N°d'ordre : 4480

THÈSE

présentée à

L'UNIVERSITÉ BORDEAUX I

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET D'INFORMATIQUE

par Robin HUART

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : MATHÉMATIQUES APPLIQUÉES

Simulation numérique d'écoulements magnétohydrodynamiques par des schémas distribuant le résidu

---

Soutenue le : 2 Février 2012

Après avis des rapporteurs :

M.	James ROSSMANITH	Rapporteur
M.	Boniface NKONGA	Rapporteur

Devant la commission d'examen composée de :

M.	Tahar AMARI	Directeur de recherche, École Polytechnique	Président du Jury
M.	Boniface NKONGA	Professeur, Université Nice Sophia Antipolis	Rapporteur
M.	James ROSSMANITH	Professeur, University of Wisconsin	Rapporteur
M.	Rémi ABGRALL	Professeur, Institut Polytechnique de Bordeaux	Directeur de thèse
M.	Guido HUIJSMANS	Senior Scientific Officer, ITER Organization	Examinateur
M.	Luc MIEUSSENS	Professeur, Institut Polytechnique de Bordeaux	Examinateur
M.	Mario RICCHIUTO	Chargé de recherche, Inria Bordeaux Sud-Ouest	Examinateur

À ma famille

Remerciements

Je tiens à remercier en premier lieu mon directeur de thèse Rémi Abgrall pour m'avoir proposé d'effectuer ce travail enthousiasmant à la fin de mon stage de fin d'études. J'ai pu grâce à lui apprendre beaucoup de choses qui me seront utiles à l'avenir. Malgré les difficultés d'emploi du temps pour travailler ensemble, ses conseils et ses remarques ont toujours été précieux.

Je remercie également les membres du jury qui ont accepté de participer à la soutenance. Parmi eux, Boniface Nkonga et James Rossmanith qui ont de plus accepté d'être rapporteurs d'un manuscrit au style d'écriture peut-être un peu lourd. Merci à eux, et spécialement à Boniface Nkonga que j'ai pu croiser régulièrement depuis 5 ans, déjà en soutenance de projet de master, et dont les péripéties le jour de la soutenance ont bien fait rire l'assemblée qui avait bien fait de venir ce jour-là. Merci aussi à Guido Huysmans avec qui j'ai eu le plaisir de collaborer quelques fois pendant ces 4 ans, pour sa sympathie et pour son regard de physicien sur mes travaux. Merci à Luc Mieussens pour sa participation au jury et à Tahar Amari pour en avoir été le président, et aujourd'hui m'accueillir à Palaiseau dans le cadre d'un postdoc. Un grand merci enfin pour Mario Ricchiuto. Outre les bons moments passés en-dehors du labo, il a toujours su se montrer disponible pour m'aider et m'apporter son expertise sur de nombreux points délicats.

Je tiens enfin à remercier toutes les personnes que j'ai eu le plaisir de côtoyer pendant ces quelques années. Guillaume (blox), mon ex-voisin, à qui je souhaite bonne chance pour la fin de sa thèse et pour la suite. Damien (Dams), ninja normand à la solde de l'Église du Flying Spaghetti Monster, à qui je souhaite également bonne chance pour la fin de sa thèse. Pascal ((censored)), maître breton du lever de coude ne pouvant être vaincu qu'à FIFA (par à peu près n'importe qui sachant tenir une manette), issu du grand cru Matmeca 2007. Bonne chance pour ta thèse. Abdou (dont le surnom ne doit pas être lu à voix haute par un mortel), maître granger au repos ou advanced marauder à l'heure du berserk, grand sage métalleux et protecteur des petits chats. Captain Orel (pink ou dark selon l'humeur), expert en fluctuations chaotiques de PSR, principal prophète connu en lien spirituel direct avec Tonton Jojo, poète et ambassadeur de l'humour fin et élégant. Manu, le Che Guevara du Midi, théoricien de l'amour universel et révolutionnaire libérateur des mœurs, un chercheur qui n'hésite pas à aller sur le terrain expérimenter (ou faire expérimenter) d'improbables techniques de séduction. Nico (netWorms), ex-cobural, spécialiste alsacien ès kung-fu shaolin et films d'horreur de série Z temporairement exilé en Suisse. Adam (Laden), ex-cobural, fanfaron soubassophoniste de talent et esthète des espaces en bijection, qui nous rappelait régulièrement avec philosophie que "la thèse c'est l'amour, mec" (merci aussi à Hubert pour cette inspiration !). Hocine, ex-cobural, le chasseur de gazelles (et quelles gazelles !) qui ne s'assumait pas pleinement, pour nos discussions tantôt scientifiques tantôt footballistiques tantôt politiques tantôt... Christelle (Cricri), patiente et diplomate, qui s'est révélée au détour d'une conférence au Portugal être une princesse ch'nordiste irrésistible, chose que Hugues savait déjà mais qu'elle aurait préféré ignorer sur le moment.

Merci à tous ceux de l'équipe super sérieuse dans laquelle j'ai été accueillie pendant mon stage et que je n'ai pas encore cités : Mathieu (mateo), Cheche, Jérémie, Benji (entraines-toi ami pongiste), Cédric, Hubert, Mathieu (patator), Guillem (le borloo invisible). Merci aux footeux qui ont survécu aux assauts d'Abdou : Julien (ex-SED), Cédric (SED), Hervé (encore un SED !), Brice, François, Yoan, ... Merci enfin à tous les autres : Pierre, François (the Scotch man, pas seulement pour ses travaux de recherche), François (ex-cobural et nouveau papa), Pascal H. (qui est maintenant à Pau), Juliette, Arnaud, Josy, Pietro, Gianluca, Dante, Maria-Giovanna, Stéphane, Virginie, Sébastien, Xavier, ...

Table des matières

Chapitre 1

Contexte général

Sommaire

1.1 Motivations

Le contexte global

A l'heure où les questions de l'exploitation de combustibles fossiles qui se raréfient et de leur impact sur l'environnement sont devenues, à juste titre, primordiales dans les opinions publiques de nombreux pays, la recherche de nouvelles sources d'énergie constitue l'un des enjeux majeurs des prochaines décennies. La disparition annoncée des réserves de pétrole et l'inquiétude suscitée par les changements climatiques ouvrent la voie à de nouvelles alternatives telles que les énergies solaire, éolienne, géothermique, marémotrice ou encore la biomasse et l'utilisation de biocarburants, entre autres. Cependant, ces solutions ne sont pas jugées suffisantes actuellement pour pouvoir remplacer un jour l'approvisionnement en charbon, pétrole, gaz et combustibles nucléaires dont nous tirons le principal de nos énergies. C'est pourquoi le recours à ces énergies devrait perdurer au détriment de l'environnement, mais c'est aussi la raison qui pousse les gouvernements à mettre en place des programmes de recherche ayant vocation à proposer des alternatives crédibles, à la fois sur le plan de la capacité de production de quantités suffisantes et sur celui de la minimisation de l'impact environnemental. Dans ce contexte, l'idée de la fusion nucléaire comme source d'énergie a pu émerger très tôt. Elle a néanmoins toujours constitué un défi véritable, qui tente de "mettre en boîte l'énergie du Soleil".

La fusion nucléaire

L'existence de la réaction de fusion nucléaire est connue depuis les années 1920 environ, et elle ne fut recréée sur Terre que lors de l'explosion des premières bombes à hydrogène, quelques années après la fin de la Seconde Guerre Mondiale. Ce phénomène n'est observé naturellement qu'au sein des étoiles. La combustion qui semble s'opérer en leur sein est le résultat de la fusion d'éléments qui la composent, et elle engendre des phénomènes d'une importance capitale : une libération d'énergie sous forme de rayonnements (tels que nous en captons du Soleil) et la création de tous les éléments chimiques dont nous disposons (ce qu'on appelle la nucléosynthèse). De fait, les quantités d'énergie nécessaires à l'établissement de ce

processus sont colossales, seulement rendues possibles dans les étoiles par l'énorme pression gravitionnelle qui y règne. Et encore, cette pression n'est capable "d'allumer" les réactions de fusion que jusqu'à un certain stade, la synthèse des atomes de fer, car plus les éléments réactifs sont lourds plus la réaction est difficile à obtenir. La synthèse de tous les éléments plus lourds que le fer ne peut se faire que par des apports d'énergie encore plus importants et elle ne peut donc avoir lieu que lors des explosions d'étoiles, les supernovae. C'est pour cette raison que pour vouloir domestiquer un tel phénomène, il faut commencer par envisager les réactions les plus simples qui puissent exister, autrement dit celles faisant intervenir les éléments les plus légers comme les isotopes de l'hydrogène.

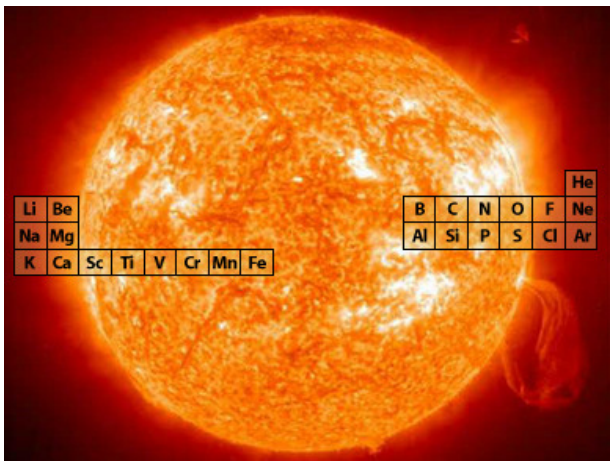


FIGURE 1.1 – La nucléosynthèse stellaire

Les origines d'ITER

Les premières recherches sur la fusion thermonucléaire contrôlée furent entreprises séparément après les premiers essais de bombes H par les principales puissances, attirées par les promesses d'une énergie propre et dont les combustibles seraient disponibles en abondance. Bien que la finalité de tels programmes soit en partie civile, l'étude des premiers générateurs de hautes puissances capables de produire les énergies requises par la fusion était alors strictement réservée aux militaires. Les pays engagés dans ces recherches décidèrent de mettre leurs résultats en commun en 1958 et ils se rendirent alors compte combien le pari qu'ils avaient fait était osé. Plusieurs configurations possibles avaient émergé, mais ce ne fut que plus tard qu'une se détacha du lot : la configuration tokamak développée en URSS grâce aux travaux d'Andréï Sakharov. Même si d'autres types de machines sont toujours étudiées, le consensus s'est rapidement établi sur le fait que le tokamak représentait la meilleure solution (bien que, récemment, les exploits réalisés par la Z-machine aux laboratoires Sandia aient pu raviver l'intérêt des configurations Z-pinch pour les expériences de fusion). Les progrès ralentirent mais sans jamais s'arrêter complètement, et les machines continuèrent de s'améliorer progressivement. Jusqu'à ce qu'il soit décidé en 1985 d'entreprendre un projet international de recherches sur la fusion contrôlée à partir des technologies des tokamaks, à l'initiative de Mikhaïl Gorbatchev. Ainsi naquit le projet ITER (pour International Thermonuclear Experimental Reactor) qui rassemblait alors seulement l'URSS, les États-Unis, l'Europe et le Japon.

La technologie dont nous allons parler est ce qu'on appelle la Fusion par Confinement Magnétique (FCM), qui regroupe entre autres les tokamaks et les stellarators (et quelques autres configurations). Elle est à distinguer de la Fusion par Confinement Intertiel (FCI), qui comprend les machines à base de lasers mais également les configurations Z-pinch et θ-pinch, et de la Fusion par Confinement Électrostatique

(FCE, une technologie moins avancée). Pour compléter ce panorama, on peut mentionner une autre voie extrêmement controversée : la fusion froide.

Les atouts de la fusion

Comme nous l'avons dit, la réaction envisagée pour la production d'électricité à partir de réacteurs à fusion est celle qui ferait intervenir les éléments les plus légers, à condition qu'on puisse en disposer en quantités suffisantes. Il se trouve que ce sont aussi les réactions qui font intervenir les éléments les plus légers qui dégagent le plus d'énergie. Les candidats retenus sont donc les réactions possibles entre les isotopes de l'hydrogène (les noyaux les plus simples qui soient) que sont le Deutérium (^2_1H) et le Tritium (^3_1H). Au premier rang, la réaction qui doit avoir lieu pour ITER est la suivante :

$$^2_1H +^3_1 H \rightarrow^4_2 He(3.52MeV) +^1_0 n(14.06MeV)$$

Il y a de plus émission d'une particule alpha dont l'énergie correspond à la masse perdue à l'issue de la réaction par la célèbre relation d'Einstein $E = mc^2$. Par chance, on dispose d'une quantité de deutérium quasiment inépuisable et son extraction n'est pas très onéreuse. En revanche, le tritium est très rare et difficile à produire. Il faudrait donc qu'ITER soir capable de générer son propre tritium en quantités suffisantes. Ce serait le rôle d'une des deux réactions suivantes, l'autre pouvant être une source alternative de tritium mais plus difficile à obtenir (moins probable) :

$$^2_1H +^2_1 H \quad \rightarrow \quad ^3_2He(0.82MeV) +^1_0 n(2.45MeV)$$

$$^2_1H +^2_1 H \quad \rightarrow \quad ^3_1H(1.01MeV) +^1_1 H(3.03MeV)$$

Un autre avantage par rapport aux centrales nucléaires à fission est la sécurité, à cause de l'élimination du risque d'emballement des réactions et de fonte du réacteur. En réalité, il faut adopter un tout autre point de vue sur le déroulement des opérations : la difficulté est de maintenir (confiner) un plasma très chaud dans les conditions du régime de fusion pendant suffisamment longtemps. C'est un équilibre fragile, difficile à obtenir, et toute perturbation entraînerait l'extinction pure et simple du plasma ! Autrement dit, la crainte de l'opérateur sera davantage la panne que le risque d'explosion.

Enfin, l'atout certainement majeur est la propreté. Les réactions que nous avons écrites plus haut montrent bien que les rejets issus de la fusion même ne sont pas toxiques ni radioactifs (on récupère de l'hélium et des neutrons), à l'inverse des déchets dûs à la fission de l'uranium ou du plutonium. Le seul élément radioctif est le tritium, mais il cesse d'émettre au bout d'une dizaine d'année et le but sera de le consommer le plus possible à l'intérieur du plasma. Son surplus éventuel pourra être stocké le temps que son activité cesse.

Les défis techniques de la FCM

Les réacteurs d'étude actuels de type tokamak, la configuration également retenue pour ITER, se heurtent néanmoins à plusieurs difficultés majeures qu'il faudra résoudre avant de pouvoir envisager une exploitation industrielle. La première, que nous avons déjà évoquée, est l'approvisionnement du plasma en tritium, un des combustibles de la réaction. Outre la réaction deutérium-deutérium dont nous avons parlée, il semble nécessaire de considérer un apport plus conséquent en utilisant du lithium (son bombardement par des neutrons énergétiques pouvant donner naissance à de l'hélium et du tritium), qui est loin d'être une ressource inépuisable.

Ensuite se pose le problème des matériaux dont seront faits les réacteurs, ceux qu'on appelle les "composants face au plasma" (CFP), qui devront pouvoir supporter des gradients de température très

importants le cas échéant, sans produire trop d'impuretés qui nuiraient au comportement du plasma et réduiraient fortement la durée de vie du réacteur. Il ne faudrait pas avoir à remplacer ces composants trop souvent, ni que cela demande trop de temps, si l'on souhaite maintenir une production continue d'électricité pour les usagers. Ce problème est aujourd'hui considéré comme le plus limitant pour les tokamaks, beaucoup de questions n'ayant pas encore trouvé de réponse.

Enfin, et c'est le sujet qui nous intéressera par la suite, la dernière problématique est celle du confinement du plasma lui-même. Comme c'est du comportement du plasma que dépend tout le reste, il faut être capable de le contrôler et en particulier de le restreindre à une partie seulement de la chambre qui l'abrite, de façon à préserver au maximum les composants face au plasma et d'empêcher la production parasite d'impuretés à leur contact. Or la physique des plasmas recèle une grande richesse de phénomènes, dont beaucoup d'instabilités qui peuvent anéantir les efforts de confinements. Les plus dangereuses pour un tokamak sont en particulier celles qui se développent au bord du plasma et peuvent donc endommager les matériaux du réacteur : les ELMs (pour *Edge Localized Modes*). Cependant, elles peuvent aussi jouer un rôle potentiellement bénéfique : nettoyer le plasma de ses impuretés. Il serait donc plus judicieux de les contrôler que de les détruire.

Contexte de nos travaux

La compréhension des mécanismes de ces instabilités reste un défi à l'heure actuelle, ainsi que l'estimation des pertes d'énergie qu'elles causent. De la même manière qu'en Mécanique des Fluides pour comprendre les phénomènes de turbulence, on a alors recours à la simulation numérique. Comme nous le verrons plus loin, l'écoulement du plasma chaud et dense régnant dans un tokamak peut se décrire par les équations de la MHD, qui voient ce plasma comme un fluide mono-espèce.

C'est dans ce cadre que le projet ANR-CIS ASTER a vu le jour, autour de la problématique de la modélisation MHD des écoulements d'un plasma dans un tokamak. Au croisement de disciplines variées telles que la physique des plasmas, les mathématiques appliquées et l'informatique, le projet a rassemblé l'équipe Bacchus de l'INRIA Bordeaux Sud-Ouest et l'institut de recherches sur la fusion magnétique (IRFM) du CEA Cadarache. Il comprenait trois axes d'études :

1. l'amélioration du code de calcul JOREK développé au CEA Cadarache,
2. l'adaptation aux problèmes MHD du code de calcul FluidBox développé dans l'équipe Bacchus,
3. les stratégies de raffinement de maillage dans le cadre des discrétisations utilisées dans JOREK et leur validation avec ce même code.

Le présent travail porte sur la réalisation du second point. Notre point de vue sera différent de celui adopté dans JOREK. Le modèle employé dans ce dernier est un modèle réduit représentatif de la physique principale décrivant les écoulements à l'origine des instabilités : les équations de la MHD "réduite" et une discrétisation spatiale 2D des surfaces poloïdales à laquelle s'ajoute une représentation en harmoniques de Fourier dans la direction toroïdale. Le passage au jeu d'équations complet de la MHD a été effectué par la suite, dans une formulation non-conservative utilisant le potentiel vecteur comme variable au lieu du champ magnétique. Nous avons emprunté une direction différente et plus généraliste en partant d'un logiciel de simulation d'écoulements fluides, FluidBox, et en adaptant ses algorithmes aux équations de la MHD, tout en prenant en compte la contrainte $\overrightarrow{\nabla} \cdot \overrightarrow{B} = 0$ sur des maillages d'éléments finis. Nous verrons tout ceci par la suite mais disons simplement ici que notre approche se base sur des maillages non structurés et des méthodes numériques non-linéaires capables d'appréhender les chocs. Par conséquent, le domaine d'application de FluidBox est plus vaste que celui des plasmas de tokamak avec leur géométrie, il traite de MHD d'une manière générale (ce qui ne signifie pas que **tous** les modèles de MHD soient

pris en charge, voir les différents modèles que nous distinguons dans la section suivante). On peut citer comme exemples d'applications alternatives les plasmas astrophysiques. En revanche, nous sommes moins avancés que JOREK sur la simulation d'instabilités comme les ELMs.

Avant d'entamer l'exposé de nos travaux, introduisons dès à présent le contexte physique dans lequel nous nous plaçons de façon à nous familiariser avec le vocabulaire qui sera employé et avec la réalité physique de ce que nous simulerons.

1.2 Plasmas et Magnétohydrodynamique

Un plasma est un milieu matériel dont les atomes, ou une partie d'entre eux, ont perdu des électrons si bien que ce milieu ne peut plus être considéré comme électriquement neutre à l'échelle locale. Cependant il garde une certaine cohésion et est **globalement** neutre. On parle souvent de "gaz ionisé" et c'est historiquement ainsi que ce 4^e état de la matière fut découvert, à travers des expériences de décharge électrique dans des tubes remplis de gaz, initiées par Crookes en 1879 (l'inventeur du tube cathodique). Ce fut Thomson en 1895 qui comprit la nature de la matière présente dans les tubes de Crookes et ce n'est qu'en 1928 que le terme "plasma" apparut, proposé par Irving Langmuir. Cependant un plasma n'est pas a priori forcément un gaz, car on peut obtenir un milieu partiellement ionisé à partir d'un état condensé, liquide ou solide.

A l'état naturel, les plasmas sont très peu présents sur Terre, ce qui permet notre existence puisque les plasmas ne sont pas vraiment propices à l'apparition de la vie. Ils sont cependant l'état de la matière de loin le plus représenté à l'échelle de l'Univers. Les étoiles, les nébuleuses, les quasars et les pulsars sont les principaux exemples de plasmas astronomiques (au moins partiels). Plus proches de nous, les plasmas se manifestent sous la forme d'éclairs ou d'aurores boréales (celles-ci étant dûes à la rencontre entre le vent solaire et l'ionosphère, deux autres plasmas). Les applications industrielles des plasmas sont encore assez peu nombreuses, mais on peut citer les néons et les téléviseurs qui se basent sur le principe de décharge électrique dans un gaz inerte, comme introduit plus haut, ou encore les torches à plasmas. La physique des plasmas reste majoritairement cantonnée aux centres de recherches, aussi bien civils que militaires. La caractérisation la plus juste se fait généralement par la densité et la température, comme on peut le voir sur la figure 1.2.

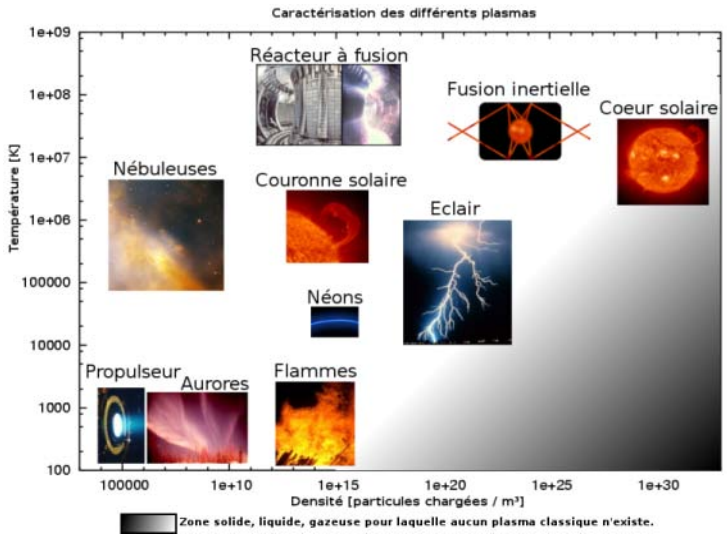


FIGURE 1.2 – Exemples de plasmas divers (source : Wikipedia)

La Magnétohydrodynamique (ci-après désignée MHD) est un modèle décrivant certains types de plasmas, ou plutôt des plasmas dans certains régimes. Le terme fut introduit par Hannes Alfvén en 1942, un des pères de la MHD, dont les travaux lui valurent le Prix Nobel en 1970. Il existe d'ailleurs différents sous-modèles de MHD qui diffèrent par les approximations qui leurs sont associées. Ils ont cependant tous la particularité d'être des modèles décrivant le plasma comme un milieu complètement ionisé et continu, autrement dit de négliger les phénomènes purement cinétiques (dont la description nécessite de distinguer les particules). Le chapitre suivant est consacré à la présentation de ce modèle d'un point de vue physique et mathématique. Disons simplement ici que le modèle utilisé sera celui d'une MHD "classique" non relativiste. Il peut être enrichi pour créer des sous-modèles comme la MHD Hall (qui prend en compte les courants de Hall), la MHD bi-fluide qui se rapproche du modèle cinétique à deux fluides, ou encore des modèles de MHD anisotropes où les constantes physiques telles que la résistivité, ou la conductivité thermique, deviennent des tenseurs. Tout dépend des applications visées. Dans les plasmas chauds tels que rencontrés dans les tokamaks, ou dans les plasmas stellaires, la MHD classique est une bonne première approche. Toutefois la prise en compte de l'anisotropie naturelle des écoulements pourrait s'avérer nécessaire au final. Les modèles Hall ou bi-fluide, ou tous ceux qui rendent compte de phénomènes cinétiques, ne doivent être considérés qu'en cas de réelle nécessité car les temps caractéristiques qu'ils mettent en jeu sont bien plus courts que ceux de la MHD classique.

La magnétohydrodynamique doit être vue comme une généralisation de la mécanique des fluides, puisqu'on y ajoute simplement le fait que, la matière étant entièrement ionisée, les champs électromagnétiques intéragissent avec l'écoulement. Si on couple les équations de Navier-Stokes avec celles de Maxwell, on donne naissance aux équations de l'électromagnétohydrodynamique (EMHD). Si on applique au plasma des champs électrostatiques ou que celui-ci en crée, et qu'on n'a pas de champs magnétique autre que celui induit, on parle d'Électrohydrodynamique (EHD). Si à l'inverse, il n'y a pas de champ électrostatique global mais seulement des champs magnétiques, on parle de MHD. C'est le cas dans les applications que nous avons citées et auxquelles nos méthodes s'appliqueront. Il s'ensuit que la richesse des phénomènes pouvant avoir lieu dans ces plasmas est beaucoup plus grande qu'en mécanique des fluides, d'où la complexité de ce domaine. Les ondes qu'on y trouve peuvent être des échanges d'informations tripartites, entre par exemple la pression, la vitesse et le champ magnétique dans le cadre de la MHD. En EMHD, il y en a encore davantage. Tout ceci ne fait qu'accroître la nécessité d'avoir recours à la simulation numérique pour prédire certains écoulements ou appréhender certains phénomènes, comme les nombreuses instabilités MHD.

Comme annoncé dans la section précédente, c'est le sujet qui nous a intéressé. Nous commencerons par étudier plus en détail le modèle de la MHD que nous utiliserons, de son origine physique à sa structure mathématique (sans prétendre être exhaustifs). Nous présenterons ensuite les schémas numériques que nous avons appliqués à notre modèle, qui ont surtout été éprouvés sur des problèmes de mécanique des fluides jusque-là, puis l'adaptation que nous en avons faite à l'instationnaire tout en prenant en compte la contrainte $\overrightarrow{\nabla} \cdot \overrightarrow{B} = 0$, ainsi que la discrétisation de la partie diffusive des équations et certaines perpectives. Enfin, nous illustrerons nos propos par des simulations permettant de tester divers aspects de nos méthodes. Nous concluerons en reprenant les perspectives évoquées et en y ajoutant des directions possibles de poursuite de nos travaux.

Chapitre 2

Présentation du modèle de la Magnétohydrodynamique

Sommaire

2.1 Description des équations

La Magnétohydodynamique est une approximation de la physique des plasmas qui considère le "bain" d'ions et d'électrons comme un seul fluide. Ce modèle est dérivé de la théorie cinétique des gaz ionisés, ce qui explique que les variables MHD soient des moyennes des quantités associées aux ions et aux électrons.

Nous ne nous attacherons pas ici à retrouver cette dérivation, mais simplement à fournir une compréhension intuitive des équations. Pour cela il suffit de considérer un volume de plasma comme un fluide classique mais dont les électrons se sont détachés des noyaux. Autrement dit, il est constitué de charges mobiles au niveau microscopique, bien qu'il soit neutre à tout niveau macroscopique, car tout déséquilibre de charges entraîne une réorganisation rapide visant à rétablir localement la neutralité, sur des échelles de temps bien inférieures à celles des phénomènes macroscopiques de la MHD. On est donc en présence de champs électromagnétiques auto-générés et le plasma répondra à tout champ électromagnétique provenant de l'extérieur. Le reste découle simplement de l'application du second principe de Newton et des

équations de Maxwell. Les applications que nous visons nous permettent de nous restreindre au cas de plasmas non relativistes.

2.1.1 Électromagnétisme

Des phénomènes électromagnétiques sont observés et étudiés depuis l'Antiquité. La plus importante évolution de notre compréhension de ces phénomènes s'est faite au cours du XIX$^{\text{e}}$ siècle, et ce fut finalement Maxwell qui sut formuler une théorie unifiant toutes les découvertes de ses prédecesseurs vers 1865. L'évolution des champs électriques et magnétiques est régie par les quatre équations de Maxwell, tandis que la force de Lorentz décrit l'effet de ces champs sur les porteurs de charge.

La force de Lorentz

A la fin du XIX^{e} siècle, environ 30 ans après Maxwell, Lorentz introduisit une théorie de la matière composée d'atomes constitués de fragments électriquement chargés, les ions et les électrons. Pour appliquer cette théorie microscopique à l'électromagnétisme de Maxwell qui est une description macroscopique, il détermina la force que subit une particule chargée q se mouvant à une vitesse $\overrightarrow{v}$ dans un champ magnétique $\overrightarrow{B}$ et un champ électrique $\overrightarrow{E}$. On la désigne à présent comme la force de Lorentz.

$$\overrightarrow{F_{Lo}} = q\left(\overrightarrow{E} + \overrightarrow{v} \wedge \overrightarrow{B}\right)$$

D'un point de vue macroscopique, en présence d'un continuum de charges de densité ρ_e la force volumique locale de Lorentz s'écrit :

$$\overrightarrow{f_{Lo}} = \rho_e\left(\overrightarrow{E} + \overrightarrow{v_e} \wedge \overrightarrow{B}\right)$$

On peut définir ici par abus de langage le "champ électromagnétique" $\overrightarrow{E_{EM}} = \overrightarrow{E} + \overrightarrow{v_e} \wedge \overrightarrow{B}$ tel que $\overrightarrow{f_{Lo}} = \rho_e \overrightarrow{E_{EM}}$. Nous avons introduit la vitesse locale des porteurs de charge $\overrightarrow{v_e}$ pour ne pas confondre avec la vitesse moyenne du volume infinitésimal de plasma local que nous serons amenés plus tard à désigner $\overrightarrow{u}$. En effet, bien qu'on modélise le plasma comme un fluide mono-espèce, la réalité est que les électrons se diffusent beaucoup plus facilement que les ions, qu'on considère même parfois immobiles par rapport au mouvement d'ensemble. C'est l'origine de la différence entre les deux vitesses et de l'apparition de courants dans le plasma. Pour éviter par la suite ce problème, on peut noter que par définition la densité de courant $\overrightarrow{j}$ est égale à $\rho_e \overrightarrow{v_e}$. Ainsi, on préfèrera écrire :

$$\overrightarrow{f_{Lo}} = \rho_e \overrightarrow{E} + \overrightarrow{j} \wedge \overrightarrow{B} \tag{2.1.1}$$

La partie magnétique de cette force est appelée la force de Laplace. En toute rigueur, il faudrait également prendre en compte l'effet Hall qui est une conséquence de la migration des électrons, créant des zones de surplus d'électrons et engendrant donc des champs électriques. La différence de densité de charges crée des champs qui doivent théoriquement être pris en compte dans la loi d'Ohm. Cependant cet effet est contrebalancé par la réorganisation rapide des charges qui cherchent en permanence à rétablir la quasi-neutralité. L'effet Hall ne peut donc se produire qu'à des échelles très courtes par rapport à celles qui sont mises en jeu en MHD. C'est pourquoi il est courant de considérer cet effet négligeable, bien que la légimité de cette approximation soit encore sujette à débat chez les physiciens même dans le cadre des expériences de fusion. L'intérêt principal de cette remarque sur la distinction entre les vitesses électronique et d'ensemble, et sur l'effet Hall que nous négligeons, est de comprendre la différence entre notre "force de Lorentz macroscopique", dont la partie magnétique travaille (c'est une force de Laplace), et la partie magnétique de la force de Lorentz microscopique dont le travail à l'échelle de la particule est toujours nul ($\overrightarrow{v} \wedge \overrightarrow{B}$ étant toujours orthogonal à $\overrightarrow{v}$!).

Cette théorie microscopique de l'électromagnétisme et les découvertes théoriques qui y sont attachées firent de Lorentz le plus célèbre physicien de son époque avant l'arrivée d'Einstein. Ses travaux préfiguraient déjà la théorie de la relativité restreinte. Nous allons maintenant revenir au système de Maxwell. Ces équations marquèrent un tournant dans la physique de la fin du XIX^e siècle, comme en témoigne Richard Feynman ([40]) : *From a long view of the history of mankind - seen from, say, ten thousand years from now - there can be little doubt that the most significant event of the 19th century will be judged as Maxwell's discovery of the laws of electrodynamics. The American Civil War will pale into provincial insignifiance in comparison with this important scientific event of the same decade.*

Les contributions de Gauss

Dans les années 1830, Gauss établit deux lois, une portant sur le champ électrique et l'autre sur le champ magnétique. La première, appelée le théorème de Gauss, exprime le fait qu'une charge fixe crée un champ électrique divergent.

$$\int_{\partial V} \overrightarrow{E} \cdot \overrightarrow{\mathrm{d}\partial V} = \frac{Q}{\epsilon_0}$$

Q représente la charge électrique totale contenue dans V et ϵ_0 la *permittivité électrique* du vide. Si ρ_e désigne la densité locale de charges (signée donc), alors $Q = \int_V \rho_e dV$. En appliquant le théorème de Green-Ostrogradski et en remarquant que le théorème est valable pour tout volume V, on déduit une forme locale, l'équation de Maxwell-Gauss :

$$\overrightarrow{\nabla} \cdot \overrightarrow{E} = \frac{\rho_e}{\epsilon_0} \tag{2.1.2}$$

A l'inverse, il n'existe pas de "charges" (on parle de monopoles) magnétiques. C'est en partie ce que traduit une seconde équation, également formulée sous forme intégrale par Gauss à la même période, et intégrée aux autres équations sous forme locale par Maxwell. On l'appelle généralement l'équation de Maxwell-Flux :

$$\overrightarrow{\nabla} \cdot \overrightarrow{B} = 0 \tag{2.1.3}$$

Cette équation est à l'origine de l'introduction du potentiel-vecteur. Puisque $\overrightarrow{\nabla} \cdot \left(\overrightarrow{\nabla} \wedge \cdot\right) = 0$, il peut exister $\overrightarrow{A}$ tel que

$$\overrightarrow{B} = \overrightarrow{\nabla} \wedge \overrightarrow{A}$$

Le potentiel-vecteur n'est pas défini de façon unique, puisqu'on peut remplacer $\overrightarrow{A}$ par $\overrightarrow{A} + \overrightarrow{\nabla} X$ où X est un scalaire différentiable quelconque. Nous reparlerons de ce problème plus loin.

L'équation de Maxwell-Ampère

Peu après les observations d'Ørsted en 1820, Ampère établit la relation entre le courant électrique parcourant un fil et le champ magnétique qu'il génère (et qui perturbe l'aiguille d'une boussole), ce qu'on appelle aujourd'hui le théorème d'Ampère. C'est un résultat fondamental en magnétostatique, qui lie la circulation du champ magnétique le long d'une boucle orientée Γ aux courants "entrelacés" I_k dans cette boucle (i.e. traversant la surface orientée S délimitée par Γ).

$$\int_\Gamma \overrightarrow{B} \cdot \overrightarrow{dl} = \mu_0 \sum_k I_k$$

μ_0 est appelé la *perméabilité magnétique* du vide. En présence d'une densité de courants $\overrightarrow{j}$ traversant S, $\sum_k I_k$ devient naturellement $\int_S \overrightarrow{j} \cdot \overrightarrow{\mathrm{d}S}$. En appliquant le théorème de Stokes au membre de gauche, on

obtient :

$$\int_S \overrightarrow{\nabla} \wedge \overrightarrow{B} \cdot \overrightarrow{\mathrm{d}S} = \mu_0 \int_S \overrightarrow{j} \cdot \overrightarrow{\mathrm{d}S}$$

Ceci étant vrai quelque soit la surface orientée S, on en déduit la forme locale :

$$\overrightarrow{\nabla} \wedge \overrightarrow{B} = \mu_0 \overrightarrow{j}$$

Toutefois, Maxwell trouvait cette équation incomplète. Pour des raisons de symétrie, il imaginait que l'inverse de l'induction magnétique, découverte par Faraday (voir plus bas), soit possible. C'était aussi l'intime conviction de Faraday, qui ne put jamais le démontrer. En d'autres termes, tout comme un champ magnétique variable induit un champ électrique, un champ électrique variable devrait pouvoir induire un champ magnétique. D'autres arguments furent avancés, mais de nos jours, on use d'une méthode simple pour trouver le terme manquant.

Il faut introduire ici une équation que nous n'avons sciemment pas encore utilisée. Ce n'est que depuis les travaux de Lorentz que nous considérons que le courant électrique est la manifestation du déplacement de porteurs de charge. Il peut être vu comme un écoulement de charges, sur lequel on peut faire un bilan de matière : la variation de charge d'un volume $V(t)$ donné est proportionnelle à la quantité de porteurs de charge quittant ou pénétrant ce volume.

$$\int_{V(t)} \frac{\partial \rho_e}{\partial t} dV = -\int_{\partial V(t)} \rho_e \overrightarrow{v_e} \cdot \overrightarrow{n}\, d\partial V$$

En appliquant le théorème de la divergence au terme de droite et en considérant que ceci est vrai pour tout volume $V(t)$, on obtient l'équation dite de conservation de la charge :

$$\frac{\partial \rho_e}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{j} = 0$$

où on rappelle que par définition, la densité de courant est $\overrightarrow{j} = \rho_e \overrightarrow{v_e}$.

Prenons maintenant la divergence de l'équation locale issue du théorème d'Ampère.

$$\overrightarrow{\nabla} \cdot \left(\overrightarrow{\nabla} \wedge \overrightarrow{B} \right) = \mu_0 \overrightarrow{\nabla} \cdot \overrightarrow{j}$$

Le membre de gauche est nul d'après (A.7), ce qui implique que $\mu_0 \overrightarrow{\nabla} \cdot \overrightarrow{j} = 0$. Ceci n'est vrai qu'en régime permanent, et c'est pourquoi le théorème d'Ampère est bien un résultat de magnétostatique. Dans le cas général, dynamique, il manque un terme pour que la conservation de la charge soit respectée. La divergence de ce terme doit être égale à $\partial_t \rho_e$. Si on fait intervenir l'équation de Maxwell-Gauss :

$$\frac{\partial \rho_e}{\partial t} = \frac{\partial}{\partial t} \left(\epsilon_0 \overrightarrow{\nabla} \cdot \overrightarrow{E} \right) = \epsilon_0 \overrightarrow{\nabla} \cdot \frac{\partial \overrightarrow{E}}{\partial t}$$

La conservation de la charge se reformule :

$$\overrightarrow{\nabla} \cdot \left(\overrightarrow{j} + \overrightarrow{j_D} \right) = 0$$

où nous introduisons la quantité $\overrightarrow{j_D} = \epsilon_0 \dfrac{\partial \overrightarrow{E}}{\partial t}$ qu'on appelle généralement le courant de déplacement. L'équation locale d'Ampère devrait donc être :

$$\overrightarrow{\nabla} \wedge \overrightarrow{B} = \mu_0 (\overrightarrow{j} + \overrightarrow{j_D})$$

On obtient finalement l'équation de Maxwell-Ampère :

$$\overrightarrow{\nabla} \wedge \overrightarrow{B} = \mu_0 \overrightarrow{j} + \frac{1}{c_0^2} \frac{\partial \overrightarrow{E}}{\partial t} \tag{2.1.4}$$

en définissant une nouvelle constante c_0 telle que $\epsilon_0\mu_0 c_0^2 = 1$. Ce fut la prédiction fondamentale de la théorie de Maxwell : sa correction de l'équation d'Ampère fait que les équations admettent comme solution des ondes électromagnétiques qui se déplacent à la vitesse c_0 dans le vide. c_0 pouvait être mesuré à partir de la mesure de μ_0 et ϵ_0 (expériences sur les charges et les courants) et on mesura que la lumière se déplaçait à la même vitesse, confirmant ainsi sa nature d'onde électromagnétique.

L'induction électromagnétique

Ce phénomène fut mis en évidence par les expériences menées par Faraday en 1831, dont il tira la loi dite de Faraday. En 1834, Lenz en compléta l'énoncé et on parle depuis de la loi de Lenz-Faraday.

$$e = -\frac{\mathrm{d}\Phi}{\mathrm{d}t}$$

Φ étant le flux magnétique traversant le circuit. C'est une loi empirique, comme toutes celles concernant l'électromagnétisme de cette époque. Dans un circuit électrique, on observe l'apparition d'une force électromotrice e en présence d'un champ magnétique qui varie dans le temps. Prenons une section S quelconque, de contour orienté $\partial S = \Gamma$, traversée par un flux magnétique. e est défini comme la circulation du champ électromagnétique total $\overrightarrow{E_{EM}} = \overrightarrow{E} + \overrightarrow{v} \wedge \overrightarrow{B}$ (introduit plus haut comme le champ associé à la force de Lorentz) sur le bord de la section.

$$e = \int_\Gamma \overrightarrow{E_{EM}} \cdot \overrightarrow{\mathrm{d}\Gamma}$$

Il s'agit donc d'une équation intégrale. La variation du flux magnétique à l'origine de l'apparition d'une force électromotrice peut avoir deux origines : une évolution de la taille du circuit et/ou un champ magnétique variable. Comme le dit par exemple Richard Feynman [40], ces deux effets sont indépendants dans leur nature mais sont regroupés au sein de la même loi.

Pour reformuler l'effet d'un champ magnétique variable, prenons une section S immobile ($\overrightarrow{v} = \overrightarrow{0}$).

$$\int_\Gamma \overrightarrow{E} \cdot \overrightarrow{dl} = -\frac{\mathrm{d}}{\mathrm{d}t} \int_S \overrightarrow{B} \cdot \overrightarrow{dS}$$

En appliquant le théorème de Stokes à gauche et en utilisant le fait que le cadre est fixe à droite, on obtient :

$$\int_S \left(\overrightarrow{\nabla} \wedge \overrightarrow{E}\right) \cdot \overrightarrow{\mathrm{d}S} = -\int_S \frac{\partial \overrightarrow{B}}{\partial t} \cdot \overrightarrow{\mathrm{d}S}$$

Cette expression étant vraie pour toute surface orientée S, les intégrandes sont nécessairement égaux, ce qui nous donne l'équation de Maxwell-Faraday :

$$\overrightarrow{\nabla} \wedge \overrightarrow{E} = -\frac{\partial \overrightarrow{B}}{\partial t} \tag{2.1.5}$$

Pour ce qui est de la partie $\overrightarrow{u} \wedge \overrightarrow{B}$ de la force électromotrice, on peut montrer qu'il existe des situations où elle est invalidée par l'expérience, tandis que la partie de e dûe à $\overrightarrow{E}$ est toujours correcte [40]. Il ne faut donc retenir que l'équation ci-dessus dans le cas général, ce qu'a fait Maxwell.

Remarques : Choix de jauge et champ électromoteur

On sait que la force d'attraction électrostatique (interaction de Coulomb) entre deux charges dérive d'une énergie potentielle, et donc que le champ électrique $\overrightarrow{E}$ lui-même dérive d'un potentiel ϕ tel que $\overrightarrow{E} = -\overrightarrow{\nabla}\phi$. Toutefois le rotationnel d'un gradient étant nul, cela signifie qu'il existe une autre composante

au champ $\overrightarrow{E}$ qui vérifie (2.1.5) : le champ induit. On en déduit que le champ électrique total s'exprime sous la forme (c'est-à-dire notamment à des constantes d'intégration près) :

$$\overrightarrow{E} = -\overrightarrow{\nabla}\phi - \frac{\partial \overrightarrow{A}}{\partial t}$$

le terme d'induction s'annulant bel et bien dans le cas statique. Nous avons vu que le potentiel-vecteur était défini à un terme $\overrightarrow{\nabla} X$ près. Le potentiel électrique doit donc être défini à un terme $-\dfrac{\partial X}{\partial t}$ près. De cette façon, quel que soit le choix de la jauge X, $\overrightarrow{E}$ est défini de manière unique et X disparaît de toutes les équations. C'est ce qu'on appelle *l'invariance de jauge de la théorie* : il n'y a pas besoin a priori de se préoccuper de la valeur de la jauge X pour résoudre les équations de Maxwell.

Revenons à la définition de la force électromotrice dans le cas général (la vitesse $\overrightarrow{v}$ étant a priori non nulle) :

$$\begin{aligned} e &= \int_\Gamma \left(\overrightarrow{E} + \overrightarrow{v} \wedge \overrightarrow{B}\right) \cdot \overrightarrow{\mathrm{d}l} \\ &= \int_S \overrightarrow{\nabla} \wedge \left(-\overrightarrow{\nabla}\phi - \frac{\partial \overrightarrow{A}}{\partial t} + \overrightarrow{v} \wedge \overrightarrow{B}\right) \cdot \overrightarrow{\mathrm{d}S} \end{aligned}$$

Or d'après (A.6), $\overrightarrow{\nabla} \wedge \left(\overrightarrow{\nabla}\phi\right) = \overrightarrow{0}$, donc nous pouvons définir le champ électromoteur E_m dont la seule composante électrique est celle induite par le champ magnétique :

$$e = \int_\Gamma \overrightarrow{E_m} \cdot \overrightarrow{\mathrm{d}l} = \int_\Gamma \left(\overrightarrow{v} \wedge \overrightarrow{B} - \frac{\partial \overrightarrow{A}}{\partial t}\right) \cdot \overrightarrow{\mathrm{d}l}$$

L'approximation MHD

Dans le cadre des applications qui nous intéressent, il est d'usage de considérer que le courant de déplacement est négligeable tant qu'on se situe dans des régimes loin d'être relativistes. Par conséquent, nous ne travaillerons pas avec l'équation (2.1.4) mais avec une version simplifiée qui se trouve être la forme originale (forme locale du théorème découvert par Ampère en magnétostatique) :

$$\overrightarrow{\nabla} \wedge \overrightarrow{B} = \mu_0 \overrightarrow{j} \tag{2.1.6}$$

Une analyse simple des ordres de grandeurs peut être effectuée ([44]) pour s'en convaincre. Prenons des quantités caractéristiques du problème Δl_0 (distance), Δt_0 (durée), B_0 (champ magnétique), E_0 (champ électrique) et $v_0 = \dfrac{\Delta l_0}{\Delta t_0}$. D'après l'équation de Maxwell-Faraday (2.1.5) :

$$\left|\overrightarrow{\nabla} \wedge \overrightarrow{E}\right| \sim \frac{E_0}{\Delta l_0} = \left|-\frac{\partial \overrightarrow{B}}{\partial t}\right| \sim \frac{B_0}{\Delta t_0}$$

$$\Rightarrow \qquad E_0 \sim B_0 v_0$$

Et si nous comparons les ordres de grandeur du courant de déplacement et du rotationnel du champ magnétique (cf. (2.1.4)) :

$$\frac{1}{c_0^2}\left|\frac{\partial \overrightarrow{E}}{\partial t}\right| \sim \frac{1}{c_0^2}\frac{E_0}{\Delta t_0} \quad \text{et} \quad \left|\overrightarrow{\nabla} \wedge \overrightarrow{B}\right| \sim \frac{B_0}{\Delta l_0}$$

$$\Rightarrow \quad \frac{1}{c_0^2}\left|\frac{\partial \overrightarrow{E}}{\partial t}\right| \sim \frac{v_0^2}{c_0^2}\frac{B_0}{\Delta l_0} \quad \ll \quad \left|\overrightarrow{\nabla} \wedge \overrightarrow{B}\right| \sim \frac{B_0}{\Delta l_0}$$

Ce qui confirme que le courant de déplacement est négligeable si $v_0 \ll c_0$, c'est-à-dire pour des vitesses d'ondes du plasma non relativistes. La même hypothèse est parfois faite en électronique et en électrotechnique, où on parle alors d'*Approximation des régimes quasi-stationnaires* (ARQS).

La résistivité

On a dit en parlant des forces de Lorentz et de Laplace que le courant était une manifestation de la diffusion des électrons. S'il y a diffusion, il est probable qu'il y ait des collisions et donc que le courant ait du mal à se propager. C'est ce que traduit la notion de résistivité du milieu, une opposition à la circulation du courant.

Les premières mesures de la résistance (équivalent macroscopique de la résistivité) furent effectuées par Cavendish en 1781 puis surtout par Ohm entre 1825 et 1827. La loi d'Ohm historique relie la différence de potentiel électrique ΔV aux bornes une résistance R à l'intensité I du courant la traversant.

$$\Delta V = RI$$

Afin d'expliquer théoriquement cette loi, Ohm s'inspira beaucoup de la loi de Fourier sur la conduction de la chaleur, tant il est vrai que les deux lois présentent de fortes similitudes (potentiel/température, intensité électrique/flux thermique, $\frac{1}{R}$/conductivité thermique).

De nos jours, la résistivité est définie de façon cinétique en considérant les collisions à l'échelle microscopique. Ainsi, lorsqu'on étudie le modèle cinétique à deux fluides de la MHD, on peut trouver une équation sur la densité de courant $\overrightarrow{j}$, qui est une équation d'évolution qu'on appelle *Loi d'Ohm généralisée*. Cependant les temps caractéristiques mis en jeu sont trop courts pour être traités dans le cadre de la MHD. On utilise alors une version simplifiée qui constitue une bonne approximation pour peu qu'on ait une bonne estimation formelle de la résistivité η, qui est déterminée en modélisant les termes de collision de l'équation (ou expérimentalement). On aboutit alors à la version la plus courante de la loi d'Ohm :

$$\overrightarrow{E} + \overrightarrow{u} \wedge \overrightarrow{B} = \eta \overrightarrow{j} \tag{2.1.7}$$

où $\overrightarrow{u}$ est bel et bien la vitesse du plasma et non la vitesse thermique des électrons. Pour une présentation détaillée de la loi d'Ohm généralisée à l'échelle de la théorie cinétique, le lecteur pourra se référer à [44].

2.1.2 Couplage avec les lois de conservation de l'hydrodynamique

Lorsqu'on travaille à la simulation de phénomènes physiques, il est très intéressant de travailler autant que possible avec des variables dont la physique nous dit qu'elle doivent être conservées (on dira *conservatives*), afin que l'on puisse faire en sorte que cela soit le cas numériquement. C'est ce qui justifie notre choix de formuler à présent le problème à résoudre en terme de variables conservatives et donc de *lois de conservation*.

Notations

Avant de commencer cette section, il est utile de préciser certaines notations que nous serons amenés à employer pour l'écriture des équations sous leur forme définitive. Suivant les règles classiques du produit matriciel (dans $\mathbb{R}^3$ par exemple), on écrira régulièrement la matrice $\overrightarrow{X}\overrightarrow{Y}^t$.

$$\overrightarrow{X}\overrightarrow{Y}^t = \begin{pmatrix} X_1Y_1 & X_1Y_2 & X_1Y_3 \\ X_2Y_1 & X_2Y_2 & X_2Y_3 \\ X_3Y_1 & X_3Y_2 & X_3Y_3 \end{pmatrix}$$

Supposons que ces deux vecteurs soient des variables en espace, nous écrirons la divergence de cette matrice de la manière suivante :

$$\overrightarrow{\nabla} \cdot \left(\overrightarrow{X} \overrightarrow{Y}^t \right) = \begin{pmatrix} \overrightarrow{\nabla} \cdot (X_1 \overrightarrow{Y}) \\ \overrightarrow{\nabla} \cdot (X_2 \overrightarrow{Y}) \\ \overrightarrow{\nabla} \cdot (X_3 \overrightarrow{Y}) \end{pmatrix} = \left(\overrightarrow{Y} \cdot \overrightarrow{\nabla} \right) \overrightarrow{X} + \overrightarrow{X} \overrightarrow{\nabla} \cdot \overrightarrow{Y} \tag{2.1.8}$$

De plus, les matrices résultant de la dérivation d'un vecteur X par un vecteur Y seront de terme général :

$$\left(\frac{\partial \overrightarrow{X}}{\partial \overrightarrow{Y}} \right)_{ij} = \frac{\partial X_i}{\partial Y_j} \tag{2.1.9}$$

avec i l'indice des lignes et j celui des colonnes. Une application directe est le calcul du gradient d'un vecteur Y qui est de terme général :

$$\left(\overrightarrow{\nabla} \overrightarrow{Y} \right)_{ij} = \left(\frac{\partial \overrightarrow{Y}}{\partial \overrightarrow{x}} \right)_{ij} = \frac{\partial Y_i}{\partial x_j}$$

Enfin, nous aurons souvent recours à des formules d'analyse vectorielle dont le rappel systématique alourdirait beaucoup le texte, et que nous avons donc placées en annexes (cf. Annexe A).

La description Lagrangienne du mouvement

Pour comprendre le sens physique des équations de conservation, le plus simple est de changer le point de vue de l'observateur. La description Lagrangienne consiste précisément à placer l'observateur dans un référentiel lié au système que l'on étudie. Ici, l'objet est un milieu continu, de type fluide, ce qui signifie que l'observateur suit une certaine quantité de matière. Ceci se traduit par le fait que la vitesse du référentiel se confond avec la vitesse moyenne du volume de matière suivi. Autrement dit, la vitesse globale du volume de matière (qui est donc a priori déformable) est nulle dans notre référentiel mobile. D'un point de vue mathématique, ce procédé fait que la position d'un volume n'est plus une donnée fixe, comme dans une description Eulérienne, mais qu'elle est une fonction du temps. Ainsi, tout le système ne dépend que d'une seule variable, le temps, et de plus la taille du volume varie suivant la contraction ou la dilatation de la matière.

Si on remonte à une date t_0 antérieure à t, on peut définir la transformation qui décrit l'évolution spatiale d'un volume infinitésimal de matière, qu'on appellera parfois "particule", entre ces deux dates. Autrement dit, on s'intéresse à la fonction décrivant les trajectoires de la matière.

$$\overrightarrow{f} : \begin{cases} \Omega_0 \times \mathbb{R}^+ & \longrightarrow & \Omega \\ \left(\overrightarrow{X}, t \right) & \longmapsto & \overrightarrow{x} \end{cases}$$

À supposer que nous excluions les zones de discontinuité dans notre analyse, f est un $\mathcal{C}^\infty$-difféomorphisme de Ω_0 dans Ω. Les régions où ceci n'est plus vrai sont toujours des hypersurfaces du domaine. Par exemple, dans notre cas comme en mécanique des fluides, il peut s'agir de fronts d'ondes (chocs, discontinuités de contact, etc...) ou de parois. Les hypothèses de régularité sur f et sur la solution sont donc toujours vraies hors de régions négligeables au sens de Lebesgue (de mesure nulle), c'est-à-dire qu'elles le sont *presque partout*. Les équations que nous allons présenter par la suite seront donc aussi à ne considérer qu'en dehors des zones de discontinuité.

La transition d'une description Lagrangienne à une description Eulérienne de l'évolution d'une quantité

U quelconque se fait donc de la manière suivante :

$$\frac{\mathrm{d}}{\mathrm{d}t}U\left(\overrightarrow{f}(\overrightarrow{X},t),t\right)=\frac{\partial}{\partial t}U\left(\overrightarrow{f}(\overrightarrow{X},t),t\right)+\frac{\partial\overrightarrow{f}}{\partial t}(\overrightarrow{X},t)\cdot\frac{\partial U}{\partial\overrightarrow{x}}\left(\overrightarrow{f}(\overrightarrow{X},t),t\right)$$

$$=\partial_t U\left(\overrightarrow{x},t\right)+\overrightarrow{v}\left(\overrightarrow{X},t\right)\cdot\overrightarrow{\nabla}_x U\left(\overrightarrow{x},t\right)$$

Les dérivées temporelles $\frac{\mathrm{d}}{\mathrm{d}t}$ qui apparaissent dans le cadre Lagrangien sont appelées entre autres des dérivées Lagrangiennes, ou encore particulaires, mais bien d'autres noms existent. La jacobienne de $\overrightarrow{f}$ est définie grâce aux hypothèses que nous avons faites, et on peut remarquer que $\frac{\partial}{\partial t}\overrightarrow{f}(\overrightarrow{X},t)=\overrightarrow{v}(\overrightarrow{X},t)$ est la vitesse Lagrangienne de la "particule".

La vitesse Eulérienne $\overrightarrow{u}$ est par définition une fonction des coordonnées Eulériennes, soit $\overrightarrow{u}\left(\overrightarrow{x},t\right)=\overrightarrow{u}\left(\overrightarrow{f}(\overrightarrow{X},t),t\right)$. À t fixé, on peut donc dire que $\overrightarrow{u}\left(\overrightarrow{f}(\cdot,t),t\right)=(\overrightarrow{u}\circ\overrightarrow{f})\Big|_t(\cdot)$ est mathématiquement semblable à une vitesse Lagrangienne comme $\overrightarrow{v}(\cdot,t)$. De fait, $\forall t$, on fera l'assimilation $(\overrightarrow{u}\circ\overrightarrow{f})\Big|_t=\overrightarrow{v}|_t$ et par conséquent, d'un point de vue instantané, on considère que $\overrightarrow{u}(\overrightarrow{x},t)=\overrightarrow{v}\left(\overrightarrow{X},t\right)$ mais que les variations de ces deux fonctions dans le temps ne seront pas les mêmes (car les deux fonctions sont attachées à deux formalismes distincts). Il est alors correct d'écrire :

$$\frac{\mathrm{d}}{\mathrm{d}t}U\left(\overrightarrow{f}(\overrightarrow{X},t),t\right)=\partial_t U\left(\overrightarrow{x},t\right)+\overrightarrow{u}\left(\overrightarrow{x},t\right)\cdot\overrightarrow{\nabla}_x U\left(\overrightarrow{x},t\right)\tag{2.1.10}$$

Lorsqu'on étudie un **continuum**, on s'intéresse à des quantités définies localement (pour toute "particule" assimilée à un point), et non sur un volume, alors que comme nous allons le voir, les raisonnements physiques s'appuient sur des volumes mobiles et déformables $V(t)$ propres à une description Lagrangienne. Or toute information $I(t)$ définie sur un volume peut s'écrire à l'aide d'une quantité locale U :

$$I(t)=\int_{V(t)}U(\overrightarrow{x}(t),t)d\overrightarrow{x}$$

Nous cherchons à donner une expression de $\frac{\mathrm{d}I}{\mathrm{d}t}$ en fonction de U. Ceci nous sera utile par la suite. I peut être réécrit via un changement de variables entre $\overrightarrow{x}$ et $\overrightarrow{X}$, qui reflète bien l'idée de conservation de la matière (nos "particules") puisqu'on définit la même quantité (*portée* par les "particules") sur le volume V_0 qu'elle occupait à une date antérieure (en supposant $t_0<t$ même si rien ne nous y oblige).

$$I(t)=\int_{V_0}U\left(\overrightarrow{f}(\overrightarrow{X},t),t\right)\det\left(\frac{\partial\overrightarrow{f}}{\partial\overrightarrow{X}}(\overrightarrow{X},t)\right)d\overrightarrow{X}$$

La jacobienne de $\overrightarrow{f}$ existe et est inversible grâce aux hypothèses que nous avons faites. On notera $\boldsymbol{J}$ le jacobien de la transformation, i.e. $\boldsymbol{J}\left(\overrightarrow{X},t\right)=\det\left(\frac{\partial\overrightarrow{f}}{\partial\overrightarrow{X}}(\overrightarrow{X},t)\right)$.

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}I(t)&=\frac{\mathrm{d}}{\mathrm{d}t}\int_{V_0}U\left(\overrightarrow{f}(\overrightarrow{X},t),t\right)\boldsymbol{J}\left(\overrightarrow{X},t\right)d\overrightarrow{X}\\&=\int_{V_0}\frac{\mathrm{d}U}{\mathrm{d}t}\left(\overrightarrow{f}(\overrightarrow{X},t),t\right)\boldsymbol{J}\left(\overrightarrow{X},t\right)d\overrightarrow{X}+\int_{V_0}U\left(\overrightarrow{f}(\overrightarrow{X},t),t\right)\frac{\mathrm{d}\boldsymbol{J}}{\mathrm{d}t}\left(\overrightarrow{X},t\right)d\overrightarrow{X}\\&=\int_{V(t)}\left(\frac{\partial U}{\partial t}+(\overrightarrow{u}\cdot\overrightarrow{\nabla}_x)U\right)(\overrightarrow{x},t)d\overrightarrow{x}+\int_{V_0}U\left(\overrightarrow{f}(\overrightarrow{X},t),t\right)\frac{\partial\boldsymbol{J}}{\partial t}\left(\overrightarrow{X},t\right)d\overrightarrow{X}\end{aligned}$$

où nous avons réutilisé la formule de changement de variable ainsi que l'égalité (2.1.10). Le dernier terme nécessite de dériver le déterminant ce qui est un peu calculatoire. Nous avons préféré mettre le détail de la démonstration dans l'annexe A.2. Le résultat à retenir est :

$$\frac{\partial \boldsymbol{J}}{\partial t}\left(\overrightarrow{X},t\right) = \boldsymbol{J}\left(\overrightarrow{X},t\right)\overrightarrow{\nabla}\cdot\overrightarrow{u}\left(\overrightarrow{x},t\right)$$

On arrive donc finalement à l'expression suivante :

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}I(t) &= \int_{V(t)}\left(\frac{\partial U}{\partial t} + (\overrightarrow{u}\cdot\overrightarrow{\nabla}_x)U + U\overrightarrow{\nabla}_x\cdot\overrightarrow{u}\right)(\overrightarrow{x},t)d\overrightarrow{x} \\ \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}I(t) &= \int_{V(t)}\left(\frac{\partial U}{\partial t} + \overrightarrow{\nabla}_x\cdot(U\overrightarrow{u})\right)(\overrightarrow{x},t)d\overrightarrow{x} \end{aligned}$$

Enfin, comme il n'y a pas d'ambiguïtés sur les dépendances de chaque variable, on se permettra aussi de les occulter en écrivant :

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{V(t)} U dV = \int_{V(t)}\left(\frac{\partial U}{\partial t} + \overrightarrow{\nabla}\cdot(U\overrightarrow{u})\right)dV \tag{2.1.11}$$

Afin d'alléger les notations, on procèdera toujours ainsi par la suite.

Conservation de la masse

Imaginons que nous suivions un volume élémentaire V de plasma. En l'absence de phénomènes de diffusion (rencontrés principalement lorsque plusieurs espèces chimiques sont présentes) ou de réactions chimiques (on a dit que le fluide était ici supposé constitué d'une seule espèce), on s'attend à ce que la quantité de matière de ce volume reste constante dans le temps. Quitte à ce que le volume se déforme en cas de contraction ou d'expansion. Or, la masse de chaque "particule" de plasma étant également invariante dans le temps, la masse m de notre volume ne peut pas varier. Ceci s'écrit :

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\int_{V(t)}\rho(t,\overrightarrow{x})dV = 0$$

Ce qui devient en utilisant la relation (2.1.11) :

$$\int_{V(t)}\left(\frac{\partial \rho}{\partial t} + \overrightarrow{\nabla}\cdot(\rho\overrightarrow{u})\right)dV = 0$$

Or ceci est vrai quel que soit le volume $V(t)$. Par conséquent, la forme locale de cette équation est également vraie et s'écrit :

$$\frac{\partial \rho}{\partial t} + \overrightarrow{\nabla}\cdot(\rho\overrightarrow{u}) = 0$$

Il s'agit de l'équation de conservation de la masse, également appelée équation de continuité, sous sa forme locale. Elle fait partie du système de la MHD idéale.

Une fois arrivés ici, il est intéressant de se rappeler que la vraie nature du plasma est multi-fluide, la MHD ne représentant qu'une approximation à grande échelle. Or dans la pratique, à l'échelle de la théorie cinétique, il existe de nombreux phénomènes turbulents dont on ne peut rendre compte à l'échelle de la MHD. Certains ont peu d'influence à grande échelle, d'autres moins. Parmi eux, il y a les modes turbulents dûs aux gradients de température ionique (ITG en anglais) dont les effets observables à l'échelle de la MHD sont semblables à une diffusion générale de la matière. Cette observation a été mise en avant par les physiciens du CEA Cadarache pour proposer une modification de l'équation de continuité prenant en compte un phénomène de diffusion qui modélise l'action des structures turbulentes ITG à petite échelle.

Naturellement, introduire une diffusion de matière revient à prendre en compte un flux de matière Φ_ρ^D supplémentaire. Le terme à rajouter doit donc déprendre de cette quantité. En comptant ce flux positif lorsqu'il est sortant (i.e. dans la même direction que $\overrightarrow{dS}$), on peut écrire la seconde loi de Fick :

$$\frac{\mathrm{d}m}{\mathrm{d}t} = -\Phi_\rho^D = -\int_{\partial V(t)} \overrightarrow{j}_\rho^D \cdot \overrightarrow{dS}$$

où $\overrightarrow{j}_\rho^D$ est la densité surfacique de flux diffusifs. En appliquant les mêmes procédés que précédemment, on trouve une écriture locale à cette nouvelle loi :

$$\frac{\partial \rho}{\partial t} + \overrightarrow{\nabla} \cdot (\rho \overrightarrow{u}) + \overrightarrow{\nabla} \cdot \overrightarrow{j}_\rho^D = 0$$

En 1855, le physiologiste allemand Adolph Fick établit une loi empirique selon laquelle le flux de diffusion de matière est proportionnel au gradient de la concentration de l'espèce concernée. C'est ce qu'on appelle depuis la première loi de Fick. En passant de la concentration à ρ (par exemple en multipliant par la masse molaire, constante, si l'unité de concentration utilise les moles), et en appliquant cette loi à notre espèce unique, on obtient alors :

$$\overrightarrow{j}_\rho^D = -D\overrightarrow{\nabla}\rho$$

où D est un coefficient de diffusion, qui est dans la pratique toujours variable. On aboutit finalement à notre équation complète de conservation de la masse :

$$\frac{\partial \rho}{\partial t} + \overrightarrow{\nabla} \cdot (\rho \overrightarrow{u}) - \overrightarrow{\nabla} \cdot \left(D\overrightarrow{\nabla}\rho\right) = 0$$

Conservation de la quantité de mouvement

Revenons à notre volume test de matière $V(t)$, de masse totale m, et définissons sa vitesse d'ensemble $\overrightarrow{v}$ telle que :

$$m\overrightarrow{v} = \int_{V(t)} \rho \overrightarrow{u}\, dV$$

En réalité, on définit la quantité de mouvement $\overrightarrow{p}(t) = \int_{V(t)} \rho \overrightarrow{u}\, dV$. La deuxième loi de Newton stipule que la variation de $\overrightarrow{p}$ dans le temps est proportionnelle à l'action des forces extérieures s'appliquant sur $V(t)$. Dans notre cas, il s'agit des contraintes appliquées sur les bords de $V(t)$, $\overrightarrow{F_\sigma}$, et de la force volumique de Lorentz $\overrightarrow{F_{Lo}}$ qui s'applique dès qu'on est en présence d'un milieu non électriquement neutre. A l'échelle d'un tokamak (de l'ordre de quelques mètres), il est clair que comparé à ces forces (ne serait-ce que l'interaction électromagnétique), l'interaction gravitationelle est tout à fait négligeable. En astrophysique, les plasmas sont généralement d'une toute autre taille et on ne peut négliger les effets gravitationnels qu'à condition de se trouver très loin des objets massifs les plus proches (planètes, étoiles, ...).

$$\frac{\mathrm{d}\,(m\overrightarrow{v})}{\mathrm{d}t} = \sum \overrightarrow{F}_\sigma + \sum \overrightarrow{F}_{Lo}$$

Commençons par rappeler l'expression de la force de Lorentz (2.1.1) :

$$\sum \overrightarrow{F_{Lo}} = \int_{V(t)} \overrightarrow{f_{Lo}} dV = \int_{V(t)} \left(\rho_e \overrightarrow{E} + \overrightarrow{j} \wedge \overrightarrow{B}\right) dV$$

où ρ_e désigne la densité de charges et $\overrightarrow{j}$ la densité de courants. La MHD modélise des plasmas totalement ionisés où les effets des champs électriques sont généralement négligeables. Cet état de fait peut s'expliquer par une estimation des ordres de grandeur, comme on l'a fait pour justifier l'approximation MHD (qui nous affranchissait du courant de déplacement). Prenons à nouveau des grandeurs caractéristiques Δl_0,

Δt_0, B_0 et $v_0 = \dfrac{\Delta l_0}{\Delta t_0}$ et souvenons-nous que nous avons déjà établi que $E_0 \sim B_0 v_0$. Pour trouver une estimation de ρ_e, utilisons l'équation de Maxwell-Gauss (2.1.2) :

$$\overrightarrow{\nabla} \cdot \overrightarrow{E} = \frac{\rho_e}{\epsilon_0}$$

$$\Rightarrow \quad \rho_e \sim \epsilon_0 \frac{E_0}{\Delta l_0}$$

$$\Rightarrow \quad \rho_e \overrightarrow{E} \sim \frac{v_0^2}{c_0^2} \frac{B_0^2}{\mu_0 \Delta l_0}$$

Nous avons à nouveau fait usage du fait que $\epsilon_0 \mu_0 = \dfrac{1}{c_0^2}$. D'un autre côté, en faisant appel à notre version simplifiée de l'équation de Maxwell-Ampère (2.1.6) :

$$\overrightarrow{\nabla} \wedge \overrightarrow{B} = \mu_0 \overrightarrow{j}$$

$$\Rightarrow \quad \overrightarrow{j} \sim \frac{B_0}{\mu_0 \Delta l_0}$$

$$\Rightarrow \quad \overrightarrow{j} \wedge \overrightarrow{B} \sim \frac{B_0^2}{\mu_0 \Delta l_0}$$

Il est donc clair que tant que $v_0 \ll c_0$ (cas non relativistes), la force de Lorentz est réduite aux forces de Laplace :

$$\int_{V(t)} \overrightarrow{f_{Lo}} dV = \int_{V(t)} \overrightarrow{j} \wedge \overrightarrow{B} dV$$

Une conséquence de ceci est que l'équation de Maxwell-Gauss ainsi que la densité de charges ρ_e deviennent totalement inutiles et peuvent être oubliées dès maintenant. Reprenons maintenant les calculs et après quelques manipulations, la densité volumique de forces de Lorentz peut être réécrite :

$$\begin{aligned}
\overrightarrow{j} \wedge \overrightarrow{B} &= \frac{1}{\mu_0} \left(\overrightarrow{\nabla} \wedge \overrightarrow{B} \right) \wedge \overrightarrow{B} \\
&\overset{(A.4)}{=} \frac{1}{\mu_0} \left((\overrightarrow{B} \cdot \overrightarrow{\nabla}) \overrightarrow{B} - \frac{1}{2} \overrightarrow{\nabla} (\overrightarrow{B} \cdot \overrightarrow{B}) \right) \\
&\overset{(2.1.8)}{=} \frac{1}{\mu_0} \left(\overrightarrow{\nabla} \cdot (\overrightarrow{B} \overrightarrow{B}^t) - \overrightarrow{B} \overrightarrow{\nabla} \cdot \overrightarrow{B} - \overrightarrow{\nabla} (\frac{\overrightarrow{B}^2}{2}) \right) \\
&\overset{(2.1.3)}{=} \frac{1}{\mu_0} \left(\overrightarrow{\nabla} \cdot (\overrightarrow{B} \overrightarrow{B}^t) - \overrightarrow{\nabla} (\frac{\overrightarrow{B}^2}{2}) \right)
\end{aligned}$$

Il s'agit de la divergence de la partie magnétique de ce qu'on appelle *le tenseur des contraintes de Maxwell*. On y distingue deux contributions.

