Sur la résolution efficace d'équations aux dérivées partielles en mécanique des fluides multiphasique et imagerie médicale

Louis Le Tarnec

Thèse de doctorat

pour obtenir le grade de

Docteur de l'école normale supérieure de Cachan

Spécialité : Mathématiques

présentée par

Louis Le Tarnec

Sur la résolution efficace d'équations aux dérivées partielles en mécanique des fluides multiphasique et imagerie médicale

Directeur de thèse : **Jean-Michel Ghidaglia**

Rapporteurs :
Julie Delon, Université Paris Descartes
Fayssal Benkhaldoun, Université Paris 13

Examinateurs :
Florian De Vuyst, ENS Cachan
Daniel Bouche, ENS Cachan
Sylvain Faure, Laboratoire de Mathématiques d'Orsay

Septembre 2014

Remerciements

Cette thèse repose sur trois années de travail, dont une partie a été effectuée au sein du centre hospitalier de l'université de Montréal, en appui à l'équipe RUBIC qui met au point de nouvelles techniques d'imagerie médicale, l'autre chez Eurobios à Gentilly, dans le cadre de contrats industriels. J'ai quitté ces lieux - ainsi que le monde de la recherche académique - à l'heure où je la rédige, mais toujours conservé la volonté d'achever ce travail, car il me tenait à cœur de valoriser ces années de recherche qui ont été d'une grande importance dans mon évolution personnelle. Mes remerciements vont principalement au Pr. Jean-Michel Ghidaglia, qui m'a encouragé à poursuivre cet effort jusqu'à son terme, et qui me permet de soutenir cette thèse à l'ENS Cachan. Je le remercie également pour le riche apprentissage professionnel qu'il m'a transmis pendant mon année chez Eurobios, dont je conserverai longtemps la trace, ainsi que pour la confiance qu'il m'a accordée et pour sa grande bienveillance à mon égard, quels qu'aient été mes choix d'orientation. Je remercie avec lui mes anciens collaborateurs au sein d'Eurobios, en particulier Sylvain Faure et Christophe Labourdette, dont les travaux préalables à mon arrivée constituent en grande partie le support de cette thèse, ainsi que Joris Costes à qui j'adresse toute mon amitié.

Je remercie vivement le Pr. Damien Garcia du centre hospitalier de l'université de Montréal, qui m'a également accordé une grande confiance et m'a formé au difficile travail de recherche, ainsi que le Dr. François Destrempes, dont la rigueur et la persévérance sans équivalents m'ont permis de rédiger et faire publier mon premier article. Je salue également l'équipe RUBIC, avec laquelle j'ai passé une très belle - bien que parfois très froide - année canadienne.

Je tiens également à remercier les Pr. Julie Delon et Fayssal Benkhaldoun, qui ont accepté de rapporter cette thèse, ainsi que Florian De Vuyst, Daniel Bouche et Sylvain Faure, membres du jury.

Résumé Ce travail s'articule en quatre parties. Les trois premières ont pour socle commun l'adaptation d'un schéma volumes finis ($VFFC$) à des situations variées, dans le but d'un gain en efficacité pour la simulation numérique d'écoulements complexes. La première partie concerne la simulation numérique efficace de la chute d'un bloc de liquide au sein d'une poche de gaz, et propose un nouveau modèle mutualisant de précédents travaux pour associer finesse des résultats et efficacité de calcul. La deuxième partie vise à la mise en place d'un schéma AMR (Adaptive Mesh Refinement) général pour la résolution par volumes finis des systèmes non conservatifs. La troisième partie a pour finalité le couplage dynamique de deux modèles représentant plus ou moins finement une physique donnée. Enfin, dans un tout autre domaine où l'efficacité de résolution des équations aux dérivées partielles revêt également une grande importance, la quatrième partie s'intéresse au problème du flot optique en imagerie - c'est à dire à la recherche d'un champ de déplacement à partir d'une suite d'image - et approfondit une méthode existante (méthode de Horn et Schunck) d'un point de vue pratique et théorique.

Abstract This work is divided into four parts. The first three parts, as a common base, aim at adapting a finite volume scheme ($VFFC$) to various situations, in order to get a better efficiency for the numerical simulation of complex flows. The first part deals with the efficient numerical simulation of a falling block of liquid in a gas pocket, and proposes a new model to combine previous work for associating precision of results and computational efficiency. The second part aims at the establishment of a general AMR (Adaptive Mesh Refinement) scheme for resolution by finite volumes of non-conservative systems. The purpose of the third part is the dynamic coupling of two models representing more or less finely a given physical system. Finally, in any other area where the efficiency of solving partial differential equations is of great importance too, the fourth part deals with the problem of optical flow in imaging - i.e. the research of a displacement field from several successive images - and deepens an existing method (Horn and Schunck method) from a practical and theoretical perspective.

Table des matières

I Mécanique des fluides numérique **9**

1 Flux de Pan **10**
1.1 Introduction . . . 10
1.2 Problème posé . . . 12
1.2.1 Problème physique . . . 12
1.2.2 Représentation du problème . . . 13
1.3 Base théorique . . . 14
1.3.1 Méthode VFFC : Volumes Finis à Flux Caractéristiques . . . 14
1.3.2 Cas du gaz seul . . . 17
1.3.3 Cas d'une tranche de liquide seule . . . 22
1.4 Méthode mise au point . . . 27
1.4.1 Présentation générale . . . 27
1.4.2 Résultats numériques . . . 31
1.5 Conclusion et perspectives . . . 40

2 AMR **42**
2.1 Introduction . . . 42
2.2 La fonction de contrôle et la notion d'équirépartition . . . 43
2.2.1 Fonction de contrôle . . . 43
2.2.2 Equirépartition . . . 43
2.3 Lissage de la fonction de contrôle . . . 43
2.3.1 Principe . . . 43
2.3.2 Choix de la fonction de lissage : . . . 44
2.3.3 Exemple . . . 45
2.4 Calcul direct du maillage équidistribué . . . 46
2.4.1 Méthode simple . . . 46
2.4.2 Algorithme de De Boor . . . 47
2.4.3 Limites du calcul direct . . . 48
2.5 Condition CFL avec maillage mobile . . . 48
2.6 Equations différentielles de maillage . . . 49
2.6.1 Théorie . . . 49
2.6.2 Mise en œuvre . . . 50
2.7 Méthode VFFC pour les modèles non conservatifs . . . 50
2.8 Résolution du système physique avec AMR . . . 52
2.9 Calcul des matrices . . . 53
2.9.1 Méthode de calcul . . . 53
2.9.2 Invariance galiléenne . . . 55
2.10 Modèles considérés . . . 55

2.10.1 Une équation . . . 56
2.10.2 Trois équations . . . 56
2.10.3 Quatre équations . . . 57
2.11 Comparaison des méthodes de remaillage . . . 60
2.12 Cas test de Sod . . . 61
2.13 Cas test de Ransom . . . 66
2.14 Conclusion et perspectives . . . 69

3 Couplage **71**
3.1 Introduction . . . 71
3.2 Modèle à 7 équations . . . 72
3.2.1 Description du modèle . . . 72
3.2.2 Résolution du modèle . . . 74
3.2.3 Invariance galiléenne et AMR . . . 76
3.3 Modèle homogène à 5 équations . . . 78
3.3.1 Description du modèle . . . 79
3.3.2 Résolution dans le cas $u_r = 0$ . . . 80
3.4 Validation et test du modèle homogène . . . 82
3.4.1 Cas test de Sod . . . 82
3.4.2 Comparaison modèle homogène / modèle à 7 équations . . . 83
3.5 Couplage modèle homogène / modèle à 7 équations . . . 89
3.5.1 Couplage à frontière fixe . . . 89
3.5.2 Couplage à frontière mobile . . . 91
3.5.3 Cas test . . . 93
3.6 Conclusion . . . 99

II Imagerie médicale **100**

4 Méthode de Horn Schuck pour le flot optique **101**
4.1 Introduction . . . 101
4.2 Analyse de la méthode Horn Schunck . . . 102
4.2.1 Description de la méthode . . . 102
4.2.2 Étude théorique de la convergence . . . 105
4.3 Horn et Schunck en échographie cardiaque . . . 106
4.3.1 Adaptation de la méthode Horn Schunck aux coordonnées polaires 107
4.3.2 Test de l'algorithme . . . 109
4.3.3 Cas réels . . . 115
4.4 Conclusion . . . 118

III Annexe : convergence des itérations de Horn Schuck **124**

Introduction générale

La première partie du travail présenté dans ce document a été réalisée au sein d'une entreprise spécialisée en informatique, sciences de la complexité et simulation numérique (*Eurobios, Gentilly, France*). Elle se positionne en appui aux industriels confrontés à des problèmes dépassant le cadre de leur $R\&D$ traditionnelle, et désireux de transmettre leurs difficultés à des chercheurs spécialistes du domaine. Cette entreprise s'aide d'un vaste réseau d'experts et fait le lien entre leurs travaux universitaires et ces problématiques industrielles. Deux sujets ont été plus particulièrement à l'origine des éléments présentés ici.

Tout d'abord, dans le cadre d'un partenariat avec une société innovante conceptrice de membranes confinantes pour le transport du méthane liquide (ou GNL), nous nous sommes intéressés à la simulation numérique efficace des impacts de vague sur une paroi, ainsi qu'à la compréhension phénoménologique des phénomènes physiques en jeu. Bien que plusieurs modèles simples $0D$ ou $1D$ représentant ce phénomène aient été à notre disposition avant ce travail [5, 10, 11], aucun d'entre eux ne prenait en compte la variation longitudinale des propriétés du bloc de liquide représentant la vague. Nous présenterons ici un modèle que nous qualifions toujours de *modèle simple* en raison de son efficacité informatique et du nombre modéré de lignes de code, mais qui permet d'introduire cette variation et, en conséquence, d'étudier la déformation longitudinale du bloc de liquide. Nous baptiserons ce modèle *Flux de Pan*, pour des raisons qui devraient s'éclaircir à la lecture du premier chapitre, dont il fait l'objet. Nous accompagnerons sa présentation d'une description théorique du schéma VFFC, ou *Volumes Finis à Flux Caractéristiques* [1], qui sert de socle aux trois premiers chapitres.

Le deuxième sujet qui nous intéresse concerne la simulation numérique fine des écoulements multiphasiques, dont l'application visée ici est la modélisation d'une explosion au sein d'un espace confinant artificiel constitué de mousse - la mousse n'étant rien d'autre, en bonne approximation, qu'un mélange d'eau liquide, d'air et de vapeur d'eau -. Ce travail devrait mener à la conception d'un outil industriel et robuste permettant la prédiction des pressions engendrées par une onde de détonation se propageant dans un tel milieu. Ce sujet intéresse les industries militaires et civiles cherchant à mieux connaître l'intérêt de l'utilisation des mousses comme protection face à des explosions. Ce travail traitant une physique complexe et ayant comme finalité la création d'un outil efficace du point de vue du temps de calcul, un travail théorique de modélisation et optimisation a dû être mené sur de nombreux aspects du problème, dont nous évoquons ici deux points qui nous ont principalement occupés.

- La mise en place de l'AMR, c'est à dire du raffinement automatique de maillage,

est une étape qui pourrait assurer un gain significatif en terme, au choix, de précision ou de temps de calcul. Si l'implémentation de cette méthode ne pose pas de problème particulier pour les modèles rigoureusement conservatifs, son adaptation à des modèles *presque* conservatifs, comme nous constaterons être le cas pour les milieux multiphasiques, est en revanche difficile et peu documentée. Nous aborderons ici les aspects théoriques de la mise en place de l'AMR pour les modèles conservatifs à partir de la méthode VFFC, proposerons une méthodologie d'adaptation aux modèles non conservatifs, et la validerons sur un cas test académique. Ce travail fait l'objet du deuxième chapitre. Nous l'accompagnerons d'une description théorique de l'adaptation du schéma VFFC aux modèles non conservatifs [6].

- Par ailleurs, toujours dans le but de rendre plus précise à temps de calcul donné, ou plus rapide à précision donnée, la simulation numérique des ondes de détonation dans les mousses, nous avons implémenté selon la méthode VFFC un modèle pré-existant [8, 9], permettant de prédire l'évolution d'un système diphasique à trois espèces de façon plus simpliste du point de vue de la physique mais plus efficace du point de vue du temps de calcul que le modèle standard. Il s'agit de ne considérer qu'une seule température et qu'une seule vitesse pour les trois espèces, approximation qui peut être grossièrement fausse au cœur d'une explosion, mais qui peut aussi s'avérer excellente à une distance suffisante de la charge. Une fois ce modèle implémenté, nous avons entrepris d'opérer un couplage dynamique entre le modèle complet et le modèle simplifié, c'est à dire un couplage dont la frontière varie continûment en fonction de l'évolution physique du système. Une présentation théorique de la méthode mise au point dans ce but, ainsi que des tests concluant sur son efficacité, sont présentés dans le troisième chapitre. Il sera aussi l'occasion de présenter en détails le modèle diphasique à trois espèces standard.

La deuxième partie du travail présenté dans ce document a été réalisée au sein du *CRCHUM* (*Centre de Recherche du Centre Hospitalier de l'Université de Montréal*), dans le domaine du traitement d'images médicales. Plus précisément, le projet concernait la détermination du flot optique, c'est-à-dire l'obtention d'un champ de déplacements à partir de deux images successives, qui est un problème difficile car mal posé. Bien que ce thème puisse sembler éloigné des problématiques présentées dans la première partie, nous allons voir que les méthodes employées ont de nombreux points communs. En effet, l'algorithme de détermination du flot optique étudié ici est basé sur la résolution d'une équation aux dérivées partielles par différences finies. Il s'agit de la méthode de Horn et Schunck [16], qui sera décrite en détails dans le quatrième et dernier chapitre. Nous verrons que cette méthode est basée sur la résolution itérative d'un système linéaire. Aucune preuve de convergence n'était proposée par les auteurs de l'article original datant de 1981, qui se contentaient d'un constat empirique, mais plus tard deux travaux de recherche indépendants ont abouti en 2004 et 2008 à la proposition de deux preuves théoriques [20, 21]. Concernant ce problème de convergence, le lecteur sera renvoyé à l'article en annexe mettant en défaut ces deux preuves et proposant une nouvelle preuve correcte et très générale.

Nous nous appliquerons enfin, toujours dans le quatrième chapitre, à appliquer l'algorithme de Horn et Schunck à des cas rencontrés en échographie cardiaque, et à tester

son efficacité. Un travail préalable d'adaptation de l'algorithme aux coordonnées polaires devra être réalisé, car dans ce domaine les données sont souvent obtenues dans un tel système. En effet, pour balayer un large domaine sans déplacer la sonde, on peut faire varier l'angle d'émission des ultrasons plutôt que de leur imposer une translation. En particulier, il était important que l'adaptation en coordonnées polaires de l'algorithme permette l'utilisation d'une méthode itérative similaire à celle employée dans le cas standard et faisant l'objet de l'article en annexe.

Première partie

Mécanique des fluides numérique

Chapitre 1

Flux de Pan

1.1 Introduction

Dans cette partie, nous nous intéressons au problème de la chute d'un bloc d'eau entouré de gaz dans un domaine fermé, et en particulier aux phénomènes physiques résultants de l'impact de ce bloc d'eau sur le bord du domaine, considéré comme un mur. Les applications principalement visées concernent l'étude des mouvements de méthane liquide (GNL) au sein du méthane gazeux dans les méthaniers. Plus précisément, nous nous intéressons à une technologie récente et innovante qui, au lieu de stocker le GNL dans des conteneurs métalliques de petite taille comme cela est fait classiquement, repose sur un entreposage direct dans une cuve parallélépipédique de plusieurs milliers de m^3. L'intérêt en terme d'optimisation de l'espace utilisé est évident, mais le risque de montée en pression par évaporation du méthane liquide et de rupture de la cuve doit être contrôlé. Le maintient à l'état liquide du méthane a pour condition un environnement à température très basse, de l'ordre de -160^oC. Les différentes membranes entourant la cuve ont donc un rôle essentiel d'isolement thermique à jouer. En conséquence, leur tenue mécanique aux perturbations rencontrées lors du transport, et la démonstration du maintient de leur intégrité, sont primordiales. C'est dans ce contexte que nous nous intéressons au phénomène d'impact sur une paroi d'une vague entourée de gaz, phénomène qui se produit continûment durant le transport en raison du ballottement et de la nécessaire évaporation d'une partie du méthane à proximité de cette paroi. Le dessin suivant présente un exemple de membrane :

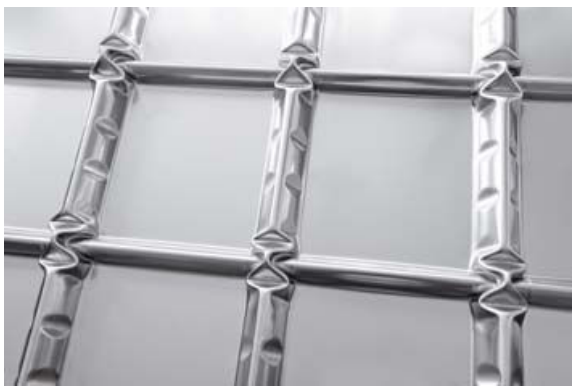


FIGURE 1.1 – Membrane pour cuve de méthanier

Le problème physique de l'impact d'une vague sur un mur a fait l'objet de nombreux travaux expérimentaux [3], travaux que nous cherchons à mieux représenter d'un point de vue numérique. Ces études expérimentales ont permis de mettre en évidence trois régimes apparaissant successivement lors de l'impact [4] :

- Tout d'abord, le bloc de liquide se dirige vers la paroi, entraîné par une vitesse initiale ainsi que par une accélération. Cette dernière peut être aussi bien due à la gravité qu'au ballottement du bateau, à l'origine d'un mouvement du référentiel d'étude. À ce stade, le liquide peut être vu comme un bloc incompressible et l'unique force qui le ralentit résulte de la compression du gaz qui le sépare de la paroi, éventuellement accompagnée de la détente du gaz situé de l'autre côté du bloc.
- Peu à peu, alors que la pression du gaz augmente, le rendant plus difficilement déformable, on entre dans une phase où la compressibilité du gaz et du liquide jouent tous deux un rôle similaire dans le ralentissement du bloc de liquide.
- Si l'impact est suffisamment violent, la pression du gaz devient telle qu'à son tour il peut être considéré comme incompressible, et c'est alors la compressibilité du liquide qui joue un rôle primordial pour absorber l'énergie de l'impact.

Une difficulté importante dans l'analyse de ces phénomènes est due à l'impossibilité de produire des essais à l'échelle 1, et de les réaliser avec du méthane, ce dernier point étant dû aux conditions extrêmes qui seules peuvent permettre son maintient à l'état liquide. Les essais sont donc réalisés à échelle réduite, en général 1 : 40, en assurant un maintient du nombre de Froude. Mais ce *Froude scaling* ne saurait assurer à lui seul l'exacte similitude entre les essais sur modèle réduit et la réalité. C'est pourquoi la simulation numérique joue également un rôle clé dans la compréhension des phénomènes physiques. Nous présenterons dans la suite un modèle $0D$ très simplifié permettant de représenter le phénomène en s'épargnant une simulation $2D$ complète, appelé modèle de *Bagnold*. Ces dernières années, différents travaux ont permis de complexifier peu à peu ce dernier modèle en prenant en compte davantage de phénomènes, comme la compressibilité du liquide ou la possible d'une fuite de gaz sur les côtés du bloc, grâce à des approches $0D$ avancées ou $1D$ [10, 11, 23]. Dans ce document, nous proposons une nouvelle extension de ces modèles, qui prend en compte une possible déformation longitudinale du bloc d'eau. L'intérêt peut être d'observer quelle quantité de gaz se trouve *prisonnière* du bloc de liquide au moment de l'impact, cette quantité pouvant influencer sensiblement les pressions maximales à la paroi résultantes de l'impact. Le modèle présenté ici, que nous avons appelé *Flux de Pan*, a l'avantage d'être très efficace du point de vue du temps de calcul et de la mémoire utilisée, et permet notamment de réaliser des simulations en quelques minutes. En revanche, il ne permet de traiter que les instants antérieurs au contact du liquide et de la paroi - si toutefois ce contact a lieu, car nous verrons des situations dans lesquelles la totalité du liquide remonte sans avoir touché la paroi -.

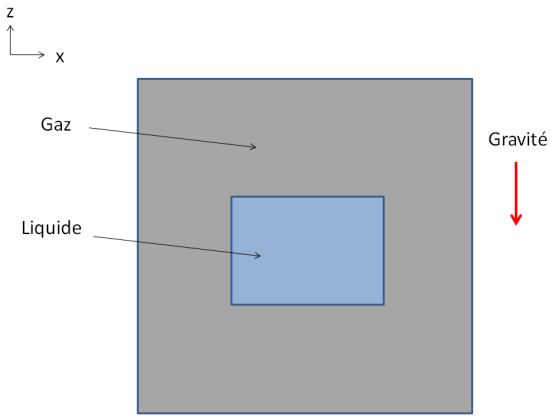
Par ailleurs ce modèle, dans sa version simplifiée car sans prise en compte de la déformation longitudinale du bloc de liquide, a permis de produire, parmi les courbes de l'article [4] faisant le point sur l'apport de la simulation numérique dans la problématique de mise à l'échelle des essais expérimentaux, celles en provenance de *ENS Cachan / Eurobios*. Ce dernier article présente en détails l'approche phénoménologique de l'impact des vagues, les difficultés liées à cette mise à l'échelle, les résultats numériques

obtenus et leur utilisation dans ce contexte. Nous nous focaliserons donc ici sur l'aspect théorique du modèle mis au point et sur quelques cas d'application.

1.2 Problème posé

1.2.1 Problème physique

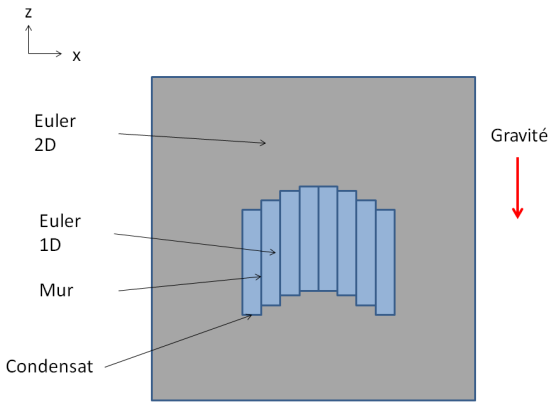
Un bloc de liquide se déplace dans le gaz, son mouvement étant dû à la gravité ainsi qu'à sa vitesse initiale. Nous souhaitons résoudre le problème - à savoir connaître l'évolution temporelle du contour du bloc de liquide et de l'état des fluides en tout point - en nous imposant comme condition d'interdire tout mélange entre le gaz et le liquide ainsi que tout mouvement horizontal de liquide.



Nous nous inspirerons de travaux déjà réalisés pour traiter cette question. Un modèle 2D complet est présenté dans [12], dont la version 1D est reprise dans la sous-section 1.3.3. Toutefois ce modèle est trop complexe et trop coûteux en temps de calcul pour nos applications. Un modèle beaucoup plus simple est proposé dans [10], pour lequel les équations d'Euler ne sont résolues que dans le gaz, et qui considère le liquide comme un bloc indéformable. Des développements de ce dernier modèle sont réalisés dans [11], et consistent par exemple à remplacer le bloc de liquide par un ressort ou à le traiter avec les équations d'Euler 1D. La limite de ces derniers modèles concerne leur incapacité à prévoir la déformation du bloc de liquide.

L'approche développée ici propose, comme compromis à ces méthodes, de traiter le problème sans recourir à de gros moyens numériques mais en prenant en compte une déformation du bloc de liquide. L'idée générale est :

- de découper le bloc de liquide en tranches verticales dont les parois sont considérées comme des murs et dont les largeurs sont fixes,
- de traiter le gaz grâce aux équations d'Euler 2D,
- de traiter l'intérieur de chaque tranche de liquide grâce aux équations d'Euler 1D,
- de traiter les interfaces horizontales entre liquide et gaz grâce à une méthode particulière, inspirée de [12] et présentée dans la sous-section 1.3.3. Cette méthode introduit notamment la notion de *condensat*, qui constitue une discrétisation spécifique pour la zone proche de l'interface.



1.2.2 Représentation du problème

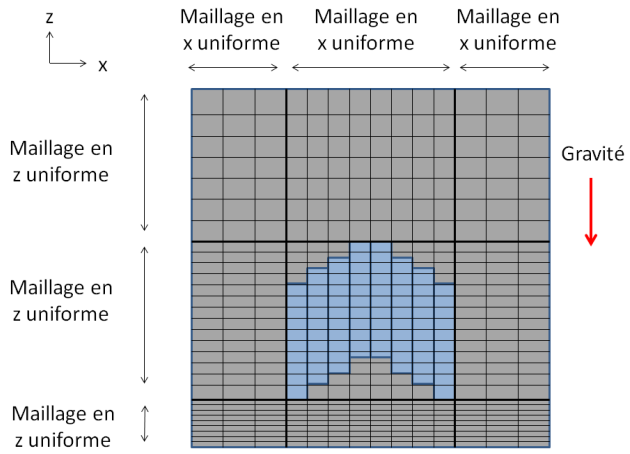
La discrétisation spatiale est réalisée sur une grille cartésienne. Dans un souci de limitation des moyens de calcul nécessaires, les cas standards se traitent avec quelques centaines de cellules, qui doivent suffire à représenter correctement la physique en jeu. Dans ce but, le maillage est mobile et réadapté en continu à la position du bloc de liquide. Plus spécifiquement, il présente les caractéristiques suivantes.

Maillage vertical

Pour décrire correctement la physique en jeu dans les applications visées - notamment pour traiter correctement d'éventuels mouvements très violents du bloc de liquide -, nous choisissons de mettre en place un maillage vertical mobile, qui conservera néanmoins un nombre de mailles fixes pour une question de simplicité. Lorsque le bloc tombe, nous souhaitons maintenir un nombre de mailles fixe dans la zone de gaz inférieure (i.e. entre l'extrémité inférieure du domaine et celle du bloc de liquide), qui pourra être amenée à devenir très petite en fonction des applications. Cet accroissement de la précision dans la zone inférieure ne doit pas se payer par une perte de précision dans la zone médiane (i.e. la zone contenant le bloc de liquide), nous maintenons donc un nombre de mailles fixe sur la hauteur du bloc de liquide. Par conséquent, la zone de gaz supérieure (i.e. au dessus du bloc de liquide) conserve également un nombre de mailles fixe. Cette zone n'étant pas une zone d'intérêt particulier lors de la chute du bloc, la perte de précision qui en résulte n'a pas beaucoup d'importance. Si le bloc rebondit et se dirige vers le haut, les mêmes règles seront appliquées, de sorte que le maillage reprend progressivement sa forme initiale.

Maillage horizontal

Il ne serait pas utile, en revanche, de rendre le maillage horizontal mobile, car les tranches de liquide sont par hypothèse de largeur constante dans le temps. Nous faisons également l'hypothèse que toutes les tranches ont la même largeur, et nous affectons une unique cellule horizontale à chaque tranche pour une question de simplicité. Nous considérons par ailleurs un maillage uniforme de part et d'autre du bloc de liquide, dont le nombre de cellules est choisi par l'utilisateur.



1.3 Base théorique

1.3.1 Méthode VFFC : Volumes Finis à Flux Caractéristiques

Cette méthode est présentée dans [1]. Il s'agit d'un schéma volumes finis à variables centrées, utilisé en premier lieu pour résoudre les systèmes d'équations conservatifs en une ou plusieurs dimensions, c'est à dire ceux qui peuvent s'écrire sous la forme suivante :

$$\partial_t\ v + div\big(F(v)\big) = 0, \tag{1.1}$$

où v est le vecteur des variables conservatives et $F(v)$ le vecteur des flux, unique pour v donné. Nous avons omis la présence d'un éventuel terme source afin d'alléger les notations.

Volumes finis.

Nous souhaitons résoudre le système (1.1) sur un domaine Ω que nous supposons décomposé en cellules polyédriques. L'intégration du système sur une cellule d'indice i, recouvrant le domaine Ω_i d'enveloppe T_i, donne :

$$\int_{\Omega_i} \Big(\partial_t\ v + div\big(F(v)\big)\Big) d\tau = \frac{d}{dt}\int_{\Omega_i} v\ d\tau + \int_{T_i} F.n\ ds.$$

Considérons maintenant que l'enveloppe T_i est constituée de m_i plans notés A_{ij} , $(1 \leq j \leq m_i)$, de sorte que A_{ij} sépare les cellules i et j (nous ne considérons pas les cellules frontalières dans ce résumé, se référer à [1] pour plus de précisions). Notons n et $n+1$ deux itérations séparées par le pas de temps Δt, et v_i la valeur moyenne de v sur Ω_i à l'itération n, soit :

$$v_i = \frac{1}{|\Omega_i|}\int_{\Omega_i} v\ d\tau,$$

alors la discrétisation temporelle de l'équation (1.1) peut s'écrire :

$$|\Omega_i|\frac{v_i^{n+1} - v_i^n}{\Delta t} + \sum_{j=1}^{m_i} |A_{ij}|\phi_{ij} = 0 \tag{1.2}$$

où

$$\phi_{ij} = \frac{1}{|A_{ij}|} \int_{A_{ij}} F(v).n_{ij}\ ds, \tag{1.3}$$

avec n_{ij} la normale au plan A_{ij} sortante de la cellule Ω_i. Pour que le schéma soit conservatif, il faut que l'on ait $\phi_{ij} = \phi_{ji}$ pour toutes les cellules i et j du domaine Ω. La méthode de calcul de ces flux est significative pour la qualité des résultats, et fait la spécificité de la méthode VFFC.

Flux caractéristiques

Considérons deux cellules i et j séparées par une face A_{ij} de normale n_{ij} allant de i vers j. Nous souhaitons calculer la valeur de ϕ_{ij} à partir de la donnée de v_i et v_j, valeurs de v moyennées sur ces deux cellules, afin d'être en mesure de résoudre l'équation discrétisée (1.2). Notons d la dimension du domaine Ω, et commençons par réécrire le système (1.1) en introduisant e_i , $(1 \leq i \leq d)$, une base canonique de $\mathbb{R}^d$:

$$\partial_t v + \sum_{i=1}^{d} \partial_i\big(F(v).e_i\big) = 0, \tag{1.4}$$

où ∂_i représente la dérivée parielle selon la $i^{ème}$ dimension d'espace. Nous supposons ici que le vecteur v est uniforme sur la face A_{ij}, et par conséquent que le flux $F(v)$ l'est également, de sorte que l'écriture de (1.4) sur la face A_{ij} peut s'écrire :

$$\partial_t v + \partial_{n_{ij}}\big(F(v).n_{ij}\big) = 0, \text{ avec } \partial_{n_{ij}} = n_{ij}.\nabla. \tag{1.5}$$

Il suffit pour s'en assurer de supposer que nous ayons choisi la base canonique e_i , $(1 \leq i \leq d)$ de telle sorte que $e_1 = n_{ij}$. Introduisons maintenant la matrice jacobienne :

$$J(v,n) = \frac{\partial\big(F(v).n\big)}{\partial v},$$

définie pour tout vecteur unitaire n. Notons par ailleurs v_{ij} la valeur du vecteur v sur la face A_{ij}. Elle est considérée égale à une moyenne pondérée des vecteurs v_i^n et v_j^n. Si nous shoisissons de pondérer par les volumes des cellules, nous obtenons :

$$v_{ij} = \frac{|\Omega_i|\ v_i^n + |\Omega_j| v_j^n}{|\Omega_i| + |\Omega_j|}.$$

En multipliant le système (1.5), valable sur la face A_{ij}, par $J(v, n_{ij})$, nous trouvons :

$$\partial_t\big(F(v).n_{ij}\big) + J(v_{ij}, n_{ij})\ \partial_{n_{ij}}\big(F(v).n_{ij}\big) = 0. \tag{1.6}$$

Soit maintenant un vecteur unitaire n et un vecteur de variables conservatives v quelconques. Nous considérerons dans la suite que le système (1.1) est hyperbolique, c'est à dire que $J(v,n)$ est diagonalisable avec des valeurs propres réelles. Nous notons $\mu_k(v,n)$ les valeurs propres de $J(v,n)$, $l_k(v,n)$ ses vecteurs propres à gauche et $r_k(v,n)$ ses vecteurs propres à droite. Ainsi nous pouvons écrire :

$${}^t J(v,n).l_k(v,n) = \mu_k(v,n)\ l_k(v,n), \tag{1.7}$$

$$J(v,n).r_k(v,n) = \mu_k(v,n)\ r_k(v,n). \tag{1.8}$$

Nous introduisons également la matrice $R(v,n)$ dont les colonnes sont les vecteurs $r_k(v,n)$, et la matrice $L(v,n)$ dont les lignes sont les vecteurs $l_k(v,n)$. Nous supposons enfin que nous avons normalisé les vecteurs propres de façon à avoir :

$$L(v,n)\ R(v,n) = Id.$$

Nous pouvons alors écrire :

$$diag\big(\mu(v,n)\big) = L(v,n)\ J(v,n)\ R(v,n), \tag{1.9}$$

où $diag\big(\mu(v,n)\big)$ désigne la matrice diagonales dont les éléments diagonaux sont les valeurs $\mu_k(v,n)$.

Rappelons que nous cherchons une valeur moyenne sur la face A_{ij} de la quantité $F(v).n_{ij}$ notée ϕ_{ij}, alors que nous ne disposons que des quantités centrées $F(v_i)$, et multiplions l'équation (1.6) par ${}^t l_k(v_{ij},n_{ij})$. Nous obtenons en vertu de (1.7) :

$$\big(\partial_t + \mu_k(v_{ij},n_{ij})\partial_{n_{ij}}\big)\Big({}^t l_k(v_{ij},n_{ij})\big(F(v).n_{ij}\big)\Big) = 0, \tag{1.10}$$

ce qui signifie que les quantités ${}^t l_k(v_{ij},n_{ij})\big(F(v).n_{ij}\big)$ sont advectés dans la direction n_{ij} selon les vitesses $\mu_k(v_{ij},n_{ij})$. Afin d'appliquer le principe intuitif consistant à chercher l'information dans le sens opposée à la vitesse d'advection - ou à considérer une moyenne si cette vitesse est nulle -, et en vertu de la remarque précédente, nous supposons que ϕ_{ij} vérifie :

$$\text{si } \mu_k(v_{ij},n_{ij}) > 0,\ \text{alors } {}^t l_k(v_{ij},n_{ij})\ \phi_{ij} = {}^t\ l_k(v_{ij},n_{ij})\big(F(v_i).n_{ij}\big), \tag{1.11}$$

$$\text{si } \mu_k(v_{ij},n_{ij}) < 0,\ \text{alors } {}^t l_k(v_{ij},n_{ij})\ \phi_{ij} = {}^t\ l_k(v_{ij},n_{ij})\big(F(v_j).n_{ij}\big), \tag{1.12}$$

$$\text{si } \mu_k(v_{ij},n_{ij}) = 0,\ \text{alors } {}^t l_k(v_{ij},n_{ij})\ \phi_{ij} = {}^t l_k(v_{ij},n_{ij})\Big(\frac{F(v_i).n_{ij} + F(v_j).n_{ij}}{2}\Big), \tag{1.13}$$

où pour $k \in \{i,j\}$, nous avons $v_k = v_k^n$ ou $v_k = v_k^{n+1}$ en fonction du choix d'une itération implicite ou explicite. Dans l'ensemble de cette thèse, à moins que ce ne soit précisé, nous nous bornerons à l'utilisation en explicite du schéma VFFC. Cet ensemble de conditions détermine exactement le flux ϕ_{ij}, et peut être réécrit de la façon suivante :

$$\phi_{ij} = \left(\frac{F(v_i)+F(v_j)}{2} - sign\big(J(v_{ij},n_{ij})\big)\ \frac{F(v_j)-F(v_i)}{2} \right).n_{ij} \tag{1.14}$$

où la matrice signe est donnée par :

$$sign\big(J(v_{ij},n_{ij})\big) = R(v_{ij},n_{ij})\ diag\Big(sign\big(\mu(v_{ij},n_{ij})\big)\Big)\ L(v_{ij},n_{ij}).$$

Par ailleurs, le pas de temps Δt est contraint par la condition CFL :

$$\Delta t < min_{\ i,j}\ \left(\frac{|\Omega_i|}{|A_{ij}|\ max_k\ |\mu_k(v_{ij},n_{ij})|} \right).$$

1.3.2 Cas du gaz seul

Equations d'Euler 2D.

Nous utilisons la base canonique (e_x, e_z), correspondant respectivement aux directions horizontales et verticales, dans laquelle nous projetons la vitesse u pour obtenir ses coordonnées u_x et u_z. Les équations d'Euler 2D s'écrivent ainsi :

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial z} = 0. \tag{1.15}$$

$$\frac{\partial \rho u}{\partial t} + \frac{\partial \rho u^2}{\partial x} + \frac{\partial \rho uv}{\partial z} + \frac{\partial p}{\partial x} = 0. \tag{1.16}$$

$$\frac{\partial \rho v}{\partial t} + \frac{\partial \rho uv}{\partial x} + \frac{\partial \rho v^2}{\partial z} + \frac{\partial p}{\partial z} = 0. \tag{1.17}$$

où nous avons noté :

- ρ la masse volumique
- p la pression.

Accompagné d'une loi $p(\rho)$ reliant de manière unique pression et masse volumique - sous l'hypothèse d'un écoulement isentropique -, ce système est fermé. Nous pouvons utiliser le formalisme suivant :

$$v = \begin{pmatrix} \rho \\ \rho u_x \\ \rho u_z \end{pmatrix}, F_x(v) = \begin{pmatrix} \rho u_x \\ \rho u_x^2 + p \\ \rho u_x u_z \end{pmatrix}, F_z(v) = \begin{pmatrix} \rho u_z \\ \rho u_x u_z \\ \rho u_z^2 + p \end{pmatrix} \tag{1.18}$$

pour réécrire le système (3.10, 3.11, 3.12) sous la forme :

$$\frac{\partial v}{\partial t} + \frac{\partial F_x(v)}{\partial x} + \frac{\partial F_z(v)}{\partial z} = 0 \tag{1.19}$$

ou encore, en notant $F(v) = \big(F_x(v), F_z(v)\big)$:

$$\frac{\partial v}{\partial t} + div\big(F(v)\big) = 0, \tag{1.20}$$

ce qui permet d'utiliser la théorie VFFC présentée dans [1] et résumée dans la sous-section 1.3.1.

Remarque Il est sous-entendu ici comme dans la suite du document que les paramètres physiques u_x, u_z, ρ et p se déterminent de manière unique à partir du vecteur v, et qu'il en va par conséquent de même de toutes les variables physiques pouvant en résulter, ainsi que du flux F. En notant $v = (\rho, \rho\ u_x, \rho\ u_z) = (v_1, v_2, v_3)$, nous avons en effet :

$$u_x = \frac{v_2}{v_1},\ u_z = \frac{v_3}{v_1},\ \rho = v_1 \text{ et } p = p(v_1).$$

Reprenons ici les grandes lignes de la théorie VFFC dans notre cas particulier à deux dimensions sur un maillage cartésien. Intégrons tout d'abord le système d'équations (3.13) sur une maille $K = [x_{-\frac{1}{2}}, x_{+\frac{1}{2}}] \times [z_{-\frac{1}{2}}, z_{+\frac{1}{2}}]$:

$$\int_K \frac{\partial v}{\partial t} + \int_K \frac{\partial F_x(v)}{\partial x} + \int_K \frac{\partial F_z(v)}{\partial z} = 0. \tag{1.21}$$

Nous notons $v_{i,j}$ la valeur moyenne de v dans la maille K, d_{xi} et d_{zj} étant les dimensions de cette maille. Alors (3.14) mène à la discrétisation suivante :

$$v_{i,j}^{n+1} = v_{i,j}^{n} - \frac{\Delta t}{d_{xi}}\big(F_{x_{i+\frac{1}{2},j}} - F_{x_{i-\frac{1}{2},j}}\big) - \frac{\Delta t}{d_{zj}}\big(F_{z_{i,j+\frac{1}{2}}} - F_{z_{i,j-\frac{1}{2}}}\big), \tag{1.22}$$

les exposants n et $n+1$ correspondant à deux itérations successives séparées par le pas de temps Δt. Il reste à définir $F_{x_{i\pm\frac{1}{2},j}}$ et $F_{z_{i,j\pm\frac{1}{2}}}$, qui correspondent aux valeurs moyennes aux faces des flux F_x et F_z. Ces flux aux faces s'expriment en fonction des flux aux centres des cellules adjacentes, eux mêmes issus des valeurs de $v_{i,j}$ correspondantes. La méthode de calcul de ces flux, présentée en détails dans [1] et résumée dans la sous-section 1.3.1, fait intervenir des termes de décentrement dépendant des gradients des flux physiques.

Plus spécifiquement dans ce cas, la grille étant cartésienne et orientée selon les deux directions e_x et e_z de la base canonique, seules les normales e_x et e_z sont considérées, et l'on a :

$$\begin{aligned}
\mu_1(v,e_x) &= u_x - c \text{ et } \mu_1(v,e_z) = u_z - c,\\
\mu_2(v,e_x) &= u_x \text{ et } \mu_2(v,e_z) = u_z,\\
\mu_3(v,e_x) &= u_x + c \text{ et } \mu_3(v,e_z) = u_z + c.
\end{aligned}$$

$$\begin{aligned}
J(v,e_x) &= \begin{pmatrix} 0 & 1 & 0\\ c^2-u_x^2 & 2\,u_x & 0\\ -u_x\,u_z & u_z & u_x \end{pmatrix} \text{ et } J(v,e_z) = \begin{pmatrix} 0 & 0 & 1\\ -u_x\,u_z & u_z & u_x\\ c^2-u_z^2 & 0 & 2\,u_z \end{pmatrix},\\
R(v,e_x) &= \begin{pmatrix} 1 & 0 & 1\\ u_x-c & 0 & u_x+c\\ u_z & 1 & u_z \end{pmatrix} \text{ et } R(v,e_z) = \begin{pmatrix} 1 & 0 & 1\\ u_x & 1 & u_x\\ u_z-c & 0 & u_z+c \end{pmatrix},\\
L(v,e_x) &= \begin{pmatrix} \frac{c+u_x}{2\,c} & \frac{-1}{2\,c} & 0\\ -u_z & 0 & 1\\ \frac{c-u_x}{2\,c} & \frac{1}{2\,c} & 0 \end{pmatrix} \text{ et } L(v,e_z) = \begin{pmatrix} \frac{c+u_z}{2\,c} & 0 & \frac{-1}{2\,c}\\ -u_x & 1 & 0\\ \frac{c-u_z}{2\,c} & 0 & \frac{1}{2\,c} \end{pmatrix}.
\end{aligned}$$

Et nous avons finalement pour tout i, le choix étant fait d'un schéma explicite :

$$\begin{aligned}
F_{x_{i+\frac{1}{2},j}} = &\frac{F_x(v_{i,j}^n) + F_x(v_{i+1,j}^n)}{2}\\
&- sign\big(J(v_{i+\frac{1}{2},j}, e_x)\big)\ \frac{F_x(v_{i+1,j}^n) - F_x(v_{i,j}^n)}{2},
\end{aligned} \tag{1.23}$$

et pour tout j :

$$\begin{aligned}
F_{z_{i,j+\frac{1}{2}}} = &\frac{F_z(v_{i,j}^n) + F_z(v_{i,j+1}^n)}{2}\\
&- sign\big(J(v_{i,j+\frac{1}{2}}, e_z)\big)\ \frac{F_z(v_{i,j+1}^n) - F_z(v_{i,j}^n)}{2},
\end{aligned} \tag{1.24}$$

avec, pour $n = e_x$ ou $n = e_z$:

$$sign\big(J(v,n)\big) = R(v,n)\ S(v,n)\ L(v,n), \tag{1.25}$$

où

$$S(v,n) = \begin{pmatrix} sign\big(\mu_1(v,n)\big) & 0 & 0 \\ 0 & sign\big(\mu_2(v,n)\big) & 0 \\ 0 & 0 & sign\big(\mu_3(v,n)\big) \end{pmatrix}.$$

Par ailleurs, les valeurs de v aux faces sont les suivantes :

$$v_{i+\frac{1}{2},j} = \frac{d_{xi}\ v^n_{i,j} + d_{xi+1}\ v^n_{i+1,j}}{d_{xi} + d_{xi+1}} \tag{1.26}$$

et

$$v_{i,j+\frac{1}{2}} = \frac{d_{zj}\ v^n_{i,j} + d_{zj+1}\ v^n_{i,j+1}}{d_{zj} + d_{zj+1}}. \tag{1.27}$$

Enfin, la condition CFL s'écrit dans ce cas :

$$\Delta t < min\ _{i,j} \left(\frac{|\Omega_i|}{|A_{ij}|\ max_k\ |\lambda_k(V_{ij}, n_{ij})|} \right)$$

$$= min\Bigg(\ min\ _{i,j} \left(\frac{d_{xi}}{|u_{x\ i,j}| + |c_{i,j}|} \right) , min\ _{i,j} \left(\frac{d_{zj}}{|u_{z\ i,j}| + |c_{i,j}|} \right) \ \Bigg).$$

Maillage mobile

Nous souhaitons reprendre l'équation (3.14), en intégrant cette fois l'équation (3.13) sur une cellule $K(t)$ dont la hauteur peut varier :

$$K(t) = [x_{-\frac{1}{2}}, x_{+\frac{1}{2}}] \times [z_{-\frac{1}{2}}(t), z_{+\frac{1}{2}}(t)].$$

Nous partons de l'expression suivante, valable pour toute fonction $f(z,t)$ sous réserve de bonnes propriétés d'intégration et d'ensemble de définition :

$$\frac{d}{dt}\int_{K(t)} f\ dz = \int_{K(t)} f_t\ dz + \int_{K(t)} (\lambda\ f)_z\ dz, \tag{1.28}$$

où λ désigne la vitesse d'évolution du domaine d'intégration, dans notre cas la vitesse de maillage. Cette propriété nous permet d'écrire :

$$\begin{aligned}
&\int_{K(t)} \frac{\partial v}{\partial t} + \int_{K(t)} \frac{\partial F_x(v)}{\partial x} + \int_{K(t)} \frac{\partial F_z(v)}{\partial z} \\
&= \int_{x_{-\frac{1}{2}}}^{x_{+\frac{1}{2}}} \left(\int_{z_{-\frac{1}{2}}(t)}^{z_{+\frac{1}{2}}(t)} \frac{\partial v}{\partial t}\ dz \right) + \int_{K(t)} \frac{\partial F_x(v)}{\partial x} + \int_{K(t)} \frac{\partial F_z(v)}{\partial z} \\
&= \int_{x_{-\frac{1}{2}}}^{x_{+\frac{1}{2}}} \left(\frac{\partial}{\partial_t} \left(\int_{z_{-\frac{1}{2}}(t)}^{z_{+\frac{1}{2}}(t)} v\ dz \right) - \int_{z_{-\frac{1}{2}}(t)}^{z_{+\frac{1}{2}}(t)} \frac{\partial}{\partial_z}(\lambda\ v)\ dz \right) \\
&+ \int_{K(t)} \frac{\partial F_x(v)}{\partial x} + \int_{K(t)} \frac{\partial F_z(v)}{\partial z} \\
&= \frac{\partial}{\partial_t} \left(\int_{K(t)} v \right) + \int_{K(t)} \frac{\partial F_x(v)}{\partial x} + \int_{K(t)} \frac{\partial \big(F_z(v) - \lambda\ v\big)}{\partial z}.
\end{aligned} \tag{1.29}$$

Notons maintenant $F_z^*(v,\lambda) = F_z(v) - \lambda v$. Le système (1.20) s'intègre sur $K(t)$ de la façon suivante :

$$\frac{\partial}{\partial_t}\left(\int_{K(t)} v\right) + \int_{K(t)} \frac{\partial F_x(v,\lambda)}{\partial x} + \int_{K(t)} \frac{\partial F_z^*(v,\lambda)}{\partial z} = 0. \tag{1.30}$$

Cela mène, en considérant toujours que les flux sont calculés en explicite, à la discrétisation suivante :

$$V_{i,j}^{n+1} = \frac{{d_z}_j^n}{{d_z}_j^{n+1}}\, V_{i,j}^n - \frac{\Delta t\ {d_z}_j^n}{d_{xi}\ {d_z}_j^{n+1}}\big(F_{x_{i+\frac{1}{2},j}} - F_{x_{i-\frac{1}{2},j}}\big) - \frac{\Delta t}{{d_z}_j^{n+1}}\big({F^*_z}_{i,j+\frac{1}{2}} - {F^*_z}_{i,j-\frac{1}{2}}\big), \tag{1.31}$$

avec ${F^*_z}_{i,j\pm\frac{1}{2}}$ se calculant de la même manière que dans le cas sans mouvement de maille. il nous faut déterminer :

$$J^*(v,e_z,\lambda) = \frac{\partial F_z^*(v,\lambda)}{\partial v},$$

ainsi que les valeurs propres $\mu_k^*(v,e_z,\lambda)$ de $J^*(v,e_z,\lambda)$ et les deux matrices $L_k^*(v,e_z,\lambda)$ et $R_k^*(v,e_z,\lambda)$ des vecteur propres à gauche et à droite, eux-mêmes notés respectivement $l_k^*(v,e_z,\lambda)$ et $r_k^*(v,e_z,\lambda)$. Tout cela est très simple : il est clair que nous avons :

$$J^*(v,e_z,\lambda) = J(v,e_z) - \lambda I_d.$$

Cela entraîne que les vecteurs propres sont inchangés, de sorte que :

$$L_k^*(v,e_z,\lambda) = L_k(v,e_z) \text{ et } R_k^*(v,e_z,\lambda) = R_k(v,e_z).$$

Par ailleurs nous avons naturellement $\mu_k^*(v,e_z,\lambda) = \mu_k(v,e_z) - \lambda$. Ainsi, reprenant les développements de la méthode VFFC sur maillage fixe, nous pouvons écrire pour tous i et j :

$$\begin{aligned} {F_z^*}_{i,j+\frac{1}{2}} = {} & \frac{F_z^*(v_{i,j}^n,\lambda_j) + F_z^*(v_{i,j+1}^n,,\lambda_{j+1})}{2} \\ & - sign\big(J^*(v_{i,j+\frac{1}{2}},e_z,\lambda_{j+\frac{1}{2}})\big)\ \frac{F_z^*(v_{i,j+1}^n,\lambda_{j+1}) - F_z^*(v_{i,j}^n,\lambda_j)}{2}, \end{aligned} \tag{1.32}$$

où

$$sign\big(J^*(v,e_z)\big) = R(v,e_z)\ S^*(v,e_z)L(v,e_z), \tag{1.33}$$

avec

$$S^*(v,e_z,\lambda) = \begin{pmatrix} sign\big(\mu_1(v,e_z)-\lambda\big) & 0 & 0 \\ 0 & sign\big(\mu_2(v,e_z)-\lambda\big) & 0 \\ 0 & 0 & sign\big(\mu_3(v,e_z)-\lambda\big) \end{pmatrix}.$$

Notant maintenant, pour j quelconque, $z_{j+\frac{1}{2}}^n$ et $z_{j+\frac{1}{2}}^{n+1}$ les positions du nœud $j+\frac{1}{2}$ aux itérations n et $n+1$. Nous écrivons :

$$\lambda_{j+\frac{1}{2}} = \frac{z_{j+\frac{1}{2}}^{n+1} - z_{j+\frac{1}{2}}^n}{\Delta t} \tag{1.34}$$

et

$$\lambda_j = \frac{\lambda_{j+\frac{1}{2}} + \lambda_{j-\frac{1}{2}}}{2}. \tag{1.35}$$

Enfin, la condition CFL s'écrit dans ce cas :

$$\Delta t < min_{\ i,j} \left(\frac{|\Omega_i|}{|A_{ij}|\ max_k\ |\lambda_k(V_{ij}, n_{ij})|} \right)$$

$$= min\Bigg(min_{\ i,j} \left(\frac{d_{xi}}{|u_{x\ i,j}| + |c_{i,j}|} \right), min_{\ i,j} \left(\frac{d_{zj}}{|u_{z\ i,j} - \lambda_j| + |c_{i,j}|} \right) \Bigg).$$

Condition de mur

Nous ne considérerons dans ce document que des conditions aux limites de mur. Nous cherchons ici à calculer les flux frontaliers, pour lesquels les précédentes expressions ne peuvent s'appliquer car elles nécessiteraient l'existence de cellules extérieures au domaine. Rappelons d'abord que l'expression générale du flux est $F = (F_x, F_y)$ avec

$$F_x = (\rho u_x \rho u_x^2 + p \rho u_x u_z) \text{ et } F_z = (\rho u_z, \rho u_x u_z, \rho u_z^2 + p).$$

Si la frontière est orthogonale à e_x, nous pouvons affirmer que la composante u_x est nulle à la frontière, ce qui permet d'écrire la composante du flux de bord parallèle à la normale e_x sous la forme $F_b = (0, p_b, 0)$ où p_b désigne une pression de bord. De même pour une frontière orthogonale à e_z, la composante du flux de bord parallèle à la normale e_z s'écrit $F_b = (0, 0, p_b)$. Il reste donc à déterminer la pression de bord p_b.

Cas d'un maillage fixe. Nous nous basons ici sur la sous-section 1.3.1. Notons que pour $n \in \{e_x, e_z, -e_x, -e_z\}$, la troisième valeur propre de $J(v, n)$ s'écrit $\mu_3(v, n) = (u.n) + c$. Cette expression a déjà été présentée dans la section 1.3.2 pour $n \in \{e_x, e_z\}$ et se vérifie également pour les normales opposées. Cette troisième valeur propre est toujours positive dans le cas des écoulements subsoniques. Nous faisons dans la suite l'hypothèse que nos écoulements sont subsoniques - et donc que cette valeur propre est positive -, ce qui mène, en reprenant l'équation (1.11), à l'écriture suivante :

$${}^t l_3(v_b, n)\ F_b = {}^t\ l_3(v_b, n)\big(F(v_{int}).n\big)$$

où v_b désigne le vecteur des variables conservatives au bord et v_{int} son équivalent au centre de la cellule de bord. En faisant l'approximation $v_b = v_{int}$, nous écrivons :

$${}^t l_3(v_{int}, n)\ F_b = {}^t\ l_3(v_{int}, n)\big(F(v_{int}).n\big),$$

ce qui, en rappelant que le flux F_b ne s'exprime qu'en fonction de la variable p_b, permet de calculer p_b de façon unique. Plus spécifiquement nous avons, pour $n \in \{e_x, -e_x\}$:

$${}^t l_3(v, n) = \Big(\frac{c - (u.n)}{2\ c}, \frac{1}{2\ c}, 0\Big) \tag{1.36}$$

et pour $n \in \{e_z, -e_z\}$:

$${}^t l_3(v, n) = \Big(\frac{c - (u.n)}{2\ c}, 0, \frac{1}{2\ c}\Big). \tag{1.37}$$

Cas d'un maillage mobile. Nous avons vu à la sous-section 1.3.2 que la matrice $J(v, e_z)$ se transformait en $J*(v, e_z, \lambda)$, et que les vecteurs propres de $J(v, e_z)$, notées $l_k(v, e_z)$ et $l_k(v, e_z)$, étaient les mêmes que ceux de $J^*(v, e_z, \lambda)$, notés $l_k^*(v, e_z, \lambda)$ et $l_k^*(v, e_z, \lambda)$. Par ailleurs nous savons que la matrice $J(v, e_x)$ reste inchangée en raison de l'absence de mouvement de mailles selon x. Enfin, nous savons par hypothèse que la vitesse de maillage λ est nulle sur la frontière, de sorte que la composante parallèle à la normale extérieure d'un flux de bord, notée F_b^* dans le cas d'un maillage mobile, a la même expression que F_b et ne dépend donc que de p_b. Les remarques précédentes permettent de déterminer p_b, et donc F_b^*, en écrivant :

$$ {}^t l_3^*(v_{int}, n)\ F_b^* = {}^t\ l_3^*(v_{int}, n)\big(F^*(v_{int}).n\big) \tag{1.38} $$

que nous pouvons mettre sous la forme :

$$ {}^t l_3(v_{int}, n)\ F_b = {}^t\ l_3(v_{int}, n)\big(F^*(v_{int}).n\big) \tag{1.39} $$

avec $F^*(v_{int}).e_z = F(v_{int}) - \lambda_{int} v$ et $F^*(v_{int}).e_x = F(v_{int})$. Nous avons noté λ_{int} la vitesse de maillage dans la cellule de bord, qui se calcule comme présenté dans la sous-section 1.3.2.

