sur une coupe verticale, vecteur sur interface **aluminium**)

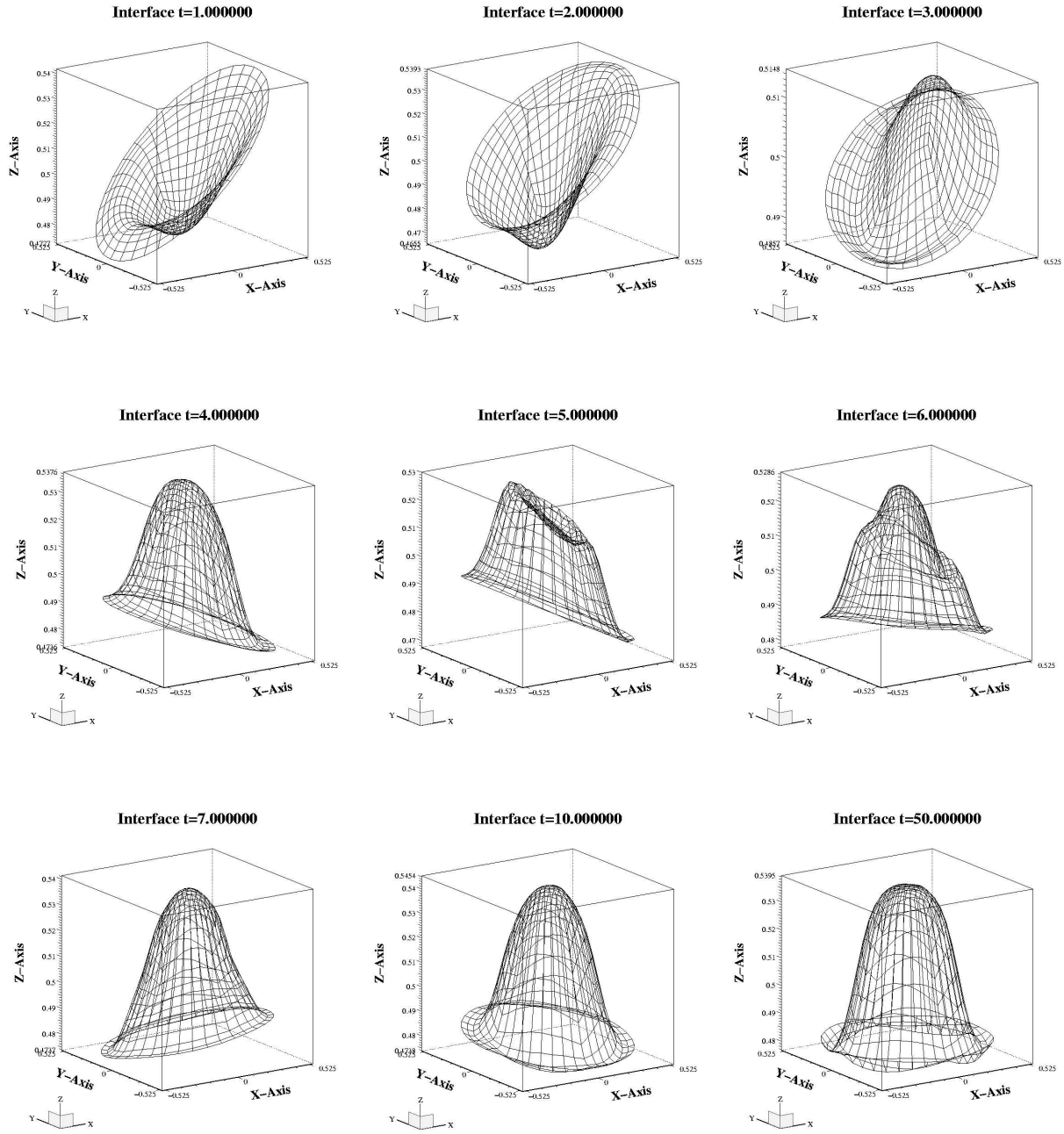


FIG. 7.7 – Évolution de la forme de l'interface jusqu'à l'état stationnaire

Stabilisation d'un rolling instable par restriction centrale de l'arrivée de courant

Portons à présent le champ magnétique vertical à 0.5, ce qui donne lieu à un rolling instable si la distribution de courant est uniforme. Lorsque nous restreignons dans ce cas la surface anodique de la même manière que précédemment en prenant $R_i = 0.35$. Ainsi, la conclusion la plus intéressante est qu' **une restriction centrale de l'arrivée du courant dans certains cas de rolling instable empêche l'interface d'exploser et conduit à un état stationnaire.**

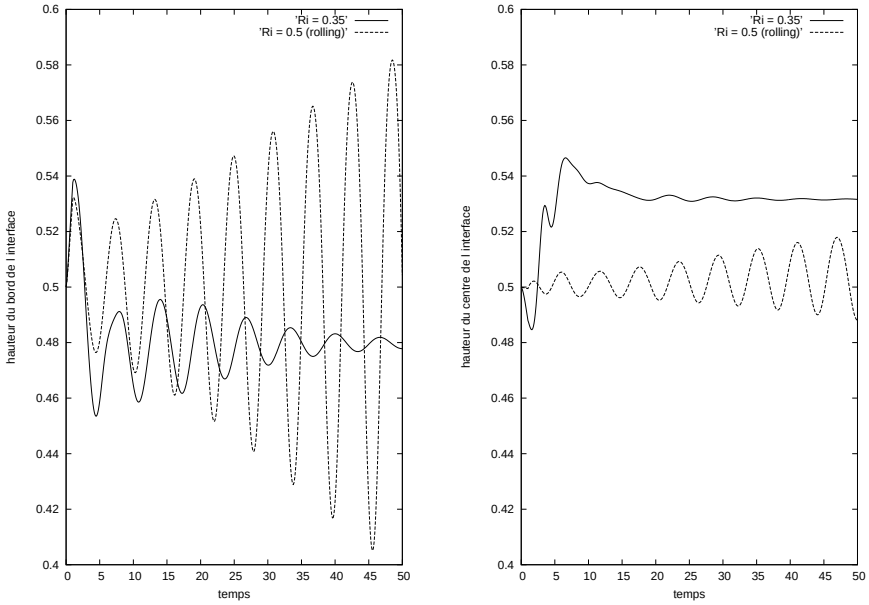


FIG. 7.8 – Stabilisation d'un rolling instable

7.3 Interprétations physiques

7.3.1 Effet des conditions aux limites magnétiques sur la déformée d'interface

L'absence de champ magnétique vertical n'empêche pas pour autant les courants électriques horizontaux d'engendrer une force de Lorentz. En effet, si l'on suppose que l'épanouissement du courant électrique sous l'anode est régi par l'équation stationnaire de diffusion du potentiel électrique ϕ (équation de Poisson), qui prend encore la forme de l'équation de Maxwell-Gauss :

$$-\operatorname{div}(\sigma\,\vec{\nabla}\phi) = 0, \quad \Leftrightarrow \quad \operatorname{div}\vec{E} = 0\,, \quad \text{avec}\,\vec{E} = -\vec{\nabla}\phi\,, \tag{7.3}$$

les propriétés de symétrie des distributions de courant considérées permettent de ramener en deux dimensions le phénomène d'épanouissement du courant. Si l'on impose un potentiel plus élevé sur la surface anodique restreinte que sur la cathode, la propriété de monotonie du laplacien nous invite à supposer que le courant électrique tend à rejoindre les zones situées au-dessous de la surface non-émettrice. En étendant le raisonnement au cas cylindrique, on trouve qu'une distribution centrale de courant engendre une orientation centrifuge de la composante radiale du vecteur J dans l'électrolyte. On justifie ainsi mathématiquement la figure intuitive 7.2. Ces considérations sont par ailleurs confirmées par des simulations de diffusion du potentiel électrique sous l'anode et à travers l'interface (voir p. 112 et FIG. 7.11). En supposant donc ce résultat valide, on peut estimer qu'une force de Lorentz dirigée vers le bas résulte de l'application d'une distribution centrale de courant :

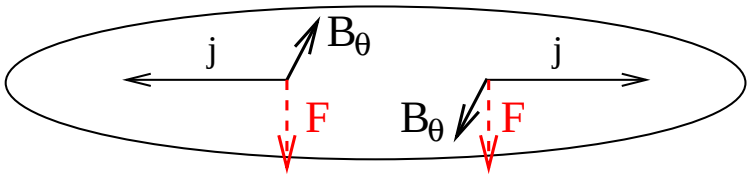


FIG. 7.9 – Force de Lorentz résultant d'une arrivée centrale du courant

Cette intuition est confirmée par des post-traitements supplémentaires :

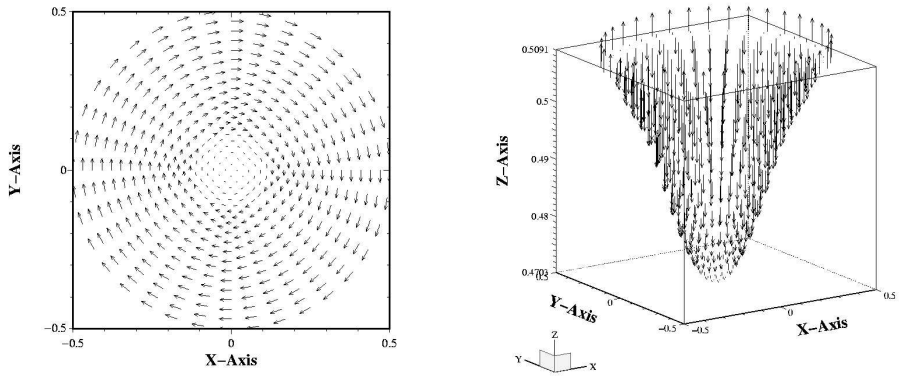


FIG. 7.10 – Champ magnétique et force de Lorentz sur l'interface **aluminium** ($R_i = 0.3$)

Par ailleurs, les conditions interfaciales portant sur le champ électrique (cf. **3.3**) imposent

$$\Big(\frac{J_1}{\sigma_1} - v_1 \times B_1\Big) \times n = \Big(\frac{J_2}{\sigma_2} - v_2 \times B_2\Big) \times n\,,$$

or $v_1 = v_2 = v$, donc

$$\frac{J_2}{\sigma_2} \times n = \frac{J_1}{\sigma_1} \times n - \Big(v \times (B_1 - B_2)\Big) \times n\,,$$

de plus, les conditions interfaciales $B_1 \cdot n = B_2 \cdot n$ et la formule $(a \times b) \times c = (a \cdot c)b - (b \cdot c)a$ impliquent :

$$J_2 \times n = \frac{\sigma_2}{\sigma_1}\, J_1 \times n - \sigma_2\, (v \cdot n)\, (B_2 - B_1)\,;$$

enfin, à l'état stationnaire, $v \cdot n = 0$, et $\sigma_2 \ll \sigma_1$ par hypothèse, donc :

$$\|J_2 \times n\| \simeq 0\,.$$

Ainsi, J_2 et n sont quasiment colinéaires, ce qui revient à dire que l'interface est quasiment perpendiculaire à J_2, d'où le creusement observé si l'on admet le profil d'épanouissement centrifuge du courant. Pour étayer cette supposition, on effectue maintenant des simulations sur l'épanouissement en question.

Un point de vue souvent adopté au sujet des déformées d'interface résultant du profil des anodes consiste à supposer que la différence de potentiel entre l'anode et l'interface est constante à l'équilibre. Ainsi, par la loi d'Ohm, l'interface "tend à se positionner" de manière à se trouver à égale distance de la source de courant en chacun de ses points. Afin de vérifier ces suppositions, nous résolvons l'équation de diffusion du potentiel électrique (7.3), à travers deux fluides de conductivités $\sigma_1 = 10^4$ et $\sigma_2 = 1$ en imposant une différence de potentiel entre l'anode et la cathode :

$$\begin{cases} -\mathrm{div}\,(\sigma \nabla \phi) = 0, & \text{dans } \Omega = [0,4] \times [0, 1 + h_2], \\ \phi = 4, & \text{sur l'anode}, \\ \phi = 0, & \text{sur la cathode}, \\ \dfrac{\partial \phi}{\partial n} = 0, & \text{sur le reste du bord}. \end{cases} \tag{7.4}$$

Ces tests sont effectués sur un rectangle, pour trois tailles de surface émettrice centrée (que nous appelons anode dans le problème (7.4), et deux hauteurs d'électrolyte différentes. On déduit de la

solution le courant électrique en résolvant le problème variationnel :

$$\int_\Omega J \cdot \psi = -\int_\Omega \sigma \nabla\phi \cdot \psi \,, \tag{7.5}$$

où ψ est une fonction test à valeur dans $\mathbb{R}^2$. Le résultat qui en ressort montre bien que le courant s'épanouit à la sortie de l'anode, et, par ailleurs, que plus la hauteur d'életrolyte est petite, plus le courant s'épanouit dans l'alumuinium (cf. FIG. 7.11). Nous renvoyons également en pages 118 et 119 pour des simulations sur l'influence de la hauteur d'électrolyte dans des configurations où la distribution anodique de courant est restreinte.

Nous pensons donc que le **creusement de l'interface observé en 7.2.1 est physiquement justifié par le profil de l'épanouissement du courant électrique sous l'anode**.