Gradient de pression magnétique $\qquad \overrightarrow{\nabla} \left(\dfrac{\overrightarrow{B}^2}{2\mu_0} \right)$

Tension magnétique $\qquad \dfrac{1}{\mu_0} \left(\overrightarrow{B} \cdot \overrightarrow{\nabla} \right) \overrightarrow{B} = \dfrac{1}{\mu_0} \overrightarrow{\nabla} \cdot \left(\overrightarrow{B} \overrightarrow{B}^t \right)$

Passons à présent aux effets des contraintes matérielles (de contact). Une contrainte de compression est comptée négative par convention, à l'inverse de la pression hydrodynamique classique. Considérant à nouveau notre volume $V(t)$ de plasma, les efforts des contraintes se manifestent aux bords de ce volume.

On introduit le tenseur des contraintes $\boldsymbol{\sigma}$ qui donne l'état de contraintes en un point selon les plans orthogonaux aux axes de référence.

$$\sum \overrightarrow{F_\sigma} = \int_{\partial V(t)} \boldsymbol{\sigma} : \overrightarrow{n}\, dS = \int_{V(t)} \overrightarrow{\nabla} \cdot \boldsymbol{\sigma} dV$$

Le tenseur des contraintes peut se décomposer en une partie sphérique et une partie déviatorique. La partie sphérique correspond à la pression hydrodynamique.

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix} = \begin{pmatrix} -p & 0 & 0 \\ 0 & -p & 0 \\ 0 & 0 & -p \end{pmatrix} + \begin{pmatrix} s_{xx} & s_{xy} & s_{xz} \\ s_{yx} & s_{yy} & s_{yz} \\ s_{zx} & s_{zy} & s_{zz} \end{pmatrix}$$

L'application de la loi de conservation du moment angulaire permet de montrer que le tenseur $\boldsymbol{\sigma}$, et donc également le déviateur des contraintes $\boldsymbol{s}$, est toujours symétrique.

Le plasma est un fluide ionisé. Tout comme un fluide classique, il ne peut pas, en l'absence de viscosité, supporter de déformations permanentes, plastiques, causées par des contraintes déviatoriques. C'est ce qui distingue ces états de l'état solide. Par conséquent, le déviateur ne peut contenir que la partie d'origine visqueuse des contraintes. Par habitude, on appellera ce tenseur $\boldsymbol{\tau}$ et non $\boldsymbol{s}$.

On modélise généralement le plasma comme un fluide newtonien, ce qui signifie que les contraintes visqueuses sont proportionnelles à la partie symétrique du tenseur des taux de déformation et s'écrit :

$$\boldsymbol{\tau} = \mu \left((\overrightarrow{\nabla}\, \overrightarrow{u}) + (\overrightarrow{\nabla}\, \overrightarrow{u})^t \right) + \lambda \overrightarrow{\nabla} \cdot \overrightarrow{u}\, \boldsymbol{I}$$

où $\boldsymbol{I}$ est la matrice identité, μ est la viscosité dynamique et λ la viscosité de volume. Généralement, on utilise en plus l'hypothèse de Stokes :

$$3\lambda + 2\mu = 0$$

Ceci est faux dans la majorité des cas mais constitue une approximation suffisamment satisfaisante pour nos travaux, dans le sens où sur les cas simples qui sont souvent abordés dans un premier temps pour valider les méthodes numériques, une telle approximation n'engendre aucune anomalie physique (visible en tout cas). Par la suite, et pour des problèmes de plus en plus réalistes, ce choix pourrait être progressivement remis en cause. Pour résumer, en appliquant les mêmes raisonnements que précédemment, nous avons :

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \int_{V(t)} \rho \overrightarrow{u}\, \mathrm{d}V &= \int_{V(t)} \overrightarrow{j} \wedge \overrightarrow{B} \mathrm{d}V + \int_{V(t)} \overrightarrow{\nabla} \cdot (-p\boldsymbol{I} + \boldsymbol{\tau}) \\ \Rightarrow \int_{V(t)} \left(\frac{\partial(\rho \overrightarrow{u})}{\partial t} + \overrightarrow{\nabla} \cdot (\rho \overrightarrow{u}\, \overrightarrow{u}^t) \right) \mathrm{d}V &= \int_{V(t)} \left(\frac{1}{\mu_0} \overrightarrow{\nabla} \cdot \left(\overrightarrow{B}\, \overrightarrow{B}^t - \frac{\overrightarrow{B}^2}{2} \boldsymbol{I} \right) + \overrightarrow{\nabla} \cdot (-p\boldsymbol{I} + \boldsymbol{\tau}) \right) \mathrm{d}V \\ &\quad - \int_{V(t)} \frac{1}{\mu_0} \overrightarrow{B}\, \overrightarrow{\nabla} \cdot \overrightarrow{B}\, dV \end{aligned}$$

Ceci est vrai pour tout volume $V(t)$, et donc aussi localement.

$$\frac{\partial(\rho \overrightarrow{u})}{\partial t} + \overrightarrow{\nabla} \cdot \left[\rho \overrightarrow{u}\, \overrightarrow{u}^t + \left(p + \frac{\overrightarrow{B}^2}{2\mu_0} \right) \boldsymbol{I} - \frac{1}{\mu_0} \overrightarrow{B}\, \overrightarrow{B}^t - \boldsymbol{\tau} \right] + \frac{\overrightarrow{B}}{\mu_0} \overrightarrow{\nabla} \cdot \overrightarrow{B} = \overrightarrow{0}$$

Il s'agit de notre équation de conservation de la quantité de mouvement.

L'évolution du champ magnétique

Les quantités conservées sont celles que l'on retrouve en mécanique des fluides : $(\rho, \rho\overrightarrow{u}, E)$. A contrario, le champ magnétique en lui-même ne se conserve pas. Seul le flux magnétique se conserve, mais ce n'est pas ce que nous cherchons. La seule équation traitant de l'évolution de $\overrightarrow{B}$ est celle de Maxwell-Faraday (2.1.5), dont on décide donc de partir :

$$\frac{\partial \overrightarrow{B}}{\partial t} = -\overrightarrow{\nabla} \wedge \overrightarrow{E}$$

$\overrightarrow{E}$ est lié au champ magnétique par la loi d'Ohm (2.1.7), ce qui nous permet d'écrire :

$$\frac{\partial \overrightarrow{B}}{\partial t} + \overrightarrow{\nabla} \wedge \left(\eta \overrightarrow{j} - \overrightarrow{u} \wedge \overrightarrow{B}\right) = \overrightarrow{0}$$

Nous cherchons une réécriture de ce rotationnel en divergence afin d'imiter la nature conservative des autres équations du système. Sans cela, nous ne pourrions pas appliquer nos algorithmes tels que nous les présenterons dans les chapitres suivants. De plus, numériquement, nous éprouverions certainement de grandes difficultés à être conservatifs, c'est-à-dire à ne pas perdre de l'information à chaque pas de temps.

Commençons par le dernier terme, le plus simple :

$$\begin{aligned}
\overrightarrow{\nabla} \wedge \left(\overrightarrow{u} \wedge \overrightarrow{B}\right) &\overset{(A.5)}{=} \overrightarrow{u}\overrightarrow{\nabla} \cdot \overrightarrow{B} - \left(\overrightarrow{u} \cdot \overrightarrow{\nabla}\right)\overrightarrow{B} - \overrightarrow{B}\overrightarrow{\nabla} \cdot \overrightarrow{u} + \left(\overrightarrow{B} \cdot \overrightarrow{\nabla}\right)\overrightarrow{u} \\
&\overset{(2.1.8)}{=} \overrightarrow{\nabla} \cdot \left(\overrightarrow{u}\overrightarrow{B}^t - \overrightarrow{B}\overrightarrow{u}^t\right)
\end{aligned}$$

Intéressons-nous à présent à la partie résistive en utilisant l'équation de Maxwell-Ampère et un peu d'analyse vectorielle.

$$\begin{aligned}
\overrightarrow{\nabla} \wedge \left(\eta \overrightarrow{j}\right) &\overset{(2.1.6)}{=} \overrightarrow{\nabla} \wedge \left(\frac{\eta}{\mu_0}\overrightarrow{\nabla} \wedge \overrightarrow{B}\right) \\
&\overset{(A.2)}{=} \overrightarrow{\nabla} \wedge \left(\overrightarrow{\nabla} \wedge \left(\frac{\eta}{\mu_0}\overrightarrow{B}\right)\right) - \overrightarrow{\nabla} \wedge \left(\overrightarrow{\nabla}\left(\frac{\eta}{\mu_0}\right) \wedge \overrightarrow{B}\right) \\
&\overset{(A.8),(A.5)}{=} \overrightarrow{\nabla}\left(\overrightarrow{\nabla} \cdot \left(\frac{\eta}{\mu_0}\overrightarrow{B}\right)\right) - \left(\overrightarrow{\nabla} \cdot \overrightarrow{\nabla}\right)\left(\frac{\eta}{\mu_0}\overrightarrow{B}\right) - \overrightarrow{\nabla}\left(\frac{\eta}{\mu_0}\right)\overrightarrow{\nabla} \cdot \overrightarrow{B} \\
&\quad + \left(\overrightarrow{\nabla}\frac{\eta}{\mu_0} \cdot \overrightarrow{\nabla}\right)\overrightarrow{B} + \overrightarrow{B}\Delta\left(\frac{\eta}{\mu_0}\right) - \left(\overrightarrow{B} \cdot \overrightarrow{\nabla}\right)\overrightarrow{\nabla}\frac{\eta}{\mu_0} \\
&\overset{(2.1.8)}{=} \overrightarrow{\nabla} \cdot \left(\overrightarrow{\nabla} \cdot \left(\frac{\eta}{\mu_0}\overrightarrow{B}\right)\boldsymbol{I} - \overrightarrow{\nabla}\left(\frac{\eta}{\mu_0}\overrightarrow{B}\right)\right) - \overrightarrow{\nabla} \cdot \left(\left(\overrightarrow{\nabla}\frac{\eta}{\mu_0}\right)\overrightarrow{B}^t - \overrightarrow{B}\left(\overrightarrow{\nabla}\frac{\eta}{\mu_0}\right)^t\right)
\end{aligned}$$

A ce stade, une simplification est possible en exprimant le gradient d'un vecteur grâce à (2.1.9). Si on se place dans $\mathbb{R}^3$ et qu'on suppose temporairement pour simplifier que $\mu_0 = 1$:

$$\begin{aligned}
\overrightarrow{\nabla}\left(\eta\overrightarrow{B}\right) &= \begin{pmatrix} \partial_x(\eta B_x) & \partial_y(\eta B_x) & \partial_z(\eta B_x) \\ \partial_x(\eta B_y) & \partial_y(\eta B_y) & \partial_z(\eta B_y) \\ \partial_x(\eta B_z) & \partial_y(\eta B_z) & \partial_z(\eta B_z) \end{pmatrix} \\
&= \eta\overrightarrow{\nabla}\overrightarrow{B} + \overrightarrow{B}\left(\overrightarrow{\nabla}\eta\right)^t
\end{aligned}$$

Ceci nous permet d'obtenir finalement :

$$\overrightarrow{\nabla} \wedge \left(\eta \overrightarrow{j}\right) = \overrightarrow{\nabla} \cdot \left[\overrightarrow{\nabla} \cdot \left(\frac{\eta}{\mu_0} \overrightarrow{B}\right) \boldsymbol{I} - \frac{\eta}{\mu_0} \overrightarrow{\nabla} \overrightarrow{B} - \overrightarrow{\nabla} \left(\frac{\eta}{\mu_0}\right) \overrightarrow{B}^t\right]$$

Notre expression finale de l'équation d'évolution du champ magnétique est donc :

$$\frac{\partial \overrightarrow{B}}{\partial t} + \overrightarrow{\nabla} \cdot \left(\overrightarrow{B} \overrightarrow{u}^t - \overrightarrow{u} \overrightarrow{B}^t\right) + \overrightarrow{\nabla} \cdot \left[\overrightarrow{\nabla} \cdot \left(\frac{\eta}{\mu_0} \overrightarrow{B}\right) \boldsymbol{I} - \frac{\eta}{\mu_0} \overrightarrow{\nabla} \overrightarrow{B} - \overrightarrow{\nabla} \left(\frac{\eta}{\mu_0}\right) \overrightarrow{B}^t\right] = \overrightarrow{0}$$

La conservation de l'énergie

L'énergie a longtemps été une quantité abstraite, car nous ne sommes sensibles qu'à ses transformations et car elle nous apparaît de différentes manières. Comme pourrait le dire Lavoisier, l'énergie ne peut être ni créée ni perdue, mais seulement s'échanger et/ou changer de forme. Par conséquent elle se conserve à une échelle globale, ou idéalement si le système étudié peut être isolé de toute interaction avec l'extérieur.

En mécanique des fluides, la formulation fondamentale de cette propriété est le premier principe de la thermodynamique :

$$\Delta E = W + Q$$

E est l'énergie totale du système, W représente le travail effectué par l'ensemble des forces agissant sur le système et Q représente l'ensemble des échanges de chaleur avec l'extérieur ou des sources.

En MHD, l'évolution de l'énergie magnétique n'est pas dictée par le travail des forces ou les échanges de chaleur, mais se déduit de l'équation d'évolution du champ magnétique. Par conséquent, **ce** premier principe reste vrai mais uniquement pour une partie de l'énergie du système, l'énergie mécanique que nous noterons E_m et qui regroupe l'énergie interne et l'énergie cinétique. Ceci s'explique par le fait que les forces de Lorentz magnétiques ne sont pas complètement conservatives : le gradient de pression magnétique est conservatif dans le sens où il dérive clairement de l'énergie magnétique, mais pas la tension magnétique (on rappelle qu'ici les forces de Lorentz fournissent du travail car elles doivent s'interpréter comme des forces de Laplace, c'est-à-dire à l'échelle macroscopique, et non simplement comme des forces agissant à l'échelle microscopique sur les porteurs de charge). L'énergie totale est donc $E = E_m + E_{mag}$.

$$\mathrm{d}E_m = \delta W + \delta Q$$

Remarque 1. *Ce fait ne doit pas être troublant : il est dû au choix de considérer le premier principe dans sa définition "mécaniste". Si on considère que la définition du premier principe tient dans les phrases d'introduction, c'est-à-dire la conservation de l'énergie sous toutes ses formes, la mise en équations est différente et revient nécessairement à ce que nous allons faire maintenant.*

Comme nous l'avons vu en établissant la conservation de la quantité de mouvement, le travail provient de la force de Lorentz et des efforts des contraintes (dont la pression hydrodynamique). Intéressons-nous à nouveau à un volume $V(t)$ de plasma sur lequel nous étudions la variation d'énergie mécanique. Les échanges de chaleur avec les volumes adjacents sont modélisés par de la conduction thermique, sous forme de flux de surface ($\overrightarrow{q_F}$) dont l'expression fut établie par Fourier au début du XIX^e siècle :

$$\overrightarrow{q_F} = -\kappa \overrightarrow{\nabla} T$$

κ représentant la conductivité thermique locale du plasma et T la température. D'autre part, il existe une source volumique de chaleur dûe à la présence de courants électriques en milieu résistif. Les courants étant des déplacements d'électrons, on conçoit aisément que les collisions de ces derniers avec les élements

moins mobiles du plasma créent de l'agitation thermique. C'est ce qu'on appelle le chauffage par effet Joule. La puissance ainsi dissipée est donnée par la première loi de Joule (1841) :

$$P_J = \eta \overrightarrow{j}^2$$

où nous reprenons les notations de la loi d'Ohm (2.1.7). Ces deux lois expriment le même phénomène de collisions, et Maxwell montra qu'elles sont équivalentes.

Nous pouvons donc revenir au premier principe et écrire :

$$\frac{\mathrm{d}}{\mathrm{dt}} \int_{V(t)} E_m dV = \int_{V(t)} \left(\overrightarrow{j} \wedge \overrightarrow{B} \right) \cdot \overrightarrow{u}\, dV + \int_{\partial V(t)} (\boldsymbol{\sigma}\, \overrightarrow{u}) \cdot \overrightarrow{n}\, d\partial V - \int_{\partial V(t)} \overrightarrow{q_F} \cdot \overrightarrow{n}\, d\partial V + \int_{V(t)} \eta \overrightarrow{j}^2 dV$$

Appliquons alors le théorème de Stokes aux intégrales de bord ainsi que (2.1.11) au premier terme, puis remarquons que l'équation obtenue est vraie $\forall V(t) \in \mathbb{R}^3$.

$$\begin{aligned} \int_{V(t)} \frac{\partial}{\partial t} E_m + \overrightarrow{\nabla} \cdot (E_m \overrightarrow{u}) &= \int_{V(t)} \left[\left(\overrightarrow{j} \wedge \overrightarrow{B} \right) \cdot \overrightarrow{u} + \overrightarrow{\nabla} \cdot (\boldsymbol{\sigma}\, \overrightarrow{u}) - \overrightarrow{\nabla} \cdot \overrightarrow{q_F} + \eta \overrightarrow{j}^2 \right] dV \\ \Rightarrow \qquad \frac{\partial}{\partial t} E_m + \overrightarrow{\nabla} \cdot (E_m \overrightarrow{u}) &= \left(\overrightarrow{j} \wedge \overrightarrow{B} \right) \cdot \overrightarrow{u} + \overrightarrow{\nabla} \cdot (\boldsymbol{\sigma}\, \overrightarrow{u}) + \overrightarrow{\nabla} \cdot \left(\kappa \overrightarrow{\nabla} T \right) + \eta \overrightarrow{j}^2 \end{aligned}$$

Nous devons maintenant réexprimer certains termes sous la forme d'une divergence. Commençons par le travail des forces de Lorentz. D'après l'équation de Maxwell-Ampère (2.1.6) et les règles du produit mixte (A.11) on a :

$$\begin{aligned} \left(\overrightarrow{j} \wedge \overrightarrow{B} \right) \cdot \overrightarrow{u} &= \overrightarrow{j} \cdot \left(\overrightarrow{B} \wedge \overrightarrow{u} \right) \\ &= \frac{1}{\mu_0} \left(\overrightarrow{\nabla} \wedge \overrightarrow{B} \right) \cdot \left(\overrightarrow{B} \wedge \overrightarrow{u} \right) \end{aligned}$$

Ayons ensuite recours à quelques manipulations d'analyse vectorielle :

$$\begin{aligned} \mu_0 \left(\overrightarrow{j} \wedge \overrightarrow{B} \right) \cdot \overrightarrow{u} &\overset{(A.4)}{=} -\overrightarrow{\nabla} \cdot \left(\left(\overrightarrow{B} \wedge \overrightarrow{u} \right) \wedge \overrightarrow{B} \right) - \overrightarrow{B} \cdot \left(\overrightarrow{\nabla} \wedge \left(\overrightarrow{u} \wedge \overrightarrow{B} \right) \right) \\ &\overset{(A.5)}{=} \overrightarrow{\nabla} \cdot \left(\left(\overrightarrow{u} \wedge \overrightarrow{B} \right) \wedge \overrightarrow{B} \right) - \overrightarrow{B} \cdot \left(\overrightarrow{\nabla} \cdot \left(\overrightarrow{u}\, \overrightarrow{B}^t - \overrightarrow{B}\, \overrightarrow{u}^t \right) \right) \\ &\overset{(A.10)}{=} \overrightarrow{\nabla} \cdot \left(\left(\overrightarrow{u} \cdot \overrightarrow{B} \right) \overrightarrow{B} - \overrightarrow{B}^2 \overrightarrow{u} \right) - \overrightarrow{B} \cdot \left(\overrightarrow{\nabla} \cdot \left(\overrightarrow{u}\, \overrightarrow{B}^t - \overrightarrow{B}\, \overrightarrow{u}^t \right) \right) \end{aligned}$$

Passons à présent au travail des contraintes. Parmi elles, on a énuméré dans la section concernant la conservation de la quantité de mouvement la pression hydrodynamique p et le tenseur des contraintes visqueuses $\boldsymbol{\tau}$. On a donc sans surprise :

$$\overrightarrow{\nabla} \cdot (\boldsymbol{\sigma}\, \overrightarrow{u}) = -\overrightarrow{\nabla} \cdot (p \overrightarrow{u}) + \overrightarrow{\nabla} \cdot (\boldsymbol{\tau}\, \overrightarrow{u})$$

Nous laissons pour le moment le terme de conduction thermique qui est déjà sous forme de divergence. Il ne reste donc plus que le terme de dissipation par effet Joule à traiter. En se reportant de nouveau à l'équation de Maxwell-Ampère simplifiée (2.1.6), on peut écrire :

$$\eta \overrightarrow{j}^2 = \frac{\eta}{\mu_0^2} \left(\overrightarrow{\nabla} \wedge \overrightarrow{B} \right)^2$$

On va maintenant reformuler ce terme de manière à l'intégrer au mieux dans l'équation finale.

$$\begin{aligned}
\mu_0^2 \eta \overrightarrow{j}^2 &\overset{(A.2)}{=} \left(\overrightarrow{\nabla} \wedge \overrightarrow{B}\right) \cdot \left[\overrightarrow{\nabla} \wedge \left(\eta \overrightarrow{B}\right) - \left(\overrightarrow{\nabla}\eta\right) \wedge \overrightarrow{B}\right] \\
&\overset{(A.4)}{=} \overrightarrow{B} \cdot \left[\overrightarrow{\nabla} \wedge \left(\overrightarrow{\nabla} \wedge \left(\eta \overrightarrow{B}\right) - \left(\overrightarrow{\nabla}\eta\right) \wedge \overrightarrow{B}\right)\right] \\
&\quad -\overrightarrow{\nabla} \cdot \left[\left(\overrightarrow{\nabla} \wedge \left(\eta \overrightarrow{B}\right)\right) \wedge \overrightarrow{B} - \left(\overrightarrow{\nabla}\eta \wedge \overrightarrow{B}\right) \wedge \overrightarrow{B}\right]
\end{aligned}$$

Intéressons-nous plus particulièrement au dernier terme. Il est de la forme $\mathrm{div}\left(\overrightarrow{T}\right)$, où $\overrightarrow{T}$ est donné par :

$$\begin{aligned}
\overrightarrow{T} &\overset{(A.9)}{=} -\overrightarrow{B} \wedge \left(\overrightarrow{\nabla} \wedge \left(\eta \overrightarrow{B}\right)\right) + \overrightarrow{B} \wedge \left(\overrightarrow{\nabla}\eta \wedge \overrightarrow{B}\right) \\
&\overset{(A.4),(A.10)}{=} -\overrightarrow{\nabla}\left(\eta \overrightarrow{B}^2\right) + \left(\eta \overrightarrow{B} \cdot \overrightarrow{\nabla}\right) \overrightarrow{B} + \eta \overrightarrow{B} \wedge \left(\overrightarrow{\nabla} \wedge \overrightarrow{B}\right) + \left(\overrightarrow{B} \cdot \overrightarrow{\nabla}\right)\left(\eta \overrightarrow{B}\right) \\
&\quad + \overrightarrow{B}^2 \overrightarrow{\nabla}\eta - \left(\overrightarrow{B} \cdot \overrightarrow{\nabla}\eta\right) \overrightarrow{B} \\
&\overset{(A.4)}{=} -\eta \overrightarrow{\nabla} \overrightarrow{B}^2 - \overrightarrow{B}^2 \overrightarrow{\nabla}\eta + \overrightarrow{B}^2 \overrightarrow{\nabla}\eta + 2\eta \left(\overrightarrow{B} \cdot \overrightarrow{\nabla}\right) \overrightarrow{B} + \overrightarrow{B}\left(\overrightarrow{B} \cdot \overrightarrow{\nabla}\eta\right) - \overrightarrow{B}\left(\overrightarrow{B} \cdot \overrightarrow{\nabla}\eta\right) \\
&\quad + \eta \left(\overrightarrow{\nabla}\left(\frac{\overrightarrow{B}^2}{2}\right) - \left(\overrightarrow{B} \cdot \overrightarrow{\nabla}\right) \overrightarrow{B}\right) \\
&= -\eta \left[\overrightarrow{\nabla}\left(\frac{\overrightarrow{B}^2}{2}\right) - \left(\overrightarrow{B} \cdot \overrightarrow{\nabla}\right) \overrightarrow{B}\right] \\
&\overset{(2.1.8)}{=} -\eta \left[\overrightarrow{\nabla} \cdot \left(\frac{\overrightarrow{B}^2}{2} \boldsymbol{I} - \overrightarrow{B}\overrightarrow{B}^t\right) + \overrightarrow{B}\overrightarrow{\nabla} \cdot \overrightarrow{B}\right]
\end{aligned}$$

Si nous utilisons de plus l'équation (2.1.3), nous arrivons finalement à la formulation suivante :

$$\begin{aligned}
\eta \overrightarrow{j}^2 &= \frac{\overrightarrow{B}}{\mu_0^2} \cdot \left[\overrightarrow{\nabla} \wedge \left(\overrightarrow{\nabla} \wedge \left(\eta \overrightarrow{B}\right) - \overrightarrow{\nabla}\eta \wedge \overrightarrow{B}\right)\right] + \overrightarrow{\nabla} \cdot \left[\frac{\eta}{\mu_0^2} \overrightarrow{\nabla} \cdot \left(\frac{\overrightarrow{B}^2}{2} \boldsymbol{I} - \overrightarrow{B}\overrightarrow{B}^t\right)\right] \\
&\overset{(A.2)}{=} \frac{\overrightarrow{B}}{\mu_0^2} \cdot \left(\overrightarrow{\nabla} \wedge \left(\eta \overrightarrow{\nabla} \wedge \overrightarrow{B}\right)\right) + \overrightarrow{\nabla} \cdot \left[\frac{\eta}{\mu_0^2} \overrightarrow{\nabla} \cdot \left(\frac{\overrightarrow{B}^2}{2} \boldsymbol{I} - \overrightarrow{B}\overrightarrow{B}^t\right)\right]
\end{aligned}$$

Nous laissons volontairement certains termes sous la forme $\overrightarrow{B} \cdot \left(\overrightarrow{\nabla} \dots\right)$ car nous reconnaissons des termes de l'équation d'évolution du champ magnétique. En effet, nous voulons à présent compléter l'équation issue du premier principe en rajoutant l'évolution de l'énergie magnétique, qui s'exprime de façon identique à la pression magnétique, c'est-à-dire :

$$E_{mag} = \frac{\overrightarrow{B}^2}{2\mu_0}$$

On remarque assez rapidement que l'évolution de cette quantité peut être obtenue en multipliant l'équation portant sur $\overrightarrow{B}$ par $\dfrac{\overrightarrow{B}}{\mu_0}$:

$$\frac{\overrightarrow{B}}{\mu_0} \cdot \partial_t \overrightarrow{B} + \frac{\overrightarrow{B}}{\mu_0} \cdot \left(\overrightarrow{\nabla} \cdot \left(\overrightarrow{B}\overrightarrow{u}^t - \overrightarrow{u}\overrightarrow{B}^t\right)\right) + \frac{\overrightarrow{B}}{\mu_0^2} \cdot \left(\overrightarrow{\nabla} \wedge \left(\eta \overrightarrow{\nabla} \wedge \overrightarrow{B}\right)\right) = 0$$

Notons que bien que nous supposerons des paramètres physiques constants plus tard, tout ce que nous faisons ici ne nécessite à aucun moment de considérer la résistivité η constante. Maintenant si nous regroupons tous les termes dont nous disposons, on peut obtenir l'évolution de l'énergie totale $E = E_m + E_{mag}$:

$$\begin{aligned}\frac{\partial E_m}{\partial t} + \overrightarrow{\nabla}\cdot(E_m\overrightarrow{u}) + \frac{\partial}{\partial t}\left(\frac{\overrightarrow{B}^2}{2\mu_0}\right) = \frac{1}{\mu_0}\Big[\overrightarrow{\nabla}\cdot\Big((\overrightarrow{u}\cdot\overrightarrow{B})\overrightarrow{B} - \overrightarrow{B}^2\overrightarrow{u}\Big) + \cancel{\overrightarrow{B}\cdot\Big(\overrightarrow{\nabla}\cdot\Big(\overrightarrow{B}\overrightarrow{u}^t - \overrightarrow{u}\overrightarrow{B}^t\Big)\Big)}\Big] \\ - \overrightarrow{\nabla}\cdot(p\overrightarrow{u}) + \overrightarrow{\nabla}\cdot(\boldsymbol{\tau}\,\overrightarrow{u}) + \overrightarrow{\nabla}\cdot\Big(\kappa\overrightarrow{\nabla}T\Big) \\ + \overrightarrow{\nabla}\cdot\left[\frac{\eta}{\mu_0^2}\overrightarrow{\nabla}\cdot\left(\frac{\overrightarrow{B}^2}{2}\boldsymbol{I} - \overrightarrow{B}\overrightarrow{B}^t\right)\right] + \cancel{\frac{\overrightarrow{B}}{\mu_0^2}\cdot\Big(\overrightarrow{\nabla}\wedge\Big(\eta\overrightarrow{\nabla}\wedge\overrightarrow{B}\Big)\Big)} \\ \cancel{-\frac{\overrightarrow{B}}{\mu_0}\cdot\Big(\overrightarrow{\nabla}\cdot\Big(\overrightarrow{B}\overrightarrow{u}^t - \overrightarrow{u}\overrightarrow{B}^t\Big)\Big)} - \cancel{\frac{\overrightarrow{B}}{\mu_0^2}\cdot\Big(\overrightarrow{\nabla}\wedge\Big(\eta\overrightarrow{\nabla}\wedge\overrightarrow{B}\Big)\Big)}\end{aligned}$$

On obtient alors la forme finale de notre équation de conservation de l'énergie :

$$\begin{aligned}\frac{\partial E}{\partial t} + \overrightarrow{\nabla}\cdot(E_m\overrightarrow{u}) + 2\overrightarrow{\nabla}\cdot\left(\frac{\overrightarrow{B}^2}{2\mu_0}\overrightarrow{u}\right) + \overrightarrow{\nabla}\cdot(p\overrightarrow{u}) - \frac{1}{\mu_0}\overrightarrow{\nabla}\cdot\Big((\overrightarrow{u}\cdot\overrightarrow{B})\overrightarrow{B}\Big) \\ -\overrightarrow{\nabla}\cdot(\boldsymbol{\tau}\,\overrightarrow{u}) - \overrightarrow{\nabla}\cdot\Big(\kappa\overrightarrow{\nabla}T\Big) - \overrightarrow{\nabla}\cdot\left[\frac{\eta}{\mu_0^2}\overrightarrow{\nabla}\cdot\left(\frac{\overrightarrow{B}^2}{2}\boldsymbol{I} - \overrightarrow{B}\overrightarrow{B}^t\right)\right] = 0\end{aligned}$$

$$\Rightarrow \quad \frac{\partial E}{\partial t} + \overrightarrow{\nabla}\cdot\left[\left(E + p + \frac{\overrightarrow{B}^2}{2\mu_0}\right)\overrightarrow{u} - \frac{\overrightarrow{B}}{\mu_0}\Big(\overrightarrow{u}\cdot\overrightarrow{B}\Big)\right] + \overrightarrow{\nabla}\cdot\left[\frac{\eta}{\mu_0^2}\overrightarrow{\nabla}\cdot\left(\overrightarrow{B}\overrightarrow{B}^t - \frac{\overrightarrow{B}^2}{2}\boldsymbol{I}\right) - \boldsymbol{\tau}\,\overrightarrow{u} - \kappa\overrightarrow{\nabla}T\right] = 0$$

Synthèse

Nous avons à présent établi toutes les équations du système MHD classique, que nous allons rappeler ici de façon synthétique pour les références ultérieures. On peut d'ores et déjà remarquer que dans les équations de conservation ainsi que dans la loi de Maxwell-Faraday, tous les termes faisant intervenir le champ magnétique sont proportionnels à $\dfrac{\overrightarrow{B}}{\sqrt{\mu_0}}$ à l'exception de ceux issus des effets résistifs. Ceci nous incite à définir une nouvelle variable de champ magnétique :

$$\overrightarrow{B}^1 := \frac{\overrightarrow{B}}{\sqrt{\mu_0}}$$

A partir de maintenant, nous ne travaillerons plus qu'avec cette variable et non avec l'originale, sauf mention contraire. Le système s'écrit donc finalement :

$$\frac{\partial \rho}{\partial t} + \overrightarrow{\nabla}\cdot\Big(\rho\overrightarrow{u} - D\overrightarrow{\nabla}\rho\Big) = 0 \tag{2.1.12a}$$

$$\frac{\partial(\rho\overrightarrow{u})}{\partial t} + \overrightarrow{\nabla}\cdot\left[\rho\overrightarrow{u}\,\overrightarrow{u}^t + \left(p + \frac{\overrightarrow{B}^2}{2}\right)Id - \overrightarrow{B}\overrightarrow{B}^t - \boldsymbol{\tau}\right] = \overrightarrow{0} \tag{2.1.12b}$$

$$\frac{\partial E}{\partial t} + \overrightarrow{\nabla} \cdot \left[\left(E + p + \frac{\overrightarrow{B}^2}{2} \right) \overrightarrow{u} - \left(\overrightarrow{u} \cdot \overrightarrow{B} \right) \overrightarrow{B} + \frac{\eta}{\mu_0} \left(\overrightarrow{\nabla} \cdot \left(\overrightarrow{B} \overrightarrow{B}^t \right) - \overrightarrow{\nabla} \left(\frac{\overrightarrow{B}^2}{2} \right) \right) - \boldsymbol{\tau} \overrightarrow{u} - \kappa \overrightarrow{\nabla} T \right] = 0 \tag{2.1.12c}$$

$$\frac{\partial \overrightarrow{B}}{\partial t} + \overrightarrow{\nabla} \cdot \left[\overrightarrow{B} \overrightarrow{u}^t - \overrightarrow{u} \overrightarrow{B}^t + \overrightarrow{\nabla} \cdot \left(\frac{\eta}{\mu_0} \overrightarrow{B} \right) \boldsymbol{I} - \frac{\eta}{\mu_0} \overrightarrow{\nabla} \overrightarrow{B} - \overrightarrow{\nabla} \left(\frac{\eta}{\mu_0} \right) \overrightarrow{B}^t \right] = \overrightarrow{0} \tag{2.1.12d}$$

$$\overrightarrow{\nabla} \cdot \overrightarrow{B} = 0 \tag{2.1.12e}$$

Nous appellerons ce système celui de la MHD résistive (2.1.12). Ceci servira à le distinguer de celui dit de la MHD idéale, qui est une approximation semblable à celle qui permet de passer des équations de Navier-Stokes à celles d'Euler en mécanique des fluides. Autrement dit, on suppose que tous les phénomènes diffusifs, qui se traduisent par des dérivées secondes en espace, sont régis par des paramètres qui s'annulent. Ceci concerne donc la conductivité thermique κ, la résistivité électrique η, le coefficient de diffusion de masse D et la viscosité dynamique μ, qui sont supposés tous nuls. Obtenues par ces simplifications, les équations qui forment le système de la MHD idéale (2.1.13) s'écrivent :

$$\frac{\partial \rho}{\partial t} + \overrightarrow{\nabla} \cdot (\rho \overrightarrow{u}) = 0 \tag{2.1.13a}$$

$$\frac{\partial (\rho \overrightarrow{u})}{\partial t} + \overrightarrow{\nabla} \cdot \left[\rho \overrightarrow{u} \overrightarrow{u}^t + (p + \frac{\overrightarrow{B}^2}{2}) Id - \overrightarrow{B} \overrightarrow{B}^t \right] = \overrightarrow{0} \tag{2.1.13b}$$

$$\frac{\partial E}{\partial t} + \overrightarrow{\nabla} \cdot \left[\left(E + p + \frac{\overrightarrow{B}^2}{2} \right) \overrightarrow{u} - (\overrightarrow{u} \cdot \overrightarrow{B}) \overrightarrow{B} \right] = 0 \tag{2.1.13c}$$

$$\frac{\partial \overrightarrow{B}}{\partial t} + \overrightarrow{\nabla} \cdot \left(\overrightarrow{B} \overrightarrow{u}^t - \overrightarrow{u} \overrightarrow{B}^t \right) = \overrightarrow{0} \tag{2.1.13d}$$

$$\overrightarrow{\nabla} \cdot \overrightarrow{B} = 0 \tag{2.1.13e}$$

2.1.3 Aspects thermodynamiques

Le système d'équations n'est pas encore fermé. Nous avons obtenu 8 équations pour 9 variables, la dernière étant la pression p. Il nous faut donc trouver une relation entre p et les variables conservatives U, ce que nous appelons une équation d'état. Cette partie est du ressort de la thermodynamique.

La thermodynamique classique étudie l'équilibre des systèmes macroscopiques. On parle d'équilibre thermodynamique global lorsque tout le système est dans un état stationnaire et uniforme, ce qui est un cas particulier peu intéressant pour nous, mais aussi d'équilibre thermodynamique local lorsque les échelles de temps et d'espace des phénomènes microscopiques transitoires (qui empêchent de considérer un équilibre local) sont petites devant les échelles macroscopiques. Or ceci est justement l'hypothèse que nous faisons lorsque nous adoptons une description fluide du plasma, c'est-à-dire lorsqu'on décrit le

problème en utilisant le modèle de la Magnétohydrodynamique. En particulier, le modèle suppose que le plasma est et demeure totalement ionisé, ce qui permet d'ignorer les phénomènes de recombinaison entre ions et électrons. Pour ce qui est de l'échelle spatiale, on peut se convaincre simplement (sans s'attarder à examiner divers ordres de grandeur) que si le plasma est suffisamment dense pour être vu comme un milieu continu, alors le fait de faire des bilans sur de petits volumes met en jeu des échelles spatiales a priori bien supérieures à celle, par exemple, du libre parcours moyen de chaque particule.

Il est entendu au demeurant que bien des phénomènes peuvent se produire au niveau microscopique et invalider cette hypothèse. L'apparition de modes turbulents à petite échelle peut avoir des effets très nets au niveau macroscopique, mais on n'en rend pas compte naturellement dans les équations de la MHD. Il y a deux remarques à bien prendre en compte à ce sujet, qui font que ce problème n'est pas forcément un obstacle infranchissable a priori. Premièrement, lorsqu'on a établi notre système de lois de conservation, le rappel a été fait que considérer un milieu comme un continuum constitue déjà une première approximation très importante. Cette dernière implique directement que les évènements microscopiques seront ignorés, ce qui revient à supposer qu'ils n'ont pas lieu (et donc restreint le nombre d'applications physiques potentielles du modèle). Sur cette question, nous avons dit que la physique des plasmas de tokamaks avait été depuis longtemps jugée modélisable par les équations de la MHD, ce qui autorise notre approche et celles de nombreux autres numériciens avant nous. C'est également le cas pour certains plasmas d'étude en astrophysique. Deuxièmement, dans les cas où les évènements miscrocopiques déstabilisants ne peuvent plus être ignorés à l'échelle macroscopique, il peut rester la possibilité de les modéliser au niveau des équations. Ceci nécessite toutefois une étude approfondie des phénomènes mis en cause à l'échelle de la théorie cinétique et rien ne garantit de parvenir à trouver une modélisation satisfaisante. Un exemple de recours à cette solution est le terme que nous avons introduit dans l'équation de conservation de la masse.

L'équation d'état

Nous considérerons par conséquent à chaque instant que nous avons un équilibre thermodynamique local qui autorise une certaine description du système. Ce qu'on entend par là est que cet état d'équilibre permet la définition de quantités intensives qui caractérisent le système, comme la température T et la pression p dans notre cas. Mais une description complète de l'état de ce système requiert des informations supplémentaires, extensives, qui peuvent être par exemple la quantité de matière n et le volume V. Cette description est celle de la mécanique des fluides, mais les phénomènes cinétiques qui ont lieu dans un plasma dense comme le nôtre sont assimilables à ce qui se passe dans un gaz. Les quantités qui suffisent à caractériser un système sont appelées des variables d'état, les quatre que nous avons citées étant les plus généralement utilisées. Cependant d'autres choix sont possibles. Une relation entre ces variables permet de fermer le système de lois de conservation, c'est l'équation d'état.

Il existe de nombreuses formules construites à partir de la théorie cinétique des gaz, qui sont de plus ou moins bonnes approximations du comportement réel du gaz. La plus simple et la plus connue est l'équation des gaz parfaits. On peut également citer les modèles de Van der Waals, de Dieterici, du Viriel et bien d'autres, dont certaines assez récemment élaborées. Cela fait partie d'un domaine très actif encore aujourd'hui en mécanique des fluides. En ce qui concerne le comportement des plasmas, on considère généralement que le modèle des gaz parfaits est suffisamment satisfaisant [44]. En particulier, une analyse d'ordres de grandeurs semble permettre de confirmer la faible corrélation des particules dans le cadre des régimes qui nous intéressent et pour des pressions qui ne sont pas excessives.

On n'utilisera donc que cette loi dans nos travaux, bien qu'il soit envisageable du point de vue numérique et mathématique de faire d'autres choix si cela s'avérait plus tard nécessaire. Bien que cette loi soit très largement connue aujourd'hui, par souci d'être complet, autorisons-nous un bref rappel à

son sujet. Un gaz parfait est, comme son nom l'indique, la plus simple description possible d'un gaz. On suppose que les particules sont des points de masse parfaitement élastiques qui n'intéragissent pas entre eux, si ce n'est lors de collisions, qui sont donc élastiques elles aussi. Le fait de négliger le volume propre des particules revient à considérer que la distance moyenne parcourue par chaque particule est bien supérieure à celui-ci. Ce qui est d'autant plus vrai pour de faibles densités et de hautes températures, cas où la loi des gaz parfaits est la plus appropriée.

Considérons un volume V de plasma contenant N particules et dont la température absolue mesurable est T. La théorie cinétique des gaz, notamment de par sa définition de la température via l'introduction de la constante universelle de Boltzmann k_B, permet d'établir ([25]) la relation entre la pression p et T :

$$pV = Nk_BT$$

Lorsqu'on a seulement l'équilibre local, ce qui est toujours le cas pour nous, la même équation est valable en tout point, chaque paramètre devenant une variable d'espace (l'obtention de cette équation étant valable pour des volumes infinitésimaux "entourant" chaque point). Ceci signifie qu'on doit remplacer la quantité $\frac{N}{V}$ par la quantité locale n qui est la densité de particules (en m^{-3}). Soit m la masse d'une "particule", alors on peut écrire :

$$p = \rho\frac{k_B}{m}T$$

Or on peut relier la constante de Boltzmann à la constante spécifique (massique) du gaz parfait : $r_s = \frac{k_B}{m}$. Ce qui nous donne finalement :

$$p = \rho r_s T$$

Notre but est de trouver une relation entre p (en Pa) et l'énergie interne $\rho\epsilon$ (en $J.m^{-3}$, ρ étant la densité et ϵ la densité massique d'énergie interne). L'énergie totale E (une des variables conservatives) est ensuite donnée par $E = \rho\epsilon + \frac{1}{2}\rho\vec{u}^2 + \frac{1}{2\mu_0}\vec{B}^2$. Il suffit ensuite de connaître l'équation d'état citée ci-dessus pour retrouver T (en vue du post-traitement uniquement). Le plus évident est de revenir au premier principe de la thermodynamique dont nous nous sommes déjà servis pour établir la conservation de l'énergie mécanique. Laissons un moment de côté l'effet des forces de Lorentz et regardons ce qui se passe dans le volume V. En assimilant le fluide à un gaz parfait, on peut montrer ([25]) que :

$$d\epsilon = c_v dT$$
$$dh = c_p dT$$

où h est la densité massique d'enthalpie, c_v est la capacité thermique massique à volume constant et c_p la capacité thermique massique à pression constante. Toutes ces quantités dépendent *a priori* de la température. On définit le rapport des deux capacités $\gamma = \frac{c_p}{c_v}$. Grâce à la connaissance de l'équation d'état et de la définition de l'enthalpie massique $h = \epsilon + \frac{p}{\rho}$, on a alors la possibilité d'écrire :

$$\begin{aligned}
& dh && = c_p dT \\
\Rightarrow\quad & d(\epsilon + \frac{p}{\rho}) && = c_p dT \\
\Rightarrow\quad & d\left(\frac{p}{\rho}\right) && = c_p dT - d\epsilon \\
\Rightarrow\quad & d\left(\frac{p}{\rho}\right) && = c_p \frac{d\epsilon}{c_v} - d\epsilon \\
\Rightarrow\quad & d\left(\frac{p}{\rho}\right) && = (\gamma - 1)d\epsilon
\end{aligned}$$

Notre relation entre la pression et l'énergie interne est donc seulement différentielle ! Si γ est supposé constant et qu'on intègre entre un instant de référence et la date courante, on obtient :

$$\frac{p}{\rho} = \frac{p_0}{\rho_0} + (\gamma - 1)(\epsilon - \epsilon_0)$$

Autrement dit, la connaissance de la pression et de la densité à l'état intial ne suffit pas à caractériser complètement le système. Si on ne se donne pas aussi l'énergie interne initiale, on ne pourra pas la calculer à un instant ultérieur quelconque. Ou sinon, y a-t-il un moyen de déterminer l'énergie interne initiale ? Elle dépend de la température (énergie cinétique microscopique) et des énergies de liaison. On néglige généralement les énergies de liaison puisqu'on ne traite pas les réactions chimiques et que donc ces énergies ne sont pas modifiées. Dans un plasma de protons et d'électrons, cette approximation est acceptable tant qu'on reste en-deçà du régime de fusion. Reste donc à considérer la part exprimée par la température. C'est à ce moment qu'on peut faire intervenir le troisième principe de la thermodynamique, qui exprime que l'entropie d'un système à la température absolue nulle est nulle. On s'autorise alors une représentation de la matière sous une forme parfaitement ordonnée, c'est-à-dire qu'on suppose que dans de telles conditions le milieu adopte une structure cristalline. La traduction de cet état est que la partie thermique de l'énergie interne est nulle. Si on omet toute autre forme d'énergie que celles déjà citées, par exemple les énergies de liaison, le potentiel gravitationnel ou encore la simple énergie de masse, parce qu'elles sont supposées ne pas être modifiées - ou de manière négligeable pour la gravité - lors des problèmes étudiés, alors on peut se permettre de travailler avec $\epsilon(T = 0) = 0$. D'ailleurs, la loi des gaz parfaits n'est de toute façon pas prévue pour rendre compte de telles situations. De la même manière, si toutes les particules avancent telles un bloc dans une même direction sans se permettre d'écarts, la pression exercée aux bords du domaine lié au fluide sera nulle (ou remarquons que l'équation d'état donne directement $p(T = 0) = 0$). On en conclut donc que pour cet état particulier, $\epsilon_0 = T_0 = p_0 = 0$. En le prenant comme état de référence, on peut alors écrire :

$$p = (\gamma - 1)\rho\epsilon$$

Tout ceci n'est vrai que si nous supposons que γ est une constante, ce qui revient à supposer que les capacités c_v et c_p ne dépendent pas de T. On parle alors de gaz parfaits de Laplace, et on a de plus :

$$\epsilon = c_v T$$

L'entropie

La physique statistique définit l'entropie d'un système comme une mesure du désordre qui y règne au niveau microscopique. Le principe fondamental qui régit l'évolution de cette quantité assez abstraite de prime abord est le second principe de la thermodynamique :

$$dS \geq \frac{\delta Q}{T}$$

S représente l'entropie et δQ l'échange de chaleur lors d'une transformation infinitésimale et T la température à laquelle elle s'effectue. L'inégalité s'explique par l'introduction de la notion de réversibilité. Une transformation est dite réversible si elle se présente comme une succession d'états d'équilibres infiniment proches effectuée de façon continue. Une telle transformation est dite quasistatique et se caractérise par le fait que la transformation inverse (comme si on inversait la flèche du temps) soit possible. Or ceci n'est jamais le cas et constitue donc un modèle idéal de transformation. Toutes les transformations réelles sont irréversibles. Dans notre cas, les sources d'irréversibilité peuvent être les phénomènes de diffusion et de dissipation que l'on trouve dans les équations de la MHD résistive et les transformations non-linéaires

brutales telles que les chocs. Par exemple, l'écoulement de chaleur d'une région chaude à une région froide, qui traduit la propriété fondamentale que les phénomènes physiques tendent toujours à établir un équilibre, n'aurait pas de sens si elle était jouée à l'envers : cela signifierait que la Nature recherche le déséquilibre ! Cependant, si on supprime toutes les sources d'irréversibilté, la variation d'entropie est seulement liée aux échanges de chaleur et l'inégalité précédente devient égalité. C'est le cas si l'écoulement est supposé idéal, comme modélisé par les équations d'Euler ou de la MHD idéale, en l'absence de phénomènes violents comme les chocs. Le principe peut se reformuler :

$$dS = \frac{\delta Q}{T} + dS_{\text{créé}}$$

où $dS_{\text{créé}}$ représente la part d'entropie créée par les processus irréversibles et est toujours positive. On peut donc néanmoins envisager deux types de transformations isentropiques :

. un modèle idéal réversible comme cité précédemment et dont les échanges avec l'extérieur sont nuls (système isolé ou pseudo-isolé et pas de source de chaleur),

. un modèle réalisable où la création d'entropie d'origine irréversible est contrebalancée par une cession de chaleur au milieu extérieur.

En mécanique des milieux continus, l'entropie devient une quantité définissable localement comme toutes les autres grandeurs. On peut alors reformuler le second principe en termes d'entropie volumique ρs (s étant la densité massique d'entropie). Ne sachant pas donner d'expression pour la création d'entropie d'origine irréversible, on ne considère souvent le second principe que sous la forme d'inégalité. Si on considère un volume $V(t)$ de plasma que l'on suit au cours du temps, on peut comptabiliser des échanges de chaleur avec l'extérieur dûs à la conduction thermique et une source volumique dûe à l'effet Joule. Si on admet qu'il y a de plus une diffusion de masse, il faut prendre en compte un échange convectif supplémentaire. Schématiquement, on a :

$$\frac{\mathrm{d}}{\mathrm{dt}} \int_{V(t)} \rho s dV \quad \geq \quad - \int_{\partial V(t)} (\overrightarrow{f} + \overrightarrow{g}) \cdot \overrightarrow{n} d\partial V + \int_{V(t)} J dV$$

où $\overrightarrow{f}$, $\overrightarrow{g}$ et J représentent, respectivement, le flux (proportionnel à D) d'entropie entrant avec un échange de matière, celui entrant par conduction thermique et donc par échange d'énergie cinétique microscopique (proportionnel à κ), et la source (proportionnelle à η) d'entropie créée par effet Joule. Cette version continue du second principe est appelée l'inégalité de Clausius-Duhem. Notons que si le modèle que nous nous attachons à résoudre est celui de la MHD idéale, puisque les phénomènes tels que la diffusion de matière ou la conduction de chaleur s'évanouissent, l'inégalité se simplifie grandement et la raison pour laquelle l'égalité n'est pas atteinte tient seulement à la présence potentielle de chocs dans l'écoulement. On se retrouve alors avec une version continue idéale du second principe :

$$\int_{V(t)} \left(\frac{\partial(\rho s)}{\partial t} + \overrightarrow{\nabla} \cdot (\rho \overrightarrow{u} s) \right) dV \geq 0$$

Autrement dit, ceci étant vrai pour tout volume $V(t)$:

$$\frac{\partial(\rho s)}{\partial t} + \overrightarrow{\nabla} \cdot (\rho \overrightarrow{u} s) \geq 0 \tag{2.1.14}$$

Enfin, pour être complets, un gaz parfait de Laplace (ou tout fluide polytrope) possède une entropie massique qui vérifie

$$ds = -\frac{\gamma}{\rho} d\rho + \frac{dp}{p}$$

et qui s'exprime donc, en intégrant :

$$s = \ln \left(\frac{p}{\rho^\gamma} \right) + C$$

où C est une constante que l'on prendra nulle (nous ne nous intéressons qu'aux variations de l'entropie).

2.2 Prise en compte de la contrainte de Maxwell-Flux

2.2.1 Quelle problématique ?

Parmi les équations de Maxwell, il en est une que nous ne prenons pas directement en compte dans la simulation (hormis celle de Maxwell-Gauss qui ne nous est plus utile) : l'équation de Maxwell-Flux. Il ne s'agit pas d'une équation d'évolution mais d'une contrainte à respecter en permanence pour avoir une solution physiquement pertinente (le champ magnétique doit théoriquement rester strictement solénoïdal).

Du point de vue continu, on peut facilement montrer que cette propriété du champ magnétique est vérifiée naturellement par les équations de la MHD. Il suffit d'estimer l'évolution en temps de la divergence du champ magnétique à partir de l'équation d'évolution de ce même champ. Même si on ne suppose pas a priori que $\overrightarrow{\nabla} \cdot \overrightarrow{B} = 0$, on a :

$$\begin{aligned} \frac{\partial}{\partial t} \overrightarrow{\nabla} \cdot \overrightarrow{B} &= \overrightarrow{\nabla} \cdot \frac{\partial \overrightarrow{B}}{\partial t} \\ &= -\overrightarrow{\nabla} \cdot \left[\overrightarrow{\nabla} \wedge \left(\eta \overrightarrow{j} - \overrightarrow{u} \wedge \overrightarrow{B} \right) \right] \\ &= 0 \end{aligned}$$

car $\overrightarrow{\nabla} \cdot \left(\overrightarrow{\nabla} \wedge \cdot \right)$ est toujours nul (A.7). Ainsi, une solution analytique est assurée de conserver $\overrightarrow{\nabla} \cdot \overrightarrow{B}$ et il suffit donc de partir d'une solution initiale dont le champ magnétique soit strictement solénoïdal.

Cependant, à partir du moment où on discrétise les équations et où on les résoud de manière approchée, rien ne garantit a priori que les "petites" erreurs de discrétisation ne résultent pas en un léger écart à la nullité de la divergence du champ magnétique. En pratique, il est bien évident que c'est au contraire toujours le cas. Or qu'il s'agisse de résoudre un problème stationnaire ou instationnaire, l'algorithme est toujours itératif, ce qui signifie que les erreurs auront le mauvais goût de s'accumuler au fil du temps. De plus, un champ magnétique divergent produit une force qui accélère le plasma dans une direction parallèle au champ magnétique et entraîne également un transport du plasma orthogonal au champ magnétique ([11],[17],[18]). Ces deux comportements ne sont pas physiquement acceptables, ils ne devraient pas avoir lieu.

2.2.2 Présentation succinte des solutions existantes

Contrairement au physicien, le numéricien ne peut pas garantir une divergence de $\overrightarrow{B}$ parfaitement nulle (ne serait-ce que dû à l'erreur d'arrondi machine pour commencer). En outre, le numéricien a tendance à se demander pourquoi il devrait être plus précis sur cette donnée que sur les autres, et préfère s'assurer que la divergence converge vers 0 lorsque les pas de discrétisation en temps et en espace tendent à s'annuler (il cherche la consistance). En réalité, il est clair que la contrainte de Maxwell-Flux demande une attention particulère qui se situe entre les deux points de vue, pour les raisons évoquées plus haut. Dès lors, deux approches existent. Soit on part d'une solution physiquement irréprochable et on n'effectue aucun traitement particulier sur la divergence en espérant que l'erreur reste raisonnablement petite, soit on adopte d'entrée une approche algorithmique permettant de contrôler l'erreur commise sur la divergence du champ magnétique. La première approche **peut** donner des résultats satisfaisants suivant le type de schémas employés. Mais les algorithmes les plus précis et les plus robustes sont généralement structurés de façon à résoudre le problème supplémentaire de trouver un champ magnétique quasi solénoïdal.

Ce sera le cas pour nous aussi, sauf mention contraire (les deux approches sont possibles dans nos algorithmes). Par conséquent, nous devons rajouter explicitement la contrainte (2.1.3) au problème pendant la résolution numérique. Plusieurs techniques existent pour cela [95].

La plus simple à décrire consiste à ne pas tenir compte de (2.1.3) lors de l'établissement des équations, contrairement à ce que nous avons fait. Apparaissent alors des termes proportionnels à $\overrightarrow{\nabla} \cdot \overrightarrow{B}$ dans le

système. Cette approche fut proposée par Powell ([74]) dans le cadre d'une utilisation avec ses solveurs de Riemann approchés. La conséquence est d'introduire une huitième onde supplémentaire au système de la MHD idéale en 1D (problèmes de Riemann aux surfaces entre les éléments). Cette onde traduit une propagation des erreurs de divergence. La méthode de Powell s'est avérée améliorer les résultats par rapport à l'approche classique. Cependant les termes introduits sont non-conservatifs, ce qui a pour effets une perte potentielle d'information et surtout des relations de saut incorrectes qui excluent de pouvoir résoudre correctement des problèmes comportant des chocs.

Une autre méthode est celle du transport contraint (CT en anglais) introduit par Evans et Hawley ([38]). Il s'agit d'utiliser un schéma aux différences finies sur un maillage cartésien, mais en définissant le champ magnétique à des emplacements différents des autres variables. Ceci permet de construire un algorithme vérifiant par construction un analogue discret de la contrainte (2.1.3) à chaque pas de temps. Le choix de la définition de cet analogue discret est dicté par la méthode. Cette méthode a évolué pour intégrer des flux numériques de type Volumes Finis plus robustes. Par extension, on peut dire que le principe de ces méthodes précède celui des méthodes d'Éléments Finis récentes où l'on construit des éléments vérifiant aussi par construction un analogue discret de la contrainte lors de la résolution. Ces méthodes se révèlent très efficaces pour garantir un champ magnétique solénoïdal, l'erreur de divergence pouvant descendre jusqu'à l'erreur d'arrondi machine ! Toutefois, il s'agit d'utiliser une formulation Galerkin, plus onéreuse en termes de stockage mais avec les mêmes problèmes que la méthode de Galerkin classique, à moins d'employer des schémas de type Galerkin Discontinu encore plus coûteux. Ce n'est pas l'approche que nous avons retenue ici.

Enfin, une autre catégorie de méthodes largement utilisée est celle des schémas utilisant une projection (imaginée originalement par Boris [16]). L'idée est simple : résoudre à chaque pas de temps le système sans se préoccuper de l'erreur de divergence commise, puis projeter le champ magnétique solution sur un sous-espace de champs magnétiques de divergence nulle, et recommencer en partant du champ projeté obtenu. La phase de projection nécessite la résolution d'une équation de Poisson sur le tout domaine, donc la résolution d'un système linéaire.

On trouve parfois, chez les physiciens notamment, des algorithmes résolvant non pas le champ magnétique mais le potentiel vecteur $\overrightarrow{A}$, ce qui permet de laisser tomber la contrainte de Maxwell-Flux. Toutefois, cette formulation fait apparaître des dérivées spatiales d'ordre supérieur (on perd un ordre de précision et on sait généralement moins bien résoudre ces problèmes) et ne s'écrit pas de façon conservative (les problèmes comportant des chocs sont a priori exclus).

2.2.3 L'approche adoptée : *Divergence cleaning*

La méthode que nous avons choisie est celle proposée par Dedner et al. [35] à partir de travaux réalisés sur la résolution des équations de Maxwell ([9], [63], [64]). Elle consiste à introduire un multiplicateur de Lagrange dans le jeu de variables à résoudre, dont l'évolution permette de contrôler la divergence du champ magnétique. Ceci se traduit par une équation supplémentaire couplée au reste du système. Pour comprendre le fonctionnement de cette méthode, attardons-nous sur les quelques étapes qui l'ont construite.

Formulation contrainte des équations de Maxwell

Revenons sur les équations de Maxwell que nous avions présentées dans la première partie de la section précédente. Ce système est composé des équations (2.1.2)-(2.1.3)-(2.1.4)-(2.1.5) :

$$\frac{1}{c_0^2}\frac{\partial \overrightarrow{E}}{\partial t} - \overrightarrow{\nabla} \wedge \overrightarrow{B} = -\mu_0 \overrightarrow{j}$$
$$\frac{\partial \overrightarrow{B}}{\partial t} + \overrightarrow{\nabla} \wedge \overrightarrow{E} = \overrightarrow{0}$$
$$\overrightarrow{\nabla} \cdot \overrightarrow{E} = \frac{\rho_e}{\epsilon_0}$$
$$\overrightarrow{\nabla} \cdot \overrightarrow{B} = 0$$

Les deux dernières équations sont des contraintes du système, dans le sens où si elles sont vérifiées intialement, elles n'ont pas besoin d'apparaître pour que la solution à n'importe quelle date soit l'unique solution les respectant. En effet, tout comme nous avons démontré en début de section que l'évolution de la divergence du champ magnétique est nulle dans le cas continu, on peut montrer que :

$$\frac{\partial}{\partial t}\left(\overrightarrow{\nabla} \cdot \overrightarrow{E} - \frac{\rho_e}{\epsilon_0}\right) \overset{(2.1.4)}{=} -\left(\frac{1}{\epsilon_0}\frac{\partial \rho_e}{\partial t} - c_0^2 \overrightarrow{\nabla} \cdot \left(\overrightarrow{\nabla} \wedge \overrightarrow{B} - \mu_0 \overrightarrow{j}\right)\right)$$
$$\overset{(A.7)}{=} -\frac{1}{\epsilon_0}\left(\frac{\partial \rho_e}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{j}\right) = 0$$

On retrouve l'équation de conservation de la charge, donc la contrainte de Maxwell-Gauss (2.1.2) est vérifiée par la solution des seules équations (2.1.4) et (2.1.5) si et seulement si l'état initial vérifie également cette contrainte. Or généralement, lorsqu'on résoud les équations de Maxwell, la densité de courant $\overrightarrow{j}$ et la densité de charges ρ_e sont des sources connues (autrement le système serait sous-déterminé) qui vérifient par conséquent nécessairement la contrainte de Maxwell-Gauss. Cependant, si ce n'est pas le cas, ce qui semble pouvoir se produire dans les simulations PIC d'après [64], ou bien surtout, si la solution numérique est telle que

$$\frac{\partial (\rho_e)_h}{\partial t} + \overrightarrow{\nabla}_h \cdot \overrightarrow{j}_h \neq 0$$

ou que les opérateurs discrets (liés au schéma employé) ne vérifient pas l'identité (A.7), i.e.

$$\overrightarrow{\nabla}_h \cdot (\overrightarrow{\nabla}_h \wedge \cdot) \neq 0$$

il n'est plus possible d'espérer obtenir une solution respectant les contraintes (2.1.2) et (2.1.3) en résolvant seulement les équations d'évolution (2.1.4) et (2.1.5). Pour remédier à ce problème, Assous et al. ([9]) introduisirent des multiplicateurs de Lagrange (φ, ψ) associés aux deux contraintes dans les équations d'évolution. Le système obtenu est :

$$\frac{1}{c_0^2}\frac{\partial \overrightarrow{E}}{\partial t} - \overrightarrow{\nabla}\varphi - \overrightarrow{\nabla} \wedge \overrightarrow{B} = -\mu_0 \overrightarrow{j} \tag{2.2.1a}$$
$$\frac{\partial \overrightarrow{B}}{\partial t} - \overrightarrow{\nabla}\psi + \overrightarrow{\nabla} \wedge \overrightarrow{E} = \overrightarrow{0} \tag{2.2.1b}$$
$$\overrightarrow{\nabla} \cdot \overrightarrow{E} = \frac{\rho_e}{\epsilon_0}$$
$$\overrightarrow{\nabla} \cdot \overrightarrow{B} = 0$$

Les mêmes auteurs formulèrent, à partir de ce système contraint, un système d'équations d'ondes pour chacun des champs $(\overrightarrow{E}, \overrightarrow{B})$ et y adjoignirent les conditions initiales et limites adéquates, avant de les résoudre dans un contexte d'éléments finis. Ces conditions sont que les contraintes doivent être respectées à l'instant initial (ils n'avaient pas le problème rencontré dans les simulations PIC), et que les multiplicateurs

de Lagrange doivent être nuls sur tout le domaine à l'instant initial et sur les bords à tout instant. Bien entendu, la formulation est consistente avec le système de Maxwell original puisque les multiplicateurs de Lagrange sont uniformément nuls si les contraintes sont respectées.

Cette méthode reste proche de la méthode de projection employée par Boris ([16]) dans la mesure où les multiplicateurs de Lagrange doivent être actualisés en résolvant des problèmes stationnaires. Munz et al. reprirent ce type de formulation et l'étendirent dans le cadre de l'électromagnétisme pur (dans le vide) et des simulations PIC (densité de charge non uniformément nulle, donnée par l'équation de Vlasov) dans des configurations qui ne nécessitaient pas de se préoccuper de la contrainte (2.1.3) sur le champ magnétique ([63], [64]). La seule contrainte à vérifier dans ce cas est celle de Maxwell-Gauss. En effet, il faut remarquer que la plupart du temps les équations de Maxwell sont résolues dans le vide, et que dans ce cas l'absence d'intéraction avec une matière ionisée préserve le champ magnétique de tout emballement. La contrainte de Maxwell-Flux en électromagnétisme pur n'a donc pas du tout la même importance qu'en MHD. L'approche que formulèrent Munz et al. est une généralisation des méthodes du même type (problème contraint au niveau des équations, avec ou sans multiplicateur explicite) utilisées jusqu'alors, et elle permet même d'élargir cette famille. Ils la désignèrent pour cette raison sous le nom de formulation GLM (pour *generalized Lagrange multiplier*, cf. [63]). Ce système s'écrit :

$$\frac{\partial \overrightarrow{E}}{\partial t} - c_0^2 \overrightarrow{\nabla} \wedge \overrightarrow{B} + c_0^2 \overrightarrow{\nabla} \varphi = -\frac{\overrightarrow{j}}{\mu_0} \tag{2.2.2a}$$

$$\frac{\partial \overrightarrow{B}}{\partial t} + \overrightarrow{\nabla} \wedge \overrightarrow{E} = \overrightarrow{0}$$

$$\mathcal{D}(\varphi) + \overrightarrow{\nabla} \cdot \overrightarrow{E} = \frac{\rho_e}{\epsilon_0} \tag{2.2.2b}$$

$$\overrightarrow{\nabla} \cdot \overrightarrow{B} = 0$$

On constate un couplage plus explicite des équations, qui dépend d'un opérateur $\mathcal{D}$. Il est de plus intéressant de considérer l'équation que vérifie le multiplicateur de Lagrange pour comprendre la façon dont la correction opère. Il suffit pour cela de procéder exactement de la même manière que pour obtenir l'équation de conservation de la charge :

$$\frac{\partial}{\partial t}\left(\mathcal{D}(\varphi) + \overrightarrow{\nabla} \cdot \overrightarrow{E}\right) = \frac{1}{\epsilon_0}\frac{\partial \rho_e}{\partial t}$$

$$\overset{(2.2.2a)}{\Rightarrow} \quad \frac{\partial \mathcal{D}(\varphi)}{\partial t} + \overrightarrow{\nabla} \cdot \left(c_0^2 \overrightarrow{\nabla} \wedge \overrightarrow{B} - c_0^2 \overrightarrow{\nabla}\varphi - \frac{\overrightarrow{j}}{\epsilon_0}\right) = \frac{1}{\epsilon_0}\frac{\partial \rho_e}{\partial t}$$

$$\overset{(A.7)}{\Rightarrow} \quad \frac{\partial \mathcal{D}(\varphi)}{\partial t} - c_0^2 \Delta \varphi = \frac{1}{\epsilon_0}\left(\frac{\partial \rho_e}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{j}\right)$$

Les différentes méthodes qu'englobe la formulation GLM

La méthode de projection dûe à Boris ([16]), qui est la plus largement utilisée, se déduit des équations (2.2.2) en prenant $\mathcal{D}(\varphi) = 0$. Ce choix résulte en une équation de Poisson (elliptique, donc) sur φ :

$$-c_0^2 \Delta \varphi = \frac{1}{\epsilon_0}\left(\frac{\partial \rho_e}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{j}\right)$$

L'idée est donc de résoudre cette équation afin d'obtenir la solution φ^* puis d'injecter celle-ci dans l'équation de Maxwell-Ampère contrainte (2.2.2a) pour corriger le champ électrique, ce qui peut effectivement être vu comme une projection de champ $\overrightarrow{E}$ sur un sous-espace vérifiant la conservation de la charge et

donc la contrainte de Maxwell-Gauss. Ceci signifie qu'il faut résoudre un système linéaire à chaque pas de temps, l'équation étant linéaire en φ.

La seconde approche contrainte qu'on peut trouver dans la littérature est celle de Marder ([61]). Elle se retrouve en considérant $\mathcal{D}(\varphi) = \frac{\varphi}{X}$. L'équation sur φ devient alors :

$$\frac{\partial \varphi}{\partial t} - X c_0^2 \Delta \varphi = \frac{X}{\epsilon_0} \left(\frac{\partial \rho_e}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{j} \right)$$

On a à présent une équation d'évolution parabolique sur le multiplicateur de Lagrange. Cependant, on ne résoud pas cette équation mais on couple les équations (2.2.2a) et (2.2.2b) pour aboutir à :

$$\frac{\partial \overrightarrow{E}}{\partial t} - c_0^2 \overrightarrow{\nabla} \wedge \overrightarrow{B} = -\frac{\overrightarrow{j}}{\epsilon_0} - X c_0^2 \overrightarrow{\nabla} \left(\frac{\rho_e}{\epsilon_0} - \overrightarrow{\nabla} \cdot \overrightarrow{E} \right)$$

Cette méthode ne nécessite pas l'inversion d'un système linéaire comme la méthode de projection. Le terme supplémentaire s'apparente à un pseudo-courant correcteur. Si cette correction est moins onéreuse, elle est aussi moins efficace. Nielsen et Drobot ([66]) montrèrent cependant que la solution ainsi calculée rejoint asymptotiquement celle de la méthode de projection, et qu'en pratique quelques itérations suffisent.