1.3.3 Cas d'une tranche de liquide seule

Sans interface

Nous considérons maintenant le cas d'une colonne verticale remplie de liquide. Les cellules sont indexées par l'indice j dans le sens des z croissants. Le formalisme 2D est toujours valable, à condition de considérer nuls les flux horizontaux $F_{x_{i\pm\frac{1}{2}},j}$ et de ne pas prendre en considération la vitesse horizontale u_x. En omettant par ailleurs les indices z et i, qui précisent respectivement la direction des flux et la position horizontale de la cellule considérée, nous pouvons écrire :

$$ v_j^{n+1} = v_j^n - \frac{\Delta t}{d_{zj}}\big(F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}}\big), \tag{1.40} $$

où les flux et variables conservatives ont la même signification que dans la section 1.3.2, et où l'on ne s'intéresse qu'aux première et troisième variables du vecteur v, à savoir ρ et $\rho\ u_z$. La vitesse u_z sera notée u dans la suite de cette section.

Le même raisonnement est valable lorsque l'on considère un maillage mobile, et nous pouvons écrire, en nous basant toujours sur la section 1.3.2 :

$$ v_j^{n+1} = \frac{{d_{zj}}^n}{{d_{zj}}^{n+1}}\ v_j^n - \frac{\Delta t}{{d_{zj}}^{n+1}}\big({F^*}_{j+\frac{1}{2}} - {F^*}_{j-\frac{1}{2}}\big). \tag{1.41} $$

Enfin, la condition CFL s'écrit dans ce cas :

$$ \Delta t < min\ _j\ \left(\frac{d_{zj}}{|u_j| + |c_j|}\right). $$

Remarque Nous aurions pu tout aussi bien écrire les équations d'Euler 1D et les résoudre selon la méthode VFFC. Néanmoins, dans le cadre du problème posé, qui suppose la résolution d'équations 2D dans le gaz et 1D dans le liquide, nous préférons pour simplifier la programmation nous baser sur un unique schéma, quitte à déterminer des termes inutiles dans le cas 1D. A titre informatif, si l'on souhaite pour des questions d'efficacité informatique écrire les équations 1D dans le liquide, elles sont les suivantes :

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u}{\partial x} = 0 \tag{1.42}$$

$$\frac{\partial \rho u}{\partial t} + \frac{\partial \rho u^2}{\partial x} + \frac{\partial p}{\partial x} = 0. \tag{1.43}$$

Nous notons alors :

$$v = \begin{pmatrix} \rho \\ \rho u \end{pmatrix} \text{ et } F(v) = \begin{pmatrix} \rho u \\ \rho u^2 + p \end{pmatrix}.$$

Concernant les matrices et valeurs propres, le calcul donne les expressions suivantes, où nous omettons la variable correspondant à la normale car il n'y a qu'une direction d'espace :

$$\mu_1(v) = u - c \text{ et } \mu_2(v) = u + c,$$

$$J(v) = \begin{pmatrix} 0 & 1 \\ c^2 - u^2 & 2\,u \end{pmatrix},\ R(v) = \begin{pmatrix} 1 & 1 \\ u - c & u + c \end{pmatrix}$$

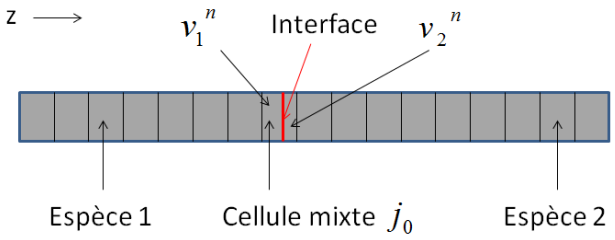
$$\text{et } L(v) = \begin{pmatrix} \frac{c+u}{2\,c} & \frac{-1}{2\,c} \\ \frac{c-u}{2\,c} & \frac{1}{2\,c}. \end{pmatrix}.$$

Ces expressions permettent d'utiliser la théorie VFFC présentée dans la section 1.3.1, puis les formules (3.16) et (1.41) avec cette fois des vecteurs à deux composantes.

Avec interface

Le problème posé est celui d'une colonne - représentée horizontalement dans les schémas qui suivront mais correspondant bien à une tranche de liquide verticale dans le cadre de notre problème - contenant une espèce 1 à gauche et une espèce 2 à droite, chacune d'entre elles respectant les équations d'Euler 1D isentropiques. Nous souhaitons connaître l'évolution de ce système et notamment l'évolution de la position de l'interface au cours du temps, l'hypothèse étant faite qu'aucun mélange ne se crée entre les deux espèces. Le traitement de la difficulté concernant l'interface est présenté ci-dessous. Nous disposons à l'itération n :

- d'une valeur du vecteur v notée v_j^n dans chaque cellule j ne contenant pas l'interface (cellules pures),
- de la position de l'interface, et de l'indice j_0 de la cellule qui la contient,
- de deux valeurs v_1^n et v_2^n du vecteur v de part et d'autre de l'interface.



Les différentes étapes pour passer de l'itération n à l'itération $n+1$ sont présentées ci-dessous. Il est important de noter que le maillage utilisé reste fixe, bien que la position de l'interface entre espèces soit variable et ne coïncide pas a priori avec un nœud de ce maillage. La méthode a pour idée générale de traiter chaque espèce par un schéma eulérien, et l'interface par un schéma lagrangien. Elle se décline ainsi :

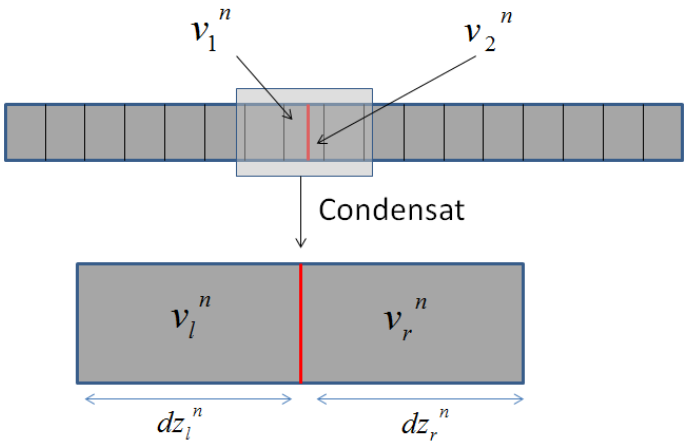
- on calcule les flux $F_{j+\frac{1}{2}}$ et $F_{j-\frac{1}{2}}$ aux faces selon la méthode VFFC de même que dans la sous-section 1.3.3, à l'exception des faces de la cellule j_0, et on détermine les valeurs v_j^{n+1} pour $j \notin \{j_0 - 1\ ,\ j_0\ ,\ j_0 + 1\}$ grâce à la formule (1.41).
- On isole la cellule j_0 ainsi que les deux cellules adjacentes, cet ensemble étant appelé *condensat*.
- On calcule deux valeurs v_l^n et v_r^n du vecteur v dans ce *condensat*, de part et d'autre de l'interface, grâce à une moyenne arithmétique pondérée par les volumes occupés. Plus précisément, si l'interface se situe à une distance d_{zint} de la face $j_0 - \frac{1}{2}$ à l'itération n, alors on note :

$$d_{zl}^{\,n} = d_{zjo-1} + d_{zint},\ d_{zr} = d_{zjo} + d_{zjo+1} - d_{zint},$$

et l'on on a :

$$v_l^n = \frac{d_{zjo-1}\ v_{j_0-1}^n + d_{zint}\ v_1^n}{d_{zl}^{\,n}}$$

$$v_r^n = \frac{d_{zjo+1}\ v_{j_0+1}^n + (d_{zj_0} - d_{zint})\ v_2^n}{d_{zr}^{\,n}}.$$



- On calcule p_{int} et u_{int}, les pression et vitesse à l'interface, d'après ([12], p.44) : on déduit du vecteur v_l^n les paramètres physiques ρ_l, c_l et p_l correspondant respectivement à la masse volumique, la vitesse du son et la pression dans la partie gauche du *condensat*. De même, on déduit du vecteur v_r^n les variables ρ_r, c_r et p_r. On calcule alors une pression et une vitesse d'interface selon les formules suivantes :

$$p_{int} = \frac{\rho_r c_r p_l + \rho_l c_l p_r}{\rho_l c_l + \rho_r c_r} + \rho_l c_l \rho_r c_r \frac{u_l - u_r}{\rho_l c_l + \rho_r c_r}$$

$$u_{int} = \frac{\rho_r c_r u_l + \rho_l c_l u_r}{\rho_l c_l + \rho_r c_r} + \rho_l c_l \rho_r c_r \frac{p_l - p_r}{\rho_l c_l + \rho_r c_r}$$

- L'étape suivante consiste à calculer un flux lagrangien en fonction de ces paramètres. En effet nous allons dans la suite déplacer l'interface à la même vitesse que fluide, par conséquent il ne faut pas prendre en compte les flux convectifs. Le flux d'interface se résume donc à :

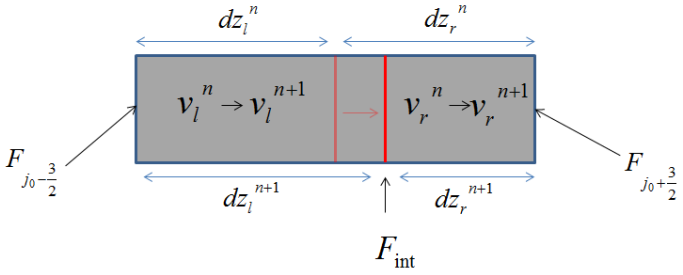
$$F_{int} = \big(0, p_{int}\big).$$

- On opère le mouvement de l'interface selon la vitesse u_{int} précédemment calculée. On note $d_{z_l}^{n+1} = d_{z_l}^{n} + \Delta t\ u_{int}$ et $d_{z_r}^{n+1} = d_{z_r}^{n} - \Delta t\ u_{int}$.

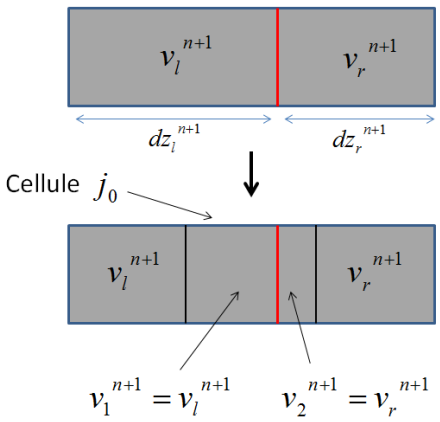
  Remarque Si la condition CFL est respectée, la vitesse du fluide multipliée par le pas de temps ne peut excéder la taille d'une cellule (cf. sous-section 1.3.2), de sorte que l'interface ne peut pas sortir du *condensat* lors de cette étape.

- Le flux d'interface permet d'actualiser naturellement les vecteurs v_l et v_r. En notant v_l^{n+1} et v_r^{n+1} les valeurs respectives de v_l et v_r à l'instant $t + \Delta t$, et en reprenant l'équation (1.41), on obtient les formules :

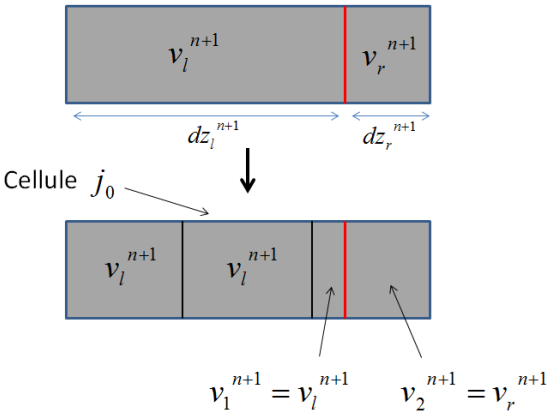
$$d_{z_l}^{n+1}\ v_l^{n+1} = d_{z_l}^{n}\ v_l^{n} - \Delta t\big(F_{j_0-\frac{3}{2}} - F_{int}\big),$$
$$d_{z_r}^{n+1}\ v_r^{n+1} = d_{z_r}^{n}\ v_r^{n} - \Delta t\big(F_{int} - F_{j_0+\frac{3}{2}}\big).$$



- Il faut maintenant projeter les valeurs v_l^{n+1} et v_r^{n+1} ainsi obtenues sur le maillage initial. Plusieurs cas sont possibles :
  - Si l'interface reste dans la même cellule j_0, on affecte à la cellule $j_0 - 1$ le vecteur v_l^{n+1} et à la cellule $j_0 + 1$ le vecteur v_r^{n+1}. La position de l'interface au sein de la cellule j_0 est gardée en mémoire, ainsi que les vecteurs v_l^{n+1} et v_r^{n+1} au sein de la cellule mixte, car on aura pour la prochaine itération $v_1^{n+1} = v_l^{n+1}$ et $v_2^{n+1} = v_r^{n+1}$.



– Si l'interface a traversé un nœud séparant deux cellules - supposons qu'il s'agisse du nœud $j_0 + \frac{1}{2}$, l'autre cas s'opérant de la même façon -, on affecte le vecteur v_l^{n+1} aux cellules j_0 et $j_0 - 1$. La cellule j_0 devient une cellule pure, tandis que la cellule $j_0 + 1$ devient une cellule mixte. Les vecteurs v_l^{n+1} et v_r^{n+1} sont gardés en mémoire au sein de la cellule mixte, car on aura pour la prochaine itération $v_1^{n+1} = v_l^{n+1}$ et $v_2^{n+1} = v_r^{n+1}$.



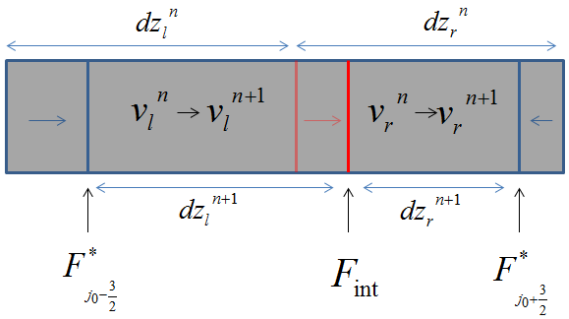
Interface et maillage mobile

Nous avons vu que la prise en compte d'un maillage mobile revenait à modifier l'expression des flux et la discrétisation temporelle. Les formules de réactualisation des variables conservatives s'écrivent cette fois : :

$$d_{z_l}^{n+1}\, v_l^{n+1} = d_{z_l}^{n}\, v_l^{n} - \Delta t\big(F^*_{j_0-\frac{3}{2}} - F_{int}\big)$$

$$d_{z_r}^{n+1}\, v_r^{n+1} = d_{z_r}^{n}\, v_r^{n} - \Delta t\big(F_{int} - F^*_{j_0+\frac{3}{2}}\big).$$

Dans ces expressions l'exposant $*$ désigne la modification du flux pour prise en compte du mouvement du maillage, comme vu dans la sous-section 1.3.2. Les valeurs de $d_{z_l}^{n+1}$ et $d_{z_r}^{n+1}$ sont calculées en prenant en compte à la fois le mouvement des interfaces et celui des extrémités du *condensat*. Nous notons que les flux aux extrémités du *condensat* prennent en compte le mouvement du maillage, mais que le flux à l'interface reste inchangé. En effet le mouvement de l'interface dépend de la physique et non de la vitesse de maillage.



La condition CFL étant écrite en prenant en compte le mouvement des mailles (cf. sous-section 1.3.2), elle interdit toujours l'interface de sortir du *condensat* au cours

d'une itération, et l'étape de projection s'effectue de la même façon que pour le cas des mailles fixes.

1.4 Méthode mise au point

1.4.1 Présentation générale

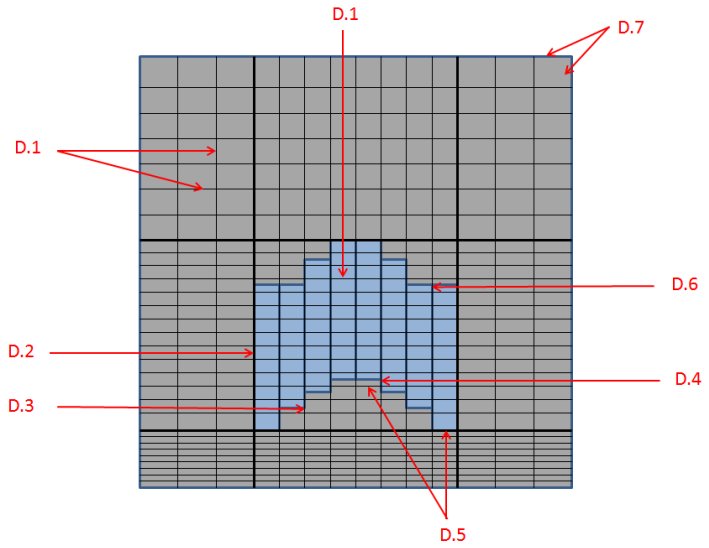
L'idée générale a été évoquée dans la partie 1.2, elle consiste à résoudre les écoulements du gaz grâce à un modèle Euler 2D sur maillage mobile, les écoulements dans chaque tranche de liquide - indépendamment les une des autres - grâce à un modèle Euler 1D sur maillage mobile, et les interfaces horizontales entre liquide et gaz grâce au modèle d'interface présenté dans la sous-section 1.3.3. Plus précisément, les différentes étapes sont les suivantes :

A. Actualisation du maillage, pour qu'il respecte les critères de position suivants : selon la verticale, l'interface la plus haute et l'interface la plus basse séparent le domaine en trois sous domaines. Dans chacun de ces sous domaines, le nombre de cellules est fixe et la taille des cellules est uniforme.

B. Calcul de la vitesse de maillage λ dans chaque cellule grâce aux formules (1.34) et (1.35).

C. Calcul des vecteurs de variables conservatives et de flux au sein de chaque cellule grâce à l'équation (1.18), la loi d'état étant différente selon qu'il s'agisse d'une cellule de liquide ou de gaz.

D. Calcul des flux aux faces des cellules. Plusieurs sous-cas se présentent, récapitulés sur la figure suivante :



D.1. Si la face est entièrement entourée de gaz ou entièrement entourée de liquide, le flux de face est calculé selon les formules (1.23) et (1.24). S'il s'agit d'une face verticale séparant deux cellules de liquide, le flux ne sera pas utilisé. Naturellement, on emploiera les lois d'état du gaz ou du liquide selon qu'il s'agisse d'une face séparant deux cellules de gaz ou deux cellules de liquide.

D.2. Si la face est une interface verticale séparant une cellule entièrement liquide et une cellule entièrement gazeuse, le flux de face est pris égal au flux de

condition de mur défini dans la sous-section 1.3.2. Les propriétés physiques considérées sont celles de la cellule de gaz. En effet ce flux servira à actualiser la cellule de gaz et ne sera pas pris en compte dans la cellule de liquide.

D.3. S'il s'agit d'une face verticale séparant une cellule mixte et une cellule de gaz, deux flux F_l et F_r sont déterminés. On utilisera le flux F_v pour actualiser la cellule de droite et le flux F_r pour actualiser la cellule de gauche. Ces flux se calculent de la façon suivante :

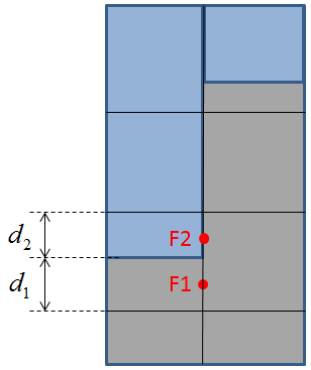
D.3.a. Un flux F_1 est calculé selon les formules (1.23) et (1.24), comme si la partie gazeuse de la cellule mixte remplissait toute sa cellule.

D.3.b. Un flux F_2 est calculé égal au flux de condition de mur défini dans la sous-section 1.3.2. Les propriétés physiques considérées sont celles de la cellule entièrement gazeuse.

D.3.c. Supposons que la cellule mixte est la cellule de gauche, l'autre cas se traitant en intervertissant les expressions. Considérons l'exemple générique représenté sur la figure suivante et écrivons :

$$F_l = F_1$$

$$F_r = \frac{d_1\ F_1 + d_2\ F_2}{d_1 + d_2}.$$



Le flux F_l sera utilisé pour actualiser la partie gazeuse de la cellule mixte de gauche, tandis que le flux F_r sera utilisé pour actualiser la cellule pure de droite.

D.4. S'il s'agit d'une face verticale séparant une cellule mixte et une cellule de liquide, le flux de face est pris égal au flux de condition de mur défini dans la sous-section 1.3.2. Les propriétés physiques considérées sont celles de la partie gazeuse de la cellule mixte. En effet ce flux servira à actualiser la partie gazeuse de la cellule mixte et ne sera pas pris en compte dans la cellule de liquide.

D.5. S'il s'agit d'une face horizontale d'une cellule mixte, ou s'il s'agit d'une face verticale séparant deux cellules mixtes, le calcul des flux n'importe pas. Ce cas sera traité spécifiquement au point G.

D.6. S'il s'agit d'une face verticale séparant deux cellules mixtes, deux flux F_l et F_r sont déterminés. De même que dans le cas $D.3$, on utilisera le flux F_v pour actualiser la cellule de gauche et le flux F_r pour actualiser la cellule de droite. Ces flux se calculent de la façon suivante :

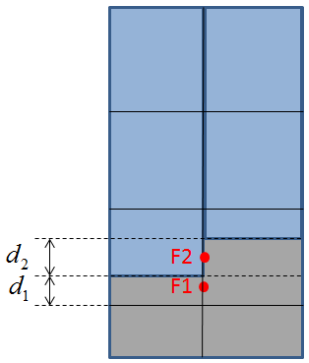
D.6.a. Un flux F_1 est calculé selon les formules (1.23) et (1.24), comme si les parties gazeuses des cellules mixtes remplissaient entièrement leurs cellules

respectives.

D.6.b. Un flux F_2 est calculé égal au flux de condition de mur défini dans la sous-section 1.3.2. Les propriétés physiques considérées sont celles de la partie gazeuse de la cellule mixte qui contient le plus grand volume de gaz.

D.6.c. Supposons que la cellule mixte qui contient le plus grand volume de gaz soit celle de droite, l'autre cas se traitant en intervertissant les expressions. Considérons l'exemple générique représenté sur la figure suivante et écrivons :

$$F_l = F_1$$
$$F_l = \frac{d_1\ F_1 + d_2\ F_2}{d_1 + d_2}.$$



Le flux F_l sera utilisé pour actualiser la partie gazeuse de la cellule mixte de gauche, tandis que le flux F_r sera utilisé pour actualiser la partie gazeuse de la cellule mixte de droite.

D.7. S'il s'agit d'une face de bord, le flux est pris égal au flux de condition de mur défini dans la sous-section 1.3.2.

E. Actualisation des vecteurs v de toutes les cellules de gaz, grâce à la formule (3.16), à l'exception des cellules mixtes et des cellules inférieures et supérieures aux cellules mixtes. Les flux utiles pour cette phase ont tous été traités au point *D*.

F. Actualisation des vecteurs v de toutes les cellules de liquide, grâce à la formule (1.41), à l'exception des cellules mixtes et des cellules inférieures et supérieures aux cellules mixtes. Les flux utiles pour cette phase ont tous été traités au point *D*.

G. Gestion des *condensats*, basée sur les sous-sections 1.3.3 et 1.3.3. On rappelle que les *condensats* sont constitués d'une cellule mixte, c'est à dire contenant une interface entre gaz et liquide, ainsi que des cellules immédiatement inférieure et supérieure à cette cellule mixte. La figure suivante représente l'exempel générique d'un *condensat* entouré de ses cellules voisines, et auquel on a affecté quatre vecteurs de variables conservatives nommés de v_1 à v_4. Les flux utiles dans la suite ont été représentés et nommés de F_1 à F_6. Ils ont tous été traités au point *D*. Les flux F_2 et F_5 ont la particularité de n'être pas propres qu'à leur face mais également à la cellule qu'ils actualisent. Dans notre exemple il s'agit de la cellule située dans la colonne centrale. Ces flux F_2 et F_5 correspondent respectivement aux cas *D*.3 et *D*.6. L'actualisation des variables conservatives se fait en deux temps :

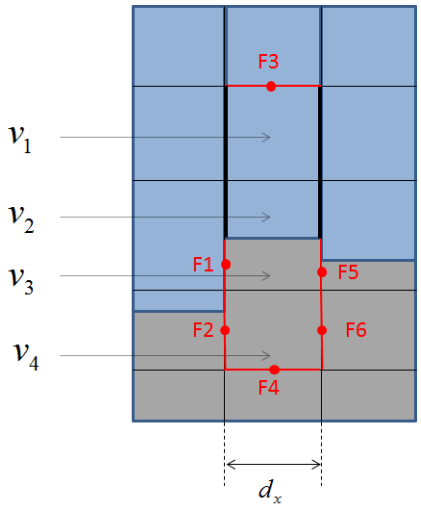
G.1. Apport des flux horizontaux pour transformer les variables à l'itération courante n en variables à l'itération intermédiaire notée $n + \frac{1}{2}$. Ainsi, pour l'exemple représenté sur la figure nous écrivons :

$$v_1^{n+\frac{1}{2}} = v_1^n$$

$$v_2^{n+\frac{1}{2}} = v_2^n$$

$$v_3^{n+\frac{1}{2}} = v_3^n + \frac{\Delta t}{d_x}(F_1 - F_5)$$

$$v_4^{n+\frac{1}{2}} = v_4^n + \frac{\Delta t}{d_x}(F_2 - F_6).$$



G.2. À partir des valeurs actualisées à l'itération $n + \frac{1}{2}$ des variables conservatives, nous appliquons la théorie de gestion d'interface présentée dans les sous-sections 1.3.3 et 1.3.3. Il va de soi que les vecteurs v_2 et v_3 caractérisent respectivement les parties liquide et gazeuse du *condensat*, tandis que les vecteurs v_1 et v_4 caractérisent les deux cellules adjacentes à la cellule mixte qui permettent de former le *condensat*. Les flux F_3 et F_4 sont naturellement utilisés comme flux extrémale du *condensat*. Ainsi les étapes de formation du *condensat*, déplacement de l'interface, actualisation des variables conservatives, traitement du mouvement des mailles et projection sont toutes exécutées comme présenté dans les sous-sections 1.3.3 et 1.3.3. La seule différence consiste à remplacer les valeurs à l'itération n du cas $1D$ présenté dans ces sous-sections par les valeurs à l'itération $n + \frac{1}{2}$ que nous venons de calculer dans notre cas 2D.

H. Il reste à prendre en compte la gravité, en retranchant aux vitesses verticales de chaque cellule - y compris les cellules mixtes - l'amplitude de la gravité g multipliée par le pas de temps Δt. Les vecteurs V^{n+1} sont modifiés en conséquence.

I. Retour à A...

Remarque 1. Ce schéma conserve la masse. Nous avons noté cependant que dans certains cas les flux ne sont pas propres à leurs faces mais dépendent de la cellule adjacente considérée, de sorte que le flux entrant dans une cellule n'est pas le même que celui sortant de la cellule voisine. Mais dans ces cas les différences entre flux n'affectent que les variables du vecteur v correspondant à l'énergie cinétique, jamais celle correspondant à la masse. Nous notons en conséquence que l'énergie cinétique, elle, n'est pas conservée.

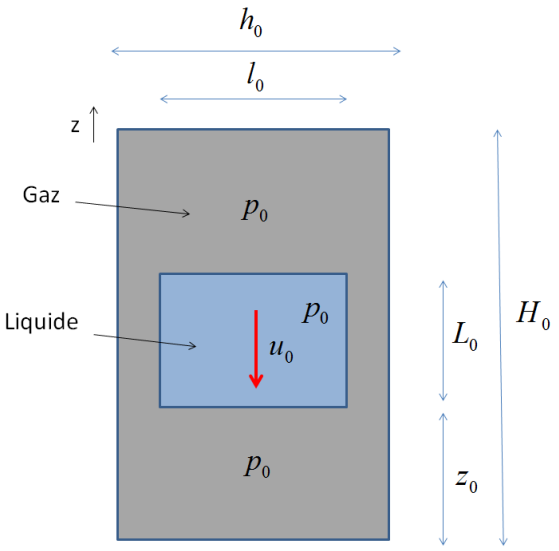
Remarque 2. Lorsqu'une face de cellule constitue l'interface entre liquide et gaz, comme cela doit toujours être le cas pour au moins deux tranches de liquide car le maillage suit les deux interfaces extrémale, nous considérons que nous sommes en présence d'une cellule mixte que nous situons arbitrairement sous l'interface. Ainsi, selon les cas, cette cellule mixte contient une quantité nulle de gaz ou de liquide, ce qui n'invalide aucune des étapes présentées ci-dessus.

Remarque 3. Il a été constaté certaines instabilités numériques engendrées par cette méthode, se caractérisant notamment par un passage des masses volumiques à des valeurs négatives à proximité des coins du bloc de liquide. En dehors du fait qu'une masse volumique négative constitue un phénomène non physique, elle mène à l'arrêt du calcul si la loi d'état utilisée ne permet pas d'en déduire une pression. La solution utilisée consiste à détecter chaque cellule dans laquelle ce phénomène intervient et à remplacer les masses volumiques de cette cellule et des huit cellules adjacentes par une moyenne des masses volumiques de ces neuf cellules, pondérée par les volumes des cellules. Ainsi nous assurons toujours la conservation de la masse et nous réglons le problème par quelques opérations de ce type dans la quasi-totalité des cas.

1.4.2 Résultats numériques

Formalisme

Nous notons p_0 la pression initiale, supposée uniforme. Nous en déduisons, à partir des lois d'état du gaz et du liquide, une densité initiale ρ_0^g dans le gaz et ρ_0^l dans le liquide. La vitesse initiale verticale du bloc de liquide est notée u_0. Afin de ne pas perturber les résultats par l'apparition d'éventuelles ondes de choc, la vitesse initiale est prise continue. Plus précisément, elle vaut uniformément u_0 entre les extrémités haute et basse du bloc de liquide, 0 aux frontières hautes et basses du domaine, et est complétée de manière affine dans les parties restantes. L'amplitude de la gravité est notée g. Par ailleurs, les longueurs L_0, z_0, H_0 et l_0 sont telles que représentées sur la figure suivante à l'instant initial. Les paramètres z_0 et L_0 sont variables en temps, et sont notés respectivement z et L en dehors de l'instant initial.



Lois d'état

Le gaz est considéré comme un gaz parfait de constante γ_g. Plus précisément, notant respectivement p et ρ les pression et masses volumique courantes, nous avons :

$$\frac{p}{p_0} = \left(\frac{\rho}{\rho_0}\right)^{\gamma_g}. \tag{1.44}$$

Le liquide suit une loi *stiffened gas* de constantes γ_l et χ_l, à savoir :

$$\frac{p}{p_0} = 1 + \frac{\left(\frac{\rho}{\rho_0}\right)^{\gamma_l} - 1}{\gamma_l \ \chi_l}. \tag{1.45}$$

Les valeurs numériques considérées sont les suivantes :

$$\gamma_g = 1.4, \ \gamma_l = \frac{1}{22500}, \text{ et } \chi_l = 7.$$

Modèle de Bagnold

Le modèle de Bagnold est un modèle analytique simplifié pour traiter le cas sans fuite, c'est à dire pour lequel $h_0 = l_0$. Il considère que la pression est uniforme dans chacune des deux poches de gaz, et que le liquide est incompressible. Sa version originale, présentée dans [5], est légèrement complétée ici pour prendre en compte le terme de gravité.

Nous appelons respectivement ρ_{haut} et ρ_{bas} les masses volumiques des poches de gaz haute et basse. La conservation de la masse dans chaque poche de gaz s'écrit alors :

$$\rho_{haut} \ (H_0 - L - z) = \rho_0^g \ (H_0 - L_0 - z_0). \tag{1.46}$$

$$\rho_{bas} \ z = \rho_0^g \ z_0, \tag{1.47}$$

et la loi de Newton appliquée au bloc de liquide s'écrit :

$$\rho_0^l \ L_0 \ \ddot{z} = -\rho_0^l \ L_0 \ g + p_0 \left(\left(\frac{\rho_{bas}}{\rho_0^g} \right)^{\gamma_g} - \left(\frac{\rho_{haut}}{\rho_0^g} \right)^{\gamma_g} \right), \tag{1.48}$$

ou encore, en utilisant les deux résultats précédents :

$$\rho_0^l \ L_0 \ \ddot{z} = -\rho_0^l \ L_0 \ g + p_0 \left(\left(\frac{z_0}{z} \right)^{\gamma_g} - \left(\frac{H_0 - L_0 - z_0}{H_0 - L - z} \right)^{\gamma_g} \right), \tag{1.49}$$

les conditions initiales s'écrivant :

$$z(0) = z_0 \text{ et } \dot{z}(0) = -u_0.$$

Notons maintenant :
$a = \frac{z_0}{H_0 - L_0 - z_0}$, $S_b = \frac{\rho_l^0 \ L_0 \ u_0^2}{p_0 \ z_0}$, $S_g = \frac{\rho_l^0 \ L_0 \ g}{p_0}$, $\tau = \sqrt{\frac{\rho_0^l \ L_0 \ z_0}{p_0}}$,
et ξ fonction du temps t telle que $z(t) = z_0 \ \xi\left(\frac{t}{\tau}\right)$.
Nous pouvons réécrire l'équation (1.49) sous forme adimensionnée :

$$\ddot{\xi} = -S_g + \left(\frac{1}{\xi} \right)^{\gamma_g} - \left(\frac{1}{1 + a \ (1 - \xi)} \right)^{\gamma_g}, \tag{1.50}$$

avec pour conditions initiales $\xi(0) = 1$ et $\dot{\xi}(0) = -S_b^{\frac{1}{2}}$. Notant respectivement p_{haut} et p_{bas} les pressions de gaz dans les poches haute et basse, il est clair que leurs équivalents adimensionnés se déduisent de la résolution de ξ :

$$\frac{p_{haut}}{p_0} = \left(\frac{1}{1 + a \ (1 - \xi)} \right)^{\gamma_g}, \tag{1.51}$$

$$\frac{p_{bas}}{p_0} = \left(\frac{1}{\xi} \right)^{\gamma_g}. \tag{1.52}$$

Il est intéressant de noter que la résolution du problème adimensionné ne dépend que de la constante des gaz parfaits γ_g et des nombres a, S_b et S_g. En particulier les nombres S_g et S_b nous serviront de référence pour désigner les importances respectives de la gravité et de la vitesse initiale dans l'impact du bloc de liquide. Nous introduisons également le nombre de Froude :

$$F_r = \frac{u_0}{\sqrt{g \ z_0}}$$

et le nombre d'impact global :

$$S = \frac{1}{2} \ S_b + S_g,$$

de sorte que nous avons :

$$S_b = \frac{2 \ F_r^2}{2 + F_r^2} \ S, \tag{1.53}$$

$$S_g = \frac{2}{2 + F_r^2} \ S. \tag{1.54}$$

L'étude analytique de l'équation (1.50), effectuée dans [23], ainsi que les expressions (1.51) et (1.52), mènent à la formule suivante reliant le nombre d'impact S et le ratio p^* de la pression maximale de gaz dans la poche du bas sur la pression initiale de gaz :

$$S = \frac{\frac{p^*}{\gamma_g - 1} + 1 - \frac{\gamma_g \ (p^*)^{\frac{1}{\gamma_g}}}{\gamma_g - 1}}{(p^*)^{\frac{1}{\gamma_g}} - \frac{2}{2 + Fr^2}}. \tag{1.55}$$

Cette expression permet de tracer la courbe $p^*(S)$. Il est intéressant de noter que cette courbe ne dépend que de la constante des gaz parfaits γ_g, du nombre de Froude F_r et du nombre d'impact S. Par ailleurs le cas présenté ici, monodimensionnel, peut être entièrement résolu à partir de la section 1.3.3, ce qui permet la comparaison des deux modèles.

Comparaison des modèles

Envisageons donc le cas pour lequel la largeur du bloc de liquide vaut la largeur du domaine, pour comparer le modèle de Bagnold au modèle *Flux de Pan* de la section 1.3.3. Comme nous l'avons vu, en vertu de l'équation (1.50), le modèle de Bagnold adimensionné est entièrement paramétré par la donnée des nombres S_b, S_g, a et de la constante des gaz parfaits γ_g. Grâce aux formules (1.53) et (1.54), la donnée de S_b et S_g peut être remplacée par celle des nombres de Froude F_r et d'impact S. En ce qui concerne le modèle *Flux de Pan*, nous fixons arbitrairement :

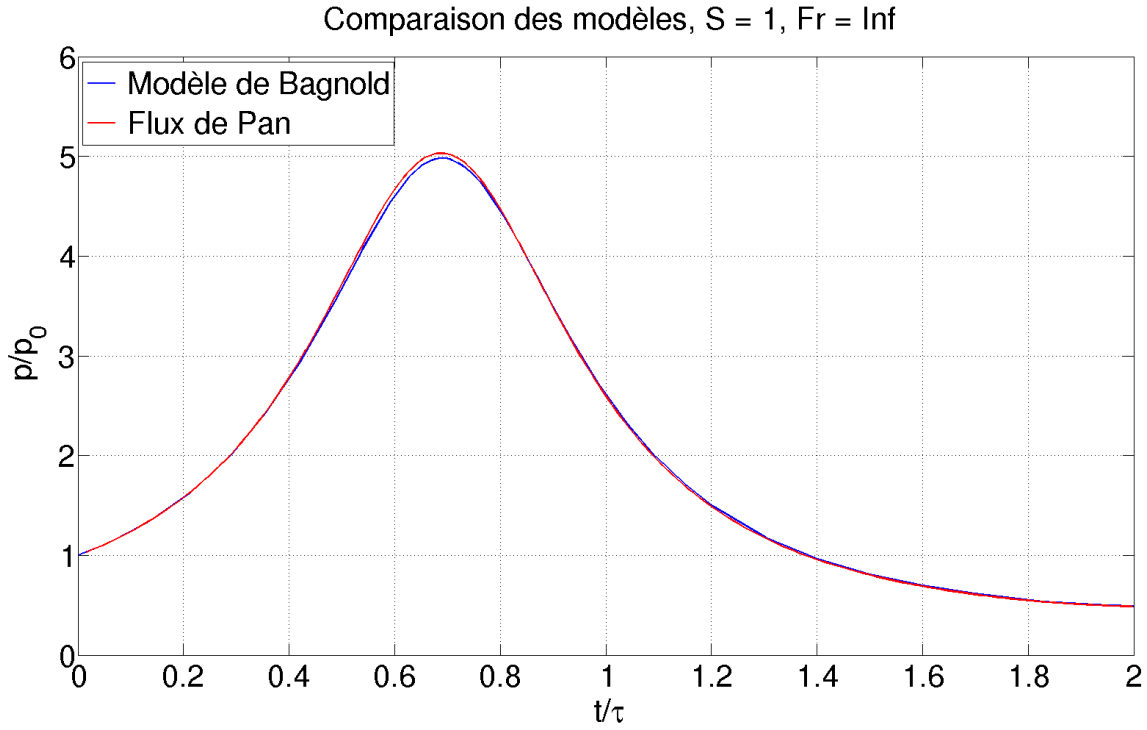
$$z_0 = l_0 = L_0 = h_0 = 1.$$

La valeur 1 considérée ici n'a pas d'importance, car elle fixe l'unité de longueur. Seule l'égalité entre ces quatre paramètres compte. Par ailleurs nous prenons $p_0 = 1$, pour assimiler la pression à son équivalent adimensionné du modèle de Bagnold. Les masses volumiques initiales sont celles des phases liquides et gazeuses de l'eau dans les conditions standards de température et de pression, à savoir :

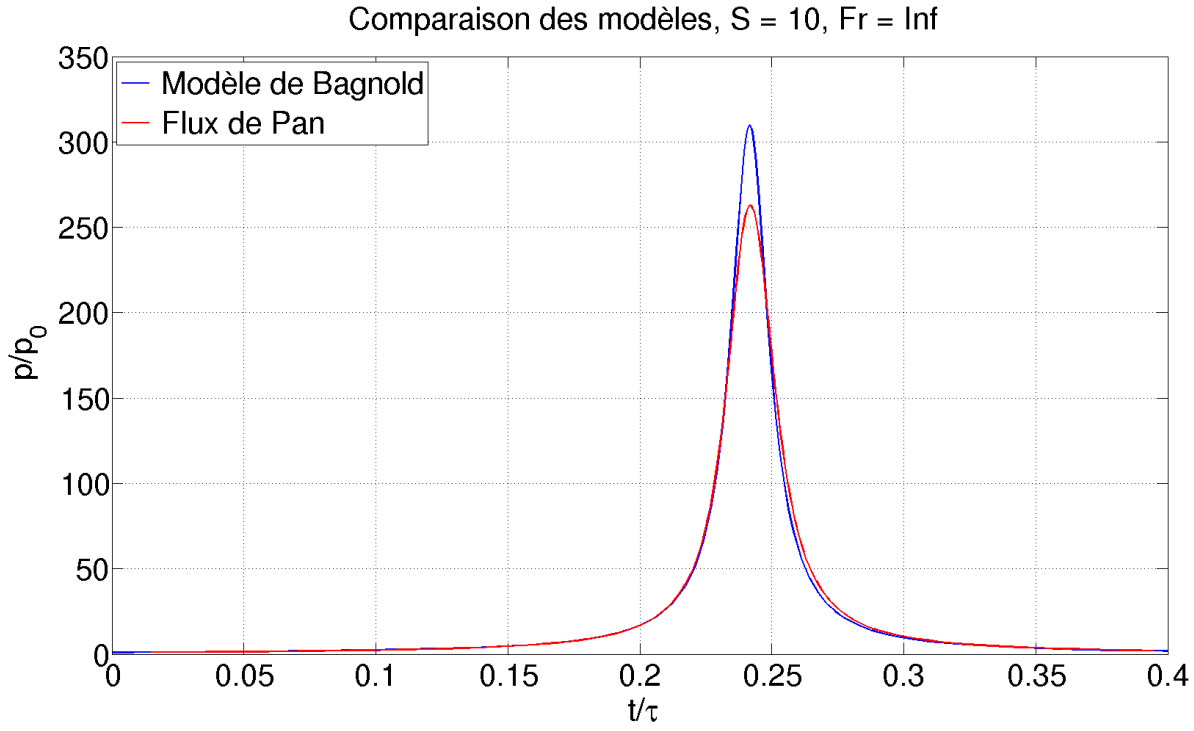
$$\rho_0^l = 1000 \text{ et } \rho_0^g = 1.2.$$

Nous notons ici que seul le ratio de ces deux valeurs a de l'importance, car ce sont les seules données contenant une unité de masse. Les données u_0 et g sont directement issues des nombres de Froude et d'impact grâce aux formules (1.53) et (1.54). La valeur de H_0 dépend directement de a. Sous ces hypothèses, les lois d'état étant fixées par ailleurs (cf. sous-section 1.4.2), les deux modèles sont régis par l'ensemble des trois nombres F_r, S et a. Nous fixerons dans la suite $a = 1$ pour nous intéresser plus spécifiquement au nombre de Froude et au nombre d'impact.

Evolution temporelle. Considérons ici un nombre de Froude infini, qui correspond à une absence de gravité, et comparons l'évolution de la pression dans la poche de gaz du dessous pour nos deux modèles, à savoir celui de Bagnold et *Flux de Pan*. En ce qui concerne le modèle *Flux de Pan*, pour lequel la pression n'est pas uniforme au sein d'une poche de gaz, nous considérons la pression maximale sur la rangée de cellules adjacente au mur du bas. Voici l'évolution temporelle de pression adimensionnée pour ces deux modèles avec un nombre d'impact pris égal à 1 :

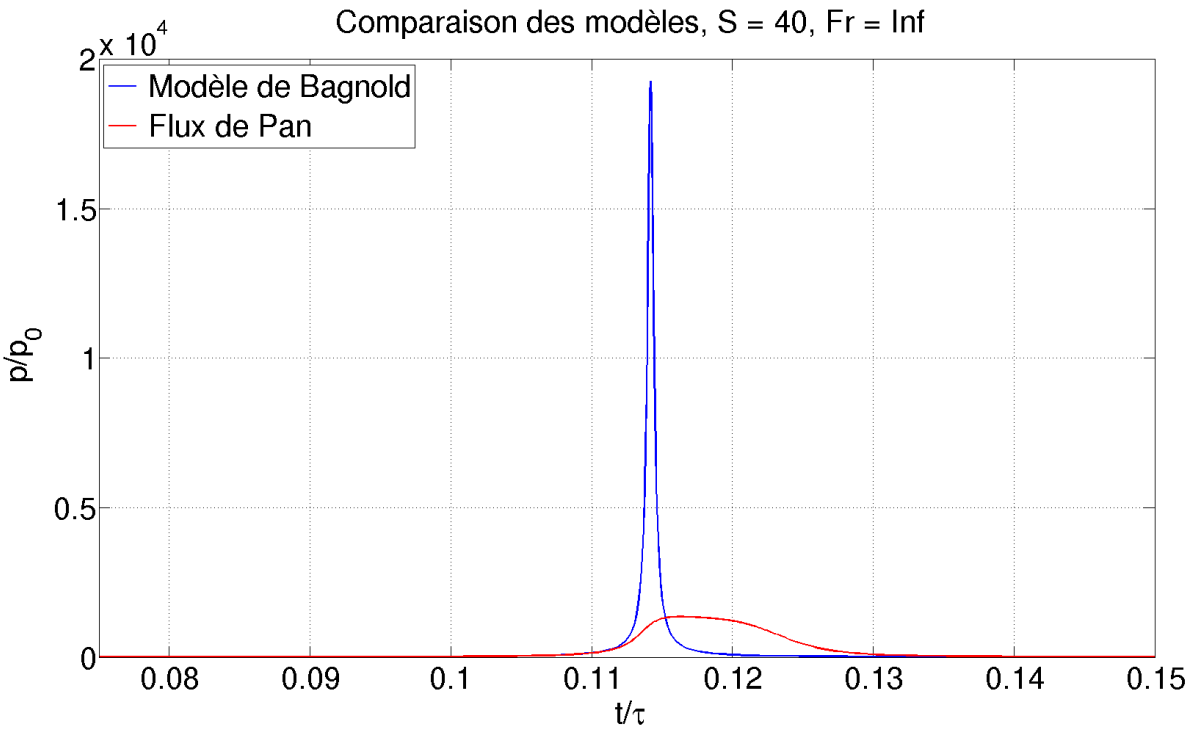


Cette excellente concordance prouve la validité du modèle de Bagnold pour les faibles nombres d'impact. En revanche, lorsque le nombre d'impact augmente, nous constatons l'apparition de différences plus importantes, par exemple ci-dessous pour $S = 10$:



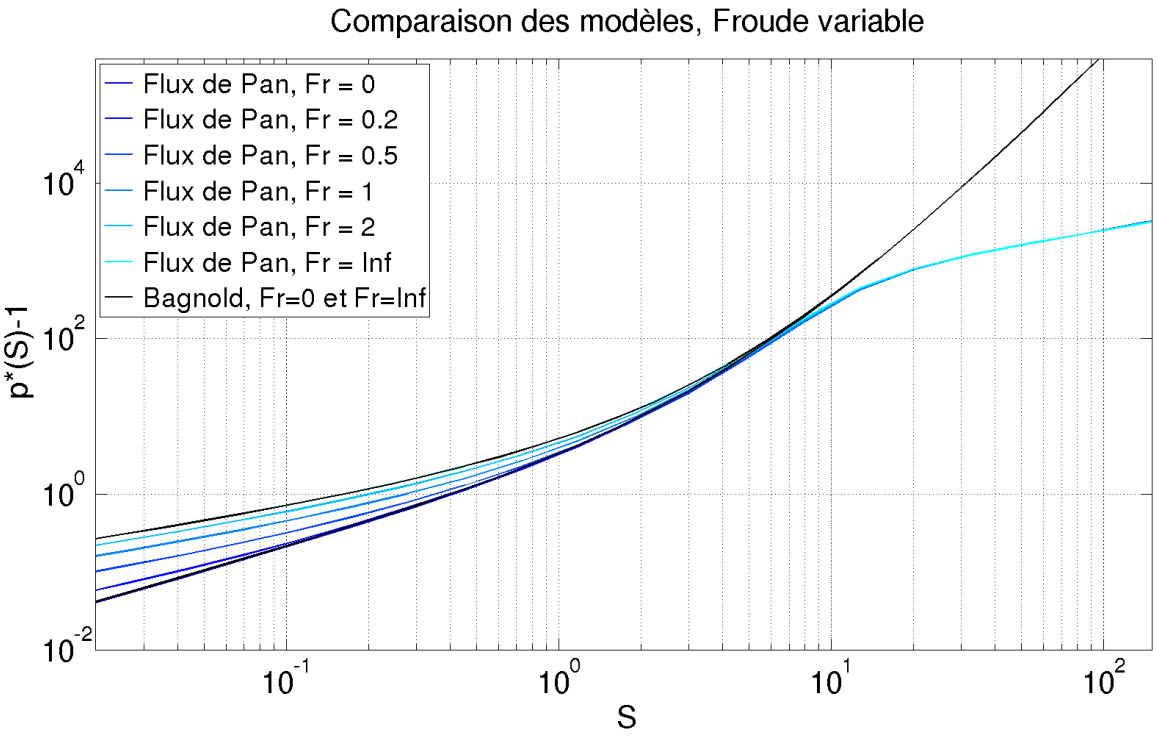
Pour des nombres d'impact encore plus importants, par exemple $S = 40$ ci-dessous,

les courbes sont très différentes, avec en particulier une pression maximale beaucoup plus faible pour le modèle *Flux de Pan* :



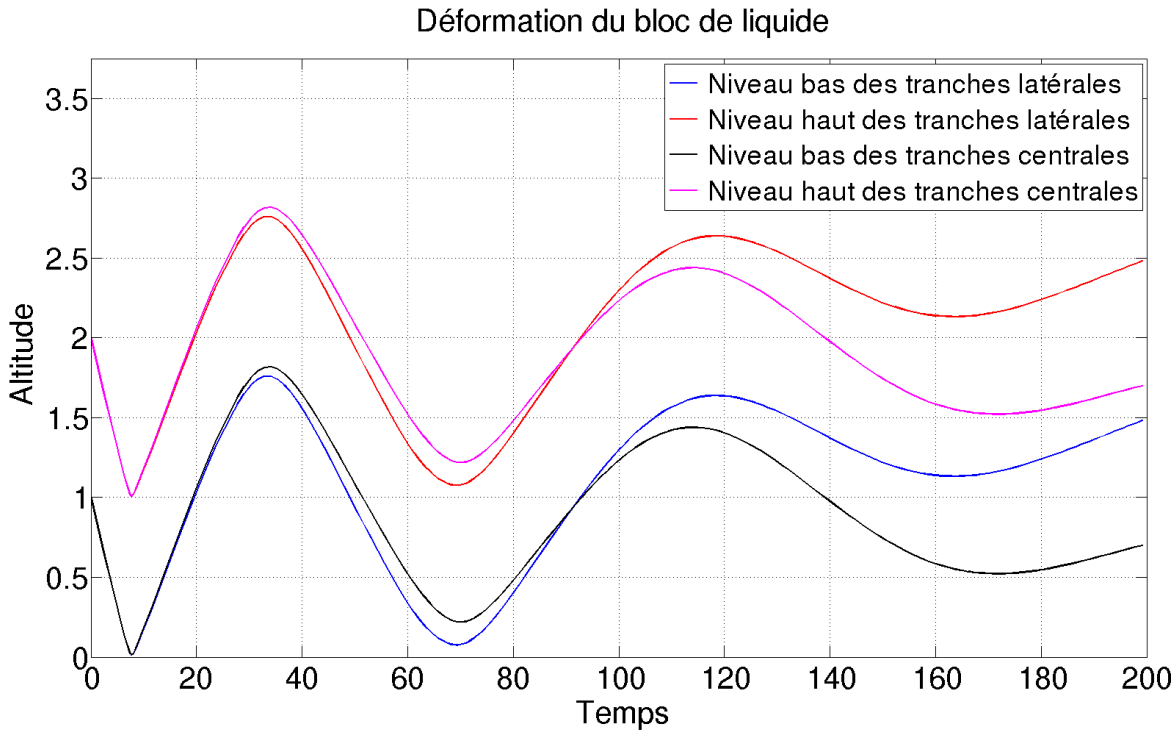
Cela s'explique par le fait que le modèle de Bagnold ne prend pas en compte la compressibilité du liquide, qui joue un rôle important lors d'un violent impact. En effet, dans ces conditions, tout se passe comme si le liquide était projeté contre le mur en l'absence de gaz, de sorte que le choc est absorbé par la compression du liquide et non plus par celle du gaz. Le modèle de Bagnold ne peut pas capturer ce phénomène et donne un résultat erroné.

Evolution de la surpression maximale en fonction de S. Nous allons maintenant comparer les deux modèles sur des courbes en unités logarithmiques donnant l'évolution de la surpression maximale en fonction du nombre d'impact. Nous rappelons que p^* désigne la valeur maximale du ratio de la pression sur la pression initiale. Nous rappelons également qu'un nombre de Froude nul signifie une chute du bloc de liquide sans vitesse initiale mais sous l'effet de la gravité, tandis qu'un nombre de Froude infini signifie une vitesse initiale en l'absence de gravité. Comme attendu au vu des résultats précédents montrant l'évolution temporelle de la pression, le modèle de Bagnold donne les bonnes valeurs de surpression maximale pour les faibles nombres d'impacts, mais donne des surpressions trop élevées pour les nombres d'impact importants :

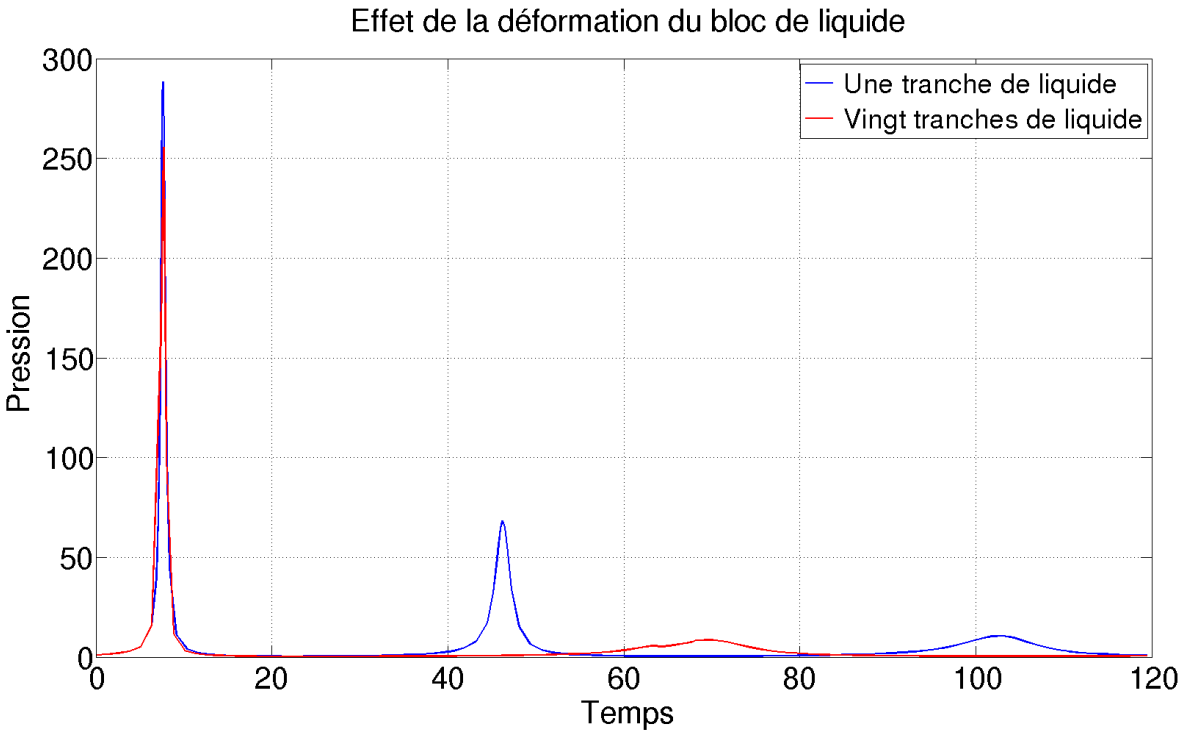


Cas avec fuite

Étude quantitative pour les cas avec fuite. Les cas où la largeur du domaine est supérieure à la largeur du bloc de liquide (i.e. $h_0 > l_0$) sont bidimensionnels, et nous pouvons nous attendre à un comportement hétérogène des tranches de liquide menant à une déformation du bloc. Nous nous proposons dans ce paragraphe d'étudier sur un cas standard cette déformation. Nous considérons le même cas que précédemment (cf. sous-section 3.4.2), pour $S = 10$ avec un nombre de Froude infini (i.e. sans gravité), à la différence près que nous imposons cette fois $h_0 = 1.01\ l_0$. La fuite est choisie suffisamment faible pour maintenir des rebonds du bloc de liquide et étudier la déformation qu'ils induisent. Par ailleurs il est clair que dans ce cas à deux dimensions, le nombre de tranches de liquide choisi influe sur le résultat. La figure ci-dessous montre l'évolution temporelle des altitudes hautes et basses des tranches de liquide centrales (i.e. celles qui sont le plus proches du milieu du domaine. Il peut y en avoir une ou deux selon que le nombre de tranches est pair ou impair) et latérales (i.e. celles qui sont le plus proches des bords gauche et droit du bloc de liquide. La symétrie du problème nous permet de ne considérer qu'un des deux côtés). Nous avons choisi ici de découper le bloc de liquide en 20 tranches.