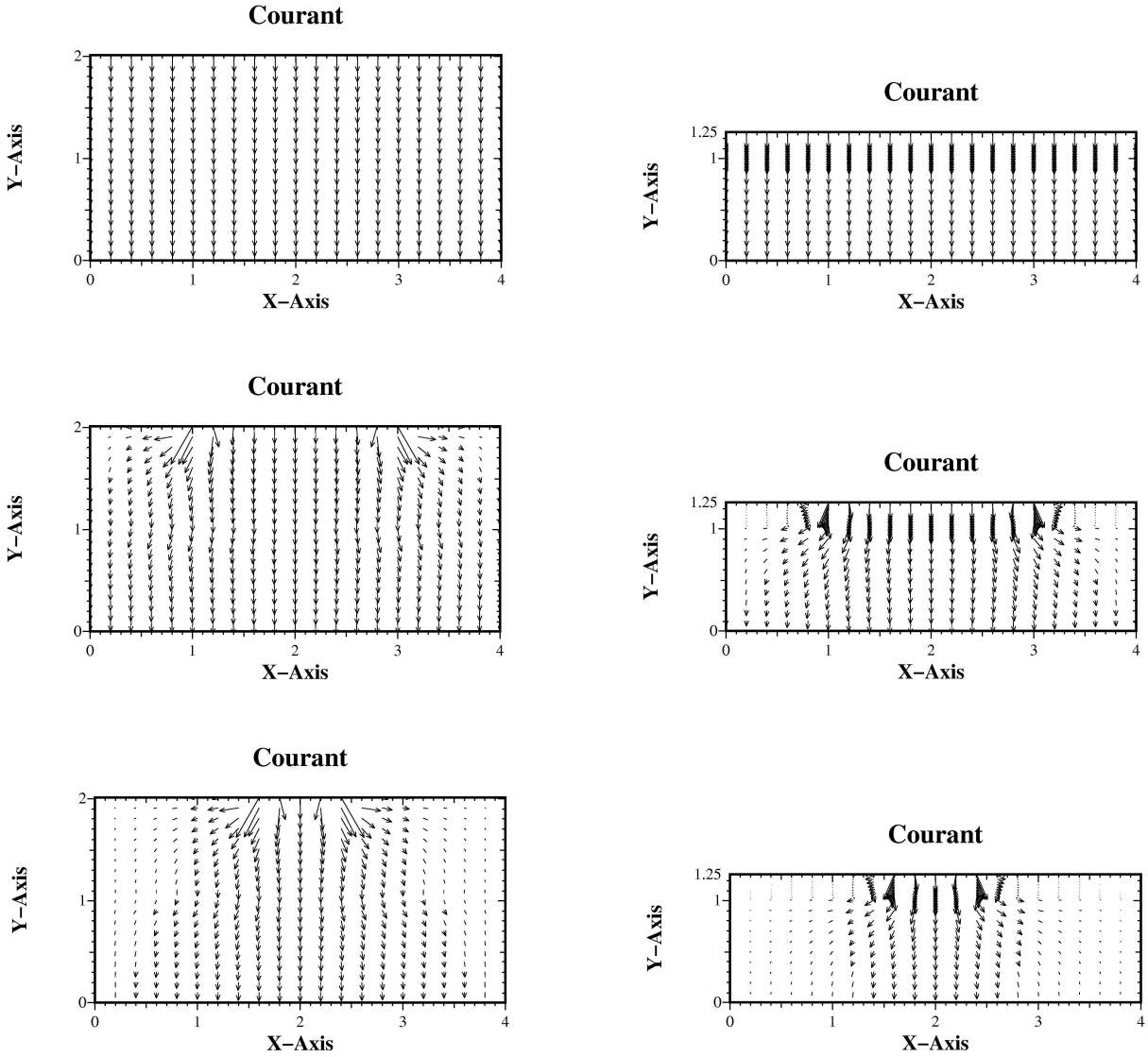


FIG. 7.11 – Calculs de profils d'épanouissement ; de haut en bas : courant uniforme ($R_i = 4$), courant central $R_i = 2$, courant central $R_i = 0.8$; à gauche : $h_2 = 1$, à droite : $h_2 = 0.25$

7.3.2 Stabilisation du phénomène de rolling par un courant central

D'après les résultats ci-dessus, on peut interpréter ce comportement en remarquant qu'une restriction centrale de l'arrivée de courant élecrique à l'anode donne lieu sous l'interface à un courant électrique horizontal centrifuge qui induit, en interagissant avec B_z par la force de Lorentz, une rotation dans l'autre sens que celui du rolling :

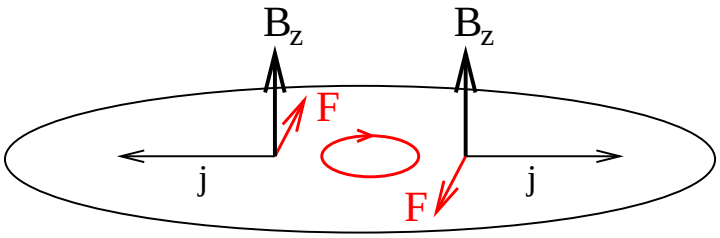


FIG. 7.12 – Sens de rotation induit par une arrivée centrale du courant

où l'on voit que le sens de rotation est opposé à celui schématisé figure 3.4. Ainsi, **une restriction centrale de l'arrivée de courant génère une force de Lorentz opposée au sens de rotation du rolling**.

7.4 Autres simulations

7.4.1 Arrivée de courant "périphérique"

On s'intéresse à un autre type de conditions aux limites magnétiques sur la surface supérieure : c'est en quelque sorte le symétrique de la configuration précédente, où l'on ne fait non plus passer le courant à l'intérieur d'un disque centré, mais exclusivement à l'extérieur :

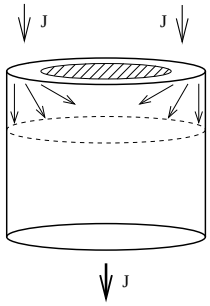


FIG. 7.13 – Restriction périphérique de l'arrivée de courant

Le même raisonnement qu'en **7.2** permet d'établir les expressions suivantes du champ magnétique horizontal à appliquer à la surface de l'anode :

$$\overline{B_\theta(r)} = \begin{cases} 0 & \text{pour} \quad 0 \le r \le R_i \\ -\dfrac{\mu_0\, I}{2\pi(R^2 - R_i^2)} \left(r - \dfrac{R_i^2}{r}\right) & \text{pour} \quad R_i \le r \le R \end{cases}, \tag{7.6}$$

7.4.1.a Sans champ magnétique vertical

Le phénomène observé est le symétrique de celui de la section **7.2.1** : l'interface, au lieu d'être creusée en son centre, se trouve creusée sur sa périphérie :

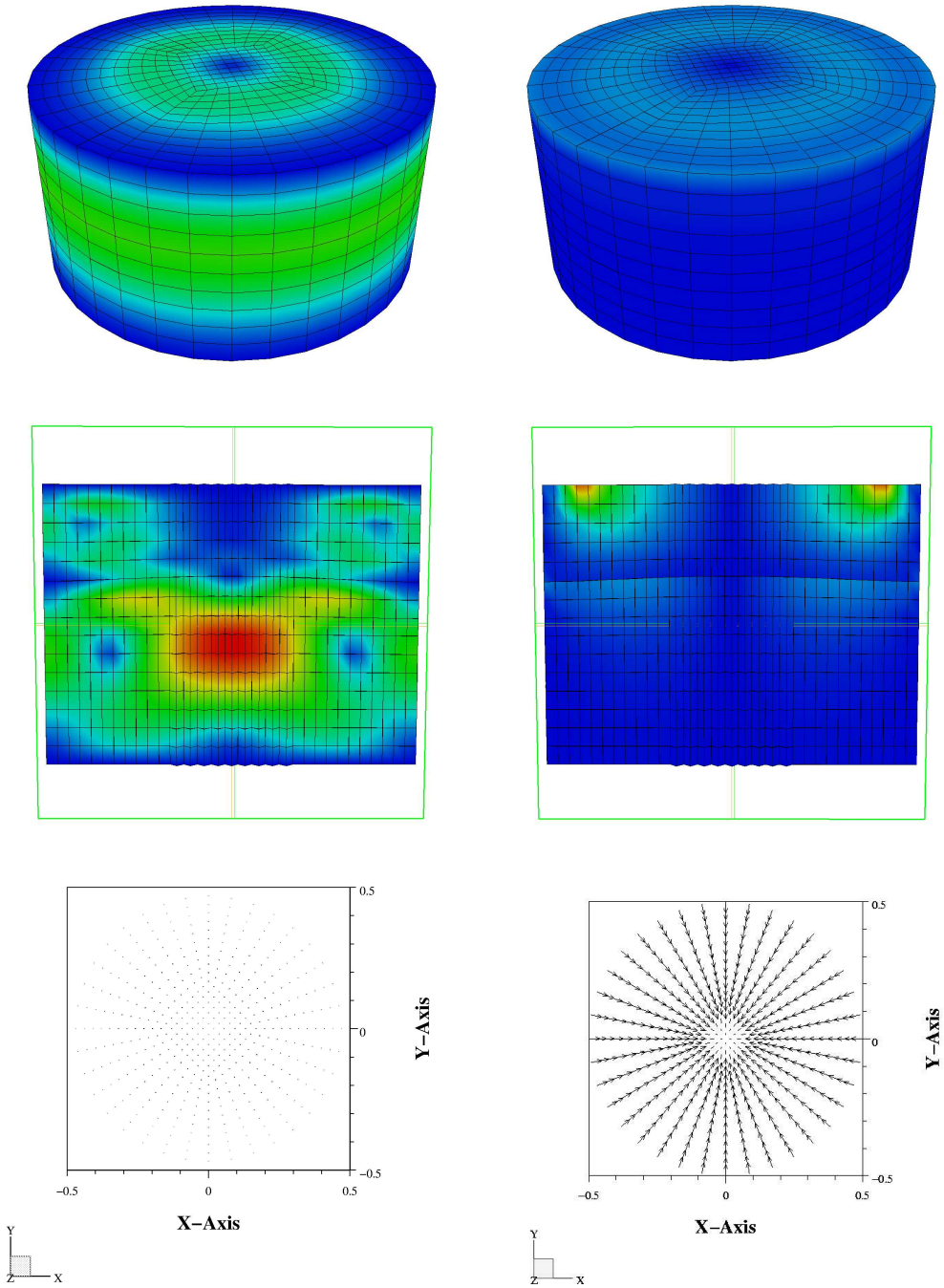


FIG. 7.14 – État stationnaire pour $R_i = 0.4$. À gauche : vitesse (norme autour de l'aluminium, norme sur une coupe verticale, vecteur sur interface). À droite : composante horizontale du courant (norme autour de l'aluminium, norme sur une coupe verticale, vecteur sur interface **aluminium**)

À l'inverse de la situation **7.2.1**, on constate que les vitesses importantes sont localisées dans l'aluminium et non dans l'électrolyte, et que les courants électriques horizontaux sont centripètes sous l'interface.

Enfin, on peut également envisager de contrôler l'intensité de la déformation en agissant sur R_i :

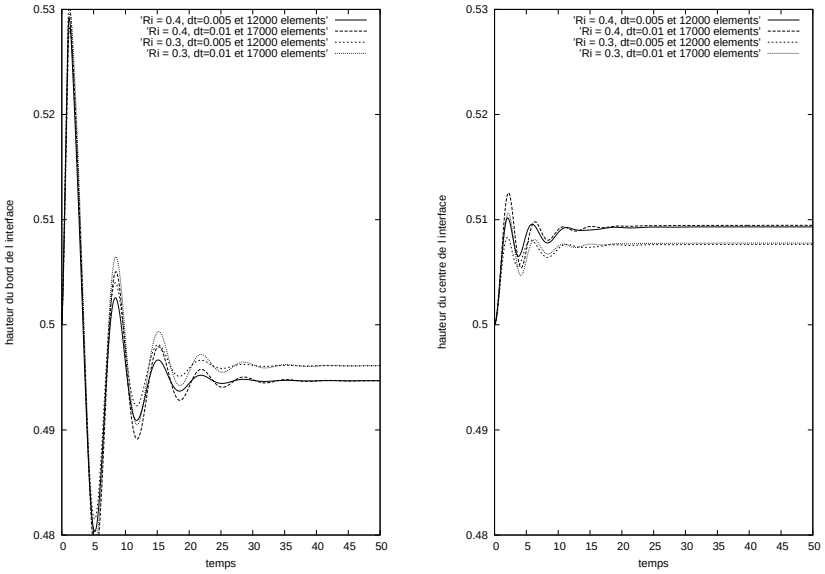
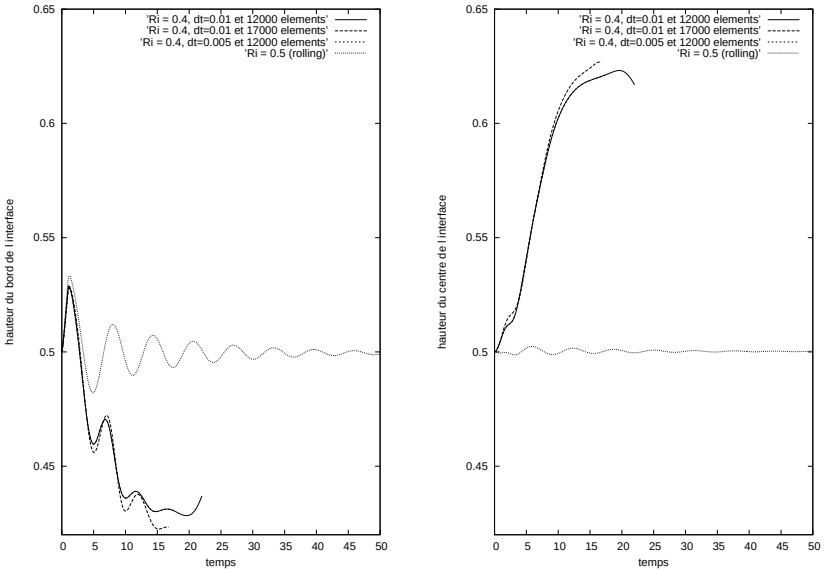