Enfin, une autre alternative est de choisir un opérateur d'évolution pour $\mathcal{D}$, soit :

$$\mathcal{D}(\varphi) = \frac{1}{X^2} \frac{\partial \varphi}{\partial t}$$

L'équation sur φ est à présent donnée par :

$$\frac{\partial^2 \varphi}{\partial t^2} - (X c_0)^2 \Delta \varphi = \frac{X^2}{\epsilon_0} \left(\frac{\partial \rho_e}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{j} \right)$$

Il s'agit d'une loi purement hyperbolique qui traduit une propagation des erreurs de divergence à la vitesse $(X c_0)$. Avec ce choix, l'équation de Maxwell-Gauss se réécrit sous la forme évolutive suivante :

$$\frac{\partial \varphi}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{E} = \frac{\rho_e}{\epsilon_0}$$

Dans ([64]), Munz et al. font remarquer que φ, c'est-à-dire l'erreur transportée, peut s'interpréter dans ce cas comme un choix de jauge des équations de Maxwell. De plus, ils prouvent que là encore la solution converge vers celle calculée par la méthode de projection. Ce système contraint peut être maintenant résolu par n'importe quel schéma en temps sans recours à un traitement particulier pour la correction.

Remarque 2. *Le fonctionnement de ces différentes méthodes repose sur quelques conditions :*

$$\varphi(\overrightarrow{x}, t = 0) = 0 \qquad \forall \overrightarrow{x} \in \Omega$$

$$\varphi(\overrightarrow{x}, t) = 0 \qquad \forall (\overrightarrow{x}, t) \in \partial\Omega \times \mathbb{R}^+$$

Dans le cas d'une méthode de projection, on peut appliquer la correction au temps initial pour corriger un éventuel défaut de départ sur la divergence de $\overrightarrow{E}$. Si on choisit la formulation hyperbolique, Munz et al. font remarquer que l'équation des ondes sur φ doit être stabilisée par la condition de radiation :

$$\frac{\partial \varphi}{\partial t} + X \overrightarrow{n} \cdot \overrightarrow{\nabla} \varphi = 0 \qquad \forall \overrightarrow{x} \in \partial\Omega$$

où $\overrightarrow{n}$ est la normale sortante du domaine.

Dans le cadre de la MHD, nous ne préoccupons plus de la contrainte de Maxwell-Gauss puisque le champ électrostatique est absent des équations. Il reste néanmoins la contrainte de Maxwell-Flux, car tous les problèmes que nous envisagerons ne la vérifieront pas de façon triviale. Or les schémas numériques que nous utiliserons ne vérifient pas a priori d'équivalent à l'identité (A.7), donc nous aurons :

$$\overrightarrow{\nabla}_h \cdot \left(\overrightarrow{\nabla}_h \wedge \cdot \right) \neq 0$$

Pour remédier à ce problème, Dedner et al. ([35]) suggérèrent d'appliquer les techniques que nous venons d'exposer pour les équations de Maxwell. La formulation GLM pour la MHD modifie donc l'équation d'induction et la contrainte de Maxwell-Flux de la façon suivante :

$$\frac{\partial \overrightarrow{B}}{\partial t} + \overrightarrow{\nabla} \cdot \left(\overrightarrow{B}\overrightarrow{u}^t - \overrightarrow{u}\overrightarrow{B}^t \right) + \overrightarrow{\nabla}\psi = \overrightarrow{0} \tag{2.2.3a}$$

$$\mathcal{D}(\psi) + \overrightarrow{\nabla} \cdot \overrightarrow{B} = 0 \tag{2.2.3b}$$

On peut alors dégager plusieurs expressions régissant la divergence du champ magnétique et le multiplicateur de Lagrange :

$$\frac{\partial(\overrightarrow{\nabla} \cdot \overrightarrow{B})}{\partial t} + \Delta\psi = 0 \tag{2.2.4a}$$

$$\frac{\partial \mathcal{D}(\overrightarrow{\nabla} \cdot \overrightarrow{B})}{\partial t} + \Delta\mathcal{D}(\psi) = 0 \tag{2.2.4b}$$

$$\frac{\partial \mathcal{D}(\overrightarrow{\nabla} \cdot \overrightarrow{B})}{\partial t} - \Delta\left(\overrightarrow{\nabla} \cdot \overrightarrow{B} \right) = 0 \tag{2.2.4c}$$

$$\frac{\partial \mathcal{D}(\psi)}{\partial t} - \Delta\psi = 0 \tag{2.2.4d}$$

On peut refaire les mêmes distinctions que précédemment selon les formes de l'opérateur $\mathcal{D}$. La méthode de projection, introduite en MHD par Brackbill et Barnes dans [18], se retrouve en prenant $\mathcal{D}(\psi) = 0$, auquel cas on résoud le système de la MHD non corrigée pour obtenir une solution temporaire $\overrightarrow{B}^*$ avant de résoudre l'équation elliptique (2.2.4a) :

$$-\Delta\psi^* = \frac{1}{\Delta t}\left(\overrightarrow{\nabla} \cdot \overrightarrow{B}^* - \overrightarrow{\nabla} \cdot \overrightarrow{B} \right)$$

et de mettre à jour le champ magnétique en ajoutant le terme de correction :

$$\overrightarrow{B}^{n+1} = \overrightarrow{B}^* - \Delta t \overrightarrow{\nabla}\psi^*$$

Le champ magnétique obtenu est la projection de $\overrightarrow{B}^*$ sur un sous-espace de champs vérifiant la contrainte de Maxwell-Flux.

Pour éviter le coût supplémentaire de la résolution de l'équation elliptique, nous avons opté pour l'approche mixte hyperbolique-parabolique présentée dans [35]. Sur la même idée que pour les équations de Maxwell, on définit l'opérateur de la manière suivante :

$$\mathcal{D}(\psi) = \frac{1}{c_h^2}\frac{\partial \psi}{\partial t} + \frac{\psi}{c_p^2}$$

ψ vérifie alors une équation d'ondes (hyperbolique) amortie (présence d'un terme parabolique), aussi appelée équation du télégraphe :

$$\frac{\partial^2 \psi}{\partial t^2} + \frac{c_h^2}{c_p^2}\frac{\partial \psi}{\partial t} - c_h^2 \Delta\psi = 0$$

Les erreurs de divergence sont donc à la fois transportées et diffusées. La contrainte de Maxwell-Flux modifiée, couplée au système MHD, devient :

$$\frac{\partial \psi}{\partial t} + c_h^2 \overrightarrow{\nabla} \cdot \overrightarrow{B} = -\frac{c_h^2}{c_p^2}\psi \tag{2.2.5}$$

Le cas purement hyperbolique est retrouvé lorsque $c_p \to \infty$, cas qui peut s'avérer utile si le schéma employé pour résoudre cette équation est déjà très diffusif (la diffusion numérique jouant alors le rôle du terme source parabolique). En ce qui concerne les conditions aux limites à employer sur ψ, Dedner et al. ([35]) en proposent trois mais ne semblent pas avoir noté de dépendance particulière au choix qui est fait parmi celles-ci. Ils imputent ce fait au caractère dissipatif très marqué de la correction mixte. En accord avec les remarques faites au sujet des équations de Maxwell corrigées, nous considèrerons par défaut une condition limite de Dirichlet uniforme sur ψ :

$$\psi(\overrightarrow{x}, t) = 0 \qquad \forall\, (\overrightarrow{x}, t) \in \partial\Omega \times \mathbb{R}^+$$

De même, nous prendrons toujours des conditions initiales telles que :

$$\overrightarrow{\nabla} \cdot \overrightarrow{B}\,(\overrightarrow{x}, 0) = 0 \qquad \forall \overrightarrow{x} \in \Omega$$

bien que ce ne soit pas une condition *sine qua non* pour que la correction soit effective.

2.3 La MHD idéale sous le regard des mathématiques

2.3.1 Système propre

Le système classique

Le système d'équations de la MHD idéale sans correction peut se mettre sous la forme quasi-linéaire suivante :

$$\begin{aligned} \frac{\partial U}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F} &= 0 \\ \Rightarrow \quad \frac{\partial U}{\partial t} + \frac{\partial \overrightarrow{F}}{\partial U} \cdot \overrightarrow{\nabla} U &= 0 \\ \Rightarrow \quad \frac{\partial U}{\partial t} + \overrightarrow{A} \cdot \overrightarrow{\nabla} U &= 0 \end{aligned}$$

avec :

$$U = \begin{pmatrix} \rho \\ \rho \overrightarrow{u} \\ E \\ \overrightarrow{B} \end{pmatrix} \quad , \quad F_i = \begin{pmatrix} \rho u_i \\ \rho \overrightarrow{u} u_i + \left(p + \dfrac{\overrightarrow{B}^2}{2} \right) \overrightarrow{\delta_i} - B_i \overrightarrow{B} \\ \left(E + p + \dfrac{\overrightarrow{B}^2}{2} \right) u_i - \left(\overrightarrow{u} \cdot \overrightarrow{B} \right) B_i \\ \overrightarrow{B} u_i - \overrightarrow{u} B_i \end{pmatrix}$$

Ici, $\overrightarrow{A} = (A_x, A_y, A_z)$ est un vecteur de matrices, ce sont les jacobiennes des flux conservatifs $\overrightarrow{F}$. Soit $\overrightarrow{n}$ une direction quelconque dans $\mathbb{R}^3$, alors la projection $\overrightarrow{A} \cdot \overrightarrow{n}$ s'exprime de la manière suivante :

$$A_n^{C,0} = \begin{pmatrix} 0 & \overrightarrow{n}^t & 0 & \overrightarrow{0}^t \\ (\gamma-1)\dfrac{\overrightarrow{u}^2}{2}\overrightarrow{n} - u_n\overrightarrow{u} & (1-\gamma)\overrightarrow{n}\overrightarrow{u}^t + \overrightarrow{u}\overrightarrow{n}^t + u_n Id & (\gamma-1)\overrightarrow{n} & (2-\gamma)\overrightarrow{n}\overrightarrow{B}^t - B_n Id - \overrightarrow{B}\overrightarrow{n}^t \\ \left(A_n^C\right)_{5,1} & \dfrac{H}{\rho}\overrightarrow{n}^t + (1-\gamma)u_n\overrightarrow{u}^t - \dfrac{B_n}{\rho}\overrightarrow{B}^t & \gamma u_n & (2-\gamma)u_n\overrightarrow{B}^t - B_n\overrightarrow{u}^t - (\overrightarrow{u}\cdot\overrightarrow{B})\overrightarrow{n}^t \\ \underline{\dfrac{B_n\overrightarrow{u} - u_n\overrightarrow{B}}{\rho}} & \underline{\dfrac{\overrightarrow{B}\overrightarrow{n}^t - B_n Id}{\rho}} & \overrightarrow{0} & u_n Id - \overrightarrow{u}\overrightarrow{n}^t \end{pmatrix}$$

où nous avons noté :

$$\left(A_n^C\right)_{5,1} = u_n\left((\gamma-1)\frac{\overrightarrow{u}^2}{2} - \frac{H}{\rho}\right) + \frac{B_n}{\rho}\overrightarrow{u}\cdot\overrightarrow{B}$$

$H = E + p + \dfrac{\overrightarrow{B}^2}{2} = \dfrac{\gamma p}{\gamma-1} + \dfrac{1}{2}\rho\overrightarrow{u}^2 + \overrightarrow{B}^2$: l'enthalpie

$c = \sqrt{\dfrac{\gamma p}{\rho}}$: la vitesse du son hydrodynamique

$u_n = \overrightarrow{u}\cdot\overrightarrow{n}$

$B_n = \overrightarrow{B}\cdot\overrightarrow{n}$

L'exposant C est utilisé pour rappeler qu'on travaille dans la base des variables conservatives U.

Cependant, ce système ne respecte pas l'invariance Galiléenne et n'est pas symétrisable, à la différence des équations d'Euler en mécanique des fluides qui sont pourtant contenues dans le système MHD. Car on peut remarquer qu'il suffit de considérer le champ magnétique uniformément nul pour les retrouver. Le problème ne peut donc provenir que de la partie magnétique des équations. Or analytiquement, le système n'a pas d'écriture unique dans le sens où on peut rajouter des termes proportionnels à $\overrightarrow{\nabla}\cdot\overrightarrow{B}$, ceux-ci étant nuls.

Partant de ce constat, Godunov ([43]) parvint à réécrire les équations pour obtenir un système symétrisable et respectant le principe d'invariance Galiléenne. Les termes ajoutés, proportionnels à $\overrightarrow{\nabla}\cdot\overrightarrow{B}$, aboutissent au système suivant :

$$\begin{aligned}
\frac{\partial\rho}{\partial t} + \overrightarrow{\nabla}\cdot(\rho\overrightarrow{u}) &= 0 \\
\frac{\partial(\rho\overrightarrow{u})}{\partial t} + \overrightarrow{\nabla}\cdot\left(\rho\overrightarrow{u}\overrightarrow{u}^t + \left(p + \frac{\overrightarrow{B}^2}{2}\right)Id - \overrightarrow{B}\overrightarrow{B}^t\right) + \overrightarrow{B}\left(\overrightarrow{\nabla}\cdot\overrightarrow{B}\right) &= \overrightarrow{0} \\
\frac{\partial E}{\partial t} + \overrightarrow{\nabla}\cdot\left(\left(E + p + \frac{\overrightarrow{B}^2}{2}\right)\overrightarrow{u} - \left(\overrightarrow{u}\cdot\overrightarrow{B}\right)\overrightarrow{B}\right) + \overrightarrow{u}\cdot\overrightarrow{B}\left(\overrightarrow{\nabla}\cdot\overrightarrow{B}\right) &= 0 \\
\frac{\partial\overrightarrow{B}}{\partial t} + \overrightarrow{\nabla}\cdot\left(\overrightarrow{u}\overrightarrow{B}^t - \overrightarrow{B}\overrightarrow{u}^t\right) + \overrightarrow{u}\left(\overrightarrow{\nabla}\cdot\overrightarrow{B}\right) &= \overrightarrow{0}
\end{aligned}$$

Les termes rajoutés ne sont ni plus ni moins que la compensation des termes proportionnels à $\overrightarrow{\nabla}\cdot\overrightarrow{B}$ qui étaient présents dans le système conservatif original, comme le fit notamment remarquer Powell ([74]) qui s'attacha à résoudre ces équations plutôt que les anciennes en proposant un solveur de Riemann adéquat. Pour s'en convaincre, il suffit de développer les termes d'origine magnétique dans la divergence en suivant la règle (2.1.8). Nous avons vus certains de ces termes à compenser lorsque nous dérivions les équations

dans la section précédente. L'inconvénient majeur de ce nouveau système est toutefois de ne plus s'écrire de façon conservative.

En revanche, il est toujours possible de l'exprimer sous forme quasi-linéaire. Si nous recalculons les jacobiennes du système symétrisable de la même manière que précédemment, nous obtenons :

$$A_n^{C,1} = \begin{pmatrix} 0 & \overrightarrow{n}^t & 0 & \overrightarrow{0}^t \\ (\gamma-1)\dfrac{\overrightarrow{u}^2}{2}\overrightarrow{n} - u_n\overrightarrow{u} & (1-\gamma)\overrightarrow{n}\overrightarrow{u}^t + \overrightarrow{u}\overrightarrow{n}^t + u_n Id & (\gamma-1)\overrightarrow{n} & (2-\gamma)\overrightarrow{n}\overrightarrow{B}^t - B_n Id \\ \left(A_n^C\right)_{5,1} & \dfrac{H}{\rho}\overrightarrow{n}^t + (1-\gamma)u_n\overrightarrow{u}^t - \dfrac{B_n}{\rho}\overrightarrow{B}^t & \gamma u_n & (2-\gamma)u_n\overrightarrow{B}^t - B_n\overrightarrow{u}^t \\ \dfrac{B_n\overrightarrow{u} - u_n\overrightarrow{B}}{\rho} & \dfrac{\overrightarrow{B}\overrightarrow{n}^t - B_n Id}{\rho} & \overrightarrow{0} & u_n Id \end{pmatrix}$$

Avec la correction

Nous allons maintenant faire exactement la même chose en prenant en compte les équations de *divergence cleaning*. Les 2 dernières équations du système conservatif deviennent donc :

$$\begin{aligned} \frac{\partial \overrightarrow{B}}{\partial t} + \overrightarrow{\nabla}\cdot\left(\overrightarrow{u}\overrightarrow{B}^t - \overrightarrow{B}\overrightarrow{u}^t + \psi Id\right) &= \overrightarrow{0} \\ \frac{\partial \psi}{\partial t} + {c_h}^2\overrightarrow{\nabla}\cdot\overrightarrow{B} &= -\frac{{c_h}^2}{{c_p}^2}\psi \end{aligned}$$

et nous redéfinissons les quantités $U = \begin{pmatrix} \rho \\ \rho\overrightarrow{u} \\ E \\ \overrightarrow{B} \\ \psi \end{pmatrix}$ et $F_i = \begin{pmatrix} \rho u_i \\ \rho\overrightarrow{u}u_i + (p + \dfrac{\overrightarrow{B}^2}{2})\delta_{ij} - B_i\overrightarrow{B} \\ u_i H - B_i(\overrightarrow{u}\cdot\overrightarrow{B}) \\ B_i\overrightarrow{u} - u_i\overrightarrow{B} + \psi\delta_{ij} \\ c_h^2 B_i \end{pmatrix}$.

Dedner et al. ([35]) ont déterminé les termes à rajouter pour que leurs équations vérifient à leur tour l'invariance galiléenne. Voici le système qui en résulte :

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \overrightarrow{\nabla}\cdot(\rho\overrightarrow{u}) &= 0 \\ \frac{\partial(\rho\overrightarrow{u})}{\partial t} + \overrightarrow{\nabla}\cdot\left(\rho\overrightarrow{u}\overrightarrow{u}^t + \left(p + \frac{\overrightarrow{B}^2}{2}\right)Id - \overrightarrow{B}\overrightarrow{B}^t\right) + \overrightarrow{B}\left(\overrightarrow{\nabla}\cdot\overrightarrow{B}\right) &= \overrightarrow{0} \\ \frac{\partial E}{\partial t} + \overrightarrow{\nabla}\cdot\left(\left(E + p + \frac{\overrightarrow{B}^2}{2}\right)\overrightarrow{u} - \left(\overrightarrow{u}\cdot\overrightarrow{B}\right)\overrightarrow{B}\right) + \overrightarrow{u}\cdot\overrightarrow{B}\left(\overrightarrow{\nabla}\cdot\overrightarrow{B}\right) + \overrightarrow{B}\cdot\left(\overrightarrow{\nabla}\psi\right) &= 0 \\ \frac{\partial \overrightarrow{B}}{\partial t} + \overrightarrow{\nabla}\cdot\left(\overrightarrow{u}\overrightarrow{B}^t - \overrightarrow{B}\overrightarrow{u}^t + \psi Id\right) + \overrightarrow{u}\left(\overrightarrow{\nabla}\cdot\overrightarrow{B}\right) &= \overrightarrow{0} \\ \frac{\partial \psi}{\partial t} + {c_h}^2\overrightarrow{\nabla}\cdot\overrightarrow{B} + \frac{{c_h}^2}{{c_p}^2}\psi + \overrightarrow{u}\cdot\left(\overrightarrow{\nabla}\psi\right) &= 0 \end{aligned}$$

Les jacobiennes projetées suivant une direction $\overrightarrow{n}$ nous donnent alors :

$$A_n^{C,2} = \begin{pmatrix} 0 & \overrightarrow{n}^t & 0 & \overrightarrow{0}^t & 0 \\ (\gamma-1)\dfrac{\overrightarrow{u}^2}{2}\overrightarrow{n} - u_n\overrightarrow{u} & (1-\gamma)\overrightarrow{n}\,\overrightarrow{u}^t + \overrightarrow{u}\,\overrightarrow{n}^t + u_n Id & (\gamma-1)\overrightarrow{n} & (2-\gamma)\overrightarrow{n}\,\overrightarrow{B}^t - B_n Id & \overrightarrow{0} \\ (K_c)_{5,1} & \dfrac{H}{\rho}\overrightarrow{n}^t + (1-\gamma)u_n\overrightarrow{u}^t - \dfrac{B_n}{\rho}\overrightarrow{B}^t & \gamma u_n & (2-\gamma)u_n\overrightarrow{B}^t - B_n\overrightarrow{u}^t & B_n \\ \dfrac{B_n\overrightarrow{u} - u_n\overrightarrow{B}}{\rho} & \dfrac{\overrightarrow{B}\,\overrightarrow{n}^t - B_n Id}{\rho} & \overrightarrow{0} & u_n Id & \overrightarrow{n} \\ 0 & \overrightarrow{0}^t & 0 & {c_h}^2\overrightarrow{n}^t & u_n \end{pmatrix}$$

Symétrisation

Généralement, pour simplifier l'expression des jacobiennes, on passe par un changement de variables. Pour commencer, on peut considérer le jeu des variables physiques :

$$V = \begin{pmatrix} \rho \\ \overrightarrow{u} \\ p \\ \overrightarrow{B} \\ \psi \end{pmatrix}$$

Le système quasi-linéaire conservatif peut se réécrire :

$$\begin{aligned} \frac{\partial U}{\partial t} + \frac{\partial \overrightarrow{F}}{\partial U}\cdot\overrightarrow{\nabla} U &= 0 \\ \Rightarrow \quad \frac{\partial U}{\partial V}\frac{\partial V}{\partial t} + \left(\frac{\partial \overrightarrow{F}}{\partial U}\frac{\partial U}{\partial V}\right)\cdot\overrightarrow{\nabla} V &= 0 \end{aligned}$$

Et on obtient donc le système quasi-linéaire physique :

$$\begin{aligned} \frac{\partial V}{\partial t} + \left(\frac{\partial V}{\partial U}\frac{\partial \overrightarrow{F}}{\partial U}\frac{\partial U}{\partial V}\right)\cdot\overrightarrow{\nabla} V &= 0 \\ \Rightarrow \quad \frac{\partial V}{\partial t} + \overrightarrow{A}^V\cdot\overrightarrow{\nabla} V &= 0 \end{aligned}$$

La matrice projetée A_n^V est donnée par :

$$A_n^V = \begin{pmatrix} u_n & \rho\overrightarrow{n}^t & 0 & \overrightarrow{0}^t & 0 \\ \overrightarrow{0} & u_n Id & \dfrac{\overrightarrow{n}}{\rho} & \dfrac{\overrightarrow{n}\,\overrightarrow{B}^t - B_n Id}{\rho} & 0 \\ 0 & \rho c^2\overrightarrow{n}^t & u_n & \overrightarrow{0}^t & 0 \\ \overrightarrow{0} & \overrightarrow{B}\,\overrightarrow{n}^t - B_n Id & \overrightarrow{0} & u_n Id & \overrightarrow{n} \\ 0 & \overrightarrow{0}^t & 0 & c_h^2\overrightarrow{n}^t & u_n \end{pmatrix}$$

La symétrisation des équations MHD non corrigées peut se trouver par exemple dans ([54]). Le jeu de variables symétrisantes pour le système corrigé est :

$$dW = \left(\frac{dp}{c^2}, \frac{\rho}{c}du, \frac{\rho}{c}dv, \frac{\rho}{c}dw, \frac{dp}{c^2} - d\rho, \frac{\sqrt{\rho}}{c}dB_x, \frac{\sqrt{\rho}}{c}dB_y, \frac{\sqrt{\rho}}{c}dB_z, \frac{\sqrt{\rho}}{c}\frac{d\psi}{c_h}\right)^t$$

Les matrices de passage se calculent alors aisément :

$$\frac{\partial W}{\partial V} = \begin{pmatrix} 0 & \overrightarrow{0}^t & \frac{1}{c^2} & \overrightarrow{0}^t & 0 \\ \overrightarrow{0} & \frac{\rho}{c} Id & \overrightarrow{0} & \overrightarrow{0}\,\overrightarrow{0}^t & \overrightarrow{0} \\ -1 & \overrightarrow{0}^t & \frac{1}{c^2} & \overrightarrow{0}^t & 0 \\ \overrightarrow{0} & \overrightarrow{0}\,\overrightarrow{0}^t & \overrightarrow{0} & \frac{\sqrt{\rho}}{c} Id & \overrightarrow{0} \\ 0 & \overrightarrow{0}^t & 0 & \overrightarrow{0}^t & \frac{\sqrt{\rho}}{c}\frac{1}{c_h} \end{pmatrix}$$

$$\frac{\partial V}{\partial W} = \begin{pmatrix} 1 & \overrightarrow{0}^t & -1 & \overrightarrow{0}^t & 0 \\ \overrightarrow{0} & \frac{c}{\rho} Id & \overrightarrow{0} & \overrightarrow{0}\,\overrightarrow{0}^t & \overrightarrow{0} \\ c^2 & \overrightarrow{0}^t & 0 & \overrightarrow{0}^t & 0 \\ \overrightarrow{0} & \overrightarrow{0}\,\overrightarrow{0}^t & \overrightarrow{0} & \frac{c}{\sqrt{\rho}} Id & \overrightarrow{0} \\ 0 & \overrightarrow{0}^t & 0 & \overrightarrow{0}^t & \frac{c}{\sqrt{\rho}} c_h \end{pmatrix}$$

On peut également travailler à partir des variables conservatives en utilisant les matrices suivantes :

$$\frac{\partial U}{\partial W} = \begin{pmatrix} 1 & \overrightarrow{0}^t & -1 & \overrightarrow{0}^t & 0 \\ \overrightarrow{u} & cId & -\overrightarrow{u} & \overrightarrow{0}\,\overrightarrow{0}^t & 0 \\ h & c\overrightarrow{u}^t & -\frac{\overrightarrow{u}^2}{2} & \frac{c\overrightarrow{B}^t}{\sqrt{\rho}} & 0 \\ \overrightarrow{0} & \overrightarrow{0}\,\overrightarrow{0}^t & \overrightarrow{0} & \frac{c}{\sqrt{\rho}} Id & 0 \\ 0 & \overrightarrow{0}^t & 0 & \overrightarrow{0}^t & \frac{c}{\sqrt{\rho}} c_h \end{pmatrix}$$

$$\frac{\partial W}{\partial U} = \begin{pmatrix} \beta\frac{\overrightarrow{u}^2}{2} & -\beta\overrightarrow{u}^t & \beta & -\beta\overrightarrow{B}^t & 0 \\ -\frac{\overrightarrow{u}}{c} & \frac{1}{c} Id & \overrightarrow{0} & \overrightarrow{0}\,\overrightarrow{0}^t & 0 \\ \beta\frac{\overrightarrow{u}^2}{2} - 1 & -\beta\overrightarrow{u}^t & \beta & -\beta\overrightarrow{B}^t & 0 \\ \overrightarrow{0} & \overrightarrow{0}\,\overrightarrow{0}^t & \overrightarrow{0} & \frac{\sqrt{\rho}}{c} Id & 0 \\ 0 & \overrightarrow{0}^t & 0 & \overrightarrow{0}^t & \frac{\sqrt{\rho}}{c}\frac{1}{c_h} \end{pmatrix}$$

où $h = \dfrac{c^2}{\gamma - 1} + \dfrac{\overrightarrow{u}^2}{2}$ et $\beta = \dfrac{\gamma - 1}{c^2}$.

A présent, il ne reste plus qu'à effectuer le changement de base de $A_n^{C,2}$ pour obtenir la matrice symétrique A_n^S :

$$A_n^S = \frac{\partial W}{\partial U} A_n^{C,2} \frac{\partial U}{\partial W} = \begin{pmatrix} u_n & c\overrightarrow{n}^t & 0 & \overrightarrow{0}^t & 0 \\ c\overrightarrow{n} & u_n Id & \overrightarrow{0} & \frac{\overrightarrow{n}\overrightarrow{B}^t - B_n Id}{\sqrt{\rho}} & 0 \\ 0 & \overrightarrow{0}^t & u_n & \overrightarrow{0}^t & 0 \\ \overrightarrow{0} & \frac{\overrightarrow{B}\overrightarrow{n}^t - B_n Id}{\sqrt{\rho}} & \overrightarrow{0} & u_n Id & c_h\overrightarrow{n} \\ 0 & \overrightarrow{0}^t & 0 & c_h\overrightarrow{n}^t & u_n \end{pmatrix}$$

A ce stade, on voit clairement que l'invariance Galiléenne est satisfaite puisque $A_n^S = u_n Id + A_n^S\big|_{u_n=0}$. La symétrie nous assure que la matrice est bien diagonalisable à valeurs propres réelles, donc que le système symétrisé est hyperbolique. Et comme on peut retrouver le système propre conservatif par changements de variables, comme nous allons le faire, le système conservatif original est lui aussi hyperbolique.

Valeurs propres et vecteurs propres du système symétrisé

À partir de maintenant, nous allons supposer que $||\overrightarrow{n}|| = 1$. Ceci ne change absolument rien à la structure des matrices utilisées jusqu'à présent, puisqu'on peut facilement remarquer que les jacobiennes calculées dans n'importe quelle base sont linéaires en la norme de $\overrightarrow{n}$. Il suffira donc par exemple, une fois les calculs terminés, de multiplier les valeurs propres par $||\overrightarrow{n}||$ et le système propre sera bien celui de la jacobienne originale.

Les valeurs propres que nous obtenons sont les suivantes :

$$\begin{aligned}
\lambda_0 &= u_n \\
\lambda_{\pm f} &= u_n \pm c_f \\
\lambda_{\pm s} &= u_n \pm c_s \\
\lambda_{\pm h} &= u_n \pm c_h \\
\lambda_{\pm a} &= u_n \pm c_a
\end{aligned}$$

où nous avons introduit les vitesses des ondes MHD (hormis la constante c_h qui est une vitesse artificielle d'advection des erreurs sur la divergence de $\overrightarrow{B}$), dans l'ordre :

. la vitesse des ondes magnétosoniques rapides :

$$c_f = \sqrt{\frac{1}{2}\left(c^2 + \frac{\overrightarrow{B}^2}{\rho} + \sqrt{\left(c^2 + \frac{\overrightarrow{B}^2}{\rho}\right)^2 - 4c^2\frac{B_n^2}{\rho}}\right)}$$

. la vitesse des ondes magnétosoniques lentes :

$$c_s = \sqrt{\frac{1}{2}\left(c^2 + \frac{\overrightarrow{B}^2}{\rho} - \sqrt{\left(c^2 + \frac{\overrightarrow{B}^2}{\rho}\right)^2 - 4c^2\frac{B_n^2}{\rho}}\right)}$$

. la vitesse des ondes d'Alfvén :

$$c_a = \frac{B_n}{\sqrt{\rho}}$$

On rappelle que c est la vitesse du son hydrodynamique, valeur propre des équations d'Euler, qui pour un gaz parfait s'exprime $c = \sqrt{\frac{\gamma p}{\rho}}$. Nous sommes à présent en mesure de calculer les vecteurs propres du système symétrisé, que nous allons présenter via la matrice R_S^0 où ils sont rangés par colonnes et dans l'ordre où nous avons présenté les valeurs propres juste au-dessus (soit r_0, r_{+f}, r_{-f}, r_{+s}, r_{-s}, r_{+h}, etc.).

$$R_S^0 = \begin{pmatrix}
0 & c & -c & c & -c & 0 & 0 & 0 & 0 \\
0 & \overrightarrow{l_f^0} & \overrightarrow{l_f^0} & \overrightarrow{l_s^0} & \overrightarrow{l_s^0} & \overrightarrow{0} & \overrightarrow{0} & \overrightarrow{l^0} & \overrightarrow{l^0} \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \overrightarrow{m_f^0} & -\overrightarrow{m_f^0} & \overrightarrow{m_s^0} & -\overrightarrow{m_s^0} & \overrightarrow{n} & \overrightarrow{n} & -\overrightarrow{l^0} & \overrightarrow{l^0} \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0
\end{pmatrix}$$

où nous avons :

$$\overrightarrow{l_f^0} = c_f \left(\overrightarrow{n} - \frac{B_n}{\rho c_f^2 - B_n^2} \overrightarrow{B_\perp} \right) \qquad \overrightarrow{l_s^0} = c_s \left(\overrightarrow{n} - \frac{B_n}{\rho c_s^2 - B_n^2} \overrightarrow{B_\perp} \right)$$

$$\overrightarrow{m_f^0} = \frac{\sqrt{\rho} c_f^2}{\rho c_f^2 - B_n^2} \overrightarrow{B_\perp} \qquad \overrightarrow{m_s^0} = \frac{\sqrt{\rho} c_s^2}{\rho c_s^2 - B_n^2} \overrightarrow{B_\perp}$$

$$\overrightarrow{l^0} = \frac{\overrightarrow{n} \wedge \overrightarrow{B}}{\sqrt{\rho}} \qquad \overrightarrow{B_\perp} = \overrightarrow{B} - B_n \overrightarrow{n}$$

L'exposant 0 signifie que les vecteurs ne sont pas normalisés. On peut vérifier aisément que tous ces vecteurs propres sont bien orthogonaux. À titre de comparaison, voilà le système propre qu'on obtient par le même procédé en l'absence de *divergence cleaning* :

$$R_S^0 = \begin{pmatrix} 0 & 0 & c & -c & c & -c & 0 & 0 \\ 0 & 0 & \overrightarrow{l_f^0} & \overrightarrow{l_f^0} & \overrightarrow{l_s^0} & \overrightarrow{l_s^0} & \overrightarrow{l^0} & \overrightarrow{l^0} \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \overrightarrow{n} & \overrightarrow{m_f^0} & -\overrightarrow{m_f^0} & \overrightarrow{m_s^0} & -\overrightarrow{m_s^0} & -\overrightarrow{l^0} & \overrightarrow{l^0} \end{pmatrix}$$

On constate que les vecteurs propres associés aux ondes rapides, lentes ainsi qu'aux ondes d'Alfvén sont les mêmes. Le premier vecteur propre est, dans les deux cas, celui associé à l'entropie. Dans la dernière matrice, on s'aperçoit immédiatement que le second vecteur propre est orthogonal à tous les autres (toutes les quantités m_i^0 et l^0 sont orthogonales à $\overrightarrow{n}$). La valeur propre qui lui est associée est $\lambda_1 = u_n = \lambda_0$. Elle correspond à la propagation des erreurs commises sur la divergence. On peut dès lors noter que le fait de rajouter la correction de la divergence a pour effet de remplacer cette "onde de progagation de la divergence" en deux nouvelles ondes se déplaçant à une vitesse relative c_h imposée et qui traduisent elles aussi le transport des erreurs de divergence à travers ψ.

Il est ensuite utile de normaliser ces vecteurs propres, puisque cela entraîne que la matrice des formes propres devient la transposée des vecteurs propres, i.e. $L_S = (R_S)^t$. La matrice R_S^0 devient donc :

$$R_S = \begin{pmatrix} 0 & c\alpha_f^s & -c\alpha_f^s & c\alpha_s^s & -c\alpha_s^s & 0 & 0 & 0 & 0 \\ 0 & \overrightarrow{l_f} & \overrightarrow{l_f} & \overrightarrow{l_s} & \overrightarrow{l_s} & \overrightarrow{0} & \overrightarrow{0} & \overrightarrow{l} & \overrightarrow{l} \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \overrightarrow{m_f} & -\overrightarrow{m_f} & \overrightarrow{m_s} & -\overrightarrow{m_s} & \overrightarrow{n} & \overrightarrow{n} & -\overrightarrow{l} & \overrightarrow{l} \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{pmatrix}$$

où les vecteurs propres sont effectivement normalisés si l'on rajoute que :

$$\alpha_f^s := \frac{\rho c_f^2 - B_n^2}{c^2 \left(\rho c_f^2 - B_n^2 \right)^2 + \overrightarrow{l_f^0}^2 + \overrightarrow{m_f^0}^2} \qquad \alpha_s^s := \frac{\rho c_s^2 - B_n^2}{c^2 \left(\rho c_s^2 - B_n^2 \right)^2 + \overrightarrow{l_s^0}^2 + \overrightarrow{m_s^0}^2}$$

$$\overrightarrow{l_f} := \frac{\overrightarrow{l_f^0}}{c^2 \left(\rho {c_f}^2 - {B_n}^2 \right)^2 + \overrightarrow{l_f^0}^2 + \overrightarrow{m_f^0}^2} \qquad \overrightarrow{l_s} := \frac{\overrightarrow{l_s^0}}{c^2 \left(\rho {c_s}^2 - {B_n}^2 \right)^2 + \overrightarrow{l_s^0}^2 + \overrightarrow{m_s^0}^2}$$

$$\overrightarrow{l} := \frac{\overrightarrow{l^0}}{\sqrt{2 \overrightarrow{l^0}^2}}$$

Cependant, il faut noter la présence de sources de singularités dans les expressions qui apparaissent aux dénominateurs, pour les vecteurs propres des ondes rapides, lentes et des ondes d'Alfvén.

Un moyen de s'affranchir de certaines singularités

Pour ce qui est des ondes d'Alfvén par exemple, les vecteurs propres normalisés ne sont plus définis dès lors que $\overrightarrow{B}$ est suivant $\overrightarrow{n}$, c'est-à-dire si $\overrightarrow{B_\perp} = \overrightarrow{0}$. Brio et Wu ([19]) ont rencontré ce problème très tôt en résolvant des problèmes unidimensionnels. Ils ont suggéré que dans ce cas, on devait pouvoir remplacer $\overrightarrow{B_\perp}$ par un vecteur arbitraire tant que celui-ci permettait de conserver une base de vecteurs propres orthonormée. Jusqu'à présent, et dans tous les tests que nous avons faits, nous n'avons jamais eu de raison de remettre en cause ce choix. Par exemple en 2D, nous avons choisi de prendre $\overrightarrow{B_\perp} = \overrightarrow{n_\perp}$ dans le plan, tel que : $\det(\overrightarrow{n}, \overrightarrow{n_\perp}) = 1$.

La partie la plus complexe est celle des ondes magnétosoniques, rapides ou lentes. Là encore, les vecteurs propres ne sont pas définis lorsque $\overrightarrow{B}$ est dirigé suivant $\overrightarrow{n}$. En s'inspirant des travaux de Brio et de Wu ([19]), Roe et Balsara ([82]) proposèrent une écriture différente en manipulant algébriquement les vecteurs propres de la jacobienne du système physique (et non de celle du système symétrisé), de façon à ce que ce soient ces vecteurs propres qui forment une base orthonormée. L'avantage de cette formulation est que sur les six situations algébriques remarquables répertoriées dans ([82]), tandis que toutes constituaient une singularité dans notre écriture des vecteurs propres du système symétrisé, seule une le reste dans le système propre physique proposé par Roe et Balsara. Certaines de ces situations illustrent bien le fait que certaines ondes peuvent être amenées à se confondre, les ondes magnétosoniques pouvant se comporter comme des ondes d'Alfvén par exemple. Quant à la dernière singularité, il s'agit du *point triple*, où les ondes rapides viennent à se confondre à la fois avec les ondes lentes et avec celles d'Alfvén. Ce sont autant de preuves que le système de la MHD, même avec la correction de la divergence qui ramène la multiplicité de la valeur propre u_n à 1, est loin d'être strictement hyperbolique, même en dimension 1 d'espace. On renvoit le lecteur à ([82]) pour plus de détails.

Nous allons maintenant présenter directement ce nouveau système propre une fois repassé en variables conservatives, c'est-à-dire celui que nous utilisons en pratique dans nos algorithmes. Les vecteurs propres et formes linéaires propres (écrites ici en colonnes et non en lignes) sont, dans l'ordre cité plus haut :

$$r_0 = \begin{pmatrix} 1 \\ \overrightarrow{u} \\ \frac{\overrightarrow{u}^2}{2} \\ \overrightarrow{0} \\ 0 \end{pmatrix} \qquad (l_0)^t = \frac{1}{c^2} \begin{pmatrix} 1 + (1-\gamma)\frac{\overrightarrow{u}^2}{2} \\ (\gamma-1)\overrightarrow{u} \\ 1-\gamma \\ (\gamma-1)\overrightarrow{B} \\ 0 \end{pmatrix}$$

$$r_{\pm f} = \begin{pmatrix} \rho\alpha_f \\ \rho\alpha_f(\overrightarrow{u} \pm c_f \overrightarrow{n}) \mp \rho\alpha_s c_s b_n \overrightarrow{b_\perp} \\ \rho\alpha_f\left(\frac{\overrightarrow{u}^2}{2} \pm u_n c_f + \frac{c^2}{\gamma-1}\right) \mp \rho\alpha_s c_s b_n \overrightarrow{b_\perp}\cdot\overrightarrow{u} + \sqrt{\rho}\alpha_s c \left\|\overrightarrow{B_\perp}\right\| \\ \sqrt{\rho}\alpha_s c \overrightarrow{b_\perp} \\ 0 \end{pmatrix}$$

$$
(l_{\pm f})^t = \frac{1}{2\rho c^2}
\begin{pmatrix}
\alpha_f \left((\gamma-1)\dfrac{\overrightarrow{u}^2}{2} \mp c_f u_n \right) \pm \alpha_s c_s b_n \overrightarrow{b_\perp} \cdot \overrightarrow{u} \\
\alpha_f \left((1-\gamma)\overrightarrow{u} \pm c_f \overrightarrow{n} \right) \mp \alpha_s c_s b_n \overrightarrow{b_\perp} \\
(\gamma-1)\alpha_f \\
\sqrt{\rho} c \alpha_s \overrightarrow{b_\perp} + (1-\gamma)\alpha_f \overrightarrow{B} \\
0
\end{pmatrix}
$$

$$
r_{\pm s} = \begin{pmatrix}
\rho \alpha_s \\
\rho \alpha_s (\overrightarrow{u} \pm c_s \overrightarrow{n}) \pm \rho \alpha_f c_f b_n \overrightarrow{b_\perp} \\
\rho \alpha_s \left(\dfrac{\overrightarrow{u}^2}{2} \pm u_n c_s + \dfrac{c^2}{\gamma-1} \right) \pm \rho \alpha_f c_f b_n \overrightarrow{b_\perp} \cdot \overrightarrow{u} - \sqrt{\rho} \alpha_f c \left\| \overrightarrow{B_\perp} \right\| \\
-\sqrt{\rho} \alpha_f c \overrightarrow{b_\perp} \\
0
\end{pmatrix}
$$

$$
(l_{\pm s})^t = \frac{1}{2\rho c^2}
\begin{pmatrix}
\alpha_s \left((\gamma-1)\dfrac{\overrightarrow{u}^2}{2} \mp c_s u_n \right) \mp \alpha_f c_f b_n \overrightarrow{b_\perp} \cdot \overrightarrow{u} \\
\alpha_s \left((1-\gamma)\overrightarrow{u} \pm c_s \overrightarrow{n} \right) \pm \alpha_f c_f b_n \overrightarrow{b_\perp} \\
(\gamma-1)\alpha_s \\
-\sqrt{\rho} c \alpha_f \overrightarrow{b_\perp} + (1-\gamma)\alpha_s \overrightarrow{B} \\
0
\end{pmatrix}
$$

$$
r_{\pm h} = \begin{pmatrix} 0 \\ \overrightarrow{0} \\ B_n \\ \overrightarrow{n} \\ \pm c_h \end{pmatrix}
\qquad
(l_{\pm h})^t = \frac{1}{2} \begin{pmatrix} 0 \\ \overrightarrow{0} \\ 0 \\ \overrightarrow{n} \\ \pm \dfrac{1}{c_h} \end{pmatrix}
$$

$$r_{\pm a} = \begin{pmatrix} 0 \\ \overrightarrow{l_a} \\ \overrightarrow{l_a} \cdot \overrightarrow{u} \\ \mp \dfrac{\overrightarrow{l_a}}{\sqrt{\rho}} \\ 0 \end{pmatrix} \qquad (l_{\pm a})^t = \frac{1}{2} \begin{pmatrix} -\overrightarrow{l_a} \cdot \overrightarrow{u} \\ \overrightarrow{l_a} \\ 0 \\ \mp\sqrt{\rho}\overrightarrow{l_a} \\ 0 \end{pmatrix}$$

où nous avons introduit les quantités suivantes :

$$b_n = \frac{B_n}{|B_n|} \qquad \overrightarrow{b_\perp} = \frac{\overrightarrow{B} - B_n \overrightarrow{n}}{\left\| \overrightarrow{B} - B_n \overrightarrow{n} \right\|}$$

$$\alpha_f = \sqrt{\frac{c^2 - c_s^2}{c_f^2 - c_s^2}} \qquad \alpha_s = \sqrt{\frac{c_f^2 - c^2}{c_f^2 - c_s^2}}$$

$$\overrightarrow{l_a} = \frac{\overrightarrow{n} \wedge \overrightarrow{B}}{\left\| \overrightarrow{n} \wedge \overrightarrow{B} \right\|}$$

Tenant compte de ce que nous avons dit plus haut, les vecteurs $\overrightarrow{b_\perp}$ et $\overrightarrow{l_a}$ sont modifiés artificiellement lorsque $\left\| \overrightarrow{b_\perp} \right\| \simeq 0$ en 1D, et en 2D on choisit aussi le plus souvent possible une direction $\overrightarrow{n}$ qui ne soit ni colinéaire ni orthogonale à $\overrightarrow{B}$ lorsque la méthode numérique nous l'autorise.

2.3.2 Formulation faible

Nous cherchons à présent à appliquer des notions bien connues sur les lois de conservation de manière générale, i.e. tout système du type :

$$\frac{\partial U}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U) = 0 \tag{2.3.1}$$

aux équations de la MHD idéale. La théorie des caractéristiques permet d'exhiber le fait que pour des flux non linéaires, il est possible que les caractéristiques entrent en contact en un temps fini. L'exemple scalaire type est l'équation de Burgers non visqueuse, pour laquelle $f(u) = \dfrac{u^2}{2}$. À partir d'un tel évènement, il n'existe plus de solution régulière, ce qui signifie qu'elle présente par la suite un caractère localement discontinu. Or si on cherche la solution de (2.3.1) directement, on suppose a priori que la solution U présente une certaine régularité (disons $\mathcal{C}^1$), qui n'autorise pas le type de discontinuités dont on connaît l'existence. Pour cette raison, on cherchera toujours la solution d'une formulation intégrale de (2.3.1), qu'on appelera une solution faible, par opposition à une solution régulière qu'on qualifierait de forte. Cette écriture nécessite l'introduction d'un espace de fonctions test à support compact dans le domaine d'étude, disons $V = \mathcal{C}_0^1(\Omega \times [0;T])$ pour une date $T > 0$ quelconque. On exprime alors :

$$\forall \varphi \in V, \quad \int_0^T \int_\Omega \varphi \left(\frac{\partial U}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U) \right) d\overrightarrow{x}\, dt = 0 \tag{2.3.2}$$

ce qui est une réécriture de (2.3.1) au sens des distributions. Toute fonction U vérifiant cette équation intégrale est appelée solution faible. Comme nous le disions, cette notion est plus générale que celle de

solution forte car elle comprend des solutions localement discontinues sur des ensembles de mesure nulle, c'est-à-dire des hypersurfaces de Ω. La figure 2.1 illustre nos propos dans le cas où Ω est un ouvert borné de $\mathbb{R}^2$. Une implication notable de ce changement de formulation est que la solution faible n'est pas définie de manière unique : on pourrait se dire qu'il existe a priori une infinité de fonctions U pouvant vérifier (2.3.2) si rien ne contrôle la discontinuité. En réalité, on peut d'ores et déjà trouver des informations

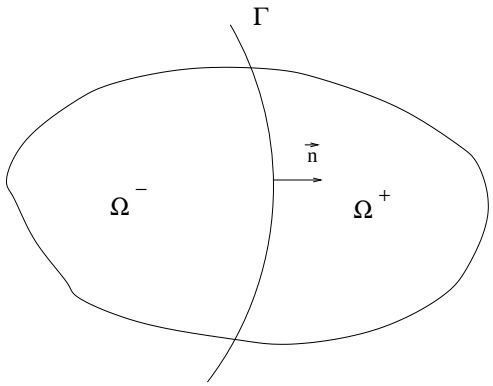


FIGURE 2.1 – Exemple d'hypersurface Γ localisant la discontinuité

sur celle-ci sans se donner d'informations supplémentaires. Ce qui fait qu'il n'existe pas une infinité de solutions, mais seulement plusieurs. Pour cela, on se base sur les notations de la figure 2.1. De part et d'autre de Γ, la solution est régulière. Si on réécrit (2.3.2) sur ce domaine, en décomposant l'intégrale sur Ω^- et Ω^+ puis en intégrant par parties, on aboutit à des conditions de saut sur les variations de U à travers la discontinuité, qui sont des équations de la forme :

$$\sigma\left(U_k^+ - U_k^-\right) = \left(\overrightarrow{F_k}(U^+) - \overrightarrow{F_k}(U^-)\right) \cdot \overrightarrow{n} \tag{2.3.3}$$

sur chaque composante U_k de U et en tout point de Γ. Ce sont les conditions de Rankine-Hugoniot (voir par exemple [36] ou [85] dans le cas général, [92] pour une application pour la mécanique des fluides et [55], [14] ou [93] pour l'application à la MHD).

Pour être complets notons qu'à l'inverse, il peut également arriver que les caractéristiques, au lieu de se concentrer en un point de l'espace-temps, se mettent à diverger sous la forme d'un faisceau à partir d'un autre point. Dans ce cas, les équations donnent lieu à une détente, c'est-à-dire une zone de régularité de la solution connectant des états distants. La méthode des caractéristiques est alors parfaitement appropriée et on étudie la solution en se basant sur le fait qu'elle est autocentrée, i.e. que $U(x,t) = R\left(\frac{x-x_0}{t}, U_g, U_d\right)$ où les états U_g et U_d sont les états connectés par la détente et où x_0 désigne le point de départ de la divergence des caractéristiques (voir [36] ou [85]).

Que les données initiales soient discontinues ou que des caractéristiques se croisent à un certain moment sont deux situations identiques : à partir d'un tel évènement on peut localement se concentrer sur l'évolution en temps d'une solution dont certaines parties présentent une discontinuité, ce qu'on appelle le problème de Riemann. De ce point de concentration naissent une ou plusieurs ondes (au plus, autant que d'équations). Ces ondes séparent des états intermédiaires à déterminer, constants ou non. La figure suivante donne une vision schématique de ce qui se passe.

Nous allons maintenant présenter un problème de Riemann unidimensionnel faisant intervenir la plupart des ondes citées jusqu'ici. Ceci nous permettra de nous appuyer sur un cadre concret pour comprendre le comportement des diverses ondes rencontrées en MHD, ainsi que la manière dont elles

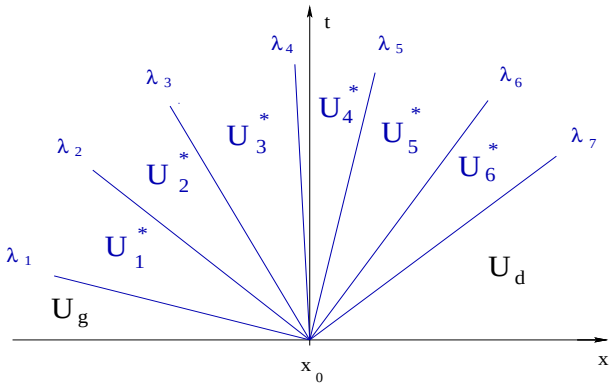


FIGURE 2.2 – Un problème de Riemann 1D générant les 7 ondes de la MHD unidimensionnelle. Les notations ne signifient en aucun cas que les ondes sont séparées par des états constants. En toute rigueur, pour des détentes, nous aurions pu dessiner des faisceaux de caractéristiques. Ce n'est qu'une illustration très schématique.

agissent sur le plasma. On note les variables conservatives ainsi :

$$U = \begin{pmatrix} \rho \\ \rho u_x \\ \rho \overrightarrow{u_\perp} \\ E \\ \overrightarrow{B_\perp} \end{pmatrix}$$

où $\overrightarrow{u_\perp} = (u_y, u_z)$ et $\overrightarrow{B_\perp} = (B_y, B_z)$. Le champ B_x n'apparaît pas car il disparaît naturellement des équations et peut donc être considéré comme une constante du problème. De plus, comme toutes les dérivées selon y et z s'annulent, la divergence du champ magnétique revient à la dérivée de B_x suivant x, qui est nulle. Ceci a pour conséquence qu'en 1D, la contrainte de Maxwell-Flux est toujours vérifiée (pour peu qu'elle le soit initialement). Le système des 7 équations restantes s'écrit donc :

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x}(U) = 0$$

avec

$$F(U) = \begin{pmatrix} \rho u_x \\ \rho {u_x}^2 + p + \frac{1}{2}\overrightarrow{B_\perp}^2 \\ \rho u_x \overrightarrow{u_\perp} - B_x \overrightarrow{B_\perp} \\ \left(E + p + \frac{1}{2}{B_x}^2 + \frac{1}{2}\overrightarrow{B_\perp}^2\right) u_x - \left(u_x B_x + \overrightarrow{u_\perp} \cdot \overrightarrow{B_\perp}\right) B_x \\ u_x \overrightarrow{B_\perp} - B_x \overrightarrow{u_\perp} \end{pmatrix}$$

Pour fixer davantage les idées, le meilleur parti est de donner un exemple de problème. Nous procèderons donc comme dans [93] et [94] en reprenant le même exemple. La donnée initiale du problème de Riemann est une solution discontinue en $x = 0$ composée de deux états différents U_g (à gauche) et U_d (à droite), définis comme suit :

$$\begin{aligned} \left(\rho_g, (u_x)_g, (\overrightarrow{u_\perp})_g, B_x, (\overrightarrow{B_\perp})_g, p_g\right) &= \left(3, 0, \overrightarrow{0}, \frac{3}{2}, (1,0)^t, 3\right) \\ \left(\rho_d, (u_x)_d, (\overrightarrow{u_\perp})_d, B_x, (\overrightarrow{B_\perp})_d, p_d\right) &= \left(1, 0, \overrightarrow{0}, \frac{3}{2}, (\cos\alpha, \sin\alpha)^t, 3\right) \end{aligned}$$

avec $\alpha = 1.5$ (presque $\frac{\pi}{2}$) et $\gamma = 5/3$. Les résultats sont donnés par la figure 2.3. Les ondes sont au nombre de 7, la taille du système, car aucune valeur propre n'est de multiplicité supérieure à 1 (hors cas particuliers qu'on ne détaillera pas). Dans cet exemple, on recense de gauche à droite, c'est-à-dire des ondes les plus rapides vers la gauche aux ondes les plus rapides vers la droite : une détente rapide, une discontinuité rotationnelle, une détente lente, une discontinuité de contact, un choc lent, une autre discontinuité rotationnelle et enfin un choc rapide, toutes ces ondes étant respectivement associées aux valeurs propres $\lambda_{-f} = u_n - c_f$, $\lambda_{-a} = u_n - c_a$, $\lambda_{-s} = u_n - c_s$, $\lambda_0 = u_n$, $\lambda_{+s} = u_n + c_s$, $\lambda_{+a} = u_n + c_a$ et $\lambda_{+f} = u_n + c_f$. Il ne s'agit pas exactement des vitesses auxquelles elles se déplacent puisque nous n'avons pas précisé en quelles valeurs de U celles-ci devaient être évaluées.

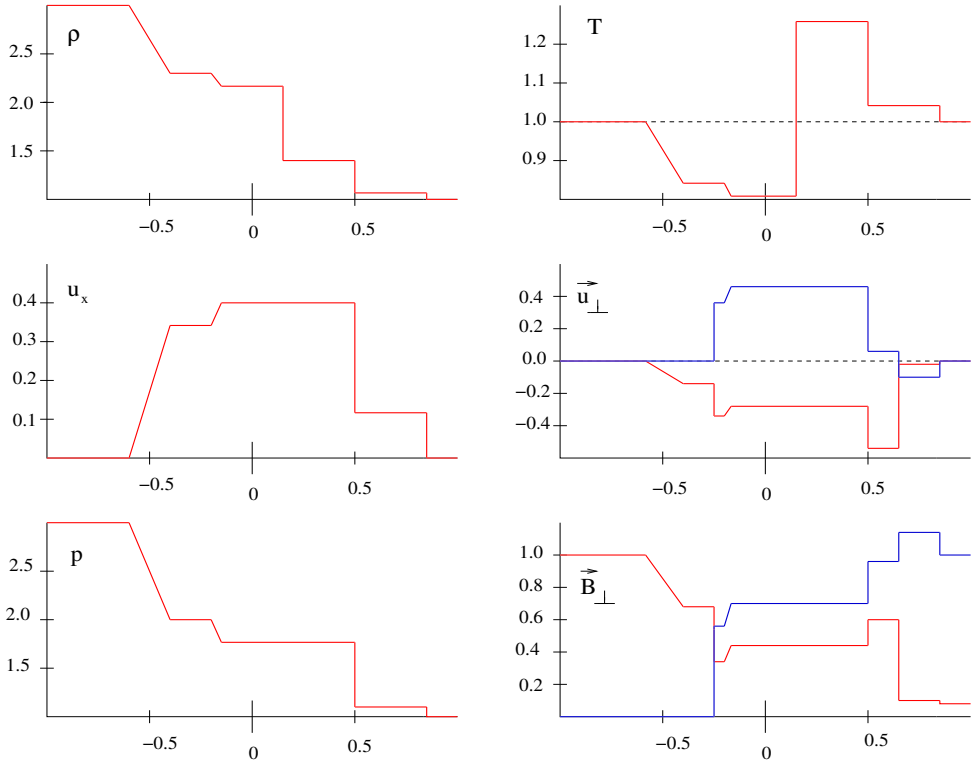


FIGURE 2.3 – Solution exacte du problème de Riemann à $t = 0.4$ selon [93]. Pour les quantités vectorielles (vitesse et champ magnétique orthogonaux), les différentes composantes sont représentées par des courbes séparées.

La recherche de la solution exacte du problème de Riemann est une tâche complexe. On ne sait pas a priori quelles ondes vont apparaître. Or pour trouver la solution, il faut le savoir car toutes ne se résolvent pas de la même façon : s'il s'agit d'une discontinuité, on dispose des conditions de Rankine-Hugoniot, sinon il faut procéder autrement. Hormis configurations particulières, qu'on ne détaillera pas ici, les ondes MHD peuvent se diviser en deux catégories : les champs linéairement dégénérés et les champs vraiment non linéaires. Pour les premiers, il s'agit d'une discontinuité dont la vitesse de propagation est par définition la valeur propre qui lui est associée, car elle prend la même valeur de part et d'autre. Les ondes concernées sont les discontinuités de contact et les ondes d'Alfvén (qui sont dans l'exemple précédent les discontinuités rotationnelles). Pour les champs vraiment non linéaires, donc les ondes magnétoacoustiques dans notre exemple, on ne saurait dire à l'avance si ce sont des détentes ou des chocs. Dans tous les cas, ces ondes séparent deux états dont un doit être connu. Comme on ne connait que les état U_g et U_d, on ne peut commencer que par formuler les états qui leur sont mitoyens, i.e. U_1^* et U_7^*. Cependant, les ondes 1 et 7 sont

toujours des champs vraiment non linéaires. Pour savoir comment ces ondes transforment l'écoulement on dispose soit de 7 invariants de Riemann (pour les détentes), soit des 7 conditions de Rankine-Hugoniot (2.3.3) (pour les chocs). Cependant, il y a une 8^e inconnue qui est la vitesse de propagation, contrairement aux champs linéairement dégénérés. Par conséquent, il existe une infinité de solutions à chacune des ondes concernées. Mais si, partant de l'état de droite par exemple, on écrit toutes les équations jusqu'à l'état de gauche, arrivé à celui-ci, on a aboutit à un système avec quelques inconnues de plus que d'équations (moins de 7), que l'on peut fermer grâce à la connaissance de l'état de gauche. Le système est donc bel et bien déterminé, voire surdéterminé si on raisonne en termes de nombre d'inconnues par rapport au nombre d'équations. Néanmoins, la solution construite comme nous le décrivons n'est pas unique. Certes, il n'y en a plus une infinité, mais les équations de fermeture que l'on choisit sont non linéaires et donc admettent plusieurs solutions. De plus, il faudrait construire ce système d'équations algériques pour tous les cas de figures, c'est-à-dire en considérant que chaque champ vraiment non linéaire peut être une détente ou bien un choc. Il manque donc visiblement un critère physique qui puisse permettre de savoir quelles ondes parmi celles envisageables peuvent se produire en réalité. Ce raisonnement suit celui du chapitre 4 de [92], qui s'applique au système des 3 équations d'Euler en dimension $d = 1$.

Remarque 3. *Ce n'est pas seulement un choix anodin que de nous être placés dans une configuration 1D pour écrire les conditions de Rankine-Hugoniot. À notre connaissance, il n'existe pas de méthode efficace pour résoudre le problème de Riemann généralisé à des hypersurfaces de $\mathbb{R}^2$ ou de $\mathbb{R}^3$. Le mieux que nous sachions généralement faire, c'est traiter le cas où l'hypersurface est un hyperplan et où la symétrie le long de celui-ci permet de se ramener en chaque point à un problème de Riemann 1D. C'est ce raisonnement qui est à la base de toutes les méthodes de type Godunov en dimension supérieure à 1, y compris les schémas Volumes Finis dont nous parlerons au chapitre suivant. Ces méthodes se basent donc fondamentalement sur des mécanismes 1D.*

2.3.3 Le rôle de l'entropie

La formulation faible permet donc d'appréhender les singularités rencontrées dans les systèmes non linéaires, au prix d'une multiplication des solutions possibles lorsque celles-ci surviennent. Or il ne peut y avoir qu'une seule solution sans cesse reproductible à un problème déterministe. Nous n'avons pas utilisé toutes les informations données par la physique et exposées dans la section 2.1. Les équations de conservation l'ont été, de même qu'en thermodynamique le premier principe et l'équation d'état, mais pas le second principe régissant l'entropie qui doit pourtant être respecté. Les solutions ne respectant pas ce principe doivent être écartées.

La solution entropique est unique

Dans la sous-section 2.1.3, nous avions formulé la version continue du second principe, c'est-à-dire l'inégalité de Clausius-Duhem ((2.1.14) pour la MHD idéale). Il se trouve que les fonctions vérifiant ce type d'inégalités sont des fonctions concaves, ce qui est donc le cas de ρs vis-à-vis des variables conservatives U. Or les mathématiciens préférant travailler avec des quantités convexes, on emploie généralement une entropie mathématique qui équivaut au signe près à ρs. La plus simple expression est

$$S(U) = -\rho s \tag{2.3.4}$$

mais d'autres pourraient être envisagées tant qu'elles restent convexes. Dans tous les cas, on requiert que les flux $\overrightarrow{G}(U)$ qui lui sont associés vérifient :

$$\frac{\partial S}{\partial t}(U) + \overrightarrow{\nabla} \cdot \overrightarrow{G}(U) \leq 0 \tag{2.3.5}$$

et permettent à cette inégalité d'être équivalente par changement de variable à celle de Clausius-Duhem. L'égalité doit de plus être atteinte dans les régions de la solution ne comportant pas de chocs. Pour S définie par (2.3.4), la physique impose que le flux associé soit $\overrightarrow{G}(S) = S\overrightarrow{u}$. D'une manière générale, si reformule l'inégalité de Clausius-Duhem par un changement de variable introduisant une nouvelle entropie mathématique, la physique impose la forme du flux qui doit lui être associé. On dit alors qu'on dispose d'un couple entropie-flux $(S, \overrightarrow{G})$.

Un autre argument physique non pris en compte est que le système de la MHD idéale est une limite du système de la MHD complète comprenant tous les effets diffusifs et dissipatifs. Autrement dit, la solution physique est forcément une limite de la solution d'un système du type

$$\frac{\partial U_\epsilon}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_\epsilon) = \overrightarrow{\nabla} \cdot (\epsilon \overrightarrow{F^v}(U_\epsilon)) \tag{2.3.6}$$

où $\overrightarrow{F^v}$ désigne les flux diffusifs proportionnels à $\overrightarrow{\nabla} U_\epsilon$, lorsque $\epsilon \to 0$. Or on peut montrer ([36],[85]) que, dans le cas d'une loi de conservation scalaire équipée de la même manière d'un couple entropie-flux, toute solution ainsi construite vérifie l'inégalité d'entropie dictée par son entropie physique, et ce pour toute entropie mathématique avec un flux compatible. On peut formaliser ce résultat par le théorème suivant.

Théorème 2.3.1.
Si $u = \lim_{\epsilon \to 0} u_\epsilon$ où u_ϵ est la solution faible d'une loi scalaire dissipative du type de (2.3.6), alors u est la solution faible entropique du système idéal correspondant (du type de (2.3.1)), c'est-à-dire la seule solution vérifiant (2.3.5) pour tout couple entropie-flux $(S, \overrightarrow{G})$.

Dans notre cas d'un système d'équations, le même résultat n'est pas démontré. Toutefois, on peut penser qu'il existe probablement, par extension, un lien (qui n'est donc pas encore formalisé) entre le respect du second principe de la thermodynamique et le fait d'être une limite du système d'équations complet. Selon [36], il est possible de trouver un tel résultat avec des modèles diffusifs contrôlés par des paramètres mathématiques évanescents, même si ce type de formulation est éloigné de la formulation physique. Toutefois, l'auteur fait remarquer qu'il est souvent possible d'adapter ces raisonnements à des modèles plus réalistes.

Revenons-en à présent aux ondes MHD. C'est le critère d'entropie qui permet de déterminer la nature de l'onde pour chaque champ vraiment non-linéaire (les discontinuités de contact préservent quant à elles l'entropie, autre cas où l'inégalité devient égalité). De plus, le respect de toutes les inégalités d'entropie pour des fonctions S convexes implique la condition de Lax [60] sur la vitesse de propagation réelle de chaque choc :

$$\forall 1 \leq i \leq p, \quad \min_{U=U_g,U_d} (\nabla_U \lambda_i(U) \cdot r_i(U)) \leq \sigma_i \leq \max_{U=U_g,U_d} (\nabla_U \lambda_i(U) \cdot r_i(U))$$

où les vecteurs colonnes $r_i(U)$ sont les vecteurs propres de $\nabla_U F_x$ (le problème de Riemann étant 1D). La condition de Lax est formulée au départ pour des lois de conservation scalaire, mais sa généralisation aux systèmes est triviale et est une conséquence du possible découplage des équations (la matrice $\nabla_U F_x$ étant diagonalisable par changement de variable) pour tout système hyperbolique en dimension $d = 1$.

On peut montrer également que les chocs expansifs ne sont pas admissibles. Les discontinuités de contact en MHD sont similaires à ce qu'elles sont en mécanique des fluides : la densité est différente de part et d'autre mais il n'y a pas de matière les traversant. Elles ne modifient pas non plus le champ magnétique de part et d'autre. Pour ce qui concerne l'action des différentes ondes sur le champ magnétique, le plus efficace est de reprendre l'illustration 2.4 issue de [93] et qui montre se qui ce passe pour l'exemple précédent de problème de Riemann, dont la solution est donnée par la figure 2.3.

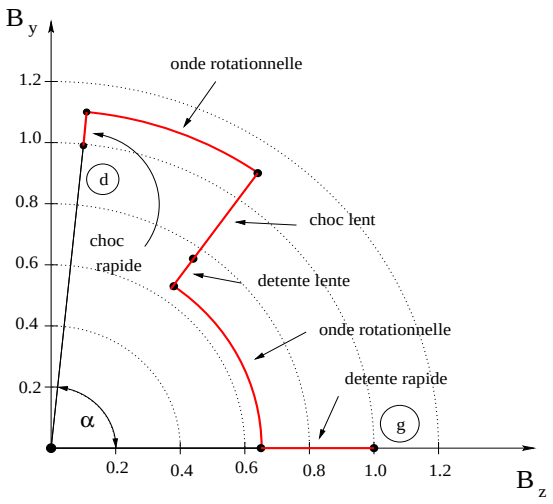


FIGURE 2.4 – Action des diverses ondes sur le champ magnétique transverse

Les chocs et les détentes rapides renforcent l'intensité du champ magnétique tandis que les ondes lentes la compriment, mais seules les discontinuités rotationnelles (ondes d'Alfvén) en modifient l'orientation (elles agissent donc aussi sur la direction de l'écoulement).

Pour conclure, il est important de faire remarquer que ce que nous avons dit sur le problème de Riemann et sa résolution, et donc les différentes ondes pouvant survenir, ne s'adapte pas à toutes les configurations. Selon l'orientation du champ magnétique par rapport à la direction de l'écoulement, des configurations très particulières peuvent émerger. En particulier de nouveaux types de chocs connectant des états physiquement très éloignés. Une bonne vision synthétique des transformations possibles est donnée par la figure 2.5 reprise de [32]. Nous reprenons aussi volontairement les commentaires associés car ils nous paraissent bien expliquer la situation.

Les transitions qualifiées d'intermédiaires font parfois encore l'objet d'un débat pour savoir si elles sont admissibles ou non, car elles n'ont été découvertes à l'origine que par les simulations numériques. Il semblerait depuis quelques années qu'un certain nombre d'entre elles aient été observées expérimentalement (voir par exemple [32] ou [93]).

Autre intérêt de l'entropie : la symétrisation des équations

L'entropie joue également occupe également un place cruciale dans l'étude des lois de conservation pour son lien étroit avec la symétrisation des équations. Nous avons déjà vu une version symétrisée des équations en recherchant le système propre. Cependant, nous avons aussi remarqué alors que les vecteurs propres pouvaient être normalisés de manière assez diverses, à en juger par la différence d'écritures entre le premier système que nous avions dérivé à la façon de Jameson [54] et celui mieux conditionné obtenu par Roe et Balsara [82]. L'entropie procure de fait un outil permettant d'obtenir une écriture encore différente du système propre, dont le conditionnement, dicté par un principe physique profond, est sans doute meilleur que celui dont nous disposons jusqu'ici.