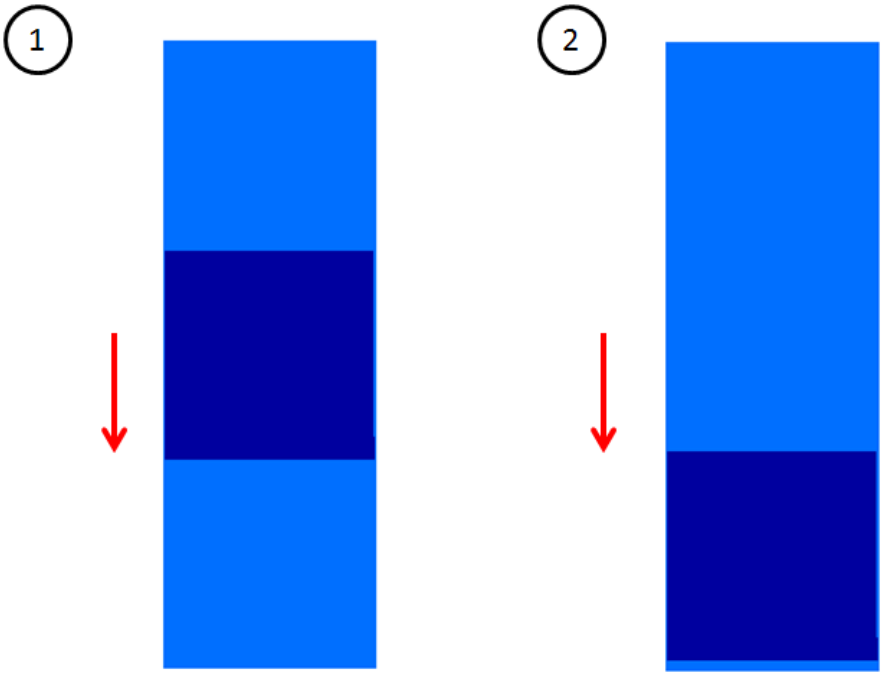


Nous remarquons que le bloc de liquide se déforme peu dans la première phase de descente, mais que par la suite il en va tout autrement. Dans cette première phase de descente, il est toutefois intéressant de signaler - ce que nous ne distinguons pas sur la figure - que ce sont les tranches latérales qui tombent le plus vite. Nous souhaitons maintenant connaître l'influence de cette déformation du bloc de liquide sur l'évolution de la pression dans la poche de gaz inférieure. Pour cela, nous avons réalisé ce même cas physique en considérant cette fois une simulation avec 20 tranches de liquide (comme précédemment) et une autre avec une seule tranche de liquide, ce dernier cas interdisant naturellement toute déformation du bloc. Les résultats concernant le profil de pression sont les suivants :

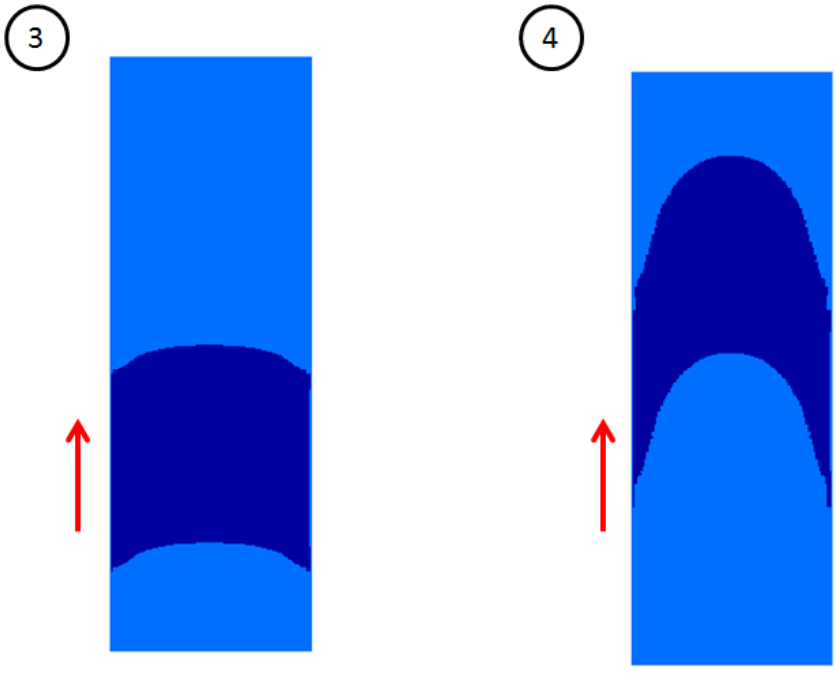


Nous remarquons deux phénomènes : tout d'abord, la très légère déformation du bloc de liquide dans la première phase de descente, que nous évoquions précédemment, induit une baisse de la pression maximale au premier impact. Par la suite, les déformations plus importantes atténuent fortement l'influence des autres impacts sur la pression.

Étude qualitative pour les cas avec fuite. Nous proposons ici un dernier cas ayant pour seul but de présenter visuellement le phénomène de rebond d'un bloc de liquide. Un grand nombre de tranches (100) a été considéré, et notre cas standard a été réalisé, pour un nombre d'impact égal à 10 et une largeur de fuite prise telle que $h_0 = 1.01\ l_0$. Les deux premières étapes présentées ci-dessous concernent (dans l'ordre chronologique) la phase de descente, pour laquelle, nous l'avons vu, la déformation du bloc de liquide est très faible :



Les deux étapes suivantes (toujours dans l'ordre chronologique) sont postérieures au rebond. Nous pouvons voir que le bloc de liquide s'incurve de façon significative :



1.5 Conclusion et perspectives

Cette partie avait pour contexte la simulation efficace de la chute d'un bloc d'eau entouré de gaz, phénomène qui intéresse en particulier les transporteurs de méthane

liquide. Deux familles de modèles existent pour traiter ce problème.

- D'une part, ceux permettant une simulation numérique complète et directe de la physique en jeu, se caractérisant par une grande précision et la nécessité d'importants moyens numériques. Le modèle bifluide à quatre équations présenté dans la sous-section 2.10.3 n'est pas un modèle pertinent pour traiter ce problème, car il considère une fraction volumique de gaz en tout point du domaine, tandis que nous souhaitons obtenir une interface nette séparant le gaz du liquide. Le moyen d'obtenir à la fois une simulation numérique complète et la séparation des zones entièrement gazeuse et entièrement liquide a été mis en place dans [12].
- D'autre part, les modèles visant l'optimisation des moyens numériques nécessaires et la limitation des phénomènes physiques à considérer, dans le double objectif de gagner en temps de calcul et de comprendre mieux ces derniers phénomènes. Ce type de modèle est notamment introduit dans [10].

Notre travail s'inscrit dans la deuxième famille de modèles, tout en visant à se rapprocher de la première. Nous nous sommes en effet inspirés des travaux présentés dans [10] et [12] pour créer un modèle efficace qui prenne en compte - ce qui n'était pas le cas dans la famille des modèles simples - la déformation du bloc d'eau. Des difficultés numériques ont dû être traitées, en particulier du fait de la nécessaire mise en place d'un mouvement de maillage pour représenter correctement la zone située en dessous du bloc de liquide, amenée à être fortement *compressée* lors des violents impacts. Le *couplage* de la gestion généralisée des interfaces gaz - liquide, présentée dans [12], et de la prise en compte d'un maillage mobile, proposée dans [10], est une nouveauté.

Ce modèle a pu être utilisé avec succès pour créer des courbes donnant la surpression maximale à la paroi inférieure en fonction de la violence de l'impact. Ces courbes, publiées dans [4], ont pour finalité la déduction des surpressions réelles à partir des surpressions mesurées sur modèle réduit, par l'intermédiaire du *Froude scaling*. On remarque que la grande quantité de simulations nécessaires pour l'obtention de telles courbes rend impossible l'utilisation de modèles fins et complets, tels celui présenté dans [12], pour cette application.

Finalement, nous nous sommes intéressés à l'étude qualitative de la déformation du bloc d'eau lors de sa chute, avec un résultat d'intérêt : en tombant, le bloc d'eau se déforme de façon à emprisonner le gaz plutôt qu'à le faire échapper. Nous ne pouvons pas prolonger la simulation au delà de l'impact du liquide sur la paroi, le modèle actuel ne permettant pas de le considérer. Toutefois, en cas de chute suffisamment violente, le liquide *rebondit*, c'est à dire qu'il remonte sans avoir touché la paroi. Nous constatons dans ce cas que lorsque la déformation du bloc de liquide est prise en compte - et donc qu'une poche de gaz est emprisonnée -, la surpression maximale est moins importante. Pour dépasser cette approche qualitative et quantifier finement ces phénomènes, il sera intéressant d'ajouter à ce modèle une prise en compte des phénomènes résultant d'un impact direct entre le liquide et la paroi.

Chapitre 2

AMR

2.1 Introduction

De façon générale, en simulation numérique, le résultat obtenu est d'autant plus précis que la description géométrique du domaine est fine, mais le temps de calcul et la mémoire utilisée croissent - souvent de façon linéaire - avec le nombre de cellules représentant la géométrie. Il y a donc un compromis à trouver entre efficacité et précision. Cependant il est possible, à moyens numériques donnés, de concentrer les *efforts* à proximité des zones les plus sensibles à la finesse de représentation, en y augmentant la *densité* du maillage. La difficulté en général rencontrée à ce stade concerne la prédiction des zones d'intérêt particulier où un raffinement local sera *rentable*, étant donné que ces zones sont naturellement amenées à évoluer en fonction de la solution du système physique. Une réponse possible est la mise en œuvre d'un raffinement localisé aux endroits où la solution est abrupte, de façon à calculer plus correctement les brusques variations spatiales que l'on rencontre par exemple dans le cas d'une onde de choc. Nous ne nous intéressons ici qu'aux problèmes en une dimension d'espace, et nous verrons que le nombre de quantités résolues a peu d'importance et s'adapte bien à toutes les méthodes que nous envisagerons. Nous nous basons sur la méthode VFFC (cf. sous-section 1.3.1), que nous adapterons aux modèles non conservatifs et aux maillages mobiles par des moyens que nous décrirons en détails. L'application visée est, à terme, l'utilisation d'un schéma AMR non conservatif robuste et fiable pour la simulation des systèmes diphasiques à trois espèces, et ce pour étudier notamment la propagation des ondes de détonation dans les mousses. À ce stade, nous avons pu obtenir des résultats convaincants sur un modèle non conservatif à deux espèces, ce qui est une nouveauté.

Deux difficultés se présentent dans la mise en place d'une telle méthode. D'une part il faut déterminer le mouvement de maillage, qui doit être par principe adapté à la solution physique et qui est de plus contraint par des questions de stabilité du schéma numérique. D'autre part, ce mouvement étant donné, il faut qu'il soit correctement pris en compte dans la résolution du système physique. L'approche que nous utilisons dans toutes les méthodes présentées ici est un découplage entre le calcul du maillage et celui de la solution. Autrement dit, le maillage servant au passage du temps n au temps $n+1$ n'est déterminé qu'à partir de la solution au temps n. D'autres méthodes consistent à résoudre un système d'équations global ayant pour inconnues la solution physique aussi bien que les coordonnées du maillage, mais la complexité des systèmes

physiques traités ici nous interdit de les envisager en première intention.

2.2 La fonction de contrôle et la notion d'équirépartition

2.2.1 Fonction de contrôle

Nous cherchons une solution $f(x,t)$ d'une équation aux dérivées partielles, où t et x sont les variables d'espace et de temps, à une dimension chacune, tandis que leur image $f(x,t)$ est un vecteur à m éléments $f_1(x,t)$, $f_2(x,t)$,..., $f_m(x,t)$. Nous commençons tout d'abord par définir une fonction $\rho(x,t)$ de $\mathbb{R}^2$ dans $\mathbb{R}$ qui soit un bon indicateur de l'amplitude des variations de f au point (x,t). Plusieurs choix sont possibles, dans ce qui suit nous prendrons suivant [7] :

$$\rho = \sqrt{\left(1 + \sum_{i=1}^{m} \frac{1}{\alpha_i}\Big(\frac{\partial f_i}{\partial x}\Big)^2\right)}, \tag{2.1}$$

avec $(\alpha_i \ , \ 1 \leq i \leq m)$ un vecteur de $\mathbb{R}^m$ servant à pondérer les importances des différentes variables résolues. On note dans la définition de ρ que la valeur de cette fonction est d'autant plus importante que les variations de f sont brusques, ce qui correspond au but recherché. On note également que $\rho(x,t)$ est toujours plus grand que 1, ce dont nous expliquons l'utilité dans le paragraphe suivant.

2.2.2 Equirépartition

Nous avons maintenant à notre disposition une fonction ρ de $\mathbb{R} \times \mathbb{R}^+$ dans $\mathbb{R}$, appelée fonction de contrôle, qui peut être calculée à chaque pas de temps à partir de la donnée de f et d'une approximation de ses gradients. Nous souhaitons alors calculer un maillage qui soit au mieux adapté à la solution f, et qui soit donc d'autant plus fin que la fonction ρ est grande. Ce critère peut être précisé : nous allons chercher un maillage dit *équiréparti*, composé des points $(x_1, x_2, ..., x_N)$ et qui vérifie :

$$\int_{x_1}^{x_2} \rho(x,t)dx = \int_{x_2}^{x_3} \rho(x,t)dx = ... = \int_{x_{N-1}}^{x_N} \rho(x,t)dx. \tag{2.2}$$

Ainsi la taille des mailles est inversement proportionnelle à la valeur de ρ. Nous voyons ici l'intérêt d'avoir une fonction ρ qui soit minorée par un nombre strictement positif, condition qui permet de donner une taille maximale aux mailles.

2.3 Lissage de la fonction de contrôle

2.3.1 Principe

Plusieurs méthodes de calcul exact ou approché du maillage équidistribué sont présentées dans ce document. Pour la plupart d'entre elles, il a été observé l'apparition de légères oscillations dans le mouvement du maillage, ayant pour conséquence des instabilités dans le calcul de la solution physique. Ce problème a pu être contourné en opérant un lissage de la fonction de contrôle préalable au calcul du nouveau maillage. Nous choisissons de remplacer la fonction ρ définie dans la section 2.2 sur $[a,b]$ par la

fonction ρ^*, définie comme le produit de convolution de ρ par une fonction g ($\mathbb{R} \to \mathbb{R}$) choisie par l'utilisateur :

$$\rho^*(x) = (\rho * g) = \int_{-\infty}^{+\infty} \rho(y)\ g(x-y)\ dy.$$

Pour calculer le produit de convolution, nous avons besoin de valeurs de ρ en dehors du domaine $[a, b]$. Nous considérons alors que pour $x < a$, $\rho(x) = \rho(a)$ et pour $x > b$, $\rho(x) = \rho(b)$. Par ailleurs le domaine d'intégration peut être restreint aux lieux où les valeurs de g sont significatives.

2.3.2 Choix de la fonction de lissage :

Divers essais ont montré que la manière la plus efficace de lisser la fonction de contrôle était la convolution par une gaussienne. Plus précisément, en notant L une longueur caractéristique choisie par l'utilisateur, nous définissons g par :

$$g(x) = \frac{1}{L\sqrt{2\pi}}\ e^{\left(-\frac{x^2}{2L^2}\right)}.$$

Cependant, le calcul du produit de convolution de ρ par une gaussienne mène à un nombre d'opérations proportionnel au carré du nombre de cellules, ce qui est rédhibitoire car toutes les autres étapes de résolution (du maillage comme de la solution physique) présentent un nombre d'opérations proportionnel au nombre de cellules. Une solution possible est d'utiliser pour g une fonction créneau :

$$x > \frac{L}{2},\ g(x) = 0.$$

$$x < -\frac{L}{2},\ g(x) = 0.$$

$$-\frac{L}{2} < x < \frac{L}{2},\ g(x) = \frac{1}{L}.$$

Cette formulation, bien que donnant une fonction ρ_* moins lisse et donc plus sujette aux oscillations de maillage, a l'avantage de permettre un calcul du produit de convolution en temps linéaire. En effet, on peut en optimiser la programmation en utilisant pour le calcul de $\rho^*(x_{i+1})$ le résultat du calcul de $\rho^*(x_i)$, seuls les termes extrémaux du domaine d'intégration changeant de l'un à l'autre.

Un bon compromis entre les deux dernières approches consiste à prendre pour g une fonction triangle :

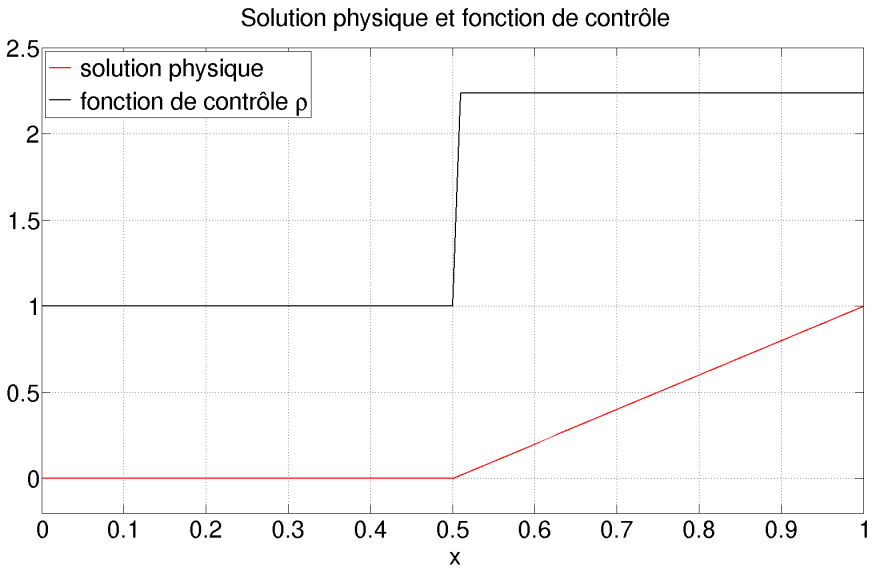
$$g(x) = \frac{1}{L}\ max\Big(0, min\big(1 + \frac{x}{L}, 1 - \frac{x}{L}\big)\Big).$$

Ce lissage est de meilleure qualité que celui par un créneau, et le produit de convolution peut être calculé en temps linéaire en remarquant que la convolution de ρ par une fonction *triangle* revient à convoluer ρ par une fonction *créneau* deux fois de suite.

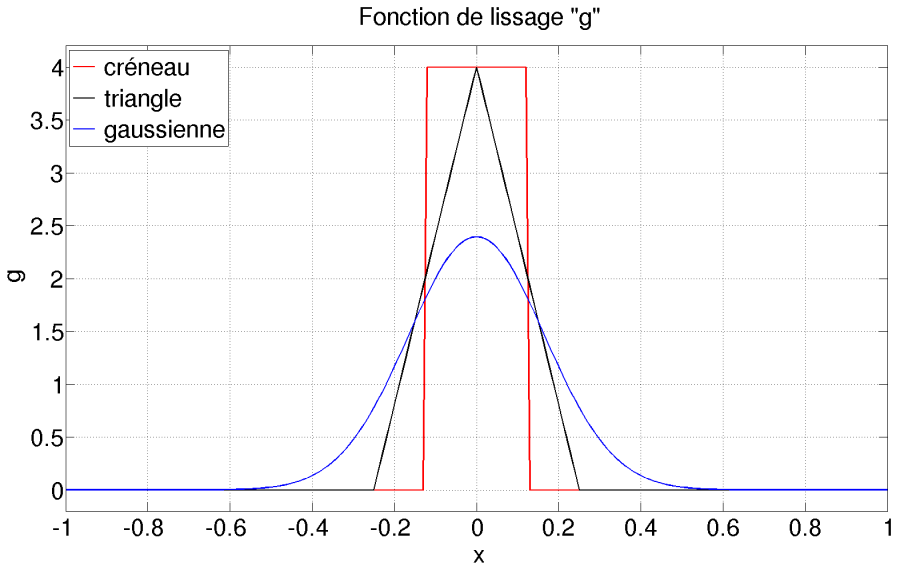
Le paramètre L, dans les expressions précédentes, est la taille caractéristique du lissage, et doit être pris du même ordre de grandeur que la taille du domaine (un quart du domaine convient bien). Il rendra la fonction ρ^* d'autant plus lisse qu'il est grand. Désormais, lorsque nous évoquerons la fonction ρ, nous sous-entendrons qu'elle a été lissée de cette manière et donc remplacée par ρ^*.

2.3.3 Exemple

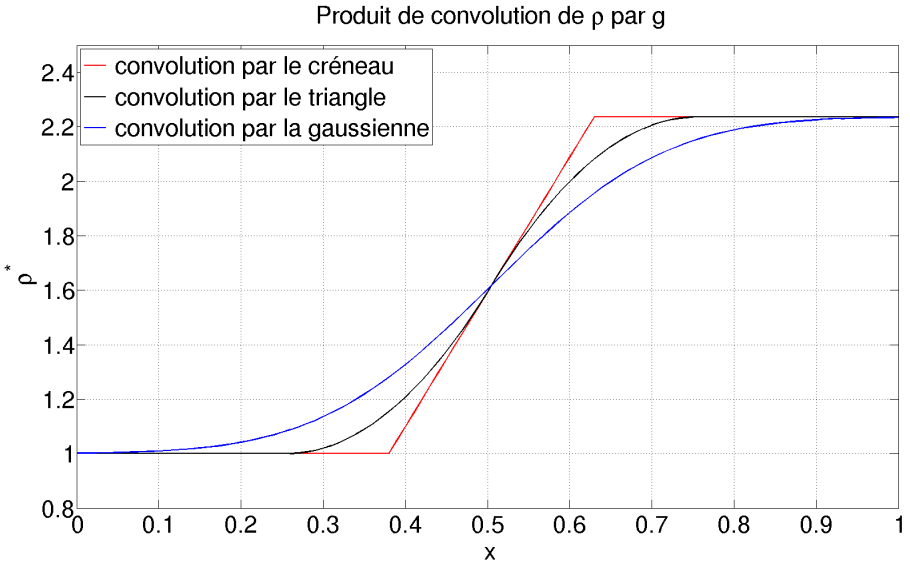
Imaginons que nous ayons obtenu une solution physique nulle sur la première partie du domaine et présentant une évolution linéaire sur la deuxième partie, alors la fonction de contrôle est la suivante (en utilisant la formule (2.1) avec $\alpha = 1$) :



Nous utilisons trois fonctions de lissage correspondant aux trois types possibles vus précédemment :



Les fonctions lissées correspondantes, une fois le produit de convolution effectué, ont la forme suivante :



Nous constatons qu'avec un lissage type *créneau*, la fonction résultante présente encore un saut dans la dérivée, ce qui peut occasionner des difficultés dans son utilisation future. Le lissage type *triangle*, en revanche, semble bien adapté. C'est celui que nous utiliserons dans la suite.

2.4 Calcul direct du maillage équidistribué

Après avoir obtenu au temps t^n, par la résolution de l'équation différentielle physique, une solution f_i^n aux nœuds $(x_i,\ 1 \leq i \leq N)$ du maillage actuel noté M_n, nous en déduisons par un calcul approximatif des gradients, au premier ou second ordre, des valeurs $(\rho_i^n,\ 1 \leq i \leq N)$ correspondant à la discrétisation de $x \to \rho(x, t^n)$ sur M_n. Dans cette section, nous exposons une méthode permettant d'approcher le maillage équidistribué correspondant. Tout commence par la réalisation d'une interpolation des valeurs connues de ρ pour obtenir une fonction $x \to \rho_{interp}^n(x)$ prenant ses valeurs sur tout l'intervalle $[x_1\ x_N]$. Nous considérons ici une interpolation linéaire, bon compromis entre simplicité et précision. La première section présente une approche simple visant à calculer directement un maillage équidistribué à partir de cette interpolation. La deuxième section expose une approche plus subtile visant une transformation de ρ_{interp}^n en une fonction affine sur les segments du maillage objectif M_{n+1}. Elle se base sur l'exposé de la première section.

2.4.1 Méthode simple

Imaginons donc que nous ayons choisi une interpolation de ρ_i^n telle que ρ_{interp}^n soit affine par morceaux sur $[x_1\ x_N]$, et plus exactement affine sur les segments $[x_i\ x_{i+1}]$ du maillage M^n. L'égalité (2.25) :

$$\int_{x_1}^{x_2} \rho_{interp}^n = \int_{x_2}^{x_3} \rho_{interp}^n = \ldots = \int_{x_{N-1}}^{x_N} \rho_{interp}^n,$$

associée à l'égalité :

$$\int_{x_1}^{x_2} \rho_{interp}^n + \int_{x_2}^{x_3} \rho_{interp}^n + ... + \int_{x_{N-1}}^{x_N} \rho_{interp}^n = \int_{x_1}^{x_N} \rho_{interp}^n,$$

nous assure que :

$$\int_{x_1}^{x_2} \rho_{interp}^n = \int_{x_2}^{x_3} \rho_{interp}^n = ... = \int_{x_{N-1}}^{x_N} \rho_{interp}^n = \frac{1}{N} \int_{x_1}^{x_N} \rho_{interp}^n.$$

Les abscisses x_1 et x_N sont connues car il s'agit des extrémités du domaine, qui est supposée inchangé au cours du temps. De plus, ρ_{interp}^n étant par hypothèse affine par morceaux, $\frac{1}{N} \int_{x_1}^{x_N} \rho_{interp}^n$ se calcule aisément. Puis, grâce à cette même hypothèse, nous déduisons facilement de l'égalité $\int_{x_1}^{x_2} \rho_{interp}^n = \frac{1}{N} \int_{x_1}^{x_N} \rho_{interp}^n$ l'abscisse x_2. Connaissant x_2 nous calculons x_3 de la même manière et, de proche en proche, nous déterminons l'intégralité du maillage équidistribué.

Cette méthode fonctionne, mais le choix d'une fonction ρ_{interp}^n figée, et dont la dérivée par rapport à x est discontinue aux nœuds du maillage M_n, rend trop dépendant le maillage M_{n+1} du maillage M_n, de sorte que des instabilités se créent et entraînent un mouvement très irrégulier du maillage au cours des itérations. Nous reviendrons dans la suite sur la nécessité d'adoucir le mouvement du maillage et sur les moyens d'améliorer ce point. Contentons nous ici d'adopter une nouvelle approche qui limite - sans pour autant échapper véritablement au problème - ce type d'instabilités.

2.4.2 Algorithme de De Boor

Cet algorithme, présenté dans [7], consiste à se rapprocher itérativement du maillage objectif en ré-interpolant à chaque étape la fonction de contrôle à partir du nouveau maillage. Nous venons de voir qu'il est possible de déterminer simplement et de façon exacte le maillage équidistribué pour une fonction de contrôle affine par morceaux. L'algorithme de De Boor part de ce constat, mais s'autorise une évolution de la fonction interpolée ρ_{interp}^n au cours du calcul. Il se déroule comme suit :

1) On considère :
   a) Une fonction ρ_0 égale à ρ_{interp}^n, donc affine sur chaque segment du maillage M_n.
   b) Un maillage initial M_n^0 égal à M_n.
   c) Un entier i égal à 0.
2) Jusqu'à convergence du maillage...
   a) $i = i + 1$
   b) On calcule le maillage équidistribué correspondant à la fonction de contrôle ρ_{i-1} par la méthode précédemment décrite, la fonction ρ_{i-1} étant affine par morceaux. Ce nouveau maillage est noté M_n^i.
   c) On introduit ρ_i, l'unique fonction affine sur les segments du nouveau maillage M_n^i qui prenne les mêmes valeurs que la fonction ρ_{interp}^n aux sommets de M_n^i.
   d) On revient à a)

Nous obtenons ainsi une suite de maillages M_n^0, M_n^1, ..., M_n^l, qui converge vers une meilleure approximation du maillage équidistribué que celle obtenue par la méthode simple présentée dans la sous-section 2.4.1, dans le sens où elle limite les oscillations

de maillage au cours des itérations physiques. Le détail de cette méthode est donné dans [7].

2.4.3 Limites du calcul direct

La méthode du calcul direct du maillage équidistribué (telle que présentée ci-dessus) permet d'obtenir, connaissant la solution physique à un instant t^n donné, un maillage qui soit bien adapté à cette solution, i.e. qui soit d'autant plus fin que la solution est abrupte. Cependant, exception faite de l'interpolation de la fonction de contrôle, le calcul d'un nouveau maillage ne dépend pas du maillage aux pas de temps précédents. Par conséquent, aucun contrôle ne peut être réalisé sur l'amplitude du mouvement du maillage entre deux itérations, ce qui pose une nouvelle difficulté, relative à la condition de stabilité pour la résolution de l'équation différentielle physique.

2.5 Condition CFL avec maillage mobile

La condition de stabilité (CFL) avec maillage mobile est une extension intuitive de la condition standard sur maillage fixe. L'interprétation classique de la condition sur maillage fixe est la suivante : une particule de fluide ne doit pas parcourir plus d'une maille en un pas de temps. Avec un maillage mobile, cette condition doit rester vraie tout en prenant en compte le mouvement du maillage, de sorte qu'une particule de vitesse nulle peut théoriquement traverser une maille en mouvement. La condition CFL sur maillage mobile est en général plus restrictive que celle sur maillage fixe, et son contrôle plus difficile à assurer.

Quantitativement, pour un système hyperbolique sous forme conservative du type :

$$v_t + F(v)_x = S(v),$$

on note :

- Λ_i le rayon spectral de la matrice jacobienne $\frac{dF(v)}{dv}$ sur le segment i du maillage. À noter que cette matrice est diagonalisable sur $\mathbb{R}$ car le système est hyperbolique.
- $\omega_{i-1/2}$ et $\omega_{i+1/2}$ les vitesses des nœuds situés aux extrémités de ce segment.

la condition CFL, qui doit être vérifiée pour tout i compris entre 2 et N, s'écrit alors :

$$\frac{\Delta t^n}{\Delta x_i^{n+1}}\ max(\Lambda_{i-1/2} - \omega_{i-1/2}, 0) - \frac{\Delta t^n}{\Delta x_{i-1}^{n+1}}\ min(\Lambda_{i-1/2} - \omega_{i-1/2}, 0) \leq 1. \tag{2.3}$$

Cette formule est démontrée dans [13]. On note que dans le cas où ω_i est nul pour tout i, cette condition est assurée par la condition CFL standard :

$$\frac{\Delta t^n\ max\ _i\ (\Lambda_i)}{min\ _i\ (\Delta x_i)} \leq 1. \tag{2.4}$$

La condition sur le pas de temps Δt^n dépend

- de la solution physique au temps t^n, par l'intermédiaire du rayon spectral Λ qui correspond à la vitesse de l'information. Dans le cas de l'équation d'Euler, cette quantité vaut $|u| + c$, où u est la vitesse de la matière et c celle du son,

- du maillage au temps t^{n+1}, et plus exactement de la taille des mailles,
- de la vitesse du maillage entre les temps t^n et t^{n+1}.

La question qui se pose maintenant est de trouver un pas de temps acceptable étant donné les maillages M^n et M^{n+1} correspondant à deux pas de temps successifs. Cela ne présenterait pas plus de difficultés que dans le cas des mailles fixes si la vitesse de maillage ω était connue en tous points, mais ce n'est pas le cas car elle dépend du pas de temps que l'on cherche précisément à calculer. En notant D_i^n le déplacement du maillage au point i (qui, lui, est connu, par simple soustraction des maillages M^{n+1} et M^n), le pas de temps Δt^n doit vérifier :

$$\frac{\Delta t^n}{\Delta x_i^{n+1}}\ max(\Lambda_{i-1/2} - \frac{D_{i-1/2}}{\Delta t^n}, 0) - \frac{\Delta t^n}{\Delta x_{i-1}^{n+1}}\ min(\Lambda_{i-1/2} - \frac{D_{i-1/2}}{\Delta t^n}, 0) \leq 1. \qquad (2.5)$$

On peut, à partir de cette expression, écrire la condition suffisante de stabilité suivante :

$$\frac{\Delta t^n\ \left(max\ _i\ \mid \Lambda_i \mid\right)}{min\ _i\ (\Delta x_i^{n+1})} + max\ _i \left(\frac{\mid D_i \mid}{min\ (\Delta x_i^{n+1}, \Delta x_{i-1}^{n+1})}\right) \leq 1.$$

Nous constatons que dans le cas où $max\ _i \left(\frac{|D_i|}{min\ (\Delta x_i^{n+1}, \Delta x_{i-1}^{n+1})}\right) \geq 1$, il n'existe pas de moyen simple d'assurer la stabilité du schéma par le choix du pas de temps. Or ce cas est d'autant plus probable que le déplacement du maillage entre deux itérations est important, ce qui invite - ainsi que pour des raisons de précision du calcul - à contrôler la vitesse d'évolution de maillage. Cela nous amène aux développements suivants.

2.6 Equations différentielles de maillage

2.6.1 Théorie

Les méthodes présentées ici sont basées sur des équations différentielles ayant pour inconnue le maillage lui même, et sont couramment appelées MMPDE (Moving Mesh Partial Differential Equations). Elles ont pour principal intérêt de permettre un rapprochement progressif du maillage vers le maillage équidistribué tout en limitant l'amplitude des mouvements, ce critère d'amplitude étant défini par l'utilisateur. Elles se basent sur un formalisme spécifique : le maillage d'un domaine $[a, b]$ est vu comme une application $x(\epsilon)$ de $[0, 1]$ dans $[a, b]$. Si N est le nombre de mailles, le i^{eme} nœud du maillage a pour coordonnée $x\left(\frac{i}{N}\right)$. Avec ces notations, on peut prouver qu'un maillage équidistribué minimise la fonctionnelle :

$$I(\epsilon) = \int_a^b \frac{1}{\rho(x)} \left(\frac{\partial \epsilon}{\partial x}\right)^2 dx.$$

La fonction de contrôle ρ a été définie dans la section 2.2. Les développements sont effectués dans [7].

Ainsi, on peut espérer se rapprocher du maillage équidistribué par une méthode de descente, qui se résume dans le cas général à faire évoluer le maillage selon la loi :

$$\frac{\partial \epsilon}{\partial t} = -\frac{P}{\tau} \frac{\partial I}{\partial \epsilon}, \qquad (2.6)$$

où τ est une constante de temps et P un opérateur linéaire défini positif. Plutôt que de chercher un maillage équidistribué à chaque pas de temps physique en faisant converger la méthode de descente, on se propose de ne résoudre la discrétisation de l'équation (2.6) que sur une itération pour calculer un nouveau maillage. Cela revient à assimiler le pas de descente au pas de temps physique, et à faire évoluer au même rythme les deux algorithmes de résolution. Lorsque la fonction de contrôle ρ change au cours du temps - c'est à dire presque toujours -, l'algorithme de descente risque de ne jamais converger vers un maillage équidistribué, mais il se dirigera toujours dans la direction de l'équidistribution tout en voyant sa vitesse de déplacement limitée par le paramètre τ. Nous avons donc bien répondu aux objectifs de contrôle de la vitesse de maillage par l'utilisateur.

2.6.2 Mise en œuvre

Le développement de l'équation (2.6) dépend du choix de P, c'est pourquoi les équations différentielles de maillage ne se ramènent pas à une méthode unique. Dans sa version la plus classique, l'équation (2.6) devient [7] :

$$\frac{\partial \epsilon}{\partial t} = \frac{1}{\tau}\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial \epsilon}{\partial x}\right). \tag{2.7}$$

La discrétisation de cette équation permet d'obtenir ϵ en fonction de x. En revenant à la définition des fonctions ϵ et x, il est clair que pour déterminer le maillage il nous faut connaître x en fonction de ϵ. Cette inversion n'est pas immédiate mais peut s'effectuer à l'aide d'une interpolation de la fonction $\epsilon(x)$ sur l'ensemble du segment $[a, b]$. Etant donné cette nécessité d'inverser la fonction $\epsilon(x)$, on note cette version MMPDE inverse.

Une autre méthode permettant de contourner le problème consiste à exprimer diféremment l'équation (2.6). Par exemple, un choix adapté de l'opérateur P donne [7] :

$$\frac{\partial x}{\partial t} = \frac{1}{\tau}\frac{\partial}{\partial \epsilon}\left(\rho\frac{\partial x}{\partial \epsilon}\right) \tag{2.8}$$

Cependant, ρ étant connu en fonction de x, cette équation est non linéaire et sa résolution s'en trouve moins précise. Par opposition à la méthode MMPDE inverse, on nomme cette version MMPDE directe.

On résout l'équation de descente de façon implicite dans le cas linéaire et semi-implicite dans le cas non linéaire. Dans les deux cas, la résolution se ramène à l'inversion d'un système tridiagonal, effectuée très efficacement par l'algorithme de Thomas [19].

Le détail de ces méthodes, aussi bien pour la théorie que pour la mise en œuvre, est exposé dans [7].

2.7 Méthode VFFC pour les modèles non conservatifs

Afin d'appliquer la méthode AMR à des systèmes d'équations non conservatifs, et avant de nous intéresser à la prise en compte du mouvement de mailles dans la

résolution des équations physiques, présentons tout d'abord la façon dont on peut prendre en compte l'apparition de termes non conservatifs dans l'application de la méthode VFFC (cf. sous-section 1.3.1) en une dimension. Nous donnons pour cadre à cette section les équations du type :

$$\frac{\partial v}{\partial t} + \frac{\partial F(v)}{\partial x} + C(v)\ \frac{\partial v}{\partial x} = S(v) \tag{2.9}$$

discrétisée sur des volumes à géométrie variable. L'inconnue v dépend du temps t et de la dimension d'espace (unique) x. Par ailleurs,

- $F(v)$ désigne un flux vectoriel de même dimension que v.
- $S(v)$ désigne un terme source de même dimension que v.
- $C(v)$ désigne une matrice carrée dont la taille est la dimension de v.

Divers exemples de systèmes de ce type issus des équations de la mécanique des fluides sont donnés dans la suite. Nous notons :

$$J(v) = \frac{\partial F(v)}{\partial v},$$

et définissons la matrice $E(v)$ telle que

$$E(v)\ J(v) = C(v).$$

L'équation (2.9) peut alors se réécrire :

$$\frac{\partial v}{\partial t} + \big(Id + E(v)\big)\ \frac{\partial F(v)}{\partial x} = S(v), \tag{2.10}$$

et son intégration sur une cellule devient :

$$\int_{K(t)} \frac{\partial v}{\partial t}\ dx + \int_{K(t)} \big(Id + E(v)\big)\ \frac{\partial F(v)}{\partial x}\ dx = \int_{K(t)} S(v)\ dx \tag{2.11}$$

que l'on peut approcher par :

$$\int_{K(t)} \frac{\partial v}{\partial t}\ dx + \big(Id + E(v)\big)\ \int_{K(t)} \frac{\partial F(v)}{\partial x}\ dx \simeq \int_{K(t)} S(v)\ dx. \tag{2.12}$$

La discrétisation donne, en utilisant la version classique de la méthode VFFC :

$$v_i^{n+1} - v_i^n \simeq \\ - \Delta t\ \big(Id + E(v_i^n)\big)\ \big(F_{i+1/2} - F_{i-1/2}\big) + \Delta t\ |K_i|\ S(v_i), \tag{2.13}$$

où les flux $F_{i-1/2}$ et $F_{i+1/2}$ se calculent de même que dans la version standard.

Remarque Dans l'écriture (2.9) on peut ajouter à $F(v)$ un flux $G(v)$ et soustraire $-\nabla_v G(v)$ à $C(v)$, de sorte que le modèle (2.9) reste inchangé. On peut utiliser cette remarque et opérer les opérations nécessaires pour assurer à la matrice J de bonnes propriétés et permettre ainsi un calcul correct de la matrice $E(v)$.

2.8 Résolution du système physique avec AMR

Dans cette section, afin d'alléger les notations, les dérivées partielles par rapport aux variables x et t sont notées de façon indicielle. Une fois mise au point la technique de déplacement du maillage, il reste à résoudre le système physique en prenant en compte ce déplacement. Cette section a pour cadre à la fois les modèles conservatifs et non conservatifs, les uns comme les autres entrant dans le formalisme introduit dans la section 2.7. Nous utilisons une relation valable pour toute fonction $f(x,t)$ sous réserve de bonnes propriétés d'intégration et d'ensemble de définition :

$$\frac{d}{dt}\int_{K(t)} f\ dx = \int_{K(t)} f_t\ dx + \int_{K(t)} (\lambda\ f)_x\ dx, \tag{2.14}$$

où λ désigne la vitesse d'évolution du domaine d'intégration, dans notre cas la vitesse de maillage. Reprenant le formalisme de la section précédente, nous allons adapter la déclinaison du schéma numérique à l'existence d'un maillage mobile. La substitution de v à f dans (2.14) et l'intégration de (2.9) sur un volume $K(t)$ donnent :

$$\frac{d}{dt}\int_{K(t)} v(t)\ dx + \int_{K(t)} (F(v) - \lambda v)_x + C(v)\ v_x\ dx = \int_{K(t)} S(v)\ dx. \tag{2.15}$$

Nous notons :

- $F^*(v,\lambda) = F(v) - \lambda\ v$,
- $J^*(v,\lambda) = \frac{\partial F^*(v,\lambda)}{\partial v} = J(v) - \lambda\ Id$,
- $E^*(v,\lambda)$ la matrice vérifiant $E^*(v,\lambda)J^*(v,\lambda) = C(v)$,
- $S^*(v,\lambda,\lambda_x) = S(v) - E^*(v,\lambda)\ \lambda_x\ v$.

Nous pouvons alors réécrire l'équation (2.15) :

$$\frac{d}{dt}\int_{K(t)} v(t)\ dx + \int_{K(t)} \big(Id + E^*(v,\lambda)\big)\ F^*(v,\lambda)_x\ dx = \int_{K(t)} S^*(v,\lambda,\lambda_x)\ dx \tag{2.16}$$

et la discrétisation donne, en utilisant la version classique de la méthode VFFC :

$$\begin{aligned} |K_i^{n+1}|\ v_i^{n+1} - |K_i^n|\ v_i^n \simeq \\ - \Delta t\ \big(Id + E^*(v_i^n, \lambda_i^n)\big)\ \big(F^*_{i+1/2} - F^*_{i-1/2}\big) + \Delta t\ |K_i|\ S^*(v_i, \lambda_i^n, \lambda_{x_i}^n). \end{aligned} \tag{2.17}$$

A moins que ce ne soit précisé, nous prenons ces valeurs à l'instant n ou $n+1$ selon le choix d'une itération implicite ou explicite. Les paramètres de déplacement de maillage se calculent très naturellement : La valeur de λ_i^n s'obtient par une moyenne pondérée des vitesses de déplacement des deux nœuds adjacents à la cellule i, ces vitesses de déplacement étant calculées en divisant par le pas de temps les position de ces nœuds aux itérations n et $n+1$. La valeur de $\lambda_{x_i}^n$ s'obtient en divisant par la largeur de la cellule i la différence de vitesse de déplacement des nœuds adjacents à la cellule i.

Il reste à calculer les flux $F^*_{i+1/2}$ et $F^*_{i-1/2}$. Multiplions l'équation (2.9) par $J^*(v)$, nous obtenons :

$$F^*(v,\lambda)_t + J^*(v,\lambda)\ F(v)_x + J^*(v,\lambda)\ E^*(v,\lambda)\ F^*(v,\lambda)_x. = J^*(v,\lambda)\ S^*(v,\lambda,\lambda_x) - \lambda_t\ v \tag{2.18}$$

La relation (2.14) dans laquelle on a substitué $F^*(v,\lambda)$ à f ainsi que l'équation (2.18) intégrée sur un volume $K(t)$ donnent :

$$\frac{d}{dt}\int_{K(t)} F^*(v,\lambda)\ dx + \int_{K(t)} J^*(v,\lambda)\ F(v)_x - \int_{K(t)} \big(\lambda\ F^*(v,\lambda)\big)_x\ dx + \int_{K(t)} J^*(v,\lambda)\ E^*(v,\lambda)\ F^*(v,\lambda)_x\ dx = \int_{K(t)} \big(J^*(v,\lambda)\ S^*(v,\lambda,\lambda_x) - \lambda_t\ v\big)\ dx. \tag{2.19}$$

Nous avons :

$$\big(\lambda\ F^*(v,\lambda)\big)_x = J^*(v,\lambda)\ (\lambda\ v)_x - \lambda_x\ J^*(v,\lambda)\ v + \lambda_x\ F^*(v,\lambda), \tag{2.20}$$

donc en notant :

$$S^{**}(v,\lambda,\lambda_x) = J^*(v,\lambda)\ S^*(v,\lambda,\lambda_x) - \lambda_t\ v - \lambda_x\ J^*(v,\lambda)\ v + \lambda_x\ F^*(v,\lambda), \tag{2.21}$$

on peut transformer l'équation (2.19) en :

$$\frac{d}{dt}\int_{K(t)} F^*(v,\lambda)\ dx + \int_{K(t)} J^*(v,\lambda)\ (Id + E^*(v,\lambda))\ F^*(v,\lambda)_x = \int_{K(t)} S^{**}(v,\lambda,\lambda_x)\ dx. \tag{2.22}$$

Cela montre que le flux $F^*(v,\lambda)$ (vu comme la moyenne de la quantité $F^*(v,\lambda)$ sur une cellule) est advecté par la matrice $\tilde{A}^*(v,\lambda) = J^*(v,\lambda)\ \big(Id + E^*(v,\lambda)\big)$, ce qui permet de calculer les flux $F^*_{i+1/2}$ et $F^*_{i-1/2}$ par la méthode VFFC.

Remarque De même que dans le cas sans AMR (cf. section 2.7), dans l'écriture (2.9) on peut ajouter à $F(v)$ un flux $G(v)$ et soustraire $-\nabla_v G(v)$ à $C(v)$, de sorte que le modèle (2.9) reste inchangé. Cette remarque sera cette fois utilisée pour assurer à $J^*(v,\lambda)$ de bonnes propriétés d'inversion à J^* et non plus à J. Ainsi, le mouvement du maillage est à prendre en compte dans le choix de $G(v)$. Nous verrons un exemple d'application de cette remarque dans le cas du modèle bi-fluide sans énergie à quatre équations.

2.9 Calcul des matrices

Nous cherchons dans cette partie à obtenir de façon simple l'expression analytique des matrices intervenant dans l'équation physique adaptée aux mouvements de mailles (matrices notées avec un exposant * dans la section 2.8). Nous partons du principe que l'expression de ces matrices est connue dans le cas des mailles fixes [1, 6]. Comme dans la section 2.8, v correspond au vecteur des variables résolues. Nous définissons également le vecteur w, de même dimension que v, qui correspond aux variables physiques. Il contient par exemple la température, la pression, la vitesse du fluide... Il est défini de telle sorte que v soit déductible de la donnée de w, et inversement.

2.9.1 Méthode de calcul

Pour un réel λ donné, qui correspondra, comme dans la section 2.8, à la vitesse de maillage, nous notons :

- f l'application qui à w associe $F(v)$,
- g la bijection qui à w associe v,
- h_λ la bijection qui à w associe le vecteur obtenu à partir de w en soustrayant λ aux composantes correspondant à des vitesses.

La traduction de ces notations pour différents modèles est donnée dans la section 2.10. Pour tout λ, nous notons $M(\lambda)$ l'application $g \circ h_\lambda \circ g^{-1}$. Nous notons que :

$$h_\lambda^{-1} = h_{-\lambda}.$$

Par conséquent :

$$M(\lambda)^{-1} = M(-\lambda).$$

Par ailleurs, nous faisons les deux hypothèses suivantes, qui seront toujours vérifiées pour nos cas :

(H1) Pour tout λ, l'application $M(\lambda)$ est linéaire.

(H2) $M(\lambda) \circ \big(f - \lambda\ g\big) = f \circ h_\lambda$.

L'hypothèse (H2) est liée aux remarques sur l'invariance galiléenne données dans la section 2.9.2. L' hypothèse (H1) est plus empirique...

Notons j l'application qui à w associe la matrice $J(v,\lambda)$, et j^* celle qui à (w,λ) associe la matrice $J^*(v,\lambda)$. Nous déduisons des résultats précédents que :

$$j^*(w,\lambda) = \frac{\partial\big(f(w) - \lambda g(w)\big)}{\partial g(w)} = \frac{\partial\big(f(w) - \lambda g(w)\big)}{\partial f\big(h_\lambda(w)\big)} \times \frac{\partial f\big(h_\lambda(w)\big)}{\partial g\big(h_\lambda(w)\big)} \times \frac{\partial g\big(h_\lambda(w)\big)}{\partial g(w)}.$$

Nous trouvons alors, en vertu de l'hypothèse (H2) :

$$j^*(w,\lambda) = M(\lambda)^{-1}\ j\big(h_\lambda(w)\big)\ M(\lambda) = M(-\lambda)\ j\big(h_\lambda(w)\big)\ M(\lambda), \tag{2.23}$$

où $M(\lambda)$ est considérée comme une matrice. Il reste à calculer la matrice $E^*(v,\lambda)$. Nous notons e l'application qui à w associe la matrice $E(v)$ et e^* l'application qui à (w,λ) associe la matrice $E^*(v,\lambda)$. On sait que l'on a pour tout (w,λ) :

$$e^*(w,\lambda)\ j^*(w,\lambda) = c(w),$$

donc :

$$e^*(w,\lambda)\ M(\lambda)^{-1}\Big(M(\lambda)\ j^*(w,\lambda)\ M(\lambda)^{-1}\Big)M(\lambda) = c(w).$$

Donc, d'après l' équation (2.23) :

$$e^*(w,\lambda)\ M(-\lambda)\ j\big(h(w)\big)\ M(\lambda) = c(w).$$

Par conséquent, en supposant que $j(w)$ est inversible :

$$e^*(w,\lambda) = c(w)\ M(-\lambda)\ j\big(h(w)\big)^{-1} M(\lambda), \tag{2.24}$$

ce qui permet d'obtenir $E^*(v,\lambda)$ à partir de l'expression analytique de $J(v)^{-1}$. Quant à la matrice $\tilde{A}^*(v,\lambda)$, elle se calcule directement grâce aux résultats précédents, étant donné que par définition $\tilde{A}^*(v,\lambda) = J^*(v,\lambda)\ (Id + E^*(v,\lambda))$.

2.9.2 Invariance galiléenne

L'objectif est ici de justifier l'hypothèse (H2) de la sous-section 2.9.1. Nous rappelons que $M(\lambda)$ définit la matrice de l'application $g \circ h_\lambda \circ g^{-1}$. Nous souhaitons maintenant comprendre pourquoi nous avons toujours le résultat : $M(\lambda) \circ \big(f - \lambda\ g\big) = f \circ h_\lambda$.

Considérons l'équation (2.9) dans laquelle le terme source et le terme non conservatif ont été annulés, et dans laquelle nous prenons pour inconnue w :

$$\frac{\partial g(w)}{\partial t} + \frac{\partial f(w)}{\partial x} = 0. \tag{2.25}$$

Considérons - comme cela doit être le cas - que l'équation (2.25) satisfait au principe d'invariance galiléenne. Dans le cas non conservatif, cette hypothèse dépend du choix de la définition du flux F, qui n'est pas unique [1]. L'invariance galiléenne signifie que pour tout λ fixé, si $w(x,t)$ est solution de l'équation (2.25), alors $h_\lambda\big(w(x+\lambda t, t)\big)$ la vérifie également. En supposant de plus que l'ensemble de définition de w recouvre tout l'espace, alors en substituant $h_\lambda\big(w(x+\lambda t, t)\big)$ à w dans (2.25) et en opérant la dérivation temporelle, on trouve la relation suivante valable pour la solution w de (2.25) :

$$\frac{\partial\big(g \circ h_\lambda\ (w)\big)}{\partial t} + \frac{\partial\big(f \circ h_\lambda\ (w) + \lambda\ g \circ h_\lambda\ (w)\big)}{\partial x} = 0. \tag{2.26}$$

En multipliant maintenant l'équation (2.25) par $M(\lambda)$, elle devient :

$$\frac{\partial\big(g \circ h_\lambda\ (w)\big)}{\partial t} + \frac{\partial\big(M(\lambda) \circ f(w)\big)}{\partial x} = 0 \tag{2.27}$$

et les équations (2.26) et (2.27) donnent :

$$M(\lambda) \circ f\ (w) = f \circ h_\lambda\ (w) + \lambda\ g \circ h_\lambda\ (w) + C(t) \tag{2.28}$$

avec $C(t)$ une constante ne dépendant que du temps, ou encore :

$$M(\lambda) \circ \big(f - \lambda\ g\big)\ (w) = f \circ h_\lambda\ (w) + C(t). \tag{2.29}$$

Cela est très cohérent avec le résultat selon lequel $M(\lambda) \circ \big(f - \lambda\ g\big) = f \circ h_\lambda$.

2.10 Modèles considérés

Dans cette partie, nous déclinons la théorie précédemment décrite pour trois modèles particuliers : le modèle d'advection simple à une équation, le modèle d'Euler à trois équations et le modèle bi-fluide sans énergie à quatre équations. Les notations sont les suivantes : α fraction volumique, ρ masse volumique, u vitesse, p pression, T température, e énergie interne massique, E énergie totale massique (i.e. $E = e + \frac{1}{2}\,\rho\,u^2$), H enthalpie totale massique (i.e. $H = E + \frac{p}{\rho}$). Les indices correspondent aux espèces considérées, par exemple : a : air, w : eau liquide.

2.10.1 Une équation

Le modèle d'advection par une vitesse fixe u d'inconnue ρ s'écrit :

$$\frac{\partial \rho}{\partial t} + u \, \frac{\partial \rho}{\partial x} = 0. \tag{2.30}$$

Il s'agit d'un modèle d'advection simple, pour lequel la condition initiale est transportée sans déformation à la vitesse u. Reprenant le formalisme de la section 2.7, nous pouvons écrire :

$$\frac{\partial v}{\partial t} + \frac{\partial F(v)}{\partial x} = 0 \tag{2.31}$$

avec :

- $v = \rho$,
- $F = \rho u$,
- $w = v$,
- h_λ est la fonction identité,
- $M(\lambda) = 1$.

Ce modèle est conservatif, en particulier la matrice $C(v)$ définie dans l'équation (2.9) est nulle. La théorie présentée ci-dessus s'applique, mais n'est pas nécessaire étant donné la simplicité du modèle.

2.10.2 Trois équations

Le modèle d'Euler à trois équations (une espèce avec énergie) s'obtient en écrivant la conservation de la masse, de la quantité de mouvement et de l'énergie. Il faut y ajouter une loi d'état reliant la pression, la masse volumique et l'énergie interne afin de rendre le problème bien posé. Cette loi d'état s'écrit en général de façon plus simple lorsqu'on la décompose en deux relations reliant d'une part l'énergie interne à la pression et à la température, d'autre part la masse volumique à la pression et à la température ($\rho(p,T)$ et $e(p,T)$). Les trois lois de conservation de ce modèle sont les suivantes :

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho u)}{\partial x} = 0. \tag{2.32}$$

$$\frac{\partial (\rho u)}{\partial t} + \frac{\partial (\rho u^2)}{\partial x} + \frac{\partial p}{\partial x} = 0. \tag{2.33}$$

$$\frac{\partial (\rho E)}{\partial t} + \frac{\partial (\rho u H)}{\partial x} = 0. \tag{2.34}$$

Reprenant le formalisme décrit de la section 2.7, nous pouvons écrire :

$$\frac{\partial v}{\partial t} + \frac{\partial F(v)}{\partial x} = 0 \tag{2.35}$$

avec :

- $v = (\rho, \rho u, \rho E)$,
- $F(v) = \big(\rho u, \rho u^2 + p, \rho u H\big)$,
- $w = (p, T, u)$,
- h_λ est la bijection qui à $w = (p, T, u)$ associe $(p, T, u - \lambda)$.

La matrice $M(\lambda)$ s'écrit :

$$M(\lambda) = \begin{pmatrix} 1 & 0 & 0 \\ -\lambda & 1 & 0 \\ \frac{\lambda^2}{2} & -\lambda & 1 \end{pmatrix}. \tag{2.36}$$

Ce modèle est conservatif, en particulier la matrice $C(v)$ définie dans l'équation (2.9) est nulle, mais toutes les étapes décrites ci-dessus restent valables (et en sont largement simplifiées).