FIG. 7.15 – Hauteurs du bord et du centre de l'interface pour $R_i = 0.3$ et $R_i = 0.4$

7.4.1.b Effet sur la stabilité du rolling

Partant du même cas de rolling de référence qu'en **7.2.2** ($B_z = 0.25$), nous appliquons un courant périphérique avec $R_i = 0.4$. Le comportement observé est une forte déstabilisation, contrairement au cas {courant central, $R_i = 0.3$} à densité de courant égale :



7.4.1.c Une configuration stable

Afin de voir si nous pouvons observer une configuration stable, nous diminuons à présent R_i. Il faut alors descendre jusqu'à $R_i = 0.2$ pour obtenir un tel cas :

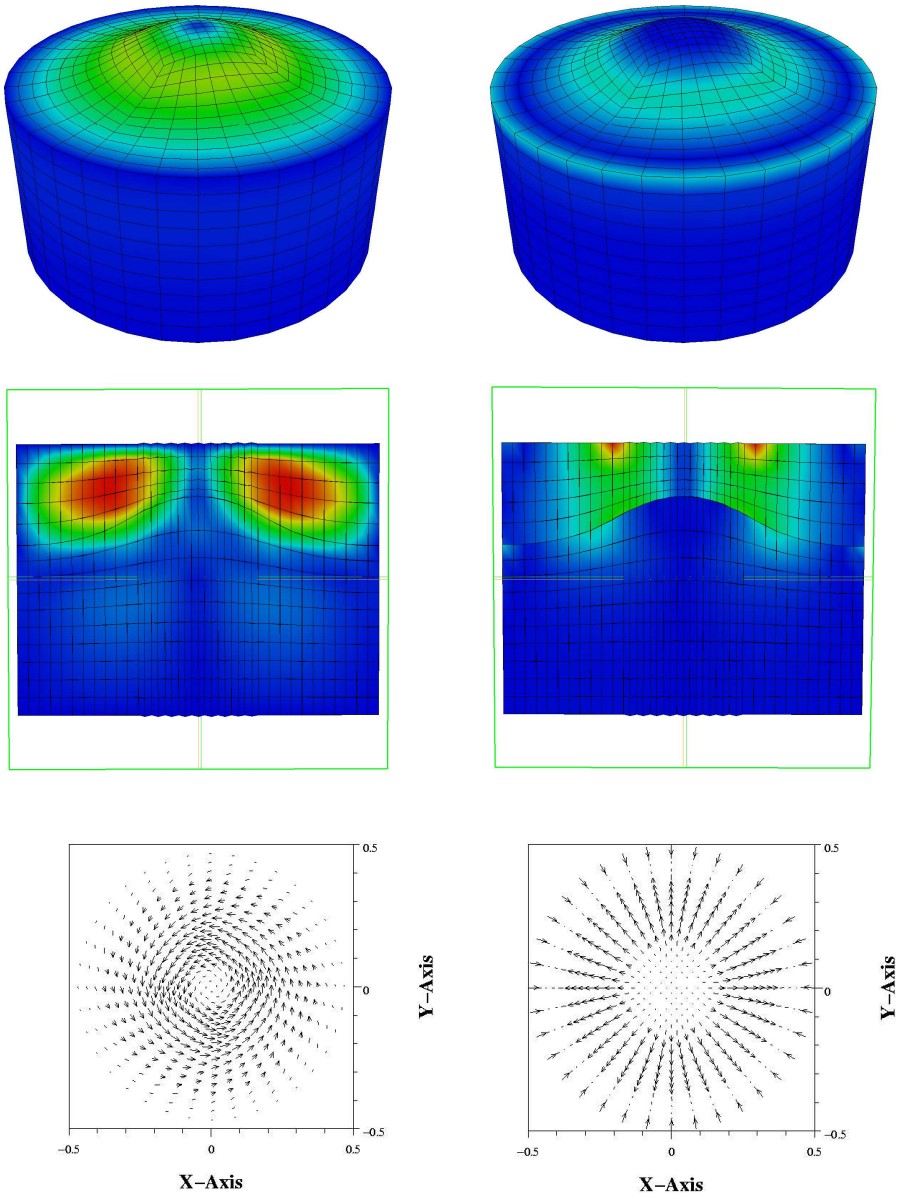


FIG. 7.16 – État stationnaire ($R_i = 0.2$). À gauche : vitesse (norme autour de l'aluminium, norme sur une coupe verticale, composante horizontale sur interface). À droite : composante horizontale du courant (norme autour de l'aluminium, norme sur une coupe verticale, vecteur sur interface **aluminium**)

On observe une aspiration de l'interface très importante, et des distributions de vitesses et courants encore jamais rencontrées.

7.4.2 Influence de la hauteur d'électrolyte

7.4.2.a Sans champ magnétique vertical

Courant central

Nous repartons ici de la configuration **7.2.1** avec R_i fixé à 0.3, et abaissons la hauteur de d'électrolyte. Ces résultats montrent une accentuation du phénomène déjà observé :

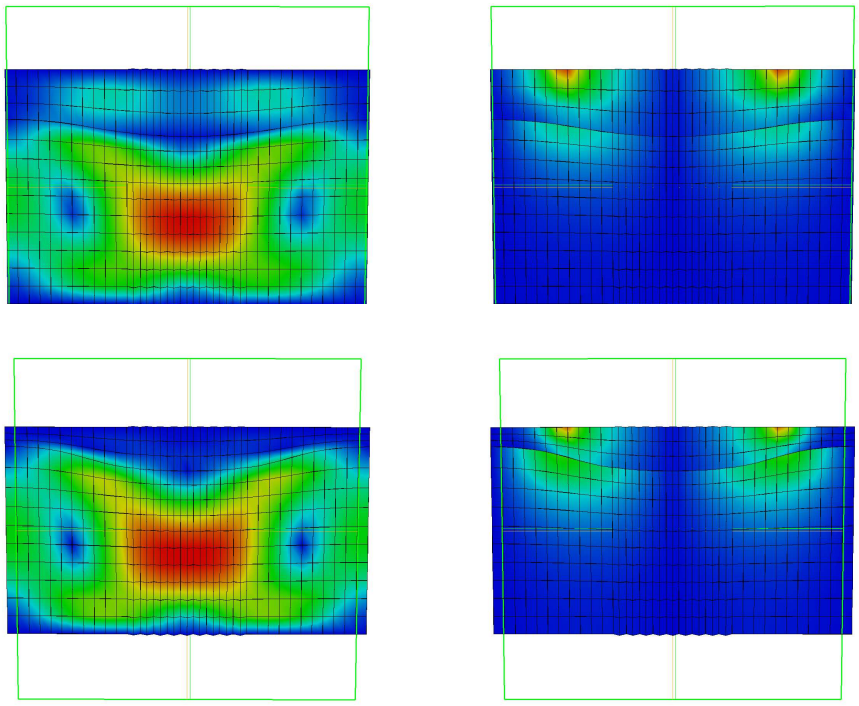


FIG. 7.17 – À gauche : norme de la vitesse sur une coupe, à droite : norme de la composante horizontale du courant électrique sur une coupe ; en haut : $h_2 = 0.15$, en bas : $h_2 = 0.075$)

Une diminution de h_2 engendre :
- une accentuation du *creusement* de l'interface,
- une augmentation des courants électriques *centrifuges* dans l'aluminium,
- une augmentation des vitesses dans l'aluminium.

Courant périphérique

Symétriquement, une diminution de h_2 engendre les mêmes phénomènes d'amplification, mais dans les directions opposées :
- une accentuation du *bombement* de l'interface,
- une augmentation des courants électriques *centripètes* dans l'aluminium,
- une augmentation des vitesses dans l'aluminium.

Dans les deux configurations, le premier point semble le plus intéressant : **pour $B_z = 0$, la diminution h_2 pour $\mathcal{S}_i$ fixé accentue la déformation de l'interface**.

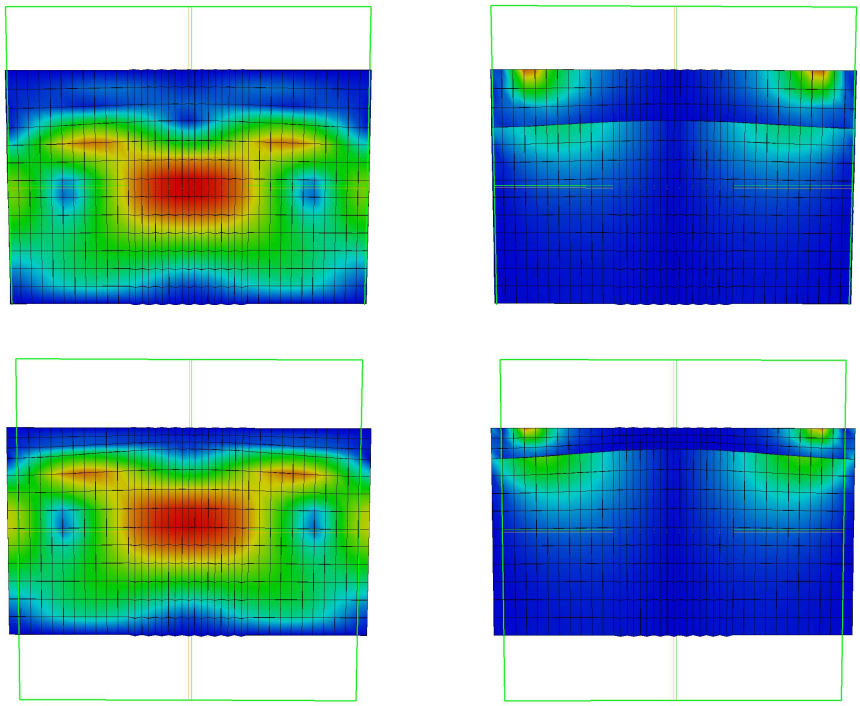


FIG. 7.18 – En haut : $h_2 = 0.15$, en bas : $h_2 = 0.075$

7.4.2.b Avec un champ magnétique vertical

La batterie de tests effectuée montre que d'une manière générale, **diminuer la hauteur d'électrolyte est stabilisant pour les phénomènes**. La déformée d'interface, quant à elle, varie à la fois en fonction de l'intensité du champ magnétique vertical et de la hauteur d'électrolyte. Nous regroupons ci-dessous les déformées d'interface à l'état stationnaire lorsque celui-là existe, les nombres de pas de temps d'atteinte de cet éventuel état, et les sens de rotation des fluides, pour des distributions de courant centrale avec $R_i = 0.3$, uniforme, et périphérique avec $R_i = 0.4$:

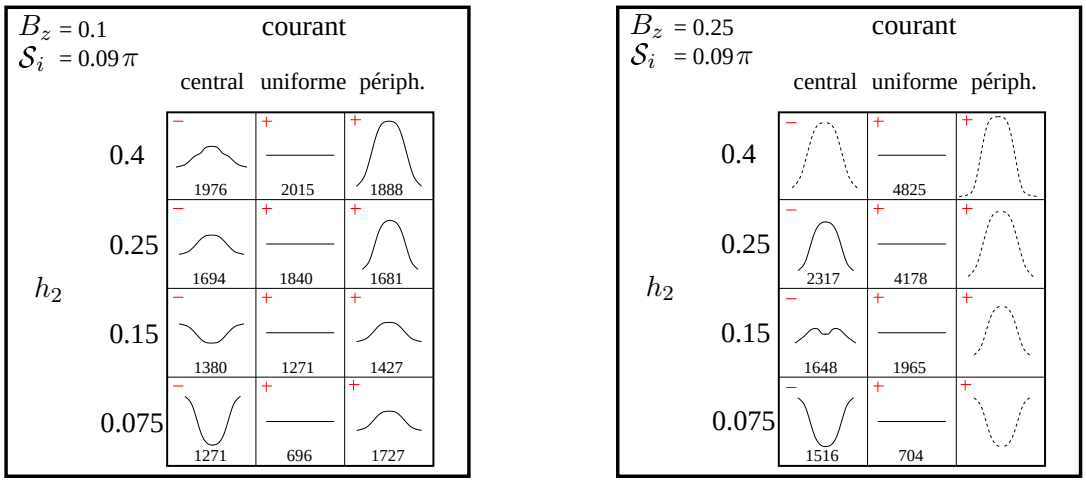


FIG. 7.19 – Synthèse

Les tableaux ci-dessus font ressortir la complexité des phénomènes : on peut par exemple contrôler l'interface à l'état stationnaire par la distribution anodique de courant, mais à condition de ne pas subir un champ magnétique vertical trop important (sinon on ne peut qu'espérer contrôler l'intensité de l'aspiration de l'interface). De plus, si on suppose cette condition remplie, la hauteur moyenne d'électrolyte influence également la forme de l'interface, et peut donc aussi jouer le rôle d'actionneur. Par ailleurs, au sujet de la stabilité, la même profusion de cas possibles se présente : d'une manière générale, restreindre l'arrivée de courant électrique au centre de la surface supérieure a des effets stabilisants sur le rolling, mais on constate que pour $B_z = 0.1$, cette situation ne se présente que pour la hauteur d'électrolyte $h_2 = 0.25$, et qu'elle n'est pas valable pour de très faibles hauteurs d'électrolyte... tout cela sans parler de la distribution périphérique de courant. Il faut donc restreindre soit le nombre d'actionneurs, soit leurs plages de variation pour donner un cadre simple aux futurs problèmes de contrôle. Plusieurs possiblités se présentent alors :

- B_z est faible et on a la capacité de décider de la déformée d'interface, au choix par B_θ ou par h_2,
- B_z est important, auquel cas il est possible de stabiliser le phénomène de rolling, et de contrôler le degré d'aspiration de l'interface en agissant sur B_θ.