Nous pouvons maintenant revenir au cas multidimensionnel ($d > 1$). Les variables W définies par $W^t = \dfrac{\partial S}{\partial U}$ sont appelées variables entropiques. La convexité de S signifie que pour tout vecteur de variables V non nul, la jacobienne de W (donc la Hessienne de S) doit être positive :

$$V^t \left(\nabla_U{}^2 S\right) V > 0$$

De plus, si nous supposons que cette jacobienne est inversible, elle est définie (son noyau se restreint à

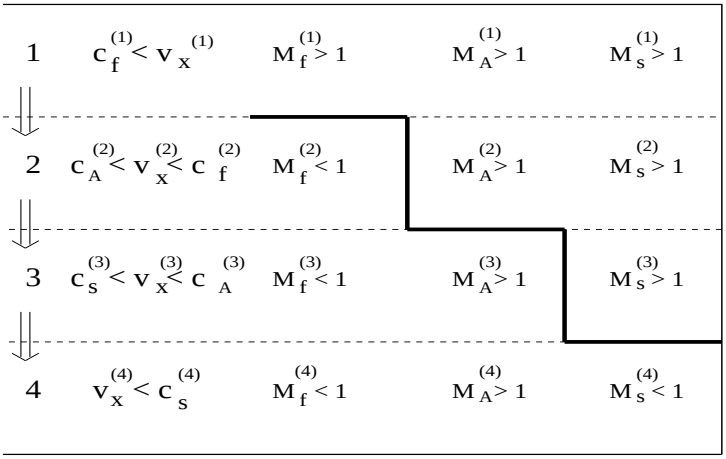


FIGURE 2.5 – États pouvant être connectés par un choc MHD. Ils sont rangés par ordre d'entropie croissante, l'état 1 possédant la plus basse. Les vitesses d'onde normales rapides, d'Alfvén et lentes sont respectivement c_f, c_A et c_s. Les nombres de Mach rapide, d'Alfvén et lent sont respectivement M_f, M_A et M_s. L'état 1 est *superrapide*, car sa vitesse de déplacement est supérieure à c_f à travers le choc. Donc tous les nombres de Mach y sont plus grands que 1. L'état 2 est *subrapide* mais *super-Alfvénique*. L'état 3 est *sub-Alfvénique* mais *superlent*. L'état 4 est *sublent*. Les transitions de type choc possibles sont 1-2 (rapide), 3-4 (lent) et 1-3, 1-4, 2-3, 2-4 (intermédiaires). Cette représentation est reprise de [32].

l'élément nul). Les dérivées croisées de S étant symétriques, on peut en conclure que

$$A_0 = \frac{\partial W}{\partial U} = \frac{\partial^2 S}{\partial U^2}$$

est symétrique définie positive. De plus, si nous nous plaçons dans une zone régulière où l'inégalité de Clausius-Duhem devient égalité, nous pouvons formuler une contrainte sur les flux d'entropie :

$$\begin{aligned}
\frac{\partial U}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U) &= 0 \\
\Rightarrow \quad \frac{\partial S}{\partial U}\frac{\partial U}{\partial t} + \frac{\partial S}{\partial U}\frac{\partial \overrightarrow{F}}{\partial U} \cdot \overrightarrow{\nabla} U &= 0 \\
\Rightarrow \quad \frac{\partial S}{\partial t} + \left(\frac{\partial S}{\partial U}\frac{\partial \overrightarrow{F}}{\partial U}\frac{\partial U}{\partial S}\right) \cdot \overrightarrow{\nabla} S &= 0 \\
&= \frac{\partial S}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{G}(U) \\
\Rightarrow \quad \left(\frac{\partial S}{\partial U}\frac{\partial \overrightarrow{F}}{\partial U}\frac{\partial U}{\partial S}\right) \cdot \overrightarrow{\nabla} S &= \frac{\partial \overrightarrow{G}}{\partial U} \cdot \overrightarrow{\nabla} U \\
\Rightarrow \quad \left(\frac{\partial S}{\partial U}\frac{\partial \overrightarrow{F}}{\partial U}\right) \cdot \overrightarrow{\nabla} U &= \frac{\partial \overrightarrow{G}}{\partial U} \cdot \overrightarrow{\nabla} U
\end{aligned}$$

où on applique partout les règles du produit matriciel usuel pour multiplier le vecteur ligne $\nabla_U S$ par chaque matrice jacobienne $\nabla_U F_i$ ou par un vecteur colonne $\partial_t U$ (produit scalaire). Par conséquent, ceci

devant être vrai pour tout U, le flux d'entropie doit toujours vérifier la condition de compatibilité :

$$\nabla_U S \nabla_U \overrightarrow{F} = \nabla_U \overrightarrow{G} \tag{2.3.7}$$

Cette relation permet en outre (voir [85]) de construire un vecteur de fonctions $\overrightarrow{H}$ telles que $\forall i, 1 \leq i \leq d$, chaque matrice

$$\frac{\partial^2 H_i}{\partial W^2} = \frac{\partial F_i}{\partial W} = \frac{\partial F_i}{\partial U}\frac{\partial U}{\partial W} = \frac{\partial F_i}{\partial U} {A_0}^{-1}$$

soit symétrique (les dérivées croisées étant symétriques), et permettant de réécrire (2.3.1) en changeant de variables :

$${A_0}^{-1}\partial_t W + \frac{\partial \overrightarrow{F}}{\partial W} \cdot \overrightarrow{\nabla} W = 0 \tag{2.3.8}$$

Or puisque A_0 est définie positive, on peut définir sa matrice racine carrée dont la décomposition s'écrit :

$${A_0}^{1/2} = R_0 \Lambda_0 L_0$$

De cette manière, on a :

$$A_0 = R_0 {\Lambda_0}^2 L_0$$

On définit également ${A_0}^{-1/2}$ par rapport à ${A_0}^{-1}$ selon le même procédé, puisque ${A_0}^{-1}$ est aussi définie positive. À partir de là, si on définit un jeu de variables W' tel que

$$dW' = {A_0}^{-1/2} dW$$

on peut réécrire (2.3.8) de la manière suivante :

$$\partial_t W' + \left[{A_0}^{1/2} \left(\frac{\partial \overrightarrow{F}}{\partial U} {A_0}^{-1} \right) {A_0}^{1/2} \right] \cdot \overrightarrow{\nabla} W' = 0$$

Il est facile de vérifier que cette matrice est symétrique, donc on a bel et bien symétrisé les équations dans la base des variables W'. Étudions une combinaison linéaire quelconque de ces d matrices avec de nouvelles notations :

$$\begin{gathered} \frac{\partial \overrightarrow{F}}{\partial U} \cdot \overrightarrow{n} = \frac{\partial \overrightarrow{F_n}}{\partial U} = K_n \\ K_n = R \Lambda L \\ M_n = {A_0}^{1/2} K_n {A_0}^{-1/2} \\ M_n = R_M \Lambda_M L_M \end{gathered}$$

Ayant calculé le système propre associé à M_n, on en déduit aisément le système propre associé à K_n.

$$R = {A_0}^{-1/2} R_M \text{ et } L = L_M {A_0}^{1/2}$$

Dans le cadre des équations de la MHD non corrigées et avec les termes sources de Powell, ces calculs ont été menés avec succès par Barth [13]. Comme l'avait découvert Godunov ([43]), les termes non-conservatifs sont indispensables pour que le système soit symétrisable, c'est-à-dire pour que le flux d'entropie imposé par l'inégalité de Clausius-Duhem soit compatible avec les lois de conservation au sens de (2.3.7).

Que se passe-t-il si on souhaite étendre ces résultats au système corrigé par la méthode de *divergence cleaning* ? Si on procède de la manière classique en définissant l'entropie mathématique par

$$S = -\frac{\rho s}{\gamma - 1}$$

alors les variables entropiques sont :

$$W = \begin{pmatrix} -\frac{-s}{\gamma-1} + \frac{\gamma}{\gamma-1} - \frac{\rho \overrightarrow{u}^2}{2p} \\ \frac{\rho \overrightarrow{u}}{p} \\ -\frac{\rho}{p} \\ \frac{\rho \overrightarrow{B}}{p} \\ 0 \end{pmatrix}$$

On voit clairement qu'il y a un problème sur la dernière composante. La jacobienne de ce vecteur n'a aucune chance d'être inversible. Pour que cela soit le cas, il faut d'une manière ou d'une autre prendre en compte ψ dans la définition de l'entropie mathématique. C'est une question très délicate. Imaginons que nous cherchions une entropie mathématique de la forme

$$S = -\frac{\rho s}{\gamma - 1} + S_\psi(U)$$

associée à un flux du type

$$\overrightarrow{G} = -\frac{\rho s}{\gamma - 1}\overrightarrow{u} + \overrightarrow{g}_\psi$$

Ici, les termes que nous introduisons ne sont pas dictés par la physique. En conséquence, il n'y a pas de conflit entre elle et la condition de compatibilité. Malgré tout, cela ne nous autorise pas à faire tout ce que nous voulons. Pour que le système soit symétrisable selon l'analyse précédente, on peut montrer par linéarité qu'il faut que le couple entropie-flux artificiel vérifie une relation de compatibilité propre :

$$\nabla_U S_\psi \nabla_U \overrightarrow{F} = \nabla_U \overrightarrow{g}_\psi$$

Le problème étant trop général, on peut aussi imposer la forme de ce flux d'entropie supplémentaire. Pour simplifier les écritures, on peut par exemple choisir

$$\overrightarrow{g}_\psi = S_\psi \overrightarrow{u}$$

de sorte qu'on puisse avoir dans les zones régulières :

$$\frac{\partial S}{\partial t} + \overrightarrow{\nabla} \cdot (S\overrightarrow{u}) = 0$$

La relation de compatibilité donne alors p équations pour p dérivées partielles de S_ψ à déterminer. À partir de là, on peut espérer construire S_ψ. Pour simplifier les expressions, il faut sans doute procéder dans la base des variables physiques V, puisqu'on montre facilement que la relation de compatibilité s'y exprime de la même manière :

$$\nabla_V S_\psi \nabla_V \overrightarrow{F} = \nabla_V \overrightarrow{g}_\psi$$

De plus, il faut préciser que les matrices à employer pour $\nabla_V \overrightarrow{F}$ ne sont pas les jacobiennes du système classique avec les termes sources de Godunov, mais celles prenant de plus en compte les termes rajoutés par Dedner et al. [35] permettant de vérifier l'invariance Galiléenne. Cependant, le problème est rendu encore plus complexe par la nécessité pour S_ψ d'être une quantité strictement convexe, c'est-à-dire de posséder une matrice Hessienne symétrique définie positive. Ceci est nécessaire pour que lorsque l'égalité n'est pas atteinte, l'entropie artificielle que nous rajoutons agisse dans le même sens que l'entropie convexe classique. De cette manière, il est cohérent d'imposer à un schéma de vérifier un équivalent discret de l'inégalité que nous construirions. Ces complications font que nous n'avons pas encore pu aboutir à la généralisation du travail de Barth, pour les équations de la MHD augmentées du *divergence cleaning* hyperbolique.

2.3.4 Quelques comparaisons avec les équations d'Euler

Nous avons dit que les équations de la MHD idéale sont une généralisation des équations d'Euler en mécanique des fluides, dans le sens où ces dernières y sont incluses. Or les équations d'Euler possèdent quelques propriétés remarquables, qui peuvent être utilisées pour construire des méthodes particulières. Voyons ici si la MHD idéale les vérifie également, ce qui nous permettrait d'adapter ces constructions à nos problèmes.

Homogénéité

Les équations d'Euler sont homogènes de degré 1, ce qui signifie que leurs flux vérifient :

$$\frac{\partial\left(\overrightarrow{F}_{\text{Euler}}\right)}{\partial U}(U)U = \overrightarrow{F}_{\text{Euler}}(U)$$

En particulier, si on considère une direction $\overrightarrow{n}$ quelconque, on peut projeter l'égalité précédente :

$$\frac{\partial\left(\overrightarrow{F}_{\text{Euler}}\cdot\overrightarrow{n}\right)}{\partial U}U = \overrightarrow{F}_{\text{Euler}}\cdot\overrightarrow{n}U \tag{2.3.9}$$

Notons à présent $\overrightarrow{F}$ les flux de la MHD idéale avec correction de la divergence et écrivons les deux termes de (2.3.9) pour savoir s'ils sont égaux ou non. Commençons par le plus simple, c'est-à-dire le second membre :

$$F_n(U) = \begin{pmatrix} \rho u_n \\ \rho u_n\overrightarrow{u} + \left(p + \dfrac{\overrightarrow{B}^2}{2}\right)\overrightarrow{n} - B_n\overrightarrow{B} \\ \left(E + p + \dfrac{\overrightarrow{B}^2}{2}\right)u_n - \left(\overrightarrow{u}\cdot\overrightarrow{B}\right)B_n \\ u_n\overrightarrow{B} - B_n\overrightarrow{u} + \psi\overrightarrow{n} \\ c_h^2 B_n \end{pmatrix} \tag{2.3.10}$$

Passons à présent au produit de la jacobienne projetée par U. Cette jacobienne doit être celle des flux non modifiés, c'est-à-dire sans rajouter les termes source de Powell. Autrement dit, par rapport à ce que nous avons écrit auparavant, nous devons travailler avec la matrice $A_n^{C,0}$ augmentée des termes dûs à la correction, et non avec $A_n^{C,2}$. Le calcul donne au final :

$$\frac{\partial F_n}{\partial U}U = \begin{pmatrix} \rho u_n \\ \rho u_n\overrightarrow{u} + \left(p + \dfrac{\overrightarrow{B}^2}{2}\right)\overrightarrow{n} + (2-\gamma)\dfrac{\overrightarrow{B}^2}{2}\overrightarrow{n} - 2B_n\overrightarrow{B} \\ \left(E + p + \dfrac{\overrightarrow{B}^2}{2}\right)u_n + (2-\gamma)\dfrac{\overrightarrow{B}^2}{2}u_n - 2B_n\overrightarrow{u}\cdot\overrightarrow{B} \\ u_n\overrightarrow{B} - B_n\overrightarrow{u} + \psi\overrightarrow{n} \\ c_h^2 B_n \end{pmatrix} \neq F_n(U)$$

On constate donc qu'à l'inverse des équations d'Euler, les termes d'origine magnétique font que (2.3.9) n'est pas respectée, autrement dit que le système de la MHD idéale n'est pas homogène de degré 1.

Mise sous forme quadratique

Un autre fait particulier concernant les équations d'Euler est qu'il est possible de les mettre sous forme quadratique à l'aide d'un changement de variable. Introduit pour la première fois par Roe ([77]), le jeu

de variables en question est appelé le vecteur paramètre (*parameter vector*). Voici son expression :

$$Z = \begin{pmatrix} \sqrt{\rho} \\ \sqrt{\rho}\,\overrightarrow{u} \\ \dfrac{E+p}{\sqrt{\rho}} \end{pmatrix} = \begin{pmatrix} Z_\rho \\ \overrightarrow{Z_u} \\ Z_E \end{pmatrix}$$

Les flux Eulériens formulés comme fonctions de Z deviennent ainsi :

$$\overrightarrow{F}_{\text{Euler}} \cdot \overrightarrow{n} = \begin{pmatrix} Z_\rho \overrightarrow{Z_u} \cdot \overrightarrow{n} \\ \overrightarrow{Z_u} \cdot \overrightarrow{n}\,\overrightarrow{Z_u} + \dfrac{\gamma-1}{\gamma}\left(Z_\rho Z_E - \dfrac{1}{2}\overrightarrow{Z_u}^2\right)\overrightarrow{n} \\ Z_E \overrightarrow{Z_u} \cdot \overrightarrow{n} \end{pmatrix}$$

Z a de plus été construit de manière à obtenir une expression simple vérifiant :

$$U = D(Z)Z \qquad \text{et} \qquad \overrightarrow{F} = \overrightarrow{R}(Z)Z$$

où les matrices $D(Z)$ et $R_i(Z)$ sont toutes linéaires en Z, et de telle manière que $D(Z)$ et $\overrightarrow{R}(Z)$ soient les jacobiennes respectives de U et de $\overrightarrow{F}$, soit :

$$D(Z) = \frac{\partial U}{\partial Z} \qquad \text{et} \qquad \overrightarrow{R}(Z) = \frac{\partial \overrightarrow{F}}{\partial Z}$$

Si nous essayons de construire un jeu de variables équivalent pour la MHD idéale, il semble judicieux de partir de celui-ci pour la partie hydrodynamique des équations. On remarque rapidement que la partie magnétique de l'équation de conservation de la quantité de mouvement, c'est-à-dire la pression magnétique et la tension magnétique, est quadratique en $\overrightarrow{B}$. Quant à la partie *divergence cleaning*, elle est linéaire en $\overrightarrow{B}$ comme en ψ et ne devrait donc pas poser de difficulté quel que soit le choix des variables, pour peu que celui-ci soit un minimum judicieux. Adoptons donc dans une première tentative le jeu de variables suivant :

$$Z = \begin{pmatrix} \sqrt{\rho} \\ \sqrt{\rho}\,\overrightarrow{u} \\ \dfrac{E + p + \frac{\overrightarrow{B}^2}{2}}{\sqrt{\rho}} \\ \overrightarrow{B} \\ \psi \end{pmatrix}$$

Les flux (2.3.10) s'écrivent alors :

$$F_n = \begin{pmatrix} Z_\rho \overrightarrow{Z_u} \cdot \overrightarrow{n} \\ \overrightarrow{Z_u} \cdot \overrightarrow{n}\,\overrightarrow{Z_u} + \left(\dfrac{\gamma-1}{\gamma}\left(Z_\rho Z_E - \dfrac{1}{2}\overrightarrow{Z_u}^2\right) + \dfrac{2-\gamma}{\gamma}\dfrac{1}{2}\overrightarrow{Z_B}^2\right)\overrightarrow{n} - \overrightarrow{Z_B} \cdot \overrightarrow{n}\,\overrightarrow{Z_B} \\ Z_E \overrightarrow{Z_u} \cdot \overrightarrow{n} - \left(\dfrac{\overrightarrow{Z_u}}{Z_\rho} \cdot \overrightarrow{Z_B}\right)\overrightarrow{Z_B} \cdot \overrightarrow{n} \\ \dfrac{1}{Z_\rho}\left(\overrightarrow{Z_u} \cdot \overrightarrow{n}\,\overrightarrow{Z_B} - \overrightarrow{Z_B} \cdot \overrightarrow{n}\,\overrightarrow{Z_u}\right) + Z_\psi \overrightarrow{n} \\ {c_h}^2 \overrightarrow{Z_B} \cdot \overrightarrow{n} \end{pmatrix}$$

Malheureusement, on s'aperçoit que non seulement certains termes ne sont pas quadratiques, mais également que les termes d'origine magnétique ne sont pas tous de même degré. Il semblerait même que cela doive être le cas quel que soit le choix fait pour $\overrightarrow{Z_B}$. Nous ne connaissons pas d'alternative.

En revanche, à défaut de pouvoir obtenir une formulation quadratique, il peut rester intéressant d'avoir une écriture sous forme polynomiale, même si de degré supérieur à 2 pour certains termes. En effet, il est plus simple de travailler avec des polynômes de fonctions qu'avec des fractions rationnelles, par exemple si on doit être amené à les intégrer... Pour cela, on remarque qu'il faut inclure les coefficients $\frac{1}{Z_\rho}$ dans une nouvelle définition des $\overrightarrow{Z_B}$, à l'instar de ce qui est fait dans [10]. On obtient alors le jeu de variables suivant :

$$Z = \begin{pmatrix} \sqrt{\rho} \\ \sqrt{\rho}\,\overrightarrow{u} \\ \dfrac{E + p + \frac{\overrightarrow{B}^2}{2}}{\sqrt{\rho}} \\ \dfrac{\overrightarrow{B}}{\sqrt{\rho}} \\ \psi \end{pmatrix}$$

Et les flux (2.3.10) s'expriment maintenant :

$$F_n = \begin{pmatrix} Z_\rho \overrightarrow{Z_B} \cdot \overrightarrow{n} \\ \overrightarrow{Z_u} \cdot \overrightarrow{n}\,\overrightarrow{Z_u} + \dfrac{\gamma - 1}{\gamma}\left(Z_\rho Z_E - \dfrac{1}{2}\overrightarrow{Z_u}^2\right)\overrightarrow{n} + \dfrac{2-\gamma}{\gamma}\dfrac{1}{2}\left(Z_\rho \overrightarrow{Z_B}\right)^2 \overrightarrow{n} - {Z_\rho}^2 \overrightarrow{Z_B} \cdot \overrightarrow{n}\,\overrightarrow{Z_B} \\ Z_E \overrightarrow{Z_u} \cdot \overrightarrow{n} - \left(\overrightarrow{Z_u} \cdot \overrightarrow{Z_B}\right) Z_\rho \overrightarrow{Z_B} \cdot \overrightarrow{n} \\ \overrightarrow{Z_u} \cdot \overrightarrow{n}\,\overrightarrow{Z_B} - \overrightarrow{Z_B} \cdot \overrightarrow{n}\,\overrightarrow{Z_u} + Z_\psi \overrightarrow{n} \\ {c_h}^2 Z_\rho \overrightarrow{Z_B} \cdot \overrightarrow{n} \end{pmatrix}$$

Cette expression est bel et bien polynomiale. Tous les termes d'origine magnétique des équations de conservation de la quantité de mouvement et de l'énergie sont de degré 4.

2.4 Adimensionnement

Revenons au système d'équations complet, prenant en compte les phénomènes diffusifs et dissipatifs :

$$\begin{aligned}
\frac{\partial \rho}{\partial t} + \overrightarrow{\nabla} \cdot (\rho \overrightarrow{u}) - \overrightarrow{\nabla} \cdot \left(D \overrightarrow{\nabla} \rho\right) &= 0 \\
\frac{\partial (\rho \overrightarrow{u})}{\partial t} + \overrightarrow{\nabla} \cdot \left(\rho \overrightarrow{u}\,\overrightarrow{u}^t + (p + \frac{\overrightarrow{B}^2}{2}) Id - \overrightarrow{B}\,\overrightarrow{B}^t\right) - \overrightarrow{\nabla} \cdot \overrightarrow{\tau} &= \overrightarrow{0} \\
\frac{\partial E}{\partial t} + \overrightarrow{\nabla} \cdot \left((E + p + \frac{\overrightarrow{B}^2}{2}) \overrightarrow{u} - (\overrightarrow{u} \cdot \overrightarrow{B}) \overrightarrow{B}\right) & \\
- \overrightarrow{\nabla} \cdot \left(\overrightarrow{\tau}\,\overrightarrow{u} - \kappa \overrightarrow{\nabla} T + \frac{\eta}{\mu_0}\left(\overrightarrow{\nabla} \cdot \overrightarrow{B}\,\overrightarrow{B}^t - \overrightarrow{\nabla} \frac{\overrightarrow{B}^2}{2}\right)\right) &= 0 \\
\frac{\partial \overrightarrow{B}}{\partial t} + \overrightarrow{\nabla} \cdot \left(\overrightarrow{u}\,\overrightarrow{B}^t - \overrightarrow{B}\,\overrightarrow{u}^t\right) + \overrightarrow{\nabla} \psi - \overrightarrow{\nabla} \cdot \left(\frac{\eta}{\mu_0} \overrightarrow{\nabla} \overrightarrow{B}\right) &= \overrightarrow{0} \\
\frac{\partial \psi}{\partial t} + {c_h}^2 \overrightarrow{\nabla} \cdot \overrightarrow{B} &= -\frac{{c_h}^2}{{c_p}^2} \psi
\end{aligned}$$

Il nous sera utile, lorsque nous nous attacherons à résoudre des problèmes de MHD non idéale, de travailler avec des quantités adimensionnées. Tout comme en mécanique des fluides, certains nombres donneront une estimation des régimes dans lesquels nous nous trouverons (prépondérance des effets résistifs, ou visqueux, ou advectifs, etc...). Pour ce faire, nous allons introduire diverses quantités sans dimension élémentaires.

$$\overline{t} = \frac{t}{t_0} \quad \overrightarrow{\overline{x}} = \frac{\overrightarrow{x}}{L_0} \quad \overline{\rho} = \frac{\rho}{\rho_0} \quad \overrightarrow{\overline{u}} = \frac{\overrightarrow{u}}{u_0} \quad \overline{E} = \frac{E}{E_0} \quad \overrightarrow{\overline{B}} = \frac{\overrightarrow{B}}{B_0}$$

$$\overline{p} = \frac{p}{p_0} \quad \overline{T} = \frac{T}{T_0} \quad \overline{D} = \frac{D}{D_0} \quad \overline{\nu} = \frac{\nu}{\nu_0} \quad \overline{\kappa} = \frac{\kappa}{\kappa_0} \quad \overline{\eta} = \frac{\eta}{\eta_0}$$

Commençons, dans l'ordre, par l'équation de continuité :

$$\partial_t \rho + \overrightarrow{\nabla} \cdot (\rho \overrightarrow{u}) - \overrightarrow{\nabla} \cdot (D \overrightarrow{\nabla} \rho) = 0$$

$$\Rightarrow \quad \frac{\rho_0}{t_0} \partial_{\overline{t}} \overline{\rho} + \frac{\rho_0 u_0}{L_0} \overrightarrow{\nabla} \cdot (\overline{\rho} \overrightarrow{\overline{u}}) - \frac{\rho_0 D_0}{L_0^2} \overrightarrow{\nabla} \cdot (\overline{D} \overrightarrow{\nabla} \overline{\rho}) = 0$$

$$\Rightarrow \quad \frac{t_u}{t_0} \partial_{\overline{t}} \overline{\rho} + \overrightarrow{\nabla} \cdot (\overline{\rho} \overrightarrow{\overline{u}}) - \frac{D_0}{L_0 u_0} \overrightarrow{\nabla} \cdot (\overline{D} \overrightarrow{\nabla} \overline{\rho}) = 0$$

$$\Rightarrow \quad S \partial_{\overline{t}} \overline{\rho} + \overrightarrow{\nabla} \cdot (\overline{\rho} \overrightarrow{\overline{u}}) - \frac{1}{Pe_\rho} \overrightarrow{\nabla} \cdot (\overline{D} \overrightarrow{\nabla} \overline{\rho}) = 0$$

avec $t_u = \dfrac{L_0}{u_0}$, le nombre de Péclet pour la diffusion de matière $Pe_\rho = \dfrac{L_0 u_0}{D_0}$, et le nombre de Strouhal $S = \dfrac{t_u}{t_0}$. L'équation de conservation de la quantité de mouvement devient quant à elle :

$$\partial_t (\rho \overrightarrow{u}) + \overrightarrow{\nabla} \cdot (\rho \overrightarrow{u} \overrightarrow{u}^t - \overrightarrow{B} \overrightarrow{B}^t) + \overrightarrow{\nabla} (p + \frac{\overrightarrow{B}^2}{2})$$

$$- \overrightarrow{\nabla} \cdot \left(\nu (\overrightarrow{\nabla} \overrightarrow{u} + \overrightarrow{\nabla} \overrightarrow{u}^t - \frac{2}{3} \overrightarrow{\nabla} \cdot \overrightarrow{u} \, Id) \right) = \overrightarrow{0}$$

$$\Rightarrow \quad \frac{\rho_0 u_0}{t_0} \partial_{\overline{t}} (\overline{\rho} \overrightarrow{\overline{u}}) + \frac{\rho_0 u_0^2}{L_0} \overrightarrow{\nabla} \cdot (\overline{\rho} \overrightarrow{\overline{u}} \overrightarrow{\overline{u}}^t) + \frac{p_0}{L_0} \overrightarrow{\nabla} \overline{p} + \frac{B_0^2}{L_0} \overrightarrow{\nabla} \cdot (\overrightarrow{\overline{B}} \overrightarrow{\overline{B}}^t - \frac{\overrightarrow{\overline{B}}}{2})$$

$$- \frac{\nu_0 u_0}{L_0^2} \overrightarrow{\nabla} \cdot \left(\overline{\nu} (\overrightarrow{\nabla} \overrightarrow{\overline{u}} + \overrightarrow{\nabla} \overrightarrow{\overline{u}}^t - \frac{2}{3} \overrightarrow{\nabla} \cdot \overrightarrow{\overline{u}}) \right) = \overrightarrow{0}$$

$$\Rightarrow \quad \frac{t_u}{t_0} \partial_{\overline{t}} (\overline{\rho} \overrightarrow{\overline{u}}) + \overrightarrow{\nabla} \cdot (\overline{\rho} \overrightarrow{\overline{u}} \overrightarrow{\overline{u}}^t) - \frac{B_0^2}{\rho_0 u_0^2} \overrightarrow{\nabla} \cdot (\overrightarrow{\overline{B}} \overrightarrow{\overline{B}}^t) + \frac{p_0}{\rho_0 u_0^2} \overrightarrow{\nabla} \overline{p} + \frac{B_0^2}{\rho_0 u_0^2} \overrightarrow{\nabla} \left(\frac{\overrightarrow{\overline{B}}^2}{2} \right)$$

$$- \frac{\nu_0}{\rho_0 u_0 L_0} \overrightarrow{\nabla} \cdot \left(\overline{\nu} (\overrightarrow{\nabla} \overrightarrow{\overline{u}} + \overrightarrow{\nabla} \overrightarrow{\overline{u}}^t - \frac{2}{3} \overrightarrow{\nabla} \cdot \overrightarrow{\overline{u}} \, Id) \right) = \overrightarrow{0}$$

$$\Rightarrow \quad S \partial_{\overline{t}} (\overline{\rho} \overrightarrow{\overline{u}}) + \overrightarrow{\nabla} \cdot (\overline{\rho} \overrightarrow{\overline{u}} \overrightarrow{\overline{u}}^t) + \frac{p_0}{\rho_0 u_0^2} \overrightarrow{\nabla} \overline{p} + \frac{1}{Al^2} \overrightarrow{\nabla} \cdot \left(\frac{\overrightarrow{\overline{B}}^2}{2} Id - \overrightarrow{\overline{B}} \overrightarrow{\overline{B}}^t \right)$$

$$- \frac{1}{Re} \overrightarrow{\nabla} \cdot \left(\overline{\nu} (\overrightarrow{\nabla} \overrightarrow{\overline{u}} + \overrightarrow{\nabla} \overrightarrow{\overline{u}}^t - \frac{2}{3} \overrightarrow{\nabla} \cdot \overrightarrow{\overline{u}} \, Id) \right) = \overrightarrow{0}$$

où nous avons défini le nombre d'Alfvén $Al = \frac{\sqrt{\rho_0}u_0}{B_0}$ (rapport entre la vitesse caractéristique de l'écoulement et la vitesse caractéristique des ondes d'Alfvén), et le nombre de Reynolds hydrodynamique $Re = \frac{\rho_0 u_0 L_0}{\nu_0}$ (identique à celui utilisé en Mécanique des fluides, mesurant l'importance des effets visqueux). Nous pouvons à présent passer à la dernière loi de conservation, celle régissant l'évolution de l'énergie totale :

$$\partial_t E + \overrightarrow{\nabla}\cdot\left((E+p+\frac{\overrightarrow{B}^2}{2})\overrightarrow{u} - (\overrightarrow{u}\cdot\overrightarrow{B})\overrightarrow{B}\right) - \overrightarrow{\nabla}\cdot(\kappa\overrightarrow{\nabla}T)$$

$$+\overrightarrow{\nabla}\cdot\left(\frac{\eta}{\mu_0}(\overrightarrow{\nabla}\times\overrightarrow{B})\times\overrightarrow{B}\right) - \overrightarrow{\nabla}\cdot\left(\overrightarrow{\tau}\overrightarrow{u}\right) = 0$$

$$\Rightarrow \quad \frac{E_0}{t_0}\partial_{\overline{t}}\overline{E} + \frac{E_0 u_0}{L_0}\overrightarrow{\nabla}\cdot(\overline{E}\,\overrightarrow{\overline{u}}) + \frac{p_0 u_0}{L_0}\overrightarrow{\nabla}\cdot(\overline{p}\,\overrightarrow{\overline{u}}) + \frac{B_0^2 u_0}{L_0}\overrightarrow{\nabla}\cdot\left(\frac{\overrightarrow{\overline{B}}^2}{2}\overrightarrow{\overline{u}} - \overrightarrow{\overline{B}}(\overrightarrow{\overline{u}}\cdot\overrightarrow{\overline{B}})\right)$$

$$-\frac{\kappa_0 T_0}{L_0^2}\overrightarrow{\nabla}\cdot\left(\overline{\kappa}\overrightarrow{\nabla}\overline{T}\right) + \frac{\eta_0 B_0^2}{\mu_0 L_0^2}\overrightarrow{\nabla}\cdot\left(\overline{\eta}(\overrightarrow{\nabla}\times\overrightarrow{\overline{B}})\times\overrightarrow{\overline{B}}\right) - \frac{\nu_0 u_0^2}{L_0^2}\overrightarrow{\nabla}\cdot\left(\overrightarrow{\overline{\tau}}\overrightarrow{\overline{u}}\right) = 0$$

$$\Rightarrow \quad S\partial_{\overline{t}}\overline{E} + \overrightarrow{\nabla}\cdot(\overline{E}\,\overrightarrow{\overline{u}}) + \frac{p_0}{E_0}\overrightarrow{\nabla}\cdot(\overline{p}\,\overrightarrow{\overline{u}}) + \frac{{B_0}^2}{E_0}\overrightarrow{\nabla}\cdot\left(\frac{\overrightarrow{\overline{B}}^2}{2}\overrightarrow{\overline{u}}\right) - \frac{B_0^2}{E_0}\overrightarrow{\nabla}\cdot\left(\overrightarrow{\overline{B}}(\overrightarrow{\overline{u}}\cdot\overrightarrow{\overline{B}})\right)$$

$$-\frac{\kappa_0 T_0}{L_0 u_0 E_0}\overrightarrow{\nabla}\cdot\left(\overline{\kappa}\overrightarrow{\nabla}\overline{T}\right) + \frac{\eta_0 B_0^2}{\mu_0 L_0 u_0 E_0}\overrightarrow{\nabla}\cdot\left(\overline{\eta}(\overrightarrow{\nabla}\times\overrightarrow{\overline{B}})\times\overrightarrow{\overline{B}}\right) - \frac{\nu_0 u_0}{L_0 E_0}\overrightarrow{\nabla}\cdot\left(\overrightarrow{\overline{\tau}}\overrightarrow{\overline{u}}\right) = 0$$

$$\Rightarrow \quad S\partial_{\overline{t}}\overline{E} + \overrightarrow{\nabla}\cdot(\overline{E}\,\overrightarrow{\overline{u}}) + \frac{p_0}{E_0}\overrightarrow{\nabla}\cdot(\overline{p}\,\overrightarrow{\overline{u}}) + \frac{{B_0}^2}{E_0}\overrightarrow{\nabla}\cdot\left(\frac{\overrightarrow{\overline{B}}^2}{2}\overrightarrow{\overline{u}} - \overrightarrow{\overline{B}}(\overrightarrow{\overline{u}}\cdot\overrightarrow{\overline{B}})\right)$$

$$-\frac{\rho_0 C_{p,0} T_0}{E_0 Pe}\overrightarrow{\nabla}\cdot\left(\overline{\kappa}\overrightarrow{\nabla}\overline{T}\right) + \frac{B_0^2}{E_0 Rm}\overrightarrow{\nabla}\cdot\left(\overline{\eta}(\overrightarrow{\nabla}\times\overrightarrow{\overline{B}})\times\overrightarrow{\overline{B}}\right) - \frac{1}{Re}\frac{\rho_0 u_0^2}{E_0}\overrightarrow{\nabla}\cdot\left(\overrightarrow{\overline{\tau}}\overrightarrow{\overline{u}}\right) = 0$$

Nous introduisons ici deux nombres sans dimension supplémentaires : le nombre de Péclet classique $Pe = \frac{L_0 u_0 \rho_0 C_{p,0}}{\kappa_0}$, et le nombre de Reynolds magnétique $Rm = \frac{\mu_0 L_0 u_0}{\eta_0}$, qui est une mesure du rapport d'intensité entre la convection et la diffusion des charges. C_p est la capacité thermique à pression constante du plasma (voir la section 2.1.3 sur la thermodynamique). Pour l'équation d'évolution du champ magnétique, le même procédé s'applique et mène aux calculs suivants :

$$\partial_t\overrightarrow{B} + \overrightarrow{\nabla}\cdot\left(\overrightarrow{B}\overrightarrow{u}^t - \overrightarrow{u}\overrightarrow{B}^t\right) + \overrightarrow{\nabla}\psi + \overrightarrow{\nabla}\times\left(\frac{\eta}{\mu_0}\overrightarrow{\nabla}\times\overrightarrow{B}\right) = \overrightarrow{0}$$

$$\Rightarrow \quad \frac{B_0}{t_0}\partial_{\overline{t}}\overrightarrow{\overline{B}} + \frac{B_0 u_0}{L_0}\overrightarrow{\nabla}\cdot\left(\overrightarrow{\overline{B}}\overrightarrow{\overline{u}}^t - \overrightarrow{\overline{u}}\overrightarrow{\overline{B}}^t\right) + \frac{\psi_0}{L_0}\overrightarrow{\nabla}\overline{\psi} + \frac{\eta_0 B_0}{\mu_0 L_0^2}\overrightarrow{\nabla}\times\left(\overline{\eta}\overrightarrow{\nabla}\times\overrightarrow{\overline{B}}\right) = \overrightarrow{0}$$

$$\Rightarrow \quad S\partial_{\overline{t}}\overrightarrow{\overline{B}} + \overrightarrow{\nabla}\cdot\left(\overrightarrow{\overline{B}}\overrightarrow{\overline{u}}^t - \overrightarrow{\overline{u}}\overrightarrow{\overline{B}}^t\right) + \frac{\psi_0}{B_0 u_0}\overrightarrow{\nabla}\overline{\psi} + \frac{\eta_0}{\mu_0 u_0 L_0}\overrightarrow{\nabla}\times\left(\overline{\eta}\overrightarrow{\nabla}\times\overrightarrow{\overline{B}}\right) = \overrightarrow{0}$$

$$\Rightarrow \quad S\partial_{\overline{t}}\overrightarrow{\overline{B}} + \overrightarrow{\nabla}\cdot\left(\overrightarrow{\overline{B}}\overrightarrow{\overline{u}}^t - \overrightarrow{\overline{u}}\overrightarrow{\overline{B}}^t\right) + \frac{\psi_0}{B_0 u_0}\overrightarrow{\nabla}\overline{\psi} + \frac{1}{Rm}\overrightarrow{\nabla}\times\left(\overline{\eta}\overrightarrow{\nabla}\times\overrightarrow{\overline{B}}\right) = \overrightarrow{0}$$

Finalement, l'équation de correction de la divergence est simple à adimensionner :

$$\partial_t \psi + c_h^2 \overrightarrow{\nabla} \cdot \overrightarrow{B} = -\frac{c_h^2}{c_p^2}\psi$$

$$\Rightarrow \quad \frac{\psi_0}{t_0}\partial_{\overline{t}}\overline{\psi} + c_h^2 \frac{B_0}{L_0}\overrightarrow{\nabla} \cdot \overrightarrow{B} = -\frac{c_h^2}{c_p^2}\psi_0\overline{\psi}$$

$$\Rightarrow \quad \frac{L_0\psi_0}{B_0 t_0}\partial_{\overline{t}}\overline{\psi} + c_h^2 \overrightarrow{\nabla} \cdot \overrightarrow{B} = -\frac{c_h^2}{c_p^2}\frac{\psi_0 L_0}{B_0}\overline{\psi}$$

Nous pouvons à présent faire le choix de notre modélisation. La façon classique de procéder est d'utiliser la vitesse d'Alfvén $V_A = \frac{B_0}{\sqrt{\rho_0}}$ comme vitesse de référence, $u_0 = V_A$, ce qui entraîne que le nombre Al devient unitaire. Par souci d'alléger le système, nous choisissons aussi $t_0 = t_u$ de manière à ce que $S = 1$. Sur le même principe, il s'ensuit que pour simplifier au maximum les expressions précédentes, on choisit $p_0 = \rho_0 u_0^2 = B_0^2$, $E_0 = p_0 = \rho_0 u_0^2 = B_0^2$, $T_0 = \frac{E_0}{\rho_0 C_{p,0}}$ et finalement $\psi_0 = B_0 u_0$. De cette manière, le paramètre β_0 du plasma, souvent utilisé pour décrire le régime du plasma (mouvement dominé par les forces de pression ou le champ magnétique), est égal à :

$$\beta_0 = \frac{2p_0}{{B_0}^2} = 2$$

Mais cela reste un β_0 qualifiant l'état de référence, potentiellement éloigné du vrai β qui, de plus, varie au cours du temps. Nous pouvons maintenant exprimer la version finale de notre système sans dimensions. La notation "$\overline{x}$" est abandonnée puisque plus aucune ambigüité n'est présente. On ne travaille plus qu'avec des quantités sans dimensions.

$$\partial_t \rho + \overrightarrow{\nabla} \cdot (\rho \overrightarrow{u}) - \frac{1}{Pe_\rho}\overrightarrow{\nabla} \cdot \left(D\overrightarrow{\nabla}\rho\right) = 0 \tag{2.4.1a}$$

$$\begin{aligned}\partial_t (\rho \overrightarrow{u}) + \overrightarrow{\nabla} \cdot \left(\rho \overrightarrow{u}\,\overrightarrow{u}^t + \left(p + \frac{\overrightarrow{B}^2}{2}\right) Id - \overrightarrow{B}\,\overrightarrow{B}^t\right) \\ -\frac{1}{Re}\overrightarrow{\nabla} \cdot \left(\nu\left(\overrightarrow{\nabla}\overrightarrow{u} + \overrightarrow{\nabla}\overrightarrow{u}^t - \frac{2}{3}\overrightarrow{\nabla} \cdot \overrightarrow{u}\, Id\right)\right) = \overrightarrow{0}\end{aligned} \tag{2.4.1b}$$

$$\begin{aligned}\partial_t E + \overrightarrow{\nabla} \cdot \left((E + p + \frac{\overrightarrow{B}^2}{2})\overrightarrow{u} - \overrightarrow{B}(\overrightarrow{u} \cdot \overrightarrow{B})\right) \\ -\frac{1}{Pe}\overrightarrow{\nabla} \cdot \left(\kappa\overrightarrow{\nabla}T\right) + \frac{1}{Rm}\overrightarrow{\nabla} \cdot \left(\eta(\overrightarrow{\nabla} \times \overrightarrow{B}) \times \overrightarrow{B}\right) - \frac{1}{Re}\overrightarrow{\nabla} \cdot \left(\overrightarrow{\tau}\overrightarrow{u}\right) = 0\end{aligned} \tag{2.4.1c}$$

$$\partial_t \overrightarrow{B} + \overrightarrow{\nabla} \cdot \left(\overrightarrow{B}\overrightarrow{u}^t - \overrightarrow{u}\overrightarrow{B}^t\right) + \overrightarrow{\nabla}\psi + \frac{1}{Rm}\overrightarrow{\nabla} \times \left(\eta\overrightarrow{\nabla} \times \overrightarrow{B}\right) = \overrightarrow{0} \tag{2.4.1d}$$

$$\partial_t \psi + \frac{c_h^2}{V_A^2}\overrightarrow{\nabla} \cdot \overrightarrow{B} + \frac{c_h^2}{c_p^2}t_0\psi = 0 \tag{2.4.1e}$$

On peut citer quelques autres nombres sans dimension parfois utilisés pour caractériser les écoulements MHD dissipatifs. Un premier, qu'on emploie également parfois pour adimensionner les équations de Navier-Stokes, est le nombre de Prandtl :

$$Pr = \frac{C_{p,0}\nu_0}{\kappa_0}$$

Il peut être intéressant pour contrôler numériquement l'importance de la conduction thermique par rapport aux efforts visqueux, d'autant plus dans le contexte de la mécanique des fluides qu'on a l'égalité :

$$Pe = Re \times Pr$$

qui permet de factoriser tous les termes diffusifs de l'équation sur l'énergie de Navier-Stokes par $\frac{1}{Re}$. La même factorisation peut être effectuée en MHD, auquel cas on voit apparaître le nombre de Prandtl magnétique :

$$Pr_m = \frac{\mu_0 \nu_0}{\eta_0}$$

Un second nombre sans dimension utile dans certains domaines d'application (comme l'astrophysique), propre à la MHD celui-ci, est le nombre de Lundquist défini par :

$$Lu = \frac{\mu_0 L_0 V_A}{\eta_0}$$

que notre choix de modélisation $u_0 = V_A$ rend strictement équivalent au nombre de Reynolds magnétique Rm. Le dernier que nous citerons est le nombre de Hartmann, souvent employé en MHD des métaux liquides et des plasmas de laboratoire (notamment pour étudier certaines instabilités), qui permet d'estimer le rapport entre l'intensité des forces de Laplace et celle des contraintes visqueuses :

$$Ha = \frac{\sqrt{\mu_0} B_0 L_0}{\sqrt{\nu_0 \eta_0}}$$

On peut voir que celui-ci s'exprime en fonction des nombres de Reynolds hydrodynamique et magnétique, et en particulier, avec notre choix de modélisation :

$$Ha = \frac{V_A}{u_0} \sqrt{Re \times Rm} = \sqrt{Re \times Rm}$$

Chapitre 3

Schémas distribuant le résidu

Sommaire

Nous allons nous intéresser dans ce chapitre à la semi-discrétisation spatiale des équations par le biais de schémas distribuant le résidu. Les schémas en temps seront présentés au chapitre suivant. Il ne sera d'ailleurs principalement question que de la résolution des problèmes stationnaires, à quelques rares expressions près qui se présenteront surtout comme des remarques.

3.1 Présentation sur des problèmes scalaires

3.1.1 Mise en place du problème discret

Le problème de Cauchy (donnée d'une solution initiale assurant l'unicité de la solution recherchée) associé à une loi de conservation hyperbolique quelconque (les flux sont supposés de classe $\mathcal{C}^1$) s'écrit de manière forte dans $\mathbb{R}^d \times \mathbb{R}^+$:

$$\begin{aligned} &\frac{\partial u}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(u) = 0 \\ &u(\overrightarrow{x}, t = 0) = u_0(\overrightarrow{x}) \end{aligned} \tag{3.1.1}$$

Ce problème va nous servir de base pour présenter le prototype des schémas distribuant le résidu, que l'on retrouve le plus souvent sous les appellations anglo-saxonnes *Residual Distribution schemes* ($\mathcal{RD}$) ou *Fluctuation Splitting schemes* (FS). Pour une étude plus poussée de ces schémas, on peut conseiller entre autres les travaux de thèse de H. Paillère [70], M. Mezine [62], M. Ricchiuto [75], C. Tavé [90] puis A. Larat [59], ainsi que [33], qui offrent une présentation assez complète de tout ce qui a été fait sur ces méthodes depuis leurs débuts.

Le domaine, borné, est découpé de façon non structurée en éléments E a priori quelconques. Le résultat est un maillage noté Ω_h. De manière générale dans ce mémoire, l'ensemble Ω_h possède un double sens : il est l'ensemble des degrés de liberté nodaux M_i du maillage, mais aussi l'espace topologique défini comme la réunion des régions délimitées par chaque élément E, chacun de ces éléments contenant une partie des degrés de liberté M_i. Ainsi, selon l'emploi qu'on veut en faire, on notera l'appartenance $M_i \in \Omega_h$ ou $M_i \in T$ pour désigner un degré de liberté (et non un point quelconque à l'intérieur de la réunion des éléments ou d'un élément en particulier) et l'inclusion $T \subset \Omega_h$ pour faire référence à un élément. Au sens topologique, Ω_h ne diffère donc de Ω que par l'approximation des bords de ce dernier. La seule exigence portant sur Ω_h est la conformité des éléments (tout sommet pour un élément doit être un sommet pour tous les éléments voisins). Le principe même de la discrétisation est de n'étudier la solution qu'à travers un nombre fini de degrés de liberté. Pour cela, on doit spécifier la forme de la solution recherchée. Dans ce chapitre, on ne se consacre qu'à l'approximation en espace. Pour nous, il s'agira toujours d'une interpolation polynomiale de cette solution à l'aide de polynômes de Lagrange. La solution est donc semi-interpolée et toutes les discrétisations, qui ne se feront également ici qu'en espace, seront dites des semi-discrétisations. D'autres types d'interpolation seraient envisageables et nécessiteraient de revisiter les concepts et les méthodes que nous allons exposer (voir par exemple [8]). L'interpolation de Lagrange ne représente la solution qu'à l'aide des valeurs de la solution en certains points du maillage, dont peuvent faire partie les sommets des éléments, qu'on appellera des noeuds. C'est pour cela que nous parlions de degrés de liberté nodaux, car les seuls degrés de liberté seront les noeuds du maillage : après partionnement approximatif du domaine, le maillage est donc décrit en rajoutant les noeuds qui servent de support à l'interpolation. L'interpolation de Hermite, qui fait appel aux dérivées de la solution en ces mêmes noeuds, pourrait s'appuyer sur le même maillage mais en faisant simplement intervenir des degrés de liberté de nature différente.

Notre première phase dans la résolution numérique de (3.1.1) est donc l'interpolation en espace. Soit φ_i le polynôme de Lagrange associé à un degré de liberté nodal M_i quelconque, alors la solution approchée

u_h se décompose ainsi :

$$\forall t \in \mathbb{R}^+,\ u_i(t) := u(\overrightarrow{x}_i, t) = u(\overrightarrow{x}(M_i), t) \text{ et } u_h(\overrightarrow{x}, t) = \sum_{M_i \in \Omega_h} u_i(t)\varphi_i(\overrightarrow{x}) \tag{3.1.2}$$

De plus, beaucoup de phénomènes physiques auxquels il est possible de se confronter, et notamment en MHD, autorisent des discontinuités. Or la formulation forte supposant une certaine régularité sur la solution (au moins différentiable), il est nécessaire de passer à une formulation plus permissive qui repose sur une écriture intégrée du problème. On autorise ainsi la solution à être discontinue sur tout ensemble de mesure nulle, comme n'importe quelle hypersurface.

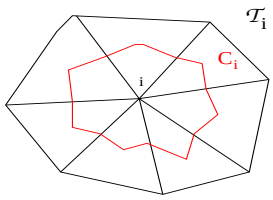


FIGURE 3.1 – Volume de contrôle C_i associé au noeud i et $\mathcal{T}_i$ la réunion des éléments contenant i

A titre d'exemple, plaçons-nous à présent sur un domaine de dimension $d = 2$ borné, auquel on a appliqué une triangulation Ω_h. On suppose également qu'on a adjoint des conditions aux limites adéquates. Comme dans les méthodes de type Volumes Finis, on définit autour de chaque degré de liberté M_i un volume de contrôle C_i (on parle aussi de cellule duale) de manière à recouvrir totalement le domaine discrétisé (on obtient un maillage dual). La figure 3.1 donne un exemple de cellule duale. L'intégration sur une de ces régions donne alors :

$$\begin{aligned}\int_{C_i} \frac{\partial u_h}{\partial t} dV &= -\int_{C_i} \overrightarrow{\nabla} \cdot \overrightarrow{F}(u_h) dV \\ &= -\sum_{T \subset \mathcal{T}_i} \int_{T \cap C_i} \overrightarrow{\nabla} \cdot \overrightarrow{F}(u_h) dV\end{aligned}$$

Le second membre peut se réécrire comme un bilan de flux sur les bords de C_i par la formule de Gauss, comme en Volumes Finis. Le principe de la distribution de résidu est cependant de considérer la contribution de chaque élément comme une part du résidu total ϕ^T sur l'élément T. On pourrait donc imaginer une **distribution** telle que :

$$\phi^T(u_h) = \int_T \overrightarrow{\nabla} \cdot \overrightarrow{F}(u_h) dV = \int_{\partial T} \overrightarrow{F}(u_h) \cdot \overrightarrow{n}\, d\partial T$$
$$\phi_i^T(u_h) = \beta_i^T \phi^T(u_h) = \int_{T \cap C_i} \overrightarrow{\nabla} \cdot \overrightarrow{F}(u_h) dV$$

Les coefficients β_i^T (ou directement les résidus partiels ϕ_i^T) définissent la distribution locale des signaux envoyés à chaque degré de liberté. Définir un schéma distribuant le résidu, c'est donc définir β_i (ou respectivement ϕ_i^T). Ensuite, pour résoudre le problème précédent, il faudrait préciser la manière dont on discrétise en temps et en espace le terme d'évolution instationnaire. Dans tout ce chapitre nous ne définirons pas la manière de procéder en temps, car celui-ci est dévolu à la discrétisation spatiale par les schémas $\mathcal{RD}$. Avant d'aller plus loin, clarifions les notations issues de (3.1.2) :

- $u(\overrightarrow{x}, t)$ et $u_h(\overrightarrow{x}, t)$ sont la solution exacte et son interpolation, i.e. des fonctions continues de l'espace et du temps,
- $\forall i \in \Omega_h$, $u_i = u_i(t) = u_h(\overrightarrow{x}_i, t)$ est une évaluation de u_h en un noeud M_i (qui se confond avec celle de u), qui est une fonction du temps seulement, si bien que sa variation en temps s'écrira $\dfrac{du_i}{dt}$

Donc, sans discrétiser le problème en temps, la semi-discrétisation en espace précédente donne le schéma suivant :

$$\int_{C_i} \frac{\partial u_h}{\partial t} dV = -\sum_{T \subset \mathcal{T}_i} \phi_i^T(u_h) = -\sum_{T \subset \mathcal{T}_i} \int_{T \cap C_i} \overrightarrow{\nabla} \cdot \overrightarrow{F}(u_h) dV \tag{3.1.3}$$

Ce n'est qu'un exemple de distribution, qui se base l'intégration directe de l'équation sur les volumes duaux que nous avons définis. Avec ce principe, si on change la constrution des cellules C_i, on change la distribution. Mais il y a bien d'autres manières de construire les résidus partiels, autrement qu'à partir de considérations géométriques.

Tout l'objet de ce chapitre va être de présenter la façon dont on peut construire ces signaux, quelles propriétés ils doivent ou peuvent vérifier, et comment élargir les définitions que nous donnerons à des configurations plus générales. On finira en exposant la manière dont sont prises en compte les conditions limites dans le cadre $\mathcal{RD}$. Durant toutes ces étapes, nous insistons une fois de plus sur le fait que nous ne ferons qu'expliquer la façon de semi-discrétiser les équations en espace avec les schémas distribuant le résidu. L'inclusion de la discrétisation temporelle du problème sera l'un des objets du prochain chapitre, et sera traité de façon indépendante. Ceci ne veut pas dire que nous n'évoquerons jamais l'instationnaire, seulement que la discrétisation en temps ne sera alors pas précisée. Mais avant toute chose, nous devons introduire d'autres notations.

Problèmes stationnaires

Le problème stationnaire s'écrit :

$$\overrightarrow{\nabla} \cdot \overrightarrow{F}(u) = 0$$

On peut difficilement résoudre ce système directement car il est en général non linéaire en u. Tout problème stationnaire étant un équilibre de phénomènes instationnaires, on cherche généralement la solution en partant d'une certaine configuration et en la faisant évoluer dans un certain temps de manière à ce que le régime permanent recherché prenne le temps de s'établir. Autrement dit, on cherche la limite du problème suivant lorsque le pseudo-temps τ tend vers $+\infty$:

$$\frac{\partial u}{\partial \tau} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(u) = 0 \tag{3.1.4}$$

Puisque τ et ∂_τ ne sont que des outils itératifs qui **imitent** le temps, sans signification physique aucune, et qu'une fois la limite atteinte on a $\partial_\tau u = 0$, il n'y a pas besoin de se soucier de la manière de discrétiser ce terme. On le discrétise donc à l'ordre 1 en pseudo-temps comme en espace (on applique un opérateur de *mass lumping* que nous verrons plus loin). Si on fait appel, comme ce sera toujours le cas, aux notations (3.1.2), on peut dire que tous les problèmes stationnaires sont généralement résolus via une méthode itérative du type :

$$u_i^{n+1} = u_i^n - \frac{\Delta\tau^n}{|C_i|} \sum_{T \subset \mathcal{T}_i} \phi_i^T(u_h) \tag{3.1.5}$$

où $\Delta\tau^n = \tau^{n+1} - \tau^n$ est le pas en pseudo-temps calculé à l'itération n et $|C_i|$ le volume de la cellule duale. La notion de partage du résidu ϕ^T implique une mise en oeuvre naturelle de ces schémas via un assemblage élément par élément des signaux envoyés à chaque degré de liberté. À convergence, (3.1.5) équivaut en chaque noeud i à l'équation suivante :

$$\sum_{T \subset \mathcal{T}_i} \phi_i^T(u_h) = 0 \tag{3.1.6}$$

De plus, si nous voulons que le schéma soit conservatif, la distribution doit respecter la propriété suivante :

$$\sum_{M_i \in T} \phi_i^T(u_h) = \sum_{M_i \in T} \beta_i^T(u_h)\phi^T(u_h) = \phi^T(u_h) \text{ i.e. } \sum_{M_i \in T} \beta_i^T(u_h) = 1 \tag{3.1.7}$$

Cette propriété est tellement souhaitable que c'est elle qui permet de décider si les signaux ϕ_i, ou les coefficients β_i, définissent un schéma $\mathcal{RD}$ ou non. Dans le cas précédent où nous définissions les résidus partiels géométriquement en intégrant sur $T \cap C_i$, cette propriété était assurée par le fait que le maillage dual recouvre tout le domaine, et qu'on ait donc $\sum_{M_i \in T} T \cap C_i = T$.

Problèmes instationnaires

Le problème instationnaire a déjà été formulé dès le départ (3.1.1). De manière analogue aux problèmes stationnaires, les résidus partiels que nous envoyons aux degrés de liberté sont des intégrales de l'équation entière, qui s'expriment forcément de la façon suivante, sur laquelle on reviendra dans le prochain chapitre :

$$\forall T \subset \Omega_h,\ \Phi_i^T(u_h) = \sum_{M_j \in T} m_{ij}^T(u_h)\frac{du_j}{dt} + \phi_i^T(u_h) \tag{3.1.8}$$

Définir un schéma $\mathcal{RD}$ instationnaire, c'est donc adjoindre un choix de matrice de masse m^T à la définition des signaux stationnaires ϕ_i sur chaque élément. Pour résoudre le problème, il faut ensuite discrétiser le tout en temps. Nous n'aborderons pas davantage ces sujets dans le présent chapitre, mais dans le suivant. Nous ne faisons que préciser les notations, et s'il peut arriver que nous parlions de problèmes instationnaires dans les prochaines pages, ce sera toujours dans un cadre général en prenant bien soin de ne préciser ni la discrétisation en temps ni la construction de la matrice de masse.

Tout comme en stationnaire, les résidus partiels doivent être soumis à une condition de conservation similaire à (3.1.7) qui s'écrit :

$$\sum_{M_i \in T} \Phi_i^T(u_h) = \Phi^T(u_h) = \int_T \frac{\partial u_h}{\partial t} dV + \int_{\partial T} \overrightarrow{F}(u_h) \cdot \overrightarrow{n}\, d\partial T \tag{3.1.9}$$

Une fois tous les éléments parcourus, le problème analogue à (3.1.6), à résoudre sur chaque degré de liberté i, devient :

$$\sum_{T \subset \mathcal{T}_i} \Phi_i^T(u_h) = 0 \tag{3.1.10}$$

Cependant, il manque la discrétisation en temps pour résoudre ce système. Ce point sera abordé au prochain chapitre. Quoiqu'il en soit, nous avons évoqué la discrétisation en espace des équations instationnaires et les notions qui s'y rapportent.

En conclusion

Le cas instationnaire étant le plus général, on en gardera les notations tout au long de ce mémoire :

- $\phi^T(u_h) = \int_T \overrightarrow{\nabla} \cdot \overrightarrow{F}(u_h) dV$ sera la **fluctuation** sur l'élément T
- les signaux $\phi_i^T(u_h)$ seront aussi appelés les fluctuations partielles
- $\Phi^T(u_h) = \int_T \left(\frac{\partial u_h}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(u_h) \right) dV$ sera le **résidu** sur l'élément T
- les signaux $\Phi_i^T(u_h)$ seront les résidus partiels

La confusion potentielle vient du fait que lorsqu'on étudie seulement des problèmes stationnaires, il arrive souvent qu'on assimile les mots "résidu" et "fluctuation", puisque les deux notions deviennent équivalentes ($\partial_t u_h = 0$). Enfin, et pour résumer, terminons en fixant la définition d'un schéma à distribution de résidu.

Définition 3.1.1. Un schéma distribuant le résidu ($\mathcal{RD}$) est défini par la donnée d'une loi de distribution vérifant la propriété de conservation (3.1.9). Sa mise en oeuvre se divise schématiquement en trois étapes :

1. Calcul du résidu $\Phi^T(u_h)$ sur chaque élément $T \subset \Omega_h$
2. Distribution des $\Phi_i^T(u_h)$ à chaque degré de liberté M_i de l'élément T selon le schéma choisi
3. Résolution de l'équation $\sum_{T \subset \mathcal{T}_i} \Phi_i^T = 0$ pour tout degré de liberté M_i du maillage

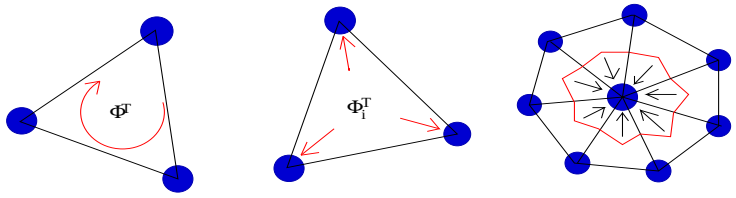


FIGURE 3.2 – Illustration des trois étapes de construction sur un maillage triangulé P_1

3.1.2 Généralisation et rapprochement avec d'autres méthodes

Lien avec les Volumes Finis - Cas 1D

Les premiers schémas $\mathcal{RD}$ sont dûs à Ni [65] et Roe [78]. C'est l'approche de ce dernier qui a par la suite mené aux schémas dont nous parlerons ici. Or Roe est également l'auteur éponyme d'un célèbre schéma Volumes Finis [77] un an plus tôt. Ce n'est donc pas un hasard si on trouve dès le départ des similitudes entre les Volumes Finis (VF) et les schémas $\mathcal{RD}$, comme il en est fait état dans [5] ou [34]. Nous allons commencer à présenter ce lien sur une équation scalaire non linéaire à résoudre dans un domaine de dimension $d = 1$ avec une méthode *upwind* (qui signifie qu'elle oriente la discrétisation selon la direction de propagation de l'information, qu'on appelle parfois le "sens du vent" par analogie avec la mécanique des fluides).

$$\begin{aligned} \frac{\partial u_h}{\partial t} + \frac{\partial f}{\partial x}(u_h) &= 0 \\ \Rightarrow \quad \frac{\partial u_h}{\partial t} + a(u_h)\frac{\partial u_h}{\partial x} &= 0 \quad \text{(forme quasi-linéaire)} \end{aligned}$$

avec $a(u_h) = f'(u_h)$. La solution u_h est supposée constante par cellule (segment) C_i et donc discontinue aux interfaces $x_{i\pm1/2}$ (voir 3.3). Le schéma Volumes Finis d'ordre 1 standard s'écrit donc :

$$|C_i|\frac{du_i}{dt} + H^-_{i+\frac{1}{2}}(u_i, u_{i+1}) - H^+_{i-\frac{1}{2}}(u_{i-1}, u_i) = 0 \tag{3.1.11}$$

L'idée du schéma *upwind* de Roe ([77]) est de définir les flux numériques H aux interfaces selon la direction de propogation de l'information (donnée par le signe de a ici). Ceci se traduit par l'expression suivante :

$$H^-_{i+1/2} = \frac{1}{2}\left(f_{i+1} + f_i - |a_{i+\frac{1}{2}}|(u_{i+1} - u_i)\right) \tag{3.1.12}$$

où on définit $a_{i+\frac{1}{2}}$ par une linéarisation conservative dite "de Roe" :

$$f_{i+1} - f_i = a_{i+\frac{1}{2}}(u_{i+1} - u_i) \tag{3.1.13}$$

Après quelques calculs, on aboutit finalement au schéma suivant :

$$|C_i|\frac{du_i}{dt} + a^-_{i+\frac{1}{2}}(u_{i+1} - u_i) + a^+_{i-\frac{1}{2}}(u_i - u_{i-1}) = 0 \tag{3.1.14}$$

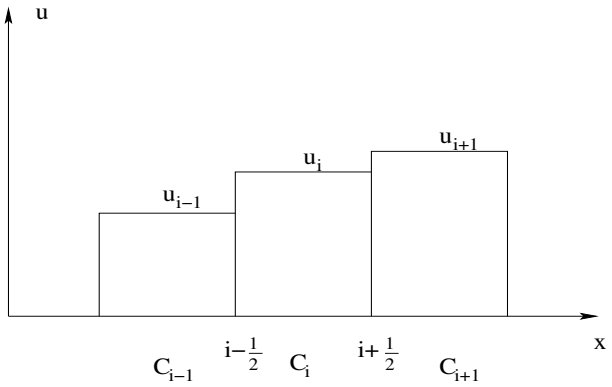


FIGURE 3.3 – Approximation constante par morceaux en Volumes Finis

où on a défini les parties positives et négatives :

$$a^+ = \frac{1}{2}(a + |a|) = \max(0, a) \text{ et } a^- = \frac{1}{2}(a - |a|) = \min(0, a)$$

En réalité, la linéarisation (3.1.13) est la contrainte pour que le schéma défini par (3.1.11) et (3.1.12) donne lieu au schéma *upwind* (3.1.14). La cellule C_i ne reçoit de l'information de droite que si a y est négatif et de gauche que si a y est positif. Passons à présent au formalisme $\mathcal{RD}$. La solution est toujours définie en chaque noeud x_i mais approchée cette fois de façon continue. On utilise en particulier une interpolation de Lagrange, linéaire pour cet exemple (éléments P_1), entre chaque noeud, ce qui fait que u_h n'est pas de classe $\mathcal{C}^1$ au passage des noeuds x_i (cf. figure 3.4).

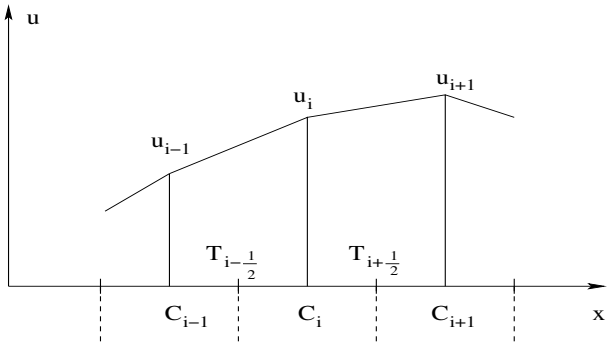


FIGURE 3.4 – Approximation polynomiale par élément en schémas aux résidus

En reprenant la description des méthodes aux résidus, le schéma doit s'écrire comme la somme des contributions des éléments $T_{i-\frac{1}{2}}$ et $T_{i+\frac{1}{2}}$:

$$\int_{C_i} \frac{\partial u_h}{\partial t} dV + \beta_i^{i+\frac{1}{2}} \phi^{i+\frac{1}{2}} + \beta_i^{i-\frac{1}{2}} \phi^{i-\frac{1}{2}}$$

Le volume de contrôle C_i est délimité par les milieux des éléments voisins, i.e. $C_i = \left[x_{i-\frac{1}{2}}; x_{i+\frac{1}{2}}\right]$, et les fluctuations sur chaque élément sont données par :

$$\phi^{i-\frac{1}{2}} = \int_{x_{i-1}}^{x_i} \frac{\partial}{\partial x} f(u_h) dx = f(u_i) - f(u_{i-1})$$

$$\phi^{i+\frac{1}{2}} = \int_{x_i}^{x_{i+1}} \frac{\partial}{\partial x} f(u_h) dx = f(u_{i+1}) - f(u_i)$$

Considérons la distribution effectuée par l'élément $T_{i+\frac{1}{2}}$. La propriété de conservation impose $\beta_i^{i+\frac{1}{2}} + \beta_{i+1}^{i+\frac{1}{2}} = 1$. Le schéma upwind de Roe peut alors être retrouvé en gardant la linéarisation de Roe pour définir le paramètre $a_{i+\frac{1}{2}}$ et en choisissant :

$$\beta_i^{i+\frac{1}{2}} = \frac{1}{2} - \frac{1}{2}\frac{a_{i+\frac{1}{2}}}{|a_{i+\frac{1}{2}}|} \text{ et } \beta_{i+1}^{i+\frac{1}{2}} = \frac{1}{2} + \frac{1}{2}\frac{a_{i+\frac{1}{2}}}{|a_{i+\frac{1}{2}}|}$$

On laisse au lecteur le soin de vérifier que l'on obtient bien l'équation suivante, proche de (3.1.14) :

$$\int_{C_i} \frac{\partial u_h}{\partial t} dV + a^-_{i+\frac{1}{2}}(u_{i+1} - u_i) + a^+_{i-\frac{1}{2}}(u_i - u_{i-1}) = 0$$

La différence tient visiblement dans la discrétisation du terme source instationnaire. Comme u_h est linéaire par morceau, et non plus constante par morceau, la version $\mathcal{RD}$ du schéma de Roe gagne un ordre de précision en espace, passant d'une précision d'ordre 1 à une précision d'ordre 2. Pour les problèmes stationnaires en revanche (et sans autre terme source), les deux schémas sont strictement équivalents à convergence.