Il est à noter que pour appliquer la méthode VFFC, l'ensemble des variables physiques ne se déduisant simplement qu'à partir de la donnée du vecteur w (contenant notamment les variables d'entrée des lois d'état, à savoir p et T), nous devons pouvoir déduire le vecteur w du vecteur v. Nous y parvenons grâce aux opérations suivantes, en notant $v = (v_1, v_2, v_3)$:

$$\rho = v_1,\ u = \frac{v_2}{v_1},\ e = \frac{v_3}{v_1} - \frac{1}{2}\left(\frac{v_2}{v_1}\right)^2$$

qui nous amènent à écrire :

$$v_1 = \rho(p, T) \tag{2.37}$$

$$\frac{v_3}{v_1} - \frac{1}{2}\left(\frac{v_2}{v_1}\right)^2 = e(p, T). \tag{2.38}$$

Ce système a priori non linéaire d'inconnues p et T peut être résolu par une méthode de Newton, ce qui nous permet d'en déduire le vecteur w. Il nous faut aussi, afin de déterminer l'expression de la matrice $J(v)$, connaître les expressions des dérivées des coordonnées du vecteur w par rapport à celles du vecteur v. En effet le vecteur $F(v)$ ne s'exprime simplement qu'en fonction des coordonnées du vecteur w. Dans ce but, il suffit de calculer et d'inverser la matrice $\frac{\partial v}{\partial w}$ contenant les dérivées des coordonnées de v par rapport à celles de w. On remarque que son expression fera intervenir les dérivées des lois d'état, ce qui constitue une difficulté lorsque celles-ci sont tabulées.

2.10.3 Quatre équations

Le modèle à quatre équations traite les problèmes à deux espèces et sans énergie. On écrit pour chacune des deux espèces une loi de conservation de la masse et une loi de conservation de quantité de mouvement. Ici nous n'ajoutons pas la conservation de l'énergie, car nous considérons que l'état du système suit une évolution isentropique, ce qui implique que les lois d'état relient directement la masse volumique à la pression $(\rho_a(p),\ \rho_w(p))$. On note que la pression est la même pour les deux espèces, et que les deux lois d'état correspondant aux deux espèces sont a priori distinctes. Les lois de conservation s'écrivent ainsi :

$$\frac{\partial(\alpha_a\rho_a)}{\partial t} + \frac{\partial(\alpha_a\rho_a u_a)}{\partial x} = 0, \tag{2.39}$$

$$\frac{\partial(\alpha_w\rho_w)}{\partial t} + \frac{\partial(\alpha_w\rho_w u_w)}{\partial x} = 0, \tag{2.40}$$

$$\frac{\partial(\alpha_a\rho_a u_a)}{\partial t} + \frac{\partial(\alpha_a\rho_a u_a^2)}{\partial x} + \alpha_a\frac{\partial p}{\partial x} = \alpha_a\ \rho_a\ g, \tag{2.41}$$

$$\frac{\partial(\alpha_w\rho_w u_w)}{\partial t} + \frac{\partial(\alpha_w\rho_w u_w^2)}{\partial x} + \alpha_w\frac{\partial p}{\partial x} = \alpha_w\ \rho_w\ g, \tag{2.42}$$

où g désigne l'accélération de pesanteur. En raison des derniers termes des deux équations de conservation de la quantité de mouvement, faisant intervenir des fractions volumiques à l'extérieur des dérivées spatiales, ce système n'est pas conservatif et nécessitera l'écriture d'une matrice $C(v)$ non nulle (cf. équation (2.9)). Comme vu dans les sections 2.7 et 2.8, le choix de cette matrice $C(v)$ n'est pas fixé mais dépend de l'expression du flux $F(v)$. De façon naturelle, on pourrait écrire :

$$F(v) = \big(\alpha_a\rho_a u_a, \alpha_w\rho_w u_w, \alpha_a(\rho_a u_a^2 + p), \alpha_w(\rho_w u_w^2 + p)\big).$$

Mais alors rien ne nous assurerait que la matrice $J(v)$ (ou $J^*(v)$ dans le cas avec AMR) soit inversible. Nous introduisons le paramètre $\pi(t)$ ne dépendant que du temps, et nous écrivons :

$$F(v) = \big(\alpha_a\rho_a u_a, \alpha_w\rho_w u_w, \alpha_a(\rho_a u_a^2 + p - \pi), \alpha_w(\rho_w u_w^2 + p - \pi)\big),$$

ce qui nous donne :

$$C(v) = -(p-\pi)\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \alpha_{a,1} & \alpha_{a,2} & \alpha_{a,3} & \alpha_{a,4} \\ \alpha_{w,1} & \alpha_{w,2} & \alpha_{w,3} & \alpha_{w,4} \end{pmatrix}, \tag{2.43}$$

où nous avons noté, pour i allant de 1 à 4, $\alpha_{a,i}$ la dérivée de α_a par rapport à la i-ème coordonnée de v, et $\alpha_{w,i}$ la dérivée de α_w par rapport à cette même coordonnée.

Calcul de π sans AMR Il nous faut choisir un paramètre $\pi(t)$ qui optimise les propriétés d'inversibilité de la matrice $J(v)$ en tout point du domaine. Nous allons dans ce but commencer par déterminer un paramètre π local en chaque cellule du maillage, qui optimisera les propriétés d'inversibilité de la matrice $J(v)$ en assurant à son déterminant une valeur suffisamment grande (les détails sont donnés dans [1]). Nous calculons dans ce but une expression analytique de π fonction des coordonnées du vecteur w, et notons π_0 la fonction qui à ω associe ce paramètre π local. De ce calcul des paramètres π locaux nous devons déduire un paramètre π global - rappelons que π ne doit dépendre que du temps -. Nous pouvons pour cela considérer le minimum des paramètres π locaux, après nous être assurés du caractère affine et décroissant de la relation liant le paramètre π local au déterminant de $J(v)$. Une fois encore, le lecteur est renvoyé à [1] pour plus de détails.

Calcul de π avec AMR Il nous faut maintenant nous assurer que $J^*(v,\lambda)$ soit inversible en suivant le même processus, ce qui fera nécessairement dépendre le paramètre π local de λ en plus de ω. Nous notons π^* la fonction qui à (ω,λ) associe le paramètre π local pour le cas avec AMR. Nous allons maintenant tenter de déduire son expression de celle de π_0, pour nous épargner le calcul formel. En vertu de (2.23), nous savons que les matrices $j^*(\omega,\lambda)$ et $j\big(h_\lambda(\omega)\big)$ sont semblables. Elles ont donc le même déterminant et les mêmes propriétés d'inversibilité. Rappelant que $\pi^*(\omega,\lambda)$ doit optimiser les propriétés d'inversibilité de $j^*(\omega,\lambda)$ en jouant sur son déterminant, et qu'il en va de même de $\pi_0\big(h_\lambda(\omega)\big)$ vis à vis de $j\big(h_\lambda(\omega)\big)$, nous pouvons alors écrire :

$$\pi^*(\omega,\lambda) = \pi_0\big(h_\lambda(\omega)\big).$$

L'expression analytique de la fonction π_0 étant d'ores et déjà connue (donnée dans [1]), la généralisation du calcul du paramètre π local aux cas avec AMR ne nécessite pas de calcul formel supplémentaire. Par ailleurs, pour les mêmes raisons que dans le cas sans AMR, le paramètre π global peut être pris égal au minimum des paramètres π locaux.

Nous écrivons finalement, après avoir calculé et fixé un paramètre π ne dépendant que de notre itération courante :

$$\frac{\partial v}{\partial t} + \frac{\partial F(v)}{\partial x} + C(v)\ \frac{\partial v}{\partial x} = 0 \tag{2.44}$$

avec :

- $v = (\alpha_a\rho_a, \alpha_w\rho_w, \alpha_a\rho_a u_a, \alpha_w\rho_w u_w)$,
- $F = (\alpha_a\rho_a u_a, \alpha_w\rho_w u_w, \alpha_a(\rho_a u_a^2 + p - \pi), \alpha_w(\rho_w u_w^2 + p - \pi))$,
- $w = (\alpha_a, u_a, u_w, p)$,
- h_λ est la bijection qui à $w = (\alpha_a, u_a, u_w, p)$ associe $(\alpha_a, u_a - \lambda, u_w, p)$.

La matrice $M(\lambda)$ s'écrit :

$$M(\lambda) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\lambda & 0 & 1 & 0 \\ 0 & -\lambda & 0 & 1 \end{pmatrix}. \tag{2.45}$$

Une fois encore, l'ensemble des variables thermodynamiques du problème ne se déduisant simplement que des coordonnées du vecteur w, il nous faut pouvoir le déterminer à partir de la donnée de v, seul vecteur réactualisé par l'application du schéma numérique. Pour cela nous notons $v = (v_1, v_2, v_3, v_4)$ et nous écrivons :

$$u_a = \frac{v_3}{v_1},\ u_w = \frac{v_4}{v_2}.$$

La relation $\alpha_a + \alpha_w = 1$ peut s'écrire :

$$\frac{v_1}{\rho_a(p)} + \frac{v_2}{\rho_w(p)} = 1, \tag{2.46}$$

équation a priori non linéaire d'inconnue p, que l'on peut résoudre par une méthode de Newton. On déduit de sa résolution, par l'intermédiaire des lois d'état, les données de ρ_a et ρ_w. Il ne nous reste plus qu'à déterminer α_a, qui s'écrit naturellement $\alpha_a = \frac{v_1}{\rho_a}$. Par ailleurs, ici aussi, afin de déterminer l'expression des matrice $J(v)$ et $C(v)$, il nous faut connaître les expressions des dérivées des coordonnées du vecteur w par rapport à celles du vecteur v. Comme pour le modèle à trois équations, il suffit de calculer et d'inverser la matrice $\frac{\partial v}{\partial w}$ contenant les dérivées des coordonnées de v par rapport à celles de w. Les expressions résultantes dépendront des lois d'état et de leurs dérivées.

Plus généralement, on peut décliner ainsi nombre de modèles, plus ou moins complexes en fonction du nombre d'espèces considérées, de leur caractère monophasique ou multiphasique, de la prise en compte ou non d'une température dans la loi d'état, de

l'homogénéité ou non des températures entre phases... L'intérêt de la théorie présentée dans la section 2.9 est de décliner *sans efforts* l'adaptation des données de ces modèles au mouvement des mailles, à partir de leurs équivalents sans mouvement de mailles. Nous avons vu en particulier, dans la section 2.9, que l'adaptation des matrices E et J était aisée, et dans ce paragraphe sur le modèle à quatre équations, que l'adaptation du paramètre π n'était pas plus compliquée. Mais au delà de cet intérêt *pratique*, l'assurance de maintenir telle quelle l'expression analytique du paramètre π, à la soustraction près de la vitesse de maillage aux vitesses physiques, nous permet d'affirmer que la transition du paramètre π local au paramètre π global est toujours possible, et demeure identique. Cette théorie peut s'appliquer à tous les modèles qui vérifient l'hypothèse d'invariance galiléenne, ce qui est en général *naturellement* le cas. En revanche, l'introduction *artificielle* du paramètre π, que nous avons vu ne pas poser de difficulté particulière pour le modèle à quatre équations, invalidera cette propriété d'invariance galiléenne pour le modèle à 7 équations que nous présenterons dans la section 3.2, ou du moins nous obligera pour la respecter à complexifier considérablement la mise en œuvre de ce modèle. Ce constat nous interdit pour l'instant d'opérer avec succès notre stratégie de mise en place de l'AMR sur ce dernier modèle. Les détails sur cette difficulté seront donnés dans la sous-section 3.2.3.

2.11 Comparaison des méthodes de remaillage

Nous utilisons le modèle d'advection présenté dans la sous-section 2.30. Il s'agit simplement de résoudre l'équation :

$$\frac{\partial \rho}{\partial t} + u \frac{\partial \rho}{\partial x} = 0,$$

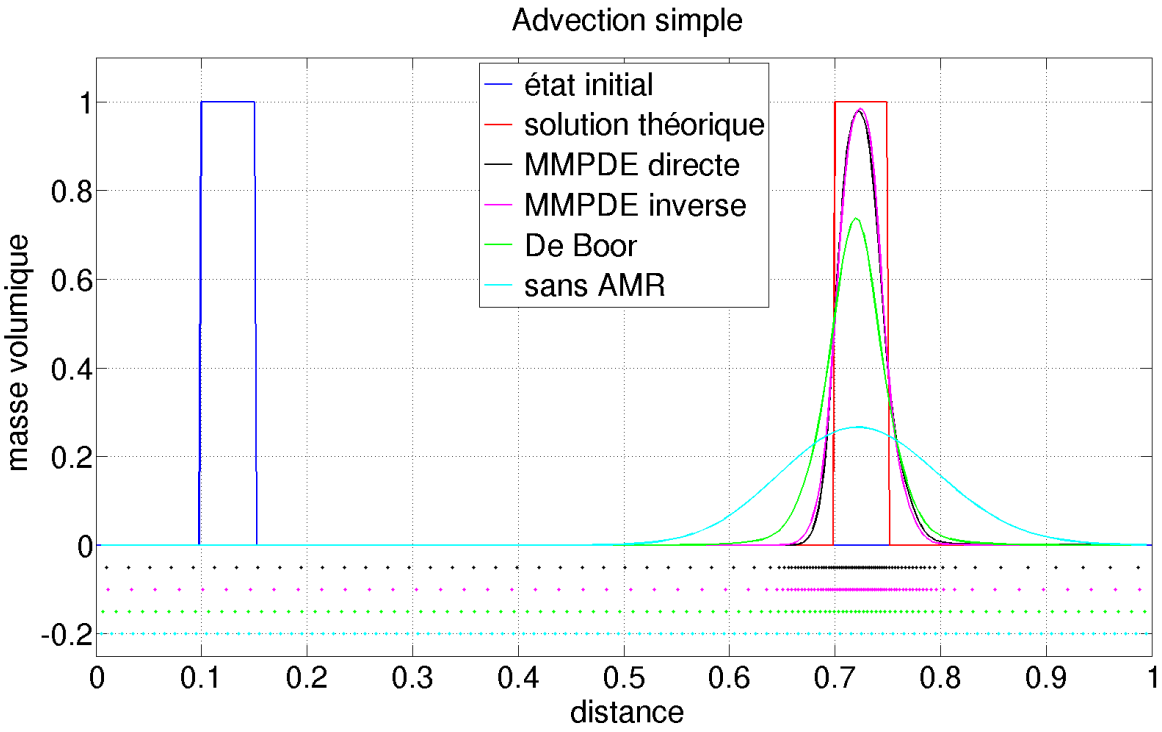
où l'unique inconnue est ρ. Nous avons appliqué ce modèle au déplacement d'un créneau entraîné par une vitesse unitaire au sein d'un domaine de longueur 1. Les différentes méthodes de calcul du mouvement du maillage, présentées plus haut, ont été testées. Nous avons en particulier obtenu les résultats suivants :

Nombre de cellules	Méthode	erreur L2	Temps CPU en s
100	Sans AMR	0.18	2.17
	De Boor	0.11	2.68
	MMPDE directe	0.078	2.67
	MMPDE inverse	0.077	2.65
500	Sans AMR	0.11	7.18
	De Boor	0.146	18.5
	MMPDE directe	0.051	65.3
	MMPDE inverse	0.058	39.4
5000	Sans AMR	0.04	404
	De Boor	0.18	436
	MMPDE directe	0.058	432
	MMPDE inverse	0.078	435

Les cas à 5000 cellules avec AMR ont dû être traités en implicite - pour la plupart des itérations - car la condition de stabilité présentée dans la section 2.5 n'était pas vérifiable, y compris en jouant sur le pas de temps. Nous remarquons que l'utilisation

de l'AMR dans ce cas ne présente plus d'intérêt, car alors les gains en terme de taille de maille deviennent faibles devant les pertes de précision liés à la modification du schéma numérique. Ce phénomène apparaît dès 500 cellules pour la méthode de De Boor.

Pour chaque méthode, nous avons utilisé les paramètres qui minimisaient l'erreur L2. Les courbes obtenues pour un maillage à 100 éléments sont les suivantes :



Au vu de ces résultats, nous utiliserons par la suite des équations différentielles de maillage (MMPDE). La meilleure qualité des résultats avec cette méthode s'explique par le fait qu'on a pu diminuer le paramètre α (équation (2.1)) de la fonction de contrôle et atténuer son lissage pour qu'elle soit plus représentative des variations de la solution physique, tout en contrôlant l'amplitude du mouvement des mailles par la gestion du paramètre τ (équation (2.6)).

2.12 Cas test de Sod

Nous avons par la suite testé cette méthodologie sur le modèle des équations d'Euler avec énergie, présenté dans la sous-section 2.10.2, qui a l'avantage d'être conservatif, ce qui signifie que la matrice C de l'équation (2.9) est nulle. Le cas test de Sod a été considéré. Il s'agit d'un domaine 1D de longueur 1, contenant initialement un matériau 1 à gauche et un matériau 2 à droite, séparés par une interface à mi-distance et tous deux constitués d'un gaz parfait de coefficient $\gamma = 1.4$. Notant ρ la masse volumique, p la pression et e l'énergie interne massique, cette loi d'état s'écrit $p = (\gamma - 1)\,\rho\, e$. Notons que rien ne nous interdit, pour rester dans le cadre théorique de la présentation du modèle à trois équations, de décomposer cette loi d'état en deux lois reliant l'énergie

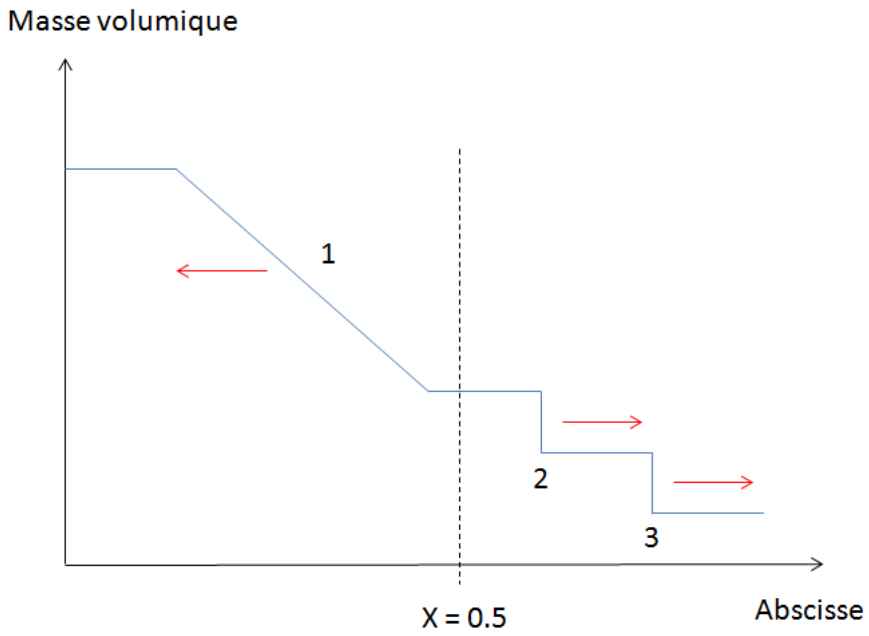
interne et la masse volumique à la pression et la température. Nous pouvons introduire par exemple le paramètre de capacité massique C_v, dont la valeur n'influera pas sur notre résultat, et écrire :

- $e(p,T) = C_v\ T$.
- $\rho(p,T) = \frac{p}{(\gamma-1)\ C_v\ T}$.

A l'état initial le matériau 1 a une masse volumique $\rho_1 = 1$, une vitesse $u_1 = 0$ et une pression $p_1 = 1$, tandis que le matériau 2 a pour propriétés $\rho_2 = 0.125$, $u_2 = 0$ et $p_2 = 0.1$. Le temps final est 0.2. Nous imposons de part et d'autre du domaine des conditions aux limites de Neumann homogènes, qui n'auront toutefois par d'importance car nous verrons que le temps final de $0.2s$ ne laisse pas le temps à l'information d'atteindre le bord du domaine depuis la discontinuité. Il s'agit d'un problème de Riemann monodimensionnel communément utilisé pour tester la précision des codes de mécanique des fluides. Il est particulièrement intéressant car malgré sa simplicité apparente, il présente trois phénomènes dont la résolution peut s'avérer difficile, et que l'on peut décrire qualitativement ainsi :

1. La discontinuité de contact se maintiendra au cours du temps, mais sa position variera. La pression initiale étant plus importante à gauche, elle se dirigera vers la droite. En dehors de l'instant initial, la pression et la vitesse seront continues à travers la surface, seules la masse volumique et l'énergie interne seront discontinues.
2. Une discontinuité correspondant à onde de choc se créera au niveau de l'interface au temps initial et parcourra le domaine vers la droite à la vitesse du son, soit plus rapidement que la discontinuité de contact. Au travers de cette onde de choc, toutes les quantités (ρ, u, p, e) sont en général discontinues.
3. Parallèlement, une onde de raréfaction (ne présentant pas de discontinuité) se propagera vers la gauche. L'extrémité gauche de cette onde se déplacera également à la vitesse du son.

Cette évolution se traduira par la présence de quatre paliers aux états thermodynamiques uniformes, comme illustré ci-dessous pour le profil de masse volumique. Les numéros correspondent à ceux de l'énumération ci-dessus.



Davantage de détails sur ce cas sont donnés dans [12]. La solution analytique est déclinée dans [14]. Les éléments suivants peuvent aider à sa compréhension :

- La vitesses de propagation de l'extrémité gauche de l'onde de raréfaction est donnée par la vitesse du son dans l'état thermodynamique initial de la partie gauche du domaine, à savoir $c_1 = \sqrt{\gamma \frac{p_1}{\rho_1}}$.
- La vitesses de propagation de l'onde de choc est donnée par la vitesse du son dans l'état thermodynamique initial de la partie droite du domaine, à savoir $c_2 = \sqrt{\gamma \frac{p_2}{\rho_2}}$.
- A droite de l'onde de choc et à gauche de l'onde de raréfaction, l'information provenant de la discontinuité initiale n'a pas encore été transmise, sa vitesse étant bornée par celle du son. Les états thermodynamiques sont donc inchangés.
- Les états thermodynamiques à droite et à gauche du choc sont liés par la relation de Rankine Hugoniot. Ainsi, notant $\Gamma = \frac{\gamma-1}{\gamma+1}$ et $\beta = \frac{\gamma-1}{2\gamma}$, ρ_g et rho_d les masses volumiques à gauche et à droite du choc, ainsi que p_g et p_d les pressions à gauche et à droite du choc, nous avons :

$$\rho_g = \rho_d \, \frac{p_g + \Gamma \, p_d}{p_d + \Gamma \, p_g}.$$

  Or en vertu du point précédent nous savons que l'état thermodynamique à droite de l'onde de choc n'a pas changé depuis l'état initial, donc :

$$\rho_g = \rho_2 \, \frac{p_g + \Gamma \, p_2}{p_2 + \Gamma \, p_g}.$$

- La discontinuité de contact ne correspondant pas à une onde de choc (sauf à l'instant initial), on a continuité de la pression de part et d'autre. On en déduit que les deux paliers à sa gauche et à sa droite ont pour pression la valeur p_g définie au point précédent.

La valeur de la pression p_g, en revanche, ne s'exprime pas par une relation simple, mais comme solution de l'équation :

$$\left(p_g - p_2\right)\sqrt{\frac{1-\Gamma}{\rho_2\ \left(p_g + \Gamma\ p_2\right)}} = \left(p_1^{\beta} - p_g^{\beta}\right)\sqrt{\frac{\left(1-\Gamma^2\right)\ p_g^{1/\gamma}}{\Gamma^2\ \rho_1}}. \tag{2.47}$$

La masse volumique (dont on déduit la pression grâce à la loi d'état) au niveau de l'onde de raréfaction est donnée par la relation suivante, où l'on a noté x l'abscisse considérée, t le temps, et x_0 la position initiale de la discontinuité de contact (dans notre cas, $x_0 = 0.5$) :

$$\frac{\rho}{\rho_1} = \left(1 - \frac{\gamma-1}{\gamma+1}\ \left(\frac{x-x_0}{c_1\ t} + 1\right)\right)^{\frac{2}{\gamma-1}},$$

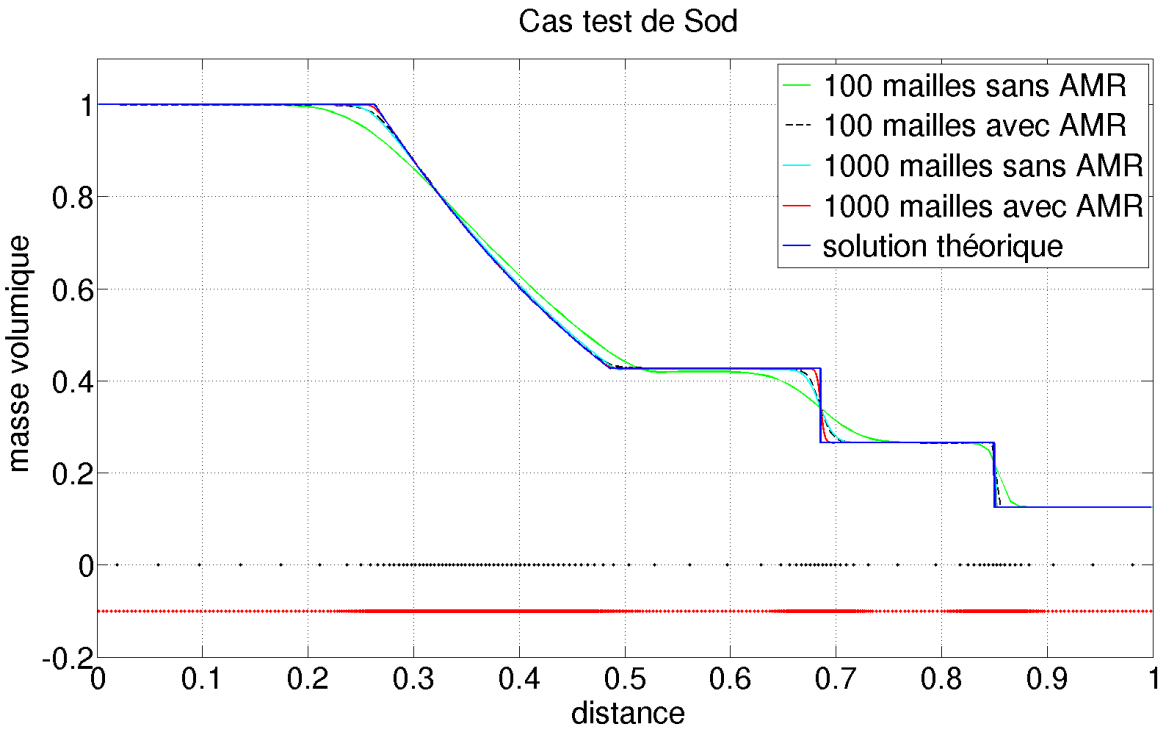
où l'on rappelle que $c_1 = \sqrt{\gamma\frac{p_1}{\rho_1}}$.

Le cas test de Sod a été résolu par la méthode VFFC, avec et sans AMR. Pour les cas avec AMR, la méthode de mouvement de maillage utilisée est une équation différentielle de maillage, plus précisément la méthode *MMPDE directe* présentée dans la sous-section 2.6.1. Par ailleurs, pour obtenir la solution analytique, l'équation (2.47) a été résolue par une méthode de Newton standard. Les résultats obtenus sont les suivants, l'erreur étant calculée par norme $L2$ adimensionnée de la différence des profils de masse volumique théoriques et calculés.

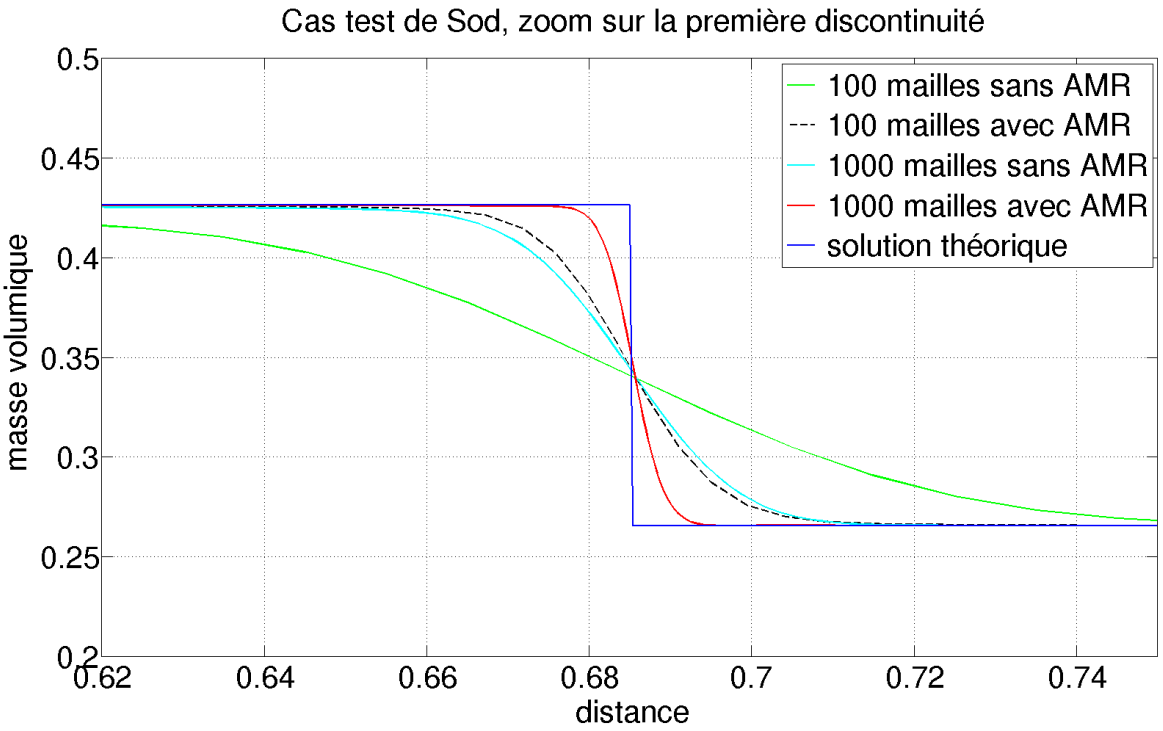
Nombre de cellules	Méthode	erreur L2	Temps CPU en s
100	Sans AMR	0.035	0.49
	Avec AMR	**0.013**	**4.63**
1000	**Sans AMR**	**0.014**	**26.5**
	Avec AMR	0.0071	434

Nous remarquons ici que les cas *AMR 1000 mailles* et *sans AMR 100 mailles* offrent une précision à peu près équivalente (cf. valeurs en gras). Nous pouvons alors conclure que la mise en place de l'AMR nous a fait gagner un facteur 6 en temps de calcul, à précision donnée. Par ailleurs les gains en mémoire consommée atteignent probablement un facteur 10, le nombre de valeurs stockées par cellule étant peu différent avec et sans AMR.

La figure suivante présente les courbes obtenues pour ces quatre cas, ainsi que l'état des maillages AMR au temps final et la solution théorique :



Afin de mieux distinguer les différences entre les cas, la courbe suivante présente les mêmes profils en centrant l'échelle autour de la première discontinuité de masse volumique :



Nous voyons ici que la mise en place de la méthode AMR offre une amélioration significative de la qualité du résultat, à nombre de mailles fixé. Nous nous attacherons

dans le prochain exemple à observer quelle augmentation du nombre d'éléments peut compenser l'absence d'utilisation de la méthode AMR, et à comparer les temps CPU correspondant.

2.13 Cas test de Ransom

Nous avons finalement testé les développements précédents sur le modèle à quatre équations, non conservatif, qui a été présenté dans la sous-section 2.10.3 et considère un mélange de deux fluides en évolution isentropique. En particulier, nous nous sommes intéressés au cas test de Ransom, dont une solution analytique approchée est connue. Il s'agit d'un tube vertical de 12 m, dans lequel un jet de mélange air/eau tombe sous l'action de la gravité. On note :

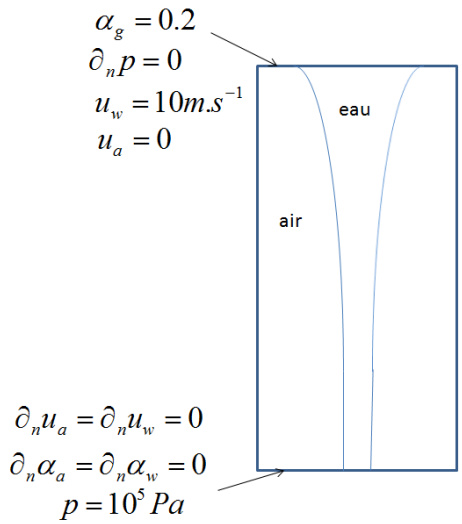
- p la pression.
- α_a et α_w les fractions volumiques d'air et d'eau ($\alpha_a + \alpha_w = 1$).
- u_a et u_w les vitesses de l'air et de l'eau.
- ρ_a et ρ_w les masses volumiques de l'air et de l'eau.

On fixe en haut du domaine une vitesse d'injection d'eau $u_w = 10\ m.s^{-1}$ (vers le bas), une vitesse nulle de l'air ($u_a = 0$), ainsi qu'une fraction volumique d'eau de 0.8 (i.e. $\alpha_a = 0.2$, $\alpha_w = 0.8$). Le fond du tube est ouvert à la pression ambiante : on lui impose $p = 10^5 Pa$. Les variables dont les conditions aux limites n'ont pas été spécifiées ici obéissent à des conditions de Neumann homogènes.

La masse volumique initiale de l'air est $\rho_{a,0} = 1.078 kg.m^{-3}$, celle de l'eau est $\rho_{w,0} = 1000 kg.m^{-3}$. La pression initiale est $p_0 = 10^5 Pa$. La fraction volumique d'eau initiale est $\alpha_{w,0} = 0.8$ (i.e. $\alpha_{a,0} = 0.2$ pour l'air).

Les lois d'état utilisées sont, pour l'eau, une masse volumique constante, et pour l'air, une loi des gaz parfaits isentropique de constante $\gamma = 1.4$. Cette dernière loi peut s'écrire, en notant p et p_0 les pressions courante et initiale, ρ et ρ_0 les masses volumiques courante et initiale : $\frac{p}{p_0} = \left(\frac{\rho}{\rho_0}\right)^{\gamma}$. Par ailleurs, l'intensité de la gravité est prise égale à $9.81 m.s^{-2}$.

Lors de l'état transitoire, l'action de la gravité sur l'eau a davantage d'effet en bas du domaine, car la condition imposée à la limite inférieure, à savoir une pression ambiante, n'offre pas de résistance à la chute. L'eau tombe donc plus vite en bas qu'en haut, ce qui crée un affinement du jet dans la partie basse, et induit le déplacement d'un volume de gaz du haut vers le bas. Finalement, un état stationnaire est atteint, pour lequel la fraction volumique d'air croît du haut vers le bas.



Il est à noter que ce cas sera traité comme un cas monodimensionnel, pour lequel les importances relatives des volumes d'eau et d'air pour une tranche horizontale du tube sont données par les variables α_a et α_w.

En faisant l'hypothèse de l'absence de variations de pression dans le liquide, ce cas dispose d'une solution analytique. Nous nous intéresserons particulièrement dans la suite aux profils de fraction volumique d'air, pour lesquels cette solution s'écrit :

$$\alpha_g(x,t) = \begin{cases} 1 - \frac{\alpha_{w,0}\ u_{w,0}}{\sqrt{u_{w,0}^2 + 2\ g\ (x - x_0)}} & \text{si} \quad x < x_0 + u_{w,0}\ (t - t_0) + \frac{g}{2}\ (t - t_0)^2 \\ 1 - \alpha_{w,0} & \text{sinon} \end{cases} \tag{2.48}$$

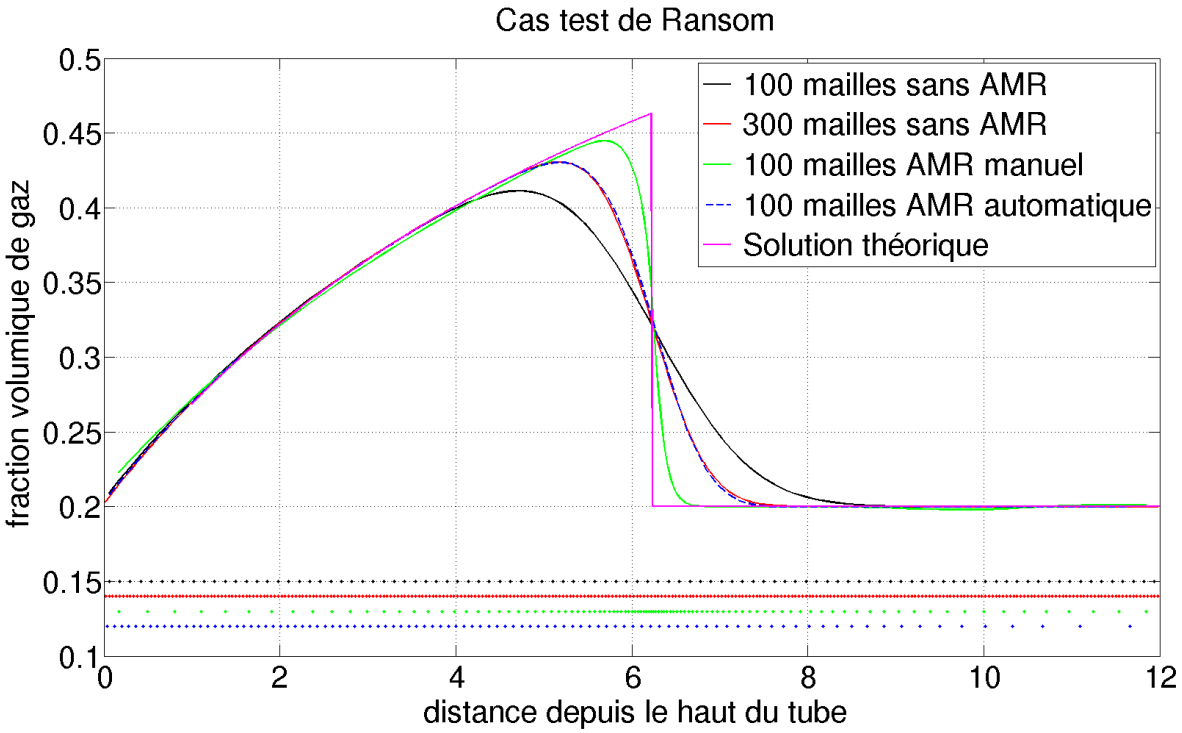
où l'on a noté x_0 l'abscisse de l'extrémité haute du tube, dans notre cas égale à 0. Plus de détails sur ce cas sont disponibles dans [6].

Des tests ont été effectués avec la méthode VFFC couplée aux équations différentielles de maillage, et plus précisément la version *MMPDE inverse* présentée dans la sous-section 2.6.1. Nous avons comparé les résultats avec ceux obtenus lorsque le mouvement des mailles est programmé pour correspondre parfaitement à la solution analytique. Ainsi dans la suite, la dénomination *AMR manuel* indique une adaptation artificielle à la solution du cas test, par opposition à *AMR automatique* qui indique la méthode standard. Pour chacun des tests réalisés, nous nous sommes arrêtés après un temps de 0.5 s. Le temps de calcul a été mesuré et l'erreur L^2 par rapport à la solution théorique calculée.

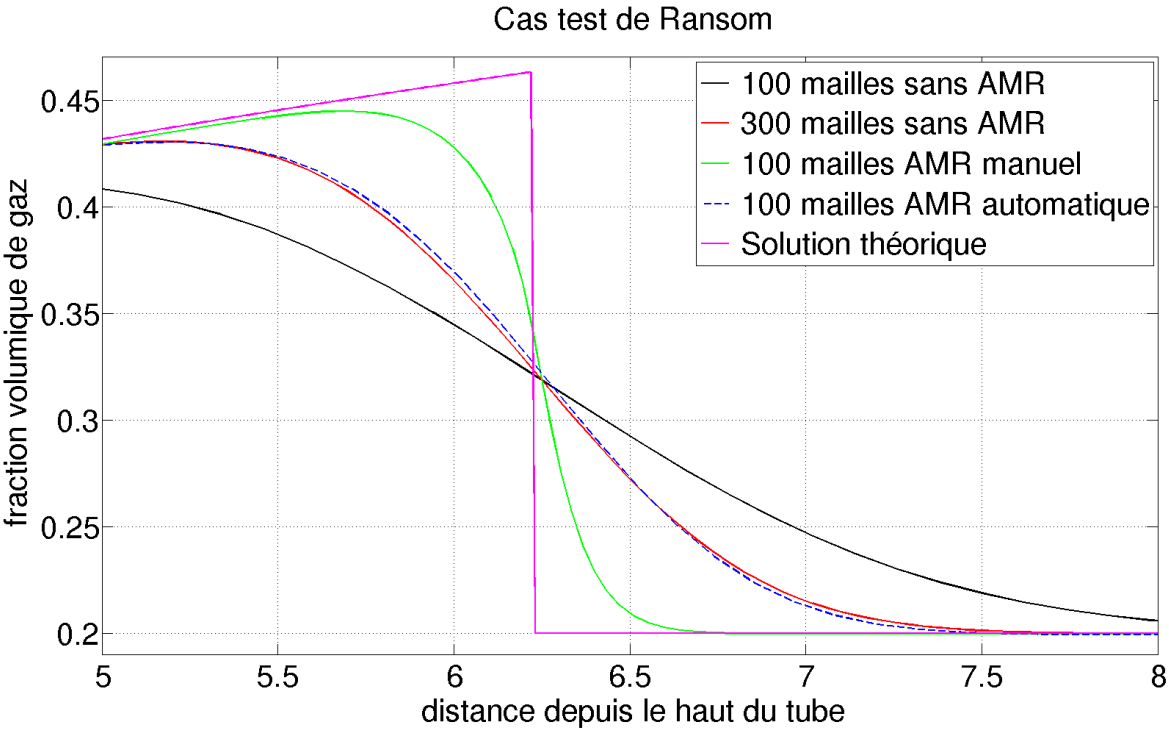
Nombre de cellules	Méthode	erreur L2	Temps CPU en s
100	Sans AMR	0.12	25
	Avec AMR automatique	**0.09**	**117**
	Avec AMR manuel	0.05	235
300	**Sans AMR**	**0.09**	**228**
	Avec AMR automatique	0.063	1332
	Avec AMR manuel	0.037	3265

Nous remarquons notamment que les cas *AMR 100 mailles* et *sans AMR 300 mailles* offrent une précision équivalente, alors que ce dernier cas consomme un temps de calcul environ deux fois plus important (cf. valeurs en gras). Quant à la mémoire utilisée, en considérant qu'elle est proportionnelle au nombre de cellules, la mise en place de l'AMR la réduit d'un facteur 3 à précision donnée. Le temps de calcul indiqué pour le cas avec programmation du mouvement du maillage (AMR manuel) n'est pas à considérer, car la méthode employée pour déterminer l'évolution de ce mouvement, très coûteuse en temps, aurait sans doute pu être optimisée. Par ailleurs la précision obtenue dans ce dernier cas est à analyser avec précaution, étant entendu qu'on ne saurait la reproduire sur un cas quelconque.

La figure ci-dessous présente le profil spatial de la fraction volumique de gaz dans les différents cas, après un temps de $0.5s$. Les différents maillages correspondant à cet instant sont également représentés.



Afin de mieux distinguer les différences entre les cas, la courbe suivante présente les mêmes profils en centrant l'échelle autour de la discontinuité :



Tout cela montre l'intérêt de la mise en œuvre de l'AMR, y compris pour les modèles non conservatifs.

2.14 Conclusion et perspectives

Cette partie avait pour objectif l'adaptation du schéma numérique $VFFC$ au cas des maillages mobiles. Deux problèmes se posent, d'une part la façon dont le maillage doit évoluer, d'autre part la façon dont le schéma numérique doit être modifié pour prendre en compte cette évolution. Pour le premier point, nous avons effectué un travail bibliographique - basé principalement sur [7] - et testé différentes méthodes, pour finalement centrer notre choix sur les équations différentielles de maillage ($MMPDE$).

L'adaptation du schéma numérique $VFFC$ au cas d'un maillage mobile est simple et intuitive dans le cas d'un modèle conservatif. Il s'agit principalement de modifier le flux traversant la face d'une cellule de façon à prendre en compte la matière qu'elle traverse en raison de son mouvement. Au prix d'une légère modification des modèles, on obtient un schéma aussi robuste que celui traitant les cas sans mouvement de maille, et on constate une grande efficacité de la résolution avec maillage mobile sur des cas tests classiques tels celui de Sod.

L'adaptation du schéma numérique $VFFC$ au cas d'un maillage mobile est en revanche plus délicate pour les modèles non conservatifs. Avant d'introduire l'approche globale et théorique présentée ci-dessus, des méthodes plus simples avaient été envisagées, mais aucune ne présentait la robustesse que l'on devait attendre d'une telle méthode. En particulier, nous n'avons pu mener à bien le cas test de Ransom autrement qu'en définissant et calculant finement les nouvelles matrices $E*$ et J^* introduites

dans le cœur de cette partie. Cela nous a amenés, afin de conserver leur *simplicité* et leurs propriétés par rapport aux cas sans maillage mobile, à produire des résultats généraux sur ces matrices, issus de la propriété d'invariance galiléenne des modèles rencontrés.

Mais au delà de ce cas test théorique, l'application que nous visions était l'utilisation d'un schéma AMR non conservatif fiable pour la propagation des ondes de détonation dans les mousses. Pour des raisons que nous avons évoquées dans la sous-section 2.10.3 et que nous reprendrons dans la sous-section 3.2.3, cette application n'a pas pu voir le jour de façon satisfaisante. Pour principale cause de ce constat, citons l'introduction *artificielle* du paramètre π (cf. sous-section 2.10.3) dans le modèle standard, opérée à des fins de stabilité numérique du schéma, qui invalide la propriété d'invariance galiléenne dont nous nous servions pour obtenir des expressions simples de E^* et J^*. Des efforts supplémentaires devront être réalisés sur ce point.

Chapitre 3

Couplage

3.1 Introduction

Ce travail a pour but la simulation efficace de la propagation des ondes de détonation dans les mousses aqueuses, sujet dont l'intérêt a vu le jour lorsqu'il a été constaté qu'un tel matériau disposait de bonnes propriétés d'atténuation, en plus d'une capacité à retenir les produits chimiques dangereux issus d'une explosion. L'écoulement au sein d'une mousse aqueuse est complexe, ce milieu étant composé de trois fluides : l'air, l'eau liquide et sa vapeur. L'atténuation de l'explosion est notamment due au phénomène suivant : l'énergie produite, initialement mécanique, est convertie en chaleur latente et entraîne l'évaporation de l'eau liquide. Alors que seuls des travaux expérimentaux avaient jusqu'à récemment été entrepris pour comprendre ces phénomènes, et plus spécifiquement déterminer la nature de la mousse à mettre en place autour d'un explosif donné afin d'assurer une protection suffisante, un travail de recherche plus approfondi - dans lequel nous nous inscrivons - a vu le jour, avec pour finalité la mise en place d'un outil de simulation précis et efficace [6]. Il s'agit de la première approche numérique visant à représenter finement les interactions entre les trois fluides en présence, y compris les changements de phase, par un modèle multi-fluides avec déséquilibre entre phases. Le modèle créé dans le cadre de ce projet de recherche, appelé dans la suite *modèle à 7 équations*, considère une vitesse et une température propres à chaque phase, les hétérogénéités étant notables dans les conditions extrêmes de température et de pression où nous nous plaçons. Cependant, à distance suffisante de la charge, on peut faire l'hypothèse d'une unique température et d'une unique vitesse. Apparaît alors l'opportunité d'utiliser un modèle simplifié dans ces régions, que nous appellerons *modèle homogène*. Cette partie vise à décrire les deux modèles évoqués, à présenter la mise en œuvre concrète du modèle homogène - qui n'avait pas été opérée avant ces travaux de thèse, contrairement au cas du modèle à 7 équations -, et à proposer une méthodologie pour coupler ces deux modèles dans le but d'une meilleure efficacité de temps de calcul et de mémoire utilisée. Une des difficultés que nous aurons à traiter sera la mobilité de la position pertinente de la frontière entre modèles.

3.2 Modèle à 7 équations

3.2.1 Description du modèle

Le modèle à 7 équations sert à décrire de manière complète et précise l'évolution d'un système composé d'air, d'eau liquide et de vapeur d'eau, le tout en déséquilibre thermodynamique. Les seules hypothèses faites sont l'unicité de la température et de la vitesse pour les formes gazeuses. La phase liquide dispose, en revanche, d'une température et d'une vitesse spécifiques. Nous nous plaçons ici dans le cas monodimensionnel. Nous notons dans la suite :

- α_a, α_w et α_s les fractions volumiques respectives de l'air, de l'eau liquide et de la vapeur d'eau.
- α_g la fraction volumique de gaz, i.e. $\alpha_g = \alpha_a + \alpha_s$.
- ρ_a, ρ_w et ρ_s les masses volumiques respectives de l'air, de l'eau liquide et de la vapeur d'eau.
- ρ_g la masse volumique du gaz, i.e. $\alpha_g\ \rho_g = \alpha_a\ \rho_a + \alpha_s\ \rho_s$.
- $\delta = \alpha_a\ \rho_a - \alpha_s\ \rho_s$.
- p la pression, commune aux trois espèces.
- e_a, e_w et e_s les énergies internes spécifiques respectives de l'air, de l'eau liquide et de la vapeur d'eau.
- E_a, E_w et E_s les énergies totales spécifiques respectives de l'air, de l'eau liquide et de la vapeur d'eau.
- H_a, H_w et H_s les enthalpies totales spécifiques respectives de l'air, de l'eau liquide et de la vapeur d'eau.
- e_g l'énergie interne spécifique du gaz, i.e. $\alpha_g\ \rho_g\ e_g = \alpha_a\ \rho_a\ e_a + \alpha_s\ \rho_s\ e_s$.
- E_g l'énergie totale spécifique du gaz, i.e. $E_g = e_g + \frac{1}{2}\ \rho_g\ u_g^2$.
- H_g l'énergie totale spécifique du gaz, i.e. $H_g = E_g + \frac{p}{\rho_g}$.
- u_w et u_g les vitesses du liquide et du gaz.
- T_w et T_g les températures du liquide et du gaz.

Conservation de la masse Elle s'écrit naturellement pour l'air :

$$\frac{\partial \alpha_a \rho_a}{\partial t} + \frac{\partial \alpha_a \rho_a u_g}{\partial x} = 0. \tag{3.1}$$

Les conservations des masses d'eau liquide et de vapeur d'eau s'écrivent quant à elles, respectivement :

$$\frac{\partial \alpha_w \rho_w}{\partial t} + \frac{\partial \alpha_w \rho_w u_w}{\partial x} = -Q_s, \tag{3.2}$$

$$\frac{\partial \alpha_s \rho_s}{\partial t} + \frac{\partial \alpha_s \rho_s u_g}{\partial x} = Q_s, \tag{3.3}$$

où Q_s désigne la production volumique par unité de temps de vapeur d'eau, sous l'effet de l'évaporation. Le terme Q_s sera dépendant des variables thermodynamiques et du modèle d'évaporation choisi. Il va de soi que si l'on additionne les deux expressions (3.2) et (3.3), ces termes sources s'annulent. Par ailleurs, si l'on additionne (3.1) et (3.3), on obtient :

$$\frac{\partial \alpha_g \rho_g}{\partial t} + \frac{\partial \alpha_g \rho_g u_g}{\partial x} = Q_s. \tag{3.4}$$

On peut également écrire par soustraction des expressions (3.1) et (3.3) :

$$\frac{\partial \delta}{\partial t} + \frac{\partial \delta u_g}{\partial x} = -Q_s. \tag{3.5}$$

Dans l'établissement du modèle, afin de rendre non dégénérés les cas où la fraction volumique de vapeur d'eau est nulle, nous substituerons aux expressions (3.1) et (3.3) les expressions (3.4) et (3.5).

Conservation de la quantité de mouvement Nous rappelons que nous considérons une unique vitesse pour les deux espèces en phase gazeuse, ce qui nous permet de n'écrire qu'une équation pour cette phase :

$$\frac{\partial \alpha_g \rho_g u_g}{\partial t} + \frac{\partial \alpha_g \rho_g u_g^2}{\partial x} + \alpha_g \ \frac{\partial p}{\partial x} = Q_s \ u_i + C_{drag} \ (u_w - u_g). \tag{3.6}$$

La variable u_i désigne une vitesse d'interface entre le gaz et le liquide, classiquement prise égale à $\alpha_g \ u_g + \alpha_w \ u_w$. Nous considérons qu'un volume de liquide en cours d'évaporation évolue à cette vitesse, ce qui explique qu'il faille ajouter la quantité $Q_s \ u_i$ à la quantité de mouvement globale du gaz. Le terme $C_{drag} \ (u_w - u_g)$ correspond à une perte ou un gain de quantité de mouvement liés au freinage ou à l'entraînement du gaz par le liquide. La valeur de C_{drag} est positive, dépend des variables thermodynamiques du problème et est en général proportionnelle à la valeur absolue de la différence entre la vitesse du gaz et la celle du liquide [6]. Nous remarquons que, de même que pour le modèle à quatre équations - dont nous avons fait connaissance dans la section 2.10 -, la fraction volumique de gaz présente en dehors de la dérivée spatiale rend le système non conservatif. De façon symétrique, nous pouvons écrire pour le liquide :

$$\frac{\partial \alpha_g \rho_g u_g}{\partial t} + \frac{\partial \alpha_g \rho_g u_g^2}{\partial x} + \alpha_g \ \frac{\partial p}{\partial x} = -Q_s \ u_i + C_{drag} \ (u_g - u_w), \tag{3.7}$$

et nous notons une fois de plus que les termes sources s'annulent naturellement lorsqu'on additionne les expressions de conservation de quantité de mouvement (3.6) et (3.7).

Conservation de l'énergie Comme pour la vitesse, nous ne considérons qu'une température pour les deux espèces en phase gazeuse, ce qui nous permet de n'écrire qu'une équation de conservation de l'énergie pour cette phase :

$$\frac{\partial \alpha_g \rho_g u_g}{\partial t} + \frac{\partial \alpha_g \rho_g \ H_g \ u_g}{\partial x} + p \ \frac{\partial \alpha_g}{\partial t} = Q_s \left(h_{is} + \frac{u_i^2}{2} \right) + C_{drag} \ (u_w - u_g) \ u_i + Q_{is}. \tag{3.8}$$

Le terme $Q_s \left(h_{is} + \frac{u_i^2}{2} \right)$ correspond à l'énergie apportée à la phase gazeuse par un éventuel volume de liquide évaporé. Le terme $\frac{u_i^2}{2}$ est lié à l'énergie cinétique de ce volume évaporé, tandis que le terme h_{is} est lié à la fois à son énergie interne et au travail mécanique engendré par l'évaporation. Il dépend des variables thermodynamiques du problème et du modèle d'évaporation considéré. Le terme h_{is} correspond en général à l'enthalpie interne massique de saturation de la vapeur d'eau, c'est à dire celle obtenue à partir des lois d'état en maintenant la pression telle quelle et en considérant la température de saturation correspondant à cette pression. Le terme $C_{drag} \ (u_w - u_g) \ u_i$ correspond à l'énergie cinétique gagnée ou perdue par l'action de la force de traînée, elle même induite par la différence de vitesse entre le liquide et le gaz. Enfin, le terme

Q_{is} correspond à un terme d'échange de chaleur entre les deux phases, dépendant des variables thermodynamiques du problème. De façon symétrique, nous pouvons écrire la conservation de l'énergie pour la phase liquide :

$$\frac{\partial \alpha_w \rho_w u_w}{\partial t} + \frac{\partial \alpha_w \rho_w \ H_w \ u_w}{\partial x} + p \ \frac{\partial \alpha_w}{\partial t} = \\ - Q_s \left(h_{iw} + \frac{u_i^2}{2} \right) + C_{drag} \ (u_g - u_w) \ u_i + Q_{iw}. \quad (3.9)$$

De même que pour la vapeur d'eau, le terme h_{iw} correspond en général à l'enthalpie interne massique de saturation de l'eau liquide, et Q_{iw} correspond à un terme d'échange de chaleur dépendant des variables thermodynamiques du problème. Ainsi le modèle à 7 équations peut finalement s'écrire sous cette forme :

$$\frac{\partial \delta}{\partial t} + \frac{\partial (\delta u_g)}{\partial x} = -Q_s, \quad (3.10)$$

$$\frac{\partial (\alpha_g \rho_g)}{\partial t} + \frac{\partial (\alpha_g \rho_g u_g)}{\partial x} = Q_s, \quad (3.11)$$

$$\frac{\partial (\alpha_w \rho_w)}{\partial t} + \frac{\partial (\alpha_w \rho_w u_w)}{\partial x} = -Q_s, \quad (3.12)$$

$$\frac{\partial (\alpha_g \rho_g u_g)}{\partial t} + \frac{\partial (\alpha_g \rho_g u_g^2)}{\partial x} + \alpha_g \frac{\partial p}{\partial x} = Q_s u_i + C_{drag}(u_w - u_g), \quad (3.13)$$

$$\frac{\partial (\alpha_w \rho_w u_w)}{\partial t} + \frac{\partial (\alpha_w \rho_w u_w^2)}{\partial x} + \alpha_w \frac{\partial p}{\partial x} = -Q_s u_i + C_{drag}(u_g - u_w), \quad (3.14)$$

$$\frac{\partial (\alpha_g \rho_g E_g)}{\partial t} + \frac{\partial (\alpha_g \rho_g H_g u_g)}{\partial x} + p \frac{\partial \alpha_g}{\partial t} = \\ Q_s \big(h_{is} + \frac{|u_i|^2}{2} \big) + C_{drag}(u_w - u_g).u_i + Q_{is}, \quad (3.15)$$

$$\frac{\partial (\alpha_w \rho_w E_w)}{\partial t} + \frac{\partial (\alpha_w \rho_w H_w u_w)}{\partial x} + p \frac{\partial \alpha_w}{\partial t} = \\ -Q_s \big(h_{iw} + \frac{|u_i|^2}{2} \big) + C_{drag}(u_g - u_w).u_i + Q_{iw}, \quad (3.16)$$

système que nous supposons fermé par trois lois d'état correspondant aux trois espèces ($\rho_a(p,T)$, $e_a(p,T)$, $\rho_w(p,T)$, $e_w(p,T)$, $\rho_s(p,T)$, $e_s(p,T)$), ainsi que par les expressions des termes sources (Q_s, C_{drag}, Q_{is}, Q_{iw}) en fonction des variables thermodynamiques du problème.