Quant à l'influence de B_z, elle est certes déstabilisante, mais on voit mal comment le champ magnétique vertical pourrait faire office de commande. Pour des raisons légèrement différentes (manque de signification physique, cf. **7.1**), il est plus sage de ne pas retenir la diminution de la hauteur d'électrolyte comme commande stabilisante, mais plutôt comme commande pour la déformée d'interface.

7.4.3 Deux fluides, deux sens de rotation opposés

Nous avons vu en **7.4.1** que, d'une part, l'injection d'un courant périphérique donne lieu à une explosion de l'interface en présence d'un champ magnétique vertical, mais qu'en augmentant la taille de la couronne par laquelle passe le courant (donc en diminuant sa densité), on arrive quand même à exhiber un état stable. La situation n'est donc pas simple, car il n'est pas possible de qualifier une telle configuration d'inconditionnellement déstabilisante... Nous terminons cette batterie de tests par l'exposé d'une situtation où la restriction périphérique du passage du courant est appliquée non seulement à l'anode, mais aussi à la cathode (bord inférieur). Avec $B_z = 0.25$ et $h_2 = 0.15$, on obtient l'état stationnaire suivant :

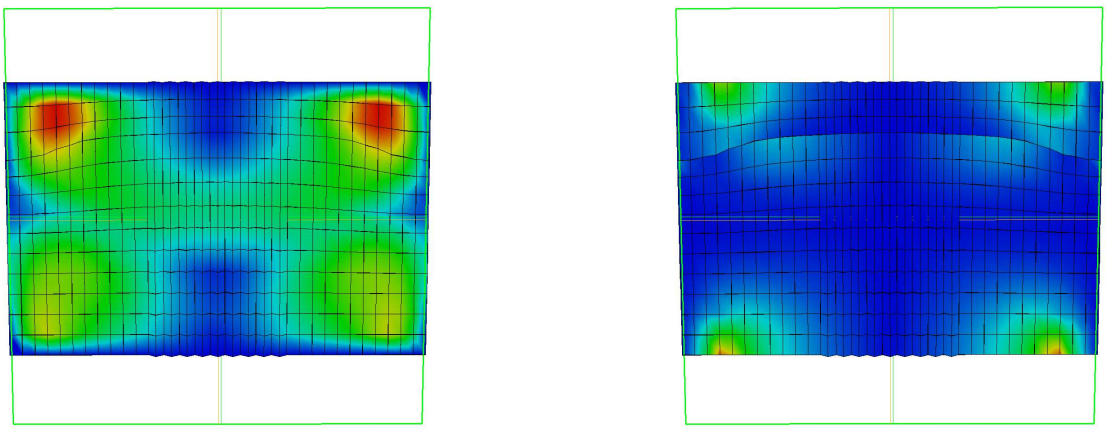


FIG. 7.20 – Normes de la vitesse et de la composante horizontale du courant

Des post-traitements sur la vitesse au niveau des première et dernière mailles en partant du bas montrent que les fluides ont des sens de rotations opposés :

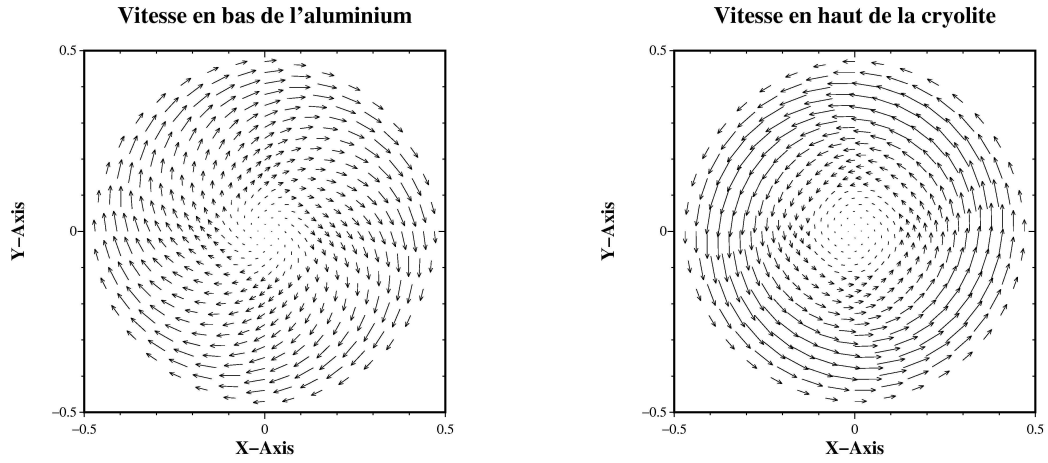


FIG. 7.21 – Vitesses horizontales dans l'aluminium et dans l'électrolyte

Cette observation rejoint celle déjà faite par D. Munger [76], qu'il qualifie d'"aspiration MHD".

7.5 Conclusion

Malgré la variété des situations possibles, la mesure de la surface émettrice de courant (qui se règle à l'aide du scalaire R_i) est un actionneur pertinent pour contrôler les deux aspects suivants :

- STATIQUE : forme de l'interface à l'état stationnaire,
- DYNAMIQUE : stabilité du système (ce qui inclut le temps d'atteinte de l'état stationnaire lorsque celui-là existe).

Le deuxième point est celui qui semble le plus intéressant : l'exemple typique est figure 7.8 page 111 : avec un courant uniforme, on explose, et on arrive à éviter cette explosion avec une arrivée de courant restreinte.

Cependant, notons que le cas statique peut être un problème de contrôle envisageable ; c'est peut-être le plus réaliste : pour de faibles champs magnétiques verticaux et hauteurs de cryolite, nous arrivons à accentuer la déformée d'interface sous la surface anodique en diminuant au choix la mesure de la surface émettrice de courant ou la hauteur d'électrolyte.

Critère / Commande	STATIQUE (forme de l'interface)	DYNAMIQUE (stabilité du système)
$B_\theta(r)$	Si B_z ou h_2 est faible (resp. élevé), un courant restreint creuse (resp. aspire) l'interface sous la surface émettrice d'autant plus que celle-là est faible.	h_2 grand : un courant central stabilise le rolling obtenu avec un courant uniforme ; h_2 petit : un courant uniforme stabilise le rolling.
h_2	Diminuer h_2 renforce le phénomène évoqué ci-dessus.	Augmenter h_2 déstabilise le système (simple phénomène d'inertie ?).
B_z	Un B_z important engendre une aspiration de l'interface.	Augmenter B_z déstabilise le système.

Chapitre 8

Optimisation de forme d'interface en hydrodynamique

L'optimisation de forme est un problème très répandu, notamment dans l'étude du design des avions (cf. B. Mohammadi et O. Pironneau [72]) ou encore en mécanique des structures (cf. G. Allaire [3]). Dans notre cas, il ne s'agit pas à proprement parler d'optimisation de forme car celle-là ne joue pas ici le rôle de commande, mais plutôt de variable d'état dont on veut qu'elle prenne une valeur particulière (problème de poursuite dans lequel le critère est la forme de l'interface). Il n'en demeure pas moins que les mêmes outils qu'en optimisation de forme sont utilisés, à savoir principalement la dérivation par rapport à un domaine.

8.1 Équations d'état et fonction coût

On considère un problème bifluide, dont on cherche à contrôler la position de l'interface par une force volumique $u(x, z)$. Le domaine volumique Ω de contour $\partial\Omega$ (de normale n et tangente t) se divise en deux sous-domaines Ω_1 et Ω_2. Ceux-là sont séparés par l'interface Σ (dont on note n_Σ la normale extérieure au fluide 1), paramétrée par la fonction $x \mapsto h(x)$. On sait que pour que le problème soit bien posé, la prise en compte de la convection est nécessaire dans le modèle (cf. D. Errate *et al.* [30]). Ainsi, on considère le problème de contrôle d'interface fluide-fluide le plus simple possible, basé sur un écoulement régi par les équations de Navier-Stokes en dimension 2.

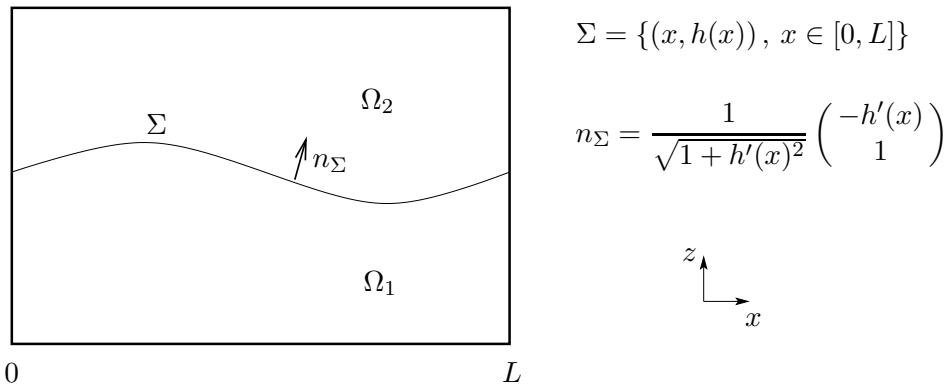


FIG. 8.1 – Définition du problème

Ainsi, densité ρ et viscosité η dépendent de h :

$$\rho(x,z,h(x)) = \begin{cases} \rho_1 & \text{si} \quad z < h(x) \\ \rho_2 & \text{si} \quad z > h(x) \end{cases}, \quad \text{et} \quad \eta(x,z,h(x)) = \begin{cases} \eta_1 & \text{si} \quad z < h(x) \\ \eta_2 & \text{si} \quad z > h(x) \end{cases}.$$

On cherche la vitesse v, la pression p et la hauteur d'interface h telles que (les dépendances des variables sont omises pour des raisons de lisibilité : v et p dépendent de x et z, h dépend de x, et Σ dépend de h) :

$$\left\{ \begin{array}{rcccc} \rho\, v \cdot \nabla v - \mathrm{div}\,[\eta\, D(v)] + \nabla p & = & \rho\, g + u & \text{dans} & \Omega\,, \\ -\mathrm{div}\, v & = & 0 & \text{dans} & \Omega\,, \\ v \cdot n & = & 0 & \text{sur} & \partial\Omega\,, \\ \eta\, D(v) n \cdot t & = & 0 & \text{sur} & \partial\Omega\,. \\ v \cdot n_\Sigma & = & 0 & \text{sur} & \Sigma\,, \\ \int_0^L h & = & V_0^{\mathrm{alu}} & \text{fixé}, & \end{array} \right. \tag{8.1}$$

où $D : v \mapsto \nabla v + (\nabla v)^T$. Ainsi, on se propose de résoudre le problème de minimisation (avec $x \mapsto h_0(x)$ fixée telle que $\int_0^L h_0 = V_0^{\mathrm{alu}}$) :

$$\underset{u \in L^2(\Omega)^2}{\mathrm{Inf}}\ J(u) = \frac{1}{2}\int_0^L (h - h_0)^2 + \frac{Q}{2}\int_\Omega \|u\|^2\,, \tag{8.2}$$

où h est donnée par les équations d'état (8.1). Noter que par la suite, la propriété $\int_0^L h_0 = V_0^{\mathrm{alu}}$ n'est pas utilisée (ce qui est étonnant).