Lien avec les Volumes Finis - Cas 2D

On peut généraliser ce résultat à tout type de schémas Volumes Finis, pour des dimensions supérieures et des systèmes d'équations. L'exemple type et dont l'écriture est la plus simple est un domaine bidimensionnel ne comportant que des triangles. On se restreint toujours à une interpolation de Lagrange linéaire par morceaux. Les normales utilisées dans les deux formalismes sont définies de manière différente (voir la figure 3.5 pour nos triangles P_1), mais ont toutes pour norme la longueur de l'arête à laquelle elles sont orthogonales, de sorte que leur somme soit nulle puisque le contour du triangle est fermé. Le volume de contrôle C_i est cependant le même dans les deux cas, délimité par les milieux des arêtes et les centres de gravité des triangles (figure 3.1). On donne l'écriture d'un schéma Volumes Finis standard d'ordre 1 :

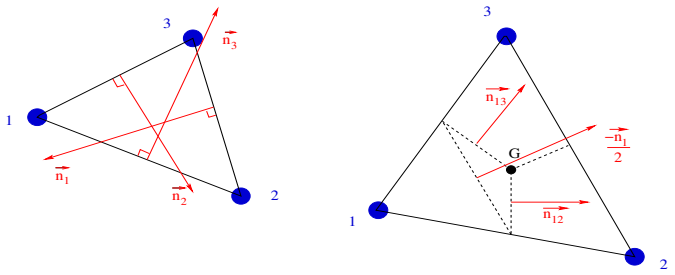


FIGURE 3.5 – Définition des normales locales à l'élément pour un triangle P_1

$$|C_i|\frac{du_i}{dt} + \int_{\partial C_i} H(\vec{n}) d\partial C_i = 0$$

$$\Leftrightarrow \quad |C_i|\frac{du_i}{dt} + \sum_{T \subset \mathcal{T}_i} \sum_{\substack{M_j \in T \\ j \neq i}} H(u_i, u_j, \vec{n}^T_{ij}) = 0$$

où le flux numérique H peut être donné sous diverses formes, dont la suivante :

$$H(u_i, u_j, \vec{n}^T_{ij}) = \frac{1}{2}\left(\vec{F}(u_i) + \vec{F}(u_j)\right) \cdot \vec{n}^T_{ij} - \frac{1}{2} Q(u_i, u_j)\,(u_i - u_j)$$

Dans cette écriture, Q est un paramètre de dissipation. Pour le cas des systèmes d'équations, par exemple les équations d'Euler, Q peut être la valeur absolue de la matrice du schéma de Roe ([77]). De nombreux

autres flux numériques existent, en particulier ceux qui s'attachent à la résolution exacte ou approchée des problèmes de Riemann aux interfaces. On parle dans ce cas des méthodes de type Godunov (le schéma de Godunov étant celui qui fait appel à la résolution exacte des problèmes de Riemann). La méthode des Volumes Finis pose comme condition que la méthode recouvre le cas d'une solution constante (si deux inconnues voisines sont égales, il n'y a pas de problème de Riemann à résoudre), i.e. pour toute direction $\overrightarrow{n}$:

$$H(u, u, \overrightarrow{n}) = \overrightarrow{F}(u) \cdot \overrightarrow{n}$$

De plus, les flux numériques doivent dans tous les cas être symétriques pour que le schéma soit conservatif, autrement dit :

$$H(u_i, u_j, \overrightarrow{n}_{ij}) = -H(u_j, u_i, -\overrightarrow{n}_{ij})$$

Pour le schéma cité en exemple, ceci signifie que Q est nécessairement symétrique ($Q(u_i, u_j) = Q(u_j, u_i)$). Dans un premier temps, si on cherche naïvement à construire un schéma $\mathcal{RD}$ à partir de la contribution de chaque élément au schéma VF, on obtient :

$$\forall T,\ \phi_i^T = \sum_{M_j \in T, j \neq i} H(u_i, u_j, \overrightarrow{n}_{ij}^T)$$

Si les flux se compensent deux à deux par symétrie sur chaque arête de C_i, alors à l'échelle d'un élément T chevauché par C_i on a :

$$\sum_{M_j \in T} \phi_j^T = \sum_{i=1}^{3} \sum_{j \neq i} H(u_i, u_j, \overrightarrow{n}_{ij}^T) = 0$$

C'est en réalité la définition même de la conservation pour les schémas Volumes Finis. La propriété de conservation (3.1.9) n'est pas respectée, donc il ne s'agit pas d'un schéma distribuant le résidu ! Maintenant, il est possible ([1]) de supprimer cette symétrie des flux tout en restant conservatif au niveau du domaine dans sa globalité, et même en préservant la contribution totale reçue par chaque degré de liberté avec le schéma VF (donc sans rien changer au résultat numérique). On utilise le fait que la somme des normales (telles qu'elles sont définies) sur le contour fermé C_i est nulle pour écrire :

$$\begin{aligned} |C_i| \frac{du_i}{dt} &= -\sum_{M_j \in \mathcal{T}_i} \sum_{T \subset (\mathcal{T}_i \cap \mathcal{T}_j)} H(u_i, u_j, \overrightarrow{n}_{ij}^T) + \sum_{M_j \in \mathcal{T}_i} \sum_{T \subset (\mathcal{T}_i \cap \mathcal{T}_j)} \overrightarrow{F}(u_i) \cdot \overrightarrow{n}_{ij}^T \\ &= -\sum_{M_j \in \mathcal{T}_i} \sum_{T \subset (\mathcal{T}_i \cap \mathcal{T}_j)} \left(H(u_i, u_j, \overrightarrow{n}_{ij}^T) - H(u_i, u_i, \overrightarrow{n}_{ij}^T) \right) \end{aligned}$$

On peut alors vouloir écrire un schéma dont la contribution apportée par un élément T est donnée par :

$$\phi_i^T = \sum_{j \neq i} \left(H(u_i, u_j, \overrightarrow{n}_{ij}^T) - \overrightarrow{F}(u_i) \cdot \overrightarrow{n}_{ij}^T \right)$$

En utilisant la correspondance entre les normales utilisées en VF et en $\mathcal{RD}$ (cf. 3.5), testons la conservation de ce nouveau schéma :

$$\begin{aligned} \phi^T &= \sum_{i=1}^{3} \sum_{j \neq i} \left(H(u_i, u_j, \overrightarrow{n}_{ij}^T) - \overrightarrow{F}(u_i) \cdot \overrightarrow{n}_{ij}^T \right) \\ &= \sum_{i=1}^{3} \left(0 - \overrightarrow{F}(u_i) \cdot \left(-\frac{\overrightarrow{n_i}}{2} \right) \right) \\ &= \frac{1}{2} \sum_{i=1}^{3} \overrightarrow{F}_i \cdot \overrightarrow{n}_i \end{aligned}$$

Il s'agit là du résidu $\phi^T(u_h)$ calculé par la règle des trapèzes :

$$\begin{aligned}\phi^T &= \frac{1}{2}(\overrightarrow{F}_1 + \overrightarrow{F}_2)\cdot(-\overrightarrow{n}_3) + \frac{1}{2}(\overrightarrow{F}_2 + \overrightarrow{F}_3)\cdot(-\overrightarrow{n}_1) + \frac{1}{2}(\overrightarrow{F}_3 + \overrightarrow{F}_1)\cdot(-\overrightarrow{n}_2) \\ &= \frac{1}{2}\left(\overrightarrow{F}_1\cdot(-\overrightarrow{n}_3 - \overrightarrow{n}_2) + \overrightarrow{F}_2\cdot(-\overrightarrow{n}_3 - \overrightarrow{n}_1) + \overrightarrow{F}_3\cdot(-\overrightarrow{n}_1 - \overrightarrow{n}_2)\right) = \frac{1}{2}\sum_{j=1}^{3}\overrightarrow{F}_j\cdot\overrightarrow{n}_j\end{aligned}$$

Ce schéma est donc conservatif, il s'agit effectivement de la formulation $\mathcal{RD}$ du schéma Volumes Finis de départ. Pour les problèmes instationnaires, on peut faire la même remarque que dans le cas monodimensionnel sur la discrétisation du terme source. La principale différence entre les schémas $\mathcal{RD}$ et les Volumes Finis est la continuité de la solution et des flux aux interfaces, ce qui est aussi vrai si on compare ces schémas aux méthodes de type Galerkin Discontinu.

Lien avec les Élements Finis

Restons dans le cadre d'éléments triangulaires de Lagrange P_1 et avec un problème hyperbolique scalaire. La première similitude dont nous avons parlée est le fait que la solution se décompose selon les fonctions de base comme en formulation Éléments Finis, plus précisément celle de Galerkin. La question est ensuite de savoir si un schéma distribuant le résidu peut s'inscrire dans le cadre d'une formulation variationnelle. Pour un problème stationnaire, la méthode de Galerkin consiste à formuler le problème au sens des distributions, en chaque degré de liberté M_i :

$$a(u_h,\varphi_i) = \int_{\Omega_h}\varphi_i\overrightarrow{\nabla}\cdot\overrightarrow{F}(u_h)dV = 0$$

L'espace des fonctions test φ_i est le même que l'espace d'approximation. Les fonctions de base de Lagrange ont la particularité d'être uniformément nulles au-delà des premiers voisins, ce qui nous permet d'écrire :

$$\begin{aligned}\forall M_i\in\Omega_h,\ \int_{\Omega_h}\varphi_i\overrightarrow{\nabla}\cdot\overrightarrow{F}(u_h)dV &= \int_{\mathcal{T}_i}\varphi_i\overrightarrow{\nabla}\cdot\overrightarrow{F}(u_h)dV \\ &= \sum_{T\subset\mathcal{T}_i}\int_T\varphi_i\overrightarrow{\nabla}\cdot\overrightarrow{F}(u_h)dV = 0\end{aligned}$$

Comparons ceci, pour le même problème, avec une des écritures possibles des schémas distribuant le résidu à l'échelle d'un élément T :

$$\phi_i^T := \beta_i^T\phi^T = \beta_i^T\int_T\overrightarrow{\nabla}\cdot\overrightarrow{F}(u_h)dV$$

Toutes ces quantités sont scalaires et le coefficient β_i^T est constant sur l'élément. On peut donc tout aussi bien écrire :

$$\phi_i^T = \int_T(\varphi_i + (\beta_i^T - \varphi_i))\overrightarrow{\nabla}\cdot\overrightarrow{F}(u_h)dV = \int_T(\varphi_i + \gamma_i^T)\overrightarrow{\nabla}\cdot\overrightarrow{F}(u_h)dV \tag{3.1.15}$$

C'est à ce stade qu'on voit que les schémas $\mathcal{RD}$, écrits sous cette forme, s'inscrivent dans une formulation variationnelle de type Éléments Finis avec des fonctions test modifiées par l'introduction d'une fonction de décentrement γ_i^T (par opposition au schéma de Galerkin classique qu'on qualifiera de centré). On modifie ainsi légèrement l'espace des fonctions test par rapport à l'espace d'approximation, et on doit alors parler d'une formulation de type Petrov-Galerkin des schémas $\mathcal{RD}$.

La méthode de Galerkin est réellement un schéma centré lorsque le problème est linéaire en u et pour des triangles P_1. Les normales que nous avons définies pour ces éléments permettent de simplifier l'écriture des gradients des fonctions de base, qui sont constants en P_1 :

$$\forall M_j\in T,\ \overrightarrow{\nabla}\varphi_j|_T = \frac{\overrightarrow{n}_j}{2|T|} \tag{3.1.16}$$

Remarque 4. *$X|_T$ désigne la restriction d'une quantité X qui varie en espace à l'élément T. On peut voir ici que les gradients des fonctions de base de Lagrange ne sont pas continus au passage d'un élément à un autre (i.e. aux bords), ce qui est vrai quel que soit le type d'élément de Lagrange considéré.*

La linéarité du problème fait apparaître la vitesse d'avection constante $\overrightarrow{\lambda}$ telle que $\overrightarrow{F} = \overrightarrow{\lambda} u$.

$$\begin{aligned} \phi_i^{T,G} &= \int_T \varphi_i \overrightarrow{\lambda} \cdot \overrightarrow{\nabla} u_h dV &&= \int_T \varphi_i dV \times \left(\sum_{M_j \in T} u_j \overrightarrow{\lambda} \cdot \overrightarrow{\nabla} \varphi_j \right) \\ \Longrightarrow \quad \phi_i^{T,G} &= \frac{|T|}{3} \sum_{M_j \in T} \frac{1}{2|T|} (\overrightarrow{\lambda} u_j) \cdot \overrightarrow{n}_j &&= \frac{1}{3} \sum_{M_j \in T} \frac{1}{2} \overrightarrow{F}_j \overrightarrow{n}_j \\ \Longrightarrow \quad \phi_i^{T,G} &= \frac{1}{3} \phi^T \end{aligned}$$

Ce schéma est connu pour être inconditionnellement instable. Enfin, notons que la conservation de la formulation Petrov-Galerkin des schémas $\mathcal{RD}$ est assurée par la non contribution de la fonction de décentrement au résidu total :

$$\left\{ \begin{array}{l} \displaystyle\sum_{M_i \in T} \beta_i^T = 1 \\ \\ \displaystyle\sum_{M_i \in T} \varphi_i|_T = 1 \end{array} \right. \Longrightarrow \sum_{M_i \in T} \gamma_i^T = 0$$

L'idée de cette forte similitude entre les deux familles de méthodes a été avancée très tôt, durant la première Lecture Series du von Karman Institute for Fluid Dynamics en 1993 et dans [23]. Bien évidemment, l'analyse précédente s'étend de façon immédiate aux problèmes instationnaires. La problématique de la matrice de masse qui se pose alors trouve les mêmes réponses qu'en Éléments Finis. On pourra en particulier avoir recours à des techniques de *mass lumping* si l'on souhaite. Mais si au contraire on veut être précis sur la matrice de masse ou toute autre quantité nécessitant de calculer des valeurs liées aux fonctions de base, il sera possible (sinon nécessaire) d'utiliser le passage à des éléments de référence sur lesquels on peut effectuer un calcul formel. Nous aurons l'occasion d'y revenir d'ici la fin de ce chapitre.

Nous allons maintenant exposer les quelques propriétés importantes que peuvent vérfier les différents schémas distribuant le résidu. Il n'est plus indispensable a priori de se restreindre au cas scalaire.

3.2 Propriétés de la distribution

3.2.1 Consistance

La première question à se poser à propos d'un schéma est de savoir s'il peut converger (en omettant dans un premier temps les problèmes éventuels de stabilité), autrement dit qu'il est consistant, de manière à ce que si les pas de discrétisation Δt et h tendent vers 0 la solution soit bien une solution faible du problème continu. On traite ici de la consistance du schéma $\mathcal{RD}$ en espace, dans le cas d'un système d'équations stationnaires à p inconnues. Pour simplifier l'exposé, et coïncider avec le cadre de la preuve exposée dans [7], le domaine de dimension d est supposé maillé à l'aide d'éléments linéaires P_1 (segments, triangles ou tétraèdres selon la valeur de d), ci-après désignés T. La résolution du problème est effectuée en utilisant une discrétisation simple à l'ordre 1 en temps (en pseudo-temps en fait, un simple outil itératif). On considère donc la méthode $\mathcal{RD}$ suivante (extension aux systèmes du modèle (3.1.5)) :

$$U_i^{n+1} = U_i^n - \frac{\Delta t^n}{|C_i|} \sum_{T \subset \mathcal{T}_i} \phi_i^T(U_h^n) \tag{3.2.1}$$

et plus précisément sa limite formelle quand $t \to \infty$, qui s'écrit naturellement :

$$\forall M_i \in \Omega_h, \quad \sum_{T \subset \mathcal{T}_i} \phi_i^T(U_h) = 0 \tag{3.2.2}$$

Cette formulation exprime l'indépendance de la solution vis-à-vis du parcours itératif. Ceci permet en particulier, pour ces problèmes stationnaires, de faire appel à une stratégie de pas de temps local. Cependant, on peut faire remarquer ici que dans la pratique, comme la convergence n'est jamais atteinte parfaitement, la qualité de la solution qu'on obtient est toujours sensible, aussi peu soit-il, au chemin suivi. C'est un autre sujet, qui requiert au préalable de nous assurer de la consistance de la limite formelle (3.2.2). Dans [7], les auteurs ont démontré un théorème de type Lax-Wendroff ainsi que des conditions de stabilité assurant la convergence des schémas distribuant le résidu. Leur travail se situe dans le cadre de systèmes de lois de conservation stationnaires et sur des éléments finis triangulaires de Lagrange P_2, mais présente le principe général de construction de schémas consistants et précis sur des éléments P_k d'ordre supérieur. Dans ce qui suit, on a donc potentiellement davantage de degrés de liberté que de sommets sur chaque élément. Nous allons simplement rappeler ici les hypothèses et le théorème en lui-même par souci d'être complet, mais nous renvoyons à [7] pour la preuve. La première hypothèse porte sur une régularité raisonnable du maillage.

Hypothèse 3.2.1. *Le maillage Ω_h est conforme et régulier. Par régulier nous entendons que tous les éléments sont grossièrement de la même taille, plus précisément qu'il existe des constantes C_1 et C_2 telles que :*

$$C_1 \leq \sup_{T \subset \Omega_h} \frac{h^d}{|T|} \leq C_2$$

où h désigne la plus grande longueur d'arête du maillage.

On suppose donc d'emblée que cette hypothèse est vérifiée par notre maillage Ω_h, et on définit de plus un maillage dual $\mathcal{C}_h$, associé à l'ensemble des degrés de liberté du maillage, dont la régularité est nécessairement contrainte par celle de Ω_h. On introduit ensuite deux espaces fonctionnels :

$$V_h^k = \left\{ v_h \in \left(\mathcal{C}^0(\mathbb{R}^d)\right)^p ; v_h|_T \in \mathbb{P}^k(T), \forall T \subset \Omega_h \right\}$$

$$X_h = \left\{ v_h ; v_h|_C \in \left(\mathbb{P}^0(C)\right)^p, \forall C \subset \mathcal{C}_h \right\}$$

où p est le nombre d'inconnues du système d'équations (U_h est à valeurs dans $\mathbb{R}^p$). Ceci nous permet de définir l'opérateur de *mass lumping* $L_h : V_h^k \to X_h$ qui à toute fonction U_h décomposable selon les fonctions de base Lagrangiennes (i.e. appartenant à V_h^k) associe une fonction constante par morceaux sur les cellules duales (qui est dans X_h), comme en Volumes Finis. C'est en particulier l'opérateur qu'on utilise pour discrétiser en espace et distribuer le terme $\partial_t U_h$ quand on recherche une solution stationnaire à un problème (la façon de discrétiser ce terme n'influant pas à convergence). Reste maintenant à préciser la régularité requise lors de la construction des résidus. Ces conditions s'appuient nécessairement sur la structure de la solution (qui est dans V_h^k, ce qu'on a vu depuis le début de ce chapitre avec $k = 1$), mais aussi sur la façon de gérer les flux et leur intégration.

Remarque 5. *Nous avons vu que l'écriture des schémas $\mathcal{RD}$ fait intervenir la fluctuation $\int_{\partial T} \overrightarrow{F} \cdot \overrightarrow{n} d\partial V$ sur chaque élément T. Généralement, on procède numériquement au calcul via une formule de quadrature ayant une certaine précision. Nous aurons l'occasion de revenir sur ce point mais disons avec un peu d'avance que cette formule doit être au moins d'une précision de degré égal à l'ordre k d'interpolation de la solution, $U_h \in V_h^k$. Si tel est le cas, on peut représenter le flux $\overrightarrow{F}$ par son interpolée $\overrightarrow{F}_h$ au même ordre*

k. En effet, la formule de quadrature à l'ordre k étant exacte pour un polynôme de degré k, le calcul serait strictement identique en interpolant d'abord la fonction à l'ordre k puis en intégrant de manière exacte. Cette représentation n'est alors en rien gênante pour une analyse de consistance puisqu'elle est au pire la moins précise que nous nous autorisions : si la consistance est avérée dans ce cas, elle le sera pour toute meilleure représentation (qu'on ne détaillerait pas à chaque fois) correspondant à une formule de quadrature plus précise. Ce sujet sera de nouveau abordé lors de l'étude de la construction systématique de schémas d'ordre élevé.

Sans aborder la représentation des flux, nous devons supposer, indépendamment, un minimum de régularité sur la distribution qu'effectue le schéma.

Hypothèse 3.2.2. *Soit Ω_h un maillage vérifiant l'hypothèse 3.2.1. Quelle que soit la constante $C \in \mathbb{R}^+$, il existe une autre constante $C'(C, \Omega_h) \in \mathbb{R}^+$, qui dépend seulement de C et du maillage, telle que pour tout $U_h \in V_h^k$ vérifiant $\forall l \leq p, \|(U_l)_h\|_{L^\infty(\mathbb{R}^d)} \leq C$, on ait :*

$$\forall T \subset \Omega_h, \forall M_i \in T, \|\phi_i^T\| \leq C'(C, \Omega_h) h \sum_{M_j \in T} \|U_j - U_i\|$$

Comme il est noté dans [7], on peut voir cette hypothèse comme une façon de requérir la continuité des résidus partiels vis-à-vis de U_h, puisqu'il s'agit en fait d'une exigence un peu plus forte : c'est un caractère lipschitzien qui est demandé. Passons enfin aux conditions de régularité sur l'approximation des flux (en gardant en mémoire la remarque précédente et le fait que l'approximation en question pourrait être une interpolation polynomiale).

Hypothèse 3.2.3. *Il existe une approximation $\overrightarrow{F}_h$ des flux $\overrightarrow{F}$ qui vérifie les contraintes suivantes :*

(i) Elle autorise la conservation.

$$\forall U_h \in V_h^k, \phi^T := \int_T \overrightarrow{\nabla} \cdot \overrightarrow{F}_h(U_h) dV = \sum_{M_i \in T} \phi_i^T$$

(ii) Elle est continue au passage d'un élément à un autre, au moins dans sa composante normale au bord.

$$\forall U_h \in V_h^k, \forall T_1, T_2 \text{ voisins, } \overrightarrow{F}_h(U_h)|_{T_1} \cdot \overrightarrow{n} = \overrightarrow{F}_h(U_h)|_{T_2} \cdot \overrightarrow{n} \text{ presque partout sur } T_1 \cap T_2$$

où $\overrightarrow{n}$ est une normale quelconque au bord $T_1 \cap T_2$.

(iii) Elle est suffisamment régulière pour que sa divergence soit une fonction continue (lipschitzienne) de U_h sur chaque élément. Autrement dit, quelle que soit $C > 0$, il existe $C'(C)$ telle que pour tout $U_h \in V_h^k$ vérifiant $\forall l \leq p, \|(U_l)_h\|_{L^\infty(\mathbb{R}^d)} \leq C$, on ait :

$$\forall T \subset \Omega_h, \text{ si on note } \overrightarrow{F}_h^T := \overrightarrow{F}_h|_T, \text{ alors } \|\overrightarrow{\nabla} \cdot \overrightarrow{F}_h^T(U_h)\| \leq \frac{C'}{h} \sum_{M_i, M_j \in T} \|U_i - U_j\| \text{ presque partout sur } T$$

(iv) Elle est elle-même consistante si U_h est aussi une approximation consistante de U. Pour toute suite $(U_h)_h$ bornée dans $\left(L^\infty(\mathbb{R}^d \times \mathbb{R}^+)\right)^p$ indépendamment de h et convergente vers U dans $\left(L^2_{loc}(\mathbb{R}^d \times \mathbb{R}^+)\right)^p$, on vérifie :

$$\lim_{h \to 0} \|\overrightarrow{F}_h(U_h) - \overrightarrow{F}(U)\|_{\left(L^1_{loc}(\mathbb{R}^d \times \mathbb{R}^+)\right)^d} = 0$$

On peut à présent exposer le théorème de type Lax-Wendroff qui s'applique aux schémas $\mathcal{RD}$.

Théorème 3.2.1. *Soit la condition initiale* $U_0 \in \left(L^\infty(\mathbb{R}^d)\right)^p$ *et* U_h *l'approximation sur laquelle se base la résolution, par un schéma quelconque dont une écriture générique est (3.2.1). Ce schéma doit vérifier la condition 3.2.2 et être cohérent avec une représentation des flux qui respecte 3.2.3. On suppose en outre qu'il existe une constante* C*, qui dépend seulement des constantes* C_1 *et* C_2 *(issues de l'hypothèse 3.2.1 sur le maillage) et de* U_0*, et une fonction* $U \in \left(L^2(\mathbb{R}^d \times \mathbb{R}^+)\right)^p$ *telles que :*

$$\sup_h \sup_{x,y,t} \|U_h(x,y,t)\|_{\mathbb{R}^p} \leq C$$

$$\lim_{h \to 0} \|U - U_h\|_{\left(L^2_{loc}(\mathbb{R}^d \times \mathbb{R}^+)\right)^p} = 0$$

Alors U *est une solution faible du système de lois de conservation (3.1.1).*

La preuve est décrite dans [7] pour le cas des triangles mais elle s'étend de façon naturelle aux éléments non linéaires (nécessite seulement une redéfinition de l'espace V_h^k), aux configurations 3D et aux ordres supérieurs si on suit certains principes de construction.

3.2.2 Principe du maximum, positivité et monotonie

Préambule sur la notion de stabilité

La consistance assure que la convergence en $(h, \Delta t)$ vers une solution faible continue est réalisable avec le schéma employé. Toutefois, outre la question de savoir si la solution obtenue à la limite est la seule admissible ou non (il faut sélectionner la solution entropique), il reste à déterminer comment se comporte la solution avant que la limite ne soit atteinte, ce qui n'arrive bien sûr jamais exactement. En particulier, on veut éviter certains cas pathologiques comme la survenue d'un état stable qui ne serait pas la solution physique ou même une divergence aboutissant à une incohérence physique (typiquement, une densité ou une énergie négative). Ce sujet est celui de la stabilité du schéma numérique : un schéma est stable s'il garantit que de telles situations n'auront pas lieu ou s'il est capable de les corriger quand elles surviennent. On peut considérer ces situations indésirables comme des perturbations artificielles de la physique du problème. Si le schéma permet à celles-ci de perdurer et de parasiter la solution, ou pire de s'emballer et de s'éloigner sans cesse de la réalité physique (donc de la "vraie" solution), on dit qu'il est instable puisqu'il il ne ramène pas de lui-même l'équilibre (que représente la situation la plus proche physiquement correcte). Mais si stabilité il y a, et qu'on sait de plus être consistant, on est assuré de converger en maillage (au sens large, c'est-à-dire en considérant la droite du temps discrétisée comme une partie du maillage). On pourrait en fait définir la stabilité comme la capacité d'un schéma consistant à converger.

La question de la stabilité dépasse le cadre purement mathématique de la discrétisation du problème. Le comportement de la solution est nécessairement contraint par les propriétés de l'équation elle-même, autrement dit par la physique qu'elle ne fait que décrire. Ce qu'on a dit est que si on s'assure qu'à chaque étape, la solution est consistante avec la physique du problème, alors il n'y a aucune raison pour que la convergence ne soit pas atteinte. Il faut donc contraindre le schéma de manière à ce qu'il imite, au moins sous certains aspects qui s'avèreraient des foyers d'instabilités numériques, l'évolution physique du problème. Si on y parvient, on est assurés d'obtenir une solution, non convergée en maillage (c'est-à-dire évidemment calculée avec un h et un Δt non nuls), qui sera en plus une approximation sensée du point de vue du physicien, puisqu'on sera sûr que la solution calculée appartient à un chemin qui mène de manière inéluctable à la limite recherchée : la solution continue presque partout.

Le principe du maximum pour une loi de conservation scalaire

Nous devons revenir à la loi de conservation scalaire (3.1.1), et nous placer dans le cas linéaire pour commencer. Le flux s'exprime alors en fonction d'une vitesse d'advection constante $\overrightarrow{\lambda}$.

$$\frac{\partial u}{\partial t} + \overrightarrow{\nabla} \cdot (\overrightarrow{\lambda} u) = \frac{\partial u}{\partial t} + \overrightarrow{\lambda} \cdot \overrightarrow{\nabla} u = 0 \tag{3.2.3}$$

Supposons que nous ayons une solution initiale présentant une discontinuité et deux états constants de part et d'autre de celle-ci. La méthode des caractéristiques permet d'affirmer que la solution initiale ne fera que subir une translation avec le temps, dont la vitesse et la direction sont données par $\overrightarrow{\lambda}$ (cf. figure 3.6).

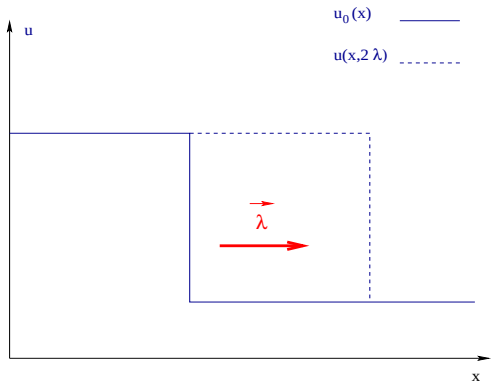


FIGURE 3.6 – Advection linéaire pure dans le cas 1D

Maintenant, si on revient au cas plus général d'un flux non linéaire, la différence est que la vitesse d'advection n'est plus constante et que les caractéristiques peuvent se croiser (donnant lieu à des chocs) ou s'écarter (formant une zone de raréfaction). En revanche, même si ce problème est beaucoup plus complexe, il y a toujours une propriété qu'on conserve, c'est le fait que la solution préserve ses bornes initiales :

$$u^0_{max} = \max_{\overrightarrow{x} \in \Omega} (u_0(\overrightarrow{x})) \text{ et } u^0_{min} = \min_{\overrightarrow{x} \in \Omega} (u_0(\overrightarrow{x}))$$

Cette propriété du problème continu est généralement désignée comme *principe du maximum*. Il se trouve qu'en général, les schémas numériques ne respectent pas ce principe et que la solution calculée dépasse les bornes théoriques. Si ce phénomène s'emballe au cours du processus itératif, il peut aboutir à des situations aberrantes et faire échouer la simulation. On est donc en présence d'une source d'instabilité numérique, et on voudrait être capable de construire un schéma préservant exactement les extrema initiaux. Mais cela s'avère trop complexe. Or on peut trouver une condition moins forte et donc plus facile à mettre en place. La majorité des phénomènes physiques modélisables à l'aide d'une loi de conservation de la forme (3.1.1) (sinon tous les phénomènes de ce type) sont en réalité des limites idéales de problèmes régis par une loi du type :

$$\frac{\partial u}{\partial t} + \overrightarrow{\nabla} \cdot \left(\overrightarrow{F}(u) \right) + \overrightarrow{\nabla} \cdot \left(\overrightarrow{\epsilon}(u) \overrightarrow{\nabla} u \right) = 0 \tag{3.2.4}$$

où le paramètre non linéaire $\epsilon(u)$ est considéré suffisamment proche de la limite $\epsilon \to 0$ pour qu'on considère que celle-ci est atteinte. Le dernier terme de cette équation, qui la rend parabolique, modélise toujours des phénomènes dissipatifs. Son existence est donc à relier à la discussion sur l'entropie et les phénomènes irréversibles en thermodynamique (voir le chapitre précédent). En particulier, on doit se rappeler le lien entre solution visqueuse et respect d'une inégalité d'entropie vu au chapitre précédent (les implications de ces deux critères physiques se rejoignent). Par rapport au cas idéal, l'ajout de ce terme

va avoir tendance à amortir les profils discontinus. La solution ressemblera donc davantage à ce qu'on voit sur la figure 3.7, qui illustre le cas 1D mais se généralise aux dimensions supérieures. Ceci signifie

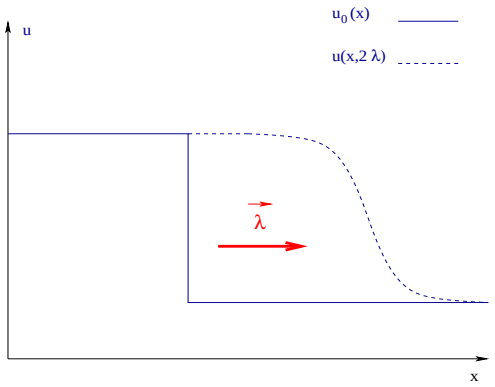


FIGURE 3.7 – Advection-diffusion linéaire dans le cas 1D avec $\epsilon > 0$

que si c'est un non-sens physique d'être au-dessus du maximum, il n'est en revanche pas aberrant de se situer en-dessous. Réciproquement pour le minimum. Par conséquent, à défaut de pouvoir construire un schéma préservant exactement les bornes, on pourra requérir seulement une inégalité sur les bornes atteintes numériquement, sans poser a priori de problème de stabilité numérique. On verra que la notion de monotonie est liée au respect d'un principe du maximum discret.

Expression pratique dans le formalisme $\mathcal{RD}$

Le principe du maximum discret que nous venons d'évoquer peut s'exprimer sous la forme globale suivante en tout degré de liberté i :

$$u^n_{min} = \min_{M_j \in \Omega_h} u^n_j \leq u^{n+1}_i \leq \max_{M_j \in \Omega_h} u^n_j = u^n_{max} \tag{3.2.5}$$

Pour analyser le comportement du schéma vis-à-vis de ce principe, il est donc nécessaire de pouvoir le réécrire comme fonction des inconnues u_j. En scalaire, tous les schémas $\mathcal{RD}$ connus peuvent se mettre sous la forme suivante :

$$\phi^T_i = \sum_{M_j \in T} c^T_{ij}(u_i - u_j) \tag{3.2.6}$$

$$\Rightarrow \quad \int_{C_i} \frac{\partial u_h}{\partial t} dV = -\sum_{T \subset \mathcal{T}_i} \sum_{M_j \in T} c^T_{ij}(u_i - u_j)$$

même si dans les cas où la loi de conservation est non linéaire, les coefficients c_{ij} dépendent de la solution et doivent être construits soigneusement. Simplifions le problème en considérant des problèmes soit stationnaires, soit instationnaires mais discrétisés à l'ordre 1, en ayant recours à l'opérateur de *mass lumping* pour approcher l'intégrale de la dérivée en temps. On peut alors écrire :

$$\begin{aligned} |C_i|\frac{du_i}{dt} &= -\sum_{T \subset \mathcal{T}_i} \sum_{M_j \in T} c^T_{ij}(u_i - u_j) \\ \Rightarrow \qquad \frac{du_i}{dt} &= -\frac{1}{|C_i|} \sum_{M_j \in \mathcal{T}_i, j \neq i} \left(\sum_{T \subset \mathcal{T}_i \cap \mathcal{T}_j} c^T_{ij} \right) (u_i - u_j) \\ &= -\frac{1}{|C_i|} \sum_{M_j \in \mathcal{T}_i, j \neq i} \widetilde{c_{ij}}(u_i - u_j) \end{aligned}$$

L'analyse qui suit est issue de [33]. Une condition forte reliée au principe du maximum peut être exprimée à ce stade. Si u_i est un maximum, on souhaite qu'il décroisse ou se conserve. Au contraire, s'il s'agit d'un minimum, on souhaite qu'il augmente ou stagne. Dans la littérature, on se réfère habituellement à cette propriété via le qualificatif *Local Extremum Diminishing* ou LED.

$$\begin{cases} \dfrac{du_i}{dt} \leq 0 & \text{si } \forall M_j \in \mathcal{T}_i,\ u_i \geq u_j \\ \\ \dfrac{du_i}{dt} \geq 0 & \text{si } \forall M_j \in \mathcal{T}_i,\ u_i \leq u_j \end{cases}$$

On peut remarquer que cette condition est automatiquement vérifiée si :

$$\widetilde{c_{ij}} = \sum_{T \subset \mathcal{T}_i \cap \mathcal{T}_j} c_{ij}^T \geq 0 \ \forall M_j \in \mathcal{T}_i, j \neq i \tag{3.2.7}$$

Une condition encore plus forte est de requérir la positivité des coefficients locaux à l'élément.

$$\forall M_j \in \mathcal{T}_i, \forall T \subset \mathcal{T}_i, c_{ij}^T \geq 0 \tag{3.2.8}$$

Dans ce cas, on parlera de schéma à coefficients positifs. Pour aller plus loin, il faut connaître la façon dont est approché le terme source. En guise d'exemple d'application, que ce soit pour résoudre un problème stationnaire ou instationnaire avec une précision globalement d'ordre 1, on peut utiliser le schéma-θ en temps :

$$u_i^{n+1} = u_i^n - \frac{\Delta t^n}{|C_i|} \sum_{T \subset \mathcal{T}_i} \left((1-\theta)\phi_i^T(U^n) - \theta \phi_i^T(U^{n+1}) \right) \tag{3.2.9}$$

Selon la valeur de θ, on peut retrouver le schéma d'Euler explicite ($\theta = 0$), de Crank-Nicolson ($\theta = \frac{1}{2}$) ou d'Euler implicite ($\theta = 1$). On peut affirmer que le principe du maximum discret (3.2.5) est vérifié par le schéma (3.2.9) si la condition (3.2.7) est respectée et sous une contrainte de type CFL sur le pas de temps :

$$\forall M_i \in \Omega_h, |C_i| - (1-\theta)\Delta t \sum_{T \subset \mathcal{T}_i} \sum_{M_j \in T, j \neq i} c_{ij}^T \geq 0$$

De plus, dans le cas purement explicite ($\theta = 0$), cette restriction assure des bornes plus étroites en chaque degré de liberté :

$$\min_{M_j \in \mathcal{T}_i} u_j^n \leq u_i^{n+1} \leq \max_{M_j \in \mathcal{T}_i} u_j^n \tag{3.2.10}$$

Pour une démonstration, on renvoit à [33]. Un tel schéma sera dit positif. Il est également possible de contraindre davantage le schéma (3.2.9) et de lui imposer d'être localement positif. Il vérifiera alors bien sûr toujours les mêmes principes du maximum discret et même (3.2.10) dans le cas purement explicite. Pour cela, il doit vérifier la propriété locale (3.2.8) et la condition sur Δt :

$$\forall M_i \in \Omega_h, \forall T \subset \Omega_h, |T \cap C_i| - (1-\theta)\Delta t \sum_{M_j \in T, j \neq i} c_{ij}^T \geq 0$$

Que le schéma soit positif localement ou non, on remarque que la contrainte sur le pas de temps disparaît dans le cas purement implicite $\theta = 1$. Si un schéma scalaire purement implicite est à coefficients positifs, alors il vérifie un principe du maximum discret global (3.2.5) de manière inconditionnelle.

Lien avec la monotonie

Définition 3.2.1. Un schéma est dit monotone si la solution obtenue vérifie à toute date t^n :

$$\forall n > 0,\ u_0 \leq v_0 \Rightarrow u^n \leq v^n$$

où l'indice $_0$ désigne la solution initiale.

L'inégalité porte sur tout le domaine d'étude, c'est-à-dire sur chaque degré de liberté si on considère les interpolations. Par conséquent, un schéma monotone conserve ses bornes initiales.

Définition 3.2.2. Un schéma préserve la monotonie si, partant de deux solutions dont la comparaison est uniforme à une date t^n, les solutions obtenues au temps t^{n+1} préservent cette relation d'ordre uniforme.

$$\forall n > 0,\ u^n \leq v^n \Rightarrow u^{n+1} \leq v^{n+1}$$

Un schéma préservant la monotonie est donc monotone. De plus, on peut montrer qu'un schéma positif (au sens que nous avons défini plus haut) préserve la monotonie. Donc un schéma positif respecte un principe du maximum discret. Dans ces cas-là, on le qualifie également de stable en norme L^∞.

En conclusion

Comment savoir si le schéma est stable désormais ? Il y deux solutions. Soit on connaît parfaitement la physique du problème et les propriétés critiques que le schéma doit imiter, et on sait s'il les respecte ou non ; soit seul le calcul permettra de l'affirmer ou non, selon l'occurence ou non d'une situation physiquement non viable identifiable par l'utilisateur. On peut d'ores et déjà affirmer que le seul critère exposé ici ne suffit pas pour assurer que **tout** schéma scalaire le vérifiant est stable. En revanche, il est important de garder à l'esprit que les aspects physiques dominants lors d'une simulation dépendent du problème. Par conséquent, il ne sert à rien de requérir la positivité pour un problème très régulier pour lequel on est sûr de ne pas rencontrer de gradient significatif.

Extension aux systèmes

On ne peut malheureusement pas étendre cette propriété aux systèmes de lois de conservation de façon triviale, c'est-à-dire équation par équation. Le couplage des équations signifie que l'information globale, bien que conservée, peut changer de formes et se transférer d'une variable à une autre. Par conséquent, à moins qu'il ne soit possible de découpler les équations, il est faux de prétendre que le principe du maximum s'applique à chaque équation individuellement. Un découplage est possible, par exemple, dans tout système hyperbolique monodimensionnel. Pour le montrer, revenons à la formulation quasi-linéaire d'une loi de conservation en une dimension d'espace :

$$\frac{\partial U}{\partial t} + A\frac{\partial U}{\partial x} = 0$$

Or le système est supposé hyperbolique, ce qui signifie que la matrice est diagonalisable ($A = R\Lambda L$) et donc :

$$\begin{aligned} \frac{\partial U}{\partial t} + R\Lambda L\frac{\partial U}{\partial x} &= 0 \\ \Rightarrow \quad L\frac{\partial U}{\partial t} + \Lambda L\frac{\partial U}{\partial x} &= 0 \end{aligned}$$

L étant inversible, il existe un jeu de variables W tel que $L = \frac{\partial W}{\partial U}$. Par conséquent, on a :

$$\frac{\partial W}{\partial t} + \Lambda\frac{\partial W}{\partial x} = 0$$

Une fois le système diagonalisé, toutes les équations sont indépendantes et peuvent être traitées comme des problèmes scalaires (le seul problème qui reste étant que cette écriture n'est pas conservative, mais il est peut-être possible de déterminer des lois de flux dont les dérivées sont les coefficients Λ_i). Les choses se compliquent dès que nous passons en dimension supérieure, car il n'est alors plus possible (hors cas remarquables sans intérêt) de diagonaliser (découpler) les équations par un simple changement de

variables, tout simplement parce que les jacobiennes dans chaque direction ne sont pas diagonalisables dans la même base. Pour que cela soit possible il faudrait que les jacobiennes soient commutantes, ce qui n'est généralement pas le cas.

En revanche, pour une loi de conservation scalaire linéaire, un principe proche de celui de la conservation du maximum peut être formulé en norme L^2 sur le domaine entier, ainsi que son affaiblissement en termes d'inégalité. Pour cela, on ignore les conditions qui s'appliquent aux bords ou on suppose qu'elles sont telles qu'elles ne permettent à aucune information de sortir. Il est alors correct d'écrire :

$$\forall t, \|u(.,t)\|_{L^2(\Omega)} = \|u_0\|_{L^2(\Omega)} \quad \text{(Conservation)}$$

$$\Rightarrow \quad \forall t, \|u(.,t)\|_{L^2(\Omega)} \leq \|u_0\|_{L^2(\Omega)} \quad \text{(Dissipation)}$$

Ceci peut se généraliser de façon immédiate aux systèmes linéaires :

$$\|U(.,t)\|^2_{(L^2(\Omega))^p} = \int_\Omega \|U(\overrightarrow{x},t)\|^2_{\mathbb{R}^p}\, dV \leq \|U_0\|^2_{(L^2(\Omega))^p} \tag{3.2.11}$$

où p représente le nombre de variables, i.e. la taille du vecteur U. On appelle généralement cette norme - induite par le produit scalaire dans $L^2(\Omega)$ - l'énergie, ce mot ne revêtant plus ici le même sens physique (il s'agit tout au plus d'une analogie).

Inégalités d'entropie

Pour des systèmes hyperboliques non linéaires, une telle inégalité en norme L^2 n'est plus valable. Néanmoins, on peut d'une certaine manière dire qu'elle vérifie un principe du maximum pour une autre norme (en quelque sorte) : l'entropie mathématique définie au chapitre précédent. Il est important de se rappeler maintenant le théorème 2.3.1 qui lie le respect des inégalités d'entropie à la limite de la solution faible d'un système dissipatif. C'est exactement le type d'argument que nous avons donné dans le cas scalaire pour justifier l'existence d'un maximum global de la solution. En réalité, dans le cas scalaire 1D, un lien peut être démontré entre l'entropie et la limite diffusive d'une loi de conservation : c'est le théorème de Kruzkov ([58]).

Théorème 3.2.2. ***(Kruzkov)***
Pour toute solution initiale u_0 bornée et mesurable sur $\mathbb{R}$, il existe une unique solution entropique de (3.1.1) dans $L^\infty(\Omega \times [0:T]) \cap \mathcal{C}([0:T]; L^1_{loc}(\mathbb{R}))$. Elle satisfait le principe du maximum :

$$\forall t > 0, \|u(.,t)\|_{L^\infty(\Omega)} = \|u_0\|_{L^\infty(\mathbb{R})}$$

Le théorème n'est pas complet, mais nous ne gardons que cette partie qui nous intéresse. Ce qu'on pourrait en déduire, c'est que la bonne généralisation de la monotonie aux systèmes peut être celle du respect d'inégalités d'entropie. Ce n'est pas tout à fait exact. Ou plutôt, tout dépend de ce qu'on entend par là. Il est vrai qu'il y a une similarité et qu'à ce titre, imposer à un schéma de vérifier des inégalités d'entropie **est** la bonne généralisation du respect d'un maximum en scalaire, **mais cela ne signifie pas pour autant qu'on empêchera**, contrairement au cas scalaire, **la solution de montrer des oscillations non physiques autour des chocs**. En particulier, cela implique qu'on ne peut plus garantir qu'un schéma préserve la positivité de la densité et de la pression sur des temps longs, simplement en se basant sur cette propriété. Néanmoins, il reste très appréciable qu'un schéma vérifie un équivalent discret de l'inégalité de Clausius-Duhem pour que la solution retenue par le schéma soit l'unique solution entropique. Ceci signifie en outre que, puisqu'on ne retient que la solution physique lors de la convergence en maillage, un raffinement dans la zone de discontinuité aura tendance à réduire les oscillations non physiques.

Nous résolvons le système de lois de conservation (2.3.1) et nous recherchons à formuler la condition sur sa solution équivalant au respect des inégalités d'entropie mathématiques. Pour cela, reprenons les notations de la section 2.3.3 du chapitre précédent :

$$\begin{array}{rrcl} & \dfrac{\partial S}{\partial t}(U) + \overrightarrow{\nabla} \cdot \overrightarrow{G}(U) & \leq & 0 \\ \Rightarrow & \dfrac{\partial S}{\partial U}\dfrac{\partial U}{\partial t} + \dfrac{\partial \overrightarrow{G}}{\partial U} \cdot \overrightarrow{\nabla} U & \leq & 0 \\ \overset{(2.3.7)}{\Rightarrow} & \nabla_U S \dfrac{\partial U}{\partial t} + \nabla_U S \nabla_U \overrightarrow{F}(U) \cdot \overrightarrow{\nabla} U & \leq & 0 \\ \Rightarrow & \left\langle W, \dfrac{\partial U}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U) \right\rangle & \leq & 0 \end{array}$$

Notons le système de lois de conservation à résoudre $R(U) = 0$, et supposons que notre méthode numérique approche la solution U par U_h et l'opérateur R par R_h. Dans ce cas, on requerra qu'une quantité du type $\langle W_h, R_h \circ U(W_h) \rangle$, que l'on appelle parfois l'énergie par analogie avec le cas précédent des systèmes linéaires, décroisse. Pour des systèmes non linéaires, on remplace donc la recherche de la monotonie par celle de la stabilité énergétique (entropique), c'est-à-dire qu'on souhaite obtenir un schéma qui consomme (dissipe) l'énergie mathématique que nous venons de définir. On préfèrera conserver la désignation 'stabilité entropique' qui nous semble plus claire. En ce sens, c'est l'entropie mathématique qui est dissipée, puisqu'étant opposée à l'entropie physique. Plus de détails peuvent être trouvés en annexe, où nous revenons sur la stabilité entropique de schémas que nous allons présenter plus loin.

Ceci est difficile à obtenir en pratique et dans le cas général. Il s'avère que les schémas scalaires monotones qu'on étend aux systèmes satisfont généralement assez bien cette propriété (en particulier la réduction des oscillations et donc la préservation de la positivité de certaines grandeurs, voir par exemple [72]). Même si ça ne suffit pas toujours pour prouver qu'ils vérifient une inégalité d'entropie discrète. C'est pourquoi on s'intéresse au cas scalaire même en sachant que le but ultime est de résoudre des systèmes de lois de conservation dans des configurations multidimensionnelles. Pour finir, on peut faire la même remarque que dans le cas scalaire, à savoir qu'on peut annoncer que le critère de stabilité entropiqe ne suffit pas à caractériser un schéma stable.

3.2.3 Précision en espace

Intéressons-nous à présent à l'erreur commise lors de la discrétisation du problème. À partir de maintenant, nous noterons h la plus grande longueur d'arête du maillage Ω_h et toute quantité indicée par $_h$ sera une interpolation polynomiale de Lagrange à un certain ordre k. Les éléments ne sont plus nécessairement des triangles, même si nous n'avons pas encore abordé la façon de mettre en oeuvre les schémas $\mathcal{RD}$ sur d'autres configurations, et un élément quelconque sera désigné par E. En revanche, on conserve les notations C_i et $\mathcal{T}_i$, par rapport à un degré de liberté M_i, comme références respectives à sa cellule duale et à l'ensemble des éléments le contenant.

Avant de parler de précision, il faut définir l'erreur à laquelle on se réfère. L'idée intuitive serait de comparer la solution U^* que nous obtenons du schéma avec la véritable solution U du problème continu et de définir l'erreur comme l'écart de ces deux solutions dans la norme d'un espace fonctionnel approprié, par exemple :

$$\mathcal{E}^* = \|U^* - U\|_{L^2(\Omega_h)}$$

Cependant nous ne connaissons ni l'une ni l'autre de ces solutions de façon formelle. Une alternative est de mesurer non pas l'écart final obtenu mais l'erreur propre au schéma par rapport à un analogue exact : c'est l'erreur de troncature. Dans ce chapitre, nous nous consacrons à la résolution de problèmes

stationnaires, ce qui amène à résoudre en chaque degré de liberté le problème suivant :

$$\forall M_i \in \Omega_h,\ |C_i|\frac{U_i^{n+1}-U_i^n}{\Delta\tau} + \sum_{E\subset\mathcal{T}_i}\phi_i^E(U_h^n) = 0$$

Comme nous ne nous intéressons qu'à l'erreur commise sur la solution finale, i.e. une fois la convergence atteinte, il revient au même d'étudier l'erreur de troncature du schéma convergé, qui est :

$$\forall M_i \in \Omega_h,\ \sum_{E\subset\mathcal{T}_i}\phi_i^E(U_h) = 0 \tag{3.2.12}$$

Quel que soit le degré de liberté M_i, notons l'analogue exact au problème associé (3.2.12) dont nous parlions plus haut, $\phi_i^{ex}(U) = 0$, où U est la solution continue, limite de U_h lorsque $h \to 0$. Comme noté dans [59], on pourrait étudier le vecteur $\overrightarrow{\mathcal{E}}$ dont chaque composante correspondant à un degré de liberté M_i vaudrait :

$$\begin{aligned}\mathcal{E}_i &= \sum_{E\subset\mathcal{T}_i}\left(\phi_i^E(U_h) - \phi_i^{ex}(U)\right)\\ &= \sum_{E\subset\mathcal{T}_i}\phi_i^E(U_h)\end{aligned}$$

Le système à résoudre étant constitué de p équations, chacune de ces composante est en réalité un vecteur de taille p. Si on procède ainsi, on ne définira que des erreurs locales, or on souhaite définir une erreur globale pour qualifier le schéma. On préfère donc définir l'erreur de troncature comme une projection de ce vecteur sur les valeurs d'une fonction régulière $\psi(\overrightarrow{x})$ quelconque appartenant à $\mathcal{C}_0^1(\Omega)$. L'avantage de cette procédure réside dans le fait que nous aboutissions ainsi à une expression de l'erreur de troncature indépendante du maillage, puisque projetée d'une certaine façon sur tout le domaine. En réalité, cette façon de définir l'erreur est celle de la méthode de Galerkin, ce qui se justifie ici par l'interprétation de type Petrov-Galerkin de nos schémas. C'est d'ailleurs l'idée qui va transparaître dans l'analyse suivante : nous considérons nos schémas comme des perturbations du schéma de Galerkin. En s'inspirant de [7], [1], [90] et [33], on va étudier la nouvelle quantité $\mathcal{E}_\psi$, définie par :

$$\mathcal{E}_\psi = \sum_{M_i\in\Omega_h}\psi_i\sum_{E\subset\mathcal{T}_i}\phi_i^E(U_h) \tag{3.2.13}$$

où on note $\psi_i = \psi(\overrightarrow{x_i})$. Le fait que cette quantité soit de taille p n'ajoute aucune difficulté à l'analyse de précision, car tout se passe exactement comme en scalaire. Pour s'en convaincre, il suffit de considérer les raisonnements qui vont suivre composante par composante.

Définition 3.2.3. Un schéma distribuant le résidu est dit précis à l'ordre k si l'erreur de troncature $\mathcal{E}_\psi$ commise est en $O(h^k)$.

Nous cherchons à construire un schéma d'ordre $k+1$ à partir d'une interpolation polynomiale de degré k de la solution. Comme nous l'avons annoncé, on découpe l'analyse en deux parties, l'erreur relative entre le schéma $\mathcal{RD}$ et la méthode de Galerkin continu (donc dûe au décentrement) et ensuite l'erreur inhérente au schéma de Galerkin lui-même. Ce qui revient à écrire :

$$\begin{aligned}\mathcal{E}_\psi &= \sum_{M_i\in\Omega_h}\psi_i\sum_{E\subset\mathcal{T}_i}\phi_i^E\\ &= \underbrace{\sum_{E\subset\Omega_h}\sum_{M_i\in E}\psi_i\left(\phi_i^E - \Psi_i^E\right)}_{I} + \underbrace{\sum_{E\subset\Omega_h}\sum_{M_i\in E}\psi_i\Psi_i^E}_{II}\end{aligned}$$

où Ψ_i^E représente le résidu partiel associé au schéma de Galerkin. Rappelons que par définition, l'opérateur de projection selon les fonctions de base de Lagrange φ_i (qui ont chacune pour support $\mathcal{T}_i$) s'écrit :

$$\pi_h^k\left(\psi(\overrightarrow{x})\right) = \psi_h(\overrightarrow{x}) = \sum_{M_i \in \Omega_h} \int_{\mathcal{T}_i} \psi_i \varphi_i(\overrightarrow{x}) dV = \sum_{M_i \in \Omega_h} \sum_{E \subset \mathcal{T}_i} \int_E \psi_i \varphi_i|_E(\overrightarrow{x}) dV$$

Gardons aussi à l'esprit que l'erreur commise en interpolant avec des polynômes de Lagrange de degré k est d'ordre $k+1$. Commençons par étudier le second terme, c'est-à-dire l'erreur associée au schéma de Galerkin :

$$\begin{aligned} II &= \sum_{E \subset \Omega_h} \sum_{M_i \in E} \psi_i \int_E \varphi_i|_E \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h) dV = \sum_{E \subset \Omega_h} \int_E \psi_h|_E \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h) dV \\ &= \int_{\Omega_h} \psi_h \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h) dV \end{aligned}$$

Pour alléger le texte, nous introduisons la notation suivante :

$$r(U_h) = \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h)$$

Si U désigne en outre la solution exacte du problème, $r(U) = 0$. Une simple manipulation algébrique montre alors que :

$$\begin{aligned} \psi_h r(U_h) &= (\psi_h - \psi)(r(U_h) - r(U)) + (\psi_h - \psi) r(U) + \psi(r(U_h) - r(U)) + \psi r(U) \\ &= (\psi_h - \psi)(r(U_h) - r(U)) + \psi(r(U_h) - r(U)) \\ &= \left(O(h^{k+1}) + \psi\right)(r(U_h) - r(U)) \end{aligned}$$

$$\Rightarrow \ \psi_h r(U_h) \underset{h \to 0}{\sim} \psi\left(r(U_h) - r(U)\right) \tag{3.2.14}$$

En appliquant ce raisonnement à II, on obtient :

$$\begin{aligned} II &\underset{h \to 0}{\sim} \int_{\Omega_h} \psi \left(\overrightarrow{\nabla} \cdot (\overrightarrow{F}(U_h) - \overrightarrow{F}(u)) \right) dV \\ &\underset{h \to 0}{\sim} \int_{\partial \Omega_h} \psi \left(\overrightarrow{F}_h(U_h) - \overrightarrow{F}(U) \right) \cdot \overrightarrow{d\partial\Omega} - \int_\Omega \overrightarrow{\nabla} \psi \cdot \left(\overrightarrow{F}(U_h) - \overrightarrow{F}(U) \right) dV \\ &= O\left(\max_{\Omega_h} \left\| \overrightarrow{F}(U_h) - \overrightarrow{F}(U) \right\|_{\mathbb{R}^d} \right) \end{aligned}$$

car ψ étant régulière (dans $\mathcal{C}_0^1(\Omega)$), $\overrightarrow{\nabla}\psi$ est borné. Les flux étant des fonctions régulières (en particulier Lipschitziennes) de la solution, on a alors une estimation du type :

$$\max_{\Omega_h} \left\| \overrightarrow{F}(U_h) - \overrightarrow{F}(U) \right\|_{\mathbb{R}^d} = \max_{\Omega_h} \left\| \overrightarrow{C}_F (U_h - U) \right\|_{\mathbb{R}^d} = O(h^{k+1})$$

où les quantités $\overrightarrow{C}_F$ sont des constantes de Lipschitz des flux. Cependant, les intégrales précédentes ne sont jamais calculées exactement, car les flux ne s'expriment pas de façon simple en U et donc a fortiori en espace. Deux solutions sont possibles : soit interpoler à nouveau $F \circ U_h$, soit opter pour une formule de quadrature. Dans le premier cas, cela introduit la quantité $F_h(U_h)$ et on a schématiquement :

$$\left\| \overrightarrow{F}_h(U_h) - \overrightarrow{F}(U) \right\| \leq \left\| \overrightarrow{F}_h(U_h) - \overrightarrow{F}(U_h) \right\| + \left\| \overrightarrow{F}(U_h) - \overrightarrow{F}(U) \right\| = O(h^{k+1})$$

à condition que cette interpolation des flux se fasse également à l'ordre $k+1$, par exemple de la même manière que pour U_h. Dans le cas d'une formule de quadrature, celle-ci doit être aussi d'ordre $k+1$. C'est cette dernière solution qu'on adopte généralement. On est alors en mesure de conclure que :

$$II = O(h^{k+1})$$

Passons à présent à l'étude de l'erreur commise par rapport au schéma de Galerkin.

$$I = \sum_{E \subset \Omega_h} \sum_{M_i \in E} \psi_i \left(\phi_i^E - \Psi_i^E \right)$$

Rappelons que le schéma de Galerkin, tout comme n'importe quel schéma distribuant le résidu, vérifie la propriété de conservation (3.1.9). Ceci nous permet d'écrire :

$$I = \sum_{E \subset \Omega_h} \frac{1}{N_s^E} \sum_{M_i \in E} \sum_{M_j \in E} (\psi_i - \psi_j) \left(\phi_i^E - \Psi_i^E \right)$$

où N_s^E désigne le nombre de degrés de liberté appartenant à l'élément E. Comme la fonction ψ est régulière (en particulier Lipschitzienne), et que les degrés de liberté au sein d'un élément sont séparés par une distance de l'ordre de h, on peut affirmer que $\psi_i - \psi_j = O(h)$. De plus, si le maillage est régulier (au sens de l'hypothèse 3.2.1), le nombre d'éléments constitutifs du maillage peut être estimé en $O(h^{-d})$.

$$I = O(h^{-d}) \times O(h) \times \left(O(\max_{E \in \Omega_h} (\phi_i^E)) + O(\max_{E \in \Omega_h} (\Psi_i^E)) \right)$$

La contribution du schéma de Galerkin est estimable immédiatement pour un élément E quelconque. En reprenant les notations et les arguments précédents, on obtient :

$$\begin{aligned}
\Psi_i^E &= \int_E \varphi_i|_E r(U_h) dV = \int_E \varphi_i|_E \left(r(U_h) - r(U) \right) dV \\
&= \int_{\partial E} \varphi_i|_E \left(\overrightarrow{F}(U_h) - \overrightarrow{F}(U) \right) \cdot \overrightarrow{n} \, d\partial E - \int_E \overrightarrow{\nabla} \varphi_i|_E \cdot \left(\overrightarrow{F}(U_h) - \overrightarrow{F}(U) \right) dV \\
&= O(h^{d-1}) \times O(h^{k+1}) + O(h^d) \times O(h^{-1}) \times O(h^{k+1}) \\
&= O(h^{k+d})
\end{aligned}$$

On remarque que si ϕ_i^E est également en $O(h^{k+d})$, l'erreur I est bien en $O(h^{k+1})$ comme souhaité (comme II). On est donc en mesure de conclure par le théorème suivant :

Théorème 3.2.3.
Un schéma distribuant le résidu basé sur une interpolation P_k de la solution est d'ordre $k+1$ si :
- *les flux $\overrightarrow{F}_h$ sont approchés avec une erreur d'ordre $k+1$*
- *les résidus partiels ϕ_i^E sont en $O(h^{k+d})$*

3.2.4 Préservation de la linéarité

Définition 3.2.4. Un schéma est dit préservant la linéarité ($\mathcal{LP}$ pour *Linearity Preserving*) si la distribution s'annule avec le résidu global :

$$\forall E \subset \Omega_h, \forall M_i \in E, \ \phi_i^E \xrightarrow[\phi^E \to 0]{} 0$$

Cette dernière notion, en apparence assez naturelle, a été introduite dans [70] et reprise par la suite dans une large partie des travaux sur les schémas $\mathcal{RD}$ (entre autres [90], [75] et [59]), pour la raison que nous exposons maintenant. Sachant que tous les schémas $\mathcal{RD}$ peuvent se réécrire sous la forme :

$$\phi_i^E = \beta_i^E \phi^E$$

une condition suffisante pour qu'un schéma soit $\mathcal{LP}$ est que les coefficients de distribution β_i^E soient uniformément bornés en U, et ce indépendamment de h. Si tel est le cas, nous serons alors assurés que l'erreur commise sur la distribution est du même ordre que l'erreur commise en évaluant le résidu. De manière un peu plus formelle, on peut écrire :

$$\text{Si } \exists C > 0, \forall U_h \in V_h^k, \forall E \subset \Omega_h, \forall M_i \in E, \|\beta_i^E(U_h)\| < C, \text{ alors } \phi^E(U_h) = O(h^{k+d}) \Leftrightarrow \phi_i^E = O(h^{k+d}) \tag{3.2.15}$$

En pratique, lorsque nous parlerons de schémas $\mathcal{LP}$, il s'agira toujours de schémas dont les coefficients sont bornés vis-à-vis de U. Ceci permet d'établir un nouveau résultat sur la précision formelle des schémas $\mathcal{LP}$:

Théorème 3.2.4.
Un schéma distribuant le résidu basé sur une interpolation P_k de la solution et qui préserve la linéarité est d'ordre $k+1$ si et seulement si :

$$\forall E \subset \Omega_h,\ \phi^E(U_h) = O(h^{k+d})$$

La différence notoire entre ce théorème et le précédent tient dans les hypothèses à vérfier. Avec un schéma dont les coefficients sont bornés, les deux hypothèses du théorème 3.2.3 deviennent liées : la première entraîne la seconde. Si on revient au cas stationnaire qui est l'objet principal de ce chapitre, on vérifie trivialement que :

$$\phi_i = \beta_i \phi = O(1) \times \int_{\partial E} \left(\overrightarrow{F}_h(U_h) - \overrightarrow{F}(U) \right) d\partial E = O(h^{d-1}) \times O(h^{k+1}) = O(h^{k+d})$$

De plus, nous avons dit plus haut (3.2.15) que la seconde hypothèse du théorème 3.2.3 équivalait à celle du théorème 3.2.4. Par conséquent, un schéma $\mathcal{LP}$ basé sur une interpolation de degré k de la solution et des flux est d'ordre $k+1$. La seule hypothèse étant désormais celle de 3.2.4, l'approximation des flux à l'ordre $k+1$ n'est plus qu'un moyen et non une condition indispensable. Pour obtenir la précision désirée, on peut de manière équivalente la remplacer par une formule de quadrature sur ϕ qui soit d'ordre $k+1$. Ceci est à mettre en relation avec la remarque 5 et le fait que certaines formules de quadrature à un certain ordre équivalent l'intégration exacte d'une certaine interpolée du même ordre.

En pratique, on cherchera donc toujours à garantir la précision du schéma en travaillant avec des distributions $\mathcal{LP}$. Pour ceux qui ne respecteraient pas la condition $\mathcal{LP}$, nous disposons d'une méthode de limitation qui consiste à projeter d'une certaine façon les coefficients β_i^E sur un sous-espace borné. Ce point sera abordé plus loin.

3.2.5 Théorème de Godunov

Quelques manipulations algébriques sur le schéma prototype scalaire (3.2.6) permettent de montrer le théorème suivant, dit de Godunov [42] :

Théorème 3.2.5. *Il n'est pas possible de construire un schéma linéaire qui soit à la fois d'ordre élevé ($\mathcal{LP}$) et qui préserve la monotonie (positif).*

On précise qu'on parle de linéarité par rapport à u. Autrement dit, on entend par schéma linéaire une distribution de type (3.2.6) dont les coefficients c_{ij} sont indépendants de u. Pour une démonstration dans le cadre du formalisme $\mathcal{RD}$, on peut se référer à [70], [62], [90] ou [59]. L'idée est de supposer que le schéma est linéaire et $\mathcal{LP}$ puis d'étudier les coefficients c_{ij} pour montrer qu'il y en a nécessairement au moins un qui est négatif (le schéma n'est alors pas positif). L'implication fondamentale de ce théorème est que tous les schémas que nous voudrons monotones et d'ordre élevé devront nécessairement être non linéaires.

3.3 Présentation des principaux schémas sur des triangles P_1

3.3.1 Préambule sur la notion de décentrement

Dans cette nouvelle partie, nous allons présenter les quatre schémas que nous avons employés pour résoudre les équations de la MHD. Deux classes se distinguent : les schémas upwind multidimensionnels ($\mathcal{MU}$ pour *Multidimensional Upwind*) et les schémas dits centrés (non pas en référence au schéma de Galerkin, i.e. dans le cadre de la formulation Petrov-Galerkin, mais dans le sens d'une équirépartion de la fluctuation). Le caractère *upwind* est la dernière notion de stabilité que nous ayons à introduire. Du point de vue physique, l'information se propage dans certaines directions avec une certaine vitesse. Or, le système étant hyperbolique, nous avons accès, ne serait-ce que de manière approchée, à ces informations. Il paraît donc judicieux d'en tenir compte, même si ce n'est a priori pas une obligation. Lorsque c'est le cas, c'est-à-dire que le schéma distribue les résidus en fonction de la propagation de l'information, on dit qu'il est *upwind*. Cette approche a d'ailleurs démontré son efficacité en se révélant être la meilleure pour résoudre les équations scalaires d'advection linéaire. La désignation $\mathcal{MU}$ renvoie plus spécifiquement au même caractère considéré pour des problèmes de dimension $d > 1$ en espace. Si on ne tient pas compte de ces informations, on peut assister à l'apparition de modes parasites d'origine purement numérique dans la solution. Toutefois, même pour des problèmes scalaires, on ne sait pas construire de façon systématique de schémas qui soient parfaitement $\mathcal{MU}$ tout en étant à la fois parfaitement monotones et d'ordre élevé. Par conséquent, chacun des schémas que nous allons présenter peut être vu comme le résultat d'un choix entre quelques-unes des propriétés évoquées jusqu'ici.

Pour une majorité d'auteurs, dans le formalisme $\mathcal{RD}$, le décentrement fait référence au caractère *upwind*. Parmi les quatre schémas de base que nous allons présenter, seuls les deux premiers sont concernés par cette notion (ce sont des schémas $\mathcal{MU}$). On se restreint à nouveau à la résolution de problèmes stationnaires.