3.2.2 Résolution du modèle

Nous avons remarqué que le modèle à 7 équations n'était pas conservatif, en raison des termes de dérivées de pression dans les équations de conservation de la quantité de mouvement et de l'énergie. Ce modèle peut se formaliser ainsi :

$$\frac{\partial v}{\partial t} + \frac{\partial F(v)}{\partial x} + \tilde{C}(v) \ \frac{\partial F(v)}{\partial x} + D(v) \ \frac{\partial v}{\partial t} = \tilde{S}(v). \quad (3.17)$$

La matrice $\tilde{C}(v)$ joue le même rôle que la matrice $C(v)$ du modèle à quatre équations (cf. sous-section 2.10.3), à savoir la prise en compte des termes $\alpha_g \frac{\partial p}{\partial x}$ et $\alpha_w \frac{\partial p}{\partial x}$ dans les

équations de conservation de la quantité de mouvement. L'introduction de la matrice $D(v)$, qui n'apparaissait pas dans la formalisation du modèle à quatre équations, est nécessaire pour prendre en compte les termes $p\frac{\partial \alpha_g}{\partial t}$ et $p\frac{\partial \alpha_w}{\partial t}$ dans les équations de conservation d'énergie. Le vecteur v s'écrit :

$$v = (\delta, \alpha_g\rho_g, \alpha_w\rho_w, \alpha_g\rho_g u_g, \alpha_w\rho_w u_w, \alpha_g\rho_g E_g, \alpha_w\rho_w E_w).$$

Le vecteur $\tilde{S}(v)$ contient les termes sources. Nous notons $\tilde{S}(v) = (S_1, S_2, S_3, S_4, S_5, S_6, S_7)$ avec :

- $S_1 = -Q_s$,
- $S_2 = Q_s$,
- $S_3 = -Q_s$,
- $S_4 = Q_s u_i + C_{drag}(u_w - u_g)$,
- $S_5 = -Q_s u_i + C_{drag}(u_g - u_w)$,
- $S_6 = Q_s\big(h_{is} + \frac{|u_i|^2}{2}\big) + C_{drag}(u_w - u_g).u_i + Q_{is}$,
- $S_7 = -Q_s\big(h_{iw} + \frac{|u_i|^2}{2}\big) + C_{drag}(u_g - u_w).u_i + Q_{iw}$.

La matrice $D(v)$ s'écrit :

$$D(v) = p \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{g,1} & \alpha_{g,2} & \alpha_{g,3} & \alpha_{g,4} & \alpha_{g,5} & \alpha_{g,6} & \alpha_{g,7} \\ \alpha_{w,1} & \alpha_{w,2} & \alpha_{w,3} & \alpha_{w,4} & \alpha_{w,5} & \alpha_{w,6} & \alpha_{w,7} \end{pmatrix}, \tag{3.18}$$

où nous avons noté, pour i allant de 1 à 7, $\alpha_{g,i}$ la dérivée de α_g par rapport à la i-ème coordonnée de v, et $\alpha_{w,i}$ la dérivée de α_w par rapport à cette même coordonnée. En ce qui concerne la matrice $\tilde{C(v)}$, de même que pour la matrice $C(v)$ dans le modèle à quatre équations, elle n'est pas unique mais dépend du choix du flux $F(v)$. De même que pour le modèle à quatre équations, nous introduirons dans le flux $F(v)$ un paramètre $\pi(t)$ ne dépendant que du temps, dont nous ajusterons la valeur après coup pour assurer de bonnes propriétés d'inversion à la matrice $J(v) = \frac{\partial F(v)}{\partial v}$. Ainsi, nous écrivons :

$$F(v) = (\delta u_g, \alpha_g\rho_g u_g, \alpha_w\rho_w u_w, \alpha_g(\rho_g u_g^2 + p - \pi), \alpha_w(\rho_w u_w^2 + p - \pi), \alpha_g\rho_g H_g u_g, \alpha_w\rho_w H_w u_w),$$

et nous avons en conséquence :

$$\tilde{C(v)} = -(p - \pi) \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{g,1} & \alpha_{g,2} & \alpha_{g,3} & \alpha_{g,4} & \alpha_{g,5} & \alpha_{g,6} & \alpha_{g,7} \\ \alpha_{w,1} & \alpha_{w,2} & \alpha_{w,3} & \alpha_{w,4} & \alpha_{w,5} & \alpha_{w,6} & \alpha_{w,7} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.19}$$

Nous allons maintenant transformer l'équation (3.17) en une équation de la forme générale évoquée dans la présentation de la méthode VFFC pour les systèmes non

conservatifs (cf. section 2.7). Pour cela, nous introduisons les matrices $C(v)$ et $S(v)$ définies par :

$$C(v) = \big[I_d + D(v)\big]^{-1}\big[J(v) + \tilde{C(v)}\big] - J(v), \tag{3.20}$$

$$S(v) = \big[I_d + D(v)\big]^{-1}\tilde{S}(v), \tag{3.21}$$

ce qui permet de réécrire l'équation (3.17) sous la forme :

$$\frac{\partial v}{\partial t} + \frac{\partial F(v)}{\partial x} + C(v)\ \frac{\partial F(v)}{\partial x} = S(v). \tag{3.22}$$

Il nous est ainsi possible de résoudre ce modèle par la méthode VFFC (cf. sous-section 1.3.1).

Mais avant toute chose il nous faut choisir un paramètre π qui rende la matrice $J(v)$ inversible, ce qui peut se faire de même que pour le modèle à quatre équations (cf. sous-section 2.10.3), en déterminant d'abord dans chaque cellule un paramètre π adapté au vecteur v local, c'est à dire assurant à la matrice $J(v)$ un déterminant de valeur absolue suffisamment grande, puis en considérant le minimum de ces paramètres π locaux pour en déduire un paramètre π global (les détails sont donnés dans [6]). Nous rappelons que la détermination d'un unique paramètre π pour l'ensemble des cellules est nécessaire, seule la dépendance en temps de ce paramètre étant autorisée pour que son introduction ne modifie pas le système d'équations initial. Comme pour le modèle à quatre équations, cette possibilité de considérer le minimum des paramètres π locaux pour en déduire un paramètre π global provient du caractère affine et décroissant de la fonction associant le déterminant de $J(v)$ au paramètre π local.

Reprenant le même formalisme que pour les précédents modèles, nous introduisons maintenant :

$$w = (\alpha_s, \alpha_a, u_g, u_w, e_w, T_g, p).$$

L'ensemble des variables thermodynamiques du problème ne se déduisant simplement que des coordonnées du vecteur w, il nous faut pouvoir le déterminer à partir de la donnée de v, seul vecteur réactualisé par l'application du schéma numérique VFFC (cf. sous-section 1.3.1). Le passage du vecteur v au vecteur w se fait par le biais d'un système d'équations non linéaires faisant intervenir les lois d'état. Les détails sont également donnés dans [6]. Par ailleurs, afin de déterminer l'expression analytique des matrice $J(v)$, $\tilde{C}(v)$ et $D(v)$, il nous faut connaître les expressions des dérivées des coordonnées du vecteur w par rapport à celles du vecteur v. Comme pour le modèle à quatre équations (cf. sous-section 2.10.3), il suffit de calculer et d'inverser la matrice $\frac{\partial v}{\partial w}$ contenant les dérivées des coordonnées de v par rapport à celles de w. Les expressions résultantes dépendront des lois d'état et de leurs dérivées.

3.2.3 Invariance galiléenne et AMR

Bien que ce ne soit pas l'objet de cette partie sur le couplage de modèles, posons nous ici la question de savoir si le modèle à 7 équations entre dans le cadre de la théorie présentée dans la section 2.8 sur la mise en place de l'AMR pour les modèles non conservatifs. Il est clair que la transformation de l'équation (3.17) en (3.22) nous permet d'écrire le schéma numérique prenant en compte le mouvement du maillage, notamment à l'aide de nouvelles matrices $J^*(v)$ et $C^*(v)$. En revanche, l'écriture analytique de ces

deux matrices à partir de leurs équivalents sans AMR, $J(v)$ et $C(v)$, pose davantage de difficultés. En effet, l'étude que nous avons réalisé dans la section 2.9, qui nous a permis d'obtenir ces écritures analytiques pour les modèles présentés dans la section 2.10, a pu être menée à bien grâce à l'hypothèse selon laquelle :

$$M(\lambda) \circ \big(f - \lambda\ g\big) = f \circ h_\lambda,$$

où nous rappelons que :

- λ désigne la vitesse de maillage.
- f désigne l'application qui à w associe F,
- g désigne la bijection qui à w associe v,
- h_λ désigne la bijection qui à w associe le vecteur obtenu à partir de w en soustrayant λ aux composantes correspondant à des vitesses,
- $M(\lambda)$ définit la matrice de l'application $g \circ h_\lambda \circ g^{-1}$.

Nous avons vu que cette propriété, vérifiée pour nos précédents modèles (cf. section 2.10), découlait de leur propriété d'invariance galiléenne. Elle n'est plus vraie pour le modèle à 7 équations présenté dans [6] et repris dans la section 3.2, pour la raison suivante : le paramètre π vient se soustraire à la pression dans les deux équations de quantité de mouvement, comme c'était le cas pour le modèle à quatre équations, mais il n'intervient pas dans les termes de pression des équations de conservation de l'énergie - équations qui n'étaient pas considérées dans le modèle à quatre équations -, ce qui crée une incohérence. La bonne manière de résoudre ce problème, et de rendre le flux consistant avec la propriété d'invariance galiléenne, est d'introduire le paramètre π partout où intervient la pression, c'est à dire d'écrire :

$$F(v) = \Big(\delta u_g, \alpha_g\rho_g u_g, \alpha_w\rho_w u_w, \alpha_g(\rho_g u_g^2 + p - \pi), \alpha_w(\rho_w u_w^2 + p - \pi),$$

$$\alpha_g\rho_g\Big(e_g + \frac{p-\pi}{\rho_g}\Big)u_g, \alpha_w\rho_w\Big(e_w + \frac{p-\pi}{\rho_w}\Big)u_w\Big).$$

Alors la matrice $M(\lambda)$ existe (cf. section 2.9), vérifie les hypothèses précédemment évoquées, et s'écrit :

$$M(\lambda) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -\lambda & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -\lambda & 0 & 1 & 0 & 0 \\ 0 & \frac{\lambda^2}{2} & 0 & -\lambda & 0 & 1 & 0 \\ 0 & 0 & \frac{\lambda^2}{2} & 0 & -\lambda & 0 & 1 \end{pmatrix}. \tag{3.23}$$

Malheureusement, cette nouvelle expression du flux complexifie considérablement le calcul du paramètre π permettant l'inversibilité de la matrice $J(v)$ (ou $J^*(v)$ pour les cas avec AMR). Nous avons toujours la propriété remarquée dans le cadre du modèle à quatre équations (cf. sous-section 2.10.3) selon laquelle :

$$\pi^*(\omega, \lambda) = \pi_0\big(h_\lambda(\omega)\big),$$

où nous rappelons que π_0 désigne la fonction associant à ω la valeur du paramètre π local dans le cas sans AMR, tandis que π^* associe ce même paramètre au couple (ω, λ)

dans le cas avec AMR. En effet le même raisonnement s'applique pour la décliner. La difficulté ne provient donc pas de l'expression de π^*, mais de celle de π_0. En effet, étant donné que nous avons introduit le paramètre π dans les termes de flux d'énergie, l'expression liant le déterminant de J au paramètre π local est nettement plus complexe, et en particulier perd son caractère affine. Ainsi, même s'il est toujours possible de déterminer localement un paramètre π assurant l'inversibilité de J, il n'existe pas à notre connaissance de méthode fiable et efficace pour déduire des paramètres π locaux un paramètre π global assurant à la matrice J de bonnes propriétés d'inversibilité sur l'ensemble du domaine.

Les considérations que nous avons faites dans les sections 2.9 et 2.10, afin de profiter des propriétés d'invariance galiléenne des modèles pour en déduire des modifications simples permettant la prise en compte de l'AMR (notamment dans le calcul des matrices E, J et du paramètre π), ne s'appliquent donc pas facilement au modèle à 7 équations. Il serait certes possible, en théorie, de conserver la version standard du modèle à 7 équations (i.e. ne contenant le paramètre π que dans les équations de conservation de la quantité de mouvement) et de décliner l'expression des matrices E^* et J^* et du paramètre π avec AMR directement, sans passer par leurs expressions sans AMR. Cependant l'apparition de la nouvelle variable λ, correspondant à la vitesse de déplacement des mailles, rendrait ces expressions très complexes. En plus de la difficulté induite par la réalisation du calcul formel, il serait à craindre que le problème lié à la détermination du paramètre π global ne soit pas pour autant résolu, la possibilité de déduire un paramètre π global des paramètres π locaux étant d'autant moins certaine que l'expression du paramètre π local est complexe.

Si ce travail sur l'invariance galiléenne ne permet pas pour l'instant la mise en place d'un méthode AMR robuste pour le modèle à 7 équation, il a permis de donner un argument pour modifier le flux $F(v)$ du modèle standard en introduisant le paramètre π dans ses deux dernières coordonnées. Cela donnerait à ce flux une propriété d'invariance galiléenne qui, si nous n'avons pas pu l'exploiter efficacement dans le cas de l'AMR, pourrait avoir d'autres intérêts encore méconnus.

3.3 Modèle homogène à 5 équations

Comme expliqué en introduction, le modèle à 7 équations considère une température et une vitesse propres à chaque phase, ce qui est nécessaire pour la modélisation correcte de la propagation d'une onde de détonation dans la mousse, étant donné les fortes hétérogénéités rencontrées dans les zones de température et pression extrêmes. Cependant, si l'on se place suffisamment loin de la charge, on retrouve une homogénéité des vitesses et des températures dans les deux phases, et le système physique peut alors être résolu beaucoup plus simplement. Nous allons voir de quelle manière on peut mettre en place un modèle simplifié pour cette zone *homogène* et comment on peut coupler les deux modèles, étant entendu qu'on ne sait pas a priori où se situe la *bonne* frontière séparant spatialement les zones nécessitant ou non la prise en compte des hétérogénéités.

3.3.1 Description du modèle

Nous allons décrire un modèle à une seule température pour le liquide et le gaz, et au nombre d'équations réduit. Ce modèle est décrit dans [8] et utilisé dans [9]. Dans un premier temps, nous considérons que les vitesses du liquide et du gaz ne sont pas nécessairement les mêmes, mais que leur différence u_r est définie en fonction des variables thermodynamiques du problème. Nous définissons :

- $\rho = \alpha_g\rho_g + \alpha_w\rho_w$
- $c_g = \frac{\alpha_g\rho_g}{\rho}$
- $c_w = \frac{\alpha_w\rho_w}{\rho}$
- $u = c_g u_g + c_w u_w$
- $E = c_g E_g + c_w E_w$
- $H = c_g H_g + c_w H_w$
- $c_r = c_g - c_w$
- $u_r = u_g - u_w$
- $h_r = h_g - h_w$.

Alors, en réarrangeant les 7 équations du modèle complet, nous pouvons écrire :

(3.10)
$$\frac{\partial\delta}{\partial t} + \frac{\partial(\delta u_g)}{\partial x} = -Q_s, \tag{3.24}$$

(3.11)+(3.12)
$$\frac{\partial\rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} = 0, \tag{3.25}$$

(3.11)-(3.12)
$$\frac{\partial(\rho c_r)}{\partial t} + \frac{\partial(\rho c_r u + \rho\frac{1-c_r^2}{2}u_r)}{\partial x} = 2Q_s, \tag{3.26}$$

(3.13)+(3.14)
$$\frac{\partial(\rho u)}{\partial t} + \frac{\partial(\rho u^2 + \rho\frac{1-c_r^2}{4}u_r^2)}{\partial x} + \frac{\partial p}{\partial x} = 0, \tag{3.27}$$

(3.15)+(3.16)
$$\frac{\partial(\rho E)}{\partial t} + \frac{\partial\big(\rho H u + \rho(h_r - \frac{c_r}{2}u_r^2 + u.u_r)\frac{1-c_r^2}{4}u_r\big)}{\partial x} = Q_s(h_{is} - h_{iw}) + Q_{is} + Q_{iw}. \tag{3.28}$$

Ainsi, nous aboutissons à un modèle à 5 équations conservatif, dont le vecteur des variables conservatives s'écrit $v = (\delta, \rho, \rho c_r, \rho u, \rho E)$ et dont le vecteur flux s'écrit :

$$F(v) = \begin{pmatrix} \delta u_g \\ \rho u \\ \rho c_r u + \rho\frac{1-c_r^2}{2}u_r \\ \rho u^2 + \rho\frac{1-c_r^2}{4}u_r^2) + p \\ \rho u H + \rho(h_r - \frac{c_r}{2}u_r^2 + u.u_r)\frac{1-c_r^2}{4}u_r) \end{pmatrix}. \tag{3.29}$$

Pour nous assurer que le modèle soit consistant, il reste à vérifier que nous pouvons bien obtenir le vecteur flux $F(v)$ à partir de la seule donnée du vecteur v. Comme pour les modèles précédents, nous allons utiliser comme intermédiaire un vecteur des variables physiques, en l'occurrence $w = (p, T, \alpha_s, \alpha_a, u)$. Considérons que nous avons à notre connaissance le vecteur w. Grâce à la relation $\alpha_w = 1 - \alpha_s - \alpha_a$, nous connaissons les fractions volumiques des trois espèces. Par ailleurs, étant donné que nous n'avons qu'une seule température, et par l'intermédiaire des lois d'état, la donnée de p et T nous donne les masses volumiques et énergies internes des trois espèces. Nous avons donc à notre disposition l'ensemble des variables thermodynamiques du problème, ce qui par hypothèse nous permet d'obtenir u_r. La variable u est présente dans le vecteur w, et la variable u_g se déduit immédiatement de u et u_r. Toutes ces variables permettent d'obtenir aisément les vecteurs v et $F(v)$. Maintenant, étant donné que les vecteurs v et w ont la même taille et que v se déduit de manière unique de w, on peut prédire - ce que nous vérifierons dans la sous-section 3.3.2 - que w, et donc $F(v)$, se déduisent de v. Nous sommes donc bien en mesure de résoudre les équations (3.24) à (3.28), qui forment un système conservatif, par la méthode VFFC (cf. sous-section 1.3.1).

3.3.2 Résolution dans le cas $u_r = 0$

Les équations deviennent :

$$\frac{\partial \delta}{\partial t} + \frac{\partial(\delta u_g)}{\partial x} = -Q_s, \tag{3.30}$$

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} = 0, \tag{3.31}$$

$$\frac{\partial(\rho c_r)}{\partial t} + \frac{\partial(\rho c_r u)}{\partial x} = 2Q_s, \tag{3.32}$$

$$\frac{\partial(\rho u)}{\partial t} + \frac{\partial(\rho u^2 + p)}{\partial x} = 0, \tag{3.33}$$

$$\frac{\partial(\rho E)}{\partial t} + \frac{\partial(\rho H u)}{\partial x} = Q_s(h_{is} - h_{iw}) + Q_{is} + Q_{iw}. \tag{3.34}$$

Cet ensemble d'équations peut se réécrire sous la forme :

$$\frac{\partial v}{\partial t} + \frac{\partial F(v)}{\partial x} = S(v) \tag{3.35}$$

$$\text{avec } v = \begin{pmatrix} \delta \\ \rho \\ \rho c_r \\ \rho u \\ \rho E \end{pmatrix}, \; F(v) = \begin{pmatrix} \delta u \\ \rho u \\ \rho c_r u \\ \rho u^2 + p \\ \rho u H \end{pmatrix} \text{ et } S(v) = \begin{pmatrix} -Q_s \\ 0 \\ 2\ Q_s \\ 0 \\ Q_s(h_{is} - h_{iw}) + Q_{is} + Q_{iw} \end{pmatrix}.$$

Le calcul formel donne, en rappelant que $J(v) = \frac{\partial F(v)}{\partial v}$:

$$J(v) = \begin{pmatrix} u & -\frac{\delta u}{\rho} & 0 & \frac{\delta}{\rho} & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & -C_r u & u & C_r & 0 \\ d_{p_0} & -u^2 + d_{p_1} & d_{p_2} & 2u + d_{p_3} & d_{p_4} \\ u\ d_{p_0} & -uH + u\ d_{p_1} & u\ d_{p_2} & H + u\ d_{p_3} & u + u\ d_{p_4} \end{pmatrix} \tag{3.36}$$

où l'on note, pour i allant de 1 à 5, v_i la i-ème composante de v, et $d_{p_i} = \frac{\partial P}{\partial v_i}$. Nous verrons dans la suite comment calculer ces dernières valeurs. Nous allons maintenant nous employer à déterminer le vecteur $w = (p, T, \alpha_s, \alpha_a, u)$ à partir du vecteur v, la donnée du vecteur w permettant d'accéder à l'ensemble des autres variables, et notamment de calculer le flux $F(v)$. Nous allons commencer par calculer la pression et la température à partir des variables conservatives, les autres variables de w s'en déduisant plus facilement. Nous notons :

$$y = (\alpha_a \rho_a, \alpha_s \rho_s, \alpha_w \rho_w, \rho e) = (y_1, y_2, y_3, y_4).$$

Alors les deux relations :

- $\alpha_a\ \rho_a\ e_a + \alpha_s\ \rho_s\ e_s + \alpha_w\ \rho_w\ e_w = \rho e$
- $\alpha_a + \alpha_s + \alpha_w = 1$

peuvent s'écrire, en introduisant la fonction :

$$F(y, p, T) = \begin{pmatrix} y_1\ e_a(p,T) + y_2\ e_s(p,T) + y_3\ e_w(p,T) - y_4 \\ \frac{y_1}{\rho_a(p,T)} + \frac{y_2}{\rho_s(p,T)} + \frac{y_3}{\rho_w(p,T)} - 1 \end{pmatrix},$$

de la manière suivante :

$$F(y, P, T) = 0. \tag{3.37}$$

Par une méthode de Newton, on peut donc obtenir les pression et température p et T en fonction de y. De plus, en notant $X = (p, T)$ et en différentiant la relation (3.37) par rapport à y_i pour i entre 1 et 4, nous obtenons :

$$\frac{\partial F}{\partial y_i} + \frac{\partial F}{\partial X}\ \frac{\partial X}{\partial y_i} = 0 \tag{3.38}$$

et l'inversion d'un système linéaire à deux équations et deux inconnues nous donne $\frac{\partial p}{\partial y_i}$ et $\frac{\partial T}{\partial y_i}$.

Par ailleurs nous avons $y = A(v)\ v$ avec $A(v) = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ -\frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2}\ u^2 & 0 & -u & 1 \end{pmatrix}$.

La vitesse u seule variable présente dans les coordonnées de $A(v)$, s'exprime très facilement en fonction de v, en divisant sa quatrième coordonnée par sa deuxième coordonnée. Cela nous donne les variables de y et par conséquent les pression et température p et T en fonction de v, ce qui permet d'obtenir facilement les variables physiques du vecteur w et, par leur intermédiaire, les coordonnées du flux $F(v)$. En effet, il ne nous manque plus pour compléter les variables physiques que les données de α_s, α_a et u.

Connaissant maintenant les pression et température p et T, les lois d'état nous donnent l'ensemble des variables thermodynamiques, en particulier les masses volumiques ρ_a, ρ_s et ρ_w. Alors les trois premières coordonnées du vecteur y - qui sont également à notre connaissance, comme nous venons de le voir - nous permettent d'en déduire immédiatement les fractions volumiques des trois espèces. Celles-ci nous permettent de calculer aisément la masse volumique globale $\rho = \alpha_a\ \rho_a + \alpha_s\ \rho_s + \alpha_w\ \rho_w$, qui elle même nous donne la vitesse u grâce à la quatrième composante du vecteur v. Nous avons donc entièrement déterminé le vecteur w à partir du vecteur des variables conservatives v.

Enfin, nous pouvons calculer les dérivées de p et T par rapport à v en utilisant :

$$\frac{\partial p}{\partial v} = A^t\ \frac{\partial p}{\partial y}, \tag{3.39}$$

$$\frac{\partial T}{\partial v} = A^t\ \frac{\partial T}{\partial y}. \tag{3.40}$$

L'équation (3.39), en particulier, nous donne d_{p_1}, d_{p_2}, d_{p_3}, d_{p_4} et d_{p_5}, et permet donc la détermination de $\frac{\partial F(v)}{\partial v}$ dont l'expression est donnée dans l'équation (3.36). Nous sommes maintenant en mesure, à partir du vecteur v, de déterminer le flux $F(v)$ et sa différentielle $J(v)$, ce qui nous permet d'utiliser la méthode VFFC (cf. sous-section 1.3.1).

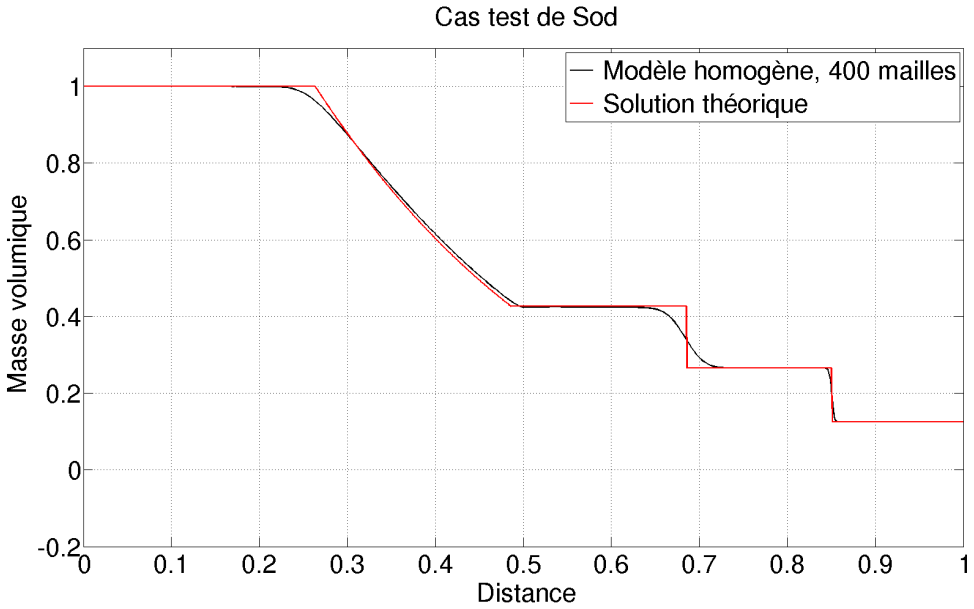
3.4 Validation et test du modèle homogène

Le modèle homogène ne considérant qu'une seule température et qu'une seule vitesse, il est nécessairement approximatif dans le cas général, de même que le modèle à 7 équations, dans une moindre mesure, ce dernier ne considérant qu'une température et qu'une vitesse pour la phase gazeuse. Deux questions sont posées ici, portant d'une part sur la correcte résolution du système d'équations par la méthode présentée ci-dessus, d'autre part par la capacité de ce modèle à approcher correctement la physique d'un cas complexe ne présentant pas de trop fortes hétérogénéités. Nous répondrons à la première grâce au cas test de Sod, à la deuxième grâce à une confrontation avec le modèle à 7 équations.

3.4.1 Cas test de Sod

Nous utilisons à nouveau le cas test de Sod, dont nous rappelons que la solution théorique est connue. Nous renvoyons le lecteur à la section 2.12 pour les détails. Ce cas ne considère qu'une seule espèce, affectée de conditions thermodynamiques très différentes de part et d'autre du milieu du domaine. Le modèle homogène considère trois espèces, mais rien n'empêche de leur définir des lois d'état identiques, ce qui devra créera la même évolution du système que celle qu'auraient donnée les équations d'Euler. La définition de trois lois d'état identiques pour les trois espèces est d'ailleurs la seule manière d'assurer que l'homogénéité des températures et des vitesses soit une hypothèse rigoureusement valable. Nous appliquons donc la méthode exposée dans la section 3.3, l'unique loi d'état considérée étant une loi des gaz parfaits de coefficient $\gamma = 1.4$. Le résultat pour 400 cellules au temps $t = 0.2$ est donné ci-dessous, et nous permet de constater - le résultat obtenu étant exactement le même que celui obtenu

avec le modèle d'Euler à trois équations - que la méthode VFFC a été correctement implémentée et que les étapes de résolution des systèmes linéaires par méthodes de Newton sont efficaces. Nous notons que le temps de calcul CPU constaté pour ce cas est de $7.5s$, tandis qu'il est de $4.8s$ pour le modèle à trois équations, différence qui nous semble très acceptable.



3.4.2 Comparaison modèle homogène / modèle à 7 équations

Nous allons maintenant comparer les deux modèles sur un même cas, qui ne sera pas paramétré a priori pour assurer une quelconque homogénéité des vitesses ou des températures, et qui intégrera toute la complexité du modèle à 7 équations. Nous partirons d'une condition initiale présentant une unique vitesse et une unique température, et observerons comment la contrainte d'homogénéité du modèle à 5 équations influera sur l'évolution globale du système.

À la différence du cas test de Sod, nous considérons trois espèces distinctes, à savoir l'air, l'eau liquide et sa vapeur. Par ailleurs nous nous plaçons dans une zone proche de l'équilibre entre le liquide et sa vapeur, et autorisons le changement de phase, qui jouera un rôle majeur dans l'évolution du système. Les lois d'état utilisées sont des lois analytiques classiquement utilisées pour ces trois espèces, à savoir :

- Pour l'air, une lois des gaz parfaits de coefficients $\gamma = 1.4$ et $C_v = 641.0 J.kg^{-1}.K^{-1}$:
  - $e(p,T) = C_v\ T$.
  - $\rho(p,T) = \frac{p}{(\gamma-1)\ C_v\ T}$.

  Notons que les quantités ρ et e sont reliées entre elles indépendamment du paramètre C_v et de la température. Cependant, à la différence du modèle à trois équations, la définition du paramètre C_v joue cette fois un rôle important, car elle contribuera à l'ajustement de la température, commune à au moins deux espèces - trois espèces pour le modèle homogène -.

- Pour la vapeur d'eau, une lois des gaz parfaits de coefficients $\gamma = 1.4$ et $C_v = 6715.8 J.kg^{-1}.K^{-1}$.
- Pour l'eau liquide, une loi *stiffened gas* - déjà évoquée dans la section 1.4 - de paramètres $\gamma = 1.4$, $C_v = 3789.2 J.kg^{-1}.K^{-1}$ et $\pi = 1.647\ 10^9 Pa$:
  - $e(p,T) = \frac{\gamma\ (p+\pi)\ C_v\ T}{\gamma\ p+\pi}$.
  - $\rho(p,T) = \frac{\gamma\ p+\pi}{(\gamma-1)\ \gamma\ C_v\ T}$.

Par ailleurs, étant donné que nous prenons en compte le changement de phase, nous avons besoin d'une courbe de saturation, elle aussi analytique, qui servira à calculer les enthalpies internes massiques de saturation :

$$p_{sat}(T) = p_0\ e^{b\ \left(\frac{1}{T_0}-\frac{1}{T}\right)} \tag{3.41}$$

avec $p_0 = 101417.98 Pa$, $T_0 = 373.15 K$, et $b = 4965.91 K$. On aura reconnu dans cette loi qu'un état de température T_0 et de pression p_0 constitue un état d'équilibre entre l'eau liquide et sa vapeur.

Modèle de changement de phase Le modèle considéré est issu de [6], qui s'inspire de [18]. Nous reprenons le formalisme introduit dans la définition du modèle à 7 équations de la section 3.2, et donnons ici le détail du calcul des termes sources en fonction des variables thermodynamiques. Nous écrivons la conservation de l'énergie globale de la façon suivante :

$$Q_s(h_{is} - h_{iw}) + Q_{is} + Q_{iw} = 0, \tag{3.42}$$

et introduisons deux coefficients d'échange thermique ω_{is} et ω_{iw} tels que

$$Q_{is} = \omega_{is}(h_{sat,s} - h_s), \tag{3.43}$$

$$Q_{iw} = \omega_{iw}(h_{sat,w} - h_w). \tag{3.44}$$

Les quantités $h_{sat,s}$ et $h_{sat,w}$ représentent respectivement les enthalpies internes de saturation de la vapeur d'eau et de l'eau liquide. Les coefficients d'échange thermique s'expriment en fonction des variables thermodynamiques et d'une constante de temps τ prise égale à $10^{-3} s$, valeur typique pour le temps caractéristique de retour à l'équilibre liquide - vapeur pour l'eau [18]. Plus précisément, ils s'expriment ainsi :

$$\omega_{is} = \frac{\alpha_s\ \alpha_w\ \rho_s}{\tau}, \tag{3.45}$$

$$\omega_{iw} = \frac{\alpha_s\ \alpha_w\ \rho_w}{\tau}. \tag{3.46}$$

Finalement, il nous reste à définir le paramètre Q_s, qui s'exprime ainsi en fonction des paramètres préalablement établis :

$$Q_s = -\frac{Q_{is} + Q_{iw}}{h_{is} - h_{iw}}. \tag{3.47}$$

Le lecteur est renvoyé à [6] pour davantage de détails sur ce modèle, cet article étant lui même inspiré du code de thermohydraulique CATHARE [18].

Force de traînée Toujours en se basant sur [6] et en reprenant le formalisme établi lors de la présentation du modèle à 7 équations, on utilisera le modèle suivant pour la force de traînée :

$$C_{drag} = A\ \theta_\rho\ \frac{\alpha_w\ \alpha_g\ \rho_w\ \rho_g}{\rho}\ |u_g - u_w| \tag{3.48}$$

avec :

$$\theta_\rho = \frac{\alpha_w}{\alpha_g} + \frac{\rho_g}{\rho_w} \tag{3.49}$$

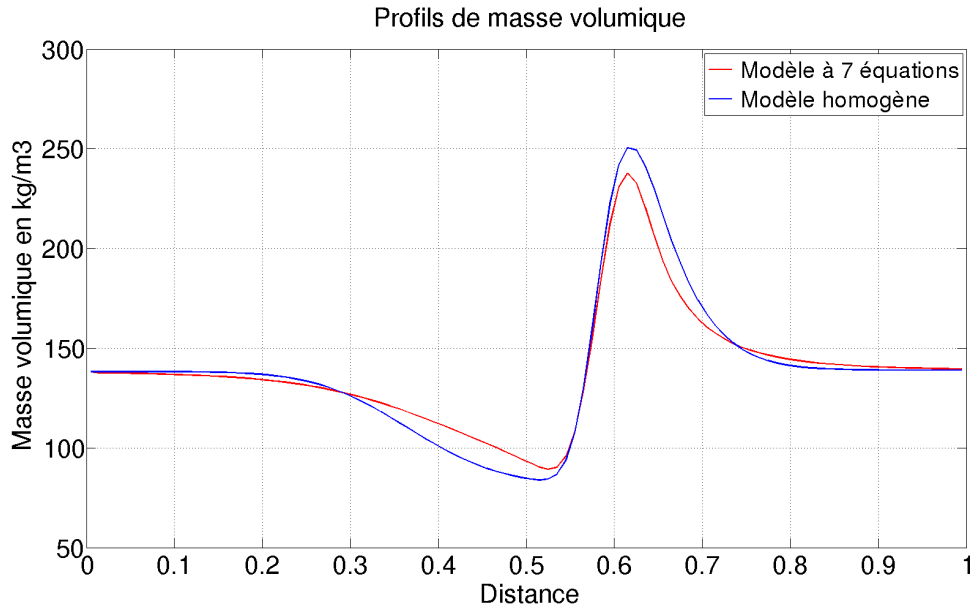
et A une constante prise égale à $10^4 m^{-1}$. Une fois de plus, les détails sur ce modèle sont donnés dans [6].

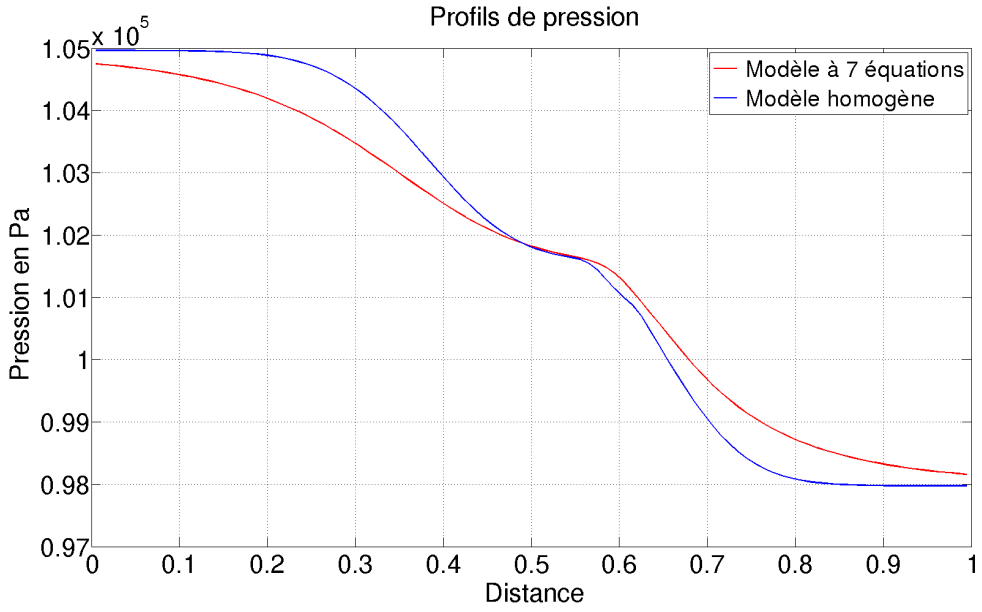
Présentation du cas Nous séparons en son milieu un domaine monodimensionnel de longueur 1, nous donnant ainsi deux sous domaines nommés 1 et 2 que nous initialiserons différemment. Nous considérons un mélange initial des trois espèces, avec à gauche comme à droite une même répartition volumique et une vitesse nulle. Plus précisément, nous fixons $\alpha_a = 0.3$, $\alpha_v = 0.5$ et $\alpha_l = 0.2$. Rappelant que l'état d'équilibre liquide/vapeur à la pression atmosphérique s'obtient en associant la température T_0 et la pression P_0 définies plus haut, et de façon à obtenir un léger déséquilibre thermodynamique dans l'ensemble du domaine avec des conditions différentes de part et d'autre de son milieu, nous écrivons :

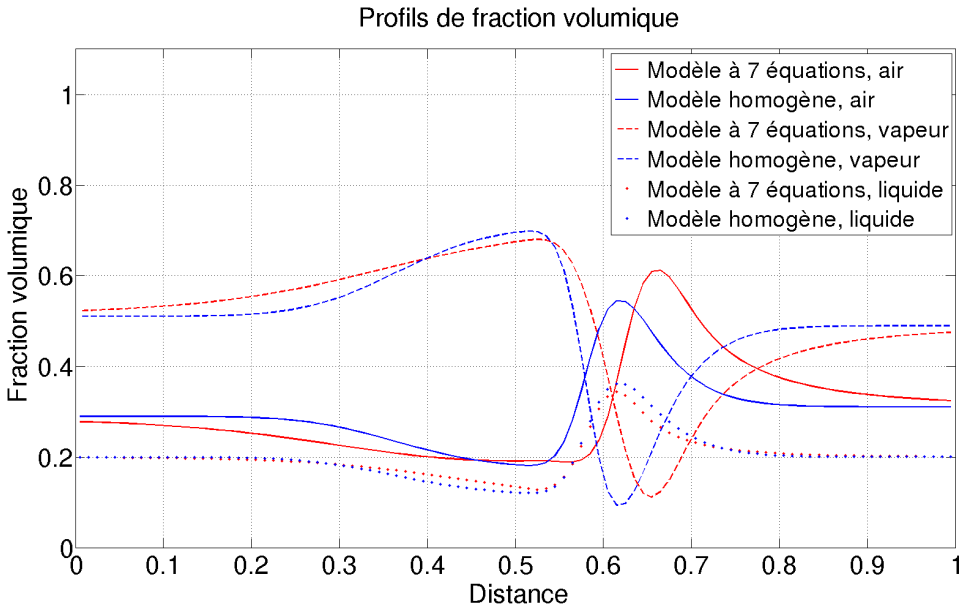
- Dans la partie gauche, $p = p_0$ et $T = T_0 + 1K$.
- Dans la partie droite, $p = p_0$ et $T = T_0 - 1K$.

Cette répartition des pressions et des températures créera, à gauche, un changement de phase du liquide vers la vapeur qui induira une montée de pression, et à droite, inversement, un changement de phase de la vapeur vers le liquide qui induire une baisse de pression. Ce déséquilibre des pressions induira un mouvement de matière de la gauche vers la droite. L'hétérogénéité entre espèces des vitesses d'écoulements sera atténuée par la force de traînée, de même que l'hétérogénéité des températures sera atténuée par les transferts thermiques entre espèces ainsi que par le changement de phase. Ainsi toute la finesse du modèle à 7 équations sera mise en œuvre dans ce test, et il sera intéressant d'observer comment se comportera le modèle homogène. Nous nous sommes arrêtés arbitrairement au bout de $0.025s$, et avons observé, pour les deux modèles, les profils résultants pour différentes variables d'intérêt. Précisons enfin que nous avons considéré des conditions aux limites de Neumann homogènes de part et d'autre du domaine.

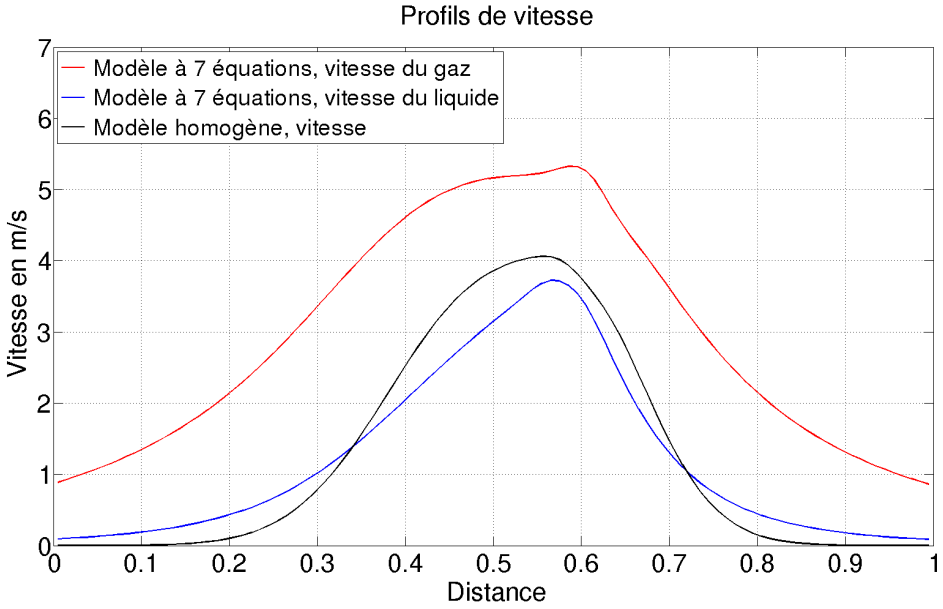
Pour commencer, les figures suivantes confirment que le modèle homogène offre une approximation intéressante des profils de masse volumique, de pression et de fractions volumiques des différentes espèces :

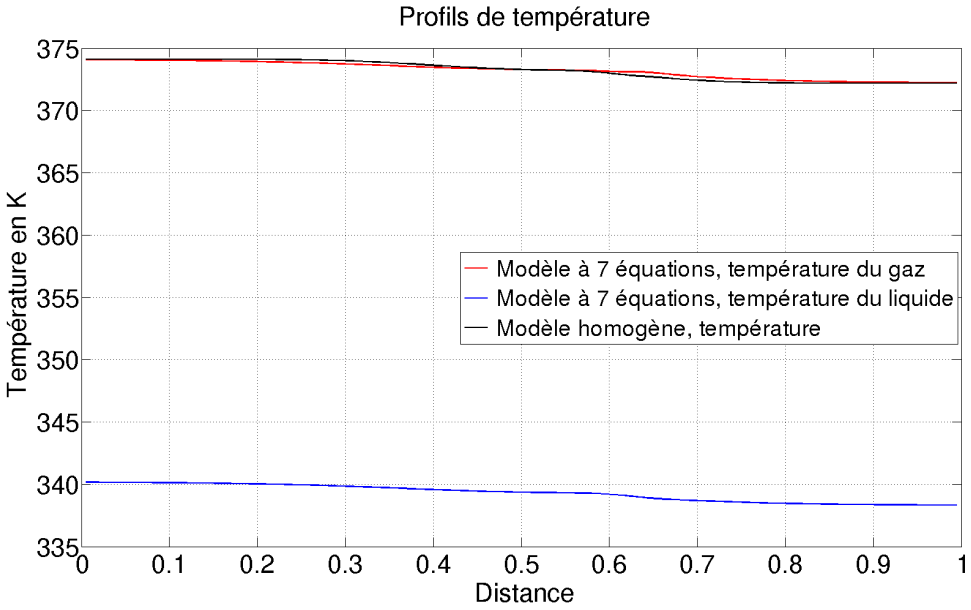

Profils de masse volumique
Modèle à 7 équations
Modèle homogène
Masse volumique en kg/m3
300
250
200
150
100
50
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Distance


Profils de pression
x 10^5
Modèle à 7 équations
Modèle homogène
Pression en Pa
1.05
1.04
1.03
1.02
1.01
1
0.99
0.98
0.97
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Distance



Il est évident que les hétérogénéités de température et de vitesse jouent un véritable rôle dans ce cas test, et nous ne pouvons prétendre à obtenir les mêmes profils avec un modèle les prenant en compte qu'avec un autre ne les considérant pas. Il est toutefois appréciable que l'idée générale de l'écoulement semble être capturée par le modèle homogène, et ce au pris d'un temps de calcul bien moindre. En effet le temps CPU correspondant à la simulation du modèle à 7 équations est de $27.2s$, tandis que celui correspondant à la simulation du modèle homogène est de $4.1s$. L'utilisation du modèle homogène fait donc gagner un facteur 6 sur le temps de calcul. Cela peut s'expliquer par le retrait de deux coordonnées à l'ensemble des vecteurs et des matrices, ainsi que par la convergence plus rapide de la méthode de Newton permettant de passer du vecteur v au vecteur w. Observons maintenant à quel point les températures et vitesses sont hétérogènes selon les phases d'après le modèle à 7 équations, et où se situent leurs équivalents pour le modèle homogène :





Ces courbes confirment que nous avons une réelle hétérogénéité des vitesses et des températures entre phases, et montrent que pour chacune de ces variables, leur équivalent pour le modèle à 7 équations ne se situe pas nécessairement entre la valeur prise par le gaz et celle prise par le liquide. Lorsque nous aborderons le couplage entre les deux modèles (section 3.5), nous ferons remarquer que la température unique (respectivement, la vitesse unique) du modèle à 5 équations, déduite des températures des deux phases (respectivement, des vitesse des deux phases) et permettant d'assurer la conservation des masses, énergie et quantité de mouvement, ne se situe pas nécessairement entre ces deux dernières températures (respectivement, ces deux dernières vitesses). Par ailleurs nous constatons que la température du modèle à 5 équations se situe largement plus près de la température du gaz que de celle du liquide. Cela s'explique par la plus grande sensibilité de la masse volumique du gaz à la température,

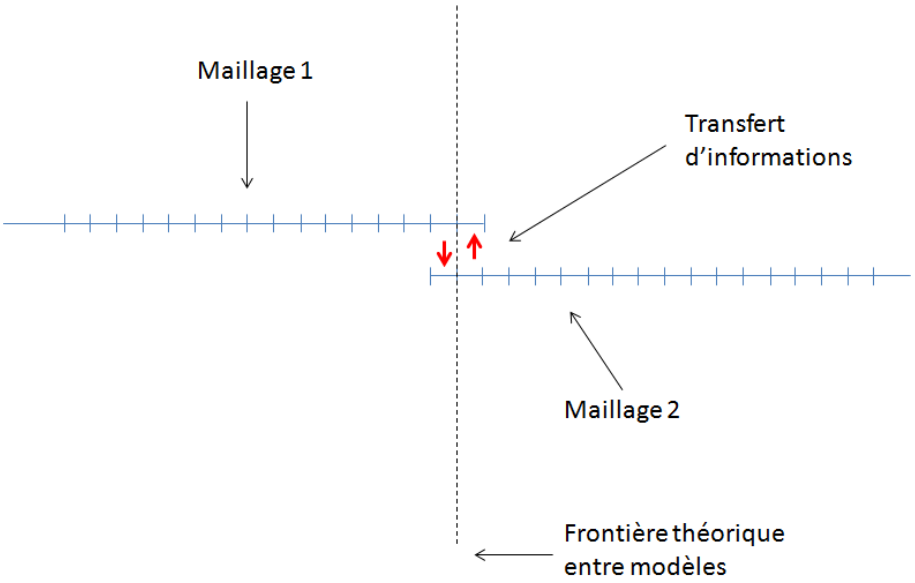
et donc par le fait qu'une température unique trop éloignée de celle du gaz ne permet pas, au sein du modèle à 5 équations, un respect approximatif des fractions volumiques des trois espèces calculées par le modèle à 7 équations.

3.5 Couplage modèle homogène / modèle à 7 équations

Nous allons voir dans cette section, comme annoncé en introduction, comment coupler le modèle à 7 équations et le modèle homogène, dans le cas où les hétérogénéités de vitesse ou de température ne seraient à considérer que dans une partie du domaine. Dans un premier temps, nous considérerons que cette partie du domaine est connue et ne varie pas avec le temps. Puis nous verrons comment adapter continument à la solution la frontière séparant les modèles.

3.5.1 Couplage à frontière fixe

Prenons le cas d'un maillage de N cellules, la frontière entre les deux modèles se situant au niveau du k-ième nœud. Nous créons alors deux maillages distincts, le premier de k cellules, le second de $N-k+2$ cellules, tels que les deux dernières cellules du premier maillage aient les mêmes positions que les deux premières cellule du second maillage. L'un de ces maillages sera utilisé avec le modèle à 7 équations - considérons qu'il s'agit du premier maillage -, l'autre avec le modèle homogène. Une itération consiste à appliquer à chaque maillage le schéma VFFC du modèle correspondant, puis à opérer un passage d'informations au niveau des cellules extrémale, comme présenté sur le schéma ci-dessous :



La méthode générale est la suivante :

A. Application du schéma VFFC à 7 équations sur le maillage 1.

B. Application du schéma VFFC à 5 équations sur le maillage 2.

C. Écrasement du vecteur des variables conservatives (vecteur v) de la dernière cellule du maillage 1. Remplacement par un nouveau vecteur calculé à partir du

vecteur v de la deuxième cellule du maillage 2. Réactualisation en conséquence des vecteurs $F(v)$ et $J(v)$.

D. Écrasement du vecteur v de la première cellule du maillage 2. Remplacement par un nouveau vecteur calculé à partir du vecteur v de l'avant-dernière cellule du maillage 1. Réactualisation en conséquence des vecteurs $F(v)$ et $J(v)$.

D. Retour à A.

Nous bornant ici à la version explicite du schéma VFFC, l'application des conditions aux limites sur l'extrémité droite du maillage 1 et l'extrémité gauche du maillage 2 n'a pas d'influence. En effet, la réactualisation des cellules extrémale opérée après chaque itération *effacera* la mémoire des conditions aux limites. Nous allons maintenant entrer dans le détail de ces réactualisations, qui devront être opérées avec rigueur pour assurer la conservativité exacte de la masse, de la quantité de mouvement et de l'énergie.

Transfert modèle homogène → modèle à 7 équations Toujours dans notre hypothèse où le modèle à 7 équations est considéré sur le maillage 1 et le modèle homogène sur le maillage 2, nous cherchons ici à transformer un vecteur de variables conservatives à 5 composantes noté v'', obtenu sur la deuxième cellule du maillage 2, en vecteur de variables conservatives à 7 composantes, noté v', à affecter à la dernière cellule du maillage 1. Le transfert d'informations dans ce sens ne présente aucune difficulté. En effet nous disposons grâce au modèle à 5 équations d'une vitesse et d'une température supposées uniques, d'une pression ainsi que des fractions massiques des trois espèces, qui suffisent pour alimenter le modèle à 7 équations. Plus précisément, du vecteur v'' à 5 composantes nous pouvons déduire, comme nous l'avons vu, un vecteur w'' des variables physiques, à savoir $w'' = (p, T, \alpha_s, \alpha_a, u)$. De ce vecteur et des lois d'état nous déduisons l'énergie interne de l'eau liquide e_w, et l'ensemble de ces données nous donnent le vecteur w' du modèle à 7 équations, à savoir $w' = (\alpha_s, \alpha_a, u_g, u_w, e_w, T_g, p)$, étant entendu que $T_g = T$ et $u_g = u_w = u$. De ce vecteur w' à 7 composantes nous déduisons un vecteur v' à 7 composantes ainsi qu'un flux $F(v')$ et que sa différentielle $J(v')$, qui seront utilisés lors de la prochaine itération du modèle à 7 équations. Cette transition conserve naturellement les masses ainsi que les quantités de mouvement et d'énergie, car l'hypothèse d'homogénéité des températures et des vitesses a été maintenue lors du passage aux vecteurs à 7 composantes.

Transfert modèle à 7 équations → modèle homogène Cette transition est plus complexe, car dans l'hypothèse où le vecteur v à 7 composantes induit des températures et vitesses différentes selon les phases, il nous faut les traduire en une unique vitesse et une unique température pour le modèle à 5 équations. La seule bonne méthode pour déterminer cette unique température et cette unique vitesse est de les calculer de façon à assurer la conservation de la masse, de la quantité de mouvement et de l'énergie au cours de cette transition. Notons $v' = (v'_1, v'_2, v'_3, v'_4, v'_5, v'_6, v'_7)$ le vecteur des variables conservatives du modèle à 7 équations, que nous cherchons à transformer en un vecteur à 5 composantes $v'' = (v''_1, v''_2, v''_3, v''_4, v''_5, v''_6, v''_7)$ compatible avec le modèle à 5 équations. Reprenant l'élaboration des équations (3.24) à (3.28), il est clair que la conservation des quantités d'intérêt s'écrit :

- $v''_1 = v'_1$
- $v''_2 = v'_2 + v'_3$
- $v''_3 = v'_2 - v'_3$
- $v''_4 = v'_4 + v'_5$

- $v''_5 = v'_6 + v'_7$.

Une fois calculé le vecteur v'', nous pouvons calculer comme vu précédemment un vecteur w'' des variables physiques à 5 composantes, dont nous déduirons notamment les composantes du vecteur des flux $F(v'')$ et de sa différentielle $J(v'')$, qui nous permettent de poursuivre le calcul.