8.2 Problème adjoint

On écrit la fonction de Lagrange $\mathcal{L}$ du problème d'optimisation (8.1)-(8.2) en définissant cinq multiplicateurs de Lagrange $(\alpha, \beta, \gamma, \mu, \nu)$. Cela en fait un de moins que le nombre d'équations d'état car on décide de prendre v dans un espace fonctionnel tel que $v \cdot n = 0$ sur $\partial\Omega$; alors, si l'on choisit le multiplicateur α (associé à la première équation d'état) tel que

$$\alpha \cdot n = 0 \quad \text{sur} \quad \partial\Omega\,, \tag{8.3}$$

il n'est pas nécessaire de tenir compte de $v \cdot n = 0$ sur $\partial\Omega$ dans le lagrangien. Par ailleurs, β varie dans Ω, γ sur $\partial\Omega$, μ sur Σ, et ν est une constante. Ainsi,

$$\begin{aligned} \mathcal{L}(v,p,h,u,\alpha,\beta,\gamma,\mu,\nu) = J(u) - \int_\Omega \alpha \cdot \Big[\rho\, v \cdot \nabla v - \mathrm{div}\,[\eta\, D(v)] + \nabla p - \rho g - u\Big] - \int_\Omega \beta\,(-\mathrm{div}\, v) \\ - \int_{\partial\Omega} \gamma\, \eta\, D(v)\, n \cdot t - \int_\Sigma \mu\, v \cdot n_\Sigma - \nu\Big(\int_0^L h - V_0^{\mathrm{alu}}\Big)\,. \end{aligned}$$

Par rapport à v : la dérivée de $\mathcal{L}$ par rapport à v doit s'annuler dans toutes les directions $\tilde{v}$ telles que $\tilde{v} \cdot n = 0$ sur $\partial\Omega$:

$$\begin{aligned} &\lim_{\varepsilon \to 0}\left[\frac{\mathcal{L}(v + \varepsilon\,\tilde{v}, \cdot) - \mathcal{L}(v, \cdot)}{\varepsilon}\right] = 0\,, \quad \forall\,\varepsilon\,\forall\,\tilde{v}\,, \quad \text{soit} \\ &\int_\Omega \Big[\rho\,[v \cdot \nabla\alpha - (\nabla v)^T\alpha] + \mathrm{div}\,[\eta\, D(\alpha)] - \nabla\beta + \delta_\Sigma\,[(\rho_2 - \rho_1) v \cdot \alpha - \mu]\, n_\Sigma\Big] \cdot \tilde{v} \\ &+ \int_{\partial\Omega} (\alpha \cdot t - \gamma)\,[\eta\, D(\tilde{v})\, n \cdot t\,] - \int_{\partial\Omega} [\eta\, D(\alpha)\, n \cdot t\,]\,(\tilde{v} \cdot t) = 0\,, \quad \forall\,\tilde{v}\,, \end{aligned} \tag{8.4}$$

par intégrations par parties. Le symbole δ_Σ désigne la mesure de Dirac sur la surface Σ :

$$\langle \delta_\Sigma \, , \varphi \rangle = \int_\Omega \delta_\Sigma \, \varphi = \int_\Sigma \varphi \, .$$

On voit alors que $\gamma = \alpha \cdot t$ sur $\partial\Omega$, ce qui permet de récrire rigoureusement la fonction de Lagrange sous la forme :

$$\mathcal{L}(v,p,h,u,\alpha,\beta,\mu,\nu) = J(u) - \int_\Omega \eta \, \nabla\alpha : [D(v) - pI] - \int_\Omega \alpha \cdot (\rho \, (v \cdot \nabla v - g) - u) - \int_\Omega \beta \, (-\mathrm{div}\, v) \\ - \int_\Sigma \mu \, v \cdot n_\Sigma - \nu \Big(\int_0^L h - V_0^{\mathrm{alu}} \Big) \, ,$$

avec v, $\alpha \in \mathbb{H}^1_n(\Omega) = \{ \tilde{v} \in H^1(\Omega)^2 \, , \, \tilde{v} \cdot n = 0 \text{ sur } \partial\Omega \}$; p, $\beta \in L^2(\Omega)$; u, h et μ suffisamment réguliers. Nous utiliserons désormais cette formulation dans toute la suite.

Par rapport à p : par un procédé similaire,

$$\partial_p \mathcal{L} = 0 \Leftrightarrow \int_\Omega \mathrm{div}\, \alpha \, \tilde{p} = 0 \, , \quad \forall \, \tilde{p} \, . \tag{8.5}$$

Par rapport à h : la dérivée de $\mathcal{L}$ par rapport à h dans toute direction $\tilde{h}$ doit s'annuler :

$$\lim_{\varepsilon \to 0} \left[\frac{\mathcal{L}(v,p,h+\varepsilon\,\tilde{h},\cdot) - \mathcal{L}(v,p,h,\cdot)}{\varepsilon} \right] = 0 \, , \quad \forall \, \varepsilon \, \forall \, \tilde{h} \, , \quad \text{soit}$$

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[\varepsilon \int_\Omega [(\partial_z \rho) \, (g - v \cdot \nabla v) - (\partial_z \eta) \, \nabla\alpha : D(v)] \, \tilde{h} \, + o(\varepsilon) \right. \\ \left. - \int_{\Sigma(h+\varepsilon\tilde{h})} \mu \, v \cdot n_\Sigma + \int_{\Sigma(h)} \mu \, v \cdot n_\Sigma + \varepsilon \int_0^L (h - h_0 - \nu) \, \tilde{h} \right] = 0 \, , \tag{8.6}$$

où les développements limités au premier ordre de η et ρ sont définis grâce à la formule des sauts, à savoir le théorème donnant la dérivée (au sens des distributions) d'une fonction de $\mathbb{R}^N$ discontinue sur une surface : si φ est une fonction localement intégrable, régulière séparément sur les domaines Ω_1 et Ω_2 séparés par la frontière régulière Σ de normale n_Σ extérieure à Ω_1,

$$\partial_i \varphi = \{ \partial_i \varphi \} + n_\Sigma \cdot e_i (\varphi_2 - \varphi_1) \, \delta_\Sigma \, ,$$

où $\{\partial_i \varphi\}$ est la fonction définie presque partout comme étant égale à la dérivée usuelle de φ sur chaque domaine séparément, et $\varphi_{\{1,2\}}$ est le prolongement par continuité de $\varphi_{|\Omega_{\{1,2\}}}$ dans $\overline{\Omega}_{\{1,2\}}$. Ainsi, dans notre cas, on peut écrire les développements limités au premier ordre :

$$\begin{cases} \eta(x,z,h+\epsilon\tilde{h}) - \eta(x,z,h) \;=\; (\partial_z \eta) \, \varepsilon \tilde{h} + o(\varepsilon) \, , \;\; \text{avec} \;\; \partial_z \eta = \dfrac{(\eta_2 - \eta_1) \, \delta_\Sigma}{\sqrt{1+h'^2}} \\ \rho(x,z,h+\epsilon\tilde{h}) - \rho(x,z,h) \;=\; (\partial_z \rho) \, \varepsilon \tilde{h} + o(\varepsilon) \, , \;\; \text{avec} \;\; \partial_z \rho = \dfrac{(\rho_2 - \rho_1) \, \delta_\Sigma}{\sqrt{1+h'^2}} \end{cases} . \tag{8.7}$$

Concernant les termes en μ, on remarque que :

$$\int_\Sigma \mu v \cdot n_\Sigma = \int_{\Omega_1} \mathrm{div}\, (\mu v) \, ,$$

et on fait appel au théorème 4.2 du livre [24] de M.C. Delfour et J.-P. Zolésio, qui nous apprend que si Ω est transporté par un champ de vitesses V, et transformé ainsi en un domaine Ω_t, alors la dérivée de la fonction :

$$J_V(t) = \int_{\Omega_t} \varphi(t,x)\,dx$$

en $t=0$ est donnée par :

$$\boxed{dJ_V(0) = \int_\Omega \varphi'(0,x)\,dx + \int_{\partial\Omega} \varphi(0,x)\,V(0,x)\cdot n\,dx\,.}$$

On renvoie à **8.4** pour une approche n'utilisant pas ce résultat. Dans notre cadre, cette formule donne (en prenant $t=\varepsilon$ et $V=(0,\tilde h)^T$) :

$$\lim_{\varepsilon\to 0}\frac{1}{\epsilon}\left[\int_{\Omega_1(h+\varepsilon\tilde h)} \mathrm{div}\,(\mu v) - \int_{\Omega_1(h)} \mathrm{div}\,(\mu v)\right] = \int_\Sigma \mathrm{div}\,(\mu v)\,\frac{\tilde h}{\sqrt{1+h'^2}}\,. \tag{8.8}$$

Il reste à injecter les formules (8.7) et (8.8) dans (8.6) pour obtenir :

$$\int_\Sigma \Big[-(\eta_2-\eta_1)\,\nabla\alpha : D(v) + \alpha\cdot(\rho_2-\rho_1)(g - v\cdot\nabla v) - \mathrm{div}\,(\mu v)\Big]\,\frac{\tilde h}{\sqrt{1+h'^2}} + \int_0^L (h-h_0-\nu)\,\tilde h = 0\,,\quad \forall\,\tilde h\,,$$

que la paramétrisation de Σ par h dans la première intégrale permet d'écrire sous la forme (en utilisant également $\mathrm{div}\, v = 0$) :

$$\int_0^L \Big[-(\eta_2-\eta_1)\,\nabla\alpha : D(v) + \alpha\cdot(\rho_2-\rho_1)(g - v\cdot\nabla v) - v_x\,\partial_x\mu + h - h_0 - \nu\Big]\,\tilde h = 0\,,\quad \forall\,\tilde h\,, \tag{8.9}$$

Précisons que v et α sont systématiquement pris au point $(x,h(x))$ lorsqu'ils apparaissent sous le signe $\int_0^L$.

On regroupe à présent les formulations fortes des trois équations (8.4), (8.5) et (8.9) obtenues par dérivation du lagrangien par rapport aux trois variables d'état et la condition (8.3) pour obtenir le problème adjoint suivant (dans les variables α, β, μ et ν) :

$$\left\{\begin{array}{rcccc}
\rho\,[(\nabla v)^T\alpha - v\cdot\nabla\alpha] - \mathrm{div}\,[\eta\,D(\alpha)] + \nabla\beta + \delta_\Sigma\,(\mu - v\cdot\alpha)\,n_\Sigma & = & 0 & \text{dans} & \Omega \\
\mathrm{div}\,\alpha & = & 0 & \text{dans} & \Omega \\
\alpha\cdot n & = & 0 & \text{sur} & \partial\Omega \\
[\eta\,D(\alpha)]\,n\cdot t & = & 0 & \text{sur} & \partial\Omega \\
(\eta_2-\eta_1)\,\nabla\alpha : D(v) - \alpha\cdot(\rho_2-\rho_1)(g - v\cdot\nabla v) + v_x\partial_x\mu\ + \nu & = & h-h_0 & \text{sur} & [0,L]
\end{array}\right. \tag{8.10}$$

(rappelons que $\gamma = \alpha\cdot t$ sur $\partial\Omega$).

8.3 Équations de sensitivité

La variation de (v,p,h) en fonction de celle de la commande u est contrainte par l'équation d'état, autrement dit pout toute perturbation $\tilde u$ de u, il existe une perturbation $(\tilde v,\tilde p,\tilde h)$ de (v,p,h) satisfaisant les équations de sensitivité (on note $\mathcal{F}=0$ l'équation d'état) :

$$\lim_{\varepsilon\to 0}\frac{\mathcal{F}(v+\varepsilon\tilde v, p+\varepsilon\tilde p, h+\varepsilon\tilde h, u+\varepsilon\tilde u) - \mathcal{F}(v,p,h,u)}{\varepsilon} = 0 \tag{8.11}$$

Comme dans la dérivation de la fonction de Lagrange par rapport à h, il sort de la première équation des termes surfaciques en raison de la discontinuité des densité et viscosité (en passant par une formulation faible de la première équation d'état). En dehors de cela, la subtilité réside dans la dérivation de la condition de non-pénétration à l'interface : $v \cdot n_\Sigma = 0$ sur Σ. Alors, dans l'état perturbé, $v + \varepsilon\tilde{v}$ est pris sur $\Sigma(h + \varepsilon\tilde{h})$ donc s'écrit

$$v[x, h(x) + \varepsilon\tilde{h}(x)] + \varepsilon\tilde{v}[x, h(x) + \varepsilon\tilde{h}(x)] = v[x, h(x)] + \partial_z v[x, h(x)]\,\varepsilon\tilde{h}(x) + \varepsilon\tilde{v}[x, h(x)] + \mathcal{O}(\varepsilon^2)\,.$$

Il reste à exprimer $n_{\Sigma(h+\varepsilon\tilde{h})}$ en fonction de $n_{\Sigma(h)}$, en remarquant que :

$$\frac{1}{\sqrt{1 + (h + \varepsilon\tilde{h})'^2}} = \frac{1}{\sqrt{1 + h'^2}} \left(1 + \varepsilon \frac{2h'\tilde{h}'}{1 + h'^2} + \mathcal{O}(\varepsilon^2)\right)^{-\frac{1}{2}}.$$