3.3.2 Le schéma Narrow (N)

Le schéma N a été développé à partir des idées de Roe ([81], [79]), et étendu aux systèmes par van der Weide et Deconinck ([96]). Plaçons-nous dans le cas bidimensionnel d'un maillage d'éléments triangulaires P_1. Notons que d'après la définition des normales (figure 3.5), chaque normale pointe vers le degré de liberté associé. Elles constituent donc un bon outil pour estimer dans quelle mesure l'information se propage vers un degré de liberté ou non. Ensuite, on utilise une jacobienne moyenne sur l'élément pour évaluer de manière approximative si chaque degré de liberté est en amont ou en aval de chaque onde. Ceci se traduit par l'introduction des matrices K_i :

$$\forall M_i \in E,\ K_i = \frac{\partial \overrightarrow{F}}{\partial U}(\overline{U}) \cdot \overrightarrow{n_i} \tag{3.3.1}$$

Une simple moyenne arithmétique sur les degrés de liberté suffit à définir $\overline{U}$. Si on diagonalise ensuite cette matrice, on obtient des valeurs propres du type $\overrightarrow{v} \cdot \overrightarrow{n_i}$ dont le signe permet de qualifier chaque

noeud d'amont ou d'aval. D'après ce que nous avons dit, l'idée d'un schéma *upwind* est de n'envoyer d'information que si cette quantité est positive. Par conséquent, la distribution doit être régie par la quantité suivante :

$$K_i^+ = R_i \Lambda_i^+ L_i \tag{3.3.2}$$

où les matrices R, Λ, L sont issues de la décomposition de chaque matrice K_i, et où la matrice diagonale Λ_i^+ est définie par :

$$\forall 1 \leq l \leq p, \quad \left(\Lambda_i^+\right)_{ll} = \max\left(0, (\Lambda_i)_{ll}\right)$$

Remarque 6. *Nous avons vu dans le chapitre précédent que le système symétrisable pour lequel le système propre a été obtenu est différent du système conservatif de la MHD idéale que nous résolvons. Toutefois, la différence tenant en des termes proportionnels à $\overrightarrow{\nabla} \cdot \overrightarrow{B}$ et à $\overrightarrow{\nabla}\psi$ que nous souhaitons maintenir faibles, on estime qu'on peut raisonnablement utiliser ce système dans nos méthodes. Ainsi, dès que nous aurons besoin de faire appel au système propre, ce sera celui du système symétrisable. Le même parti pris avait été adopté par [30], sans l'approche divergence cleaning.*

La distribution du schéma N appliqué aux systèmes est une généralisation du cas scalaire où il préserve la monotonie (sous certaines conditions sur lesquelles nous reviendrons plus loin). Elle s'écrit donc comme fonction explicite des inconnues U_i :

$$\phi_i^{\mathrm{N}} = K_i^+ \left(U_i - \tilde{U}\right) \tag{3.3.3}$$

Comme noté dans [31], si on souhaite éviter d'avoir à gérer des erreurs de conservation, la quantité moyenne $\tilde{U}$ doit être définie de manière à assurer (3.1.9) :

$$\begin{aligned} \phi &= \sum_{i=1}^{N_s} \phi_i^{\mathrm{N}} \\ &= \sum_{i=1}^{N_s} K_i^+ U_i - \left(\sum_{i=1}^{N_s} K_i^+\right) \tilde{U} \end{aligned}$$

où N_s désigne le nombre de degrés de liberté dans l'élément. On obtient donc :

$$\begin{aligned} \tilde{U} &= N \left(\sum_{i=1}^{N_s} K_i^+ U_i - \phi\right) \\ \text{avec} \quad N &= \left(\sum_{i=1}^{N_s} K_i^+\right)^{-1} \end{aligned}$$

Le point sensible dans la construction de ce schéma est donc l'existence de la matrice N. Une singularité se produit lorsqu'une valeur propre de la matrice à inverser devient nulle. Dans [1], il est montré que pour un système symétrisable, cela ne peut se produire que pour tout vecteur propre commun à toutes les matrices K_i. Si on se réfère au chapitre précédent où le système propre est exposé, on se rend compte que de la même manière qu'en mécanique des fluides, cela ne concerne en MHD que le vecteur propre r_0 associé à l'entropie car les normales utilisées pour définir les matrices K_i ne sont pas colinéaires deux à deux. Ainsi, si la valeur propre associée à r_0 s'annule pour toutes les matrices K_i la matrice N ne peut plus être calculée. Comme $\lambda_0 = u_n$ et que les normales $\overrightarrow{n}_i$ sont linéairement indépendantes deux à deux, ceci ne peut se produire que lorsque $\overrightarrow{u} = \overrightarrow{0}$. Même dans ce cas, il est montré qu'il est toujours possible de définir la distribution de manière équivalente sans passer par l'inversion de N (le schéma possède une limite finie lorsque $\overrightarrow{u} \to \overrightarrow{0}$). Toutefois en pratique, nous ne modifions pas la définition du schéma pour ce cas particulier. On préfère appliquer systématiquement une légère correction "entropique", qui s'inspire

d'une correction de Harten ([47]), sur les valeurs propres λ_k^+ (cela consiste à régulariser la fonction partie positive à l'approche de 0 d'une certaine manière, telle que $f(0) = \varepsilon > 0$).

On peut montrer que ce schéma offre un taux minimal de diffusion numérique croisée, pour la catégorie des schémas stables en norme d'énergie dans le cas de flux linéaires. Le lecteur intéressé pourra consulter par exemple [1] ou [75]. Nous y revenons en annexe, ou nous évoquons par la même occasion la question de la stabilité entropique, qui est d'intérêt pour nos problèmes non linéaires.

3.3.3 Le schéma Low Diffusion A (LDA)

Le schéma LDA est le second schéma $\mathcal{MU}$. Sur des triangles P_1, il distribue la fluctuation selon les paramètres K_i^+.

$$\phi_i^{\text{LDA}} = K_i^+ N \phi$$

La matrice N est celle définie pour le schéma précédent. D'après sa définition, on voit clairement que les coefficients β_i (qui sont désormais des matrices) sont bornés. Ce schéma est donc $\mathcal{LP}$, ce qui signifie que sur un triangle P_1, l'interpolation étant de degré 1 et d'après la proposition 3.2.3, si les flux sont interpolés de façon linéaire, il est d'ordre 2. La conservation se vérifie de façon immédiate. En revanche, le schéma LDA ne préserve généralement pas la monotonie dans le cas scalaire et n'est pas stable en énergie une fois appliqué aux systèmes linéaires. Il n'est donc clairement approprié qu'aux problèmes réguliers, c'est-à-dire ne faisant pas intervenir de discontinuités. Dans ces situations, pour des flux linéaires, il offre même un taux de diffusion d'énergie plus bas que celui du schéma N (voir par exemple [75]).

3.3.4 Le schéma de Lax-Friedrichs (LxF)

Nous passons à présent aux schémas dits centrés. Le schéma aux résidus de Lax-Friedrichs provient d'une reformulation du schéma de Rusanov en Volumes Finis, qui, sur des triangles P_1, s'écrit :

$$\phi_i^{\text{VFRus}} = \frac{1}{2}\left(\overrightarrow{F}_1 + \overrightarrow{F}_2\right)\cdot \overrightarrow{n}_{12} + \frac{1}{2}\alpha_{12}(U_1 - U_2) + \frac{1}{2}\left(\overrightarrow{F}_1 + \overrightarrow{F}_3\right)\cdot \overrightarrow{n}_{13} + \frac{1}{2}\alpha_{13}(U_1 - U_3)$$

où, l'opérateur $\rho(.)$ désignant le rayon spectral d'une matrice :

$$\alpha_{ij} = \max_{l\in\{i,j\}}\left(\rho\left(\frac{\partial \overrightarrow{F}}{\partial U}(U_l)\cdot \overrightarrow{n}_{ij}\right), \rho\left(\frac{\partial \overrightarrow{F}}{\partial U}(U_l)\cdot \overrightarrow{n}_{ji}\right)\right)$$

C'est la plus grande valeur propre calculable à l'interface entre les zones C_i et C_j. On a vu qu'on pouvait reformuler de manière équivalente le schéma Volumes Finis en schéma distribuant le résidu, en ajoutant un terme :

$$\phi_i^{\text{RDRus}} = \phi_i^{\text{VFRus}} - \overrightarrow{F}_1\cdot(\overrightarrow{n}_{12} + \overrightarrow{n}_{13}) = \phi_i^{\text{VFRus}} - \overrightarrow{F}_1\cdot\overrightarrow{n}_1$$

Cependant, on ne souhaite pas nécessairement garder exactement le schéma VF de Rusanov. Tout ce qu'on souhaite, c'est conserver le caractère diffusif du schéma de Rusanov avec une expression simple. Le schéma $\mathcal{RD}$ qu'on construit alors, qu'on peut appeler de Rusanov comme de Lax-Friedrichs (de manière rigoureuse ce n'est exactement ni l'un ni l'autre dans aucun autre formalisme), s'assure d'être diffusif en majorant tous les α_{ij} précédents par un α^T constant sur l'élément et remplace la partie des flux par une expression centrée consistante, plus simple et permettant clairement de rester conservatif dans l'optique du passage à des configurations plus complexes (on y reviendra) :

$$\phi_i^{\text{LxF}} = \frac{1}{3}\phi + \alpha\sum_{j\neq i}(U_i - U_j) \tag{3.3.4}$$

avec

$$\alpha := h \sup_{\overrightarrow{n} \in \mathbb{R}^d} \left[\sup_{\overrightarrow{x} \in T} \left(\rho \left(\frac{\partial \overrightarrow{F}}{\partial U}(U_h(\overrightarrow{x})) \cdot \frac{\overrightarrow{n}}{\|\overrightarrow{n}\|} \right) \right) \right]$$

où h est pris comme étant la plus grande longueur d'arête de l'élément. Le terme diffusif ne permet pas au schéma de Rusanov d'être $\mathcal{LP}$, donc il n'est que d'ordre 1, comme le schéma N. Il est en outre beaucoup plus diffusif que ce dernier. Bien que ceci puisse être un défaut, cela lui confère une grande robustesse. Comme pour le schéma N, on peut montrer très facilement qu'en scalaire, il préserve la monotonie pour toute loi de flux non-linéaire. En effet, si celle-ci est au moins Lipshitzienne, il se base sur une majoration de la constante de Lipschitz sur l'élément. Dans le cas de systèmes linéaires, pour la même raison, il dissipe l'énergie. Pour des systèmes non-linéaires en revanche, l'analyse de stabilité entropique est moins évidente (voir par exemple [89]). Toutefois, il convient de remarquer que le paramètre α, qui contrôle la diffusion numérique, peut être pris aussi grand que nécessaire pour s'assurer que cette diffusion parvienne à lisser les discontinuités, sans créer d'oscillations, c'est-à-dire qu'il soit bien LED (variable par variable, α pouvant être vu comme une matrice diagonale). Cette façon de procéder est tout à fait consistante avec la résolution approchée d'un problème de Riemann en Volumes Finis. En effet, le flux Volumes Finis de Rusanov est une solution approchée à 1 état intermédiaire constant du problème de Riemann. Cette solution est construite de manière symétrique et ses vitesses d'ondes majorent celles de la résolution exacte du problème de Riemann (figure 3.8).

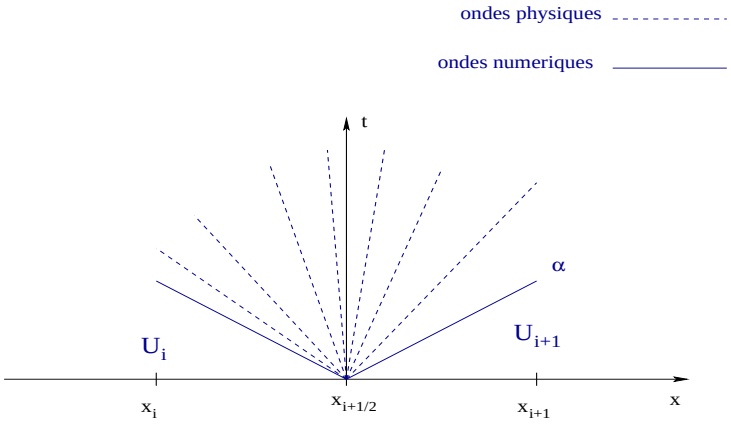


FIGURE 3.8 – Illustration 1D du solveur de Riemann approché de Rusanov sur un élément $T_{i+\frac{1}{2}}$

Par conséquent, si on imagine que le schéma N puisse théoriquement être pris en défaut en ce qui concerne sa positivité, on a de très bonnes raisons de penser que le schéma de Lax-Friedrichs pourra toujours être construit de manière à être positif. En pratique, le schéma préserve très bien la monotonie en MHD idéale, qualité à mettre en perspective avec, encore une fois, son taux de diffusion très important.

3.3.5 Le schéma Streamline Upwind (SU)

Enfin, le dernier schéma centré proposé est issu des méthodes d'Éléments Finis. Tout comme nous avons dit que le schéma de Galerkin était un schéma $\mathcal{RD}$, on peut inclure le célèbre schéma SUPG (*Streamline Upwind Petrov-Galerkin*), introduit par Brooks et Hughes [20], dans le formalisme $\mathcal{RD}$. Le schéma de Galerkin, purement centré, étant inconditionnellement instable et efficace presque exclusivement lors de la résolution de problèmes elliptiques, il fallait trouver autre chose pour pouvoir aborder des problèmes hyperboliques non-linéaires. L'idée pour le stabiliser fut alors de le décentrer (en lui donnant un caractère *upwind* pour véritablement améliorer la stabilité), en modifiant les fonctions tests mais sans sortir du

cadre Éléments Finis d'une formluation variationnelle, ce qui constitue l'approche Petrov-Galerkin. Mais comme ceci entraîne un surplus de diffusivité numérique, le choix adopté dans [20] fut de modifier les fonctions test de façon à n'introduire le décentrement que selon les "lignes de courant" (d'où le terme *Streamline*, l'application visée étant la mécanique des fluides). Pour plus de précisions, on pourra aussi consulter [98]. C'est ce qui aboutit au schéma SUPG, dont l'écriture en formalisme $\mathcal{RD}$ serait :

$$\phi_i^{\text{SUPG}} := \int_E \left(\varphi_i \overrightarrow{\nabla} \cdot \overrightarrow{F} + (\overrightarrow{\lambda} \cdot \overrightarrow{\nabla} \varphi_i) \tau r(U_h) \right) dV \tag{3.3.5}$$

où le vecteur de matrices $\overrightarrow{\lambda}$ désigne les jacobiennes :

$$\overrightarrow{\lambda}(\overrightarrow{x}) := \frac{\partial \overrightarrow{F}}{\partial U}(U(\overrightarrow{x}))$$

Remarque 7. *Nous faisons usage à dessein de la notation $r(U_h)$ temporairement, pour mettre en évidence le fait que c'est toute l'équation qui doit être intégrée. Par conséquent, l'application de ce schéma aux cas instationnaires doit être réalisée avec :*

$$r(U_h) = \frac{\partial U_h}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h)$$

et ce, en prenant ensuite en compte la façon dont est discrétisé le problème en temps. La formulation SUPG est en fait à interpréter comme une méthode de Galerkin modifée par une formulation moindres carrés qui doit s'appliquer à toute l'équation. On retrouve aussi ces méthodes sous le nom de GLS (pour Galerkin Least Squares, voir [51] et [73] pour plus de précisions).

Bien qu'on ait parlé de lignes de courant, ce terme n'est qu'une illustration. La méthode fonctionne pour tout système hyperbolique. La même remarque aurait pu être faite à propos du terme *upwind* qui signifie "amont", emprunté à un vocabulaire évoquant plutôt la mécanique des fluides, mais qui se généralise au déplacement d'une information de nature quelconque.

On vérifie trivialement qu'il s'agit bien d'un schéma distribuant le résidu car $\sum_{M_i \in E} \varphi_i = 1$ et donc $\sum_{M_i \in E} \overrightarrow{\nabla} \varphi_i = \overrightarrow{0}$, ce qui permet d'écrire :

$$\sum_{M_i \in E} \phi_i^{\text{SUPG}} = \int_E \sum_{M_i \in E} \varphi_i \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h) dV + \int_E \left(\overrightarrow{\lambda} \cdot \overrightarrow{\nabla} \left(\sum_{M_i \in E} \varphi_i \right) \right) \tau r(U_h) dV = \phi^E$$

Cependant, puisque nous ne sommes pas tenus de "coller" à la formulation variationnelle en formalisme $\mathcal{RD}$, ou plutôt comme la formulation variationnelle dans laquelle nous nous inscrivons n'est pas celle de Galerkin, mais une de Petrov-Galerkin qui dépend du schéma, on peut définir une version autrement centrée et plus simple à évaluer. Pour cela, on se ramène au cas de triangles P_1 et, de plus, on interpole les flux de façon linéaire (ce qui ne change pas l'ordre du schéma d'après la proposition 3.2.3). Dans ce cas, les signaux envoyés par le schéma SUPG s'écrivent :

$$\begin{aligned}
\phi_i^{\text{SUPG}} &= \int_T \varphi_i \overrightarrow{\nabla} \cdot \sum_{M_j \in T} \overrightarrow{F}(U_j) \varphi_j dV + \int_T \left(\overrightarrow{\lambda} \cdot \overrightarrow{\nabla} \varphi_i \right) \tau r(U_h) dV \\
&= \frac{|T|}{3} \sum_{j \in T} \overrightarrow{F}_j \cdot \frac{\overrightarrow{n_j}}{2|T|} + \int_T \left(\lambda \cdot \overrightarrow{\nabla} \varphi_i \right) \tau r(U_h) dV \\
&= \frac{1}{3} \phi_h(U_h) + \int_T \left(\lambda \cdot \overrightarrow{\nabla} \varphi_i \right) \tau r(U_h) dV
\end{aligned}$$

où nous avons noté ϕ_h la fluctuation calculée de manière approchée (en linéarisant les flux). Notons aussi que la paramètre matriciel τ est constant par élément. Nous allons revenir sur sa définition, mais avant cela disons qu'on peut définir le schéma SU (*Streamline Upwind*) par analogie avec le calcul précédent :

$$\phi_i^{\mathrm{SU}} = \frac{1}{3}\phi + \int_T \left(\lambda \cdot \overrightarrow{\nabla}\varphi_i\right) \tau r(U_h) \tag{3.3.6}$$

Sur des triangles P_1, la seule différence est que nous nous permettons de calculer ϕ autrement qu'en linéarisant les flux, par exemple en intégrant sur les bords de l'élément avec une méthode des trapèzes ou du point milieu qui donnent la précision requise par le théorème 3.2.3, ou d'autres encore. Comme ce schéma est clairement $\mathcal{LP}$, cette remarque est à mettre en relation avec le théorème 3.2.4, et donc la possibilité d'utiliser comme alternative une formule de quadrature directe sur ϕ. L'avantage se trouve surtout dans le passage à des ordres supérieurs, comme on le verra plus loin.

Pour terminer la description du schéma SU, il nous faut préciser τ. Une analyse dimensionnelle rapide de 3.3.6 permet de trouver ses dimensions :

$$[h^d][\lambda][h^{-1}][\tau]\frac{[\phi_h]}{h^d} = [\phi_h]$$

$$\Rightarrow \qquad [\tau] = \frac{[h]}{[\lambda]}$$

Dans le cas scalaire, τ a la dimension d'un temps caractéristique de maille et est de la forme $\frac{h}{\|\lambda\|}$. En systèmes, il devient une matrice dont chaque composante a cette dimension. Dans tous les cas, il exprime la quantité de diffusion à apporter selon chaque direction. Parmi les travaux effectués sur les méthodes d'Éléments Finis stabilisés, on trouve plusieurs définitions pour τ (voir [91] ou [98]). Celle que nous retenons est la suivante :

$$\tau = h^d N \tag{3.3.7}$$

où N est toujours la matrice définie par le schéma éponyme. Cette définition est consistance à l'analyse précédente du fait que $[N] = \frac{1}{[\lambda][h]^{d-1}}$ (car chaque normale selon laquelle on projette les jacobiennes, et qui sert donc au calcul de N, est de norme égale à la taille du bord qui lui est associé, arête ou face donc).

3.3.6 Sur le calcul de la fluctuation

On a vu que le calcul de la fluctuation ϕ est crucial pour le développement de ces schémas. Si on reste sur des éléments P_1, en ne requérant que l'ordre 2, celui-ci est mis en oeuvre très simplement par une linéarisation des flux, ou de manière équivalente une formule de quadrature des trapèzes ou du point milieu. Comme nous le referons remarquer plus loin, le fait d'intégrer les flux avec une précision d'ordre 2 n'est qu'une approximation qui vient se superposer à celle de l'interpolation des inconnues. C'est juste la méthode la plus simple pour assurer l'ordre recherché, en accord avec les théorèmes de précision que nous avons présentés. Toutefois, cela revient à supposer que les flux sont linéaires en U, ce qui n'est pas le cas. Or il pourrait théoriquement s'avérer que la représentation des flux par des fonctions linéaires ne rende pas suffisamment compte de l'état qui règne dans l'élément. Cela pourrait suffir dans une majorité de cas, mais pas toujours. En particulier, il serait envisageable qu'un élément E ne soit pas en état d'équilibre et que, par coïncidence algébrique, on obtienne une fluctuation $\phi^E = 0$. Si cela se produit, alors tous les schémas $\mathcal{LP}$ ne distribueront rien en provenance de cet élément. Si le maillage n'est pas structuré, ce problème est généralement mineur puisque les éléments voisins viendront perturber la coïncidence aux itérations suivantes. Cependant, sur un maillage structuré, on peut raisonnablement craindre que cette situation se retrouve sur les éléments voisins, ce qui peut alors poser un problème réel de convergence. Ceci illustre le fait qu'on peut considérer qu'il y a un enjeu de stabilité constitué par la représentation des

flux. C'est avec le terme de stabilité que nous reviendrons sur ce problème, lors du passage à des éléments autres que les triangles P_1. La solution à ce problème est clairement d'approcher les flux avec davantage de précision. Dans le cas des triangles P_1, au lieu d'utiliser une formule des trapèzes sur chaque arête, on peut utiliser une formule de Simpson en reconstruisant la solution au point milieu. Si on reprend la définition des normales de la figure 3.5, qu'on désigne par Γ_i l'arête opposée au noeud i, et qu'on numérote sur chaque Γ_i les sommets 1 et 3, et le point milieu qui reste à construire 2, on écrit :

$$\phi^T = -\sum_{\Gamma_i \subset \partial T} \frac{1}{6}\left(\overrightarrow{F}(U_1) + 4\overrightarrow{F}(U_2) + \overrightarrow{F}(U_3)\right) \cdot \overrightarrow{n}_i$$

avec

$$U_2 = U_1\varphi_1\left(\frac{1}{2}(\overrightarrow{x}_1 + \overrightarrow{x_3})\right) + U_3\varphi_3\left(\frac{1}{2}(\overrightarrow{x_1} + \overrightarrow{x_3})\right) = \frac{1}{2}\left(U_1 + U_3\right)$$

Nous avons également développé une autre alternative, qui pourrait être généralisée aux ordres supérieurs mais sur laquelle nous ne reviendrons pas, qui est d'interpoler les variables physiques V. L'avantage est que les flux sont polynomiaux en V, de degré au plus 4, donc faciles à intégrer exactement. En reprenant les notations ci-dessus, nous allons donner le résultat de l'intégration de $F_n(U_h)$ sur les bords Γ_i constitutifs de ∂T. Pour fixer les idées, prenons un des termes de la fluctuation :

$$\int_{\Gamma_i} \rho\,\overrightarrow{u}\cdot\overrightarrow{n}_i\,\overrightarrow{u}\,d\Gamma = \int_{\Gamma_i}\sum_{j\in\Gamma_i}\rho_j\varphi_j(\overrightarrow{x})\sum_{j\in\Gamma_i}\overrightarrow{u}_j\cdot\overrightarrow{n}_i\varphi_j(\overrightarrow{x})\sum_{j\in\Gamma_i}\overrightarrow{u}_j\varphi_j(\overrightarrow{x})d\Gamma$$

En P_1, il n'y a que deux degrés de liberté par Γ_i, numérotés localement 1 et 2. Clairement, l'expression de la fluctuation va reposer sur quelques intégrales seulement, que nous donnons ici :

$$\int_{\Gamma_i}\varphi_1(\overrightarrow{x})d\Gamma = \int_{\Gamma_i}\varphi_2(\overrightarrow{x})d\Gamma = \frac{|\Gamma_i|}{2} \qquad \int_{\Gamma_i}{\varphi_1}^2(\overrightarrow{x})d\Gamma = \int_{\Gamma_i}{\varphi_2}^2(\overrightarrow{x})d\Gamma = \frac{|\Gamma_i|}{3}$$

$$\int_{\Gamma_i}\varphi_1(\overrightarrow{x})\varphi_2(\overrightarrow{x})d\Gamma = \frac{|\Gamma_i|}{6} \qquad \int_{\Gamma_i}{\varphi_1}^3(\overrightarrow{x})d\Gamma = \int_{\Gamma_i}{\varphi_2}^3(\overrightarrow{x})d\Gamma = \frac{|\Gamma_i|}{4}$$

$$\int_{\Gamma_i}{\varphi_1}^2(\overrightarrow{x})\varphi_2(\overrightarrow{x})d\Gamma = \int_{\Gamma_i}{\varphi_2}^2(\overrightarrow{x})\varphi_1(\overrightarrow{x})d\Gamma = \frac{|\Gamma_i|}{12} \qquad \int_{\Gamma_i}{\varphi_1}^4(\overrightarrow{x})d\Gamma = \int_{\Gamma_i}{\varphi_2}^4(\overrightarrow{x})d\Gamma = \frac{|\Gamma_i|}{5}$$

$$\int_{\Gamma_i}{\varphi_1}^3(\overrightarrow{x})\varphi_2(\overrightarrow{x})d\Gamma = \int_{\Gamma_i}{\varphi_2}^3(\overrightarrow{x})\varphi_1(\overrightarrow{x})d\Gamma = \frac{|\Gamma_i|}{20} \qquad \int_{\Gamma_i}{\varphi_1}^2(\overrightarrow{x}){\varphi_2}^2(\overrightarrow{x})d\Gamma = \frac{|\Gamma_i|}{30}$$

À noter que les longueurs $|\Gamma_i|$ sont les normes respectives des normales $\overrightarrow{n}_i$ définies par 3.5. Si on note,

sur chaque Γ_i, $\overrightarrow{r}_k \cdot \overrightarrow{n_i} = r_{n_k}$, alors la fluctuation sur l'élément T se calcule par :

$$\phi^T = \sum_{\Gamma_i \subset \partial T} \begin{pmatrix} \frac{1}{3}\left(\rho_1 u_{n_1} + \rho_2 u_{n_2} + \frac{1}{2}\left(\rho_1 u_{n_2} + \rho_2 u_{n_1}\right)\right) \\ \frac{1}{12}(\rho_1+\rho_2)(u_{n_1}+u_{n_2})(\overrightarrow{u}_1+\overrightarrow{u}_2) + \frac{1}{6}(\rho_1 u_{n_1}\overrightarrow{u}_1 + \rho_2 u_{n_2}\overrightarrow{u}_2) + \\ \left(\frac{1}{2}(p_1+p_2) + \frac{1}{6}(\overrightarrow{B}_1^2 + \overrightarrow{B}_2^2 + \overrightarrow{B}_1\cdot\overrightarrow{B}_2)\right)\overrightarrow{n}_i - \frac{1}{3}(B_{n_1}\overrightarrow{B}_1 + B_{n_2}\overrightarrow{B}_2 + \frac{1}{2}(B_{n_1}\overrightarrow{B}_2 + B_{n_2}\overrightarrow{B}_1)) \\ \frac{\gamma}{\gamma-1}\frac{1}{6}\left((p_1+p_2)(u_{n_1}+u_{n_2}) + p_1u_{n_1} + p_2u_{n_2}\right) + \frac{1}{60}\left((\rho_1+\rho_2)(u_{n_1}+u_{n_2})(\overrightarrow{u}_1+\overrightarrow{u}_2)^2\right. \\ \left. +(\rho_1 u_{n_1} + \rho_2 u_{n_2})\overrightarrow{u}_1\cdot\overrightarrow{u}_2) + \frac{1}{120}\left((\rho_1+\rho_2)(u_{n_1}\overrightarrow{u}_1^2 + u_{n_2}\overrightarrow{u}_2^2) + u_{n_1}\rho_2\overrightarrow{u}_2^2 + u_{n_2}\rho_1\overrightarrow{u}_1^2\right)\right. \\ \left. +\frac{3}{40}\left(\rho_1 u_{n_1}\overrightarrow{u}_1^2 + \rho_2 u_{n_2}\overrightarrow{u}_2^2\right) + \frac{1}{12}\left(\overrightarrow{B}_1 + \overrightarrow{B_2}\right)^2(u_{n_1}+u_{n_2}) + \frac{1}{6}\left(\overrightarrow{B}_1^2 u_{n_1} + \overrightarrow{B}_2^2 u_{n_2}\right)\right. \\ \left. -\frac{1}{12}(B_{n_1}+B_{n_2})(\overrightarrow{u}_1+\overrightarrow{u}_2)\cdot\left(\overrightarrow{B}_1+\overrightarrow{B_2}\right) - \frac{1}{6}\left(B_{n_1}\overrightarrow{u}_1\cdot\overrightarrow{B}_1 + B_{n_2}\overrightarrow{u}_2\cdot\overrightarrow{B}_2\right)\right. \\ \frac{1}{6}\left((u_{n_1}+u_{n_2})\left(\overrightarrow{B}_1+\overrightarrow{B}_2\right) + u_{n_1}\overrightarrow{B}_1 + u_{n_2}\overrightarrow{B}_2 - (B_{n_1}+B_{n_2})(\overrightarrow{u}_1+\overrightarrow{u}_2)\right. \\ \left. -B_{n_1}\overrightarrow{u}_1 - B_{n_2}\overrightarrow{u}_2) + \frac{\psi_1+\psi_2}{2}\overrightarrow{n}_i \\ \frac{{c_h}^2}{2}(B_{n_1}+B_{n_2}) \end{pmatrix}$$

Pour le passage à des éléments non linéaires ou d'ordre supérieur, les calculs devraient être assez longs, au moins sur papier. Mais l'expression factorisée qu'on obtient au final, bien qu'impressionnante, ne se traduit peut-être pas forcément par un coût de calcul prohibitif. Par exemple, sur un maillage 40×40 d'éléments triangulaires P_1, la différence en temps CPU entre ce calcul de ϕ et une formule des trapèzes est à peine visible. Pour ce qui est de la qualité des résultats, elle est identique : à l'ordre 2 sur un maillage totalement non structuré, nous n'avons jamais rencontré de problème de stabilité lié à la qualité de la représentation des flux.

3.4 Extension des schémas d'ordre 1 à l'ordre 2

Dans cette section, nous allons voir comment construire de façon systématique un schéma formellement d'ordre élevé à partir de schémas d'ordre 1 dont on souhaite conserver les propriétés. Pour ce faire, nous nous restreignons encore à un maillage 2D d'éléments triangulaires P_1.

3.4.1 Limitation des résidus

Idéalement, on souhaiterait construire un schéma d'ordre arbitrairement élevé qui préserve la monotonie, c'est-à-dire capable d'appréhender les chocs en toute circonstance sous certaines contraintes de pas de temps. Comme on travaille sur des triangles P_1, on cherche à obtenir l'ordre 2. La méthode présentée ici permet de partir d'un schéma d'ordre 1 préservant la monotonie et de le rendre $\mathcal{LP}$ en appliquant une limitation de ses coefficients (matrices) de distribution. Commençons par exposer le principe dans le cas scalaire.

Cas scalaire

Les schémas N et Lax-Friedrichs peuvent se mettre sous la forme suivante :

$$\phi_i^T = \beta_i^T \phi^T$$

tout simplement en définissant :

$$\beta_i^T = \begin{cases} \dfrac{\phi_i^T}{\phi^T} & \text{si } \phi^T \neq 0 \\ 0 & \text{sinon} \end{cases}$$

Les coefficients β_i^T n'étant pas bornés, ces schémas ne sont pas $\mathcal{LP}$. Il faut cependant remarquer qu'il est possible d'appliquer une transformation à ces coefficients, à partir du moment où celle-ci permet de rester conservatif :

$$\mathcal{F} : \begin{cases} \mathbb{R} & \longrightarrow & \mathbb{R} \\ \{\beta_i\}_{1\leq i\leq N_s} & \longmapsto & \{\beta_i^*\}_{1\leq i\leq N_s} \end{cases}$$

De plus, on veut conserver la positivité des schémas d'ordre 1. Pour cela on procède comme dans [90]. Si $\phi \neq 0$ et $\beta_i \neq 0$, on peut écrire :

$$\begin{aligned} \beta_i^* \phi &= \beta_i^* \phi \frac{\phi_i}{\phi} \frac{\phi_i}{\phi} \\ &= \frac{\beta_i^*}{\beta_i} \phi_i \\ &= \frac{\beta_i^*}{\beta_i} \sum_{M_j \in T} c_{ij}(u_i - u_j) \\ &= \sum_{M_j \in T} c_{ij}^*(u_i - u_j) \end{aligned}$$

Par conséquent, pour que les coefficients c_{ij}^* soient positifs, il faut qu'en tout degré de liberté β_i et β_i^* soient de même signe. La fonction de limitation $\mathcal{F}$ doit donc respecter trois critères :

- préservation de la conservation :

$$\sum_{M_j \in T} \beta_j^* = 1$$

- préservation de la positivité :

$$\forall M_i \in T,\ \beta_i \beta_i^* \geq 0$$

- obtention d'une distribution bornée :

$$\exists C^* > 0, \forall u_h \in V_h^k, \forall T \subset \Omega_h, \forall M_i \in T,\ |\beta_i^*(u_h)| < C^*$$

Présentation de quelques limiteurs

Nous allons maintenant présenter des exemples de fonctions $\mathcal{F}$ qui satisfont ces conditions. La première a été introduite dans le cadre des schémas $\mathcal{RD}$ par Struijs ([88]) qui l'a appliquée avec succès au schéma N sur éléments P_1, ce qui a donné un schéma qu'on a alors appelé PSI (*Positive Streamline Invariant*) car il restait $\mathcal{MU}$ une fois limité. Elle fut ensuite utilisée avec les autres schémas monotones d'ordre 1. Il s'agit en fait de l'équivalent en formalisme $\mathcal{RD}$ du limiteur Volumes Finis *minmod* dû à Roe ([80]).

$$\forall M_i \in T,\ (\beta_i^*)^{\text{minmod}} = \frac{\beta_i^+}{\sum_{M_j \in T} \beta_j^+} \tag{3.4.1}$$

où nous faisons de nouveau usage de la fonction partie positive définie par :

$$x^+ = \max(0, x) \tag{3.4.2}$$

De plus, il est intéressant de remarquer que dans le cadre de triangles P_1, ce procédé de limitation peut s'interpréter de façon géométrique. En effet, on peut voir les coefficients β_i comme des coordonnées barycentriques puisque $\beta_1 + \beta_2 + \beta_3 = 1$. Les coefficients limités vérifiant aussi cette relation de conservation, on peut considérer la fonction $\mathcal{F}$ comme une projection des β_i initiaux sur un sous-espace borné. La figure

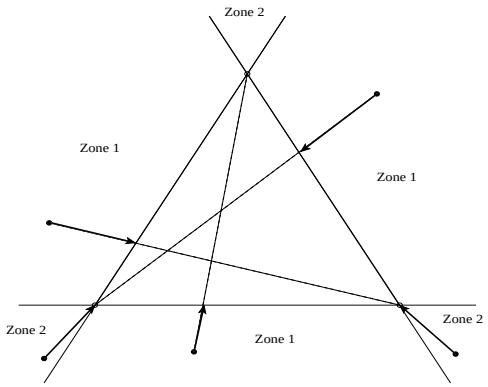


FIGURE 3.9 – Projection par la fonction *minmod* sur triangles P_1

3.9 illustre la projection opérée par le limiteur *minmod*. Les β_i sont représentés comme les coordonnées barycentriques d'un point par rapport à un certain triangle équilatéral. Si ce point se situe à l'intérieur du triangle, il n'est pas modifié. En revanche, s'il est dans une zone 1 il est projeté sur le bord le plus proche en direction du sommet opposé, et s'il est dans une zone 2, il est projecté sur le sommet le plus proche. Cette représentation géométrique permet d'imaginer facilement d'autres fonctions de limitation. Par exemple on peut citer la projection orthogonale sur les arêtes du triangle équilatéral exposée dans [6], ou encore la projection sur le cercle circonscrit à ce triangle en direction du centre de gravité. Cette dernière offre la possibilité d'obtenir des coefficients limités négatifs, à l'inverse des deux autres qui ramènent tout β_i négatif à 0.

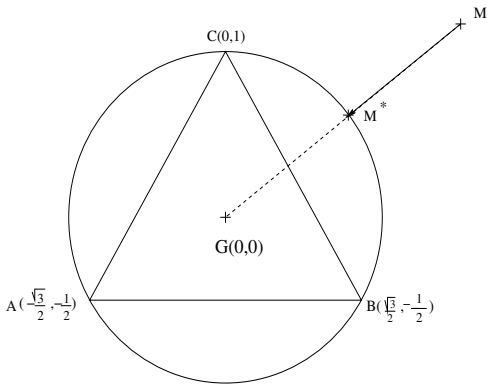


FIGURE 3.10 – Projection sur le cercle circonscrit

Pour commencer, on doit se fixer un cadre de travail, en l'occurence celui du triangle illustré par la figure 3.10. Le rayon de ce cercle est donc $r = 1$. Les coordonnées du point M dans la base barycentrique sont $(\beta_1, \beta_2, \beta_3)$ avec $\beta_1 + \beta_2 + \beta_3 = 1$ et les indices $(1, 2, 3)$ étant respectivement relatifs aux points (A, B, C). La seule chose à remarquer, c'est qu'on peut se permettre de travailler à partir de là en

coordonnées cartésiennes pour ensuite revenir aux barycentriques car le changement de coordonnées est bijectif. Donc la première étape est d'évaluer la distance $s = \|\overrightarrow{GM}\|$ séparant M du centre G.

$$\begin{aligned} s &= \sqrt{\left(\beta_1 x_A + \beta_2 x_B + \beta_3 x_3\right)^2 + \left(\beta_1 y_A + \beta_2 y_B + \beta_3 y_C\right)^2} \\ &= \frac{1}{2}\sqrt{3\left(\beta_2 - \beta_1\right)^2 + \left(1 - \beta_3\right)^2} \end{aligned}$$

Ensuite, il y a deux possibilité. Si $s \leq r$ on ne fait rien, sinon on projette en faisant une simple homothétie sur $\overrightarrow{GM}$:

$$\overrightarrow{GM^*} = \frac{r}{s}\overrightarrow{GM}$$

Ceci aboutit à un système linéaire de 3 équations à 3 inconnues :

$$\begin{cases} \beta_1^*\left(-\dfrac{\sqrt{3}}{2}\right) + \beta_2^*\dfrac{\sqrt{3}}{2} = \dfrac{r}{s}\left[\beta_1\left(-\dfrac{\sqrt{3}}{2}\right) + \beta_2\dfrac{\sqrt{3}}{2}\right] \\ \beta_1^*\left(-\dfrac{1}{2}\right) + \beta_2^*\left(-\dfrac{1}{2}\right) + \beta_3^* = \dfrac{r}{s}\left[\beta_1\left(-\dfrac{1}{2}\right) + \beta_2\left(-\dfrac{1}{2}\right) + \beta_3\right] \\ \beta_1^* + \beta_2^* + \beta_3^* = 1 \end{cases}$$

dont la solution se trouve très facilement et peut s'exprimer ainsi :

$$\forall M_i \in T,\ \ \left(\beta_i^*\right)^{\text{cercle}} = \frac{1}{3}\left(1 - \frac{r}{s}\right) + \frac{r}{s}\beta_i$$

On laisse volontairement r sans le remplacer par 1 car de fait, on pourrait aussi choisir un cercle centré en G mais de rayon supérieur.

Faisons dès à présent quelques remarques. Si on augmente trop le rayon du cercle dans la technique précédente, par exemple à partir de $r = 2$, l'expérience montre qu'on autorise une trop large plage de β_i à rester tels que définis par le schéma d'ordre 1, ce qui n'est pas bon. Par exemple, si on utilise le schéma de Lax-Friedrichs, on observe que la diffusion numérique reste importante dans de grandes régions de la solution. On ne devrait donc pas augmenter r trop au-delà de 1, à voir selon chaque cas test. L'autre remarque importante est que les différences observables entre les solutions obtenues par différents limiteurs sont minimes, hors un cas exceptionnel. Il s'agit du fait que si nous n'autorisons pas de coefficients négatifs, il est tout à fait possible qu'un degré de liberté ne reçoive aucune information de tous les éléments auxquels il appartient, surtout si le schéma à limiter n'est pas lui-même *upwind* d'une manière quelconque. C'est un cas potentiellement rare mais envisageable. Une méthode pour corriger ce défaut a été imaginée par Mezine [62] mais elle ne se généralise pas aux systèmes. La limitation sur le cercle l'élimine naturellement puisque le seul cas où un coefficient est ramené à 0 se produit lorsqu'il est aligné à un sommet et au centre de gravité (ce qui s'interprète comme le fait qu'il existe 2 des 3 sommets pour lesquels $\beta_j = \beta_k$, et dans ce cas la projection sur le 3^e sommet signifiera $\beta_j^* = \beta_k^* = 0$). Or il est très improbable que ce cas puisse se produire en un noeud et simultanément sur chaque élément contenant ce noeud.

Extension aux systèmes

Lorsqu'on résout un système de lois de conservation, les fluctuations partielles ϕ_i deviennent des vecteurs à p composantes et les coefficients β_i des matrices $p \times p$. Si on décide d'appliquer les méthodes de limitation scalaires composante par composante, le schéma sera effectivement $\mathcal{LP}$ mais on altère le couplage naturel des équations. Ceci se traduit par des oscillations importantes lors des essais numériques. Pour atténuer ce phénomène, l'idéal serait de pouvoir découpler les équations au moyen d'un changement de variables, ce que nous avons dit être impossible dans le cas général.

C'est toutefois l'idée qui va nous guider. Sur chaque élément T, on définit un vecteur de jacobiennes moyennes qu'on projette suivant un vecteur $\overrightarrow{\xi}$:

$$\overline{K} = \frac{\partial \overrightarrow{F}}{\partial U}\left(\overline{U}\right) \cdot \overrightarrow{\xi}$$

où $\overline{U}$ est défini comme une simple moyenne arithmétique sur les valeurs de U dans T. L'expérience montre que le vecteur $\overrightarrow{\xi}$ ainsi que le type de moyenne employé n'ont pas un grand impact sur le résultat. $\overrightarrow{\xi}$ peut donc être pris comme étant la vitesse moyenne ou le champ magnétique moyen. Cependant, afin d'éviter certaines situations particulières que nous avons évoquées au chapitre 2, comme un champ magnétique aligné avec $\overrightarrow{\xi}$ ou orthogonal à celui-ci, on préfère le définir de la façon suivante :

$$\overrightarrow{\xi} = \frac{1}{2}\left(\overrightarrow{\overline{B}} + \overrightarrow{\overline{B}}_{\perp}\right) \tag{3.4.3}$$

en le normalisant juste après. Ensuite, on calcule les matrices du système propre de $\overline{K}$, i.e. les matrices R, Λ, L telles que $\overline{K} = R\Lambda L$. Le principe de notre méthode est alors de projeter les signaux ϕ_i sur la base des variables caractéristiques en utilisant les formes propres de $\overline{K}$:

$$\phi_i^W = L\phi_i$$

puis de limiter composante par composante ce signal projeté avec un limiteur scalaire. Une fois ceci fait, on retourne dans la base des variables conservatives via la matrice des vecteurs propres :

$$\phi_i^* = R\left(\phi_i^W\right)^*$$

On utilisant ce procédé, on applique des limitations scalaires sur un système projeté dont le couplage est le plus réduit possible. Dans la pratique, on constate que cela permet bel et bien de supprimer les oscillations liées au non respect du couplage des équations conservatives.

Remarque 8. *Dans le cas instationnaire, la limitation doit s'effectuer sur les résidus partiels Φ_i de la même manière.*

3.4.2 Stabilisation du schéma

Introduction d'un opérateur de dissipation *upwind*

Parmi les 4 schémas de base que nous avons présenté, seul le schéma de Lax-Friedrichs n'est pas du tout *upwind*. Il n'est sensible qu'au gradient de la solution. Si on le limite pour obtenir de l'ordre élevé, comme la limitation n'est pas non plus *upwind*, la solution peut présenter un défaut de stabilité. Typiquement, le cas pathologique que nous avons évoqué où aucun noeud ne reçoit d'information a le plus de chance de se produire avec cette méthode. De plus, on peut remarquer que dans les zones régulières de la solution, le schéma tend à être le schéma de Galerkin en Éléments Finis, qui est réputé pour être instable. La limitation dans ce cas ne joue pas de rôle particulier puisque les coefficients de distribution sont clairement petits. Lors de la résolution de problèmes stationnaires, la convergence ne parvient souvent pas jusqu'au bout car près de la limite, les ondes numériques résiduelles empêchent la solution de se stabiliser autour du zéro recherché. Ces modes parasites se retrouvent aussi bien sûr dans les problèmes instationnaires.

Pour ces raisons, Abgrall [2] a proposé de rajouter aux schémas limités un terme de stabilisation, qui consiste en un opérateur de dissipation selon la direction de l'écoulement. Il s'avère que ce terme revient à celui qu'on trouve dans le schéma Éléments Finis SUPG (et donc dans le schéma SU), où il remplit exactement le même office. Cependant, ce terme a tendance à détruire la préservation de la

monotonie d'un schéma. Par conséquent, on souhaite ne l'appliquer que dans les solutions régulières, où les oscillations apparaissent, et préserver la résolution des chocs. Pour cela on utilise un senseur de choc θ, constant par élément, dont nous préciserons la construction plus loin. Une distribution stationnaire limitée et stabilisée s'écrira donc sous la forme suivante (où l'indice $_S$ signifie "stabilisé") :

$$\left(\phi_i^T\right)_S^* = {\beta_i}^*\phi^T + \theta^T\mathcal{S}_i^T(U_h) = {\beta_i}^*\phi^T + \theta^T\int_T\left(\overrightarrow{\lambda}(U)\cdot\overrightarrow{\nabla}\varphi_i\right)\tau\left(\overrightarrow{\nabla}\cdot\overrightarrow{F}(U)\right)dV \tag{3.4.4}$$

où $\mathcal{S}_i^T$ désigne l'opérateur de stabilisation associé au couple (i,T) et $\overrightarrow{\lambda}(U)$ représente le vecteur des matrices jacobiennes des flux $\overrightarrow{F}(U)$. La question de la construction de la matrice τ, constante par élément, trouve les mêmes réponses que ce que nous avons dit au sujet du schéma SU. Tout comme pour le schéma SU, nous utilisons toujours $\tau = h^d N$.

Remarque 9. *Autorisons-nous ici à parler de l'extension en instationnaire sans préciser comment la discrétisation temporelle est effectuée, de façon à respecter la logique de ce chapitre. On rappelle que la distribution en instationnaire s'écrit de manière générale comme (3.1.8). La distribution instationnaire limitée et stabilisée s'écrit :*

$$\begin{aligned}\left(\Phi_i^T\right)_S^* &= {\beta_i}^*\Phi^T + \theta^T\int_T\left(\frac{\partial\varphi_i}{\partial t} + \overrightarrow{\lambda}\cdot\overrightarrow{\nabla}\varphi_i\right)\tau\left(\frac{\partial U_h}{\partial t} + \overrightarrow{\nabla}\cdot\overrightarrow{F}(U)\right)dV\\ &= {\beta_i}^*\Phi^T + \theta^T\int_T\left(\overrightarrow{\lambda}\cdot\overrightarrow{\nabla}\varphi_i\right)\tau\left(\frac{\partial U_h}{\partial t} + \overrightarrow{\nabla}\cdot\overrightarrow{F}(U)\right)dV\end{aligned}$$

car les fonctions de base φ_i n'interpolent la solution qu'en espace (c'est un choix que nous avons fait).

Dans le cas d'un système linéaire symétrisable, ce terme de stabilisation sans l'application de τ est dissipatif et renforce la stabilité énergétique du schéma :

$$\begin{aligned}\sum_{M_i\in T}\left\langle U_i, \widetilde{\mathcal{S}}_i(U_h)\right\rangle &= \sum_{M_i\in T}U_i\int_T\left(\overrightarrow{\lambda}\cdot\overrightarrow{\nabla}\varphi_i\right)\left(\overrightarrow{\lambda}\cdot\overrightarrow{\nabla}U\right)dV\\ &= \int_T\left(\overrightarrow{\lambda}\cdot\overrightarrow{\nabla}U\right)^2dV \geq 0\end{aligned}$$

Le paramètre τ permet de normaliser la dissipation introduite (il adimensionne globalement le terme). Nous prenons le même que pour le schéma SU, c'est-à-dire celui avec la matrice N. Que l'on choisisse $\tau = h^d N$ ou $\tau = \dfrac{h^d}{\alpha}Id$ (α étant le paramètre de dissipation de Lax-Friedrichs, un autre choix possible pour τ mais qui réduit davantage l'effet dissipatif), τ est une matrice définie positive. Dans le cas de la matrice N ceci est dû au fait que nous appliquons une correction "entropique". Par conséquent, pour un système hyperbolique linéaire symétrisable, le terme de stabilisation $\mathcal{S}_i(U_h)$ serait bel et bien dissipatif :

$$\sum_{M_i\in T}\langle U_i, \mathcal{S}_i(U_h)\rangle = \left\langle\overrightarrow{\lambda}\cdot\overrightarrow{\nabla}U_h, \tau\left(\overrightarrow{\lambda}\cdot\overrightarrow{\nabla}U_h\right)\right\rangle \geq 0$$

Pour des systèmes non linéaires cette propriété est conservée, mais ce n'est pas tout à fait ce qui nous intéresse. Concernant son apport en termes de stabilité entropique, qui est la propriété physiquement cohérente pour le cas non linéaire, l'analyse n'est pas aussi claire. Sous cette forme, elle n'est pas garantie, mais on pourrait la modifier pour obtenir une inégalité d'entropie. Nous y revenons dans l'annexe B. En implicite, la stabilisation offre de plus l'avantage de mieux conditionner le système linéaire à résoudre (sans plus de détails, la discrétisation en temps n'étant pas encore abordée). Pour plus de précisions, on renvoie à [2]. Enfin, ce terme ne nuit aucunement à la conservation puisqu'il est facile de voir que :

$$\sum_{M_i\in T}\int_T\left(\overrightarrow{\lambda}(U)\cdot\overrightarrow{\nabla}\varphi_i\right)\tau\left(\overrightarrow{\nabla}\cdot\overrightarrow{F}(U)\right)dV = \int_T\left(\overrightarrow{\lambda}\cdot\sum_{M_i\in T}\overrightarrow{\nabla}\varphi_i\right)\tau\left(\overrightarrow{\nabla}\cdot\overrightarrow{F}(U)\right)dV = 0$$

Sur des triangles P_1, on utilise parfois une écriture simplifiée pour le terme de stabilisation que nous allons désormais appeler $\mathcal{S}_i(U_h)$. Elle consiste à moyenner le vecteur (de matrices) d'advection $\overrightarrow{\lambda}$ dans le premier facteur. On obtient alors :

$$\begin{aligned}\mathcal{S}_i(U_h) &= \int_T \left(\overrightarrow{\lambda}(\bar{U}) \cdot \frac{\overrightarrow{n_i}}{2\,|T|}\right) |T|\, N \left(\overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h)\right) dV \\ &= K_i N \int_T \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h) dV \\ &= K_i N \phi\end{aligned}$$

où les matrices K_i sont définies comme $\frac{1}{2}\overrightarrow{\lambda}(\bar{U}) \cdot \overrightarrow{n_i}$, avec un état moyen pris par une simple moyenne arithmétique, et où on a repris la définition des normales de la figure 3.5. Le fait de moyenner est une approximation grossière, mais qui ne se justifie que par la réduction du coût de calcul. Autrement, comme ce qui nous importe est seulement de conserver un opérateur qui dissipe l'énergie dans le cas de flux linéaires, on pourrait très bien envisager de moyenner $\overrightarrow{\lambda}$ dans les deux termes. La même remarque est faite en annexes sur le terme modifié que nous y construisons. Dans tous les cas, le terme de stabilisation doit être approché avec au moins la même précision que la fluctuation ϕ^T pour ne pas détruire la précision globale du schéma.

Couplages

Une autre stratégie pourrait être envisagée pour passer à des ordres supérieurs. Le fait de chercher à appréhender les chocs à des ordres élevés n'est pas judicieux, et dans les zones de régularité de la solution, les schémas naturellement $\mathcal{LP}$ dont nous disposons sont très efficaces. Par conséquent, une idée serait de coupler deux schémas dans chaque région, un qui soit stable (ou presque) en norme d'énergie dans les zones de forts gradients et un $\mathcal{LP}$ dans les zones régulières. Pour les mêmes raisons que pour l'application du terme de stabilisation, ce couplage passe donc par la définition d'un senseur de choc, ce que nous verrons juste après. Supposons que nous disposions de θ tel que $\theta = 0$ dans les zones de chocs et $\theta \to 1$ lorsque $\overrightarrow{\nabla} U \to \overrightarrow{0}$ (nous y revenons plus loin). Prenons de plus comme schémas à coupler le schéma de Lax-Friedrichs non limité (d'ordre 1) et le schéma SU. Le schéma sera donc de la forme :

$$\phi_i^T = (1 - \theta^T)\phi_i^{LxF} + \theta^T \phi_i^{SU}$$

De la même manière, on pourrait coupler les deux schémas $\mathcal{MU}$, N et LDA, ce qui donne naissance au schéma parfois appelé schéma B (pour *blended*, voir le *B scheme* évoqué dans [1], [75], [83], [88]).

Le schéma B est une bonne solution tant que le schéma N est capable de résoudre les chocs, ce que nous avons dit ne pas être évident. De plus, les propriétés des deux schémas sont formellement perdues dans les zones de transition où θ oscille autour de $\frac{1}{2}$. Si on choisit de tester le couplage entre les schémas centrés LxF et SU, la forte diffusion du premier accentue le conflit entre les deux schémas dans les zones de transition. Mais les chocs sont résolus de manière à garantir la positivité de la densité et de la pression, au prix d'un étalement important du front de la discontinuité (chocs résolus à l'ordre 1).

Le senseur de choc

La qualité du senseur de choc est cruciale pour les deux techniques que nous avons présentées. Il est nécessairement non linéaire, reste à savoir quelle quantité est la plus sensilbe et donc la plus appropriée. Si on pourrait choisir de travailler directement sur le gradient de U, on préfère utiliser l'entropie qui est la quantité qui varie le plus lors du passage d'un choc. Il existe deux manières de procéder alors : soit évaluer un gradient d'entropie supérieur ou égal au maximum sur l'élément, soit projeter le résidu total de l'élément sur le vecteur propre d'entropie évalué en un état moyen (quantité interprétable également

comme un gradient d'entropie sur l'élément). Une discussion sur l'effet de cette dernière option sur la stabilité énergétique (entropique) peut être trouvée dans [1].

La première définition donne un senseur de choc simple du type :

$$\theta_1 = 1 - \Delta_1 s$$

avec

$$\begin{aligned} \Delta_1 s &= \frac{\max_{M_i \in T} |s_i - \bar{s}|}{\max_{M_i \in T} s_i} \\ \bar{s} &= \frac{1}{N_s} \sum_{M_i \in T} s_i \end{aligned}$$

où N_s désigne le nombre de degrés de libertés de l'élément. La seconde est un peu plus complexe, et donc coûteuse. Prenons une direction $\overrightarrow{\xi}$ quelconque constante sur l'élément et projetons une jacobienne moyenne selon celle-ci. On déduit de la matrice K_ξ ainsi obtenue un système propre $R\Lambda L$. On sélectionne ensuite le vecteur propre r_1 associé à l'entropie. À partir de là, on peut projeter la fluctuation selon r_1 (interpréter le résultat comme une fluctuation entropique) et le diviser par le volume de l'élément pour définir une variation d'entropie $\Delta_2 s$:

$$\Delta_2 s = \frac{\langle r_1, \phi \rangle}{|T|}$$

Cette quantité n'étant pas comprise entre 0 et 1, on lui applique une fonction qui doit seulement requérir $\theta_2(0) = 1$, $\theta_2 \leq 1$ et $\lim_{\pm\infty} \theta_2(\Delta_2 s) = 0$. Un exemple est donné par :

$$\theta_2 = 1 - \frac{1}{\cosh\left(\frac{1}{\Delta_2 s}\right)}$$

Dans la pratique, comme la solution semble assez indifférente au choix de la direction $\overrightarrow{\xi}$, on a coutume d'employer (3.4.3) pour minimiser les risques de singularité du système propre (i.e. de rencontrer d'éventuels points triples).

3.5 Discussion du passage à d'autres configurations

Nous n'avons pas utilisé tous les éléments que nous allons maintenant présenter durant nos travaux. Toutefois nous abordons le sujet de façon générale pour présenter la marche à suivre, et éclairer le lecteur sur les possibilités d'extension des schémas $\mathcal{RD}$ à tous les Éléments Finis de Lagrange.

3.5.1 Rappels sur les Éléments Finis de Lagrange

La méthode des Éléments Finis de Lagrange s'appuie sur de façon générale sur des polygones (en 2D) ou des polyèdres (en 3D). Sans entrer dans les détails que l'on pourra retrouver dans tout bon livre traitant des Éléments Finis (voir par exemple [27] et [26]), exposons ici les concepts qui permettent l'extension de nos méthodes. Ces entités géométriques qui constituent nos maillages servent de support à l'interpolation de la solution par des polynômes de Lagrange (puisque tel est notre choix ici). Les éléments que nous considèrerons sont les triangles P_k et les quadrangles Q_k en dimension $d = 2$ en espace, et leurs extensions naturelles en dimension $d = 3$ à savoir les tétraèdres P_k et les hexaèdres Q_k. Dans ce contexte, k représente le degré des polynômes d'interpolation φ_i, qui font aussi office de fonctions test si notre méthode est réellement une méthode de Galerkin. Pour joindre de façon conforme des tétraèdres à des hexaèdres, il est de plus nécessaire de faire appel à des éléments pyramidaux. Ceux-ci étant encore mal maîtrisés, nous n'en parlerons pas ici (pour des informations à ce sujet, on pourra consulter à profit [15]).

Nous supposerons toujours que les éléments P_k et Q_k que nous utilisons sont convexes. De cette manière, il existe toujours un $\mathcal{C}^1$-difféomorphisme permettant de transformer chaque élément du maillage E en un élément de référence noté $\widehat{E}$. On appelera cette transformation $\mathcal{F}_E^{-1}$. À chaque degré de liberté M_i contenu dans un élément, est donc associée une fonction de base φ_i (i.e. un polynôme de Lagrange donc), comme nous l'avons vu précédemment pour les triangles P_1. Elles servent à interpoler la solution mais permettent également de définir un nouveau repérage dans chaque élément, de telle sorte que :

$$\overrightarrow{x} = \sum_{M_i \in E} \varphi_i^E \left(\overrightarrow{x}\right) \overrightarrow{x}_i$$

Définir les polynômes d'interpolation sur les éléments du maillage est une tâche lourde et onéreuse. On préfère utiliser le fait que le $\mathcal{C}^1$-difféomorphisme $\mathcal{F}_E$ est unique pour définir les fonctions de base sur l'élément de référence puis ensuite les transporter vers l'élément du maillage. Si on note $\overrightarrow{X}$ les coordonnées au sein de l'élément de référence et $\widehat{\varphi}_i$ les fonctions de base sur ce même élément, alors :

$$\forall \overrightarrow{X} \in \widehat{E}, \exists ! \overrightarrow{x} \in E,\ \varphi_i\left(\overrightarrow{x}\right) = \varphi_i \circ \mathcal{F}_E(\overrightarrow{X}) = \widehat{\varphi}_i\left(\overrightarrow{X}\right)$$

où on a donc :

$$\mathcal{F}_E : \left\{ \begin{array}{ccc} \widehat{E} & \rightarrow & E \\ \overrightarrow{X} & \mapsto & \overrightarrow{x} = \displaystyle\sum_{M_i \in E} \varphi_i^{E_1}\left(\overrightarrow{x}\right) \overrightarrow{x}_i = \sum_{M_i \in E} \widehat{\varphi}_i^{E_1}\left(\overrightarrow{X}\right) \overrightarrow{X}_i \end{array} \right.$$

Les coordonnées $\overrightarrow{X}_i$ sont associées aux degrés de liberté $\widehat{M_i}$, qui sont bien sûr ceux de $\widehat{E}$ ainsi que les antécédents des $\overrightarrow{x}_i$ par la bijection $\mathcal{F}_E$. Les fonctions de base $\varphi_i^{E_1}$ et $\widehat{\varphi}_i^{E_1}$ sont celles des éléments P_1 ou Q_1, même si l'interpolation est de degré $k > 1$. En effet, quel que soit le degré de l'interpolation, le passage de l'élément du maillage à l'élément de référence est complètement décrit par les fonctions de base correspondant à tout autre degré d'interpolation. On prend donc les fonctions de plus bas degré ($k = 1$) parce qu'elles simplifient certains calculs que nous verrons plus bas (celui des jacobiens). Donnons à présent quelques exemples d'éléments de référence sur lesquels nous appuyer. Les éléments P_k 2D et 3D sont respectivement représentés par les figures 3.11 et 3.12 pour les premières valeurs de k. De même, les éléments Q_k 2D et 3D sont illustrés sur les figures 3.13 et 3.14 respectivement. Nous avons repris par défaut les conventions de numérotation de [59], qui facilitent l'écriture des algorithmes.

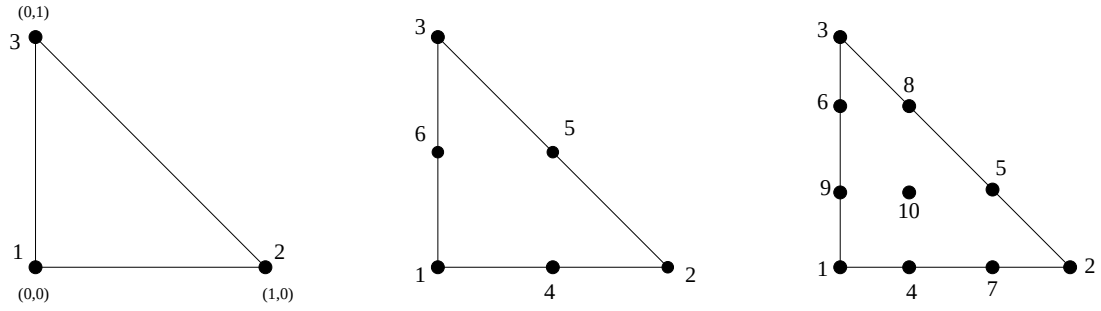


FIGURE 3.11 – Éléments P_k 2D. De gauche à droite : triangles $\widehat{T}$ de référence pour $k = 1, 2, 3$.

Afin de ne pas trop alourdir cette partie déjà un peu en marge des travaux que nous avons réalisés, nous omettons le détail des fonctions de base pour tous les éléments de référence présentés par les figures 3.11 à 3.14. Leurs expressions sont bien plus simples que dans un élément quelconque du maillage. Dans le cas des quadrangles Q_1 par exemple, les restrictions des fonctions de base de l'élément de référence aux bords de celui-ci sont linéaires. Grâce à la transformation $\mathcal{F}_E$, en transportant ces fonctions de base vers l'élément du maillage, on conserve cette propriété. Si $\mathcal{F}_E$ est facilement déterminée par les $\widehat{\varphi}_i^{E_1}$ et

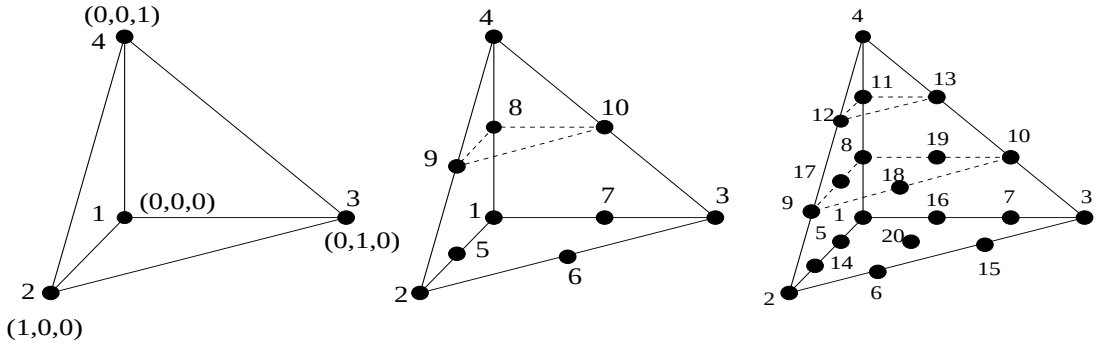


FIGURE 3.12 – Éléments P_k 3D. De gauche à droite : tétraèdres $\widehat{T}$ de référence pour $k = 1, 2, 3$. Pour mieux voir la construction, on peut noter que chaque étage l (en descendant depuis le sommet 4) forme un triangle P_l dessiné en pointillés.

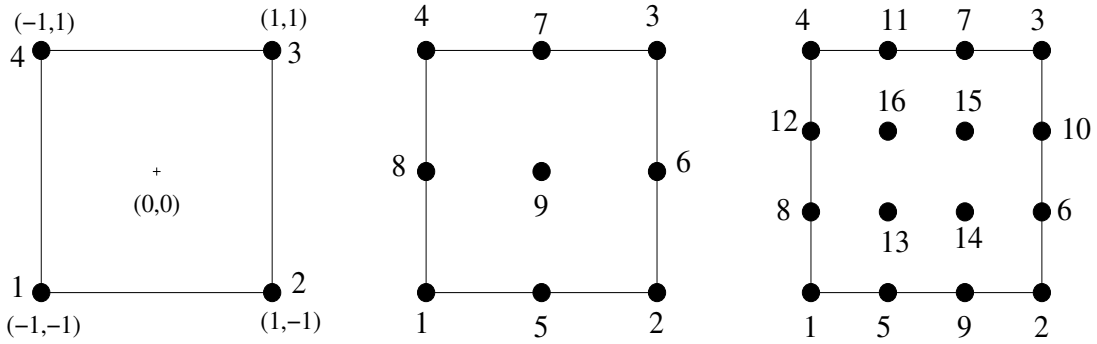


FIGURE 3.13 – Éléments Q_k 2D. De gauche à droite : quadrangles $\widehat{Q}$ de référence pour $k = 1, 2, 3$. Pour k quelconque, on dénombre $(k+1)^2$ degrés de liberté.

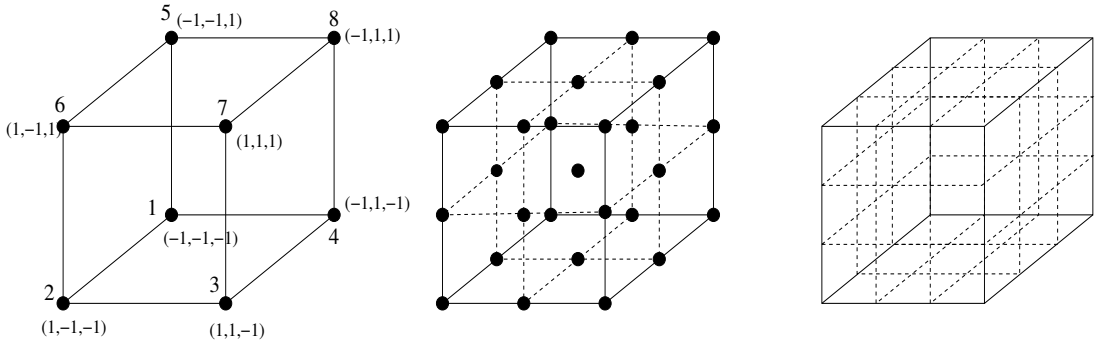


FIGURE 3.14 – Éléments Q_k 3D. De gauche à droite : hexaèdres $\widehat{Q}$ de référence pour $k = 1, 2, 3$. La construction étant difficile à représenter sur un petit espace, on omet la numérotation qui est un mélange entre celle des tétraèdres (mouvement en spirale ascendante) et celle du quadrangle Q_1. Pour l'élément Q_3, le placement des degrés de liberté se fait aux croisements des lignes dessinées (pleines et/ou discontinues), si on omet ceux situés à l'intérieur qui forment un cube similaire à Q_1. Le principe pour k quelconque est en réalité simple : identifier une face comme un élément 2D Q_k et la translater k fois de plus, de sorte qu'on ait au final $(k+1)^3$ degrés de liberté équirépartis.

les noeuds M_i, trouver sa réciproque $\mathcal{F}_E^{-1}$ demande certains efforts de calculs à faire au cas par cas (i.e. par type d'élément rencontré). Comme on en aura besoin par la suite, mais qu'une dérivation complète

pour plusieurs éléments serait bien inutilement lourde au regard de ce que nous avons réellement utilisé, on suppose dès à présent cette réciproque connue.

Remarque 10. *Nous avons choisi ici de placer les points de l'interpolation de Lagrange de façon équirépartie sur les éléments. Il est bien connu que cette méthode d'interpolation diverge pour des hautes valeurs de k, disons à partir de $k = 6$. C'est le phénomène de Runge. Comme ces valeurs n'ont encore jamais été testées pour des schémas $\mathcal{RD}$, nous ne nous étendons pas sur ce sujet. On peut simplement noter que pour des ordres aussi élevés, comme rappelé dans [56], il faut relocaliser les degrés de liberté sur chaque élément de manière symétrique (pour conserver un maillage conforme). Une façon judicieuse de procéder est de choisir comme degrés de liberté les points de Gauss-Lobatto qui permettent l'emploi direct de formules de quadrature d'ordre très élevé (l'ordre optimal pour un nombre de points de quadrature donné, tandis que l'interpolation reste de degré k). Cette méthode a déjà été proposée pour d'autres types de problèmes (voir par exemple [48]).*

3.5.2 Mise en oeuvre des schémas $\mathcal{RD}$ sur ces éléments

Cette partie a été l'objet de la thèse [59]. Nous rappelons ici le principe général. Voyons dans un premier temps comment calculer les intégrales entrant en jeu dans les différents schémas, c'est-à-dire en fait seulement la fluctuation ϕ^E sur un élément et le terme de stabilisation $\mathcal{S}_i^E$ qui apparaît également dans le schéma SUPG. Ensuite nous verrons comment étendre les principes des 4 distributions présentées dans la section 3.3.