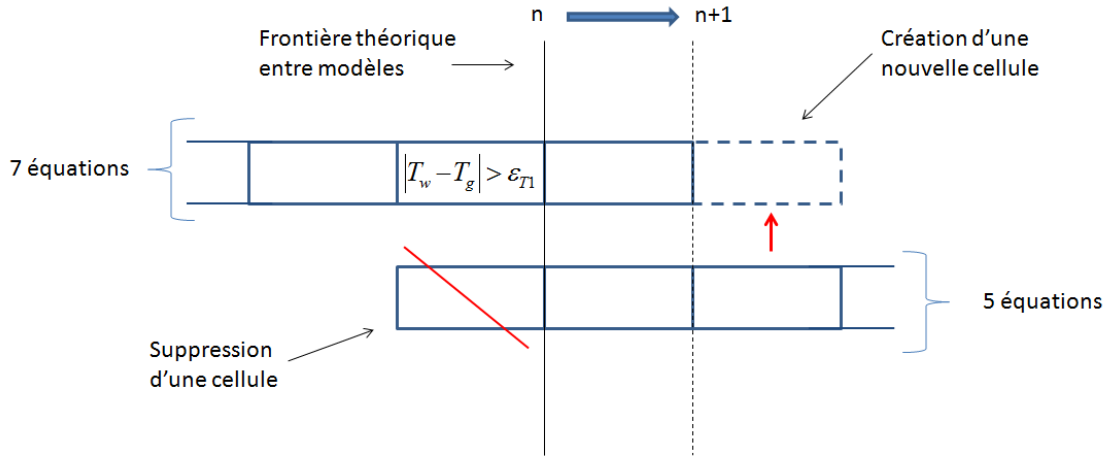
Remarque Dans la transition du modèle à 7 équations vers le modèle homogène, il est intéressant de noter que la température unique (respectivement, la vitesse unique) déduite des deux températures du liquide et du gaz (respectivement, des deux vitesses du liquide et du gaz) n'est pas nécessairement comprise entre ces deux températures (respectivement, entre ces deux vitesses). De même, bien que les modèles à 5 et 7 équations disposent chacun d'une unique valeur de pression, cette pression peut changer lors de la transition entre les deux modèles.

3.5.2 Couplage à frontière mobile

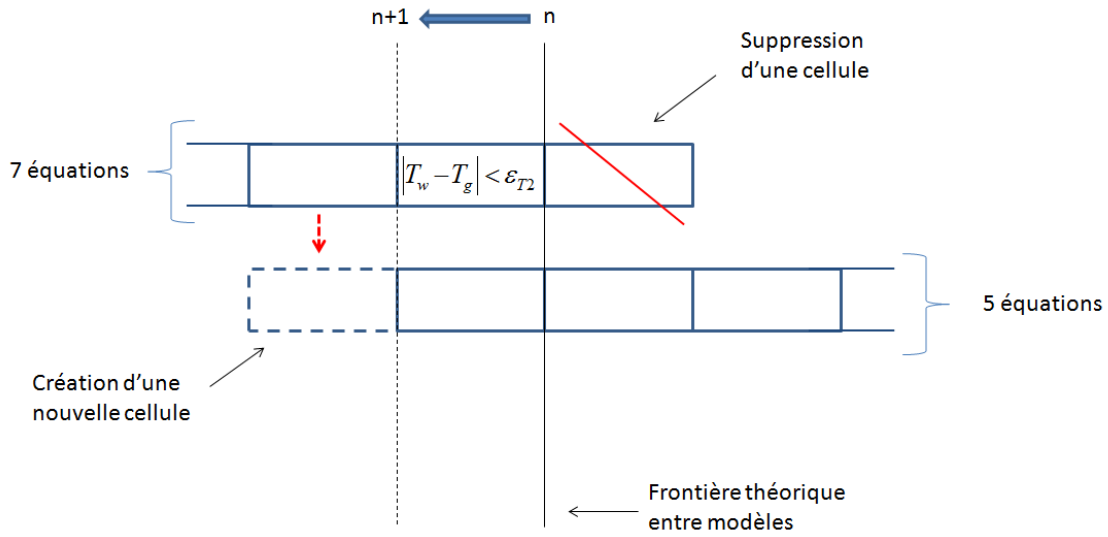
Nous pouvons faire en sorte que la position de la frontière soit sans cesse réadaptée au profil de température et de vitesse, de façon à se trouver toujours à l'endroit opportun. Pour cela nous ajoutons une étape précédant celles décrites ci-dessus - c'est à dire s'opérant avant l'application du schéma VFFC sur chacun des deux maillages -, afin de déterminer la nouvelle frontière. Nous nous plaçons donc entre deux itérations physiques notées n et $n+1$, et nous donnons tout d'abord deux critères de proximité des températures, homogènes à une température, notés ϵ_{T1} et ϵ_{T2}, et deux critères de proximité des vitesses, homogènes à une vitesse, notés ϵ_{u1} et ϵ_{u2}. Nous imposons, pour des raisons qui seront explicitées, $\epsilon_{T1} > \epsilon_{T2}$ et $\epsilon_{u1} > \epsilon_{u2}$. Plaçons nous à la fin d'une itération, lorsque les transitions entre vecteurs des variables conservatives à 5 et 7 composantes, présentées dans la sous-section 3.5.1, ont été faites. Appelons T_{g0}, T_{w0}, u_{g0} et u_{w0} les température du gaz, température du liquide, vitesse du gaz et vitesse du liquide dans la dernière cellule du maillage 1, c'est à dire la dernière cellule contenant des vecteurs de variables conservatives à 7 composantes. Nous sommes alors dans un des trois cas suivants :

A. Si $|T_{g0} - T_{w0}| < \epsilon_{T1}$ et $|u_{g0} - u_{w0}| < \epsilon_{u1}$ et $\Big[|T_{g0} - T_{w0}| > \epsilon_{T2}$ ou $|u_{g0} - u_{w0}| > \epsilon_{u2}\Big]$, alors la frontière ne bouge pas à cette itération, et tout se passe exactement comme dans le cas avec frontière fixe présenté dans la section précédente.

B. Si $|T_{g0} - T_{w0}| > \epsilon_{T1}$ ou $|u_{g0} - u_{w0}| > \epsilon_{u1}$, alors nous considérons que la vitesse ou la température deviennent trop hétérogène à proximité de la frontière entre modèles, que nous décalons donc d'une cellule pour renforcer l'emprise du modèle à 7 équations. Cela impose de supprimer la première cellule du deuxième maillage, et de créer une nouvelle cellule prolongeant le premier maillage. Cette nouvelle cellule est initialisée à l'aide de la cellule du deuxième maillage ayant la même position, selon la méthode de transition des variables conservatives des 5 composantes vers les 7 composantes décrite dans la sous-section 3.5.1. Le processus est explicité sur le schéma suivant, où nous n'avons représenté que la condition sur la température pour une question de concision :



C. Si $|T_{g0} - T_{w0}| < \epsilon_{T2}$ et $|u_{g0} - u_{w0}| < \epsilon_{u2}$, alors nous considérons que la vitesse et la température deviennent suffisamment homogènes à proximité de la frontière entre modèles, pour que l'on puisse se permettre de la décaler d'une cellule et renforcer ainsi l'emprise du modèle à 5 équations. Cela impose de supprimer la dernière cellule du premier maillage, et de créer une nouvelle cellule prolongeant le deuxième maillage. Cette nouvelle cellule est initialisée à l'aide de la cellule du premier maillage ayant la même position, selon la méthode de transition des variables conservatives des 7 composantes vers les 5 composantes décrite dans la sous-section 3.5.1. Le processus est explicité sur le schéma suivant, où nous n'avons représenté que la condition sur la température pour une question de concision :



Si nous nous sommes trouvés dans le cas C, il est possible que nous nous retrouvions à nouveau dans le cas C une fois l'opération de décalage de la frontière effectué. Dans ce cas, le processus est réitéré jusqu'à ce que l'on se trouve dans le cas A. Si nous nous sommes trouvés dans le cas B, le processus de décalage de la frontière nous a nécessairement ramenés dans le cas A.

3.5.3 Cas test

Nous élaborons ici un cas test mettant en œuvre le processus de couplage avec frontière mobile. Le paramétrage du modèle à 7 équations est différent de celui utilisé pour sa comparaison avec le modèle homogène, étant donné que nous choisissons un terme source ne permettant pas le changement de phase. Reprenant le formalisme introduit dans la description du modèle à 7 équations (cf. section 3.3), ce terme source, issue de [6], est le suivant : le paramètres Q_s, quantifiant la création de vapeur, est pris nul. Les paramètres Q_{is} et Q_{iw}, quantifiant les échanges d'énergie entre les deux phases, sont proportionnelles aux différences de température. Nous avons plus précisément $Q_{is} = Q\ (T_g - T_w)$ et $Q_{iw} = Q\ T_w - T_g$, avec :

$$Q = \frac{3\ \lambda\ \alpha_w\ \rho_w}{(\alpha_g\ \rho_g + \alpha_w\ \rho_w)\ R_{mean}^2},$$

où la constante λ est prise égale à $0.033087 W.K.m^{-1}$, et R_{mean}, homogène à une distance, vaut :

$$R_{mean} = R_0\ \left(\frac{\rho_w}{\rho_w^0}\right)^{-\frac{1}{3}}.$$

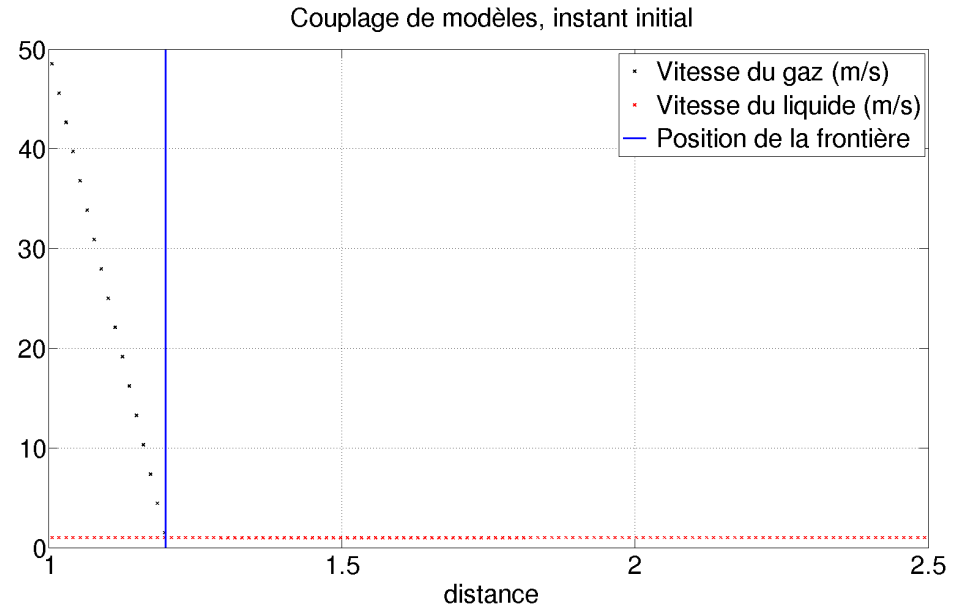
Dans cette dernière expression, ρ_w^0 est une masse volumique de référence choisie égale à $1 kg.m^{-3}$, et R_0 est une taille caractéristique des gouttes d'eau, prise égale à $10^{-4} m$. Davantage de détails sur ce modèle sont donnés dans [6].

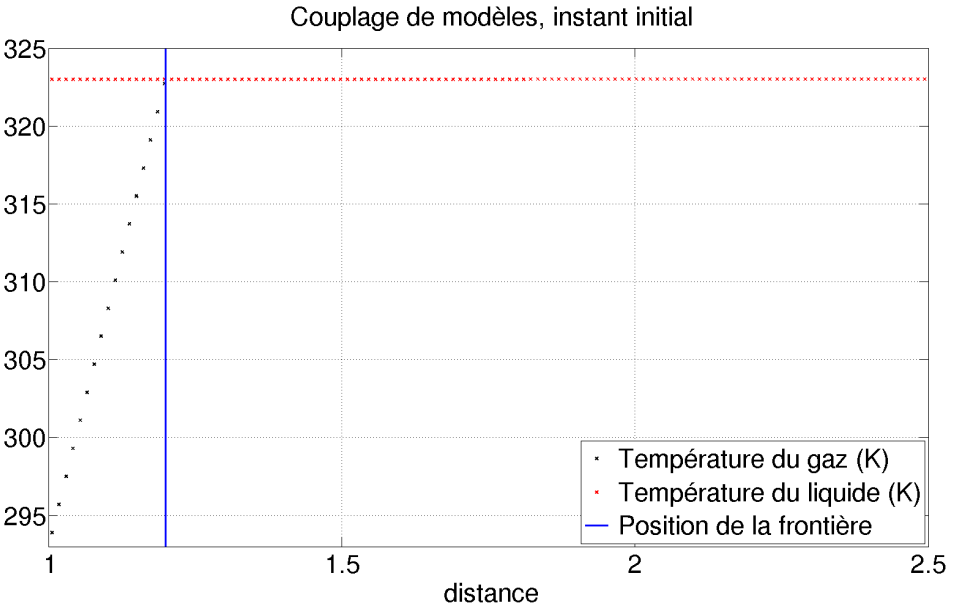
En dehors de cette absence de changements de phase, le paramétrage des modèles homogène et à 7 équations a été pris identique à celui présenté pour la comparaison des deux modèles (cf. sous-section 3.4.2), par exemple en ce qui concerne les lois d'état et le coefficient de la force de traînée. La condition initiale, en revanche, a changé. Nous avons besoin ici d'un problème initialement hétérogène, tant pour la température que pour la vitesse, sur une partie bien délimitée du domaine, et pour lequel cette hétérogénéité va se propager progressivement sur tout le domaine. Pour cela nous considérons un domaine de longueur $1.5m$, son abscisse x variant de $1m$ à $2.5m$. Nous imposons sur l'ensemble de ce domaine la pression atmosphérique $P_0 = 101417.98 Pa$. Pour la température et la vitesse, nous séparons ce domaine en deux zones :

- pour $x > 1.2$, $T_w = T_g = 323K$, et $u_w = u_g = 1 m.s^{-1}$
- pour $x < 1.2$, $T_w = 323K$, $u_w = 1 m.s^{-1}$, $T_g = 323K + \frac{(293K - 323K)\ (1.2m - x)}{0.2m}$ et $u_g = 1 m.s^{-1} + \frac{(50 m.s^{-1} - 1 m.s^{-1})\ (1.2m - x)}{0.2m}$.

Par ailleurs nous ne considérons que deux espèces, l'eau liquide et l'air, dont les fractions volumiques sont initialement égales et uniformes (i.e. $\alpha_g = \alpha_a = \alpha_w = 0.5$, $\alpha_s = 0$). Le reste des variables des vecteurs w et v se déduisent directement des précédentes, par l'intermédiaire des lois d'état. Précisons enfin que nous considérons des conditions aux limites de Neumann homogènes.

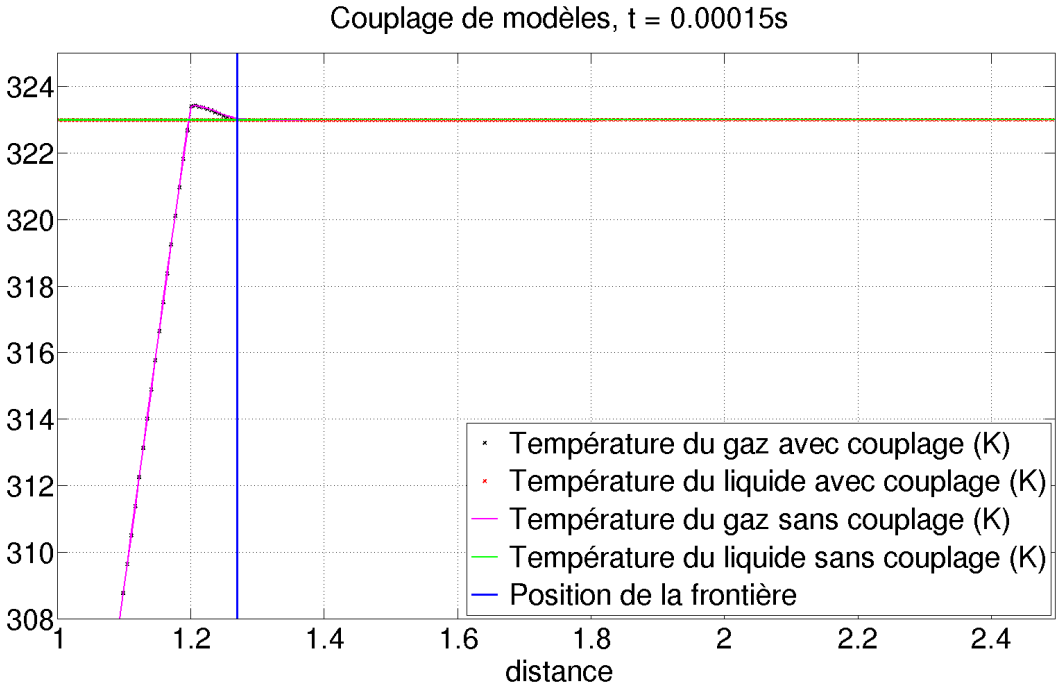
Nous avons donc, à l'état initial, une homogénéité des températures et des vitesses pour x supérieur à 1.2, abscisse qui constitue donc naturellement notre position initiale de la frontière entre modèles. À gauche de cette frontière, le modèle à 7 équations est initialisé d'après les descriptions ci-dessus. À droite de cette frontière, c'est le modèle homogène qui est initialisé. Son unique température est la température commune imposée au liquide et au gaz dans notre description du cas, et il en va de même pour la vitesse. Voici donc comment se présente l'état initial :

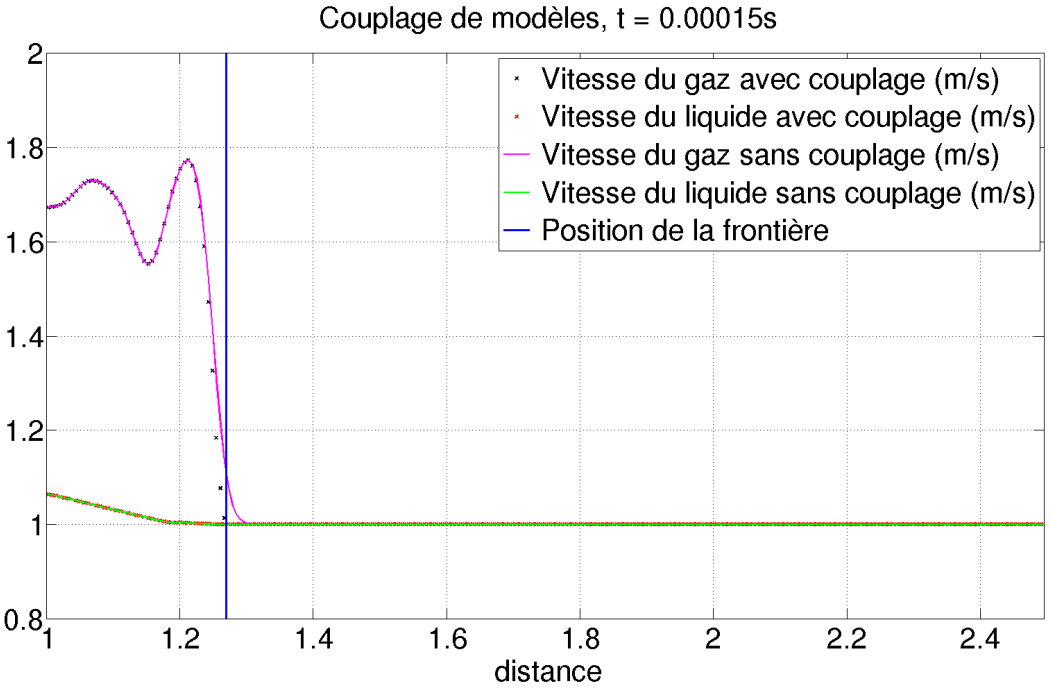




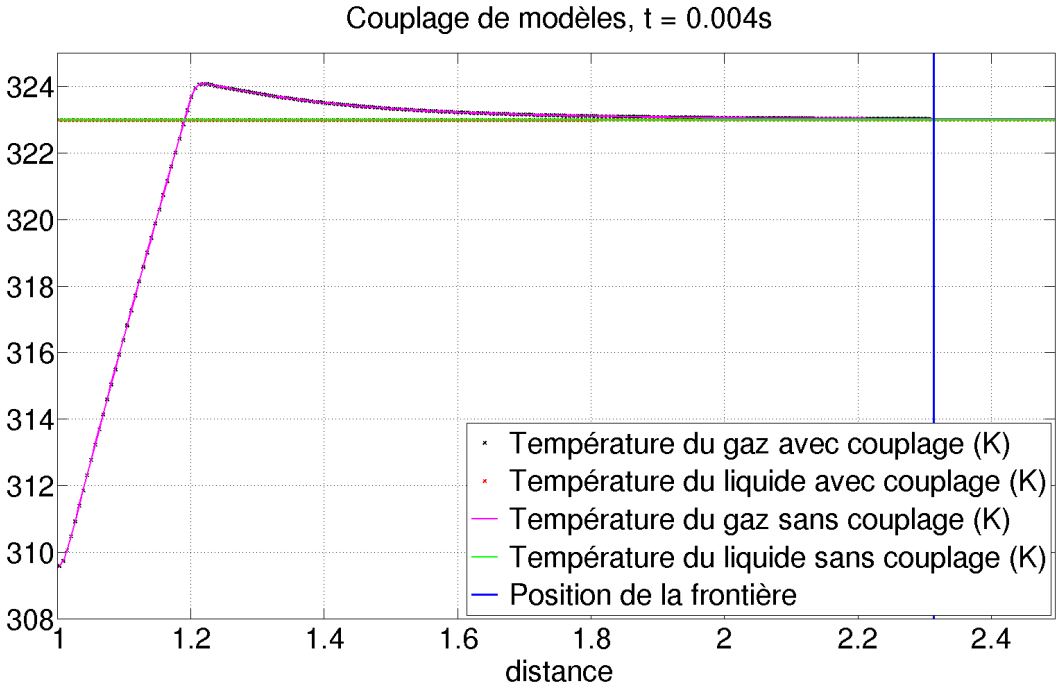
À partir de cet état initial, plusieurs phénomènes vont se produire : tout d'abord, la vitesse du gaz dans la partie gauche du domaine étant bien plus importante que celle du gaz, la force de traînée induira un fort ralentissement du gaz - ainsi qu'une accélération de l'eau, dans une bien moindre mesure étant donné sa masse volumique très supérieure -. Parallèlement, la différence de température entre gaz et liquide induira, quant à elle, un échange de chaleur entre les deux phases qui tendra à augmenter la température du gaz. Cet effet va se cumuler à celui de l'échauffement du gaz issu de la transformation de son énergie cinétique en énergie interne, elle même due au ralentissement du gaz que nous venons d'évoquer. C'est pourquoi la température du gaz va localement devenir supérieure à celle du liquide. Ces perturbations ne manqueront pas de se propager progressivement du côté où l'état initial est homogène.

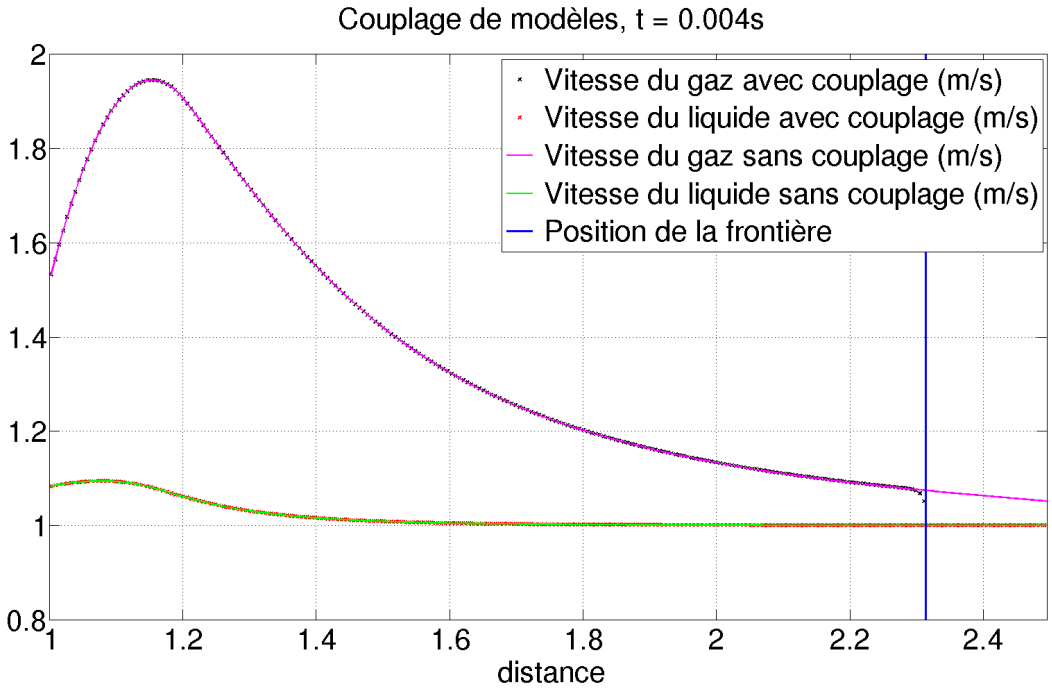
Nous allons maintenant mettre en œuvre la stratégie de couplage dynamique de modèles définie plus haut. Il est clair que pour obtenir une simulation correcte des phénomènes que nous venons d'évoquer, et pour tirer parti de la bien meilleure efficacité du modèle homogène en terme de temps de calcul, la frontière entre les deux modèles va constamment devoir s'adapter à la physique du système. Concernant le paramétrage du couplage de modèles, après quelques essais, nous avons choisi - en reprenant le formalisme introduit lors de la description du couplage de modèles, sous-section 3.5.2 - de fixer $\epsilon_{T1} = 10^{-3}\ T_{ref}$, $\epsilon_{T2} = 10^{-4}\ T_{ref}$, $\epsilon_{u1} = 10^{-3}\ u_{ref}$ et $\epsilon_{u2} = 10^{-4}\ u_{ref}$ avec $u_{ref} = 10m.s^{-1}$ et $T_{ref} = 293K$. Les profils de température et de vitesse observés au temps $t = 1.510^{-4}s$, que nous avons accompagnés de la position de la frontière séparant le modèle à 7 équations du modèle homogène, ainsi que par leurs équivalents obtenus à partir du modèle à 7 équations seul, sont les suivants :



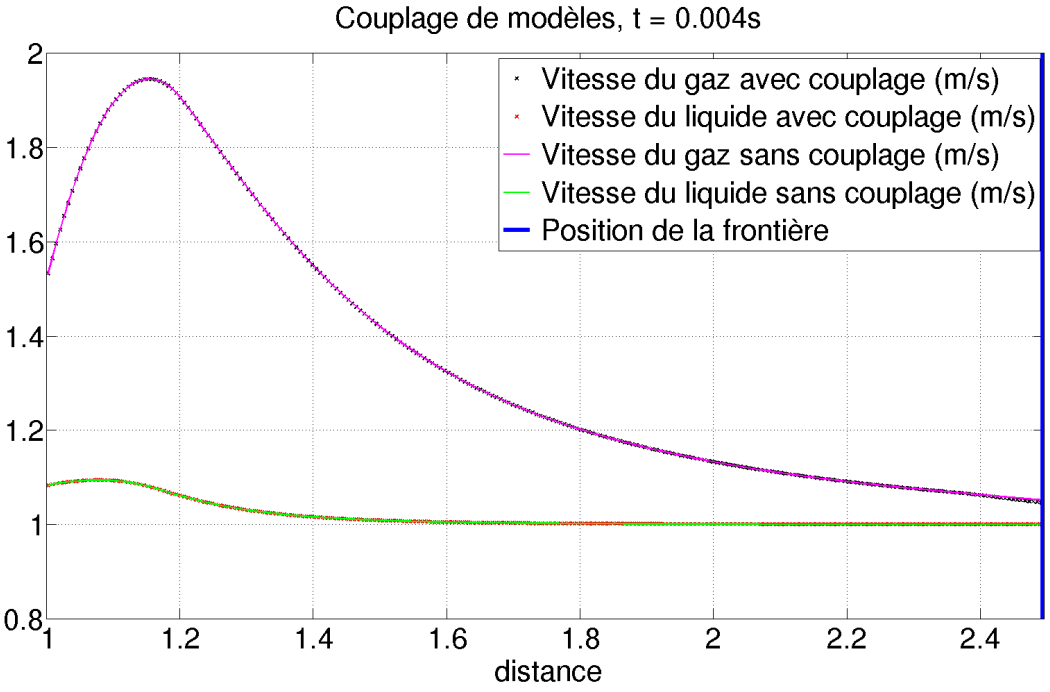


Une étude de convergence en maillage a par ailleurs été réalisée pour nous assurer que le résultat obtenu sans couplage - c'est à dire à partir du modèle à 7 équations seul - correspondait, en très bonne approximation, à la résolution exacte de ce modèle. Nous remarquons que le profil de température obtenu avec couplage est très correctement calculé, ce qui s'explique par le fait que c'est l'hétérogénéité de vitesse qui se propage le plus rapidement et positionne donc la frontière entre les deux modèles. Le profil de vitesse obtenu avec couplage est également bien calculé, mais le retard nécessairement pris par la position de la frontière entre modèles par rapport à l'évolution de la physique a pour conséquence une chute trop rapide de la vitesse du gaz à sa proximité. Rappelons en effet qu'à droite de la frontière, nous n'avons qu'une vitesse et qu'une température pour représenter le problème. Un peu plus tard, au temps $t = 4.10^{-3}s$, les profils de température et de vitesse sont les suivants :





Ici encore, le profil de température est bien calculé dans le cas avec couplage, mais on constate que l'erreur sur le profil de vitesse s'st accentuée au cours du temps, la frontière entre modèles ne s'étant pas propagée suffisamment rapidement. Afin d'obtenir une meilleure précision, nous avons augmenté la sensibilité du mouvement de frontière aux faibles hétérogénéités, en divisant par 10 les paramètres ϵ_{T1}, ϵ_{T2}, ϵ_{u1} et ϵ_{u2}. Voici alors le profil de vitesses que nous obtenons :



Nous voyons ici que le résultat est bien meilleur, et que la frontière a parcouru une distance plus importante et se trouve à la fin du calcul à l'extrémité du domaine. On peut donc, au prix d'un temps de calcul plus important, jouer sur le paramétrage du couplage pour améliorer la précision du calcul. Voici finalement un récapitulatif des temps de calcul et précisions obtenues pour ces différents cas. L'erreur calculée est l'erreur relative en norme L_2 sur la vitesse du gaz, choisie après qu'il ait été constaté que cette variable était la plus problématique en ce qui concernait le couplage. La référence pour le calcul de cette erreur est le profil obtenu avec le modèle à 7 équations seul, ce qui ne donne pas un écart avec la solution théorique - inconnue pour ce cas - mais qui montre bien la différence de profil induite par le couplage. Par ailleurs dans le tableau suivant, nous entendons par *précision standard* le premier paramétrage du couplage que nous avons considéré - à savoir le choix de ϵ_{T1}, ϵ_{T2}, ϵ_{u1} et ϵ_{u2} donné plus haut -, et par *précision accrue* le résultat obtenu après avoir divisé par 10 ces quatre derniers paramètres.

	erreur L2	Temps CPU en s
Modèle à 7 équations seul	0	110
Couplage, précision standard	0.016	81
Couplage, précision accrue	0.0036	88

Nous confirmons donc que le couplage de modèles permet d'obtenir un meilleur temps de calcul avec une précision contrôlée. Il est clair que son intérêt se ferait d'autant plus sentir que la zone présentant un caractère homogène serait large et pour une durée importante, et que l'on s'accorderait une grande marge d'erreur sur les conséquences physiques des hétérogénéités de vitesse et de température.

3.6 Conclusion

Cette partie visait à coupler deux modèles représentant la même physique - à savoir l'évolution cinématique et thermodynamique d'un mélange de trois espèces dans deux phases distinctes -, mais sous des hypothèses différentes. L'un - le modèle homogène - considère que les échanges thermiques et les échanges d'énergie mécanique entre espèces s'opère suffisamment rapidement pour qu'on puisse considérer, localement, l'unicité des vitesses et des températures. L'autre - le modèle à 7 équations - considère cette même hypothèse en ce qui concerne les deux espèces gazeuses - il aurait 9 équations dans le cas contraire - mais considère une température et une vitesse spécifiques pour chaque phase. Il va de soi que le premier modèle est plus efficace, tandis que le second est plus précis. Dans le cas de la propagation d'une onde de détonation dans un mélange d'eau liquide, de vapeur d'eau et d'air, chacun de ces modèles trouve sa pertinence dans une sous-partie du domaine. Le premier *loin* de la charge, le second *à proximité*. Cette partie avait pour but de mettre en œuvre le modèle homogène selon la méthode $VFFC$, d'opérer un couplage rigoureusement conservatif, de définir ce que nous entendons par *loin* et *à proximité* de la charge, et de faire évoluer dynamiquement la frontière pour qu'elle se situe en permanence à l'endroit le plus approprié, selon des critères définis par l'utilisateur.

Le modèle homogène, dont la théorie pré-existait [8], a pu être programmé selon la méthode $VFFC$ et validé grâce au cas test de Sod. Il a par la suite été envisagé la réalisation d'un cas test complexe, faisant intervenir des hétérogénéités entre phases, pour observer le comportement du modèle homogène par comparaison avec son équivalent à 7 équations, et contrôler que les écoulements résultants étaient, dans une certaine mesure, similaires. Si ce constat a pu être fait, il est clair que le modèle à 7 équations capture mieux cette physique hétérogène et, en plus de donner davantage d'informations - deux vitesses et deux température au lieu d'une -, calcule plus finement les données indépendantes de la phase, et notamment la pression, qui constitue notre principale variable d'intérêt dans le cas de la simulation des ondes de détonation.

Un couplage à frontière fixe a par la suite été mis en place qui, de façon classique, se base sur un prolongement des deux maillages correspondant aux deux modèles, pour leur créer une partie commune permettant les échanges d'informations, sur le même principe que les cellules du *halo* dans la mise en œuvre du parallélisme ou de conditions aux limites périodiques. La différence est ici que les données physiques ne sont pas traduites de la même manière selon le modèle considéré. En particulier, l'état physique est entièrement représenté par 5 variables scalaires dans le cas du modèle homogène, alors qu'il en faut 7 pour le modèle hétérogène à 7 équations. Cet échange a été défini de telle sorte qu'aucune perte de masse, quantité de mouvement ou énergie ne soit induite.

Finalement, nous avons mis en œuvre une méthode de déplacement de cette frontière, pour qu'elle s'adapte en permanence à la solution du système physique. Ainsi, à chaque itération, la frontière peut se décaler d'une cellule dans l'un ou l'autre sens si le besoin s'en fait sentir. L'efficacité de cette méthode, ainsi que la liberté de l'utilisateur dans le choix du compromis entre efficacité et précision, ont pu être contrôlées sur un cas test représentatif des applications visées.

Deuxième partie

Imagerie médicale

Chapitre 4

Méthode de Horn Schuck pour le flot optique

4.1 Introduction

Le contenu de ce chapitre a été en grande partie obtenu au sein d'un laboratoire rattaché à un centre de recherche hospitalier. Ce laboratoire poursuit sa mission de $R\&D$ dans plusieurs secteurs des hautes technologies de la santé associées à l'imagerie médicale. Les applications proposées sont liées aux pathologies vasculaires et aux désordres rhéologiques de la circulation sanguine. Plus précisément, l'équipe à laquelle nous avons appartenu travaille sur l'analyse de l'écoulement sanguin dans le système cardiovasculaire et l'évaluation mécanique de maladies cardiaques par imagerie ultrasonore. Ses objectifs principaux sont de développer de nouveaux paramètres cliniques, non effractifs et mesurables par échocardiographie, afin d'améliorer le diagnostic clinique de ces maladies. Une grande partie des travaux de recherche auxquels nous avons contribué a pour but d'exploiter au mieux les données issues de mesures échographiques. Le principe physique de cette technique d'imagerie repose sur les variations du comportement des ondes ultrasonores en fonction des propriétés intrinsèques du milieu qu'elles parcourent. Un ensemble de transducteurs émet des ondes ultrasonores dans différentes directions et enregistre les ondes rétrodiffusées. Afin de reconstituer l'image, on convertit en distance le temps écoulé entre l'émission et la réception, puis on fixe l'intensité du pixel correspondant proportionnelle au logarithme de l'amplitude de l'écho reçu. La gamme de fréquences habituellement utilisée va de 2 à $15MHz$. Un modèle de propagation rectiligne perturbé par des phénomènes de réflexion, réfraction, diffusion et absorption décrit correctement l'efficacité de l'imagerie par échographie : l'intensité lumineuse d'un pixel dépend directement de l'importance de ces phénomènes, et nous renseigne donc sur la texture du tissu environnant. Des détails sur cette technique d'imagerie sont donnés dans [15].

D'un point de vue théorique, cette partie porte sur la détermination du flot optique, c'est-à-dire l'obtention d'un champ de déplacements à partir de deux images successives, qui est un problème difficile car mal posé. Dans le domaine médical, le calcul du flot optique permet d'analyser les mouvements des organes, comme les contractions cardiaques. Bien que ce thème puisse sembler éloigné des problématiques liées à la simulation numérique des écoulements, nous allons voir que les méthodes employées sont très similaires. En effet, l'algorithme de détermination du flot optique étudié ici

est basé sur la résolution d'une équation aux dérivées partielles par différences finies. Il s'agit de l'algorithme de Horn et Schunck, une des premières méthodes mises au point pour traiter ce problème, très largement utilisée en raison notamment de sa simplicité. Nous présenterons en premier lieu le problème mathématique de convergence lié à cet algorithme.

Par ailleurs, l'algorithme de Horn et Schunck, tout comme nombre de méthodes de résolution du flot optique, est adapté à des images rectangulaires, discrétisées sur des grilles à pas uniforme. Or en échographie, en particulier en échographie cardiaque, les données sont souvent obtenues en coordonnées polaires. En effet, pour balayer un large domaine sans déplacer la sonde, on peut faire varier l'angle d'émission des ultrasons plutôt que de leur imposer une translation. Nous verrons dans un second temps comment adapter l'algorithme de Horn et Schunck aux coordonnées polaires en s'épargnant une interpolation sur une grille cartésienne, étudierons son efficacité et présenterons quelques cas d'application.

4.2 Analyse de la méthode Horn Schunck

Cette méthode vise à résoudre le problème du flot optique en cherchant un déplacement qui corresponde au mieux à la suite d'images tout en lui imposant une contrainte de lissage. Le problème de minimisation qui en découle se ramène à la résolution d'un système linéaire qui peut être de très grande taille, le nombre d'inconnues étant le double du nombre de pixels de l'image pour une image en deux dimensions. Une méthode itérative a été suggérée par les auteurs pour le résoudre efficacement, sans toutefois qu'une démonstration de convergence soit présentée. Bien plus tard, deux articles sont parus, présentant deux preuves de convergence différentes pour cette méthode itérative, preuves qui se trouvent toutes deux être fausses. Nous présenterons ici une preuve correcte.

4.2.1 Description de la méthode

Le flot optique est le déplacement des points d'une image, qui se traduit en terme de variations du champ d'intensité lumineuse. Les méthodes de calcul du flot optique partent toutes du principe selon lequel chaque pixel conserve son intensité initiale au cours de son déplacement. Nous notons dans la suite $I(x, y, t)$ l'intensité au point (x, y) à l'instant t. Supposons que le pixel situé au point (x, y) à l'instant t ait parcouru une distance δ_x selon l'axe des abscisses et δ_y selon l'axe des ordonnées, pendant une durée Δt. On obtient alors :

$$I(x + \delta_x, y + \delta_y, t + \Delta t) = I(x, y, t). \tag{4.1}$$

Nous notons maintenant $I_x = \frac{\partial I}{\partial x}$, $I_y = \frac{\partial I}{\partial y}$ et $I_t = \frac{\partial I}{\partial t}$. Par ailleurs les vitesses d'un pixel selon l'axe des abscisses et des ordonnées sont respectivement notées u et v. La limite de l'équation (4.1) divisée par Δ_t quand Δ_t tend vers 0 s'écrit alors :

$$I_x \; u + I_y \; v + I_t = 0. \tag{4.2}$$

L'équation (4.2) est appelée *équation du flot optique*. On remarque que l'équation (4.2) écrite en chaque pixel ne permet pas de déterminer le champ de vitesses, car on a une

équation scalaire pour deux inconnues scalaires. Le problème est donc par essence mal posé. Les méthodes de flot optique ont pour finalité de rendre ce problème bien posé en ajoutant des hypothèses supplémentaires, puis de le résoudre. Une des méthodes les plus populaires est la méthode de Horn et Schunck, dont le but est de chercher une solution approchée de (4.2) qui soit la plus régulière possible. Pour cela, on cherche le champ de vitesses qui minimise l'intégrale :

$$\int_V \Big[\big(I_x\ u + I_y\ v + I_t\big)^2 + \alpha\big(||\nabla u||^2 + ||\nabla v||^2\big)\Big] \tag{4.3}$$

où V désigne le domaine de l'image, et α un réel strictement positif indépendant de l'image. Le premier terme de cette intégrale doit être le plus petit possible en raison de l'équation (4.2). Le deuxième terme est d'autant plus petit que le champ de vitesses est lisse. Cette méthode se base donc sur une hypothèse de régularité du champ de déplacement. On se place maintenant dans l'espace fonctionnel $H^1(V)$. Soit (u, v) un champ de vitesses qui minimise la fonctionnelle (4.3) dans cet espace. Alors la différentielle de (4.3) s'annule en (u, v), ce qui s'écrit, pour toutes fonctions (ϕ_1, ϕ_2) de $H^1(V)$:

$$\int_V \Big[2\ \big(I_x\ u + I_y\ v + I_t\big)\ \big(I_x\ \phi_1 + I_y\ \phi_2\big) + 2\ \alpha\big(\nabla u.\nabla\phi_1 + \nabla v.\nabla\phi_2\big)\Big] = 0. \tag{4.4}$$

Puis, la formule de Green permet d'écrire :

$$\int_V \Big[2\ \big(I_x\ u + I_y\ v + I_t\big)\ \big(I_x\ \phi_1 + I_y\ \phi_2\big) - 2\ \alpha\big(\Delta u\ \phi_1 + \Delta v\ \phi_2\big)\Big] + 2\ \alpha \int_{\partial V} \Big[\frac{\partial u}{\partial n}\phi_1 + \frac{\partial v}{\partial n}\phi_2\Big] = 0. \tag{4.5}$$

Considérons d'abord les fonctions ϕ_1 et ϕ_2 s'annulant au bord du domaine, pour lesquelles le dernier terme du membre de gauche s'annule. Nous pouvons écrire :

$$\int_V \Big[2\ \big(I_x\ u + I_y\ v + I_t\big)\ \big(I_x\ \phi_1 + I_y\ \phi_2\big) - 2\ \alpha\big(\Delta u\ \phi_1 + \Delta v\ \phi_2\big)\Big] = 0. \tag{4.6}$$

L'équation (4.6) étant vraie pour toutes fonctions ϕ_1 et ϕ_2 s'annulant au bord du domaine, elle permet naturellement d'écrire :

$$I_x\ \big(I_x\ u + I_y\ v + I_t\big) - 2\ \alpha\ \Delta u = 0, \tag{4.7}$$

$$I_y\ \big(I_x\ u + I_y\ v + I_t\big) - 2\ \alpha\ \Delta v = 0. \tag{4.8}$$

Maintenant, en reprenant (4.5) avec des fonctions ϕ_1 et ϕ_2 quelconques, et en utilisant (4.7) et (4.8), on trouve :

$$\int_{\partial V} \Big[\frac{\partial u}{\partial n}\phi_1 + \frac{\partial v}{\partial n}\phi_2\Big] = 0. \tag{4.9}$$

Ceci étant vrai pour toutes fonctions ϕ_1 et ϕ_2 de $H^1(V)$, on en déduit :

$$\frac{\partial u}{\partial n} = 0, \tag{4.10}$$

$$\frac{\partial v}{\partial n} = 0. \tag{4.11}$$

Nous allons discrétiser les équations (4.7), (4.8), (4.10) et (4.11) sur une grille cartésienne composée des nœuds entiers (i, j) avec $1 \leq i \leq N_1$, $1 \leq j \leq N_2$, N_1 et N_2 représentant les nombres de pixels selon les deux coordonnées. Nous prenons donc comme unité de longueur le pas de discrétisation. Nous partons de l'hypothèse selon laquelle nous disposons de deux images successives séparées du pas de temps Δt. Le calcul discrétisé de la dérivée I_t de l'intensité lumineuse par rapport au temps ne pose alors aucune difficulté. En appelant I et I' ces deux images successives, nous écrivons en tout nœud (i, j) :

$$I_t = I' - I,$$

où l'unité de temps est prise égale au décalage temporel entre les deux images. Nous pourrions utiliser plus de deux images pour obtenir une meilleure approximation de la dérivée temporelle, mais nous nous limitons ici à la méthode la plus simple. Le calcul des gradients I_x et I_y n'est pas plus complexe et nous prendrons en général, au second ordre, $I_x(i,j) = \frac{1}{2}\left[I(i+1,j) - I(i-1,j)\right]$ et $I_y(i,j) = \frac{1}{2}\left[I(i,j+1) - I(i,j-1)\right]$ où h représente notre pas de discrétisation, supposé valoir 1. Le calcul du Laplacien se fait par différences finies. Nous utilisons la discrétisation classique de la dérivée seconde d'une fonction ϕ :

$$\frac{\partial^2 \phi}{\partial x^2}(i,j) = \frac{\phi(i+1,j) + \phi(i-1,j) - 2\ \phi(i,j)}{h^2} \tag{4.12}$$

La définition du Laplacien permet alors d'écrire :

$$\begin{aligned}\Delta\phi(i,j) &= \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} \\ &= \phi(i+1,j) + \phi(i-1,j) + \phi(i,j+1) + \phi(i,j-1) - 4\phi(i,j) \\ &= -4\ \left[\phi(i,j) - M(\phi)(i,j)\right] \end{aligned} \tag{4.13}$$

où $M(\phi)(i,j)$ représente une moyenne de ϕ sur les 4 plus proches voisins du nœud (i,j), et où nous rappelons que notre pas de discrétisation h vaut 1. Comme la définition du Laplacien ne dépend pas du repère cartésien considéré, nous pouvons généraliser ce raisonnement : si $M(\phi)$ représente la moyenne de ϕ sur 4 points équidistants du point courant et formant deux directions orthogonales, alors d'après (4.12), une approximation du Laplacien est donnée par :

$$\Delta\phi = -\frac{4}{d^2}\ \left[\phi - M(\phi)\right] \tag{4.14}$$

avec d la distance séparant le point courant des points intervenant dans la moyenne. Par exemple, en notant :

$$M(\phi)(i,j) = \frac{1}{4}\big[\phi(i+1,j+1) + \phi(i+1,j-1) + \phi(i-1,j+1) + \phi(i-1,j-1)\big], \tag{4.15}$$

alors $\Delta\phi = -2\ \left[M(\phi) - \phi\right]$ est une expression consistante du Laplacien, car alors $d = \sqrt{2}$. Enfin, il est possible de réaliser la moyenne de Laplaciens obtenus avec différents groupes de points pour calculer un Laplacien plus robuste. C'est le choix fait par Horn et Schunck dans leur article original. Ainsi, nous obtenons à partir des deux discrétisations présentées ci-dessus :

$$\Delta\phi(i,j) = 3\ \left[M(\phi)(i,j) - \phi(i,j)\right] \tag{4.16}$$

avec :

$$M(\phi)(i,j) = \frac{1}{6}\Big[\phi(i,j+1) + \phi(i,j-1) + \phi(i+1,j) + \phi(i-1,j)\Big] + \frac{1}{12}\Big[\phi(i+1,j+1) + \phi(i+1,j-1) + \phi(i-1,j+1) + \phi(i-1,j-1)\Big]. \quad (4.17)$$

Nous retiendrons dans la suite que $\Delta\phi = K\ \big[M(\phi) - \phi\big]$, avec K une constante strictement positive et $M(\phi)$ une moyenne de la fonction ϕ sur les voisins du point courant. En notant $\lambda = 2\alpha\ K$, les équations (4.7) et (4.8) se réécrivent :

$$I_x\ \big(I_x\ u + I_y\ v + I_t\big) - \lambda\ (M(u) - u) = 0, \quad (4.18)$$

$$I_y\ \big(I_x\ u + I_y\ v + I_t\big) - \lambda\ (M(v) - v) = 0. \quad (4.19)$$

La prochaine étape est d'exprimer u et v en fonction de $M(u)$ et $M(v)$, ce qui se fait par simple inversion matricielle. Nous obtenons :

$$u = M(u) - \frac{I_x\ \big(I_x\ M(u) + I_y\ M(v) + I_t\big)}{\lambda + I_x^2 + I_y^2}, \quad (4.20)$$

$$v = M(v) - \frac{I_y\ \big(I_x\ M(u) + I_y\ M(v) + I_t\big)}{\lambda + I_x^2 + I_y^2}. \quad (4.21)$$

Cette écriture mène aux itérations de Horn et Schunck sur l'entier k, écrites en tout point (i,j) :

$$u^{k+1} = M(u)^k - \frac{I_x\ \big(I_x\ M(u)^k + I_y\ M(v)^k + I_t\big)}{\lambda + I_x^2 + I_y^2}, \quad (4.22)$$

$$v^{k+1} = M(v)^k - \frac{I_y\ \big(I_x\ M(u)^k + I_y\ M(v)^k + I_t\big)}{\lambda + I_x^2 + I_y^2} = 0 \quad (4.23)$$

où la moyenne des points frontaliers, pour respecter (4.10) et (4.11), se calcule en considérant que les valeurs de déplacement en dehors de la frontière sont égaux aux plus proches déplacements faisant partie du domaine. Ainsi ces itérations, si elles convergent, convergent vers la solution du système (4.7), (4.8), (4.10), (4.11) discrétisé.

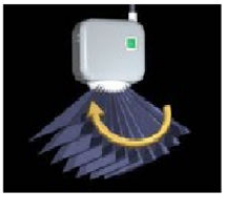
4.2.2 Étude théorique de la convergence

L'article original de Horn et Schunck ne contient pas de preuve de convergence. Cependant, l'utilisation très répandue de cet algorithme pendant trente ans a montré empiriquement que la convergence était systématique - l'article original de Horn et Schunck, datant de 1981, est cité près de 10.000 fois -. Deux preuves de convergence différentes ont par la suite été publiées, en 2004 [20] et 2008 [21]. L'article donné en annexe a été publié par la revue *SIAM journal on medical imaging*. Il commence par généraliser l'algorithme de Horn et Schunck en dimension quelconque, ce qui a du sens au vu des développements récents de l'imagerie 3D. Dans la deuxième partie, il montre que les preuves de convergence antérieures sont fausses et a priori irrattrapables. Puis des hypothèses sont faites sur le schéma numérique utilisé, une condition nécessaire et suffisante pour rendre le problème bien posé est proposée, et une démonstration de convergence est donnée en dimension quelconque sous cette condition. Une discrétisation très générale du Laplacien est proposée dans la section suivante, dont il est démontré qu'elle satisfait les hypothèses introduites dans le cadre de

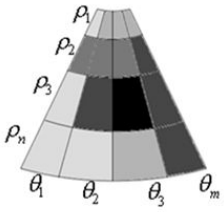
la preuve de convergence. L'article montre par la suite que les itérations de Horn et Schunck sont - sauf en ce qui concerne les points de frontière - des itérations de Jacobi par bloc, mais que les conditions générales de convergence de cette méthode ne sont pas vérifiées dans le cas de Horn et Schunck. On pourra se référer à [22] pour les détails et résultats généraux sur la méthode de Jacobi par blocs. Enfin, il est prouvé que dans les cas où le problème est bien posé, la convergence des itérations de Horn et Schunck entraîne celles d'autres méthodes itératives, à savoir *Gauss Seidel* et *SOR*. Ce dernier point était déjà connu avant notre travail, étant donné que la matrice du système linéaire de Horn et Schunck était déjà démontrée être définie positive par d'autres auteurs, avec cette fois une preuve correcte [2]. Il est toutefois intéressant d'obtenir ce même résultat par une aproche très différente. Tous les détails et références sont à consulter dans l'article en annexe.

4.3 Horn et Schunck en échographie cardiaque

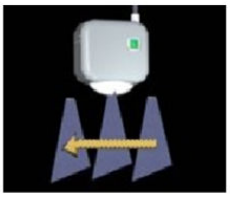
Maintenant que nous sommes convaincus de la validité des itérations de Horn et Schunck pour aborder le problème du flot optique, cette section vise à l'appliquer dans notre domaine, à savoir l'échographie du cœur. Comme alternative à la méthode de Horn et Schunck ou à celles qui en dérivent, dites *globales* car le champ de déplacement calculé en chaque point prend en compte l'ensemble des pixels, on trouve des approches dites *locales*, en général issues des travaux de Lucas et Kanade [17]. Le principe de ces dernières méthodes est de séparer l'image en fenêtres dont la taille peut varier, à condition de ne pas être réduite à un pixel. Dans l'approche originale, on considère que les déplacements sont égaux au sein d'une fenêtre, ce qui traduit l'hypothèse de régularité également considérée dans les approches globales. Ainsi, l'écriture des équations du flot optique (4.7), (4.8) en chaque point d'une fenêtre donne un système linéaire surdéterminé, que l'on résout au sens des moindres carrés. Parmi les avantages offerts par cette méthode, on note la rapidité de calcul et la robustesse par rapport aux valeurs aberrantes et aux variations rapides de mouvement, qui peuvent survenir en certains points de l'image. A contrario, un avantage de la méthode de Horn et Schunck par rapport à celle de Lucas et Kanade est qu'elle donne un champ de déplacement dense et lisse. Les approches globales semblent particulièrement bien adaptées à l'imagerie du cœur car cet organe ne présente pas a priori de variations rapides dans les déplacements, ce qui valide l'hypothèse de régularité sur l'ensemble du domaine de calcul. L'algorithme précédemment décrit, cependant, ne s'applique pas directement au cas des images issues de l'échographie cardiaque. En effet, comme expliqué en introduction, il est fréquent que les données issues de mesures échographiques soient obtenues en coordonnées polaires. Cela dépend en fait du type de sonde utilisé : on appelle sonde sectorielle une sonde qui fait varier l'angle du faisceau d'ultrasons, par opposition à la sonde linéaire pour laquelle une translation est opérée sur ce faisceau. Ce principe est représenté sur le dessin suivant, où l'on distingue ces deux types de sondes :



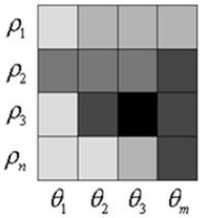
Sonde sectorielle



Type d'image obtenu



Sonde linéaire



Type d'image obtenu

Il nous faut donc, pour pouvoir traiter l'ensemble de ces situations, adapter la méthode de Horn et Schunck aux coordonnées polaires.