Le développement limité au premier ordre de l'expression entre parenthèses conduit à :

$$n_{\Sigma(h+\varepsilon\tilde{h})} = n_{\Sigma(h)} \left(1 - \varepsilon\, \frac{h'\tilde{h}'}{1 + h'^2}\right) - \varepsilon\, \frac{\tilde{h}'\, e_x}{\sqrt{1 + h'^2}} + \mathcal{O}(\varepsilon^2)\,,$$

et comme $v \cdot n_\Sigma = 0$ sur Σ, (8.11) prend la forme :

$$\left\{\begin{array}{rcl} \rho(v \cdot \nabla\tilde{v} + \tilde{v} \cdot \nabla v) - \mathrm{div}\,[\eta\, D(\tilde{v})] + \nabla\tilde{p} & & \\ -\Big\{\mathrm{div}\,[(\eta_2 - \eta_1)\, D(v)] + (\rho_2 - \rho_1)\,(g - v \cdot \nabla v)\Big\}\, \dfrac{\delta_\Sigma\, \tilde{h}}{\sqrt{1 + h'^2}} & = & \tilde{u} \quad \text{dans} \quad \Omega \\ -\mathrm{div}\,\tilde{v} & = & 0 \quad \text{dans} \quad \Omega \\ \tilde{v} \cdot n & = & 0 \quad \text{sur} \quad \partial\Omega \\ \eta\, D(\tilde{v}) n \cdot t & = & 0 \quad \text{sur} \quad \partial\Omega \\ \partial_z v \cdot n_\Sigma\, \tilde{h} + \tilde{v} \cdot n_\Sigma - \dfrac{v_x \tilde{h}'}{\sqrt{1 + h'^2}} & = & 0 \quad \text{sur} \quad \Sigma(h) \\ \displaystyle\int_0^L \tilde{h} & = & 0 \end{array}\right. \tag{8.12}$$

8.4 Gradient de la fonction coût

Nous calculons le gradient $\nabla_u J$ du critère, en explicitant la dérivée directionnelle du critère par rapport à u, sachant que les directions $\tilde{v}$, $\tilde{p}$ et $\tilde{h}$ sont liées à u par (8.12) :

$$D_u J = \lim_{\varepsilon \to 0} \frac{J(v + \varepsilon\tilde{v}, p + \varepsilon\tilde{p}, h + \varepsilon\tilde{h}, u + \varepsilon\tilde{u}, \cdot) - J(v, p, h, u, \cdot)}{\varepsilon} = \int_0^L (h - h_0)\, \tilde{h} + Q \int_\Omega u \cdot \tilde{u}$$

On combine cette expression avec (8.9) puis on intègre $\nabla\alpha$ et on utilise les première et dernière équations de sensitivité (cf. (8.12)) :

$$\begin{aligned} D_u\, J &= \int_0^L [\,(\eta_2 - \eta_1)\, \nabla\alpha : D(v) - \alpha \cdot (\rho_2 - \rho_1)(g - v \cdot \nabla v) + v_x\, \partial_x\mu + \nu\,]\, \tilde{h} + Q \int_\Omega u \cdot \tilde{u} \\ &= \int_\Omega \alpha \cdot \Big[- \rho(v \cdot \nabla\tilde{v} + \tilde{v} \cdot \nabla v) + \mathrm{div}\,[\eta\, D(\tilde{v})] - \nabla\tilde{p} + \tilde{u}\Big] + \int_0^L v_x\, \partial_x\mu\, \tilde{h} + Q \int_\Omega u \cdot \tilde{u}\,, \end{aligned}$$

$$\text{or} \quad \int_0^L v_x\, \partial_x\mu\, \tilde{h} = - \int_0^L \mu\, \partial_x\, (v_x\, \tilde{h}) \quad \text{où, rappelons-le, } v_x \text{ est pris au point } (x, h(x))\,,$$

et $\partial_x[v_x(x,h(x)] = \partial_x v_x(x,h(x)) + \partial_z\, v_x(x,h(x))\, h'(x) = -\partial_z[-v_x(x,h(x))\, h'(x) + v_z(x,h(x)]$ (étant donné que $\operatorname{div} v = 0$), donc

$$\int_0^L v_x\, \partial_x \mu\, \tilde{h} = \int_0^L \mu\, [\partial_z v \cdot n_\Sigma\, \tilde{h}\, \sqrt{1+h'^2} - v_x\, \tilde{h}']\,.$$

Alors, d'après l'équation posée sur Σ dans (8.12),

$$\int_0^L v_x\, \partial_x \mu\, \tilde{h} = -\int_\Sigma \mu\, \tilde{v} \cdot n_\Sigma\, = -\int_\Omega \delta_\Sigma \mu\, n_\Sigma \cdot \tilde{v}\,.$$

On intègre les dérivées sur $\tilde{v}$ et $\tilde{p}$, les termes de bord disparaissant du fait de (8.12) et (8.10) :

$$\begin{aligned} D_u\, J \;\; &= \;\; \int_\Omega \Big[\rho[v\cdot\nabla\alpha - (\nabla v)^T\alpha] + \operatorname{div}\,[\eta\, D(\alpha)] + \delta_\Sigma (v\cdot\alpha) n_\Sigma\Big]\cdot\tilde{v} + \int_\Omega \operatorname{div}\alpha\, \tilde{p} + \int_\Omega \alpha\cdot\tilde{u} \\ &\quad - \int_\Omega \delta_\Sigma \mu\, n_\Sigma\cdot\tilde{v} + Q\int_\Omega u\cdot\tilde{u}\,. \end{aligned}$$

Alors, la prise en compte des équations volumiques de (8.10) permet d'écrire :

$$D_u\, J \;\; = \;\; \int_\Omega (\nabla\beta + \delta_\Sigma\, \mu\, n_\Sigma)\cdot\tilde{v} - \int_\Omega \delta_\Sigma \mu\, n_\Sigma\cdot\tilde{v} + \int_\Omega (\alpha + Q\, u)\cdot\tilde{u}\,.$$

et (en intégrant le terme en β et en tenant compte une dernière fois des équations (8.12)) :

$$D_u J = \int_\Omega (\alpha + Qu)\cdot\tilde{u}\,, \;\; \text{soit} \;\; \boxed{\nabla_u\, J = \alpha + Q\, u\,.}$$

Remarque

On peut calculer par une méthode plus classique, dans la dérivée du lagrangien par rapport h (voir **8.2**), le terme contenant l'intégrale sur $\Sigma(h)$. Par le changement de variable $x \mapsto h(x) = z$, de telle sorte qu'on se ramène sur le segment $[0, L]$, on peut écrire :

- $\displaystyle\int_{\Sigma(h)} \mu\, v\cdot n_\Sigma \;\; = \;\; \int_0^L \mu(x)\, [-v_x(x,h(x))\, h'(x) + v_z(x,h(x))]\, dx\,,$

et on procède de même sur $\Sigma(h+\varepsilon\tilde{h})$:

$$\begin{aligned} \bullet\;\; \int_{\Sigma(h+\varepsilon\tilde{h})} \mu\, v\cdot n_\Sigma \;\; &= \;\; \int_0^L \mu(x)\, [-v_x(x,(h+\varepsilon\tilde{h})(x))\, (h+\varepsilon\tilde{h})'(x) + v_z(x,(h+\varepsilon\tilde{h})(x))]\, dx \\ &= \;\; \int_0^L \mu(x)\Big[[-v_x(x,h(x)) - \partial_z v_x(x,h(x))\, (\varepsilon\tilde{h}(x)) + \mathcal{O}(\varepsilon^2)]\, [h'(x) + \varepsilon\tilde{h}'(x)] \\ &\qquad\qquad + v_z(x,h(x)) + \partial_z v_z(x,h(x))\, (\varepsilon\tilde{h}(x)) + \mathcal{O}(\varepsilon^2)\Big]\, dx \\ &= \;\; \int_{\Sigma(h)} \mu\, v\cdot n_{\Sigma(h)}\; - \varepsilon\int_0^L \mu(x)\, v_x(x,h(x))\, \tilde{h}'(x)\, dx \\ &\quad + \varepsilon\int_0^L \mu(x)[-\partial_z v_x(x,h(x))\, h'(x) + \partial_z v_z(x,h(x))]\tilde{h}(x)\, dx + \mathcal{O}(\varepsilon^2)\,, \end{aligned}$$

donc (8.6) s'écrit :

$$\int_\Sigma \alpha \cdot \Big\{ \operatorname{div}\left[(\eta_2-\eta_1)\, D(v)\right] + (\rho_2-\rho_1) g \Big\} \frac{\tilde h}{\sqrt{1+h'^2}} - \int_{\Sigma(h)} \mu\, \partial_z v \cdot n_{\Sigma(h)}\, \tilde h + \int_0^L \mu\, v_x\, \tilde h' + \int_0^L (h - h_0 - \nu)\, \tilde h = 0\,, \quad \forall\, \tilde h$$

or, rappelons que (cf. **8.4**) :

$$\int_0^L v_x\, \partial_x \mu\, \tilde h = \int_0^L \mu\, [\partial_z v \cdot n_\Sigma\, \tilde h\, \sqrt{1+h'^2} - v_x\, \tilde h']\,,$$

ce qui permet de retrouver (8.9) à l'aide de la paramétrisation de Σ par $h(x)$.

8.5 Une piste pour implémenter le problème adjoint

Tandis qu'on peut résoudre l'équation d'état par une méthode ALE (cf. **3.3.1**) en temps long, il semble ne pas exister de méthode standard pour résoudre le problème adjoint (8.10), qui met en jeu quatre inconnues α, β, μ et ν. Cependant, on remarque que ν est un scalaire, ce qui implique que le reste de la dernière équation du problème adjoint doit être constant. On propose donc de chercher μ de sorte que cette l'expression

$$\psi(\alpha,\mu) = -\alpha(x,h(x)) \cdot (\rho_2-\rho_1)\,(g - v\cdot\nabla v) + v_x(x,h(x))\partial_x\mu(x) - (h(x)-h_0(x))\,, \tag{8.13}$$

soit égale à sa moyenne spatiale. Pour cela, on cherche à résoudre le sous-problème de contrôle

$$\left\{ \begin{array}{l} \operatorname*{Inf}_\mu J(\alpha,\mu)\,, \quad \text{avec} \quad J(\alpha,\mu) = \dfrac{1}{2}\displaystyle\int_0^L \left[\psi(\alpha(x,h(x)),\mu(x)) - \frac{1}{L}\int_0^L \psi(\alpha(x,h(x)),\mu(x))\,dx\right]^2 dx \\ \qquad \left\{ \begin{array}{rcl} \rho\,[(\nabla v)^T\alpha - v\cdot\nabla\alpha] - \operatorname{div}[\eta\, D(\alpha)] + \nabla\beta + \delta_\Sigma\,(\mu - v\cdot\alpha)\, n_\Sigma &=& 0 \quad \text{dans} \quad \Omega \\ \operatorname{div}\alpha &=& 0 \quad \text{dans} \quad \Omega \\ \alpha\cdot n &=& 0 \quad \text{sur} \quad \partial\Omega \\ [\eta\, D(\alpha)]\, n\cdot t &=& 0 \quad \text{sur} \quad \partial\Omega \end{array} \right. \end{array} \right. . \tag{8.14}$$

La dérivée !directionnelle de J par rapport à μ dans la direction $\tilde\mu$ prend la forme

$$\left\langle \nabla_\mu J, \tilde\mu \right\rangle = \int_0^L (\psi - \bar\psi)\, \tilde\psi \qquad (\text{où } \bar\psi = \frac{1}{L}\int_0^L \psi\ )\,, \tag{8.15}$$

avec

$$\tilde\psi = -\tilde\alpha\cdot(\rho_2-\rho_1)(g - v\cdot\nabla v) + v_x\,\partial_x\tilde\mu\,, \tag{8.16}$$

où $\tilde\alpha$ est tel que (équations de sensitivité) :

$$\left\{ \begin{array}{rcl} \rho\,[(\nabla v)^T\tilde\alpha - v\cdot\nabla\tilde\alpha] - \operatorname{div}[\eta\, D(\tilde\alpha)] + \nabla\tilde\beta + \delta_\Sigma\,(\tilde\mu - v\cdot\tilde\alpha)\, n_\Sigma &=& 0 \quad \text{dans} \quad \Omega \\ \operatorname{div}\tilde\alpha &=& 0 \quad \text{dans} \quad \Omega \\ \tilde\alpha\cdot n &=& 0 \quad \text{sur} \quad \partial\Omega \\ [\eta\, D(\tilde\alpha)]\, n\cdot t &=& 0 \quad \text{sur} \quad \partial\Omega \end{array} \right. . \tag{8.17}$$

Pour trouver le sous-problème adjoint, on définit la sous-fonction de Lagrange (m_v et m_p sont les sous-multiplicateurs de Lagrange) :

$$L(\alpha,\beta,\mu,m_v,m_p) = J(\alpha,\mu) - \int_\Omega m_v\cdot[\rho((\nabla v)^T\alpha - v\cdot\nabla\alpha)] - \int_\Omega \eta\,\nabla m_v : [D(\alpha)-\beta I_d] \\ -\int_\Omega m_v\cdot\delta_\Sigma(\mu - v\cdot\alpha)\,n_\Sigma + \int_\Omega m_p\,\mathrm{div}\,\alpha\,,$$

et on cherche les multiplicateurs m_v et m_p qui annulent les dérivées directionnelles de L par rapport à α et β. On obtient le sous-problème adjoint :

$$\left\{\begin{array}{rcc} \rho(v\cdot\nabla m_v + m_v\cdot\nabla v) - \mathrm{div}\,[\eta D(m_v)] + \nabla m_p + \delta_\Sigma(v\cdot m_v)n_\Sigma & = & \delta_\Sigma(\psi-\bar\psi)\,\frac{\rho_2-\rho_1}{\sqrt{1+h'^2}}\,(v\cdot\nabla v - g) \\ \mathrm{div}\,m_v & = & 0 \\ m_v\cdot n & = & 0 \\ [\eta\,D(m_v)]\,n\cdot t & = & 0 \end{array}\right. \tag{8.18}$$