La fluctuation

Sur les éléments d'un maillage 2D, que ce soit pour des triangles ou pour des quadrangles, la définition des fonctions de base sur l'élément de référence permet à celles-ci, une fois transportées sur l'élément réel, de posséder des restrictions aux bords que l'on peut assimiler à des polynômes d'interpolation 1D sur des segments. Par exemple, comme remarqué dans [4] et [59], pour des éléments Q_1 la restriction de chaque fonction de base sur chaque arête est linéaire. Cette propriété est extrêmement appréciable pour le calcul de la fluctuation, puisqu'elle peut se formuler comme une intégrale de contour. On rappelle que $\forall E \subset \Omega_h$:

$$\phi^E = \int_{\partial E} \overrightarrow{F}(U_h) \cdot \overrightarrow{n} d\partial E$$

où

$$U_h(\overrightarrow{x}) = \sum_{M_i \in E} U_i \varphi_i(\overrightarrow{x})$$

est, pour un élément 2D P_k ou Q_k, un polynôme de degré k sur chaque arête de E. Si on souhaite procéder de manière rigoureuse, il faut exprimer les divers termes de $\overrightarrow{F}(U_h)$ en fonction des φ_i et les intégrer exactement. Selon la nature des flux, cela peut s'avérer très complexe. Pour des flux polynomiaux, par exemple en interpolant le jeu de variables Z ou les variables physiques V à la place de U, le calcul est possible. Nous l'avons fait pour les variables physiques V sur des éléments P_1 et Q_1. Le passage à des éléments 2D d'ordre supérieur ne devrait pas être beaucoup plus difficile. En revanche, pour la MHD idéale comme pour les équations d'Euler, l'expression des flux en U fait intervenir des fractions rationnelles de polynômes de Lagrange difficiles à intégrer de manière exacte. On se contente donc d'un calcul approché. On peut considérer par exemple que les termes qui composent les flux sont au plus quadratiques en U, et dès lors deux solutions sont envisageables :

1. appliquer une formule de quadrature exacte pour un polynôme de degré $2k$ (U_h étant de degré k)
2. interpoler les flux avec des polynômes de Lagrange $\mathbb{P}^{2k}$ de degré $2k$

Une telle approximation est permise, et même plus précise que nécessaire, en vertu des théorèmes 3.2.3 et 3.2.4. Ce dernier, plus particulièrement, montre qu'il n'est pas nécessaire d'être précis à un ordre supérieur à $k+d$ sur le calcul de ϕ^E pour qu'un schéma $\mathcal{LP}$ basé sur une interpolation de degré k soit formellement d'ordre $k+1$. Ainsi, les deux solutions envisagées ci-dessus peuvent être affaiblies à un degré k (le d supplémentaire venant de l'intégration sur l'élément). C'est la stratégie que nous adoptons généralement. Pour des éléments 2D P_1 et Q_1, la quadrature consiste en une formule des trapèzes sur chaque bord, et pour des éléments P_2 ou Q_2 il s'agit d'une formule de Simpson. La discussion précédente n'est néanmoins pas sans intérêt. Si notre démarche a fait ses preuves sur les ordres de précision faibles, il semblerait que l'expérience en Éléments Finis tende à montrer que ce choix devrait être remis en cause pour des ordres très élevés (bien que le schéma reste formellement d'ordre $k+1$, des problèmes de stabilité pourraient apparaître à cause d'une approximation trop grossière des flux). Requérir une précision plus élevée que nécessaire sur les flux est d'ailleurs une des idées avancées par une autre méthode, celle des schémas $P_N P_M$ introduits dans [37]. Dans ces méthodes, c'est (entre autres choses) la seconde solution qui mise en avant : interpoler les flux avec des fonctions de forme de degré $M > N$, N étant le degré d'interpolation de la solution (dans notre exemple motivant l'exposé des deux solutions, la comparaison aurait lieu avec un schéma $P_k P_{2k}$. Quelle que soit la façon de procéder parmi les deux présentées, dès que l'on souhaite une précision d'ordre supérieur à $k+1$ sur les flux, il est nécessaire d'évaluer U_h en un certain nombre de points supplémentaires. Lesquels ? Il y a là encore deux méthodes envisageables : soit évaluer U_h en des points équirépartis de manière à conserver le plus possible les degrés de liberté déjà disponibles (par la première méthode, on aboutit alors à des formules de quadrature de Newton-Cotes d'ordre très élevé), soit prendre des points de Gauss-Lobatto de manière à formuler une quadrature d'ordre bien plus élevé avec moins de points au total. Comme on le remarquait plus haut, on pourrait même étendre l'idée en stockant d'emblée les inconnues aux points de quadrature de Gauss-Lobatto qui incluent les sommets des éléments (on pourrait donc lire le maillage et en conserver les noeuds en début de simulation, puis rajouter les noeuds internes dans un second temps). La conformité du maillage serait préservée puisque les points seraient répartis de manière symétrique, et cela aurait l'avantage pour des ordres très élevés de supprimer le phénomène de Runge, même si des oscillations moindres pourraient toujours apparaître.

Pour les éléments 3D en revanche, les calculs sont plus difficiles. À notre connaissance, il n'existe pas de mailleur capable de générer des maillages hydrides non-structurés 3D et qui ne fournisse que des hexaèdres à faces planes. Si tel était le cas et que nous utilisions cette propriété, on pourrait ramener le calcul de la fluctuation à une intégrale de contour comme en 2D, et utiliser alors le fait que les restrictions des fonctions de base de l'hexaèdre sur chaque bord sont équivalentes aux fonctions de base d'un quadrangle 2D du même ordre. De là, le problème reviendrait à approcher les flux sur une surface 2D, et à projeter le résultat sur une normale constante. À notre connaissance, aucun mailleur n'est capable de générer de tels maillages de façon automatique : soit on peut y parvenir "à la main" en pavant autant que possible des parties de l'intérieur du domaine de manière structurée et donc en obtenant des hexaèdres et des pyramides à faces quadrangulaires planes, soit il faut recombiner un maximum de tétraèdres et on ne peut plus espérer obtenir de faces quadrangulaires planes. Dans ce dernier cas, le principe général reste néanmoins le même, sauf qu'au lieu d'exprimer la fluctuation comme une intégrale de contour, puisque les normales ne sont plus constantes, on peut aussi bien considérer la fluctuation comme une intégrale de volume, et la question de la précision avec laquelle approcher les flux (donc la fluctuation) trouve les mêmes réponses que précédemment. Les solutions décrites pour les éléments 2D s'étendent naturellement en 3D (formules de quadrature d'ordre très élevé ou interpolation des flux, etc.). Pour des ordres très élevés, il suffit de connaître l'erreur commise lors d'une simple approximation d'ordre $k+d$, i.e. de pouvoir estimer la précision idéale de la formule de quadrature. Cela signifie qu'il faut connaître le degré maximal des monômes formés par la divergence des flux appliquée à une solution interpolée, dans le cas où les flux

sont polynomiaux (ou en n'interpolant pas U mais d'autres variables). Si on interpole U directement et que les flux (comme ceux de la MHD) sont non polynomiaux (ils sont rationnels), il faut choisir un degré arbitraire, qui correspondrait à un de ceux avec lequel on pourrait interpoler les flux. On peut ensuite se ramener à l'élément de référence pour mener les calculs.

$$\begin{aligned}
\phi^E(U_h) &= \int_E \overrightarrow{\nabla}_x \cdot \overrightarrow{F}(U_h(\overrightarrow{x}))dx \\
&= \int_{\widehat{E}} \sum_{l=1}^{d} \frac{\partial F_l}{\partial \overrightarrow{X}} \cdot \frac{\partial \overrightarrow{X}}{\partial x_l} \left| \det \left(\frac{\partial \mathcal{F}_E}{\partial \overrightarrow{X}} \right) \right| dX \\
&= \int_{\widehat{E}} \sum_{l=1}^{d} \overrightarrow{\nabla}_X F_l(U_h(\overrightarrow{X})) \cdot \partial_{x_l} \mathcal{F}_E^{-1} \det \left(\frac{\partial \mathcal{F}_E}{\partial \overrightarrow{X}} \right) dX
\end{aligned}$$

avec d la dimension en espace et, pour rappel,

$$U_h(\overrightarrow{X}) = \sum_{M_i \in E} U_i \widehat{\varphi}_i(\overrightarrow{X})$$

La positivité du jacobien de $\mathcal{F}_E$ est assurée par le fait que nous numérotons localement les sommets de l'élément du maillage et de l'élément de référence selon la même convention d'orientation : le sens direct. Connaissant $\mathcal{F}_E$ et $\mathcal{F}_E^{-1}$ de façon formelle, on peut alors déterminer le degré de l'intégrande ci-dessus et, selon l'ordre de précision requis et la nature des flux, trouver le meilleur compromis entre une appoximation d'ordre k et une approximation d'ordre beaucoup plus élevé de manière à pallier les problèmes de stabilité éventuels. L'approche de l'intégrale volumique cause néanmoins un défaut de conservation de l'ordre de l'erreur de troncature. Une intégration de contour, sur les faces, peut donc être préférée pour des tétraèdres, car les faces sont planes. Pour les hexaèdres, la difficulté qui nous a motivés à considérer l'intégrale de volume est que les normales aux faces ne sont pas toujours constantes, et dans ce cas, leur transport sur une face plane de référence n'est pas encore clair à nos yeux.

Le terme de stabilisation

Il s'agit cette fois d'une réelle intégrale de volume dont on rappelle l'expression dans le cas stationnaire :

$$\mathcal{S}_i^E = \int_E \left(\overrightarrow{\lambda}(U_h) \cdot \overrightarrow{\nabla} \varphi_i \right) \tau \left(\overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h) \right) dx$$

où τ est une matrice constante sur l'élément. D'après les théorèmes sur la précision formelle du schéma, cet opérateur peut être approché avec la même précision que la fluctuation. Les remarques faites au sujet du calcul de la fluctuation restent néanmoins vraies pour le terme de stabilisation. Des travaux en cours sur les équations de Navier-Stokes tendent à montrer qu'une approximation d'ordre supérieur de ce terme améliore son effet sur la convergence des problèmes stationnaires. Pour avoir une idée de la précision avec laquelle l'approcher, il faut là aussi le reformuler sur l'élément de référence :

$$\mathcal{S}_i^E = \int_{\widehat{E}} \left[\sum_{l=1}^{d} \lambda_l(U_h) \frac{\partial \widehat{\varphi}_i}{\partial \overrightarrow{X}} \cdot \frac{\partial \overrightarrow{X}}{\partial x_l} \right] \tau \left[\sum_{l=1}^{d} \frac{\partial F_l}{\partial \overrightarrow{X}} \cdot \frac{\partial \overrightarrow{X}}{\partial x_l} \right] \det \left(\frac{\partial \mathcal{F}_E}{\partial \overrightarrow{X}} \right) dX$$

Même dans le cas de flux linéaires, cette expression peut clairement être de degré supérieur à k.

La distribution

On cherche à préserver les propriétés des schémas construits sur des triangles P_1 lors du passage à d'autres éléments. Clairement, la stabilité entropique est très difficile à conserver pour tout élément. Cela représente toujours un défi majeur à l'heure actuelle. Néanmoins on peut formuler des généralisations "naturelles" des distributions présentées.

Les schémas centrés se reformulent de manière triviale sur tous les éléments :

$$\begin{aligned}\phi_i^{\text{LxF}} &= \frac{1}{N_s}\left(\phi^E + \alpha^E \sum_{M_j \in T} (U_i - U_j)\right) \\ \phi_i^{\text{SU}} &= \frac{1}{N_s}\phi^E + \int_E \left(\overrightarrow{\lambda}\cdot\overrightarrow{\nabla}\varphi_i\right)\tau\left(\overrightarrow{\nabla}\cdot\overrightarrow{F}(U_h)\right)dV\end{aligned}$$

où N_s désigne le nombre de degrés de liberté sur l'élément. L'approximation du terme SUPG vient d'être abordée. On peut vérifier facilement que ces distributions restent conservatives. Quel que soit le degré d'interpolation et l'ordre de l'erreur commise sur le calcul de la fluctuation ϕ^E, le schéma de Lax-Friedrichs reste d'ordre 1. On rappelle que le terme "centré" possède un sens différent en formalismes $\mathcal{RD}$ et Galerkin :

$$(\phi_i{}^c)^{\mathcal{RD}} = \frac{1}{N_s}\int_E \overrightarrow{\nabla}\cdot\overrightarrow{F}(U_h)dV = \frac{1}{N_s}\phi^E \quad \neq \quad (\phi_i{}^c)^{\text{Gal}} = \int_E \varphi_i \overrightarrow{\nabla}\cdot\overrightarrow{F}(U_h)dV$$

Les schémas $\mathcal{RD}$ s'insérant dans un cadre Petrov-Galerkin, les confusions qui pourraient apparaître doivent être levées en précisant la définition utilisée. Pour nous, c'est bien sûr la première.

En ce qui concerne les distributions décentrées ($\mathcal{MU}$), les seules choses qu'on souhaite vraiment préserver sont la conservation et le caractère *upwind*. La première chose à faire est donc nécessairement de définir sur quelles directions se baser. Connaissant le centre de gravité G (l'isobarycentre, puisque tous les éléments considérés sont convexes) de chaque élément, une idée assez simple et intuitive serait de définir la direction d'*upwinding* pour chaque degré de liberté i comme étant :

$$\overrightarrow{\nu_i} = \overrightarrow{GM_i}$$

On pourrait ensuite formuler les distributions LDA et N exactement de la même manière qu'en P_1, en sachant que la conservation serait assurée par la définition de la matrice N. Ce faisant, cela signifie qu'on pondère la distribution par deux informations : bien sûr, l'alignement de chaque degré de liberté avec la propagation de chaque onde, mais aussi son éloignement du centre de gravité. Or, cela signifie que dans le cas d'éléments d'ordre très élevé, certains noeuds de l'intérieur de l'élément ne recevront pas d'information, en particulier un éventuel noeud se superposant au centre de gravité comme dans le cas Q_2 ou P_3 par exemple. Ces degrés de liberté n'étant pas partagés avec d'autres éléments, il arriverait donc qu'ils ne reçoivent aucune contribution. La figure 3.15 illustre un cas pathologique sur des éléments Q_2.

Cette méthode n'est donc pas viable. On se rend en fait compte que la définition d'un schéma *upwind* n'est pas adaptée, fondamentalement, à une montée en ordre de ce genre. Par conséquent, la seule solution est d'imiter les distributions sur les éléments P_1 et Q_1, ce qui est possible en "découpant" les éléments d'ordre supérieur rencontrés dans le maillage. En effet, les distributions N et LDA vues dans les sections 3.3.2 et 3.3.3 pour le cas 2D P_1 s'étendent facilement aux quadrangles Q_1. La formule est identique, de même que le calcul de la fluctuation sur chaque bord (sauf qu'il y en a 4). La seule chose à préciser dans le cas Q_1 qui change par rapport au P_1 est la définition des directions d'*upwinding*, qui ne sont plus des normales des côtés. Deux solutions, illustrées dans la figure 3.16 : ou bien notre choix précédent en partant du centre de gravité, ou bien le choix fait dans [4] et [59] en prenant les normales aux diagonales du quadrangle. Ce sont ces directions qui permettent la définition des matrices K_i et donc de la matrice N assurant la conservation des deux schémas précités. Concernant les éléments 3D, pour les tétraèdres P_1 les normales aux faces sont proportionnelles aux gradients des fonctions de base, qui sont constants, et sont de bons candidats pour définir les directions d'*upwinding*, comme dans le cas des triangles P_1. Pour les hexaèdres Q_1, c'est exactement la même chose que pour les quadrangles Q_1 en 2D. Revenons aux éléments de degré supérieur. La figure 3.17 montre la subdivision de quelques éléments en guise

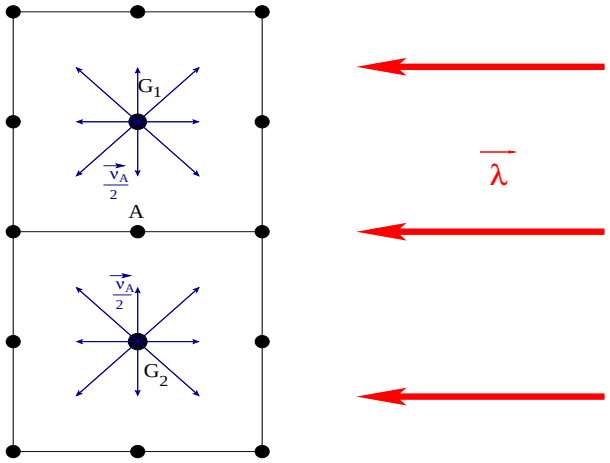


FIGURE 3.15 – Exemple de configuration singulière. Si on suppose une loi de flux linéaire, i.e. $\overrightarrow{\lambda}$ constant, et qu'on définit une distribution *upwind* de type LDA par $\phi_i^{LDA} = K_i^+ \left(\sum_{j \in E} K_j^+ \right)^{-1} \phi^Q$, avec $K_i = \overrightarrow{\lambda} \cdot \overrightarrow{\nu_i}$, alors on peut voir que : a) $K_{G_1} = 0$ puisque $\overrightarrow{\nu_{G_1}} = \overrightarrow{0}$ par définition, de même pour G_2 sur l'élément voisin Q_2, b) $\phi_A^{Q_1} = \phi_A^{Q_2} = 0$ car, dans les deux éléments, $\overrightarrow{\lambda} \cdot \overrightarrow{\nu_A} = 0$, donc A ne reçoit aucune contribution

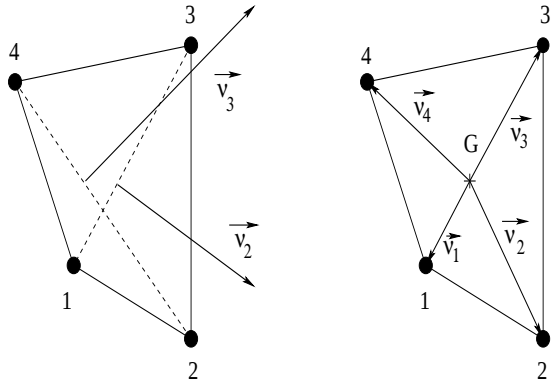


FIGURE 3.16 – Définition des directions d'*upwinding* sur un élément Q_1 : 2 possibilités

d'exemple. Quel que soit le schéma, la seule contrainte indispensable est la conservation à l'échelle de l'élément E. Or pour tout E découpé en N_E sous-éléments E^m (de type P_1 pour un élément P_k et Q_1 pour un élément Q_k), on peut définir des fluctuations locales vérifiant :

$$\phi^E = \int_E \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h) dV = \sum_{m=1}^{N_E} \int_{\partial E^m} \overrightarrow{F} \cdot \overrightarrow{n} \, d\partial E^m \tag{3.5.1}$$

$$= \int_{\partial E} \overrightarrow{F} \cdot \overrightarrow{n} \, dV = \sum_{m=1}^{N_E} \int_{\partial E^m \cap \partial E} \overrightarrow{F} \cdot \overrightarrow{n} \, d\partial E^m \tag{3.5.2}$$

$$= \sum_{m=1}^{N_E} \phi^{E^m} \tag{3.5.3}$$

La première idée pourrait donc être de distribuer dans chaque sous-élément la "sous-fluctuation" avec un schéma N ou LDA formulé selon le même principe qu'en P_1 (et qu'en Q_1). Comme ce qui nous intéresse est la conservation à l'échelle de E, on pourrait se baser aussi bien sur (3.5.1) que sur (3.5.2) pour définir

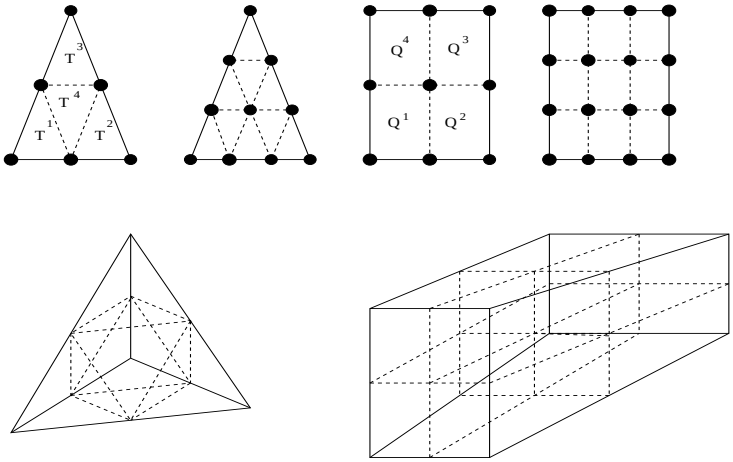


FIGURE 3.17 – Présentation schématique de quelques découpages en éléments de degré 1 pour des éléments facilement représentables. En haut et de gauche à droite : éléments 2D P_2, P_3, Q_2 et Q_3. En bas et de gauche à droite : éléments 3D, P_2 et Q_2. Pour le tétraèdre P_2, si on enlève les 4 sous-tétraèdres liés aux 4 sommets, il reste (en tirets) un octaèdre central découpable en 8 sous-tétraèdres. Concernant l'hexaèdre Q_2, notons que si ses faces ne sont pas planes, celles des sous-hexaèdres ne le seront pas non plus.

les "sous-fluctuations" Φ^{E^m}. Seulement, dans ces deux cas, cela implique d'intégrer sur chaque bord de sous-élément un polynôme au moins de degré k (si les flux sont linéarisés) avec seulement 2 points. Il faut donc construire des degrés de liberté supplémentaires sur chaque sous-élément en y calculant la valeur de U_h le temps d'approcher l'intégrale. Ce procédé est donc onéreux. Il existe une alternative citée dans [59]. Plutôt que d'appliquer le schéma *stricto sensu* sur chaque sous-élément, on pourrait formuler le schéma sans requérir la conservation locale, en travaillant à partir de la fluctuation globale et en s'assurant de sa conservation finale. Donnons les expressions qui en résultent pour fixer les idées :

$$\forall m \leq N_E, \quad \left\{ \begin{array}{rcl} \phi_i^{N,E^m} & = & \left(K_i^{E^m}\right)^+ \left(U_i - \widetilde{U}\right) \\ \phi_i^{LDA,E^m} & = & \left(K_i^{E^m}\right)^+ N\phi^E \end{array} \right.$$

avec

$$\left(K_i^{E^m}\right)^+ = \left(\frac{\partial \overrightarrow{F}}{\partial U}\left(\overline{U}^{E^m}\right) \cdot \overrightarrow{\nu_i}\right)^+ \quad \text{(cf. (3.3.2))}$$

et la construction des $\overrightarrow{\nu_i}$ dépendant du type de sous-élément (comme on l'a dit plus haut). Il faut noter que le schéma LDA distribue alors la fluctuation totale de l'élément dans chaque sous-élément. Pour le schéma N, cette approche implique que $\widetilde{U}$ soit construit de manière globale, et non par sous-élément. Dans le cas du schéma LDA, c'est la matrice N qui permet d'assurer la conservation, de la manière suivante :

$$\begin{array}{rcl} \phi^E & = & \displaystyle\sum_{m=1}^{N_E} \sum_{j \in E^m} \phi_j^{E^m} \\ & = & \displaystyle\left(\sum_{j \in E^m} \sum_{m=1}^{N_E} \left(K_j^{E^m}\right)^+ \right) N\phi^E \end{array}$$

La matrice N est donc définie comme :

$$N = \left(\sum_{j \in E^m} \sum_{m=1}^{N_E} \left(K_j^{E^m}\right)^+ \right)^{-1}$$

Pour le schéma N, c'est alors la quantité $\widetilde{U}$ qui doit permettre la conservation avec cette définition de N, à l'image de ce qui est fait dans [31] pour des éléments P_1 :

$$\begin{aligned}
\phi^E &= \sum_{m=1}^{N_E} \sum_{j \in E^m} \phi_j^{E^m} \\
&= \sum_{j \in E^m} \sum_{m=1}^{N_E} \left(K_j^{E^m}\right)^+ \left(U_j - \widetilde{U}\right) \\
\Longrightarrow \quad \left(\sum_{j \in E^m} \sum_{m=1}^{N_E} \left(K_j^{E^m}\right)^+\right) \widetilde{U} &= \sum_{j \in E^m} \sum_{m=1}^{N_E} \left(K_j^{E^m}\right)^+ U_j - \phi^E \\
\Longrightarrow \quad \widetilde{U} &= N \left(\sum_{j \in E^m} \sum_{m=1}^{N_E} \left(K_j^{E^m}\right)^+ U_j - \phi^E\right)
\end{aligned}$$

Ces distributions sont plus simples et donc moins onéreuses que des distributions rigoureuses des sous-fluctuations, dans la mesure où $\widetilde{U}$ et la matrice N sont construits globalement sur l'élément. Elles introduisent néanmoins un défaut potentiel dont l'impact serait à étudier : s'il est vrai que les contributions sont distribuées de manière *upwind* (les schémas sont bien $\mathcal{MU}$), elles contiennent en revanche de l'information provenant des noeuds situés en aval de chaque onde. Ce phénomène concerne visiblement tous les degrés de liberté autres que les sommets.

Pour finir, on peut discuter de l'intérêt de ces généralisations. Quel que soit l'élément, la distribution LDA conserve sa propriété $\mathcal{LP}$ et reste donc utilisable pour tout problème suffisamment régulier. En ce qui concerne le schéma N, les propriétés qu'il peut démontrer P_1 sont formellement perdues sur tout autre élément. Son intérêt devient alors sans doute incertain, même si nous ne l'avons jamais testé conjointement à une interpolation de degré supérieur à 1. Pour les problèmes comportant de forts gradients voire des chocs, il faudra peut-être lui préférer le schéma de Lax-Friedrichs malgré la très forte diffusion numérique de ce dernier. Sur le plan théorique, tous les schémas peuvent donc être étendus aux Éléments Finis de tout ordre moyennant une quantité d'efforts dépendant du schéma : raisonnable pour les schémas centrés (selon l'approximation de l'opérateur de stabilisation s'il est employé), davantage pour les schémas $\mathcal{MU}$.

Limitation

Le procédé de limitation en P_1 était réinterprétable dans un cadre géométrique simple. Ce principe est certainement généralisable aux tétraèdres P_1 et on doit, par exemple, pouvoir trouver aisément une projection sur la sphère circonscrite. Pour tout autre élément et/ou des ordres plus élevés, une telle réinterprétation géométrique simple ne nous est pas connue. Même si cela pourrait être un sujet de recherches, le bénéfice éventuel serait peut-être maigre au regard des efforts fournis puisque la limitation *minmod* s'étant à tous les éléments envisageables, et puisque notre expérience en P_1 est que les différentes limitations ne changent à vrai dire pas grand-chose aux résultats. On se contente donc généralement d'étendre le principe de la limitation *minmod* (voir section 3.4.1 : la relation scalaire (3.9) en remplaçant seulement T par E pour généraliser, et le principe de l'extension aux systèmes).

3.6 Prise en compte des conditions limites

3.6.1 Imposition au sens faible

Pour les problèmes de MHD idéale, nous n'avons que deux types de conditions limites : le bord est soit une paroi glissante d'un conducteur parfait (aucun phénomène dissipatif ne rentre en jeu dans les modèles idéaux), soit une ouverture permettant de communiquer avec un domaine extérieur dont on connaît l'état.

Pour imposer ces conditions limites, on procède de la même manière que dans les méthodes Éléments Finis, puisque nos schémas sont interprétables dans un formalisme Petrov-Galerkin. Autrement dit, on les impose faiblement. Avec une méthode d'Éléments Finis de Lagrange, la formulation variationnelle se décompose en chaque degré de liberté $M_i \in \Omega_h$ de la manière suivante :

$$\int_{\Omega_h} \varphi_i \left(\frac{\partial U_h}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h) \right) dV = \int_{\Omega_h} \varphi_i \frac{\partial U_h}{\partial t} dV - \int_{\Omega_h} \overrightarrow{\nabla} \varphi_i \cdot \overrightarrow{F}(U_h) dV + \int_{\partial \Omega_h} \varphi_i \overrightarrow{F}(U_h) \cdot \overrightarrow{n} d\partial\Omega_h = 0 \tag{3.6.1}$$

Nous plaçons la formulation directement sur le maillage Ω_h, et non sur Ω, car nous n'avons pas utilisé de méthode pour améliorer la représentation des bords, que ce soit par une description isoparamétrique ou des éléments courbes. Nous ne nous soucions donc pas, dans nos travaux, de l'erreur commise en approchant la géométrie. Supposons maintenant que nous travaillions à résoudre la formulation du second membre. Cela signifie que pour des noeuds suffisamment éloignés du bord, grâce à la compacité des fonctions test, nous approcherions des expressions du type :

$$\int_{\Omega_h} \varphi_i \frac{\partial U_h}{\partial t} dV - \int_{\Omega_h} \overrightarrow{\nabla} \varphi_i \cdot \overrightarrow{F}(U_h) dV = 0 \tag{3.6.2}$$

Dans ce cadre, imposer les conditions limites que nous considérerons par la suite de façon faible consisterait à remplacer les flux apparaissant dans le terme de bord par des flux $\overrightarrow{H}$ dont le rôle est de contraindre la solution à se comporter comme on le souhaite. On écrirait donc en tout degré de liberté i :

$$\int_{\Omega_h} \varphi_i \frac{\partial U_h}{\partial t} dV - \int_{\Omega_h} \overrightarrow{\nabla} \varphi_i \cdot \overrightarrow{F} dV + \int_{\partial \Omega_h} \varphi_i \overrightarrow{H} \cdot \overrightarrow{n} d\partial\Omega_h = 0 \tag{3.6.3a}$$

$$\iff \int_{\Omega_h} \varphi_i \frac{\partial U_h}{\partial t} dV - \int_{\Omega_h} \overrightarrow{\nabla} \varphi_i \cdot \overrightarrow{F} dV + \int_{\partial \Omega_h} \varphi_i \overrightarrow{F} \cdot \overrightarrow{n} d\partial\Omega_h + \int_{\partial \Omega_h} \varphi_i \left(\overrightarrow{H} - \overrightarrow{F} \right) \cdot \overrightarrow{n} d\partial\Omega_h = 0 \tag{3.6.3b}$$

Si on décidait en revanche de ne pas résoudre la formulation (3.6.2) mais la formulation originale non intégrée par parties, la façon d'imposer les conditions limites resterait la même, à savoir qu'on écrirait :

$$\int_{\Omega_h} \varphi_i \left(\frac{\partial U_h}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h) \right) dV + \int_{\partial \Omega_h} \varphi_i \left(\overrightarrow{H}(U_h) - \overrightarrow{F}(U_h) \right) \cdot \overrightarrow{n} d\partial\Omega_h = 0 \tag{3.6.4}$$

Les schémas $\mathcal{RD}$ étant interprétables dans une formulation de type Petrov-Galerkin, l'imposition des conditions limites au sens faible se fait exactement de la même façon. Autrement dit, en suivant les mêmes étapes de calcul, on aboutit à :

$$\int_{\Omega_h} (\varphi_i I_p + \gamma_i) \left(\frac{\partial U_h}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h) \right) dV + \int_{\partial \Omega_h} (\varphi_i + \gamma_i) \left(\overrightarrow{H}(U_h) - \overrightarrow{F}(U_h) \right) \cdot \overrightarrow{n} d\partial\Omega_h = 0 \tag{3.6.5}$$

où I_p est la matrice identité de taille p, le nombre de variables. On rappelle que γ_i est la matrice de décentrement associée au schéma, introduite en début de chapitre dans le cas scalaire, et dont l'écriture est $\gamma_i = \beta_i - \varphi_i I_p$. Cette formulation n'est pas toujours utilisable en pratique car il arrive que nous ne connaissions pas explicitement les matrices β_i (pour des systèmes d'équations, on ne les connaît qu'avec le schéma LDA et, selon l'approximation du terme de stabilisation, le schéma SU). Dans ce cas, puisque γ_i n'est a priori jamais une fonction bulle, la seule option qu'il nous reste est de ne pas en tenir compte, ce qui revient à un schéma de Galerkin sur le bord :

$$\int_{\Omega_h} (\varphi_i I_p + \gamma_i) \left(\frac{\partial U_h}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h) \right) dV + \int_{\partial \Omega_h} \varphi_i \left(\overrightarrow{H}(U_h) - \overrightarrow{F}(U_h) \right) \cdot \overrightarrow{n} d\partial\Omega_h = 0 \tag{3.6.6}$$

L'approximation reste consistante, mais on ne compense pas exactement la totalité des contributions du schéma dans les éléments de bord. Par conséquent, en chaque degré de liberté $M_i \in \Omega_h$, un schéma $\mathcal{RD}$

s'écrira si possible :

$$\sum_{T\subset\mathcal{T}_i}\phi_i^T=\sum_{T\subset\mathcal{T}_i}\beta_i^T\phi^T+\sum_{\Gamma\subset(\partial T\cap\partial\Omega_h)}\beta_i\int_\Gamma\left(\overrightarrow{H}(U_h)-\overrightarrow{F}(U_h)\right)\cdot\overrightarrow{n}\,d\Gamma=\sum_{T\subset\mathcal{T}_i}\beta_i^T\phi^T+\phi_i^\Gamma=0 \qquad (3.6.7)$$

ou bien, par défaut :

$$\sum_{T\subset\mathcal{T}_i}\phi_i^T=\sum_{T\subset\mathcal{T}_i}\beta_i^T\phi^T+\sum_{\Gamma\subset(\partial T\cap\partial\Omega_h)}\int_\Gamma\varphi_i\left(\overrightarrow{H}(U_h)-\overrightarrow{F}(U_h)\right)\cdot\overrightarrow{n}\,d\Gamma=\sum_{T\subset\mathcal{T}_i}\beta_i^T\phi^T+\phi_i^\Gamma=0 \qquad (3.6.8)$$

En pratique, on utilise toujours cette dernière expression pour éviter de stocker les matrices β_i. Dans tous les cas, on voit d'après les écritures ci-dessus qu'il faut approcher l'intégrale de bord avec au moins la même précision que celle du schéma employé. Pour cela, on peut apporter les mêmes remarques que dans la section précédente, c'est-à-dire qu'il faut en général recourir à une formule de quadrature ou à une interpolation d'ordre élevé des flux de bord.

3.6.2 Paroi glissante parfaitement conductrice

La paroi glissante est une condition clasique en mécanique des fluides, qui stipule que les composantes tangentielles du champ de vitesses ne sont pas affectées par la paroi, à l'inverse de la composante normale qui s'annule (aucun flux de matière ne peux traverser la paroi). La qualité de conducteur parfait du matériau de bord a exactement les mêmes effets sur le champ magnétique. Mathématiquement, si $\overrightarrow{n}$ désigne localement la normale unitaire sortante du domaine, cette condition se traduit donc par :

$$\forall\overrightarrow{x}\in\partial\Omega_h,\ \overrightarrow{u}(\overrightarrow{x})\cdot\overrightarrow{n}(\overrightarrow{x})=\overrightarrow{B}(\overrightarrow{x})\cdot\overrightarrow{n}(\overrightarrow{x})=0$$

Il suffit donc d'injecter ces égalités dans les flux physiques pour savoir à quoi doit correspondre physiquement le flux réel s'appliquant sur la paroi. Si on note F_n le flux projeté selon une normale sortante sur un bord, on obtient alors directement :

$$H_n=\begin{pmatrix}0\\\left(p+\dfrac{\overrightarrow{B}^2}{2}\right)\overrightarrow{n}\\0\\\psi\overrightarrow{n}\\0\end{pmatrix}$$

Autrement dit, le flux total qui apparait sous l'intégrale de bord dans (3.6.7) ou (3.6.8) est :

$$(H_n-F_n)=\begin{pmatrix}-\rho u_n\\-\rho u_n\overrightarrow{u}+B_n\overrightarrow{B}\\-\left(E+p+\dfrac{\overrightarrow{B}^2}{2}\right)u_n+\left(\overrightarrow{u}\cdot\overrightarrow{B}\right)B_n\\-u_n\overrightarrow{B}+B_n\overrightarrow{u}\\-{c_h}^2B_n\end{pmatrix}$$

Il ne reste ensuite plus qu'à approcher (3.6.7) ou (3.6.8) avec la précision souhaitée.

3.6.3 Entrées/sorties avec état imposé à l'infini

Les conditions d'entrée ou de sortie sont un peu plus délicates à imposer. La représentation mathématique que nous en faisons est celle d'une ouverture délimitant deux régions : une région interne,

appartenant au domaine de calcul, où la solution est calculée à chaque pas de temps, et une région externe où nous supposons que tout est constant. Sur chaque bord, cela revient donc à définir un problème de Riemann multidimensionnel. Comme nous l'avons déjà évoqué au chapitre 2, nous ne savons malheureusement pas résoudre le problème de Riemann au-delà du cas 1D. Supposons néanmoins que nous soyons en mesure de le faire, et notons $\mathcal{R}_n$ le flux de Riemann exact à travers un bord Γ contenu dans un triangle T et de normale $\overrightarrow{n}$. On écrira alors par exemple, dans le cas de (3.6.8) :

$$\phi_i^\Gamma = \beta_i^T \int_\Gamma \left(\mathcal{R}_n(U_h, U_\infty) - \overrightarrow{F}(U_h) \cdot \overrightarrow{n} \right) dV$$

Cette expression devra être approchée soit par une formule de quadrature, soit en développant les flux suivant des fonctions de base d'ordre élevé ou très élevé. Ainsi, le calul ne fera intervenir que certains degrés de liberté appartenant à Γ. Par exemple, en utilisant une formule de quadrature, on aura une expression du type :

$$\phi_i^\Gamma = \sum_{j \in \Gamma} \omega_j \left(\mathcal{R}_n(U_j, U_\infty) - F_n(U_j) \right) \tag{3.6.9}$$

où les ω_j sont les poids de la formule. Notre façon de procéder est alors de supposer que la solution du problème de Riemann multidimensionnel est assimilable localement à celle du problème Riemann unidimensionnel local, entre les états U_i et U_∞ et le long d'une normale $\overrightarrow{n}(\overrightarrow{x})$ évaluée localement. À partir de là, on connait les techniques de résolution et on peut choisir parmi les différents solveurs de Riemann approchés existant dans la littérature (voir par exemple [92]).

En mécanique des fluides, on utilise généralement (à l'instar de [1]) un flux numérique de Steger et Warming, qui utilise le fait que les équations d'Euler sont homogènes de degré 1 en U (voir [41]). Cette stratégie offre une façon simple de trier les caractéristiques entrantes et sortantes. Comme nous l'avons montré dans la section 2.3.4 du second chapitre, les équations de la MHD ne vérifient pas une telle propriété. On doit donc faire appel à un autre flux numérique. Notre choix s'est porté sur un schéma de Roe. Le flux numérique associé est issu d'une généralisation aux systèmes d'équations du schéma de Roe scalaire que nous avons vu en début de chapitre (section 3.1.2), et donc de (3.1.12). Projeté selon une direction $\overrightarrow{n}$ arbitraire, il s'écrit :

$$H_n(U_1, U_2) = \frac{1}{2} \left(F_n(U_1) + F_n(U_2) - \left| \overline{A_n} \right| (U_2 - U_1) \right) \tag{3.6.10}$$

où $\overline{A_n}(U_1, U_2)$ est une extension de la linéarisation conservative scalaire (3.1.13) qui doit vérifier trois propriétés :

1. $F_n(U_2) - F_n(U_1) = \overline{A_n}(U_2 - U_1)$ qu'on note $\Delta F_n = \overline{A_n} \Delta U$
2. $\overline{A_n}(U_1, U_1) = A_n(U_1) = \dfrac{\partial F_n}{\partial U}(U_1)$
3. $\overline{A_n}$ est diagonalisable et possède un jeu complet de vecteurs propres

La notation $|.|$ appliquée à une matrice diagonalisable renvoie au *splitting* de ladite matrice. Si on définit, selon le modèle (3.3.2) :

$$\left(\overline{A_n}\right)^+ = R_A \Lambda_A^+ L_A \text{ et } \left(\overline{A_n}\right)^- = R_A \Lambda_A^- L_A$$

de telle sorte que $\overline{A_n} = \left(\overline{A_n}\right)^+ + \left(\overline{A_n}\right)^-$, alors on a :

$$\left|\overline{A_n}\right| = \left(\overline{A_n}\right)^+ - \left(\overline{A_n}\right)^- \tag{3.6.11}$$

On appelle généralement $|\overline{A_n}|$ la matrice de Roe. Brio et Wu [19] l'ont déterminée dans notre cas d'une équation d'état de gaz parfait, pour le cas particulier $\gamma = 2$. Leur résultat a été généralisé par Cargo et Gallice [24] à toute valeur de γ. C'est de ces derniers travaux que nous sommes partis. Le détail de $\overline{A_n}$ et de son système propre, qui permet de construire la matrice de Roe, sont donnés dans l'annexe C. Une fois cette matrice connue, il ne reste plus qu'à l'évaluer en chaque point requis par (3.6.9) puisque le principe est de remplacer les flux de Riemann exacts $\mathcal{R}_n(U_j, U_\infty)$ par les flux de Roe $H_n(U_j, U_\infty)$ (3.6.10).

3.6.4 Conditions de Dirichlet - Imposition forte

Les conditions de type Dirichlet sont les plus simples qui soient. Elles consistent à imposer directement la valeur de la solution U aux noeuds situés sur la frontière du domaine Ω_h. C'est donc ce qu'on fait généralement, ignorant par là même purement et simplement les valeurs calculées par le schéma. On parle alors d'imposition forte, puisqu'on peut assigner la valeur de la solution sans passer par la formulation faible du problème. Si on initialise la solution aux bords concernés avec les valeurs que requiert la condition de Dirichlet, alors il suffit par la suite, à chaque pas de temps, d'annuler les contributions envoyées à ces degrés de liberté. Le problème de cette méthode est que le fait d'annuler brutalement certaines contributions peut faire perdre de l'information, ce qui pose en ce cas des problèmes de conservation. Ceci arrive dès lors que le schéma ignore l'existence de la condition limite, ce qui est (pour nous tout du moins) toujours le cas. Généralement, ce phénomène est numériquement négligeable (i.e. d'un point de vue quantitatif) et c'est pourquoi on met en oeuvre les conditions de Dirichlet exactement de la manière que nous venons de décrire. Si, en revanche, on souhaite être rigoureux, il est toujours possible de le faire moyennant des efforts de calcul supplémentaires. Au niveau théorique, la correction de ce défaut est très simple : dans chaque élément dont au moins un degré de liberté est situé sur une frontière de Dirichlet, la distribution du résidu (de la fluctuation dans le cas stationnaire) ne doit se faire qu'entre les autres degrés de liberté, qui appartiennent à l'intérieur du domaine.

Dans un problème physique, réel, on peut se demander quand il est utile de faire appel à ce type de condition limite. Il y a en fait un cas dans lequel les conditions de Dirichlet sont très souvent utilisées en mécanique des fluides : les conditions d'entrée supersoniques. En effet, dans ce cas, on peut prédire avant calculs que toutes les caractéristiques issues du problème de Riemann entrent dans le domaine. La même chose se produit en MHD lorsque la vitesse d'écoulement est rentrante et supérieure aux ondes rapides : l'entrée *superrapide* (voir la classification établie par la figure 2.5 du chapitre 2). Sachant cela, on connaît de façon triviale la solution du problème de Riemann, puisqu'il s'agit de U_∞. En d'autres termes, on a dans (3.6.9) :

$$\mathcal{R}_n(U_j, U_\infty) = \mathcal{R}_n(U_\infty, U_\infty) = F_n(U_\infty)$$

et donc $\phi_i^\Gamma = 0$ dans (3.6.7) et (3.6.8), pour tout degré de liberté sur un bord concerné. Cela signifie qu'il n'y a rien à faire pour de tels degrés de liberté, sinon s'assurer qu'ils gardent la valeur U_∞. Il s'agit donc d'une condition de Dirichlet stipulant que sur le bord concerné, $U = U_\infty$. À noter que dans le cas d'une sortie *superrapide*, comme toutes les caractéristiques sortent du domaine, c'est l'extérieur qui ne doit rien imposer aux degrés de liberté de bord : seul le schéma, qui propage les informations depuis l'intérieur, entre en jeu. Il suffit donc de ne rien faire.

Enfin, il faut citer un autre intérêt à ce type de condition : forcer le multiplicateur de Lagrange ψ à être nul sur le bord quand on emploie la correction de la divergence. La procédure est identique.

Chapitre 4

Résolution temporelle et phénomènes dissipatifs

Sommaire

Nous avons vu dans le chapitre précédent la discrétisation spatiale par les schémas $\mathcal{RD}$ des problèmes stationnaires et homogènes (sans termes de sources). On s'intéresse maintenant à des extensions pratiques de ces schémas à des problèmes plus complexes : instationnaires, non homogènes, augmentés de phénomènes diffusifs traduits par des termes en dérivées spatiales secondes. Notre choix d'avoir découpé ces questions ainsi, à l'aide de ces deux chapitres, est le reflet de deux approches que nous avons adoptées, assez communément répandues par ailleurs :

1. Nous avons systématiquement recours à une semi-discrétisation en temps, indépendamment de ce qui est fait en espace. L'interpolation ne sera donc faite qu'en espace. Ceci ne change pas le fait que la dérivée en temps doive être discrétisée et distribuée en espace, selon une méthode qui peut dépendre de la discrétisation de la divergence des flux et qui peut s'appliquer à tout terme de source.

2. De la même façon, notre distribution des termes diffusifs ne conditionne pas celle des termes d'advection. Au mieux, elle pourrait s'appuyer sur cette dernière.

En partant des équations corrigées du chapitre 2, et en utilisant les concepts développés au chapitre 3, ce chapitre a donc deux objectifs. D'une part, présenter la méthode discrétisation en temps en commençant par préciser comment les termes supplémentaires (ne pouvant pas être formulés sous la forme d'une divergence) sont discrétisés en espace. L'étape de la discrétisation temporelle elle-même se divise ensuite en deux parties : l'implicite et l'explicite. On pourra noter au passage la manière dont la correction de la divergence s'intègre à chaque méthode. Les questions de stabilité numérique, telles qu'abordées au chapitre précédent (stabilités énergétique et entropique notamment), ne seront pas étendues dans ce chapitre. D'autre part, le second objectif est la présentation de la méthode de discrétisation de la partie diffusive des équations (résistivité, conduction de chaleur, viscosité et diffusion de matière) en expliquant les raisons de nos choix. Ceci sera l'occasion d'évoquer les problèmes numériques posés par le couplage de cette partie avec les équations de la MHD idéale.

Dans tout ce chapitre, les questions de stabilité numérique, telles qu'abordées au chapitre précédent (stabilités énergétique et entropique notamment), ne seront pas étendues. Il reste de fait beaucoup de zones d'ombre à ce sujet, et nous n'avons pas encore poussé les investigations dans cette direction suffisamment loin. De plus, mais cette fois surtout pour alléger le texte, nous omettrons régulièrement d'écrire les conditions limites, même dans la formulation faible (Petrov-Galerkin) des équations. Ceci ne signifie pas que leur intégration dans les méthodes en temps soit un problème, car elle se fait en fait très naturellement.

Remarques préalables sur la prise en compte du *divergence cleaning*

Si le fait d'ignorer les erreurs commises sur la divergence peut donner des résultats satisfaisants, lorsque cela n'est plus le cas, on peut aboutir à des situations non physiques, si bien que la simulation diverge. On renvoie à la section 2.2.1 pour plus de détails. Ceci dépend en premier lieu de la difficulté du problème abordé, en termes de géométrie et de raideur. Le schéma employé est naturellement le second facteur déterminant le comportement de cette erreur. En particulier, une limitation composante par compossante, en plus de ne pas respecter le couplage naturel entre les équations, détruit le caractère solénoïdal du champ magnétique. A contrario la limitation avec projection sur une base de vecteurs propres n'a pas un impact critique sur l'erreur de divergence. Nous avons aussi remarqué que le terme SUPG que nous rajoutons pour stabiliser le schéma dans les zones régulières a tendance à détériorer, assez légèrement, le caractère solénoïdal du champ magnétique.

En programmant nos algorithmes, nous avons choisi de laisser l'utilisateur libre le choix d'employer ou non la correction de la divergence, en supposant que celui-ci soit averti. Mais par défaut dans ce chapitre, nous supposerons toujours que nous résolvons les équations en faisant appel à la correction hyperbolique (ou mixte) de la divergence. Les algorithmes sans correction en découlent de façon triviale. Lorsque nous aurons besoin de déroger à cette règle et de faire apparaître un traitement particulier pour le champ magnétique, nous essaierons de le faire sans ambiguïté.

4.1 Systèmes instationnaires non homogènes

4.1.1 Mise en place du problème continu

Lorsqu'on étudie des lois de conservation, les problèmes dits non homogènes sont ceux qui font intervenir des termes qui ne sont pas inclus dans la divergence (i.e. qui ne font pas partie des flux), et qui ne sont pas la dérivée en temps de U. De tels termes modélisent typiquement des sources et sont donc

appelés termes de source. Des exemples classiques sont les forces de gravité dans l'équation de conservation de la quantité de mouvement et leur travail dans celle de l'énergie, ou encore une source ou un puit volumique de chaleur dans l'équation de l'énergie, modélisant un apport ou une cession d'énergie thermique par des phénomènes non pris en compte dans les équations. Le modèle de la MHD néglige traditionnellement la gravité. Quant aux sources ou puits de chaleur envisageables en MHD, on pourrait citer par exemple la perte ou l'absorption d'énergie par rayonnement. Là encore, ce sont des phénomènes qui sont généralement négligés pour les applications que nous visons. En réalité, le seul terme de source qui puisse apparaître dans nos équations est le terme parabolique qui fait partie de la correction mixte (2.2.5), i.e. de la dernière équation de notre système, celle contrôlant le multiplicateur de Lagrange ψ.

Quoi qu'il en soit, nous abordons ici le problème de façon générale en notant le problème ainsi :

$$r(U) = \frac{\partial U}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U) + g(U) = 0 \tag{4.1.1}$$

D'après ce que nous avons dit, les termes de source se réduiront pour nous à :

$$g(U) = \begin{pmatrix} 0 \\ \overrightarrow{0} \\ 0 \\ \overrightarrow{0} \\ -\frac{{c_h}^2}{{c_p}^2}\psi \end{pmatrix} \tag{4.1.2}$$

avec c_h et c_p considérées comme des constantes du problème. Pour l'analyse mathématique, l'expression "termes de source" est couramment utilisée qu'il y ait des sources, des puits ou des deux à la fois. Dans cette section, avant d'aborder la question de la discrétisation des équations en temps, nous voulons présenter celle, en espace, des termes autres que la divergence des flux (on considère cette dernière traitée à l'aide des schémas présentés dans le chapitre précédent). Cela concerne donc le terme d'évolution $\partial_t U$ et les termes de source regroupés dans $g(U)$.

4.1.2 Une discrétisation possible dans le formalisme $\mathcal{RD}$

La première étape est la même que dans le chapitre précédent, et est toujours la même quel que soit le problème à résoudre avec nos méthodes : définir une certaine représentation de la solution, ici une interpolation polynomiale par des polynômes de Lagrange qui s'appuient sur les degrés de liberté (nodaux) du maillage. Cela signifie simplement qu'on se donne une forme de la solution qu'on recherche, forme qu'on sait être une bonne approximation (dans le sens où on la sait consistante et où on sait étudier l'erreur). Le problème à résoudre est alors :

$$r(U_h) = \frac{\partial U_h}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h) + g(U_h) = 0 \tag{4.1.3}$$

Notons T le temps final que la simulation doit atteindre, et revenons sur nos notations pour ce chapitre :

- $U_h = U_h(\overrightarrow{x}, t)$ est la solution interpolée en espace composante par composante par les mêmes fonctions de forme (polynômes de Lagrange s'appuyant sur les valeurs nodales de chaque composante). C'est donc une fonction continue de l'espace et du temps, et son taux de variation en temps est donc une dérivée partielle : $\partial_t U_h$.
- $\forall M_i \in \Omega_h$, $U_i = U_i(t)$ est une localisation de chaque composante la solution interpolée. C'est donc une fonction du temps seulement, si bien que son taux de variation en temps est une dérivée classique $d_t U_i$.
- $\forall n \in \mathbb{N}, t^n \leq T$, $U_h^n = U_h(\overrightarrow{x}, t^n)$ est une évaluation en temps de l'interpolation cette fois-ci. Chacune des p composantes est une fonction continue en espace.

– $\forall M_i \in \Omega_h, \forall n \in \mathbb{N}, t \leq T$, $U_i^n = U_h(\overrightarrow{x}_i, t^n)$ est une double évaluation, en temps et en espace : c'est un vecteur constant de taille p.

Ces précisions parfois évidentes sont surtout faites pour justifier les écritures des dérivées en temps :

$$\forall E \subset \Omega_h,\ \left.\frac{\partial U_h}{\partial t}\right|_E (\overrightarrow{x}, t) = \sum_{M_i \in E} \varphi_i|_E(\overrightarrow{x}) \frac{dU_i}{dt}(t)$$

Le principe des schémas $\mathcal{RD}$ reste ensuite le même, à savoir intégrer les équations sur un élément et distribuer le résidu ainsi obtenu aux degrés de liberté. Comme l'intégrale d'une somme est la somme des intégrales, on peut légitimement opérer sur chaque terme des équations (terme d'évolution, divergence des flux et sources) de façon indépendante. Tout en restant très général, la distribution de chaque terme peut se faire via une combinaison (potentiellement non linéaire) d'évaluations aux degrés de liberté M_i. Si on tient compte du fait supplémentaire qu'on souhaite conserver l'écriture ϕ_i pour les schémas stationnaires distribuant le bilan de flux sur l'élément, on obtient l'écriture générique suivante :

$$\forall E \subset \Omega_h, \forall M_i \in E,\ \Phi_i^E(U_h) = \sum_{j \in E} m_{ij}^1(U_h) \frac{dU_i}{dt} + \phi_i^E(U_h) + \sum_{s=1}^{N_g} m_{is}^2 g(\mathcal{L}_s(U_h))$$

où les N_g quantités $\mathcal{L}_s(U_h)$ sont d'autres combinaisons non linéaires des valeurs nodales de U_h, pour les cas où g serait non linéaire. La matrice m^1 devant le terme d'évolution est traditionnellement appelée la matrice de masse. Même si à ce stade nous ne requérons rien, pas même la conservation, cette écriture n'est pas la plus générale. Les distributions de chaque terme pourraient a priori comporter d'autres termes quelconques, qui n'auraient rien à voir avec les autres, voire sans aucun sens précis (même si évidemment, ça ne semblerait pas judicieux).

Ce n'est qu'en imposant quelques contraintes qu'on peut choisir une formulation qui reste raisonnablement générale. La première d'entre elles, celle qui définit les schémas $\mathcal{RD}$, est la conservation. Pour l'utiliser, il faut commencer par définir le résidu :

$$\forall E \subset \Omega_h,\ \Phi^E = \int_E \frac{\partial U_h}{\partial t} dV + \int_E \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h) + \int_E g(U_h) dV$$

La question délicate la plus délicate est celle de la discrétisation des termes de source. Pour calculer l'intégrale de g de façon approchée, on peut user soit d'une interpolation d'ordre suffisant $g_h(U_h)$, soit d'une formule de quadrature d'un ordre également suffisant, comme pour le calcul de la fluctuation quelle que soit son expression (intégrale de contour ou de volume). Nous avons tendance à privilégier les formules de quadrature, qui peuvent (relativement) facilement être complexifiées pour exprimer une meilleure représentation des flux. Nous calculons donc G^E à l'aide de N_q points de quadrature construits simplement en évaluant U_h. On utilise donc en pratique :

$$\forall E \subset \Omega_h,\ \Phi^E = \sum_{M_i \in E} \frac{dU_i}{dt} \int_E \varphi_i dV + \phi^E + \sum_{q=1}^{N_q} \omega_q g\left(U_h(\overrightarrow{x}_q)\right) \tag{4.1.4}$$

La propriété de conservation impose donc que soient vérifiées les conditions suivantes :

$$\forall E \subset \Omega_h,\ \sum_{M_i \in E} \sum_{M_j \in E} m_{ij}^1 \frac{dU_j}{dt} = \sum_{M_j \in E} \int_E \varphi_j dV \frac{dU_j}{dt} \text{ et } \sum_{M_i \in E} \sum_{s=1}^{N_g} m_{is}^2 g(\mathcal{L}_s(U_h)) = \sum_{q=1}^{N_q} \omega_q g\left(U_h(\overrightarrow{x}_q)\right)$$

Parmi les solutions possibles pour les obtenir, l'une des plus simples est de prendre :

$$\forall E \subset \Omega_h, N_g = N_q \text{ et } \forall q \leq N_q,\ \sum_{M_i \in E} m_{iq}^2 = \omega_q \text{ et } \mathcal{L}_q(U_h) = U_h(\overrightarrow{x}_q)$$

ce qui nous donne des résidus partiels de la forme :

$$\forall E \subset \Omega_h, \forall M_i \in E, \Phi_i^E = \sum_{M_j \in E} m_{ij}^1 \frac{dU_j}{dt} + \beta_i^E \phi^E + \sum_{q=1}^{N_q} m_{iq}^2 g(U_h(\overrightarrow{x}_q)) \tag{4.1.5}$$

tout en choisissant de requérir sur les matrices m^1 et m^2, pour la conservation :

$$\forall M_j \in E, \sum_{M_i \in E} m_{ij}^1 = \int_E \varphi_j dV \text{ et } \forall q \leq N_q, \sum_{M_i \in E} m_{iq}^2 = \omega_q \tag{4.1.6}$$

Dans le cas particulier où les sources sont linéaires, ce qui est le cas pour notre modèle puisque g est donné par (4.1.2), la discrétisation rejoint celle du terme d'évolution (l'opérateur ∂_t étant lui aussi linéaire). En effet, le terme dans (4.1.3) peut se réécrire :

$$g(U_h) = A_g U_h$$

avec A_g une matrice $p \times p$ constante. Son intégrale sur l'élément, qui se retrouve dans (4.1.4), est alors :

$$\int_E g(U_h) dV = A_g \sum_{M_i \in E} U_i \int_E \varphi_i dV = \sum_{M_i \in E} g(U_i) \int_E \varphi_i dV$$

et une façon simple, limite du cas non linéaire, de formuler la partie correspondante dans le schéma est de prendre $N_g = N_q = N_s$ (N_s étant pour rappel le nombre de degrés de liberté dans l'élément) et :

$$\forall M_i \in E, \Phi_i^E = \sum_{M_j \in E} m_{ij}^1 \frac{dU_j}{dt} + \beta_i^E \phi^E + \sum_{M_j \in E} m_{ij}^2 g(U_j) \tag{4.1.7}$$

avec un choix de contraintes pour la conservation qui devient :

$$\forall M_j \in E, \sum_{M_i \in E} m_{ij}^1 = \sum_{M_i \in E} m_{ij}^2 = \int_E \varphi_j dV \tag{4.1.8}$$

On voit que les deux matrices se ressemblent potentiellement beaucoup, si bien en fait qu'on peut être tenté de choisir $m^2 = m^1$. Même si cela n'est pas le cas, cette similarité peut nous encourager à désigner m^2 comme une seconde matrice de masse.

Dans tout ce qui suivra, on considèrera que notre schéma spatial est du type (4.1.7)-(4.1.8). Outre la conservation, une autre propriété qui pourrait contraindre les matrices de masse est la consistance. Au-delà de celle-ci, on préfère aborder directement les conditions pour obtenir l'ordre élevé. C'est ce que nous allons voir maintenant.

4.1.3 Principe de construction des schémas d'ordre élevé

Cette partie nous sera également utile dans les sections concernant la discrétisation en temps, car pour ne pas avoir à la reprendre on va inclure une hypothèse cruciale sur le schéma en temps, et donc travailler sur une discrétisation complète sans vraiment la spécifier. Pour cela, on va procéder de la même manière que dans la section 3.2.3. En particulier, on prend une définition de l'erreur de troncature proche de (3.2.13), qui s'évalue au temps t^{n+1} que nous cherchons à calculer (on se considère au temps t^n, avec la solution U_h^n calculée après n itérations en temps). Pour toute fonction $\psi \in \mathcal{C}_0^1(\Omega_h)$, on étudiera l'erreur de troncature instantanée définie comme étant :

$$\mathcal{E}_\psi^{n+1} = \sum_{M_i \in \Omega_h} \psi_i \sum_{E \subset \mathcal{T}_i} \Phi_i^{n+1}$$

avec Φ_i^{n+1} donné par la relation (4.1.7) évaluée un certain nombre de fois en un certain nombre de dates (t^n, t^{n+1}, voire des solutions intermédiaires) qui dépendent du schéma en temps mis en oeuvre. On ne détaillera pas cet aspect mais ici, mais on peut d'ores et déjà dire que l'analyse qui suit reste valable quelle que soit cette construction temporelle. Pour donner un exemple, si le schéma est purement implicite, il suffit dévaluer (4.1.7) au temps $n+1$, ce qui donne :

$$\mathcal{E}_\psi^{n+1} = \sum_{M_i \in \Omega_h} \psi_i \sum_{E \subset \mathcal{T}_i} \left(\sum_{M_j \in E} m_{ij}^1 \left(\frac{dU_j}{dt} \right)^{n+1} + \beta_i^{n+1} \phi^{n+1} + \sum_{M_j \in E} m_{ij}^2 g(U_j^{n+1}) \right) \tag{4.1.9}$$

Cet exemple pratique, que nous reverrons dans la section sur les méthodes implicites, va nous servir de base pour exposer les conditions d'obtention de l'ordre élevé.

Condition sur les matrices de masse

La seule façon que nous connaissons de construire des schémas d'ordre élevé est d'avoir une formulation du type :

$$\Phi_i^{n+1} = \beta_i^{n+1} \Phi^{n+1} = \int_E \left(\varphi_i I_p + \gamma_i^{n+1} \right) r(U_h^{n+1}) dV$$

où les matrices de distribution β_i^{n+1} et donc les fonctions de décentrement γ_i^{n+1} sont bornées (que le schéma les fournisse naturellement ou qu'on ait besoin de recourir à une limitation). La présence de la matrice identité I_p rappelle que les p variables sont interpolées à l'aide des mêmes polynômes de Lagrange φ_i. Le résidu s'exprime ainsi :

$$\Phi^{n+1} = \sum_{M_j \in E} \int_E \varphi_j dV \left(\frac{dU_j}{dt} \right)^{n+1} + \phi^{n+1} + \sum_{M_j \in E} \int_E \varphi_j dV g(U_j^{n+1})$$

ce qui signifie que les signaux Φ_i^{n+1} doivent pouvoir se réécrire :

$$\Phi_i^{n+1} = \sum_{M_j \in E} \beta_i^{n+1} \int_E \varphi_j dV \left(\frac{dU_j}{dt} \right)^{n+1} + \beta_i^{n+1} \phi^{n+1} + \sum_{M_j \in E} \beta_i^{n+1} \int_E \varphi_j dV g(U_j^{n+1})$$

Cela impose de fait aux matrices de masse définies par (4.1.7) d'avoir la forme suivante :

$$\forall M_i, M_j \in E,\ m_{ij}^1 = m_{ij}^2 = \beta_i^{n+1} \int_E \varphi_j dV \tag{4.1.10}$$

Cette contrainte est plus forte que (4.1.8). On peut reformuler ces matrices dans un formalisme Petrov-Galerkin de la manière suivante :

$$\forall M_i, M_j \in E,\ m_{ij}^1 = m_{ij}^2 = \int_E \omega_i^{n+1} \varphi_j dV = \int_E \left(\varphi_i + \gamma_i^{n+1} \right) \varphi_j dV \tag{4.1.11}$$

Il existe une infinité de fonctions tests ω_j^{n+1} et les fonctions de décentrement γ_i^{n+1} vérifiant cette propriété, mais ce sont seulement les intégrales précédentes qui nous intéressent, et celles-ci sont uniques. Par conséquent, les diverses formulations de matrices de masses étudiées dans [76], et dans les références qu'on y trouve, ne sont pas applicables dans le cas général. Dans ces références, leur emploi n'est rendu possible que par les simplifications algébriques particulières qu'apporte l'interpolation linéaire sur des éléments P_1, et dans le cas de flux linéaires en U. L'ordre élevé sera toujours atteint grâce aux matrices de masse (4.1.10), ou de manière équivalente (4.1.11) avec par exemple :

$$\forall M_i \in E,\ \omega_i^{n+1} = \beta_i^{n+1} \text{ et } \gamma_i^{n+1} = \beta_i^{n+1} - \varphi_i$$

Analyse de précision

Au lieu de préciser la discrétisation en temps, ce que nous verrons dans les sections suivantes, nous faisons juste une hypothèse générale qui devra être vérifiée par ceux-ci.

Hypothèse 4.1.1. *Soit $U^{ex}(\overrightarrow{x},t)$ la solution continue exacte du problème (4.1.1). On note alors $U^{ex,n}(\overrightarrow{x}) = U^{ex}(\overrightarrow{x},t^n)$ et $U_h^{ex,n}$ l'interpolée de celle-ci, de telle sorte que $\forall t>0,\ r(U^{ex,n}(.)) = r(U^{ex}(.,t)) = 0$. On suppose que la solution U_h^{n+1}, obtenue après la $(n+1)^{ième}$ itération avec le schéma en temps considéré, engendre une erreur temporelle qui permet d'écrire, uniformément en espace et variable par variable :*

$$r(U_h^{n+1}) - r(U_h^{ex,n}) = O(\Delta t^{k_t}) \tag{4.1.12}$$

où k_t est l'ordre de précision de la méthode en temps.

Revenons maintenant à l'erreur de troncature qui, d'après la condition (4.1.10), s'écrit maintenant :

$$\mathcal{E}_\psi^{n+1} = \sum_{M_i\in\Omega_h}\psi_i\sum_{E\subset\mathcal{T}_i}\beta_i^{n+1}\Phi^{n+1} = \sum_{E\subset\Omega_h}\sum_{M_i\in E}\psi_i\int_E\left(\varphi_i+\gamma_i^{n+1}\right)r(U_h^{n+1})dV$$

Avec l'hypothèse précédente, et en repassant par les mêmes étapes de calcul que dans la section (3.2.3), et comme $r(U^{ex,n})=0$, on obtient immédiatement :

$$\begin{aligned}
\mathcal{E}_\psi^{n+1} &= \sum_{E\subset\Omega_h}\sum_{M_i\in E}\psi_i\int_E\left(\varphi_i+\gamma_i^{n+1}\right)\left(r(U_h^{ex,n})+O(\Delta t^{k_t})\right)dV \\
&= \underbrace{\int_{\Omega_h}\psi_h\left(r(U_h^{ex,n})-r(U^{ex,n})\right)dV}_{I} + \underbrace{\sum_{E\subset\Omega_h}\sum_{M_i\in E}\left(\psi_i-\overline{\psi}\right)\int_E\gamma_i^{n+1}\left(r(U_h^{ex,n})-r(U^{ex,n})\right)dV}_{II} \\
&\quad + \underbrace{\int_{\Omega_h}\psi_h O(\Delta t^{k_t})dV}_{III} + \underbrace{\sum_{E\subset\Omega_h}\sum_{M_i\in E}\left(\psi_i-\overline{\psi}\right)\int_E\gamma_i^{n+1}O(\Delta t^{k_t})dV}_{IV}
\end{aligned}$$

avec $\overline{\psi}$ une moyenne arithmétique des valeurs de ψ sur chaque élément : on utilise le fait que la somme des γ_i^{n+1} sur les noeuds d'un élément est nulle. Nous reprendrons ici le résultat (3.2.14) démontré dans la section 3.2.3, qui nous donne ici :

$$\psi_h r(U_h^{ex,n}) = O\left(\psi(r(U_h^{ex,n})-r(U^{ex,n}))\right)$$

Le lecteur peut aisément vérifier que le changement de définition de l'opérateur r n'empêche en rien cette extension de (3.2.14). Ceci va nous permettre d'estimer le terme I :

$$I = O\left(\|\psi\|_{L^\infty}\int_{\Omega_h}\left(\left(\frac{\partial(U_h-U)}{\partial t}\right)^{ex,n} + \overrightarrow{\nabla}\cdot\left(\overrightarrow{F}(U_h^{ex,n})-\overrightarrow{F}(U^{ex,n})\right) + A_g\left(U_h^{ex,n}-U^{ex,n}\right)\right)dV\right)$$

Que le terme du milieu soit approché, élément par élément, via une formule de quadrature ou une interpolation, celle-ci doit être d'ordre $k+1$ au moins pour avoir (en supposant que l'indice $_h$ désigne, sinon l'interpolation, une représentation consistante avec la quadrature utilisée) :

$$\int_{\Omega_h}\overrightarrow{\nabla}\cdot\left(\overrightarrow{F}_h(U_h^{ex,n})-\overrightarrow{F}(U^{ex,n})\right) = \int_{\partial\Omega_h}\left(\overrightarrow{F}_h(U_h^{ex,n})-\overrightarrow{F}(U^{ex,n})\right)\cdot\overrightarrow{n}\,d\partial\Omega_h = \int_{\partial\Omega_h}O(h^{k+1})d\partial\Omega_h$$

(on renvoit à la section 3.2.3 pour plus d'explications), auquel cas on obtient globalement :

$$I = O(h^{k+1})$$

Le terme II exige également une remarque supplémentaire :

$$\begin{aligned}\forall E \subset \Omega_h, \quad (r(U_h^{ex,n}) - r(U^{ex,n}))\,|_E &= \frac{\partial}{\partial t}\left(U_h^{ex,n} - U^{ex,n}\right) + \overrightarrow{\nabla} \cdot \left(\overrightarrow{F_h}(U_h^{ex,n}) - \overrightarrow{F}(U^{ex,n})\right) \\ &\quad + A_g(U_h^{ex,n} - U^{ex,n}) \\ &= O(h^{k+1}) + O(h^k) + O(h^{k+1})\end{aligned}$$

Pour se convaincre de l'estimation du terme approchant la divergence des flux, on peut se contenter de remarquer (pour ne pas trop alourdir le texte) que dériver une interpolation fait perdre un ordre de précision (les fonctions de forme φ_i ayant pour support des zones $\mathcal{T}_i$, leurs gradients sont en $O(h^{-1})$).

À présent, on utilise les arguments suivantes : la taille de Ω_h est indépendante de h, ψ étant de classe $\mathcal{C}_0^1$ on a $\psi_i - \overline{\psi} = O(h)$ dans chaque élément, et le nombre d'éléments dans Ω_h est estimable en $O(h^{-d})$ tandis que bien sûr la taille de chaque élément est en $O(h^d)$. Les différentes estimations forment alors :

$$\begin{aligned}\mathcal{E}_\psi^{n+1} &= \underbrace{O(h^{k+1})}_{I} + \underbrace{O(h^{-d}) \times O(h) \times O(h^d) \times \left(O(h^{k+1}) + O(h^k) + O(h^{k+1})\right)}_{II} \\ &\quad + \underbrace{O(\Delta t^{k_t})}_{III} + \underbrace{O(h^{-d}) \times O(h) \times O(h^d) \times O(\Delta t^{k_t})}_{IV} \\ &= \underbrace{O(h^{k+1})}_{I} + \underbrace{O(h^{k+1})}_{II} + \underbrace{O(\Delta t^{k_t})}_{III} + \underbrace{O(h\Delta t^{k_t})}_{IV}\end{aligned} \tag{4.1.13}$$

Synthèse

En conclusion, nous avons présenté de façon détaillée une possibilité d'extension semi-discrète des schémas $\mathcal{RD}$ aux problèmes du type (4.1.1). Bien que l'exemple sur lequel nous nous soyons basé soit un schéma implicite particulier, nous espérons que l'analyse a été faite de manière suffisamment compréhensible pour pouvoir être reprise sur des schémas en temps différents. C'est une démarche un peu fastidieuse, mais dont les raisonnements ont été exposés ici. La seule simplification que nous ayons faite est de considérer que les termes de source éventuels sont linéaires, puisque c'est notre cas d'après (4.1.2), ce qui amène à les distribuer de manière analogue au terme instationnaire, c'est-à-dire via des matrices de masses. Ensuite, nous avons formulé deux conditions pour pouvoir atteindre l'ordre élevé dans le cadre qui nous est fixé. La première porte sur la forme des matrices de masse que nous devons employer, donnée par (4.1.10). La seconde précise l'erreur que doit introduire la discrétisation temporelle, ainsi que la forme sous laquelle on doit la quantifier pour qu'elle s'insère dans l'analyse que nous avons menée : c'est l'hypothèse 4.1.1. Si cette forme est adaptée à notre exemple implicite d'étude, elle devrait être reconsidérée à chaque nouveau schéma en temps (ce qui ne veut pas dire qu'elle ne serait pas pertinente, seulement qu'il faut systématiquement s'en assurer). Nous allons maintenant nous consacrer à la présentation de plusieurs schémas en temps, ainsi qu'à leur mise en oeuvre pratique, ce durant les deux prochaines sections de ce chapitre.