4.3.1 Adaptation de la méthode Horn Schunck aux coordonnées polaires

Nous repartons des équations (4.7), (4.8), (4.10), (4.11), que nous allons discrétiser sur une grille en coordonnées polaires, c'est à dire sur un ensemble de points de la forme $\{(r\ cos(\theta), r\ sin(\theta))|r = i\ d_r, \theta = j\ d_\theta\}$, avec $1 \leq i \leq N_1$, $1 \leq j \leq N_2$, N_1 et N_2 représentant les nombres de pixels selon les deux coordonnées. Notons que nous considérons toujours deux directions principales orthogonales x et y sur lesquelles sont projetés les gradients et vitesses. Ces directions, dans le cas des coordonnées polaires, sont choisies arbitrairement, mais nous choisirons en général pour la direction x la médiane de l'angle formé par le faisceau ultrasonore. Le calcul de la dérivée temporelle I_t ne pose pas plus de difficultés que dans le cas d'une grille cartésienne, et se fait toujours par soustraction de deux images successives. Les gradients I_x et I_y, eux, se calculent différemment. Nous notons (e_r, e_θ) la base conventionnelle orthonormée en coordonnées polaires, et (e_x, e_y) la base fixe en coordonnées cartésiennes. Les gradients sont donnés par :

$$\nabla I = \frac{\partial I}{\partial r} e_r + \frac{1}{r}\ \frac{\partial I}{\partial \theta} e_\theta, \tag{4.24}$$

dont une approximation au premier ordre est :

$$\nabla I(r,\theta) = \frac{I(r + d_r, \theta) - I(r,\theta)}{d_r}\ e_r + \frac{I(r, \theta + d_\theta) - I(r,\theta)}{r\ d_\theta}\ e_\theta. \tag{4.25}$$

Il faut ensuite projeter ces gradients dans la base fixe (e_x, e_y). En notant I_r et I_θ les coordonnées de ∇I dans la base (e_r, e_θ), on a classiquement :

$$\begin{pmatrix} I_x \\ I_y \end{pmatrix} = \begin{pmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix} \begin{pmatrix} I_r \\ I_\theta \end{pmatrix}. \tag{4.26}$$

Nous avons besoin maintenant d'une expression du Laplacien en coordonnées polaires. Il est bien connu que :

$$\Delta\phi = \frac{\partial^2\phi}{\partial r^2} + \frac{1}{r^2}\,\frac{\partial^2\phi}{\partial\theta^2} + \frac{1}{r}\,\frac{\partial\phi}{\partial r}. \tag{4.27}$$

Nous pouvons donc utiliser le schéma de discrétisation suivant :

$$\begin{aligned}\Delta\phi(r,\theta) = {} & \frac{\phi(r+d_r,\theta) + \phi(r-d_r,\theta) - 2\ \phi(r,\theta)}{d_r^2} \\ & + \frac{1}{r^2}\,\frac{\phi(r,\theta+d_\theta) + \phi(r,\theta-d_\theta) - 2\ \phi(r,\theta)}{d_\theta^2} \\ & + \frac{1}{r}\,\frac{\phi(r+d_r,\theta) - \phi(r-d_r,\theta)}{2\ d_r},\end{aligned} \tag{4.28}$$

qui peut se réécrire :

$$\begin{aligned}\Delta\phi(r,\theta) = K(r)\ \Big[& \lambda_1(r)\ \phi(r+d_r,\theta) + \lambda_2(r)\ \phi(r-d_r,\theta) \\ & + \lambda_3(r)\ \phi(r,\theta+d_\theta) + \lambda_4(r)\ \phi(r,\theta-d_\theta) - \phi(r,\theta)\Big]\end{aligned} \tag{4.29}$$

avec :

$$K(r) = \frac{2}{d_r^2} + \frac{2}{(r\ d_\theta)^2}, \tag{4.30}$$

$$\lambda_1(r) = \frac{1}{K(r)}\ \Big(\frac{1}{d_r^2} + \frac{1}{2\ r\ d_r}\Big), \tag{4.31}$$

$$\lambda_2(r) = \frac{1}{K(r)}\ \Big(\frac{1}{d_r^2} - \frac{1}{2\ r\ d_r}\Big), \tag{4.32}$$

$$\lambda_3(r) = \lambda_4(r) = \frac{1}{K(r)\ (r\ d_\theta)^2}. \tag{4.33}$$

Maintenant, pour tout point de la grille, de coordonnées r et θ, nous notons :

$$\begin{aligned}M(\phi)(r,\theta) = {} & \lambda_1(r)\ \phi(r+d_r,\theta) + \lambda_2(r)\ \phi(r-d_r,\theta) \\ & + \lambda_3(r)\ \phi(r,\theta+d_\theta) + \lambda_4(r)\ \phi(r,\theta-d_\theta)\end{aligned} \tag{4.34}$$

et $K(r,\theta) = K(r)$. Alors l'expression du Laplacien en coordonnées polaires s'écrit en tout point :

$$\Delta\phi(r,\theta) = K(r,\theta)\ \big[M(\phi)(r,\theta) - \phi(r,\theta)\big]. \tag{4.35}$$

Finalement, les équations (4.18) et (4.19) restent valables avec cette nouvelle définition de la fonction moyenne M, et l'on peut toujours écrire les itérations (4.22) et (4.23) pour résoudre le système linéaire qui en découle. Nous notons que, malgré l'utilisation des coordonnées polaires, nous avons conservé comme inconnues les projections u et v du déplacement sur la base fixe (e_x, e_y). Attention toutefois, il existe deux différences notables avec le cas des coordonnées cartésiennes :

- Pour le calcul de la moyenne $M(\phi)$, différents poids sont affectés aux différents voisins du point courant, c'est à dire le point où l'on calcule le Laplacien, et la distribution de ces poids est variable en fonction de la position du point courant.
- Le coefficient K n'est pas fixe mais varie selon la position du point courant.

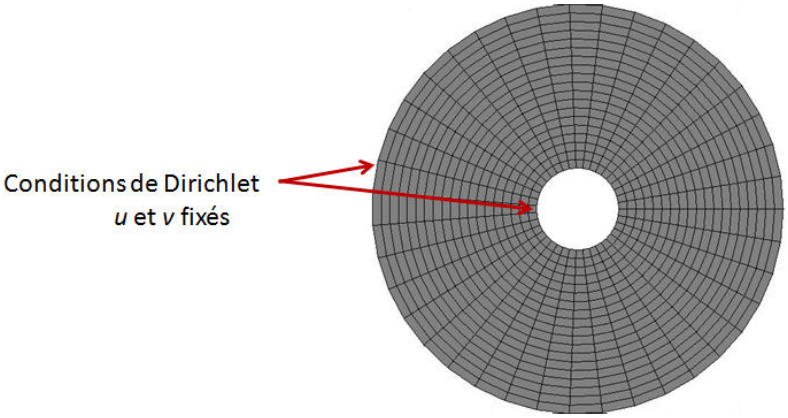
En conséquence, la preuve de convergence des itérations de Horn et Schunck présentée en annexe ne s'applique pas au cas des coordonnées polaires. Nous nous contenterons donc du constat empirique selon lequel la convergence est toujours vraie dans ce cas.

4.3.2 Test de l'algorithme

De nombreuses questions concernant la validité de l'adaptation de la méthode de Horn et Schunck en coordonnées polaires, à la différence des coordonnées cartésiennes, n'ont pas pu être résolues mathématiquement, à savoir l'existence d'une solution unique pour le problème discrétisé en coordonnées polaires, la convergence des itérations de Jacobi, la convergence de la solution discrète vers la solution exacte, et l'efficacité de cette méthode pour approximer le déplacement réel. Plusieurs cas tests ont donc été réalisés afin de contrôler tous ces points.

Convergence des itérations de Jacobi Le test consiste à tirer des images au hasard (l'intensité en chaque point est choisie selon une variable aléatoire uniforme) et à observer le comportement du schéma itératif. Ce test a été réalisé un grand nombre de fois, et la convergence est systématique. On remarque que la vitesse de convergence croît largement lorsque le facteur de lissage diminue, phénomène qui existait déjà dans le cas des coordonnées cartésiennes et qui peut être compris par lecture de la preuve de convergence présentée en annexe.

Convergence de la solution discrète vers la solution continue Il existe un cas facilement exprimable en coordonnées polaires pour lequel la solution exacte du problème de Horn et Schunck, (4.18) et (4.19), est connue. On considère une couronne, sur laquelle on va imposer des conditions de Dirichlet :



Il est bien connu que la solution de l'équation de Laplace ($\Delta\phi = 0$) sur ce domaine est de la forme :

$$\phi(r,\theta) = A\ ln(r) + B.$$

Les coefficients A et B se calculent facilement à partir des valeurs imposées aux bords. Nous rappelons que le système à résoudre s'écrit :

$$I_x\ \big(I_x\ u + I_y\ v + I_t\big) - \lambda\ (M(u) - u) \quad = \quad 0, \tag{4.36}$$
$$I_y\ \big(I_x\ u + I_y\ v + I_t\big) - \lambda\ (M(v) - v) \quad = \quad 0. \tag{4.37}$$

La méthodologie est alors la suivante : on considère les champs de déplacement (u^*, v^*) solutions de l'équation de Laplace. On tire au hasard des gradients I_x et I_y en chaque point, puis on calcule des gradients I_t de telle sorte que la relation $I_x\ u^* + I_y\ u^* + I_t = 0$ soit vérifiée partout. Il est alors clair que (u^*, v^*) est le champ de déplacement solution du problème de Horn et Schunck.

Nous pouvons maintenant contrôler la convergence de la solution du problème discrétisé vers la solution exacte du problème continu. Des tests ont été réalisés sur une couronne de rayon intérieur égal à 0.1 et de rayon extérieur égal à 1. Les conditions de Dirichlet imposées étaient $u = 0$ et $v = 2$ sur le bord intérieur, $u = 1$ et $v = 4$ sur le bord extérieur. La plage d'angles allant de 0 à 2 π a été divisée en 10 pixels, tandis que la plage de distance allant de 0.1 à 1 a été divisée en 30 pixels. La solution attendue étant invariante par rotation, le pas de discrétisation en angle n'est pas important. Il a été choisi très grand dans un souci de gain de temps. La courbe suivante montre l'évolution des solutions théoriques et calculées en fonction du rayon, pour un angle fixé. Le coefficient de lissage α a été pris égal à la moyenne des carrés des intensités, de façon à équilibrer les deux termes de l'équation à résoudre. Une étude sera réalisée par la suite sur l'influence de ce coefficient de lissage sur les résultats.

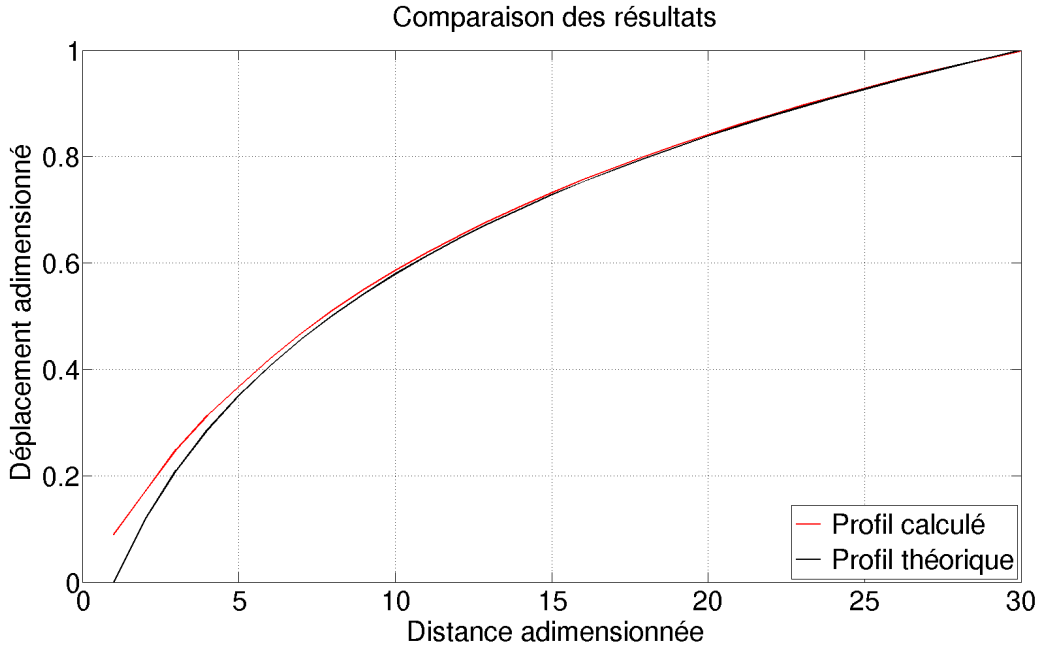


FIGURE 4.1 – Profils théorique et calculé par la méthode de Horn Schunck polaire

Nous constatons visuellement que la solution calculée est proche de la solution théorique, alors que le nombre de pixels est faible. La courbe suivante montre la convergence vers la solution exacte lorsque le pas de discrétisation tend vers zéro. Le pas de discrétisation en angle d_θ n'a pas été modifié, tandis que le pas de discrétisation en distance d_r a été progressivement diminué. R représente le rayon extérieur de la couronne. Dans la suite de ce chapitre, l'erreur relative est définie à partir de la norme L_2 sur l'ensemble du domaine de calcul.

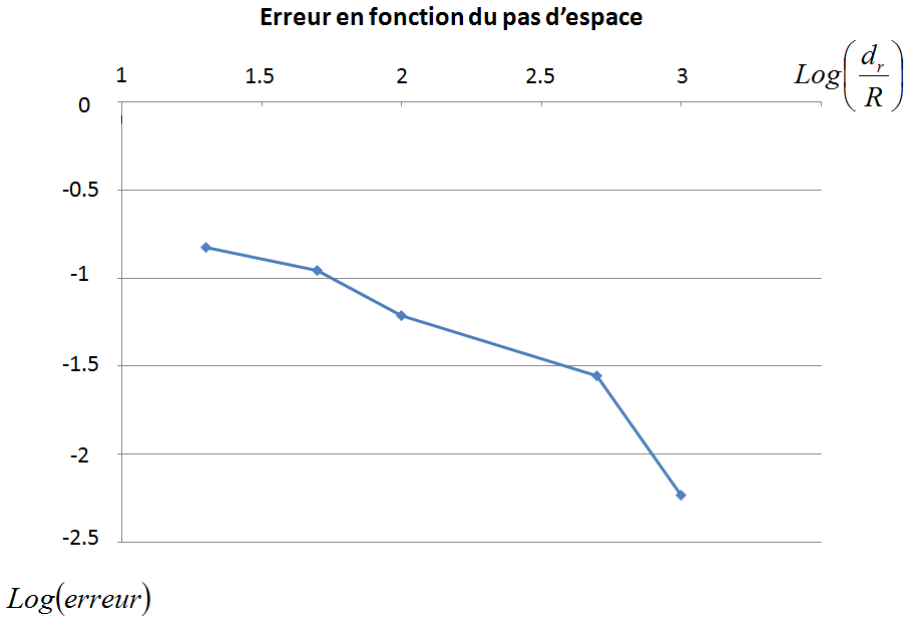


FIGURE 4.2 – Mesures d'erreurs pour l'algorithme de Horn Schunck polaire : influence du pas d'espace.

Ce cas test a également permis de faire un constat important : lors des itérations de Jacobi, le test d'arrêt sur la différence entre deux résultats successifs doit être très strict. Autrement dit, Il faut attendre une différence relative très faible entre deux itérations successives pour espérer avoir atteint la solution. Pour les tests présentés ci-dessus, il fallait attendre une erreur relative de l'ordre de 10^{-7} entre deux itérations afin d'obtenir un bon résultat. Par ailleurs, si la décroissance de cette erreur relative est assez rapide lors des premières itérations, elle devient nettement plus lente lorsqu'on approche de la condition d'arrêt. Ainsi, la convergence de la méthode de Jacobi est assez lente, ce qui est un point négatif de la méthode Horn Schunck polaire par rapport à la méthode Horn Schunck classique.

Efficacité de la méthode L'étape suivante a été de définir un champ de déplacement correspondant à une réalité physique, de déformer une image selon ce champ de déplacement, et d'observer dans quelle mesure la méthode de Horn et Schunck polaire permettait de retrouver le déplacement initialement fixé. La déformation de l'image a été faite de la manière suivante :

- Calcul des gradients I_x et I_y en tous points.
- Choix d'un champ de déplacement (u, v).
- Calcul des dérivées temporelles I_t en supposant l'équation du flot optique ($I_x\ u + I_y\ v + I_t = 0$) vérifiée.
- Calcul de l'image déformée I' à partir de l'image connue I grâce à la relation $I_t = I' - I$.

Le déplacement choisi est celui d'un vortex, c'est-à-dire un déplacement de corps rigide autour d'un point central. Notons que ce point central n'est pas l'origine de notre repère polaire. Ce cas a été choisi car il est relativement simple à mettre en œuvre, et parce que des vortex peuvent être détectés en échographie cardiaque, lorsqu'un agent de contraste a été mélangé au sang. Ce déplacement est détaillé sur le schéma suivant : le

point O représente le point central du vortex, le point M un point quelconque dans le domaine de calcul, $\overrightarrow{W} = (u, v)$ est le vecteur déplacement au point M et $\overrightarrow{n}$ un vecteur unitaire orthogonal au domaine de calcul. On fixe alors $\overrightarrow{W} = \overrightarrow{n} \otimes \overrightarrow{OM}$.

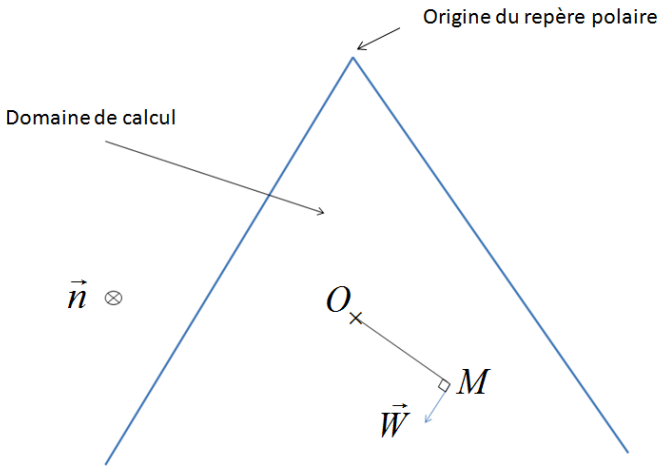


FIGURE 4.3 – Domaine de calcul

Nous visualisons ci-dessous l'image de départ. Il s'agit d'un ventricule gauche en *court axe*, c'est à dire en vue horizontale. L'angle varie de $-\frac{\pi}{6}$ et $\frac{\pi}{6}$ sur 158 pixels, et la distance de 1 à 13 sur 198 pixels.

FIGURE 4.4 – Ventricule gauche en court axe

Les figures suivantes montrent le vortex théorique utilisé pour déformer l'image et le champ de déplacement retrouvé grâce à la méthode précédemment développée, au prix d'un temps CPU avoisinant une minute. Nous constatons visuellement que l'algorithme Horn Schunck polaire semble fonctionner. Ici encore, le coefficient de lissage a été pris du même ordre de grandeur que les carrés des gradients d'intensité.

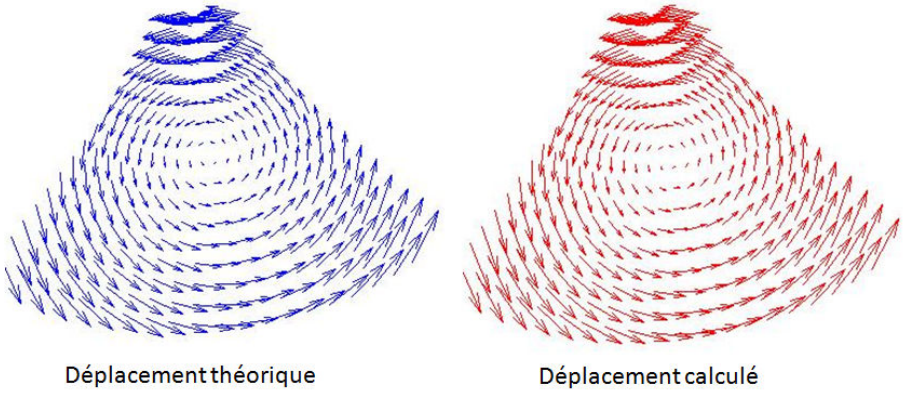


FIGURE 4.5 – Comparaison des résultats

L'erreur obtenue en norme L^2 est de 0.0018. Il est intéressant de constater une telle efficacité. Si N représente le nombre de pixels, nous avons retrouvé avec une erreur très faible un champ de déplacement comprenant 2 N valeurs scalaires, alors que nous ne disposions que de N équations. C'est l'hypothèse de régularité du champ de déplacement qui nous a permis d'y parvenir.

Influence des imperfections Dans notre déformation de l'image, au paragraphe précédent, nous avons assuré artificiellement un exact respect de l'équation du flot optique, ce qui n'est pas le cas dans la pratique. Les principales perturbations qui distordent l'équation du flot optique sont les erreurs de mesure et les erreurs de discrétisation dans le calcul des gradients. Cette dernière erreur est d'autant plus faible que les déplacements sont petits, et donc que la cadence d'acquisition des images est élevée. Plus précisément, nous allons considérer que les gradients réels suivent des lois normales de valeurs moyennes I_x et I_y (gradients calculés) et d'écarts types $\epsilon\ I_x$ et $\epsilon\ I_y$, où ϵ désigne un *petit* paramètre. Ainsi, la relation fondamentale du flot optique devient dans la pratique :

$$I_t = -(I_x\ u + I_y\ v)\ N(1, \epsilon^2), \tag{4.38}$$

où $N(1, \epsilon^2)$ représente la distribution normale de moyenne 1 et d'écart type ϵ. Toujours à partir de la même première image, on peut ainsi construire une deuxième image - ce qui revient à construire un champ de dérivées temporelles - qui prenne en compte ces imperfections. Notons que l'équation (4.38) peut se réécrire :

$$I_t = -(I_x\ u + I_y\ v)\ \Big[1 + \epsilon\ N(0, 1)\Big], \tag{4.39}$$

de sorte que l'écart type ϵ peut être vue comme l'importance relative du bruit par rapport au signal. Dans la suite, le paramètre ϵ sera donc identifié à cette importance relative du bruit. Nous constatons une progression à peu près linéaire de l'erreur en fonction du bruit. Lorsque le bruit atteint 10% du signal, nous retrouvons les déplacements théoriques avec une erreur relative inférieure à 6%, ce qui semble largement acceptable. Des tests ont été réalisés pour différentes valeurs de ϵ. Les résultats sont représentés sur la courbe suivante :

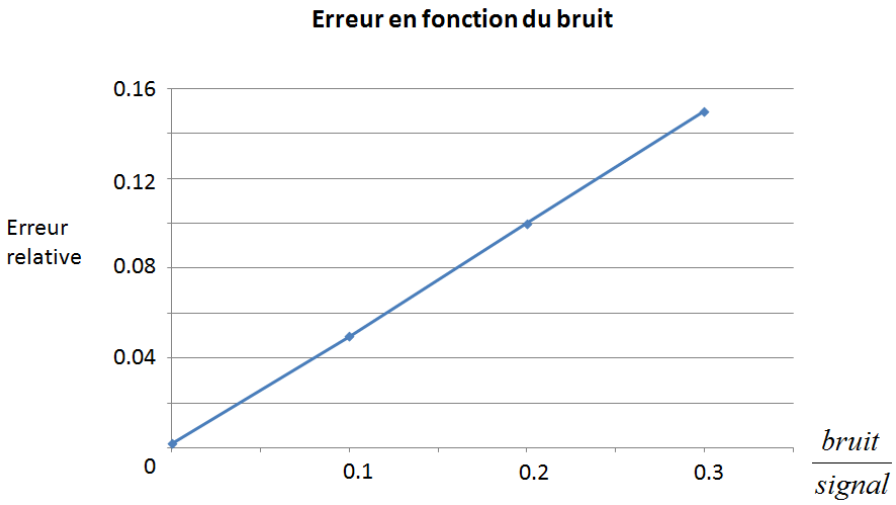


FIGURE 4.6 – Mesures d'erreurs pour l'algorithme de Horn Schunck polaire : influence du bruit.

Influence du coefficient de lissage Le coefficient de lissage - noté α dans la présentation de la méthode - est défini arbitrairement par l'utilisateur et influence le résultat, ce qui est en défaveur de l'efficacité de cette méthode pour retrouver le déplacement réel. Il est intéressant d'observer les différences obtenues dans les résultats pour différents coefficients de lissage. Pour le cas test du vortex avec ajout de 5% de bruit, nous observons qu'une multiplication par 1000 du facteur de lissage crée une différence relative de 7% seulement dans les résultats, ce qui est rassurant. Une étude plus approfondie a été réalisée : les courbes suivantes montrent l'évolution de l'erreur en fonction du coefficient de lissage, pour différentes amplitudes de bruit. Nous constatons la présence d'un minimum, c'est à dire d'un coefficient de lissage optimal, mais il est malheureusement a priori impossible de le connaître à l'avance. Nous remarquons également que ce coefficient de lissage optimal est plus petit pour une image moins bruitée. Cela n'est pas surprenant, une des finalités du facteur de lissage étant d'éliminer le bruit. Nous notons ici I_0^2 la moyenne des carrés des gradients d'intensité.

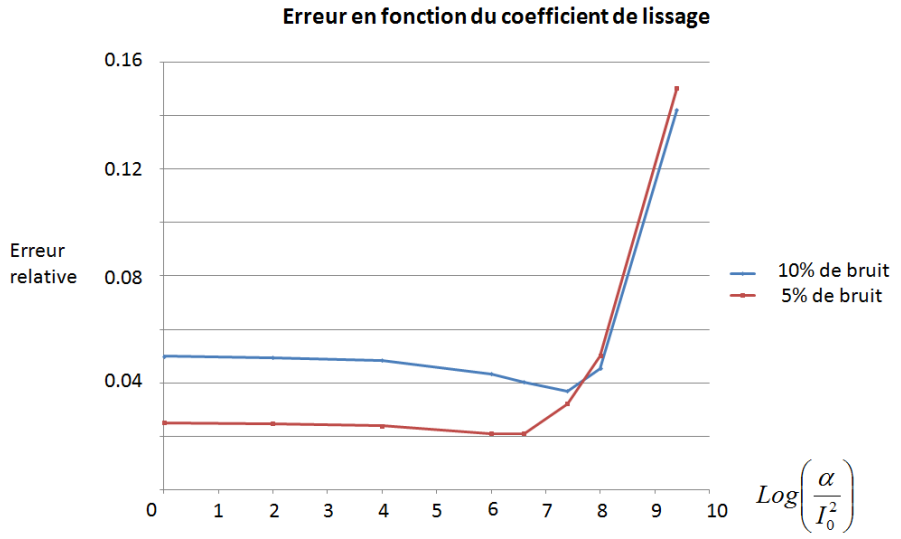


FIGURE 4.7 – Mesures d'erreurs pour l'algorithme de Horn Schunck polaire : influence du coefficient de lissage.

4.3.3 Cas réels

La dernière étape de validation est l'application de cette méthode de calcul du flot optique à des échocardiographies réelles, pour nous assurer que les résultats obtenus ressemblent visuellement aux déplacements réels du muscle cardiaque. Comme le montrent les images ci-dessous, qui superposent les images échographiques aux champs de déplacement calculés, les résultats sont encourageants :

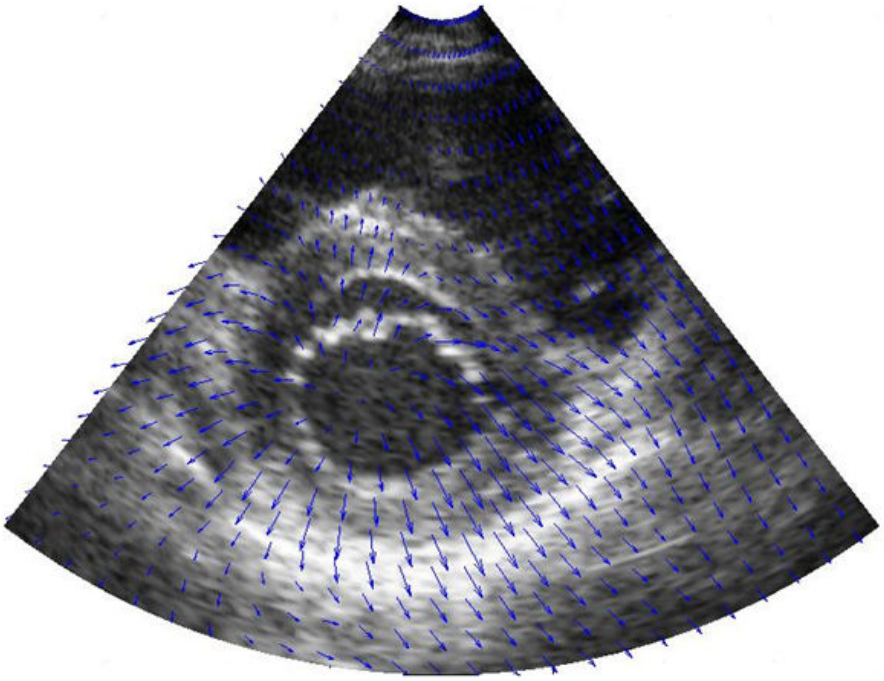


FIGURE 4.8 – Échographie du ventricule gauche en court axe avec champ de déplacement calculé. Phase de diastole.

Cette image du ventricule gauche a été acquise lors de la diastole. Il s'agit de la période au cours de laquelle le cœur se relâche pour se remplir de sang. Notre algorithme permet d'observer ce relâchement et de deviner le mouvement de la paroi du ventricule.

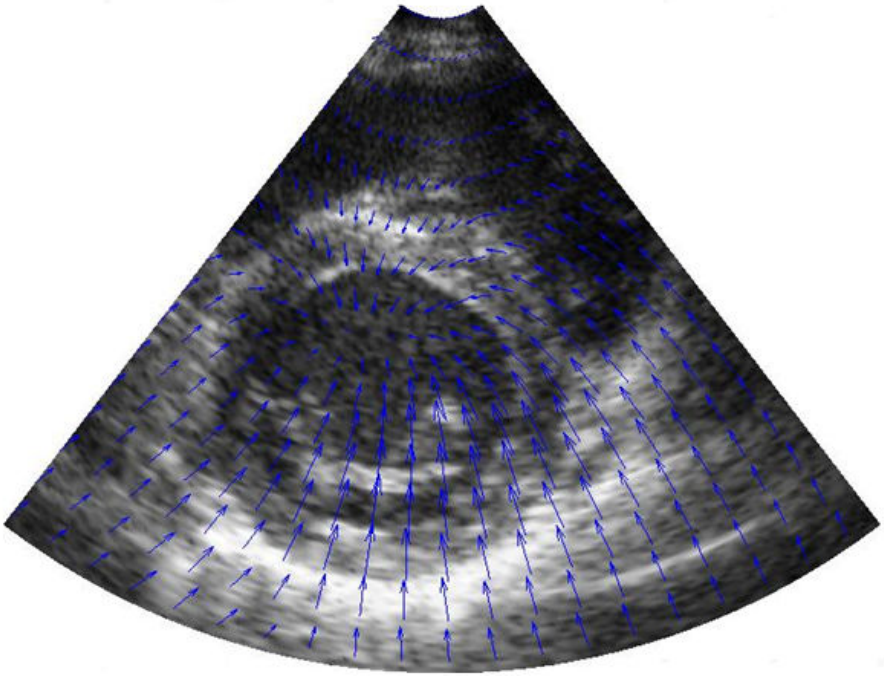


FIGURE 4.9 – Échographie du ventricule gauche en court axe avec champ de déplacement calculé. Phase de systole.

À l'inverse, cette image du même ventricule gauche a été acquise lors de la systole, c'est-à-dire lors de la contraction des chambres du cœur. Ici également, l'algorithme de Horn et Schunck polaire permet de deviner la contraction du ventricule. Notons que dans ces deux images, les flèches situées à l'intérieur du ventricule ne correspondent pas aux déplacements réels. En effet cette zone contient du sang et n'est donc pas échogène (i.e. réfléchit peu les ultrasons). Par conséquent, les déplacements obtenus dans cette zone doivent être vus comme un prolongement lissé des mouvements extérieurs à cette zone.

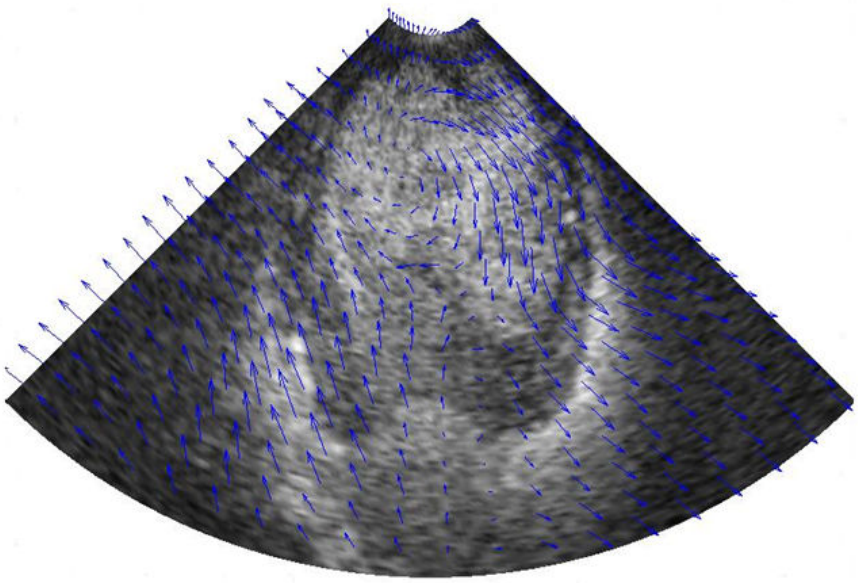


FIGURE 4.10 – Échographie du ventricule gauche en long axe avec champ de déplacement calculé. Agent de contraste dans le sang.

Cette dernière image est issue d'une autre échographie, avant laquelle un agent de contraste a été mélangé au sang. Cela permet de rendre le sang plus échogène et d'étudier son mouvement à l'intérieur du cœur. Il s'agit toujours du ventricule gauche, vu cette fois en *long axe*, c'est à dire en vue verticale. On peut voir l'arrivée du sang à gauche et la création d'un vortex en haut, ce qui correspond à ce qu'on observe en faisant défiler rapidement les images successives. La comparaison des déplacements calculés avec ce que l'on perçoit en regardant le film est la seule validation possible, et nous la considérons pertinente, le cerveau résolvant inconsciemment le problème du flot optique - par quelle méthode ? - pour nous rendre capables de deviner les déplacements.

4.4 Conclusion

Cette partie avait pour cadre la méthode de Horn et Schunck pour le flot optique, qui a été étudiée sous deux aspects :

- D'abord dans un contexte général et mathématiques, nous avons produit une preuve de convergence pour la résolution itérative du système linéaire découlant de cette méthode. Cette preuve, ainsi que des résultats annexes sur les propriétés de ce système et la possibilité d'utiliser d'autres méthodes itératives pour le résoudre, font l'objet de l'article proposé en annexe.
- Puis dans le contexte de l'imagerie médicale, et en particulier de l'échographie cardiaque, nous avons testé la pertinence d'utiliser cet algorithme pour résoudre le flot optique. Une étape intermédiaire a été de l'adapter à des images obtenues en coordonnées polaires, ce qui est presque toujours le cas en échographie cardiaque. Puis nous l'avons testé sur des cas allant de l'image analytique à l'image réelle, et avons conclu sur son efficacité.

Sur ce deuxième point, la comparaison avec des méthodes dites locales, comme celle de Lucas et Kanade [17] a été faite - à l'avantage de la méthode de Horn et Schunck - mais n'a pas été présentée, car ces méthodes locales peuvent être paramétrées et améliorées de multiples manières, de sorte qu'on ne pourrait conclure à leur inefficacité en se contentant de leur mise en œuvre la plus simple. C'est d'ailleurs ces méthodes locales qui, une fois optimisées, sont le plus souvent mises en œuvre au sein des outils d'imagerie industriels. Cependant, les algorithmes utilisés par les entreprises du secteur étant sous secret industriel, nous n'avons pu opérer la comparaison avec notre méthode. Cette partie avait donc pour but, plus que de comparer la méthode simple et ancienne de Horn et Schunck à des méthodes industrielles optimisées et éprouvées, de montrer :

- que l'on pouvait appliquer la méthode de Horn et Schunck, et plus précisément la famille des méthodes globales - cette dernière méthode pouvant elle aussi être largement améliorée -, à l'échographie cardiaque,
- qu'en dépit de la difficulté induite par l'acquisition en coordonnées polaire des données issues de l'échographie cardiaque, une interpolation sur une grille cartésienne - et la nécessaire perte d'informations qui en découle - n'était pas nécessaire, la méthode de Horn et Schunck pouvant être adaptée avec succès à ce système de coordonnées.

Ainsi, notre travail ne prétend pas reconsidérer la manière dont doivent être programmés les outils industriels, qui emploient des méthodes largement améliorées depuis les travaux de Horn et Schunck datant de 1980, et des outils de résolution des systèmes linéaires optimisés. Cependant, la méthode de Horn et Schunck a servi de socle à bon nombre des méthodes actuelles, et a encore aujourd'hui une utilité pédagogique évidente, de sorte que l'écriture d'une preuve de convergence rigoureuse nous a semblé intéressante - d'autant plus que d'autres travaux prétendaient à tort avoir répondu au problème -. Par ailleurs, son application réussie à un domaine spécifique de l'imagerie montre que l'on peut obtenir une bonne approximation du flot optique grâce à un programme simple et des moyens numériques modestes, sans nécessairement recourir à des méthodes industrielles complexes.

Conclusion générale

Comme point commun aux travaux présentés ici, nous pouvons citer la recherche d'une plus grande efficacité en temps de calcul dans la résolution d'équations aux dérivées partielles.

Le premier chapitre avait pour cadre la simulation numérique des impacts de vague. Des modèles bifluides $2D$ fins et complets existaient avant ce travail et permettaient une résolution précise du problème. Nous avons cherché à créer le modèle le plus simple et le plus efficace qui soit capable de considérer la déformation longitudinale du bloc d'eau. En plus de l'intérêt évident apporté par la décroissance des temps de calcul, nous pensons qu'en isolant les processus prépondérants intervenant dans l'évolution d'un phénomène physique, de façon à y limiter les efforts de simulation, nous allons de plus dans le sens d'une meilleure compréhension de ce phénomène. Nous savons maintenant que les extrémités latérales d'un bloc d'eau tombant dans le gaz descendent plus vite que son centre, et ont donc tendance à emprisonner une poche de gaz, ce qui n'est pas sans conséquence sur la tenue de la paroi au choc. Le modèle ayant mené à ce constat étant relativement simple par comparaison à un code bifluide $2D$ complet, nous pouvons, de plus, facilement décrypter les étapes qui mènent le système dans cette situation.

Le deuxième chapitre se voulait plus général et avait pour objectif la mise en place d'une méthodologie efficace de raffinement automatique de maillage, afin notamment d'optimiser le temps de calcul à précision donnée. L'apparition de termes non conservatifs dans les modèles à plusieurs espèces a entraîné des difficultés théorique qui, afin de les traiter de la manière la plus générique possible, nous ont amené à produire des résultats généraux sur les conséquences de l'invariance galiléenne pour un modèle conservatif. Cela conduit à une remise en question, dans le modèle à trois espèces et deux phases (7 équations), de la façon dont est corrigé le flux pour assurer à certaines matrices du système de bonnes propriétés mathématiques. Ce dernier point n'a toutefois pas pu s'accompagner d'une alternative réellement satisfaisante permettant la prise en compte des maillages mobiles, et nous n'avons pas pu élargir la méthode élaborée à ce dernier modèle.

Le troisième chapitre, plus spécifique, visait toujours à optimiser un temps de calcul, pour la simulation numérique des systèmes composé d'eau liquide, d'air et de vapeur d'eau. Nous basant sur des travaux existants portant sur :

- l'élaboration d'un modèle fin et complet pour ce type de systèmes, considérant notamment une température et une vitesse pour chaque phase,

- l'élaboration d'un modèle basé sur une hypothèse d'homogénéité des vitesses et des températures entre les phases, ayant le considérable avantage d'être conservatif,

nous avons opéré un couplage permettant de situer chaque modèle dans la zone géométrique où l'on peut tirer parti de ces points forts, qui sont respectivement la précision et l'efficacité de calcul. Il s'agissait donc, là aussi, d'optimiser la ressource numérique utilisée à phénomène physique donné.

Le quatrième chapitre concernait un autre domaine pour lequel la résolution efficace d'équations aux dérivées partielles est un enjeu important : l'imagerie. En particulier, le problème du flot optique, c'est à dire la recherche d'un déplacement à partir d'une suite d'images, nous a intéressés. Nous nous sommes concentrés sur un algorithme très connu de calcul du flot optique, celui de Horn et Schunck, qui consiste en substance à résoudre un système linéaire par une méthode itérative. La problématique de résolution efficace de ce système est primordiale étant donné sa taille qui peut devenir très importante dans les applications réelles, en particulier avec l'essor de l'imagerie $3D$. Une preuve originale de convergence de cet algorithme a été proposée ainsi qu'une généralisation de la méthode en dimension quelconque. Nous recentrant par la suite sur les applications, nous avons adapté cet algorithme au type d'images rencontré en échographie cardiaque, et avons validé son efficacité dans ce contexte.

Ainsi, sur ces différents sujets, nous avons pu contribuer à une meilleure compréhension et une utilisation plus efficace de divers outils permettant de résoudre les équations aux dérivées partielles. Si les capacités des calculateurs, en termes de mémoire et de temps d'exécution, croissent de façon vertigineuse depuis quelques décennies, la problématique de l'optimisation mathématique des méthodes de résolution n'en demeure pas moins un sujet primordial. D'une part, parce que c'est la conciliation des progrès technologiques et algorithmiques qui permet d'entrevoir sans cesse de nouvelles applications, rendant calculable en un temps raisonnable ce qui ne l'était pas auparavant. D'autre part - et c'est surtout sur ce point que s'est concentré notre travail -, parce que la création d'outils simples fonctionnant rapidement sur des machines standards est un enjeu pour nombre d'industriels, qui ne disposent pas des moyens ou compétences pour s'engager dans la voie du calcul haute performance, et qui souhaitent des réponses simples à des questions complexes.

Bibliographie

[1] Jean-Michel Ghidaglia, Anela Kumbaro, Gérard Le Coq, *On the numerical solution to two fluid models via a cell centered finite volume method.*
Eur. J. Mech. B - Fluids, volume 20, issue 6 (2001), pp. 841–867.

[2] C. Schnorr, *Determining optical flow for irregular domains by minimizing quadratic functionals of a certain class.*
Int. J. Comput. Vis., volume 6 (1991), issue 1, pp. 25–38.

[3] W. Lafeber , L. Brosset , and H. Bogaert, *Elementary Loading Processes (ELP) involved in breaking wave impacts : findings from the Sloshel project.*
Proceedings of the Twenty-second International Offshore and Polar Engineering Conference, International Society of Offshore and Polar Engineers (2012), Rhodes, Greece, June 17–22, 2012.

[4] Laurent Brosset, Jean-Michel Ghidaglia, Pierre-Michel Guilcher and Louis Le Tarnec, *Generalized Bagnold Model.*
Proceedings of the Twenty-third International Offshore and Polar Engineering, International Society of Offshore and Polar Engineers (2013), Anchorage, Alaska, USA, June 30 – July 5, 2013.

[5] Bagnold R., *Interim report on wave-pressure research.*
J. Inst. Civil Engineers, volume 12 (1939), pp. 201-226.

[6] Sylvain Faure, Jean Michel Ghidaglia, *Violent flows in aqueous foams I : Physical and numerical models.*
European Journal of Mechanics B - Fluids, volume 30 (2011), pp. 341–359.

[7] Weizhang Huang, Robert Russell, *Adaptive Moving Mesh Methods.*
Springer, 2010.

[8] A. Bernard-Champmartin, O.Poujade , J. Mathiaud, J.M.Ghidaglia, *Modelling of a balanced homogeneous mixture model (HEM).*
Acta Appl Math, volume 129 (2014), pp. 1-21.

[9] F. De Vuyst, J.M. Ghidaglia and G. Le Coq, *On the numerical simulation of multiphase water flows with changes of phase and strong gradients using the Homogeneous Equilibrium Model.*
International Journal of Finite Volumes, volume 2 (2005), pp. 1-36.

[10] Saad Benjelloun, *Etude mathématique et numérique de modèles de mécanique des fluides et de systèmes gaz-particules.*
Thèse de doctorat soutenue à l'ENS Cachan (2012).

[11] Alexandre Balmont, *Simulation numérique d'un problème de mécanique des fluides. Étude des effets de la compressibilité dans le problème du piston.*
Rapport de stage soutenu à l'ENS Cachan (2012).

[12] Jean-Philippe Brauenig, *Sur la simulation d'écoulements multi-matériaux par une méthode eulérienne directe avec capture d'interfaces en dimensions 1, 2 et 3.*
Thèse de doctorat soutenue à l'ENS Cachan (2007).

[13] Maud Meriaux, Serge Piperno, *Adaptation dynamique de maillage pour les lois de conservation hyperboliques en une dimension.*
Rapport de recherche INRIA numéro 4696 (2003).

[14] Toro, Euleterio F., *Riemann Solvers and Numerical Methods for Fluid Dynamics.*
Springer, 1999.

[15] B.A.J. Angelsen, *Ultrasound imaging : waves, signals and signal processing.*
Emantec, 2000.

[16] B.K.P. Horn et B.G. Schunck, *Determining optical flow.*
Technical Symposium East, International Society for Optics and Photonics (181), pp. 319-331.

[17] B.D. Lucas et T.Kanade, *An Iterative Image Registration Technique with an Application to Stereo Vision.*
Proceedings of imaging understanding workshop, volume 81 (1981), pp. 674-679.

[18] D. Bestion, *The physical closure laws in the CATHARE code.*
Nuclear Engineering Design, volume 124 (1990), pp. 229–245.

[19] Thomas, L.H. (1949), *Elliptic problems in linear differentiel equations over a network.*
Waston Sci. Comput. Lab report, Columbia University, New York (1949).

[20] A. Mitiche et A. Mansouri, *On convergence of the Horn and Schunck optical flow method.*
IEEE transactions on image processing, volume 13 (2004), no. 6, pp. 848-852.

[21] Y. Kameda, A. Imiya et N. Ohnishi, *A convergence proof of the Horn and Schunck optical flow computation scheme using neighborhood decomposition.*
Combinatorial Image Analysis, Springer (2008), pp. 262-273.

[22] P.G. Ciarlet, *Introduction à l'analyse numérique matricielle et à l'optimisation.*
Masson, 1988.

[23] Jean-Michel Ghidaglia, *Surrogate Models for Sloshing and Scaling.*
Communication personnelle (2013).

Troisième partie

Annexe : convergence des itérations de Horn Schuck

SIAM J. IMAGING SCIENCES
Vol. 7, No. 1, pp. 277–293

A Proof of Convergence of the Horn–Schunck Optical Flow Algorithm in Arbitrary Dimension*

Louis Le Tarnec†, François Destrempes‡, Guy Cloutier§, and Damien Garcia¶

Abstract. The Horn–Schunck (HS) method, which amounts to the Jacobi iterative scheme in the interior of the image, was one of the first optical flow algorithms. In this paper, we prove the convergence of the HS method whenever the problem is well-posed. Our result is shown in the framework of a generalization of the HS method in dimension $n \geq 1$, with a broad definition of the discrete Laplacian. In this context, the condition for the convergence is that the intensity gradients not all be contained in the same hyperplane. Two other works ([A. Mitiche and A. Mansouri, *IEEE Trans. Image Process.*, 13 (2004), pp. 848–852] and [Y. Kameda, A. Imiya, and N. Ohnishi, *A convergence proof for the Horn-Schunck optical-flow computation scheme using neighborhood decomposition*, in Combinatorial Image Analysis, Springer, Berlin, 2008, pp. 262–273]) claimed to solve this problem in the case $n = 2$, but it appears that both of these proofs are erroneous. Moreover, we explain why some standard results about the convergence of the Jacobi method do not apply for the HS problem, unless $n = 1$. It is also shown that the convergence of the HS scheme implies the convergence of the Gauss–Seidel and successive overrelaxation schemes for the HS problem.

Key words. optical flow, Horn–Schunck algorithm, Jacobi iterations

AMS subject classifications. 68U10, 11D04, 05C50, 65F10, 65N22, 65F35

DOI. 10.1137/130904727

1. Introduction.

Optical flow refers to the distribution of apparent movement of intensity patterns in an image caused by the relative motion between an observer and the scene. The Horn–Schunck (HS) method was one of the first optical flow algorithms used to determine a displacement field from several successive images [10]. The original HS method is based on a global approach and introduces a quadratic prior term of smoothness in the classical equation of the optical flow. This algorithm is especially adapted to speckled or diffuse images like

*Received by the editors January 4, 2013; accepted for publication (in revised form) October 14, 2013; published electronically January 30, 2014. This work was supported by the Canadian Institutes of Health Research (Operating grant MOP-106465) and the Natural Sciences and Engineering Research Council of Canada (Discovery grant 138570-2011 and Strategic grant STPGP-381136-09).
http://www.siam.org/journals/siims/7-1/90472.html

†RUBIC (Research Unit of Biomechanics and Imaging in Cardiology), University of Montreal Hospital Research Center (CRCHUM), Montréal, QC H2L-2W5, Canada (louisletarnec@gmail.com).

‡Laboratory of Biorheology and Medical Ultrasonics, University of Montreal Hospital Research Center (CRCHUM), Montréal, QC H2L-2W5, Canada (francois.destrempes@crchum.qc.ca).

§Laboratory of Biorheology and Medical Ultrasonics, University of Montreal Hospital Research Center (CRCHUM), Montréal, QC H2L-2W5, Canada, Department of Radiology, Radio-Oncology and Nuclear Medicine, University of Montreal, Montréal, QC H3T-1J4, Canada, and Institute of Biomedical Engineering, University of Montreal, Montréal, QC H3T-1J4, Canada (guy.cloutier@umontreal.ca).

¶RUBIC (Research Unit of Biomechanics and Imaging in Cardiology), University of Montreal Hospital Research Center (CRCHUM), Montréal, QC H2L-2W5, Canada, Department of Radiology, Radio-Oncology and Nuclear Medicine, University of Montreal, Montréal, QC H3T-1J4, Canada, and Institute of Biomedical Engineering, University of Montreal, Montréal, QC H3T-1J4, Canada (garcia.damien@gmail.com).

those encountered in several modalities where a displacement field without discontinuity or significantly high gradients is expected [24, 16, 14]. Thus, the HS method and its derived forms remain of high interest in some areas of motion imaging. Other complex and very proficient estimators for optical flow, however, now exist in the context of natural scenes [2, 3] to take into account discontinuities at object edges.

Based on a discretization of the differential operators appearing in the HS optical flow formulation, the HS method results in a linear system that can be solved with direct or iterative methods. In comparison to direct methods, the iterative solvers have the advantage of needing lower computational data storage and to be easily programmable. It is well known that the matrix involved in the HS linear system is symmetric positive definite, as a consequence of the V-ellipticity of the HS functional [18]. This ensures, for example, the efficiency of the direct Cholesky decomposition and of the iterative Gauss–Seidel or successive overrelaxation (SOR) solvers. However, the positive definiteness does not permit one to conclude about the method proposed in Horn and Schunck's initial paper, which is an iterative 2×2 blockwise solver [10] and corresponds to the Jacobi solver for the interior points of the image only. In fact, it is shown in this work that the positive definiteness is implied by the convergence of the HS scheme. The Gauss–Seidel and SOR solvers are known to converge at least twice as fast as the Jacobi solver [4, Theorem 5.3-4]. These iterative solvers can be made parallelizable using, for instance, a special red-black reordering of the unknowns in the linear system [6, 25]. The Jacobi iterative solver, however, has the advantage of being directly parallelizable since it does not use values computed in the current iteration step [20, 21].

One known general result about the convergence of the Jacobi method concerns strictly (block) diagonally dominant matrices, which is not the case here. Another result concerns (block) irreducible and weakly dominant matrices, an assumption which is not satisfied for images of dimension greater than 1 under the appropriate Neumann boundary conditions. Whether the iterative method for the HS linear system with the Neumann boundary conditions converges still remains unsolved. Indeed, the paper of Horn and Schunck did not include a proof of convergence [10]. Two proofs of convergence have been published since then, in [17] and [13], both for 2-dimensional images. However, as far as we can tell, these two proofs are erroneous. There is also a short argument in [23, p. 249] based on diagonally dominant matrices (without blocks), for the convergence of the pointwise Jacobi method, that is erroneous.

Under a general perspective, there are three main points in an optical flow algorithm: (1) the formulation of the continuous energy (functional) to be minimized; (2) the discretization scheme; and (3) the solver used to minimize the energy. The scope of this work is to present a proof that the HS iterative solver (and hence the Gauss–Seidel and SOR solvers) converges for the original quadratic HS functional under a generic discretization scheme adopted in this paper.

In section 2, we state a generalization of the HS method in dimension n. In section 3, we explain why the previous proofs are erroneous and cannot be fixed. In section 4, we define some hypotheses about the discrete Laplacian, propose a necessary and sufficient condition for the linear system of Horn and Schunck to be invertible, and state our convergence result. The proof is presented in section 5. In section 6, we define a general way of calculating a discrete Laplacian in dimension n. In section 7, we show that our general discrete Laplacian satisfies the hypotheses imposed to get the convergence result. In Appendix A, the HS iterative

scheme is derived in detail from the discretization of the HS problem. In Appendix B, it is explained why the coefficient matrix of the HS scheme is not strictly (block) diagonally dominant matrices, nor (block) irreducible and weakly dominant matrices under the appropriate Neumann boundary conditions for images of dimension greater than 1. A result is shown in Appendix C that implies the convergence of the Gauss–Seidel and SOR iterative schemes whenever the Jacobi method converges, under appropriate conditions. This result also implies that the Gauss–Seidel and SOR methods converge for the HS problem, as a consequence of the convergence of the HS iterative scheme. In Appendix D, details are given to explain why the proofs of [17, 13] are erroneous.

2. Statement of the problem. The optical flow problem is usually applied to 2-dimensional images of a moving scene [10]. Optical flow has also been used to analyze motion in one, three, or four dimensions [22, 9, 5]. In this work, we investigate the convergence of the HS optical flow problem in the generalized case of dimension $n \geq 1$. We thus consider an orthotope $V \subset \mathbb{R}^n$, i.e., a parallelotope whose edges are all mutually perpendicular (a segment if $n = 1$, a rectangle if $n = 2$, or a cuboid if $n = 3$). In the optical flow problem, each element of V generally corresponds to an intensity or brightness that varies over time. Given the intensity field over two or more successive instants, the aim of the HS method is to determine the corresponding displacement field. As we propose here a proof of convergence in the context of n-dimensional arrays, we first state an n-dimensional generalization of the HS method (for the classical 2-dimensional formulation, we refer the reader to [10]).

Let I denote the intensity field on V, I_t its derivative with respect to time t, ∇I its gradient with respect to position, and $\mathbf{u}$ the displacement field. We start from the well-known optical flow identity:

$$\nabla I \cdot \mathbf{u} + \mathbf{I_t} = \mathbf{0}, \tag{2.1}$$

which means that a given (apparently moving) point of V keeps its initial intensity during its displacement. Then, a regularization method is employed to impose low spatial variations in the displacement field. By definition, the HS method consists in minimizing the unconstrained functional:

$$J(\mathbf{u}) = \int_V \{(\nabla I \cdot \mathbf{u} + I_t)^2 + \mu \, \|\nabla \mathbf{u}\|^2\} \, dV, \tag{2.2}$$

where $\mu > 0$ is a positive real number and $\| \cdot \|$, in the entire paper, represents the Euclidean norm. The Euler–Lagrange equation corresponding to this minimization problem reads as follows:

$$\mu \triangle \mathbf{u} = (\nabla I \cdot \mathbf{u} + I_t) \nabla I = [\nabla I \, \nabla I^T] \, \mathbf{u} + I_t \nabla I \quad \text{on } V, \tag{2.3}$$

$$\frac{\partial \mathbf{u}}{\partial \mathbf{n}} = 0 \quad \text{on } \partial V. \tag{2.4}$$

Here, $\frac{\partial}{\partial \mathbf{n}}$ is the differentiation operator in the direction of the normal $\mathbf{n}$ to the boundary ∂V, and the superscript T denotes transposition of matrices. The displacements without the superscript T are considered to be column vectors (in $\mathbb{R}^n$). Note that the Neumann boundary conditions (2.4) arise naturally from the unconstrained minimization problem (2.2).

Now, we will discretize the expression of (2.3) on a lattice Λ covering the orthotope V. The restriction of the intensity field on the lattice Λ can be viewed as a (discretized) image. In what follows, we assume that there are $N \geq 2$ elements in the lattice Λ. Then, a discretized displacement field is of the form $\mathbf{u} = (\mathbf{u}_i)_{i\in\Lambda}$, where $\mathbf{u}_i = (u_{i1}, \ldots, u_{in})^T$ denotes the displacement vector at the point i. In the following, we will say that a displacement field $\mathbf{u}$ is uniform if all the displacement vectors $\mathbf{u}_i$ are identical. Now, (2.3) can be written for $i \in \Lambda$ as

$$\mu \triangle(\mathbf{u})_i = [\nabla I \, \nabla I^T]_i \, \mathbf{u}_i + I_{t,i} \, \nabla I_i, \tag{2.5}$$

where $I_{t,i}$ denotes the partial derivative of the intensity I with respect to t evaluated at the point i. In (2.5), $\triangle(\mathbf{u})_i$ is a discretized Laplacian that depends linearly on the vectors $\mathbf{u}_i$ for $i \in \Lambda$. Hence, the consideration of (2.5) for $i \in \Lambda$ yields a linear system of $n\,N$ equations and $n\,N$ unknowns, where N is the number of points in Λ. The discretization of the Laplacian can classically be written as

$$\triangle(\mathbf{u})_i = \kappa \, \{M(\mathbf{u})_i - \mathbf{u}_i\}, \tag{2.6}$$

where $\kappa > 0$ is a positive real number and M a linear transformation (on the vector space of displacement fields) that returns for each point an average of the displacement field over its neighbors:

$$M(\mathbf{u})_i = \sum_{j\in\Lambda} \lambda_{ij} \, \mathbf{u}_j, \tag{2.7}$$

where λ_{ij}, for $i, j \in \Lambda$, are nonnegative real numbers. We will adopt in section 6 a general expression of this operator. In the following, we denote for notational convenience

$$\alpha = \mu \, \kappa, \tag{2.8}$$

where μ is the regularization weight of (2.2), so that the coefficient α is a positive real number. In order to solve the linear system (2.5), Horn and Schunck [10] proposed an iterative method that is assumed to converge to the solution. Let P be the linear transformation (on the vector space of displacement fields) defined by $P(\mathbf{u})_i = P_i \, \mathbf{u_i}$ for $i \in \Lambda$, where $P_i = \mathcal{I}_n - \frac{[\nabla I \, \nabla I^T]_i}{\alpha + \|\nabla I_i\|^2}$ and $\mathcal{I}_n$ is the $n \times n$ identity matrix. Let $\mathbf{d}$ be the displacement field defined by $\mathbf{d}_i = -\frac{I_{t,i} \nabla I_i}{\alpha + \|\nabla I_i\|^2}$ for $i \in \Lambda$. Then, the HS iterative scheme is expressed as follows:

$$\mathbf{u}^{k+1} = P \, M\big(\mathbf{u}^k\big) + \mathbf{d}. \tag{2.9}$$

See Appendix A for a derivation of (2.9). Also, it is shown in Appendix B that the HS iterative scheme of (2.9) amounts to the Jacobi iterative scheme in the interior of the orthotope V, but never at its boundary points. Moreover, in that appendix, we explain why standard results (based on block diagonally dominant matrices) on the convergence of the Jacobi iterative scheme do not apply in this context, due to the natural Neumann boundary conditions (2.4). We also explain why the short argument of [23, p. 249] based on diagonally dominant matrices (without blocks) for the pointwise Jacobi method is erroneous for the HS problem.