La première équation de (8.18) s'écrit sous forme variationnelle :

$$\int_0^L (\psi-\bar\psi)\,[(\rho_2-\rho_1)\,(v\cdot\nabla v - g)]\cdot\tilde\alpha \\ -\int_\Omega\Big[\rho(v\cdot\nabla m_v + m_v\cdot\nabla v) - \mathrm{div}\,[\eta D(m_v)] + \nabla m_p + \delta_\Sigma(v\cdot m_v)n_\Sigma\Big]\cdot\tilde\alpha = 0 \tag{8.19}$$

Alors, d'après (8.15), (8.16) et (8.19), on obtient comme expression pour la dérivée !directionnelle du critère :

$$\Big\langle\nabla_\mu J,\tilde\mu\Big\rangle = \int_\Omega\Big[\rho(v\cdot\nabla m_v + m_v\cdot\nabla v) - \mathrm{div}\,[\eta D(m_v)] + \nabla m_p + \delta_\Sigma(v\cdot m_v)n_\Sigma\Big]\cdot\tilde\alpha \\ +\int_0^L [(\psi-\bar\psi)\,v_x](x,h(x))\,\partial_x\,\tilde\mu(x)\,dx$$

Le premier terme s'exprime en fonction de $\tilde\mu$ en utilisant les équations (8.17) et (8.18), et le deuxième s'intègre par parties ; ainsi

$$\Big\langle\nabla_\mu J,\tilde\mu\Big\rangle = -\int_\Sigma m_v\cdot n_\Sigma\,\tilde\mu + \int_0^L\Big\{(\partial_x\psi\,v_x)(x,h(x)) + (\psi-\bar\psi)(x,h(x))\,\partial_x[v_x(x,h(x))]\Big\}\,\tilde\mu(x)\,dx\,,$$

soit (en utilisant $\mathrm{div}\,v = \mathrm{div}\,\alpha = 0$) :

$$\Big\langle\nabla_\mu J,\tilde\mu\Big\rangle = \int_\Sigma\Big\{\Big[-m_v + [\partial_x\mu\,\partial_z v + (\rho_2-\rho_1)g\,\partial_x\alpha]\,v_x + (\psi-\bar\psi)\,\partial_z v\Big]\cdot n_\Sigma + \frac{h' - v_x\partial_x^2\mu}{\sqrt{1+h'^2}}\Big\}\,\tilde\mu\,.$$

Ainsi, on arrive à l'expression suivante pour le gradient du critère J par rapport à μ :

$$\nabla_\mu J = \Big[-m_v + [\partial_x\mu\,\partial_z v + (\rho_2-\rho_1)g\,\partial_x\alpha]\,v_x + (\psi-\bar\psi)\,\partial_z v\Big]\cdot n_\Sigma + \frac{h' - v_x\partial_x^2\mu}{\sqrt{1+h'^2}}\,.$$

Notations

$\lvert\cdot\rvert$	module (valeur absolue dans $\mathbb{R}$)
$(\cdot,\cdot)$	produit scalaire dans $\mathbb{R}^n$ ($n \in \mathbb{N}$) : $(X,Y) = \sum_{i=1}^{n} X_i\, Y_i$
$\lVert\cdot\rVert$	norme euclidienne dans $\mathbb{R}^n$: $\lVert X\rVert = \sqrt{(X,X)}$
I_n	matrice identité de taille n
I	application identité (dans un espace fonctionnel)
$\overset{\circ}{E}$ (resp. $\bar{E}$)	où E est un ensemble : intérieur (resp. adhérence) de E
d	dimension d'espace
x	variable spatiale
t	variable temporelle
Ω	ouvert de $\mathbb{R}^d$
$\bar{\Omega}$	adhérence de Ω
∂X	avec X un ouvert de $\mathbb{R}^d$: contour orienté de X
$n_{\lvert\partial X}$	normale sortante à ∂X
$t_{\lvert\partial X}$	vecteur tangent à ∂X
$\left.\begin{array}{c}\partial_x^k f \\ \partial^k f/\partial x^k \\ f_x^{(\mathrm{k})}\end{array}\right\}$	dérivée partielle $k^{\text{ème}}$ de f par rapport à x
$\left.\begin{array}{c}d_x^k f \\ d^k f/dx^k\end{array}\right\}$	dérivée totale $k^{\text{ème}}$ de f par rapport à x
$\dot{f}$	$d_t f$
∇	gradient
Δ	laplacien
div	divergence
rot	rotationnel
$\left.\begin{array}{c}\cdot \\ \times \\ \otimes\end{array}\right\}$	produit $\left\{\begin{array}{ll} \text{scalaire} & \\ \text{vectoriel} & \text{dans } \mathbb{R}^d. \\ \text{tensoriel} & \end{array}\right.$
$\mathcal{M}_{n,p}(\mathbb{R})$	ensemble des matrices réelles à n lignes et p colonnes
$\mathcal{M}_n(\mathbb{R})$	$\mathcal{M}_{n,n}(\mathbb{R})$

Lorsque A est une matrice

A^T	transposée de A
$\mathrm{Rg}\, A$	rang de A
$\lVert A\rVert$	avec $A \in \mathcal{M}_n(\mathbb{R})$: $\sup_{X\in\mathbb{R}^n,\, X\neq 0} \dfrac{\lVert AX\rVert}{\lVert X\rVert}$
$\mathrm{Det}\, A$	déterminant de A
$\mathrm{Sp}(A)$	spectre de A (ensemble des valeurs propres)
$\mathrm{SEP}_\lambda(A)$	sous-espace propre de A relatif à la valeur propre λ

Lorsque V et W sont des espaces vectoriels

$\mathcal{L}(V,W)$	ensemble des applications linéaires de V dans W
$\mathcal{L}(V)$	$\mathcal{L}(V,V)$
V'	$\mathcal{L}(V,\mathbb{R})$ = espace dual de V (formes linéaires continues sur V)
$\lvert\cdot\rvert_V$	norme sur V
$(\cdot,\cdot)_V$	produit scalaire sur V
$\langle\cdot,\cdot\rangle_{V',V}$	produit de dualité sur V
$\lVert G\rVert_{\mathcal{L}(V)}$	avec $G\in\mathcal{L}(V)$: $\sup\limits_{\varphi\in V,\,\varphi\neq 0}\dfrac{\lvert G(\varphi)\rvert_V}{\lvert\varphi\rvert_V}$

Espaces fonctionnels

$\mathbb{P}^k(K)$	avec K un fermé borné non vide de $\mathbb{R}^n$: espace des polynômes de degré au plus égal à k sur K
$\mathcal{C}^m(\Omega)$	fonctions sur Ω m fois dérivables et de dérivée $i^{\text{ème}}$ continue $\forall i\leq m$
$\mathcal{C}^0_b(I)$	avec I un intervalle de $\mathbb{R}$: fonctions continues et bornées sur I
$\mathcal{D}(\Omega)$	fonctions $\mathcal{C}^\infty$ à support compact dans Ω
$\mathcal{D}'(\Omega)$	distributions sur Ω
$H^p(\Omega)$	espace de Sobolev $W^{p,2}(\Omega)$. Voir aussi **2.1.2**.
$L^2(\Omega)$	$H^0(\Omega)$
$L^2(0,T;X)$	avec $T\in\mathbb{R}^+$, et X un espace fonctionnel sur Ω normé complet (espace de Banach) : classes de distributions ψ qui à presque tout $t\in[0,T]$ associent $\psi(t)\in X$, telles que $\lvert\psi\rvert_{L^2([0,T],X)}=\displaystyle\int_0^T\lvert\psi(t)\rvert_X^2\,dt<\infty$
$\mathcal{C}^0(0,T;X)$	avec $T\in\mathbb{R}^+$, et X un espace de Banach : fonctions continues à valeurs dans X
$\mathcal{C}^0(0,T;X_w)$	avec $T\in\mathbb{R}^+$, et X un espace de Sobolev : fonctions faiblement continues (au sens du produit scalaire sur X) à valeurs dans X
$H^1_0(\Omega)$	$\{v\in H^1(\Omega)\,\vert\, v_{\vert\partial\Omega}=0\}$
$H^1_n(\Omega)$	$\{v\in H^1(\Omega)^d\,\vert\, v\cdot n_{\vert\partial\Omega}=0\}$
$H^{\text{div}}_n(\Omega)$	$\{v\in L^2(\Omega)^d\,\vert\,\operatorname{div}v\in L^2(\Omega)\,,\ v\cdot n_{\vert\partial\Omega}=0\,,\ \operatorname{div}v=0\}$
$\mathcal{S}(\mathbb{R}^n)$	fonctions $\mathcal{C}^\infty$ telles que $\lvert x\rvert^j\,\partial^\alpha\varphi(x)\underset{\lvert x\rvert\to\infty}{\to}0\,,\ \forall j\in\mathbb{N}\,,\ \forall\alpha\in\mathbb{N}^n\,.$
$\mathcal{S}'(\mathbb{R}^n)$	distributions tempérées

Transformation de Fourier

$\widehat{\psi}(\xi)$	$\displaystyle\int_{\mathbb{R}^n}\psi(x)\,\mathrm{e}^{-2i\pi x\cdot\xi}\,dx\,,$ avec $\psi\in\mathcal{S}'(\mathbb{R}^n)$
$*$	produit de convolution

Bibliographie

[1] Y. Achdou and O. Pironneau. *Computational Methods for Option Pricing*. SIAM, Philadelphia, 2005.

[2] V. Alexéev, V. Tikhomirov, and S. Fomine. *Commande optimale*. Mir, Moscou, 1982.

[3] G. Allaire. *Concepion optimale de structures*. Springer-Verlag, 2007, to appear.

[4] S.N. Antontsev, A.V. Kazhikov, and V.N. Monakhov. *Boundary Value Problems in Mechanics of Nonhomogeneous Fluids*. North-Holland, 1993.

[5] D. Arnold, F. Brezzi, and M. Fortin. A Stable Finite Element for the Stokes Equations. *Calcolo*, 21(4) : 337–344, 1984.

[6] I. Babuška. The Finite Element Method with Lagrange Multipliers. *Numerische Mathematik*, 20 : 179–192, 1973.

[7] R. Bellman. *Dynamic Programming*. Princeton University Press, 1957.

[8] M. Bercovier and O. Pironneau. Error Estimate for Finite Element Solution of the Stokes Problem in Primitive Variables. *Numerische Mathematik*, 33 : 211–224, 1979.

[9] C. Bernardi and Y. Maday. *Approximations spectrales de problèmes aux limites elliptiques*. Springer-Verlag, 1992.

[10] D. Bernardin. *Introduction à la Dynamique des Milieux Continus*. École de Printemps. GDR Matériaux Vitreux, Nancy, 2003. http ://www.lmcp.jussieu.fr/impmc/Associations/GDR-verres/html/Cinema_D_1.pdf.

[11] P. Boily. Application des capteurs thermiques implantés pour la détection du profil de gelée dans la cuve d'électrolyse. Master's thesis, Université du Québec à Chicoutimi, 2001. http ://www.collectionscanada.ca/obj/s4/f2/dsk3/ftp05/MQ65269.pdf.

[12] V. Bojarevics. Nonlinear Waves with Electromagnetic Interactions in Aluminium Electrolysis Cells. In H. Branover and Y. Unger, editors, *Progress in Fluid Flow Research : Turbulence and Applied MHD, volume 182 of AIAA Series ; Progress in Astronautics and Aeronautics*, pages 833–848, 1998.

[13] V. Bojarevics and M.V. Romerio. Long Waves Instability of Liquid Metal-Electrolyte Interface in Aluminium Electrolysis Cells : a Generalization of Sele's criterion. *Eur. J. Mech. B*, 13 : 33–56, 1994.

[14] J. Bonnans, J.-C. Gilbert, C. Lemaréchal, and C. Sagastizabal. *Optimisation numérique*. Springer-Verlag, 1997.

[15] H. Brézis. *Analyse fonctionnelle. Théorie et applications*. Dunod, Paris, 2000.

[16] F. Brezzi. On the Existence, Uniqueness and Approximation of Saddle-Point Problems Arising from Lagrange Multipliers. *RAIRO*, R.2 : 129–151, 1974.

[17] M. Burger. *Infinite-Dimensional Optimization and Optimal Design.* Lectures Notes 285J, Departement of Mathematics, UCLA, 2003. http ://www.math.ucla.edu/˜martinb/.

[18] J.-M. Coron. Local Controllability of a 1-D Tank Containing a Fluid Modeled by the Shallow-Water Equations. *ESAIM Control Optim. Calc. Var.*, 8 : 513–554, 2002. A Tribute to J.-L. Lions.

[19] R. Courant, K.O. Friedrichs, and H. Lewy. Über die partiellen Differenzengleichungen der Matematischen Physik. *Math. Ann.*, 100 : 32–74, 1928.

[20] M.G. Crandall and P.-L. Lions. Viscosity Solutions of Hamilton-Jacobi Equations. *Trans. Amer. Math. Soc.*, 277 : 1–42, 1983.

[21] I. Danaila, F. Hecht, and O. Pironneau. *Simulation numérique en C++.* Dunod, Paris, 2003.

[22] R. Dautray and J.-L. Lions. *Analyse mathématique et calcul numérique pour les sciences et les techniques.* Masson, Paris, 1984.

[23] P.A. Davidson and R.I. Lindsay. Stability of Interfacial Waves in Aluminium Reduction Cells. *Journal of Fluid Mechanics*, 362 : 273–295, 1998.

[24] M.C. Delfour and J.-P. Zolésio. *Shapes and Geometries. Analysis, Differential Calculus, and Optimization.* SIAM Advances in Design and Control, 2001.

[25] J. Descloux, M. Flueck, and M.V. Romerio. Modelling for Instabilities in Hall-Héroult Cells : Mathematical and Numerical Aspects. *Magnetohydrodynamics in Process Metallurgy*, pages 107–110, 1991.

[26] R.J. DiPerna and P.-L. Lions. Ordinary Differential Equations, Transport Theory and Sobolev Spaces. *Invent. Math.*, 98(3) : 511–547, 1989.

[27] J. Dongarra, A. Lumsdaine, R. Pozo, and K.A. Remington. *IML++ Reference Guide*, 1996. http ://math.nist.gov/iml++/.

[28] G. Duvaut and J.-L. Lions. Inéquations en Thermoélasticité et Magnétohydrodynamique. *Arch. of Rat. Mech. Anal.*, 46 : 241–279, 1972.

[29] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements.* Springer-Verlag, 2004.

[30] D. Errate, M.J. Esteban, and Y. Maday. Couplage fluide-structure : un modèle simplifié en dimension 1. *C. R. Acad. Sci. Paris, Série I*, 318 : 275–281, 1994.

[31] P. Faurre. *Analyse numérique. Notes d'optimisation.* École Polytechnique. Ellipses , Paris, 1988.

[32] C.L. Fefferman. *Existence and Smoothness of the Navier-Stokes Equation.* Clay Mathematics Institute, 2000.
http ://www.claymath.org/millennium/Navier-Stokes_Equations/navierstokes.pdf.

[33] M.J. Flynn. Very High-Speed Computing Systems. *Proceedings of the IEEE*, 54 : 1901–1909, 1966.

[34] L. Formaggia, J.-F. Gerbeau, F. Nobile, and A. Quarteroni. Numerical Treatment of Defective Boundary Conditions for the Navier-Stokes Equations. *SIAM Journal on Numerical Analysis*, 40(1) : 376–401, 2002.

[35] L.P. Franca and S.L. Frey. Stabilized Finite Element Methods : II. The Incompressible Navier-Stokes Equations. *Computer Methods in Applied Mechanics and Engineering*, 99 : 209–233, 1992.

[36] A.V. Fursikov. *Optimal Control of Distributed Systems. Theory and Applications*. American Mathematical Society, 2000.

[37] J.-F. Gerbeau. *Problèmes mathématiques et numériques posés par la modélisation de l'électrolyse de l'aluminium*. PhD thesis, École Nationale des Ponts et Chausées, ParisTech, 1998.

[38] J.-F. Gerbeau and C. Le Bris. Existence of Solution for a Density-Dependent Magnetohydrodynamic Equation. *Advances in Differential Equations*, 2(3) : 427–452, 1997.

[39] J.-F. Gerbeau, C. Le Bris, and T. Lelièvre. *Mathematical methods for the Magnetohydrodynamics of liquid metals*. Oxford University Press, 2006.

[40] J.-F. Gerbeau, C. Le Bris, T. Lelièvre, A. Orriols, and T. Tomasino. Linear Versus Nonlinear Approaches for the Stability Analysis of Aluminium Production Cells. In *Proceedings of the European Conference on Computational Fluid Dynamics (ECCOMAS)*, 2006.

[41] V. Girault and P.-A. Raviart. *Finite Element Methods for the Incompressible Navier-Stokes Equations*. Springer-Verlag, 1986.

[42] J.-P. Givry. Computer Calculation of Magnetic Effects in the Bath of Aluminium Cells. *Transactions of the Metallurgical Society of Aime*, 239 : 1161–1166, 1967.

[43] R. Glowinski and J.-L. Lions. Exact and Approximate Controllability for Distributed Parameter Systems (part I). *Acta Numerica*, pages 269–378, 1994.

[44] R. Glowinski and J.-L. Lions. Exact and Approximate Controllability for Distributed Parameter Systems (part II). *Acta Numerica*, pages 159–333, 1996.

[45] M.D. Gunzburger. *Perspectives in Flow Control and Optimization*. SIAM Advances in Design and Control, 2002.

[46] M.D. Gunzburger, A.J. Meir, and J.S. Peterson. On the Existence, Uniqueness, and Finite Element Approximation of Solutions of the Equations of Stationary, Incompressible Magnetohydrodynamics. *Mathematics of Computation*, 56(194) : 523–563, 1991.

[47] É. Guyon, J.-P. Hulin, and L. Petit. *Hydrodynamique physique*. EDP Sciences / CNRS Éditions, Paris, 2001.

[48] F. Hecht. A non-conforming P^1 basis with free divergence in R^3. *RAIRO série analyse numérique*, 1983.

[49] E. Hille and R.S. Phillips. *Functional Analysis and Semi-Groups*. AMS Coll. Pub., 1957.

[50] C.W. Hirt, A.A. Amsden, and J.L. Cook. An Arbitrary Lagrangian Eulerian Computing Method for all Flow Speed. *Journal of Computational Physics*, 14(3) : 227–253, 1974.

[51] P. Hood and G. Taylor. *Navier-Stokes Equations Using Mixed Interpolation*. Oden ed., UAH Press, 1974.

[52] L.S. Hou and A.J. Meir. Boundary Optimal Control of MHD Flows. *Appl. Math. Optim.*, 32 : 143–162, 1995.

[53] T.J.R. Hughes, L.P. Franca, and M. Balestra. A New Finite Element Formulation for Computational Fluid Dynamics : V. Circumventing the Babuška-Brezzi Condition : a Stable Petrov-Galerkin Formulation of the Stokes Problem Accomodating Equal-Order Interpolations. *Computer Methods in Applied Mechanics and Engineering*, 59 : 85–99, 1986.

[54] R.E. Kalman. On the General Theory of Control Systems. In *Proc. 1st IFAC Congress, Moscow, 1960*. Butterworth, London, vol. 1 pages 481-492, 1961.

[55] G. Karypis and V. Kumar. *METIS : A Software Package for Partitioning Unstructured Graphs, Partitioning Meshes, and Computing Fill-Reducing Orderings of Sparse Matrices*, 1998. http ://www-users.cs.umn.edu/˜karypis/metis/metis/files/manual.pdf.

[56] H.W. Kuhn and A.W. Tucker. Nonlinear Programming. In *Proceedings of Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492, 1961.

[57] B.A. Kupershmidt. *The Variational Principles of Dynamics*. Advanced Series in Mathematical Physics, World Scientific, 1992.

[58] L. Landau and E. Lifchitz. *Électrodynamique des milieux continus*. Mir, Moscou, 1990.

[59] B.E. Launder and D.B. Spalding. *Mathematical Models of Turbulence*. Academic Press, 1972.

[60] P. Lax and N. Milgram. Parabolic equations. *Contributions to the Theory of Partial Differential Equations*, 1954.

[61] J. Leray. Étude de diverses équations intégrales non linéaires et de quelques problèmes que pose l'hydrodynamique. *J. Math. Pures Appl.*, 12 : 1–82, 1933.

[62] X. Li and J. Yong. *Optimal Control Theory for Infinite Dimensional Systems*. Birkhäuser, Boston, 1995.

[63] J.-L. Lions. *Quelques méthodes de résolution des problèmes aux limites non linéaires*. Dunod et Gauthier-Villars, Paris, 1969.

[64] J.-L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations*. Springer-Verlag, 1971.

[65] J.-L. Lions. *Exact Controllability, Stabilization and Perturbations for Distributed Systems*. Springer-Verlag, 1994.

[66] J.-L. Lions and E. Magenes. *Non-Homogeneous Boundary Value Problems and Applications*. Springer-Verlag, 1971.

[67] P.-L. Lions. *Mathematical Topics in Fluid Mechanics*. Oxford University Press, 1996.

[68] D.G. Luenberger. *Linear and Nonlinear Programming*. Springer-Verlag, 2nd edition, 2003.

[69] A. Lumsdaine, R. Pozo, and K.A. Remington. *SparseLib++ Reference Guide*, 1996. http ://math.nist.gov/sparselib++/.

[70] J.A. Meijerink and H.A van der Vorst. Guidelines for the Usage of Incomplete Decomposition in Solving Sets of Linear Equations as They Occur in Practical Problems. *Journal of Computational Physics*, 44(10) : 134–155, 1981.

[71] Message Passing Interface Forum. *MPI : A Message-Passing Interface Standard (version 1.1)*, Technical Report, 1995. http ://www.netlib.org/mpi.

[72] B. Mohammadi and O. Pironneau. *Applied Shape Optimization for Fluids*. Oxford University Press, 2001.

[73] R. Moreau. *Magnetohydrodynamics*. Kluwer Academic Publishers, 1990.

[74] R. Moreau and J.W. Evans. An Analysis of the Hydrodynamics of Aluminium in Reduction Cells. *J. Electrochem. Soc. : Electrochem. Sci. Tech.*, 131(10) : 2251–2259, 1984.

[75] R. Moreau and D. Ziegler. Stability of Aluminium Cells : a New Approach. *Light Metals*, pages 359–364, 1986.

[76] D. Munger. Simulation numérique des instabilités magnétohydrodynamiques dans les cuves de production d'aluminium. Master's thesis, Université de Montréal, 2004. http ://mhd.selfip.info/com/memoire-rectoverso.pdf.

[77] J.-C. Nédélec. A New Family of Mixed Finite Elements in R^3. *Numer. Math.*, 50 : 57–81, 1986.

[78] N.M. Newmark. A Method of Computation for Structural Dynamics. In *Proceedings of ASME Conference*. ASME, 1959.

[79] A. Orriols. Algorithmes de contrôle d'interface libre. Implémentation sur un modèle MHD linéaire. *Congrès National d'Analyse Numérique (CANUM)*, 2006.

[80] A. Orriols, J.-F. Gerbeau, C. Le Bris, and T. Lelièvre. Simulations numériques sous Mistral & Problèmes de contrôle d'interface libre. Technical report, Alcan Primary Metal Group, 2005.

[81] O. Pironneau. *Finite Element Methods for Fluids*. Wiley, 1989.

[82] L.S. Pontryagin, V.G. Boltyanski, R.V Gramkrelidze, and E.F. Mischenko. *Mathematical Theory of Optimal Processes*. Wiley, 1962.

[83] A. Quarteroni and A. Valli. *Numerical Approximation of Partial Differential Equations*. Springer-Verlag, 1997.

[84] P.-A. Raviart and J.-M. Thomas. *Introduction à l'analyse numérique des équations aux dérivées partielles*. Dunod, Paris, 1998.

[85] Y. Saad. *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company, Boston, 1996. http ://www-users.cs.umn.edu/˜saad/books.html.

[86] Y. Saad and M.H. Schultz. GMRES : A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems. *SIAM J. Sci. Statist. Comput.*, 7(3) : 856–869, 1986.

[87] Y. Safa. *Simulation numérique des phénomènes thermiques et magnétohydrodynaiques dans une cellule de Hall-Héroult*. PhD thesis, École polytechnique fédérale de Lausanne, 2005.

[88] J. Salençon. *Handbook of Continuum Mechanics*. Springer-Verlag, 2001.

[89] Scilab. http ://www.scilab.org.

[90] M. Segatz and C. Droste. Analysis of Magnetohydrodynamic Instabilities in Aluminium Reduction Cells. *Light Metals*, pages 313–322, 1994.

[91] T. Sele. Instabilities of the Metal Surface in Electrolytic Cells. *Light Metals*, pages 7–24, 1977.

[92] M. Sermange and R. Temam. Some Mathematical Questions Related to the MHD Equations. *Communications in Pure and Applied Mathematics*, XXXVI : 635–664, 1983.

[93] A.D. Sneyd. Stability of Fluid Layers Carrying a Normal Electric Current. *Journal of Fluid Mechanics*, 156 : 223–236, 1985.

[94] A.D. Sneyd and A. Wang. Interfacial Instability Due to MHD Mode Coupling in Aluminium Reduction Cells. *Journal of Fluid Mechanics*, 263 : 343–359, 1994.

[95] R. Temam. *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*. Springer-Verlag, 1988.

[96] E. Trélat. *Contrôle optimal. Théorie et applications*. Vuibert, Paris, 2005.

[97] R.S. Tuminaro, M. Heroux, S.A. Hutchinson, and J.N. Shadid. *Official Aztec User's Guide*, 1999. http ://www.cs.sandia.gov/CRF/Aztec_pubs.html.

[98] N. Urata. Magnetics and Metal Pad Instabilities. *Light Metals*, pages 581–591, 1985.

[99] K. Yosida. *Functional Analysis*. Springer-Verlag, 1965.

[100] O. Zikanov, H.A. Sun, and D.P. Ziegler. Shallow Water Model of flows in Hall-Héroult Cells. *Light Metals*, pages 335–340, 2004.

Index