In Appendix C, it is shown that the convergence of the Gauss–Seidel and SOR iterative schemes is implied by the convergence of the Jacobi method, under appropriate conditions, based on a result about symmetric positive definite matrices. In particular, this result implies that the Gauss–Seidel and SOR methods converge for the HS problem, as a consequence of the convergence of the HS iterative scheme.

It would be straightforward to prove the convergence of the Jacobi solver in the presence of Dirichlet boundary conditions since the matrix would be block irreducible and weakly block diagonally dominant in that case. However, we recall that the Neumann conditions are intrinsically related to the minimization of the cost function (2.2).

3. Previous proofs. As stated in the introduction, the two existing proofs of convergence of the Jacobi solver in the context of the HS problem ([17] and [13]) appear to be erroneous. Let us now see in detail where the errors occurred and why we think that they cannot be fixed.

In [17], the cornerstone of the proof of convergence of the Jacobi method for solving the HS linear system relies on [17, eq. (16)], which states that the function defined by the matrix "P" of [17, eq. (9)] (not to be confused with the linear transformation P of (2.9) of the present paper) is contracting for the norm defined by [17, eq. (10)], for any image. However, it appears that the only case for which this can be true is if the image is uniform, as explained in detail in Appendix D.

In [13, eq. (20)], we notice that no condition is given for the convergence of the HS iterations. However, in view of Theorem 4.1 below, that assertion is false (a condition on the image gradients is needed to make the HS problem well-posed). Thus, the proof in [13] must be erroneous. In Appendix D, we give further details on intermediate statements that are false in [13].

Finally, in [23, p. 249], the special case of the HS problem amounts to $\Psi'(s^2) = 1$ (see also [23, p. 247, second column]). In that case, the iterations of [23, eqs. (14) and (15)] amount to the Jacobi iterative scheme for the system (2.5), but without considering $n \times n$ blocks. It is asserted that "If the discrete image gradient does not vanish at one point, the system matrix of these equations is irreducibly diagonally dominant. This guarantees the existence of a unique solution of the linear system and global convergence of the Jacobi iterations [26]". But, as shown in Appendix B, the coefficient matrix of the system is not even diagonally dominant, except in a special case. Thus, that argument is also erroneous.

4. Statement of the main result. First, the operator M of (2.6) and (2.9) comes from the Laplacian discretization and returns, for each point, an average of the displacement field over its neighbors as in (2.7). We will have to define several hypotheses about M. In what follows, we will assume the following:

(H1) For all points i and j of Λ, $\lambda_{ij} = \lambda_{ji}$.

(H2) At every point i of Λ, $\sum_{j\in\Lambda} \lambda_{ij} = 1$.

Intuitively, (H1) comes from the isotropy property of the smooth Laplacian, and (H2) is necessary in order to have a null Laplacian when the displacement field is uniform. As we will see in section 7, these hypotheses are verified with the general discretization scheme of section 6. In order to state the last hypothesis, we have to define the graph G by its set of vertices $V(G) = \Lambda$ and its set of edges $E(G) = \{(i,j) \in \Lambda^2 : \lambda_{ij} \neq 0\}$. If $(i,j) \in E(G)$, we write $i \sim_G j$.

We assume that the lattice Λ is of the form $\{(i_1, i_2, \ldots, i_n) : i_\ell$ is an integer ranging from 0 to $N_\ell - 1$ for $1 \leq \ell \leq n\}$, where $N_\ell \geq 1$ for $1 \leq \ell \leq n$. Thus, the number of points in Λ is equal to $N = \prod_{\ell=1}^{n} N_\ell$. From (H1), this graph is undirected (i.e., $i \sim_G j$ if $j \sim_G i$). Let us now recall that an undirected graph G is connected if, for any two vertices i and j of G, there exists a path from i to j in G. We can now state the last hypothesis:

(H3) The graph G is connected.

We will see in section 7 that (H3) is also true with the general discretization scheme of section 6. Actually, this is an immediate consequence of the fact that the closest neighbors of a point are taken into account in the average calculation at this point. In the following, we will call *rank of* (∇I_i) the dimension of the subspace of $\mathbb{R}^n$ that is spanned by the vectors ∇I_i, $i \in \Lambda$.

Theorem 4.1. *Under hypotheses* (H1), (H2), *and* (H3), *the following hold:*

- *If the rank of* (∇I_i) *is* n, *the linear system* (2.5) *has a unique solution and the iterations* (2.9) *converge to this solution.*
- *If the rank of* (∇I_i) *is not* n, *the problem is ill-posed; i.e., the linear system* (2.5) *does not have a unique solution.*

Let us notice that the rank of (∇I_i) is different from n if and only if the intensity gradients are all contained in the same hyperplane. In that case, the image is invariant along the direction orthogonal to this hyperplane. The fact that this condition makes the problem ill-posed is not surprising, as it is clear that a displacement along this particular direction cannot be detected by studying the variations of intensity over time.

5. Proof of the main result. The linear transformation M and the coefficients λ_{ij} are defined in (2.7). The linear transformation P and the matrix P_i are defined in section 2 before (2.9). We define the norm of a displacement field $\mathbf{u}$ by $\|\mathbf{u}\| = \left(\sum_{i\in\Lambda} \|\mathbf{u}_i\|^2\right)^{1/2}$, where $\|\mathbf{u}_i\|$ is the Euclidean norm on $\mathbb{R}^n$.

Lemma 5.1. *Under hypotheses* (H1) *and* (H2), *the following hold:*

- *For every displacement field* $\mathbf{u}$, $\|M(\mathbf{u})\| \leq \|\mathbf{u}\|$.
- *If equality holds, then for any two points* $i \sim_G j$, *we have* $M(\mathbf{u})_i = \mathbf{u}_j$.

Proof. For each point i, we get by hypotheses (H1) and (H2) that $\sum_{j\in\Lambda} \lambda_{ij} = \sum_{j\in\Lambda} \lambda_{ji} = 1$. Then, for each direction ℓ, Jensen's inequality [12] applied to the strictly convex function $x \to x^2$ yields

$$\left[\sum_{j\in\Lambda} \lambda_{ij}\, u_{j\ell}\right]^2 \leq \sum_{j\in\Lambda} \lambda_{ij}\, u_{j\ell}^2. \tag{5.1}$$

We can now write

$$\begin{aligned}
\|M(\mathbf{u})\|^2 &= \sum_{\ell=1}^{n} \sum_{i\in\Lambda} \left[\sum_{j\in\Lambda} \lambda_{ij}\, u_{j\ell}\right]^2 \\
&\leq \sum_{\ell=1}^{n} \sum_{i\in\Lambda} \sum_{j\in\Lambda} \lambda_{ij}\, u_{j\ell}^2 = \sum_{\ell=1}^{n} \sum_{j\in\Lambda} \sum_{i\in\Lambda} \lambda_{ij}\, u_{j\ell}^2 \\
&= \sum_{\ell=1}^{n} \sum_{j\in\Lambda} u_{j\ell}^2 \left[\sum_{i\in\Lambda} \lambda_{ij}\right] = \sum_{\ell=1}^{n} \sum_{j\in\Lambda} u_{j\ell}^2 = \|\mathbf{u}\|^2. \qquad (5.2)
\end{aligned}$$

Moreover, let us suppose that $\|M(\mathbf{u})\| = \|\mathbf{u}\|$. Then, for each point i and each ℓ, the equality in (5.1) is reached. Thus, the coordinates $u_{j\ell}$ associated with the nonvanishing coefficients λ_{ij} are all identical (cf. [8]). This means that $\mathbf{u}_j = \mathbf{u}_{j'}$ for any two points $i \sim_G j$ and $i \sim_G j'$. Therefore, it follows that $M(\mathbf{u})_i = \sum_{j' \in \Lambda} \lambda_{ij'} \mathbf{u}_{j'} = (\sum_{j' \in \Lambda} \lambda_{ij'}) \mathbf{u}_j = \mathbf{u}_j$, where j is any point such that $i \sim_G j$. ∎

Lemma 5.2. *For every displacement field* $\mathbf{u}$*, we have* $\|P(\mathbf{u})\| \leq \|\mathbf{u}\|$*. The equality holds if and only if* $\mathbf{u}_i$ *is orthogonal to the gradient* ∇I_i *at any point* i *of* Λ*. In that case,* $P(\mathbf{u})_i = \mathbf{u}_i$ *at any point* i *of* Λ*.*

Proof. Let $\mathbf{u}$ be a displacement field and i a point of the image. There exist a vector $\vec{a}$ of $\mathbb{R}^n$ and a real number b such that $\nabla I_i^T \vec{a} = 0$ and $\mathbf{u}_i = \vec{a} + b \nabla I_i$. From the expression of P_i, we find $P_i \vec{a} = \vec{a}$ (because $\nabla I_i^T \vec{a} = 0$) and $P_i \nabla I_i = \big(1 - \frac{\|\nabla I_i\|^2}{\alpha + \|\nabla I_i\|^2}\big) \nabla I_i$. Thus, $P_i \mathbf{u}_i = \vec{a} + b \nabla I_i \big(1 - \frac{\|\nabla I_i\|^2}{\alpha + \|\nabla I_i\|^2}\big)$. We notice here that $\|\mathbf{u}_i\|^2 = \|\vec{a}\|^2 + b^2 \|\nabla I_i\|^2$ and $\|P_i \mathbf{u}_i\|^2 = \|\vec{a}\|^2 + \big(1 - \frac{\|\nabla I_i\|^2}{\alpha + \|\nabla I_i\|^2}\big)^2 b^2 \|\nabla I_i\|^2$. So, we get that $\|P_i \mathbf{u}_i\| \leq \|\mathbf{u}_i\|$, and that the equality holds if and only if $b = 0$ or $\nabla I_i = \vec{0}$ (i.e., $\mathbf{u}_i = \vec{a}$), which means if and only if $\nabla I_i^T \mathbf{u}_i = 0$. Finally, $\|P(\mathbf{u})\|^2 = \sum_{i \in \Lambda} \|P_i \mathbf{u}_i\|^2 \leq \sum_{i \in \Lambda} \|\mathbf{u}_i\|^2 = \|\mathbf{u}\|^2$, with equality if and only if $\nabla I_i^T \mathbf{u}_i = 0$ at every point i of Λ. In that case, it is clear that $P_i \mathbf{u}_i = \mathbf{u}_i$ at every point i of Λ. ∎

Let us recall that the HS iterations read as $\mathbf{u}^{k+1} = P\,M(\mathbf{u}^k) + \mathbf{d}$. We can now show the convergence of these iterations, under our condition on the intensity field. From Lemmas 5.1 and 5.2, we find that $\|PM(\mathbf{u})\| \leq \|\mathbf{u}\|$ for any displacement field $\mathbf{u}$. A feature of the following proof consists in showing that $\|(PM)^N(\mathbf{u})\| < \|\mathbf{u}\|$ for any nonzero displacement field $\mathbf{u}$, where N is the number of points in Λ.

Proof of Theorem 4.1. We still suppose that (H1), (H2), and (H3) are verified. Let us assume that the rank of (∇I_i) is n. Let us assume, by contradiction, that there is a displacement field $\mathbf{u} \neq \mathbf{0}$ such that $\|(PM)^N(\mathbf{u})\| = \|\mathbf{u}\|$. So, there is a point $i_* \in \Lambda$ such that $\mathbf{u}_{i_*} \neq \vec{0}$. Let i be any point of Λ. We claim that there is a path from i to i_* in the graph G of length $1 \leq L \leq N$ (N is the number of elements in G). Indeed, if $i \neq i_*$, then a minimal path will do; if $i = i_*$, then the path $i_0 = i \sim_G i_1 \sim_G i_* = i$ will do, where i_1 is any neighbor of i in the graph G. Let this path be of the form $i_0 = i \sim_G i_1 \sim_G i_2 \sim_G \cdots \sim i_L = i_*$. From Lemmas 5.1 and 5.2, the assumption $\|(PM)^N(\mathbf{u})\| = \|\mathbf{u}\|$ implies that $\|(PM)^L(\mathbf{u})\| = \|M(PM)^{L-1}(\mathbf{u})\| = \|(PM)^{L-1}(\mathbf{u})\| = \cdots = \|\mathbf{u}\|$. Moreover, again from Lemmas 5.1 and 5.2, we have $(PM)^L(\mathbf{u})_{i_0} = M(PM)^{L-1}(\mathbf{u})_{i_0} = (PM)^{L-1}(\mathbf{u})_{i_1} = \cdots = \mathbf{u}_{i_L}$. Also, from Lemma 5.2, we have that $M(PM)^{L-1}(\mathbf{u})_{i_0}$ is orthogonal to ∇I_{i_0} and thus that $\mathbf{u}_{i_L} = \mathbf{u}_{i_*}$ is orthogonal to $\nabla I_{i_0} = \nabla I_i$. Since the point i is arbitrary, we deduce that the space spanned by the gradient vectors ∇I_i is orthogonal to the nonzero vector $\mathbf{u}_{i_*}$, which is a contradiction. Thus, under the condition of convergence stated in the theorem and for $\mathbf{u} \neq \mathbf{0}$, we have $\|(P\,M)^N(\mathbf{u})\| < \|\mathbf{u}\|$.

We now consider the function $\mathbf{u} \to \|(P\,M)^N(\mathbf{u})\|$ defined on the hypersphere $\{\mathbf{u} \,|\, \|\mathbf{u}\| = 1\}$. This function is continuous and defined on a compact set, i.e., a bounded closed subset of the vector space of displacement fields. Therefore, the function is bounded and reaches its maximal value. This ensures that there exists $\beta < 1$ such that for every displacement field $\mathbf{u}$, $\|(P\,M)^N(\mathbf{u})\| \leq \beta \|\mathbf{u}\|$. Since, moreover, $\|PM(\mathbf{u})\| \leq \|\mathbf{u}\|$, the conclusion about the existence of a solution for the linear system (2.5), its uniqueness, and the convergence of the iterations (2.9) to this solution is then a classical result (see [19, p. 101], for example).

We now suppose that the rank of (∇I_i) is less than n. In this case, the intensity gradients are all contained in the same hyperplane. Let us consider a displacement field $\mathbf{u}^*$ that is uniform, different from zero, and orthogonal to this hyperplane. Because of hypothesis (H2), which imposes $\sum_{j\in\Lambda}\lambda_{ij}=1$ at each point i, and because $\mathbf{u}^*$ is uniform, we get $M(\mathbf{u}^*)=\mathbf{u}^*$. Moreover, because $\nabla I_i^T\,\mathbf{u}_i^*=0$ at each point i, Lemma 5.2 says that $P(\mathbf{u}^*)=\mathbf{u}^*$. Thus, $P\,M(\mathbf{u}^*)=\mathbf{u}^*$. This shows that the linear system $\mathbf{u}=P\,M(\mathbf{u})+\mathbf{d}$ (equivalent to the linear system (2.5)) has a nonzero solution when $\mathbf{d}=\mathbf{0}$, so that the coefficient matrix of the linear system (2.5) is not invertible. ∎

6. The discrete Laplacian in dimension n.

6.1. Description of a general scheme. Recall that the lattice Λ is assumed to be of the form $\{(i_1,i_2,\ldots,i_n)\ :\ i_\ell$ is an integer ranging from 0 to $N_\ell-1$ for $1\le\ell\le n\}$, where $N_\ell\ge 1$ for $1\le\ell\le n$. The lattice Λ is viewed as a subset of the Cartesian product $\mathbb{Z}^\ell$. In the following, the norms L^1 and L^∞ are denoted by $\|(i_1,i_2,\ldots,i_n)\|_{L^1}=\sum_{\ell=1}^n|i_\ell|$ and $\|(i_1,i_2,\ldots,i_n)\|_{L^\infty}=\max_{1\le\ell\le n}|i_\ell|$. We will now define a general way of calculating a discrete Laplacian in dimension n, based on [15]. As proposed in [15], we consider the n-dimensional finite-difference stencil S_i around a point i, consisting of the 3^n-1 points $k\in\mathbb{Z}^\ell$ that verify $\|i-k\|_{L^\infty}=1$. Then, we divide these stencil points into the sets $S_i^{(r)}$ $(1\le r\le n)$ of points $k\in\mathbb{Z}^\ell$ that verify $\|i-k\|_{L^1}=r$. As explained in [15], it turns out that for each r in $\{1,\ldots,n\}$, a discretization of the Laplacian can be constructed from the Taylor expansions of the points of $S_i^{(r)}$ about the point i. The remaining part of this section concerns only interior points of the lattice Λ; the boundary cases are discussed in section 6.2. So, if i is not a boundary point, the discretization of the Laplacian is given in [15, formula (2.2)]:

$$\triangle^{(r)}(\mathbf{u})_i=\kappa_r\sum_{k\in S_i^{(r)}}(\mathbf{u}_k-\mathbf{u}_i), \tag{6.1}$$

where $\kappa_r=\frac{2\,n}{r\ \mathrm{Card}(S_i^{(r)})}$. Based on the definition of $S_i^{(r)}$, it is clear that $\mathrm{Card}(S_i^{(r)})=\binom{n}{r}2^r$, where $\binom{n}{r}=\frac{n!}{r!(n-r)!}$. Thus, κ_r is independent of the point i. Then, a general way to calculate a global discrete Laplacian at the point i is to make a weighted average of the Laplacians obtained for the different sets $S_i^{(r)}$. Such a discretization can be written as

$$\triangle(\mathbf{u})_i=\sum_{r=1}^n w_r\,\triangle^{(r)}(\mathbf{u})_i, \tag{6.2}$$

where the weights $w_r\ge 0$ are nonnegative real numbers such that $\sum_{r=1}^n w_r=1$. We also denote $\kappa=\sum_{r=1}^n w_r\,\kappa_r\,\mathrm{Card}(S_i^{(r)})$, independent of i, and $\gamma_r=\frac{w_r\,\kappa_r}{\kappa}$, so that

$$\triangle(\mathbf{u})_i=\kappa\left\{\left(\sum_{r=1}^n\sum_{k\in S_i^{(r)}}\gamma_r\,\mathbf{u}_k\right)-\mathbf{u}_i\right\}. \tag{6.3}$$

We notice here that the coefficients γ_r are nonnegative and verify

$$\sum_{r=1}^n\sum_{k\in S_i^{(r)}}\gamma_r=\sum_{r=1}^n\gamma_r\,\mathrm{Card}\left(S_i^{(r)}\right)=\sum_{r=1}^n\frac{w_r\,\kappa_r}{\kappa}\,\mathrm{Card}\left(S_i^{(r)}\right)=1. \tag{6.4}$$

In the following, we will impose $w_1 \neq 0$. This is a natural hypothesis because it means that $\triangle^{(1)}(\mathbf{u})_i$, which is calculated from the closest neighbors of i, is taken into account in the Laplacian calculation at i. Therefore, we have $\gamma_1 \neq 0$. Note that a simple way of calculating a discrete Laplacian in dimension n is to set $w_1 = 1$. The *dimension independent Laplacian* given in [15] is obtained by setting $w_r = \binom{n-1}{r-1} 2^{1-n}$. These coefficients are chosen so that some properties of the smooth Laplacian are kept with the discrete Laplacian (see [15] for more details). The scheme chosen by Horn and Schunck [10] in the 2-dimensional case, detailed below, is obtained by setting $w_1 = w_2 = \frac{1}{2}$ (which is the dimension independent Laplacian in the case $n = 2$).

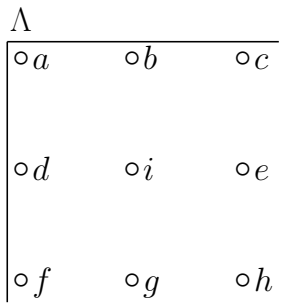


Figure 1. *The 2-dimensional finite-difference stencil for the Laplacian calculation at an interior point i of a lattice Λ.*

In Figure 1, as an example, the stencil S_i is composed of the points $\{a, b, c, d, e, f, g, h\}$. We have $S_i^{(1)} = \{b, d, e, g\}$ and $S_i^{(2)} = \{a, c, f, h\}$. The scheme of Horn and Schunck [10] is $\triangle(\mathbf{u})_i = \kappa \{\frac{1}{6}(\mathbf{u}_b + \mathbf{u}_d + \mathbf{u}_e + \mathbf{u}_g) + \frac{1}{12}(\mathbf{u}_a + \mathbf{u}_c + \mathbf{u}_f + \mathbf{u}_h) - \mathbf{u}_i\}$, with $\kappa = 3$. Here, the coefficient γ_1 associated with $S_i^{(1)}$ is $\frac{1}{6}$, and the coefficient γ_2 associated with $S_i^{(2)}$ is $\frac{1}{12}$.

Finally, from (2.4), the boundary conditions considered here are that the normal derivatives vanish at the boundary of the image. In [10], Horn and Schunck explained how to deal with these conditions: when a point outside the image is needed, the displacement of the closest point inside the image is copied.

The description of the discretization scheme given above is sufficient for programming the HS algorithm: just choose some coefficients w_r, calculate the corresponding coefficients γ_r, use (2.9) with $M(\mathbf{u})_i = \sum_{r=1}^{n} \sum_{k \in S_i^{(r)}} \gamma_r \mathbf{u}_k$ (cf. (2.6) and (6.3)), and apply the boundary conditions when necessary.

6.2. Determination of the weights in the average calculation. We will now give an expression of the coefficients λ_{ij} defined in (2.7), in order to verify hypotheses (H1), (H2), and (H3) in the next section.

Let us give a rigorous definition of the boundary conditions. We denote $\Lambda' = \{(k_1, k_2, \ldots, k_n) : -1 \leq k_\ell \leq N_\ell,\ 1 \leq \ell \leq n\}$ and define the function f from Λ' to Λ such that $f\{(k_1, k_2, \ldots, k_n)\} = (j_1, j_2, \ldots, j_n)$, where, for each ℓ in $\{1, \ldots, n\}$, the following hold:

- If $0 \leq k_\ell \leq N_\ell - 1$, then $j_\ell = k_\ell$.
- If $k_\ell = -1$, then $j_\ell = 0$.
- If $k_\ell = N_\ell$, then $j_\ell = N_\ell - 1$.

Then, our discretization scheme can be written at each point i of Λ, even if i is a boundary point:

$$\triangle(\mathbf{u})_i = \kappa \left\{ \left(\sum_{r=1}^{n} \sum_{k \in S_i^{(r)}} \gamma_r \, \mathbf{u}_{f(k)} \right) - \mathbf{u}_i \right\}. \tag{6.5}$$

Now, given two points i and j of Λ and an integer r in $\{1, \ldots, n\}$, we denote $A_{ij}^{(r)}$ the set of points defined by

$$A_{ij}^{(r)} = \{k \in S_i^{(r)} \subset \Lambda' : f(k) = j\}. \tag{6.6}$$

Then, for two points i and j of Λ, we set

$$\lambda_{ij} = \sum_{r=1}^{n} \mathrm{Card}\left(A_{ij}^{(r)}\right) \gamma_r. \tag{6.7}$$

It is clear that at each point i of Λ and for every displacement field u

$$\sum_{j \in \Lambda} \lambda_{ij} \, \mathbf{u}_j = \sum_{r=1}^{n} \sum_{k \in S_i^{(r)}} \gamma_r \, \mathbf{u}_{f(k)}. \tag{6.8}$$

Thus, from (6.5), our discretization scheme can be written as in (2.6) and (2.7): $\triangle(\mathbf{u})_i = \kappa \{M(\mathbf{u})_i - \mathbf{u}_i\}$, with $M(\mathbf{u})_i = \sum_{j=\in\Lambda} \lambda_{ij} \, \mathbf{u}_j$.

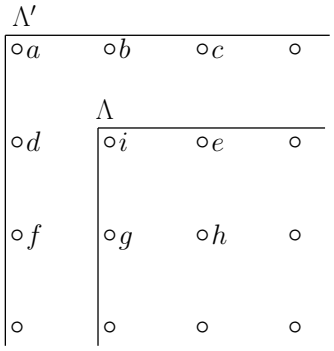


Figure 2. *The 2-dimensional finite-difference stencil for the Laplacian calculation at a boundary (corner) point i.*

In Figure 2, as a complement to the example of Figure 1, the stencil S_i of the boundary point i (located at a corner of the lattice Λ) is composed of the points $\{a, b, c, d, e, f, g, h\}$. As for Figure 1, we have $S_i^{(1)} = \{b, d, e, g\}$ and $S_i^{(2)} = \{a, c, f, h\}$. Based on the definition of (6.6), one obtains $A_{i,i}^{(1)} = \{b, d\}$ and $A_{i,i}^{(2)} = \{a\}$; $A_{i,e}^{(1)} = \{e\}$ and $A_{i,e}^{(2)} = \{c\}$; $A_{i,g}^{(1)} = \{g\}$ and $A_{i,g}^{(2)} = \{f\}$; $A_{i,h}^{(1)} = \emptyset$ and $A_{i,h}^{(2)} = \{h\}$. The corresponding scheme of Horn and Schunck [10] is $\triangle(\mathbf{u})_i = \kappa \{\frac{1}{6}(2\mathbf{u}_i + \mathbf{u}_e + \mathbf{u}_g) + \frac{1}{12}(\mathbf{u}_i + \mathbf{u}_e + \mathbf{u}_g + \mathbf{u}_h) - \mathbf{u}_i\} = \kappa \{\frac{1}{12}(5\mathbf{u}_i + 3\mathbf{u}_e + 3\mathbf{u}_g + \mathbf{u}_h) - \mathbf{u}_i\}$, with $\kappa = 3$. Again, the coefficient γ_1 associated with $S_i^{(1)}$ is $\frac{1}{6}$, and the coefficient γ_2 associated with $S_i^{(2)}$ is $\frac{1}{12}$. The case of a boundary point not located at the corner of Λ (such as the point e in Figure 2) can be treated similarly.

7. Verification of the hypotheses. We now have to verify that the general n-dimensional scheme described in section 6 fulfills the hypotheses of section 4:

(H1) For all points i and j of Λ, $\lambda_{ij} = \lambda_{ji}$.
(H2) At every point i of Λ, $\sum_{j=\in\Lambda} \lambda_{ij} = 1$.
(H3) The graph G is connected.

Proposition 7.1. *With the discretization scheme of section 6,* (H1) *is satisfied.*

Proof. Let i and j be two points of Λ, and let r be an integer in $\{1, \dots, n\}$. As in section 6.2, we denote $A_{ij}^{(r)}$ the set of points k belonging to $S_i^{(r)}$ and satisfying $f(k) = j$. By definition, a point k belongs to $A_{ij}^{(r)}$ if and only if the following hold:

- $\|k - i\|_{L^\infty} = 1$;
- $\|k - i\|_{L^1} = r$;
- $f(k) = j$.

Let us now define the function g_{ij} from $\mathbb{N}^n$ to $\mathbb{N}^n$ by $g_{ij}(k) = k + i - j$. We will show that $g_{ij}(A_{ij}^{(r)}) \subset A_{ji}^{(r)}$. We denote $i = (i_1, i_2, \dots, i_n)$ and $j = (j_1, j_2, \dots, j_n)$. Let $k = (k_1, k_2, \dots, k_n)$ be a point of $A_{ij}^{(r)}$, and let ℓ be an integer in $\{1, \dots, n\}$.

- If $k_\ell = -1$, then $j_\ell = 0$ (because $f(k) = j$) and $i_\ell = 0$ (because $\|k - i\|_{L^\infty} \leq 1$ and i belongs to Λ), so that $k_\ell + i_\ell - j_\ell = -1$.
- If $k_\ell = N_\ell$, similarly, $j_\ell = i_\ell = N_\ell - 1$ and $k_\ell + i_\ell - j_\ell = N_\ell$.
- In the other cases, $k_\ell = j_\ell$ (because $f(k) = j$), so that $k_\ell + i_\ell - j_\ell = i_\ell$.

First, from the definition of the function f, the three previous cases applied to each coordinate l of $\{1, \dots, n\}$ yield $f\big(g_{ij}(k)\big) = f(k + i - j) = i$. Moreover, in each of these three cases, it is clear that $|(k_\ell + i_\ell - j_\ell) - j_\ell| = |k_\ell - i_\ell|$ (either because $j_\ell = i_\ell$ or because $k_\ell = j_\ell$). Thus, from $\|k - i\|_{L^\infty} \leq 1$ and $\|k - i\|_{L^1} = r$, we obtain $\|g_{ij}(k) - j\|_{L^\infty} = 1$ and $\|g_{ij}(k) - j\|_{L^1} = r$. This permits us to conclude that $g_{ij}(A_{ij}^{(r)}) \subset A_{ji}^{(r)}$.

Now, as g_{ij} is a translation, it is injective, so that $\mathrm{Card}(A_{ij}^{(r)}) \leq \mathrm{Card}(A_{ji}^{(r)})$. Then, as we did not impose any hypothesis about i and j, we can exchange them and write $\mathrm{Card}(A_{ji}^{(r)}) \leq \mathrm{Card}(A_{ij}^{(r)})$, so that $\mathrm{Card}(A_{ij}^{(r)}) = \mathrm{Card}(A_{ji}^{(r)})$. Finally, (6.7) imposes that $\lambda_{ij} = \lambda_{ji}$, so that (H1) is verified. ∎

Proposition 7.2. *With the discretization scheme of section 6,* (H2) *is satisfied.*

Proof. Let i be a point of Λ. Equation (6.8) applied to a displacement field u that is uniform and different from zero yields $\sum_{r=1}^{n} \sum_{k \in S_i^{(r)}} \gamma_r = \sum_{j=\in\Lambda} \lambda_{ij}$. Then, (6.4) imposes $\sum_{j=\in\Lambda} \lambda_{ij} = 1$, so that (H2) is verified. ∎

Proposition 7.3. *With the discretization scheme of section 6,* (H3) *is satisfied.*

Proof. Let i be a point of the image lattice Λ. By hypothesis, we have that $\gamma_1 \neq 0$. So, from (6.7), we have that for two close neighbors i and j of Λ, i.e., such that $\|i - j\|_{L^1} = 1$, $\lambda_{ij} \neq 0$. Indeed, in that case, j belongs to $A_{ij}^{(1)}$, so that $\mathrm{Card}(A_{ij}^{(1)}) \neq 0$. So, two close neighbors are always linked in G, and the connectedness of G becomes obvious. ∎

So, (H1), (H2), and (H3) are fulfilled with the discretization scheme of section 6, and we are under the conditions of Theorem 4.1.

8. Conclusion. The proposed convergence result was shown using a general definition of the discrete Laplacian. That definition includes the classical scheme of Horn and Schunck

in dimension 2 and a general scheme (see section 6) for n-dimensional Laplacians. In this context, a necessary and sufficient condition for the problem to be well-posed (i.e., to have a unique solution) is that the intensity gradients not all be contained in the same hyperplane. Under that condition, the HS iterations converge to the solution. It was also shown that the convergence of the HS iterative scheme implies the convergence of the Gauss–Seidel and SOR solvers for the HS problem.

Appendix A. Here the details of the derivation of (2.9) in dimension $n \geq 1$ are presented. From (2.6) and (2.8), equation (2.5) is equivalent to

$$(\alpha\,\mathcal{I}_n + [\nabla I\,\nabla I^T]_i)\,\mathbf{u}_i - \alpha\,M(\mathbf{u})_i = -I_{t,i}\,\nabla I_i, \tag{A.1}$$

where $\mathcal{I}_n$ denotes the $n \times n$ identity matrix. Let us now notice that

$$[\nabla I\,\nabla I^T]_i^2 = \nabla I_i\,[\nabla I^T\,\nabla I]_i\,\nabla I_i^T = \|\nabla I_i\|^2\,[\nabla I \nabla I^T]_i. \tag{A.2}$$

Thus,

$$\begin{aligned}&(\alpha\,\mathcal{I}_n + [\nabla I\,\nabla I^T]_i)\,(\alpha\,\mathcal{I}_n + \|\nabla I_i\|^2\mathcal{I}_n - [\nabla I\,\nabla I^T]_i)\\ &\quad = \alpha\,(\alpha + \|\nabla I_i\|^2)\mathcal{I}_n.\end{aligned} \tag{A.3}$$

So,

$$\begin{aligned}(\alpha\,\mathcal{I}_n + [\nabla I\,\nabla I^T]_i)^{-1} &= \frac{\alpha\,\mathcal{I}_n + \|\nabla I_i\|^2\,\mathcal{I}_n - [\nabla I\,\nabla I^T]_i}{\alpha\,(\alpha + \|\nabla I_i\|^2)}\\ &= \alpha^{-1}\,\mathcal{I}_n - \alpha^{-1}\,\frac{[\nabla I\,\nabla I^T]_i}{\alpha + \|\nabla I_i\|^2}.\end{aligned} \tag{A.4}$$

We also have

$$[\nabla I\,\nabla I^T]_i\,I_{t,i}\,\nabla I_i = I_{t,i}\,\nabla I_i\,[\nabla I^T\,\nabla I]_i = \|\nabla I_i\|^2\,I_{t,i}\,\nabla I_i. \tag{A.5}$$

Now, from (A.4) and (A.5), the expression (A.1) can be rewritten as

$$\mathbf{u}_i = \left(\mathcal{I}_n - \frac{[\nabla I\,\nabla I^T]_i}{\alpha + \|\nabla I_i\|^2}\right) M(\mathbf{u})_i - \frac{I_{t,i}\,\nabla I_i}{\alpha + \|\nabla I_i\|^2}. \tag{A.6}$$

This equality leads us to write the general HS iterations for an n-dimensional image:

$$\mathbf{u}_i^{k+1} = \left(\mathcal{I}_n - \frac{[\nabla I\,\nabla I^T]_i}{\alpha + \|\nabla I_i\|^2}\right) M(\mathbf{u}^k)_i - \frac{I_{t,i}\,\nabla I_i}{\alpha + \|\nabla I_i\|^2}. \tag{A.7}$$

With the notation introduced in section 2, we thus obtain (2.9).

Appendix B. We discuss the condition of block diagonally dominant matrices in the context of the Jacobi solver for the HS problem. We refer the reader to [11, 1] for results on the convergence of the Jacobi method for strictly diagonally dominant matrices or irreducible and weakly diagonally dominant matrices, as well as [7] for the corresponding notions in the case of block matrices.

First, using (2.5), (2.6), (2.7), and (2.8), we observe that (2.5) can be rewritten in the form (see (A.1))

$$\{\alpha\,\mathbf{u}_i + [\nabla I\,\nabla I^T]_i\,\mathbf{u}_i\} - \sum_{j=1}^{N} \alpha\,\lambda_{ij}\,\mathbf{u}_j = -I_{t,i}\,\nabla I_i. \tag{B.1}$$

Let A_{ij}, for $i, j \in \Lambda$, be the $n \times n$ matrices defined by

$$A_{ij} = -\alpha\,\lambda_{ij}\mathcal{I}_n, \quad i \neq j; \tag{B.2}$$

$$A_{ii} = \big(\alpha\,\mathcal{I}_n + [\nabla I\,\nabla I^T]_i\big) - \alpha\,\lambda_{ii}\mathcal{I}_n. \tag{B.3}$$

Then, the Jacobi iteration is expressed as

$$\mathbf{u}_i^{k+1} = A_{ii}^{-1}\Big(-\sum_{j\neq i} A_{ij}\,\mathbf{u}_j^k - I_{t,i}\,\nabla I_i\Big). \tag{B.4}$$

Lemma B.1. *Assume that* $0 \leq \lambda_{ii} < 1$. *Then, the inverse matrix of the block* A_{ii} *is equal to* $\frac{1}{\alpha(1-\lambda_{ii})}\,P_i'$, *where* $P_i' = \mathcal{I}_n - \frac{[\nabla I\,\nabla I^T]_i}{\alpha(1-\lambda_{ii})+\|\nabla I_i\|^2}$.

Proof. The lemma follows directly from (A.4) upon replacing α by $\alpha' = \alpha(1-\lambda_{ii})$. ∎

So, let i be a point in the interior of Λ. From section 6, λ_{ii} is then equal to 0. Then, $P_i' = P_i$ and the Jacobi iteration for the point i of (B.4) reads as

$$\mathbf{u}_i^{k+1} = \alpha^{-1}P_i\Big(\sum_{j\neq i}\alpha\lambda_{ij}\mathbf{u}_j^k - I_{t,i}\,\nabla I_i\Big) \tag{B.5}$$

$$= P_i M(\mathbf{u}^k)_i + \mathbf{d}_i, \tag{B.6}$$

which amounts to the HS iteration (2.9). On the other hand, since $\lambda_{ii} \neq 0$ if i is a boundary point, the Jacobi iteration is never the HS iteration at boundary points.

Let $||P_i'||$ be the norm of the matrix P_i' defined by $\max_{\mathbf{u}_i\neq 0}\frac{||P_i'\mathbf{u}_i||}{||\mathbf{u}_i||}$ based on any norm of $\mathbb{R}^n$.

Lemma B.2. *Let* $n \geq 2$, *and consider a vector* $\mathbf{u}_i$ *in* $\mathbb{R}^n$ *that is orthogonal to* ∇I_i. *Then,* $P_i'(\mathbf{u}_i) = \mathbf{u}_i$. *Therefore,* $||P_i'|| \geq 1$ *no matter the norm used on* $\mathbb{R}^n$.

Proof. This result follows directly from the proof of Lemma 5.2. ∎

Lemma B.3. *If* $n \geq 2$ *and hypothesis* (H2) *is fulfilled, then* $||A_{ii}^{-1}||^{-1} \leq \sum_{j\neq i}||A_{ij}||$, *for any* i, *no matter the norm used on* $\mathbb{R}^n$.

Proof. From the definition (B.2) of A_{ij}, $j \neq i$, we have $\sum_{j\neq i}||A_{ij}|| = \alpha\sum_{j\neq i}\lambda_{ij}$. Then, from Lemmas B.1 and B.2 and hypothesis (H2), we have $||A_{ii}^{-1}||^{-1} = \alpha(1-\lambda_{ii})\,||P_i'||^{-1} \leq \alpha(1-\lambda_{ii}) = \alpha\sum_{j\neq i}\lambda_{ij}$. ∎

From Lemma B.3, one concludes that the matrix A is never weakly (or strictly) block diagonally dominant if $n \geq 2$ under hypothesis (H2).[1] On the other hand, if one uses the Euclidean norm on $\mathbb{R}^n$, one can easily show that $||A_{ii}^{-1}||^{-1} = \sum_{j\neq i}||A_{ij}||$, for any i, because

[1] Recall that a matrix A is weakly (or strictly) block diagonally dominant if $||A_{ii}^{-1}||^{-1} \geq \sum_{j\neq i}||A_{ij}||$ for any i and if that inequality is strict for some (or any, respectively) i.

$||P_i'|| = 1$ for that norm. So, in that case, A is block diagonally dominant (i.e., $||A_{ii}^{-1}||^{-1} \geq \sum_{j \neq i} ||A_{ij}||$ for any i), but the inequality is never strict.

Next, we show that the matrix A defined by (B.2) and (B.3) is not diagonally dominant if $n \geq 2$ (here, the matrix is not viewed as a block matrix), except in very special cases. We first treat the case $n = 2$. The absolute values of the diagonal elements of the matrix A_{ii} are equal to $\alpha(1-\lambda_{ii})+I_{x,i}^2$ and $\alpha(1-\lambda_{ii})+I_{y,i}^2$, whereas the sum of the absolute values of the elements off the diagonal for the corresponding rows of the matrix A are equal to $\sum_{j \neq i} \alpha\lambda_{ij} + |I_{x,i}||I_{y,i}|$ and $\sum_{j \neq i} \alpha\lambda_{ij} + |I_{y,i}||I_{x,i}|$. Using the identity $\sum_j \lambda_{ij} = 1$, diagonal dominance is then equivalent to $I_{x,i}^2 \geq |I_{x,i}||I_{y,i}|$ and $I_{y,i}^2 \geq |I_{y,i}||I_{x,i}|$, which implies that $|I_{x,i}| = |I_{y,i}|$ for each i such that $I_{x,i}$ and $I_{y,i}$ are both different from 0. This is a very special case, so that the assertion that the matrix A is diagonally dominant (in general) is false. If $n > 2$, the absolute values of the diagonal elements of the matrix A_{ii} are equal to $\alpha(1-\lambda_{ii})+I_{x_\ell,i}^2$ for $1 \leq \ell \leq n$, whereas the sum of the absolute values of the elements off the diagonal for the corresponding rows of the matrix A are equal to $\sum_{j \neq i} \alpha\lambda_{ij} + |I_{x_\ell,i}| \sum_{\ell' \neq \ell} |I_{x_{\ell'},i}|$ for $1 \leq \ell \leq n$. Therefore, the diagonal dominance of A implies that $\sum_{\ell=1}^n |I_{x_\ell,i}| \geq (n-1) \sum_{\ell=1}^n |I_{x_\ell,i}|$, which implies that $\nabla I_i = \vec{0}$ for each point i. So, again, the assertion that the matrix A is diagonally dominant (in general) is false. Therefore, it appears that the short argument given in [23, p. 249] for the convergence of the pointwise Jacobi method is erroneous.

Remarks.

1. The HS iterative scheme amounts to the Jacobi iterative scheme at the interior points of the image, but never at its boundary points. But then we believe that it is usually the HS scheme that is implemented rather than the Jacobi method. Indeed, it is easy to implement (cf. the end of section 6.1), still fully parallelizable, and it is the original method proposed by Horn and Schunck. The difference between the two schemes is due to the Neumann boundary conditions (because then $\lambda_{ii} \neq 0$ at a boundary point).
2. The Neumann boundary conditions (2.4) that come from the unconstrained minimization problem are very important. In particular, they imply that the Laplacian of a uniform displacement field vanishes, i.e., $\Delta(\mathbf{u})_i = \kappa(M(\mathbf{u})_i - \mathbf{u}_i) = \kappa(\sum_{j \in \Lambda} \lambda_{ij} - 1)\mathbf{u}_i$, so that we must have $\sum_{j \in \Lambda} \lambda_{ij} = 1$.
3. Due to this condition, known convergence results of the (block) Jacobi and Gauss–Seidel methods do not apply, unless $n = 1$. The result [1, Theorem 1, (a)] assumes that the matrix A is strictly diagonally dominant, which is not the case here. Also, the result [1, Theorem 1, (b)] assumes that A is irreducible and weakly diagonally dominant, which is not the case either. Note that one can generalize [1, Theorem 1] using the notion of block diagonally dominant matrices [7]; namely, one can prove along the lines of [1] that if A is strictly block diagonally dominant or if it is block irreducible and weakly diagonally dominant, then both the block Jacobi and the Gauss–Seidel solvers converge. But again, these hypotheses never hold for the HS problem, unless $n = 1$.
4. On the other hand, if one wants to relax the boundary condition (2.4) and allow $\sum_{j \in \Lambda} \lambda_{ij} < 1$ at a boundary point, then one can show that A is weakly block diagonally dominant for the Euclidean norm and block irreducible (based on the connectedness of the graph G, i.e., hypothesis (H3)), so that both the block Jacobi and the Gauss–Seidel

solvers then converge. This may happen if one considers a minimization problem with constraints, for instance if the displacement is known at some points of the image.

Appendix C. In this appendix, we discuss the implications of Theorem 4.1 (i.e., the convergence of the HS method) on the convergence of the Gauss–Seidel and SOR iterative schemes through the property of positive definiteness of the coefficient matrix of the HS problem. We also present a more general result that states conditions under which the convergence of the Gauss–Seidel and SOR methods is implied by the convergence of the Jacobi method. In what follows, $\rho(A)$ denotes the spectral radius of a square matrix A.

Proposition C.1. *Let $\tilde{B}$ and $\tilde{C}$ be real symmetric matrices of the same dimensions such that $\tilde{B}$ is positive definite and $\rho(\tilde{B}^{-1}\tilde{C}) < 1$. Then, the matrix $\tilde{B}+\tilde{C}$ is symmetric positive definite.*

Proof. Since the matrix $\tilde{B}$ is symmetric positive definite, it can be expressed in the form LL, where L is a symmetric invertible matrix. Indeed, one can write $\tilde{B} = R\Psi R^T$, where $RR^T = \mathcal{I}$ ($\mathcal{I}$ is the identity matrix) and Ψ is a diagonal positive definite matrix; thus, $\tilde{B} = LL$, where $L = R\Psi^{1/2}R^T$. Then, $A = \tilde{B}+\tilde{C} = L(\mathcal{I}+L^{-1}\tilde{C}L^{-1})L$. Since L is symmetric and invertible, then A is positive definite if and only if the symmetric matrix $A' = \mathcal{I}+L^{-1}\tilde{C}L^{-1}$ is positive definite. Now, one has that $\rho(L^{-1}\tilde{C}L^{-1}) = \rho(L^{-1}L^{-1}\tilde{C}L^{-1}L) = \rho(\tilde{B}^{-1}\tilde{C}) < 1$. Therefore, the real symmetric matrix $L^{-1}\tilde{C}L^{-1}$ can be written as $Q^T\Lambda Q$, where $Q^TQ = \mathcal{I}$ and Λ is a diagonal matrix such that $\rho(\Lambda) = \rho(L^{-1}\tilde{C}L^{-1}) < 1$. It follows that $A' = Q^T(\mathcal{I}+\Lambda)Q$, where $\mathcal{I}+\Lambda$ is a diagonal positive definite matrix (because any eigenvalue λ of Λ is such that $|\lambda| < 1$). Thus, A' is a symmetric positive definite matrix, and so is A. ∎

Corollary C.2. *Let $Ax = \mathbf{b}$ be a linear system, where A is a real symmetric matrix. Let A be written in the form $D-B-C$, where D, B, and C are block diagonal, block upper triangular, and block lower triangular matrices, respectively. Assume that D is positive definite. Then, the convergence of the Jacobi iterative scheme $\mathbf{x}^{k+1} = D^{-1}((B+C)\mathbf{x}^{k+1}+\mathbf{b})$ implies the convergence of the Gauss–Seidel and SOR iterative schemes. In fact, the matrix A is positive definite under the assumptions.*

Proof. Let $\tilde{B} = D$ and $\tilde{C} = -B-C$. The convergence of the Jacobi iterative scheme is equivalent to $\rho(D^{-1}(B+C)) < 1$. Thus, from Proposition C.1, the matrix A is positive definite. Henceforth, the Gauss–Seidel and SOR methods converge; see, for instance, [4, Theorem 5.3-2]. ∎

Corollary C.3. *Under hypotheses* (H1), (H2), *and* (H3), *assume that the rank of (∇I_i) is n. Then, the coefficient matrix A of* (B.1), *with blocks defined by* (B.2) *and* (B.3), *is symmetric positive definite. In particular, the Gauss–Seidel and SOR iterative schemes converge under these conditions.*

Proof. Let $\tilde{B} = \alpha P^{-1}$ and $\tilde{C} = -\alpha M$, where P and M are as in (2.9). Then, $\tilde{B}$ is the block diagonal matrix with diagonal matrix entries $A'_{ii} = \alpha\mathcal{I}_n + [\nabla I\,\nabla I^T]_i$, as follows from Appendix A. Moreover, the eigenvalues of A'_{ii} are α with multiplicity $n-1$ and $\alpha+||\nabla I_i||^2$ with multiplicity 1. Thus, the symmetric matrix $\tilde{B}$ is positive definite. Also, Theorem 4.1 implies that $\rho(\tilde{B}^{-1}\tilde{C}) = \rho(PM) < 1$. Finally, $A = \alpha P^{-1}-\alpha M = \tilde{B}+\tilde{C}$, using Appendix A. The statement on the positive definiteness of the matrix A now follows from Proposition C.1 since M is symmetric. Hence, the Gauss–Seidel and SOR iterative schemes converge under these conditions, as in the proof of Corollary C.2. ∎

Remark. The positive definiteness of the coefficient matrix of the HS problem has been proved directly in [17]. Moreover, as mentioned in section 1, the V-ellipticity of the HS functional [18] implies the positive definiteness of the coefficient matrix of the HS problem. Thus, Corollary C.3 is not a new result. However, the more general result, Corollary C.2, might be of interest to further understand the convergence of the Jacobi, Gauss–Seidel, and SOR methods.

Appendix D. In this appendix, we give more details to explain why we think the proofs presented in [17, 13] are erroneous. We show that the matrix "P" of [17, eq. (9)] (denoted here by P_* to avoid confusion with the linear transformation P of (2.9)) is not contracting for the norm defined by [17, eq. (10)], for any nonuniform image. Indeed, let i_0 be a point where $\nabla I_{i_0}^T = (I_{x,i_0}, I_{y,i_0}) \neq (0,0)$. We consider the displacement field $\mathbf{u}$ defined by $u_{2\,i-1} = I_{y,i_0}$ and $u_{2\,i} = -I_{x,i_0}$ if $i \in N_{i_0}$ (the set of four neighbors of i_0), and $u_{2\,i-1} = u_{2\,i} = 0$ otherwise. The norm defined in [17, eq. (10)], denoted by $\|.\|_*$ here to avoid any confusion, reads as $\|\mathbf{u}\|_* = \max_{1\leq i\leq N}(u_{2\,i-1}^2 + u_{2\,i}^2)^{\frac{1}{2}}$. In that case, we obtain $\|\mathbf{u}\|_* = (I_{x,i_0}^2 + I_{y,i_0}^2)^{\frac{1}{2}}$. Moreover, we find that $P_*(\mathbf{u})_{2\,i_0-1} = I_{y,i_0}$ and $P_*(\mathbf{u})_{2\,i_0} = -I_{x,i_0}$. Therefore, $\|P_*(\mathbf{u})\|_* \geq (I_{x,i_0}^2 + I_{y,i_0}^2)^{\frac{1}{2}}$, so that $\|P_*(\mathbf{u})\|_* \geq \|\mathbf{u}\|_*$. Thus, P_* is not contracting, due to this counterexample. We think that the error occurred in [17, formula (13)]: a factor c_i should be added in the second member to take into account that the sum in the first term includes all the neighbors of i. Thus, in the inequality [17, formula (15)], one should use the factor $\sqrt{2}$ instead of 1, which makes that proof break down.

In [13, eq. (20)], the Laplacian corresponding to the Neumann boundary conditions (which usually correspond to the HS problem) is denoted by L_2. The matrix N_2 is defined by the relation $N_2(\mathbf{u}) = L_2(\mathbf{u}) + \mathbf{u}$ (cf. [13, eq. (22)]).[2] Since that Laplacian operator vanishes on uniform displacement fields, any such displacement field is an eigenvector of the matrix N_2 for the eigenvalue 1. Therefore, the assertion after [13, eq. (23)] that the spectral radius $\rho(N_2)$ of the matrix N_2 (i.e., the maximal modulus of the eigenvalues of N_2) is less than 1 is erroneous. Incidentally, in [13, formula (22)], a factor $\frac{1}{2}$ is missing to get a correct expression of the average. In [13, formulas (38) and (40)], the authors also assert that $\rho\big(I_d - F^{-1}\mathrm{Diag}(S_{ij})\big) < 1$. But, at every point, the determinant of the 2×2 matrix S_{ij} is null.[3] Then, the matrices S_{ij} are singular, and so is $F^{-1}\mathrm{Diag}(S_{ij})$. Thus, 1 is an eigenvalue of $I_d - F^{-1}\mathrm{Diag}(S_{ij})$, and so the assertion is flawed. Thus, the two main intermediate results of [13] are both erroneous.

Acknowledgment. The authors are grateful to the anonymous reviewers for their comments that helped improve the presentation and the content of this work.

REFERENCES

[1] R. Bagnara, *A unified proof for the convergence of Jacobi and Gauss–Seidel methods*, SIAM Rev., 37 (1995), pp. 93–97.

[2] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski, *A database and evaluation methodology for optical flow*, Int. J. Comput. Vis., 92 (2011), pp. 1–31.

[2] In our notation, $L_2 = \triangle^{(1)}$ of (6.1) and $N_2 = M$ of (2.7).

[3] In our notation, "(i,j)" corresponds to a point i and "S_{ij}" corresponds to the $n\times n$ matrix $[\nabla I \nabla I^T]_i$. "I_d" corresponds to the $n\times n$ identity matrix $\mathcal{I}_n$.

[3] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.
[4] P. G. Ciarlet, *Introduction à l'analyse matricielle et à l'optimisation*, Masson, Paris, 1982.
[5] Q. Duan, E. D. Angelini, A. Lorsakul, S. Homma, J. W. Holmes, and A. F. Laine, *Coronary occlusion detection with 4d optical flow based strain estimation on 4d ultrasound*, in Functional Imaging and Modeling of the Heart, Lecture Notes in Comput. Sci. 5528, N. Ayache, H. Delingette, and M. Sermesant, eds., Springer, Berlin, Heidelberg, 2009, pp. 211–219.
[6] D. J. Evans, *Parallel S.O.R. iterative methods*, Parallel Comput., 1 (1984), pp. 3–18.
[7] D. G. Feingold and R. S. Varga, *Block diagonally dominant matrices and generalizations of the Gerschgorin circle theorem*, Pacific J. Math., 12 (1962), pp. 1241–1250.
[8] I. Franjic, S. Khalid, and J. Pecaric, *On the refinements of the Jensen-Steffensen inequality*, J. Inequal. Appl., no. 12 (2011).
[9] T. Guerrero, G. Zhang, T.-C. Huang, and K.-P. Lin, *Intrathoracic tumour motion estimation from CT imaging using the 3D optical flow method*, Phys. Med. Biol., 49 (2004), pp. 41–47.
[10] B. K. P. Horn and B. G. Schunck, *Determining optical flow*, Artificial Intelligence, 17 (1981), pp. 185–203.
[11] K. R. James, *Convergence of matrix iterations subject to diagonal dominance*, SIAM J. Numer. Anal., 10 (1973), pp. 478–484.
[12] J. Jensen, *Sur les fonctions convexes et les inégalités entre les valeurs moyennes*, Acta Math., 30 (1906), pp. 175–193.
[13] Y. Kameda, A. Imiya, and N. Ohnishi, *A convergence proof for the Horn-Schunck optical-flow computation scheme using neighborhood decomposition*, in Combinatorial Image Analysis, Lecture Notes in Comput. Sci. 4958, Springer, Berlin, 2008, pp. 262–273.
[14] C. Kirisits, L. F. Lang, and O. Scherzer, *Optical flow on evolving surfaces with an application to the analysis of 4D microscopy data*, in Scale Space and Variational Methods in Computer Vision, Springer, Berlin, Heidelberg, 2013, pp. 246–257.
[15] A. Kumar, *A discretization of the n-dimensional Laplacian for a dimension-independent stability limit*, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci., 457 (2001), pp. 2667–2674.
[16] T. Liu and L. Shen, *Fluid flow and optical flow*, J. Fluid Mech., 614 (2008), pp. 253–291.
[17] A. Mitiche and A. Mansouri, *On convergence of the Horn and Schunck optical flow method*, IEEE Trans. Image Process., 13 (2004), pp. 848–852.
[18] C. Schnörr, *Determining optical flow for irregular domains by minimizing quadratic functionals of a certain class*, Int. J. Comput. Vis., 6 (1991), pp. 25–38.
[19] L. Schwartz, *Cours d'analyse*, Hermann, Paris, 1967.
[20] J. Thiyagalingam, D. Goodman, J. A. Schnabel, A. Trefethen, and V. Grau, *On the usage of GPUs for efficient motion estimation in medical image sequences*, Int. J. Biomed. Imag., 2011 (2011), 137604.
[21] J. Treibig, G. Wellein, and G. Hager, *Efficient multicore-aware parallelization strategies for iterative stencil computations*, J. Comput. Sci., 2 (2011), pp. 130–137.
[22] D.-M. Tsai and H.-Y. Tsai, *Low-contrast surface inspection of mura defects in liquid crystal displays using optical flow-based motion analysis*, Mach. Vis. Appl., 22 (2011), pp. 629–649.
[23] J. Weickert and C. Schnörr, *Variational optic flow computation with a spatio-temporal smoothness constraint*, J. Math. Imaging Vision, 14 (2001), pp. 245–255.
[24] R. P. Wildes, M. J. Amabile, A.-M. Lanzillotto, and T.-S. Leu, *Recovering estimates of fluid flow from image sequence data*, Comput. Vis. Image Understand., 80 (2000), pp. 246–266.
[25] D. M. Young, *Iterative Solution of Large Linear Systems*, Dover, Mineola, NY, 2003.