4.2 L'approche implicite

Cette partie traite des discrétisations implicites des équations instationnaires. Dans l'optique de simulations complexes comme le sont celles portant sur les instabilités des plasmas de tokamaks, le développement de méthodes implicites permettant des pas de temps très supérieurs à ceux de la limite de

stabilité en explicite est à privilégier. Elles impliquent un coût supplémentaire qui reste acceptable tant qu'elles convergent en un temps raisonnable.

4.2.1 Quelques généralités sur les semi-discrétisations en temps

Revenons au problème continu (4.1.1). La discrétisation spatiale consiste en une interpolation suivie de distributions locales d'intégrales des équations, et s'interprète comme une méthode de Petrov-Galerkin. Pour la semi-discrétisation en temps, son nom indiquant qu'elle ne n'est nullement dépendante du schéma $\mathcal{RD}$ en espace, nous n'avons pas encore fixé de principe de construction. Il n'y a essentiellement que deux méthodes considérées dans nos travaux : l'intégration (I) sur un certain intervalle de temps d'une certaine formulation du problème, ou l'évaluation (E) en une certaine date de cette même formulation. Un second paramètre entre en compte, qui conditionne l'expression du schéma final : l'ordre dans lequel temps et espace sont abordés. Le sens que nous utilisons le plus souvent, noté (1), est de commencer par appliquer la méthode en temps au problème continu (4.1.1). L'autre sera donc noté (2) et consiste à appliquer la méthode en temps au schéma $\mathcal{RD}$ formulé de façon continue en temps. Schématiquement, on obtient donc quatre méthodes possibles :

(I1) À chaque instant t^n, étant donné un intervalle de temps $[t_1; t_2]$ contenant au moins t^{n+1}, y intégrer les équations de manière approchée :

$$\int_{t_1}^{t_2} \left(\frac{\partial U}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U) + g(U) \right) dt = 0$$

puis interpoler en espace les différentes solutions U^k qui apparaissent et distribuer avec un schéma $\mathcal{RD}$ de son choix, ce qui forme des signaux Φ_i^{n+1}.

(I2) À chaque instant t^n, étant donné un intervalle de temps $[t_1; t_2]$ contenant t^{n+1}, y intégrer numériquement le problème déjà semi-discrétisé en espace via un schéma $\mathcal{RD}$ quelconque.

$$\forall M_i \in \Omega_h, \sum_{E \subset \mathcal{T}_i} \int_{t_1}^{t_2} \left(\beta_i^E(t) \int_E r(U_h(.,t)) dV \right) dt = 0$$

(E1) Évaluer le système continu à une date t^e quelconque :

$$\left(\frac{\partial U}{\partial t} \right)^e + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U^e) + g(U^e) = r(U^e) = 0$$

et estimer la dérivée en temps numériquement $(\delta_t U^e)$ en faisant intervenir U^{n+1}, puis interpoler en espace les différentes solutions U^k qui apparaissent et distribuer avec un schéma $\mathcal{RD}$ de son choix, ce qui forme des signaux Φ_i^e.

(E2) Évaluer le système semi-discrétisé en espace à une date t^e quelconque, et estimer la dérivée en temps numériquement $(\delta_t U_h^e)$ en faisant intervenir U_h^{n+1} :

$$\forall M_i \in \Omega_h, \sum_{E \subset \mathcal{T}_i} \beta_i^E(t^e) \int_E \left[\left(\frac{\delta U_h}{\delta t} \right)^e + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h^e) + g(U_h^e) \right] dV = 0$$

On remarque rapidement que les méthodes (E1) et (E2) sont souvent équivalentes, car avec la méthode (E1) on distribue généralement au temps t^e. On notera donc $(E1) = (E2) = (E)$. Cela n'est pas le cas des méthodes (I1) et (I2), comme nous allons le voir plus loin à travers un exemple à l'ordre 2.

Pour être complet, on pourrait citer une dernière famille de méthodes qui consiste à évaluer le problème sur un certain intervalle de temps au sens des distributions. On aboutirait donc à une formulation faible en espace et en temps, et si les fonctions test sont prises égales aux fonctions de l'interpolation espace-temps,

à une méthode de Galerkin (on peut aussi modifier ces fonctions test, et créer ainsi des schémas $\mathcal{RD}$ dans le formalisme Petrov-Galerkin). Le problème est qu'a priori, cette méthode lie le traitement en espace et celui en temps. Il ne s'agit donc pas d'une voie naturelle pour obtenir une semi-discrétisation. Cependant, il est sans doute possible d'y parvenir en choisissant des fonctions d'interpolation qui découplent les représentations spatiale et temporelle. Un exemple peut être trouvé en utilisant les polynômes de Lagrange :

$$U_h(\overrightarrow{x},t) = \sum_{t^n \in [t_1;t_2]} \sum_{M_i \in \Omega_h} U_i^n \varphi_i(\overrightarrow{x}) l^n(t)$$

où $l^n(t)$ est le polynôme de Lagrange 1D associé au degré de liberté temporel t^n. Bien que cette méthode puisse permettre de construire un schéma $\mathcal{RD}$ semi-discrétisé (au sens où une modification du schéma en temps n'affecte pas formellement la distribution en espace, et vice-versa), nous ne l'avons pas envisagée dans nos travaux et nous l'évoquons ici plutôt comme une perspective.

Dans les méthodes citées plus haut, on qualifie a posteriori d'implicites celles qui font intervenir une évaluation des flux (ou des sources) à des dates encore inconnues, et qui nécessitent, pour calculer celles-ci, d'inverser un système linéaire. Les autres sont les méthodes explicites.

4.2.2 Exemples de schémas implicites et mise en oeuvre

Pour illustrer les idées précédentes, concentrons-nous sur la construction effective d'un schéma implicite précis à l'ordre 2 en temps comme en espace, sur un domaine 2D de triangles P_1 (pour simplifier). Ce sera l'occasion de montrer les différentes étapes de la construction du problème à résoudre. Nous allons commencer par l'approche (E), puis viendront (I1) et (I2).

Évaluation et Différences Finies d'ordre 2 : le schéma de Gear

On se considère au temps t^n, donc nous supposons connue la solution U_h^n. Pour obtenir une formulation implicite via une méthode de type (E), on peut évaluer les équations au temps inconnu suivant, t^{n+1}. Pour être précis, cela revient à appliquer la méthode (E1). On part alors dans un premier temps de :

$$r(U^{n+1}) = \left(\frac{\partial U}{\partial t}\right)^{n+1} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U^{n+1}) + g(U^{n+1}) = 0$$

Ensuite, on souhaite évaluer la dérivée en temps avec un schéma Différences Finies (DF) d'un certain ordre, généralement le même ordre qu'en espace. Les schémas ainsi construits forment la famille des schémas BDF (*Backward Difference Formula*). Pour obtenir l'ordre 2, on doit faire appel à deux dates connues, généralement les plus récentes donc t^n et t^{n-1}. En maniant des développement de Taylor en temps de U, on obtient tous calculs faits :

$$\left(\frac{\partial U}{\partial t}\right)^{n+1} = \frac{\delta U}{\delta t}^{G} + O\left((\max(\Delta t^{n-1}, \Delta t^n))^2\right)$$

avec

$$\frac{\delta U}{\delta t}^{G} = \left(\frac{1}{\Delta t^n} + \frac{1}{\Delta t^{n-1} + \Delta t^n}\right)\left(U^{n+1} - U^n\right) - \frac{\Delta t^n}{\Delta t^{n-1}\left(\Delta t^{n-1} + \Delta t^n\right)}\left(U^n - U^{n-1}\right)$$

où $\Delta t^n = t^{n+1} - t^n$ et $\Delta t^{n-1} = t^n - t^{n-1}$, et l'exposant ^{G} est introduit pour signifier qu'il s'agit en réalité d'un schéma bien connu, celui de Gear. On peut aisément concevoir qu'à chaque montée en ordre de tels schémas, il faudra faire intervenir une date antérieure de plus : t^{n-2}, puis t^{n-3}, etc. Si cette démarche n'implique quasiment aucun coût de calcul supplémentaire, elle est en revanche très gourmande en mémoire, puisqu'elle nécessite la sauvegarde des solutions aux dates requises par le schéma DF (soit (p variables $\times$ le nombre total de degrés de liberté) pour chaque date).

Le second temps est celui de la discrétisation en espace via un schéma $\mathcal{RD}$. Pour être cohérent avec l'ordre en temps, on cherche à atteindre l'ordre 2. L'interpolation U_h est donc linéaire par triangle T. Comme chaque distribution s'exprime en fonction du résidu, on commence par calculer celui-ci :

$$\begin{aligned}
\forall T \subset \Omega_h, \quad \Phi^{n+1} &= \int_T \left(\frac{\partial U_h}{\partial t}\right)^{n+1} dV + \int_{\partial T} \overrightarrow{F}(U_h^{n+1}) \cdot \overrightarrow{n} d\partial T + \int_T g(U_h^{n+1}) dV \\
&= \frac{|T|}{3} \sum_{M_i \in T} \frac{\delta U_i}{\delta t}^G \left(+O(h^2 \Delta t^2)\right) + \phi^{n+1} + \frac{|T|}{3} \sum_{M_i \in T} g(U_i^{n+1}) \qquad (4.2.1) \\
\phi^{n+1} &= \frac{1}{2} \sum_{M_i \in T} \overrightarrow{F}_i^{n+1} \cdot \overrightarrow{n_i} \ \left(+O(h^3)\right)
\end{aligned}$$

où on a utilisé, pour calculer la fluctuation ϕ^{n+1}, une formule des trapèzes avec la définition des normales de la figure 3.5. L'intégrale des termes de source est exacte puisqu'on rappelle que g est linéaire. Ensuite on établit une distribution Φ_i^{n+1} avec un schéma quelconque. Parce qu'il n'y a pas de raison d'aller vers des complications en faisant appel aux dates antérieurs pour définir les paramètres du schéma, on fait tout au temps t^{n+1}. C'est justement puisqu'on raisonne toujours ainsi que nous avons assimilé les approches (E1) et (E2). Des exemples de distributions sont alors, $\forall T \subset \Omega_h$ et $\forall M_i \in T$:

$$\begin{aligned}
\left(\Phi_i^{n+1}\right)^{\mathrm{LxF}} &= \frac{|T|}{3} \frac{\delta U_i}{\delta t}^G + \frac{1}{3}\phi^{n+1} + \frac{|T|}{3} g(U_i^{n+1}) + \alpha(U_h^{n+1}) \left(U_i^{n+1} - U_j^{n+1}\right) \\
\left(\Phi_i^{n+1}\right)^{\mathrm{N}} &= \frac{|T|}{3} \frac{\delta U_i}{\delta t}^G + K_i^+(U_h^{n+1}) \left(U_i^{n+1} - \widetilde{U_h}^{n+1}\right) + \frac{|T|}{3} g(U_i^{n+1}) \\
\left(\Phi_i^{n+1}\right)^{\mathrm{LDA}} &= \frac{|T|}{3} \frac{\delta U_i}{\delta t}^G + K_i^+(U_h^{n+1}) N(U_h^{n+1}) \phi^{n+1} + \frac{|T|}{3} g(U_i^{n+1}) \\
\left(\Phi_i^{n+1}\right)^{\mathrm{SU}} &= \frac{|T|}{3} \frac{\delta U_i}{\delta t}^G + \frac{1}{3}\phi^{n+1} + \frac{|T|}{3} g(U_i^{n+1}) \\
&\quad + \frac{1}{6} \sum_{M_j \in T} \left(\frac{\partial F}{\partial U}(U_j^{n+1}) \cdot \overrightarrow{n_i}\right) N(U_h^{n+1}) \left(\frac{\delta U_i}{\delta t}^G + \frac{1}{2|T|} \sum_{M_k \in T} \overrightarrow{F}_k^{n+1} \cdot \overrightarrow{n}_k + g(U_h^{n+1})\right)
\end{aligned}$$

le dernier terme du schéma SU correspondant à une intégration numérique d'ordre 2 du terme de stabilisation (ce n'est qu'une façon de faire parmi toutes celles qui préserveraient l'ordre 2). On a fait exprès, dans ces exemples, de choisir toujours la même matrice de masse décentrée aux noeuds. C'est l'occasion de revenir sur ce sujet. Les schémas précédents ne sont pas d'ordre 2 mais seulement d'ordre 1, car ils ne respectent pas la condition (4.1.10). Il faut donc appliquer à toutes ces distributions une limitation qui permette d'obtenir un schéma du type :

$$\Phi_i^{n+1} = \beta_i^* \Phi^{n+1}$$

À ce stade, il faut comprendre que quel que soit le choix des matrices de masse, si celles-ci ne respectent pas (4.1.10), on a recours à une limitation. L'impact de leur choix sur le schéma final est donc assez léger. Pour certains de ces schémas, comme quel que soit le choix des matrices de masse, ils ne sont pas $\mathcal{LP}$, il aurait de toute façon été nécessaire de recourir à une limitation pour avoir l'ordre 2. Pour ces schémas-là, le choix de la matrice de masse n'a donc pas a priori une très grande importance (à l'ordre 2, ceci a été largement confirmé par les simulations). Quant au schéma LDA, il fait figure d'exception : il est le seul à être $\mathcal{LP}$ en stationnaire. Si le premier choix pour les deux matrices de masse est (4.1.10), on évite donc

l'étape de limitation qui est assez onéreuse. C'est d'ailleurs tout l'intérêt du schéma LDA de proposer une distribution naturellement $\mathcal{LP}$ et *upwind*, et il serait donc peu judicieux d'utiliser d'autres matrices de masses. Le schéma LDA sera donc toujours :

$$\left(\Phi_i^{n+1}\right)^{\text{LDA}} = K_i^+(U_h^{n+1})N(U_h^{n+1})\Phi^{n+1}$$

avec le résidu Φ^{n+1} donné par (4.2.1). Avec le même choix de matrices de masse, i.e. (4.1.10), les autres schémas deviendraient :

$$\begin{aligned}
\left(\Phi_i^{n+1}\right)^{\text{LxF}} &= \frac{1}{3}\Phi^{n+1} + \alpha(U_h^{n+1})\left(U_i^{n+1} - U_j^{n+1}\right) \\
\left(\Phi_i^{n+1}\right)^{\text{N}} &= K_i^+(U_h^{n+1})\left(U_i^{n+1} - \widetilde{U}^{n+1}\right) \\
\left(\Phi_i^{n+1}\right)^{\text{SU}} &= \frac{1}{3}\Phi^{n+1} + \frac{1}{6}\sum_{M_j\in T}\left(\frac{\partial F}{\partial U}(U_j^{n+1})\cdot\overrightarrow{n_i}\right)N(U_h^{n+1})\left(\frac{\delta {U_i}^G}{\delta t} + \frac{1}{2|T|}\sum_{M_k\in T}\overrightarrow{F}_k^{n+1}\cdot\overrightarrow{n}_k + g(U_h^{n+1})\right)
\end{aligned}$$

pour le schéma N, la différence réside dans la définition de la quantité $\widetilde{U}^{n+1}$ qui conserve à présent le résidu Φ^{n+1}, alors qu'elle ne conservait auparavant que la fluctuation ϕ^{n+1}. Comme nous le disions, aucun de ces schémas n'est $\mathcal{LP}$. Cependant, on peut remarquer que le schéma SU est le seul parmi ces trois qui puisse se passer d'une limitation, si le terme de stabilisation est calculé avec une précision suffisante. Pour s'en assurer, il suffit de reprendre l'analyse de précision faite précédemment en remarquant, après quelques calculs qu'on ne détaillera pas ici, que si U^{n+1} désigne la solution exacte au temps t^{n+1}, on a bien :

$$\mathcal{S}_i(U^{n+1}) = 0 \Longrightarrow \left\|\mathcal{S}_i^h(U_h^{n+1}) - \mathcal{S}_i(U^{n+1})\right\| = O(h^{k+1})$$

Enfin, si le schéma employé est le N ou le Lax-Friedrichs, une fois la limitation effectuée, il ne reste plus qu'à rajouter l'opérateur de stabilisation du schéma SU avec un senseur de choc, en suivant les modèles de la section 3.4.2 (juste en rajoutant que les diverses quantités sont prises au temps t^{n+1}).

Nous avons donc vu en détail différents schémas d'ordre 2 constructibles à partir : de l'approche (E), des schémas stationnaires du chapitre précédent, et dans une moindre mesure, du choix des matrices de masse parmi toutes celles qui seraient envisageables en restant consistantes (les possibilités étant infinies). Nous verrons plus loin comment en déduire U_h^{n+1}, c'est-à-dire comment résoudre, en chaque degré de liberté, le problème posé par le rassemblement des contributions de tous les éléments voisins. Voyons maintenant, de manière un peu plus synthétique pour ne pas continuer à trop alourdir le texte, comment obtenir des schémas d'ordre 2 issus des méthodes (I1) ou (I2).

Intégration en temps d'ordre 2 : le schéma de Crank-Nicolson

Nous allons commencer par l'approche (I1). On reste dans le même contexte d'un domaine 2D maillé uniquement à l'aide de triangles P_1, car on recherche l'ordre 2 seulement. La première étape consiste à intégrer les équations sur un intervalle de temps. Le plus approprié ici est de choisir $\left[t^n; t^{n+1}\right]$, puisque t^n est l'état où la solution est connue et que t^{n+1} correspond à l'état que nous recherchons. Aucune autre date n'est nécessaire pour obtenir l'ordre 2. On écrit alors :

$$\int_{t^n}^{t^{n+1}}\left(\frac{\partial U}{\partial t} + \overrightarrow{\nabla}\cdot\overrightarrow{F}(U) + g(U)\right)dt = U^{n+1} - U^n + \int_{t^n}^{t^{n+1}}\left(\overrightarrow{\nabla}\cdot\overrightarrow{F}(U) + g(U)\right)dt = 0$$

$$\Longrightarrow \quad U^{n+1} - U^n + \frac{\Delta t^n}{2}\left(\overrightarrow{\nabla}\cdot\overrightarrow{F}(U^{n+1}) + \overrightarrow{\nabla}\cdot\overrightarrow{F}(U^n) + g(U^{n+1}) + g(U^n)\right) = 0\left[+O\left((\Delta t^n)^3\right)\right]$$

où la formule de quadrature utilisée est la même que dans la partie précédente, à savoir une formule des trapèzes. Le schéma en temps ainsi obtenu est celui de Crank-Nicolson. Pour harmoniser la programmation des différentes méthodes, on utilise plutôt par convention :

$$\frac{U^{n+1}-U^n}{\Delta t^n}+\frac{1}{2}\left(\overrightarrow{\nabla}\cdot\overrightarrow{F}(U^{n+1})+\overrightarrow{\nabla}\cdot\overrightarrow{F}(U^n)+g(U^{n+1})+g(U^n)\right)=0\left[+O\left((\Delta t^n)^2\right)\right]$$

qui équivaut à linéariser en temps et à évaluer au temps $t^{n+1/2}$. On aurait d'ailleurs pu construire un autre schéma que Crank-Nicolson, ayant la même précision, en remplaçant la formule des trapèzes par une évaluation au point milieu.

Il ne reste ensuite plus qu'à discrétiser tout ceci en espace. Comme précédemment, l'interpolation est linéaire par triangle, et la fluctuation au temps t^{n+1} s'exprime donc encore par (4.2.1). Pour ce qui est du résidu total, si on intègre l'équation précédente sur un élément T quelconque, on obtient une quantité qu'on décide d'appeler :

$$\Phi^{n+\frac{1}{2}}=\frac{1}{2}\left(\Phi^{n+1}+\Phi^n\right)$$

avec Φ^{n+1} donné par (4.2.1) en remplaçant seulement la dérivée en temps :

$$\frac{\delta U_i}{\delta t}^G\mapsto\frac{U_i^{n+1}-U_i^n}{\Delta t^n}=\frac{\delta U_i}{\delta t}^{CN}$$

Pour le résidu Φ^n, c'est la même chose au temps t^n. La différence avec la méthode (E) est qu'ici, la date à laquelle la distribution doit se faire est encore moins claire. En réalité, comme on n'analyse pas la stabilité des schémas instationnaires, on peut se permettre plusieurs choix. Suivant les modèles vus pour le schéma de Gear, on peut dire que la distribution finale, qu'il y ait limitation ou non, peut s'écrire :

$$\beta_i^{n+1}\Phi^{n+\frac{1}{2}}\text{ ou }\beta_i^{n+\frac{1}{2}}\Phi^{n+\frac{1}{2}}\text{ ou }\beta_i^n\Phi^{n+\frac{1}{2}}$$

avec bien sûr :

$$\beta_i^{n+1}=\beta_i(U_h^{n+1}),\ \beta_i^{n+\frac{1}{2}}=\beta_i\left(\frac{U_h^{n+1}+U_h^n}{2}\right),\ \beta_i^n=\beta_i(U_h^n)$$

Quel que soit le choix parmi ceux-ci, les β_i étant bornés, le schéma est bel et bien d'ordre élevé. Cet aspect arbitraire pourrait peut-être être levé par une étude de stabilité pointant le meilleur choix.

Voyons ce que donnerait l'approche (I2). Cette fois-ci, nous commençons par discrétiser en espace. On se considère donc à une date t variable à laquelle tous les calculs se font, et on aboutit, après limitation si besoin, à la quantité :

$$\Phi_i(t)=\beta_i^*(t)\Phi(t)$$

avec :

$$\Phi(t)=\frac{|T|}{3}\sum_{M_i\in T}\frac{dU_i}{dt}(t)+\frac{1}{2}\sum_{M_i\in T}\overrightarrow{F}(U_i(t))\cdot\overrightarrow{n_i}+\frac{|T|}{3}\sum_{M_i\in T}g(U_i(t))$$

Cette expression doit ensuite être intégrée, ici encore entre t^n et t^{n+1}, de façon à ce que le système final à résoudre soit en chaque degré de liberté M_i :

$$\sum_{T\subset\mathcal{T}_i}\int_{t^n}^{t^{n+1}}\Phi_i(t)dt=0$$

À présent si on applique la méthode des trapèzes, on obtient :

$$\frac{\Delta t^n}{2}\sum_{T\subset\mathcal{T}_i}\left(\beta_i^{n+1}\Phi^{n+1}+\beta_i^n\Phi^n\right)=0$$

On voit bien que cette formulation est assez différente de (I1). Certes, l'ambigüité sur la date d'évaluation des β_i disparait, mais cela n'apporte aucun bénéfice particulier. En revanche, chaque résidu fait intervenir une dérivée en temps à approcher, puisque celle-ci n'est plus intégrée directement. Il faut donc recourir en plus à des schémas Différences Finies en temps pour évaluler ces dérivées dans Φ^{n+1} et Φ^n, à la manière des méthodes (E). Ces efforts supplémentaires, qui rendent en plus l'écriture du schéma plus lourde, n'apportent aucune bonne propriété en contrepartie. C'est la raison pour laquelle on privilégie toujours l'approche (I1).

En conclusion, dans la pratique, nous commençons toujours par discrétiser en temps, que la méthode soit de type (E) ou (I). On a vu deux cas précis qui tentent d'exhiber la démarche de la construction de schémas $\mathcal{RD}$ implicites pour les problèmes instationnaires et non homogènes, dans le cas simple de l'ordre 2. L'approche (E) amène naturellement au schéma de Gear, tandis que (I) se traduit par un schéma de Crank-Nicolson. Dans le cadre de la résolution des équations différentielles ordinaires, ces deux schémas implicites sont connus pour être incondionnellement stables.

Remarque sur l'extension aux ordres supérieurs

Bien que nous ne soyons pas allé au-delà de l'ordre 2 dans nos simulations implicites, nous pouvons faire remarquer que le passage à l'ordre très élevé peut amener une réflexion sur la meilleure façon de procéder. Un façon d'étendre les méthodes (I1) implicites serait d'utiliser des formules de quadrature faisant intervenir davantage de degrés de liberté entre t^n et t^{n+1}, auxquels la solution devra être reconstruite au préalable, puis stockée. Des exemples de telles méthodes sont celles de Runge-Kutta implicites. L'alternative est d'élargir le domaine d'intégration au-delà de $\left[t^n; t^{n+1}\right]$ de manière à ce que les points de quadrature soient les dates auxquelles la solution est de toute façon calculée (ce qui nécessite de les sauvegarder, et ne fait donc rien économiser en mémoire). Ceci reviendra à peu de choses près à l'approche (E) où nous avons d'emblée choisi d'appliquer les différences finies en faisant appel aux dates antérieures (U^{n-1}) et non pas en construisant des solutions intermédiaires (on aurait pu construire un $U^{n+1/2}$). Une méthode compacte en temps nécessite de construire des degrés de liberté supplémentaires, et si celle-ci est implicite, chacun d'entre eux entraîne la résolution d'un système linéaire. Cela serait donc clairement prohibitif en temps de calcul. Et comme il faut bien stocker ces solutions intermédiaires pour calculer les suivantes, on ne gagnerait rien en mémoire. En espace, un avantage de la montée en ordre compacte est la facilitation de la parallélisation, ce qui n'a aucun intérêt en temps. Donc, pour résumer, lors de la montée en ordre, on devrait sacrifier la compacité du schéma en temps pour gagner en performances. Les seuls arguments pouvant aller contre cet avis porteraient sur la stabilité. Nos remarques sur l'ordre élevé sont plutôt à visée prospective, et nous n'avons pas eu ni l'occasion ni le besoin d'étudier cette stabilité pour nos travaux.

Ce qu'il faut résoudre

Au final, quelle que soit la méthode utilisée, on aboutira à un problème qui aura pris la forme générale, en tout degré de liberté M_i du maillage :

$$\sum_{E \subset \mathcal{T}_i} \left(\Phi_i^E\right)^{n+1} = 0 \tag{4.2.2}$$

Pour pouvoir faire ce calcul et en déduire U_h^{n+1}, l'étape suivante demanderait d'évaluer la divergence des flux et les termes de source au temps t^{n+1}, ce qui n'est bien sûr pas possible. Le système formé par (4.2.2) doit alors être reconsidéré avec un point de vue plus large : on cherche à annuler une certaine fonction non linéaire de U^{n+1}. Une méthode toute désignée pour résoudre ce genre de problèmes est celle de Newton-Raphson, dont nous allons à présent expliquer la mise en oeuvre.

4.2.3 Résolution itérative

Linéarisation du système : la méthode de Newton-Raphson

Pour un problème scalaire, le principe consiste à se placer en un point initial arbitraire (mais pas trop éloigné du 0 recherché si cela est possible) puis à s'approcher de la solution en calculant la coordonnée du point d'intersection entre la tangente à la fonction et l'axe des abscisses, selon la formule :

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

La tangente est décrite par un développement limité à l'ordre 1, c'est une linéarisation. On le voit mieux en l'écrivant ainsi :

$$f(x_{k+1}) = f(x_k) + f'(x_k)(x_{k+1} - x_k)$$

avec par définition $f(x_{k+1}) = 0$. La méthode de Newton-Raphson convergera soit vers un extremum local, soit vers le zéro recherché, selon ce qu'elle rencontrera en premier. D'où la nécessité de partir d'une solution pas trop éloignée de ce zéro. En pratique, dans les tests instationnaires, ce critère est toujours assez bien vérifié grâce à la physique sous-jacente, deux états 'proches' dans le temps ne pouvant pas être 'trop' différents l'un de l'autre. Dans le cas d'un système d'équations, on fait la même chose avec une fonction de $\mathbb{R}^p$ dans $\mathbb{R}^p$. Les inconnues de notre problème sont les valeurs de U en chaque degré de liberté au temps t^{n+1}. Comme on l'a vu dans le chapitre précédent tous les schémas distribuant le résidu ne font intervenir que les premiers voisins du degré de liberté à metre à jour. Notons $\mathcal{T}_i$ l'ensemble de ces premiers voisins de M_i augmenté de M_i lui-même, et $f_i = \sum_{E \subset \mathcal{T}_i} \left(\Phi_i^E\right)$ l'ensemble des contributions de chaque élément contenant M_i. Enfin, pour distinguer la k-ième itération de Newton de la date physique t^n à laquelle on se situe, on utilisera l'exposant ^{n,k}. Une itération de la méthode de Newton s'écrira alors :

$$\sum_{M_j \in \mathcal{T}_i} \left(\frac{\partial f_i^{n+1}}{\partial U_j^{n+1}}\right)^{n,k} \left(U_j^{n,k+1} - U_j^{n,k}\right) = -f_i^{n,k} = -\sum_{E \subset \mathcal{T}_i} \left(\Phi_i^E\right)^{n,k} \tag{4.2.3}$$

Les matrices de gauche sont les jacobiennes du schéma par rapport à chaque U_j^{n+1}, mais évaluées en $\left\{U_j^n\right\}_{M_j \in \mathcal{T}_i}$. On a donc un système linéaire de la taille du nombre de variables multiplié par le nombre de degrés de liberté sur le maillage. Il doit être assemblé et résolu à chaque itération, ce qui explique le coût très élevé des méthodes implicites. Son utilisation n'a donc d'intérêt que si elle permet de considérer des pas de temps grands devant celui imposé par la limite de stabilité en explicite et si elle parvient à converger raisonnablement bien pour ces pas de temps élevés.

La matrice obtenue a une structure par blocs carrés de la taille du nombre de variables. Du fait de la compacité des schémas distribuant le résidu, l'assemblage ne remplit dans chaque ligne qu'autant de blocs qu'il y a de degrés de liberté contenus dans $\mathcal{T}_i$ (en comptant un bloc diagonal donc). La matrice est donc clairement très creuse. Nous utilisons dans FluidBox un stockage CSR de cette matrice, et nous disposons de plusieurs algorithmes pour résoudre le système linéaire (Jacobi, Gauss-Seidel, GMRES, ...). Il est aussi possible d'utiliser des solveurs externes, tels que PaStiX ou HiPS.

On peut améliorer cette méthode en voyant ce qui est résolu par la méthode de Newton comme un problème stationnaire. Dans ce cas, on a coutume d'introduire une dépendance en ce qu'on appellera un "pseudo-temps" τ et de s'approcher de la limite de convergence théorique $\tau \to +\infty$ à laquelle correspond l'équilibre recherché pour peu que les conditions initiales et aux limites y conduisent. On définit ainsi une "méthode de Newton modifiée" dont on verra qu'elle agit comme une méthode de relaxation. Revenons d'abord sur nos notations. Nous avons défini le vecteur f dont chaque composante correspond à un degré de liberté du maillage et prend pour valeur la somme des contributions envoyées par les degrés de liberté

voisins (l'assemblage se faisant par élément) :

$$f_i^n = \sum_{T, i \in T} \Phi_i^T \left(U^n \right)$$

A chaque noeud i est associée une cellule de contrôle notée C_i (à la manière des Volumes Finis) et l'ensemble de ses degrés de liberté voisins a été nommé $\mathcal{T}_i$. Définissons un nouveau vecteur d'inconnues $V(\tau)$ tel que $\forall i \in \Omega, V_i(0) = U_i^n$, et vérifiant :

$$\int_{C_i} \frac{\partial V_i}{\partial \tau} dx + f_i \left(V(\tau) \right) = g_i \left(V(\tau) \right) = 0$$

Si nous appliquons maintenant la méthode de Newton-Raphson à la nouvelle fonction g, nous écrirons :

$$\begin{aligned}
\sum_{j \in \mathcal{T}_i} \frac{\partial g_i^{k+1}}{\partial V_j^{k+1}} (V^k) \left(V_j^{k+1} - V_j^k \right) &= -g_i(V^k) \\
\Rightarrow \quad \sum_{j \in \mathcal{T}_i} \left[\cancel{\int_{C_i} \frac{\partial}{\partial \tau} \left(\frac{\partial V_i}{\partial V_j} \right)^k dx} + \frac{\partial f_i^{k+1}}{\partial V_j^{k+1}} (V^k) \right] \left(V_j^{k+1} - V_j^k \right) &= -\int_{C_i} \frac{\partial V_i^k}{\partial \tau} dx - f_i(V^k) \\
\Rightarrow \quad \int_{C_i} \left(\frac{\partial V_i}{\partial \tau} \right)^k dx + \sum_{j \in \mathcal{T}_i} \frac{\partial f_i^{k+1}}{\partial V_j^{k+1}} (V^k) \left(V_j^{k+1} - V_j^k \right) &= -f_i^k
\end{aligned}$$

A convergence, lorsque $\tau \to +\infty$, $\dfrac{\partial V_i}{\partial \tau} \to 0$ et $V_j^{k+1} - V_j^k \to 0$ car l'équilibre existe et est solution du problème (posé avec ses conditions, en particulier $V(0) = U^n$). C'est bien pour cette raison que nous savons que nous convergeons vers un zéro de la fonction f. Il est clair aussi que la précision de la discrétisation en pseudo-temps n'a pas d'impact sur la précision à convergence. Nous nous contenterons donc d'une discrétisation à l'ordre 1 du premier terme :

$$\int_{C_i} \left(\frac{\partial V_i}{\partial \tau} \right)^k dx \approx \int_{C_i} \frac{V_i^{k+1} - V_i^k}{\Delta \tau^k} dx$$

Puisque nous avons vu que la jacobienne de ce terme est nulle, nous avons en fait établi un schéma implicite linéarisé en τ qui est une approximation de :

$$\int_{C_i} \frac{V_i^{k+1} - V_i^k}{\Delta \tau^k} dx + f_i^{k+1} = 0$$

Par conséquent, la méthode est A-stable pour de grands pas $\Delta \tau^k$, ce que nous voudrions exploiter pour atteindre une convergence en un nombre d'itérations raisonnable. Nous prendrons $\Delta \tau^k$ croissant avec k, jusqu'à des valeurs très élevées. Nous forcerons donc encore davantage le premier terme de la méthode à être négligeable près de la convergence, ce qui fait que plus nous nous rapprocherons de la convergence plus la méthode sera en réalité purement celle de Newton-Raphson (4.2.3).

Suivant la remarque précédente concernant la précision du schéma à convergence, nous discrétiserons le premier terme à l'ordre 1 également en espace. Revenons enfin à la notation $U^{n,k} = V^k$ pour obtenir :

$$\sum_{j \in \mathcal{T}_i} \left[\frac{|C_i|}{\Delta \tau^k} \delta_{ij} + \frac{\partial f_i^{n+1}}{\partial U_j^{n+1}} (U^{n,k}) \right] \left(U_j^{n,k+1} - U_j^{n,k} \right) = -f_i^{n,k} \tag{4.2.4}$$

Nous pouvons donc contrôler la quantité que nous mettons sur la diagonale de la matrice, de la même manière que nous le ferions avec un schéma de relaxation. Si nous prenons $\Delta \tau^k$ petit devant $|C_i|$, comme nous nous l'autorisons sur les premières itérations, la diagonale de la matrice sera largement dominante

et le schéma sera proche d'un schéma explicite en pseudo-temps. A l'inverse, à mesure que l'équilibre s'établira, le fait de prendre $\Delta\tau^k$ grand devant $|C_i|$ donnera un schéma proche de l'implicite linéarisé classique (4.2.3). On retrouve ces méthodes dans la littérature sous la dénomination "dual time stepping schemes" (voir par exemple [50], qui est en libre accès sur la page personnelle du premier auteur).

Difficultés rencontrées

Dans les tests que nous avons effectués, nous nous sommes très souvent heurtés à un défaut de convergence de la méthode, rendant très discutable le choix d'une stratégie de résolution implcite. Il y a potentiellement deux raisons qui peuvent l'expliquer.

La première difficulté réside dans la façon d'évaluer les jacobiennes du schéma. En premier lieu, il faut remarquer que les schémas que nous avons présentés ne sont pas a priori différentiables, même s'il est possible de les régulariser. En second lieu, même si nous supposions que la différentiabilité soit acquise par une reformulation des schémas, le calcul des jacobiennes exactes serait très lourd, à la fois formellement et en temps de calcul. Par exemple, les schémas multidimensionnels upwind (N et LDA) font intervenir une fonction max qui peut être régularisée en la remplaçant par $\overset{1+\epsilon}{\max}$ avec $\epsilon > 0$... Mais il resterait alors à calculer les jacobiennes des matrices du système propre. Le calcul est sans doute possible, mais s'avère très complexe et pour un résultat dont on ne sait pas a priori s'il justifierait le supplément en temps de calcul (il paraît toutefois normal de dire que la méthode de Newton-Raphson doit converger si le schéma est régulier et dérivé de façon exacte).

Deux solutions alternatives ont été envisagées :

. soit le recours à des jacobiennes formelles approximatives, c'est-à-dire qu'on ne dérive qu'en partie le schéma,

. soit l'évaluation de jacobiennes approchées numériquement par une méthode des différences finies.

Pour donner un exemple simple, les jacobiennes que nous utilisons pour le schéma de Rusanov limité et stabilisé sont celles du schéma d'ordre 1, où on ne dérive pas α. Si on considère la contribution d'un seul élément à un de ses noeuds i, écrite avec une distribution des sources décentrée aux noeuds, sa jacobienne approximative s'écrira :

$$\frac{\partial \Phi_i^{n+1}}{\partial U_j^{n+1}}(U^n) = \frac{1}{N_s}\left(\frac{\partial \left(\partial_t U_h\right)_i^{n+1}}{\partial U_i^{n+1}}\delta_{ij} + \sum_{j\in T}\frac{\partial \phi}{\partial U_j}(U^n) + \alpha(U^n)\left(N_s\delta_{ij}-1\right)\right)$$

Par conséquent, si on prend par exemple une matrice de masse décentrée aux noeuds sur des éléments triangulaires P1 et avec une fluctuation calculée par la règle des trapèzes, la matrice implicite A à une itération k de la méthode de Newton sera donnée par :

$$\begin{aligned}
A_{ii}(U^{n,k}) &= \sum_{T,i\in T}\left[\frac{|T|}{3\Delta t^n}Id + \frac{1}{3}\left(\frac{K_i^T(U_i^{n,k})}{2} + 2\alpha^T(U^{n,k})Id\right)\right] \\
&= \frac{|C_i|}{\Delta t^n}Id + \frac{1}{3}\sum_{T\ni i}\left(\frac{K_i^T}{2} + 2\alpha^T Id\right) \\
A_{ij}(U^{n,k}) &= \frac{1}{3}\sum_{T\ni i}\left(\frac{K_j^T}{2} - \alpha^T Id\right) \qquad \forall j\in\mathcal{T}_i, j\neq i
\end{aligned}$$

L'ensemble $\mathcal{T}_i$ désigne tous les premiers voisins du degré de liberté i. On ne peut pas dire grand-chose sur le conditionnement de cette matrice à cause de sa structure par blocs, ou alors de manière un peu qualitative.

On revient alors au problème scalaire, comme on peut raisonnablement espérer que le comportement du schéma soit globalement proche dans les deux cas. Les matrices K_i^T deviennent des coefficients $k_i^T = \overrightarrow{\lambda}(u_i) \cdot \overrightarrow{n}_i^T$ tels que $|k_i^T| \leq \alpha^T$. La matrice A est donc clairement une L-matrice, à savoir que $A_{ii} > 0$ et $A_{ij} \leq 0$. On fait une approximation supplémentaire en linéarisant les jacobiennes, c'est-à-dire qu'on a en réalité $k_i = \overrightarrow{\lambda}(\overline{u}) \cdot \overrightarrow{n}_i$ et donc $\sum_{j \in T} k_j^T = 0$ (la somme des normales étant nulle). Sur chaque ligne de la matrice A, on aura alors :

$$\begin{aligned}
|A_{ii}| - \sum_{j \in \mathcal{T}_i} |A_{ij}| &= \sum_{T, i \in T} \frac{1}{3} \left(\left| \frac{|T|}{\Delta t^n} + \frac{k_i^T}{2} + 2\alpha^T \right| - \left| k_j^T - \alpha^T \right| - \left| k_k^T - \alpha^T \right| \right) \\
&= \frac{1}{3} \sum_{T, i \in T} \left(\frac{|T|}{\Delta t^n} + \frac{k_i^T}{2} + 2\alpha - (\alpha - k_j^T) - (\alpha - k_k^T) \right) \\
&= \frac{|C_i|}{\Delta t^n} > 0
\end{aligned}$$

A est donc à diagonale strictement dominante. On peut montrer le même résultat pour le schéma N dans la même configuration ([6]). La résolution du système linéaire à chaque itération de Newton ne pose donc aucun problème, même si la jacobienne est très éloignée de ce qu'il y a dans le second membre d'un point de vue formel. En revanche, cet écart de signification entre la matrice et le second membre est un gros problème pour la convergence de la méthode de Newton : la direction de descente est potentiellement très différente de la direction théorique.

La deuxième difficulté est due au fait que si nous essayons d'autres manières de construire les jacobiennes, de façon plus exacte en régularisant certaines fonctions au besoin ou avec une estimation numérique comme le font les Différences Finies, le schéma ainsi que le problème à résoudre peuvent donner une matrice globale mal conditionnée. Si c'est là le problème majeur, sans doute est-il inévitable de chercher des techniques de préconditionnement, peut-être même au cas par cas suivant la nature du test à réaliser. Jusqu'à aujourd'hui, le seul préconditionnement utilisé est la multiplication de la matrice par l'inverse de sa diagonale. Il pourrait être intéressant de tester des préconditionnements plus évolués grâce aux solveurs externes que nous avons cités. Si cela ne suffit pas, il faudra alors soit entreprendre de trouver un préconditionnement s'inspirant de la physique des cas tests abordés, soit chercher à améliorer la régularité du schéma et la dérivation formelle de ses jacobiennes.

Nous avons principalement utilisé une construction approximative des jacobiennes, et très peu les Différences Finies car dans tous les cas, nous n'avons jamais réussi à obtenir une convergence aboutie de la méthode de Newton (une division du résidu par au moins 10^5) en un nombre d'itérations acceptable (moins de 50 par pas de temps par exemple). L'utilisation d'un faible pseudo pas de temps pour forcer une diagonale plus fortement dominante durant les premières itérations n'a pas permis d'amélioration le comportement de la méthode de Newton de façon suffisante. En pratique, on utilise cependant toujours un pseudo pas de temps régi par un nombre CFL qui augmente au fur et à mesure que le résidu diminue. Ceci revient à faire de l'explicite (en pseudo-temps) durant les premières itérations puis à passer progressivement à du Newton-Raphson pur au fur et à mesure que le résidu tend vers 0 et que $\Delta\tau$ tend vers l'infini.

Un autre problème se pose en implicite lorsque les cas que nous abordons mettent en jeu des chocs très forts, comme nous en présenterons dans le prochain chapitre. Si le schéma n'est pas monotone, il est clair qu'on ne pourra pas obtenir une bonne convergence et que nous finirons par obtenir des pressions négatives. Pour les schémas non linéaires appliqués à un système d'équations en plusieurs dimensions d'espace, nous ne disposons pas d'une condition formelle de stabilité entropique ou energétique. Si on peut espérer que le schéma soit suffisamment robuste pour une certaine plage de pas de temps, en se basant

sur le cas scalaire où une condition de positivité théorique peut être trouvée, rien ne garantit qu'on puisse utiliser les pas de temps élevés qui justifient l'emploi des méthodes implicites. Lorsque cette observation se traduit concrètement dans les simulations par des conditions CFL très restrictives, il devient clairement préférable d'envisager le recours à des méthodes explicites moins onéreuses.

4.3 L'alternative explicite : Runge-Kutta pour des schémas $\mathcal{RD}$

Comme les méthodes implicites que nous avons présentées ne peuvent parfois pas être mises en oeuvre de façon efficace, nous nous sommes tournés vers des méthodes explicites. De la même manière qu'en implicite, il y a potentiellement deux méthodes pour construire une discrétisation temporelle explicite : pour résumer, soit la construction repose sur le résultat d'une intégration approchée sur un certain intervalle de temps, soit sur une évaluation des équations au temps à la date courante et une approximation de la dérivée en temps faisant intervenir l'état à calculer. Avant d'aller plus loin, nous préférons resituer la problématique dans un cadre plus large. Du fait du découplage entre les traitements en temps et en espace, il faut voir que le problème de la résolution temporelle seule revient à résoudre une équation différentielle ordinaire en temps du premier ordre. La complexité des flux fait qu'il n'est pas possible de procéder à une résolution analytique. Nous devons alors avoir recours à l'une des nombreuses méthodes d'approximation qui existent, dont un panorama très complet est livré dans [21]. Nous aurions pu faire cette remarque également dans la section précédente sur les schémas implicites, mais on s'y concentre davantage sur la capacité à faire converger la méthode de Newton. Ici, le faible coût des méthodes explicites nous permet d'envisager plus sereinement le passage à des méthodes plus complexes (ou qui le deviennent lors de la montée en ordre).

4.3.1 Généralités sur les méthodes de Runge-Kutta

Les méthodes de Runge-Kutta font partie de celles qui sont construites par intégration. Leur analyse est bien trop complexe pour pouvoir être résumée ici en quelques lignes. Pour une description détaillée de ces méthodes et de leurs propriétés, on renvoie le lecteur intéressé à [21]. Retenons simplement qu'une méthode d'ordre k en temps se met en oeuvre à travers k étapes. Chacune de ces étapes correspond au calcul de ce qu'on appelle parfois un prédicteur-correcteur, c'est-à-dire une solution intermédiaire. Considérons une équation différentielle ordinaire du type :

$$\frac{dU}{dt} = f(t, U(t))$$

Une méthode de Runge-Kutta d'ordre k, appliquée sur un intevalle de temps $\left[t^n; t^{n+1}\right]$, s'écrira de manière synthétique :

$$\forall 1 \leq l \leq k,\ U^l = U^n + \Delta t \sum_{s=1}^{l} b_s k_s \tag{4.3.1}$$

où U^k sera égal à U^{n+1} et où les quantités k_s sont définies à chaque étape en fonction des $l-1$ intermédiaires de calcul précédents. Plus précisément, on a :

$$\begin{aligned}
k_1 &= f(t^n, U^n) \\
k_2 &= f(t^n + c_1\Delta t, U^n + a_{21}k_1) \\
k_3 &= f(t^n + c_1\Delta t, U^n + a_{31}k_1 + a_{32}k_2) \\
&\dots \\
k_l &= f(t^n + c_1\Delta t, U^n + a_{l1}k_1 + a_{l2}k_2 + \ldots + a_{l,l-1}k_{l-1})
\end{aligned} \tag{4.3.2}$$

Lorsque la méthode est écrite sous cette forme, on donne généralement les divers coefficients sous la forme d'un tableau de Butcher. Pour les méthodes explicites que nous venons de présenter, il a toujours la forme triangulaire suivante :

$$\begin{array}{c|ccccc} 0 & & & & & \\ c_2 & a_{21} & & & & \\ c_3 & a_{31} & a_{32} & & & \\ \vdots & \vdots & & \ddots & & \\ c_l & a_{l1} & a_{l2} & \cdots & a_{l,l-1} & \\ \hline & b_1 & b_2 & \cdots & b_{l-1} & b_l \end{array} \tag{4.3.3}$$

Les équations que nous avons à résoudre forment une équation différentielle autonome (en omettant la dépendance en espace puisque le traitement est découplé), c'est-à-dire que la variable temporelle n'apparait pas explicitement dans :

$$\frac{\partial U}{\partial t} = f(U(\cdot,t)) = -\overrightarrow{\nabla} \cdot \overrightarrow{F}(U(\cdot,t)) - g(U(\cdot,t)) \tag{4.3.4}$$

Cela fait que les coefficients c_s dans le tableau précédent ne nous sont d'aucune utilité.

D'autres schémas de type Runge-Kutta ont été construits sous une forme a priori différente, et de façon à préserver les propriétés de monotonie des schémas en espace sur des problèmes scalaires non linéaires. Il s'agit des schémas de Runge-Kutta TVD (pour *Total Variation Diminishing*) initiés par Shu et Osher [86] (pour plus de détails, voir également [45], [46], [84], [87]). Dans leur formulation originale, ils diffèrent de (4.3.1)-(4.3.2) et s'expriment, avec les mêmes indices :

$$\forall 1 \leq l \leq k,\ U^l = \sum_{s=0}^{l-1} \left(\alpha_{ls} U^s + \Delta t \beta_{ls} f(U^s)\right) \tag{4.3.5}$$

avec $U^0 = U^n$, $U^k = U^{n+1}$ et $\sum_{s=0}^{l-1} \alpha_{ls} = 1$ pour la consistance, et f identifiée par (4.3.4). Toutefois, il semblerait naturel que les deux approches fassent partie d'un même tout. C'est bien le cas : le rapprochement du formalisme de Shu et Osher avec la représentation de Butcher (4.3.1)-(4.3.2)-(4.3.3) a été fait récemment ([39],[49]).

Exemples de schémas TVD d'ordre 2 et 3

Pour chaque ordre recherché, il existe plusieurs méthodes possibles. Celle qu'on définit comme optimale est celle qui préserve la monotonie avec le nombre CFL le plus élevé. Pour l'ordre 2, la méthode optimale est donnée dans [86] :

$$\begin{aligned} U^1 &= U^n + \Delta t f(U^n) \\ U^{n+1} &= \frac{1}{2}U^n + \frac{1}{2}U^1 + \frac{\Delta t}{2} f(U^1) = U^n + \frac{\Delta t}{2}\left(f(U^n) + f(U^1)\right) \end{aligned} \tag{4.3.6}$$

avec un nombre CFL maximal égal à 1. Pour l'ordre 3, le schéma optimal a été trouvé dans [45] :

$$\begin{aligned} U^1 &= U^n + \Delta t f(U^n) \\ U^2 &= \frac{3}{4}U^n + \frac{1}{4}U^1 + \frac{\Delta t}{4} f(U^1) = U^n + \frac{\Delta t}{4}\left(f(U^n) + f(U^1)\right) \\ U^{n+1} &= \frac{1}{3}U^n + \frac{2}{3}U^2 + \frac{2\Delta t}{3} f(U^2) = U^n + \frac{\Delta t}{6}\left(f(U^n) + f(U^1) + 4f(U^2)\right) \end{aligned} \tag{4.3.7}$$

et le nombre CFL maximal est encore égal à 1. L'obtention de schémas optimaux pour les ordres strictement supérieurs à 4 est un sujet de recherches toujours actif. Il ne reste ensuite plus qu'à identifier la fonction f avec (4.3.4) et à appliquer les schémas $\mathcal{RD}$ en espace. C'est ce que nous allons voir à présent.

4.3.2 Mise en oeuvre dans le contexte $\mathcal{RD}$

Voie directe : le problème des matrices de masse

On rappelle que nous avons donné, dans la première section de ce chapitre, une écriture générique des schémas $\mathcal{RD}$ pour les systèmes instationnaires avec termes de source linéaires. Avec l'emploi d'une méthode de Runge-Kutta TVD à k étapes, on peut reformuler de manière appropriée la distribution en repartant de (4.3.5). En réutilisant les notations vues précédemment, la distribution devient alors :

$$\forall l \leq k,\ \left(\Phi_i^E\right)^l = \sum_{M_j \in E} m_{ij}^1 \frac{\delta U_j}{\delta t}^l + \beta_i^l \phi^{RK(l)}(U_h) + \sum_{M_j \in E} m_{ij}^2 \mathcal{G}^l(U_j) \tag{4.3.8}$$

où il est nécessaire que nous précisions les notations :

$$\frac{\delta U_j}{\delta t}^l = \frac{1}{\Delta t}\left(U^l - \sum_{s=0}^{l-1} \alpha_{ls} U^s\right) \tag{4.3.9}$$

$$\phi^{RK(l)}(U_h) = \sum_{s=0}^{l-1} \beta_{ls} \int_E \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h^s) dV$$

$$\mathcal{G}^l(U_j) = \sum_{s=0}^{l-1} \beta_{ls} g(U_j^s)$$

Même si on ne précise pas les matrices de masse, il reste vrai que pour que le schéma soit d'ordre élevé en espace, celles-ci doivent être prises égales à (4.1.10). À chaque étape l de la construction, et si en chaque degré de liberté $M_i \in \Omega_h$, on assemble toutes les contributions des éléments voisins, le système final à résoudre est donc :

$$\forall M_i \in \Omega_h, \forall l \leq k,\ \sum_{E \subset \mathcal{T}_i} \sum_{M_j \in E} m_{ij}^1 \frac{\delta U_j}{\delta t}^l = -\sum_{E \subset \mathcal{T}_i} \left(\beta_i^l \phi^{RK(l)}(U_h) + \sum_{M_j \in E} m_{ij}^2 \mathcal{G}^l(U_j)\right) \tag{4.3.10}$$

Ces k problèmes à résoudre en chaque degré de liberté M_i (k étant le nombre d'étapes de la méthode de Runge-Kutta, égal à l'ordre de précision jusqu'à l'ordre 4) pourraient être résolus tels quels à chaque pas de temps. Cependant, on constate la présence d'une matrice de masse à gauche, qui a donc besoin d'être inversée. Nous pourrions alors être tentés de procéder à un *mass lumping*, c'est-à-dire à la transformation de cette matrice pleine en une matrice diagonale, comme cela est souvent fait avec un certain succès en Éléments Finis. Toutefois, l'impact qu'aurait cette procédure sur le comportement des schémas $\mathcal{RD}$ n'est pas bien connu. De plus, nous avons déjà dit qu'il existait plusieurs constructions de matrices de masses envisageables avec nos méthodes, contrairement aux méthodes de Galerkin, et que la question portant sur les meilleurs choix restait ouverte (même si,il faut aussi le rappeler, l'importance de ce choix paraît faible). Dans une perspective d'étude, on souhaite donc conserver la matrice de masse pleine, sans la modifier. Dans le cadre de l'ordre élevé, un autre argument est tout simplement la préservation de l'ordre de précision qui requiert une forme bien précise, la condition (4.1.10) qui devient :

$$\forall l \leq k,\ m_{ij}^1 = m_{ij}^2 = \beta_i^l \int_E \varphi_j dV$$

Il est donc impératif de conserver des matrices de masse pleines. Or il faut noter que celles-ci dépendent la plupart du temps de U_h, et dans ce cas-là l'inversion devrait être effectuée k fois par pas de temps, ce qui rendrait l'approche explicite totalement inabordable puisque sa rapidité est déjà restreinte par la contrainte de stabilité sur le pas de temps. Le problème est que vouloir conserver une matrice de masse pleine et ne pas avoir à l'inverser semblent a priori deux choix tout à fait inconciliables. Grâce à l'interprétation Petrov-Galerkin des schémas $\mathcal{RD}$, une idée a été développée dans [76], qui permet de transférer la matrice de masse dans le second membre et de ne conserver dans le premier que la matrice de masse du schéma de Galerkin. Or, le *mass lumping* est connu pour donner des résultats satisfaisants avec les méthodes de Galerkin, donc nous y ferons appel pour le premier membre. C'est cette reformulation que nous allons voir maintenant.

Une altération possible du schéma

On repart de (4.3.8), en supposant (pour l'ordre élevé) que la condition (4.1.10) soit remplie à chaque étape l, que ce soit grâce à une limitation ou non. On peut alors simplifier l'écriture du schéma :

$$\forall l \leq k, \forall E \subset \Omega_h, \forall M_i \in E, \ \left(\Phi_i^E\right)^l = \left(\beta_i^E\right)^l \left(\Phi^E\right)^l = \int_E \left(\varphi_i + \left(\gamma_i^E\right)^l\right) r^l(U_h) dV$$

où nous définissons les opérateurs semi-discrets r^l ainsi :

$$r^l(U_h) = \frac{\delta {U_h}^l}{\delta t} + \sum_{s=0}^{l-1} \beta_{ls} \left(\overrightarrow{\nabla} \cdot \overrightarrow{F}({U_h}^s) + g({U_h}^s)\right) \tag{4.3.11}$$

en faisant appel à (4.3.9). En particulier, à la dernière itération du schéma de Runge-Kutta, on suppose qu'on obtient l'estimation d'erreur suivante :

$$r(U_h^{n+1}) := r^k(U_h) = \left(\frac{\partial U_h}{\partial t}\right)^n + \overrightarrow{\nabla} \cdot \overrightarrow{F}({U_h}^n) + g({U_h}^n) + O\left(\Delta t^{k_t}\right) = r(U_h^n) + O\left(\Delta t^{k_t}\right) \tag{4.3.12}$$

où k_t désigne l'ordre de précision du schéma. Si $k \leq 4$, alors on a $k_t = k$. Au-delà, il faut davantage d'étapes pour un même gain en précision. Cela revient à supposer que la solution obtenue à chaque pas de temps par un schéma de type Runge-Kutta vérifie l'hypothèse 4.1.1. Dans ces conditions, toute l'analyse qui a été faite dans la section 4.1.3 est applicable à notre schéma (on suppose en particulier que les matrices de masse soient données par (4.1.10)). Une démonstration est menée dans [76], mais qui reste assez partielle dans la mesure où certaines questions sont ignorées. Il subsiste une part d'ombre due notamment à l'absence de définition de l'erreur à laquelle se rapporte la précision en temps (pour une estimation a posteriori), qui doit en particulier tenir compte de l'accumulation des erreurs à chaque pas de temps (tout comme en espace, on estime la somme sur les éléments en $O(h^{-d})$...), et apparaître sous une forme compatible avec l'énoncé de l'hypothèse 4.1.1. Ceci étant souligné, nous allons continuer en nous basant sur l'idée que cette hypothèse est valable.

Revenons sur l'erreur établie dans la section 4.1.3 :

$$\mathcal{E}_\psi^{n+1} = \sum_{E \subset \Omega_h} \sum_{M_i \in E} \psi_i \int_E \left(\varphi_i + \gamma_i^k\right) r(U_h^{n+1}) dV$$

On peut isoler deux parties dans cette définition, l'erreur du schéma de Galerkin et celle due au décentrement :

$$\mathcal{E}_\psi^{n+1} = \left[\sum_{E \subset \Omega_h} \sum_{M_i \in E} \psi_i \int_E \varphi_i r(U_h^{n+1}) dV\right] + \left[\sum_{E \subset \Omega_h} \sum_{M_i \in E} \psi_i \int_E \gamma_i^k r(U_h^{n+1}) dV\right]$$

Ensuite, pour chacune, en considérant d'abord l'erreur en temps d'après l'hypothèse 4.1.1, puis l'erreur en espace, nous aboutissions à l'estimation terme à terme (4.1.13). Le regroupement de ces termes dans la logique précédente s'écrit :

$$\mathcal{E}_{\psi}^{n+1} = \left[\underbrace{O(h^{k_x+1})}_{I} + \underbrace{O(\Delta t^{k_t})}_{III}\right] + \left[\underbrace{O(h^{k_x+1})}_{II} + \underbrace{O(h\Delta t^{k_t})}_{IV}\right]$$

Or, dans toutes les méthodes explicites dont font partie celles de Runge-Kutta présentées ici, la contrainte de stabilité impose que le pas d'espace et le pas de temps soient du même ordre de grandeur, ce qui peut s'écrire $\Delta t = O(h)$. Par conséquent, on en déduit directement :

$$\mathcal{E}_{\psi}^{n+1} = \left[\underbrace{O(h^{k_x+1})}_{I} + \underbrace{O(h^{k_t})}_{III}\right] + \left[\underbrace{O(h^{k_x+1})}_{II} + \underbrace{O(h^{k_t+1})}_{IV}\right]$$

Ceci impose donc, pour avoir un schéma d'ordre $k_x + 1$ qui est l'ordre de l'erreur d'interpolation, que le schéma en temps soit précis à l'ordre $k_t = k_x + 1$, auquel cas nous avons :

$$\mathcal{E}_{\psi}^{n+1} = \left[\underbrace{O(h^{k_x+1})}_{I+III}\right] + \left[\underbrace{O(h^{k_x+1})}_{II} + \underbrace{O(h^{k_x+2})}_{IV}\right] = O(h^{k+1})$$

Mais surtout, ce qu'il est très intéressant pour nous de remarquer à ce stade, c'est que la précision que nous obtenons sur la partie décentrée du schéma et dont l'origine est le schéma en temps, autrement dit le terme IV, est plus grande que ce dont nous avons réellement besoin. C'est à partir de cette constatation que l'idée a été imaginée, dans [76], qu'il devait être possible de **modifier le schéma en temps uniquement sur la partie décentrée du schéma** $\mathcal{RD}$, tout en restant précis à l'ordre $k_x + 1$ globalement. On a une marge d'altération à cet endroit qui est d'un ordre de précision, pas plus. Pour l'exploiter, il faut redéfinir à chaque étape l le schéma à l'aide d'une expression du type :

$$\int_E \varphi_i r^l(U_h) dV + \int_E \gamma_i^l \widetilde{r}^l(U_h) dV \tag{4.3.13}$$

avec $\widetilde{r}^l$ donné par la l-ième itération d'un schéma à k étapes mais de degré $k_t' = k_t - 1 = k_x$. Trouver un tel schéma est plus aisé qu'un schéma Runge-Kutta traditionnel, moins de contraintes impliquant une plus grande variété de constructions possibles. Parmi les candidats potentiels, on souhaite ne retenir que ceux qui ont un décalage par rapport à la méthode de Runge-Kutta d'ordre classique : à chaque itération l, le schéma altéré ne doit pas faire intervenir U^l. C'est de cette manière seulement que nous pourrons faire passer les matrices de masse dans le second membre. De plus, à la manière de [76], pour simplifier l'expression finale du schéma (en factorisant la partie "évaluations des flux" dans les deux intégrandes de (4.3.13)), on conserve les coefficients β_{ls} du schéma d'origine, en ne modifiant donc que les α_{ls}, autrement dit l'incrément en temps. On sait que de tels schémas existent, qu'on peut voir comme des schémas "retardés" par rapport à ceux d'ordre classique. Certains ont été proposés dans [76], pour les ordres 2 et 3, et il est certainement relativement simple de reprendre le procédé aux ordres supérieurs, à condition que nous connaissions le schéma d'ordre optimal $(k_x + 1)$. L'expression générique de notre opérateur $\widetilde{r}^l$ est donc :

$$\widetilde{r}^l(U_h) = \widetilde{\frac{\delta U_h}{\delta t}}^{\,l} + \sum_{s=0}^{l-1} \beta_{ls} \left(\overrightarrow{\nabla} \cdot \overrightarrow{F}({U_h}^s) + g({U_h}^s)\right)$$

où le nouvel incrément en temps est une autre combinaison que (4.3.9), qui ne fait pas intervenir celle à calculer (U_h^l) :

$$\widetilde{\frac{\delta U_h}{\delta t}}^{\,l} = \frac{1}{\Delta t^n} \sum_{s=0}^{l-1} \widetilde{\alpha}_{ls} U_h^s$$

On peut ensuite bel et bien vérifier que comme annoncé, si on reprend l'analyse de précision avec le schéma (4.3.13), seul le terme IV de l'erreur est modifié :

$$\begin{aligned}\mathcal{E}_{\psi}^{n+1} &= \left[\underbrace{O(h^{k_x+1})}_{I}+\underbrace{O(\Delta t^{k_t})}_{III}\right]+\left[\underbrace{O(h^{k_x+1})}_{II}+\underbrace{O(h\Delta t^{k'_t})}_{IV}\right]\\ &= \underbrace{O(h^{k_x+1})}_{I+III}+\underbrace{O(h^{k_x+1})}_{II+IV}=O(h^{k_x+1})\end{aligned}$$

Résolution du système modifié

Il ne reste plus qu'à voir dans quelle mesure cette reformulation permet d'écarter l'inversion de la matrice de masse. Pour cela, il faut rappeler qu'une fois le nouveau schéma décrit, le système à résoudre devient :

$$\forall l\leq k, \forall M_i\in\Omega_h,\ \mathcal{P}_i^l=\sum_{E\subset\mathcal{T}_i}\left(\int_E\varphi_i r^l(U_h)dV+\int_E\gamma_i^l\widetilde{r}^l(U_h)dV\right)=0 \tag{4.3.14}$$

En réutilisant l'interprétation Petrov-Galerkin (4.3.13), quelques manipulations mettent en évidence l'intérêt de ne pas modifier les coefficients β_{ls}, car on peut alors obtenir la formulation suivante :

$$\begin{aligned}\forall M_i\in\Omega_h,\quad \mathcal{P}_i^l &= \sum_{E\subset\mathcal{T}_i}\left[\int_E\varphi_i\left(\frac{\delta U_h}{\delta t}^l+\sum_{s=0}^{l-1}\beta_{ls}\left(\overrightarrow{\nabla}\cdot\overrightarrow{F}(U_h^s)+g(U_h^s)\right)\right)dV\right.\\ &\quad\left.+\int_E\gamma_i^l\left(\widetilde{\frac{\delta U_h}{\delta t}}^l+\sum_{s=0}^{l-1}\beta_{ls}\left(\overrightarrow{\nabla}\cdot\overrightarrow{F}(U_h^s)+g(U_h^s)\right)\right)dV\right]\\ &= \sum_{E\subset\mathcal{T}_i}\left[\int_E\varphi_i\left(\frac{\delta U_h}{\delta t}^l-\widetilde{\frac{\delta U_h}{\delta t}}^l\right)dV+\int_E\varphi_i\widetilde{\frac{\delta U_h}{\delta t}}^l dV+\int_E\gamma_i^l\widetilde{\frac{\delta U_h}{\delta t}}^l dV\right.\\ &\quad\left.+\int_E\left(\varphi_i+\gamma_i^l\right)\sum_{s=0}^{l-1}\beta_{ls}\left(\overrightarrow{\nabla}\cdot\overrightarrow{F}(U_h^s)+g(U_h^s)\right)dV\right]\\ &= \int_{\mathcal{T}_i}\varphi_i\frac{\delta U_h}{\delta t}^l dV-\int_{\mathcal{T}_i}\varphi_i\widetilde{\frac{\delta U_h}{\delta t}}^l dV+\\ &\quad\sum_{E\subset\mathcal{T}_i}\int_E\left(\varphi_i+\gamma_i^l\right)\left(\widetilde{\frac{\delta U_h}{\delta t}}^l+\sum_{s=0}^{l-1}\beta_{ls}\left(\overrightarrow{\nabla}\cdot\overrightarrow{F}(U_h^s)+g(U_h^s)\right)\right)dV\end{aligned}$$

qui permet d'en déduire que le problème (4.3.14), si on isole dans le premier membre les seuls termes contenant l'inconnue U_h^l, se réécrit au final :

$$\forall l\leq k, \forall M_i\in\Omega_h,\ \int_{\mathcal{T}_i}\varphi_i\frac{\delta U_h}{\delta t}^l dV=\int_{\mathcal{T}_i}\varphi_i\widetilde{\frac{\delta U_h}{\delta t}}^l dV-\sum_{E\subset\mathcal{T}_i}\widetilde{\left(\Phi_i^E\right)}^l \tag{4.3.15}$$

où nous définissons les résidus partiels altérés :

$$\widetilde{\left(\Phi_i^E\right)}^l=\int_E\left(\varphi_i+\gamma_i^l\right)\widetilde{r}^l(U_h)dV=\beta_i^l\widetilde{\Phi^E}^l$$

avec :

$$\widetilde{\Phi^E}^l = \int_E \left(\widetilde{\frac{\delta U_h}{\delta t}}^l + \sum_{s=0}^{l-1} \beta_{ls} \left(\overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h^s) + g(U_h^s) \right) \right) dV = \int_E \widetilde{\frac{\delta U_h}{\delta t}}^l dV + \phi^{RK(l)} + \int_E \mathcal{G}^l(U_h) dV$$

Le membre de gauche dans (4.3.15) est le terme instationnaire du schéma de Galerkin, qui se traduit par l'apparition d'une matrice de masse m^G à inverser, de terme général :

$$\forall E \subset \Omega_h, \forall M_i, M_j \in E, m_{ij}^G = \int_E \varphi_i \varphi_j dV$$

Une pratique courante en Éléments Finis consiste à appliquer à ce terme un opérateur de *mass lumping* simple, en condensant les termes extra-diagonaux sur la diagonale. Formellement, en accord avec ce que nous disions dans la section 4.1.3 en commentant [76], pour des problèmes non linéaires et/ou sur des éléments également non linéaires, ce type de *mass lumping* détériore la précision du schéma. Toutefois, l'expérience montre généralement des résultats suffisamment satisfaisants pour que cette technique soit conservée, d'autant plus si l'on ne dispose pas d'autre alternative (qui ne nécessite pas d'inverser de matrice). Le résultat de l'agrégation des termes sur la diagonale donne une matrice de terme général :

$$\widetilde{m}_{ij}^G = \delta_{ij} \int_E \varphi_i dV$$

Le système à résoudre associé à chaque degré de liberté (4.3.15) devient alors :

$$\sum_{E \subset \mathcal{T}_i} \sum_{M_j \in E} \widetilde{m}_{ij}^G \frac{\delta U_j}{\delta t}^l = \sum_{E \subset \mathcal{T}_i} \int_E \varphi_i dV \frac{\delta U_i}{\delta t}^l = \int_{\mathcal{T}_i} \varphi_i dV \frac{\delta U_i}{\delta t}^l = \sum_{E \subset \mathcal{T}_i} \sum_{M_j \in E} m_{ij}^G \widetilde{\frac{\delta U_j}{\delta t}}^l - \sum_{E \subset \mathcal{T}_i} \widetilde{\left(\Phi_i^E\right)}^l \quad (4.3.16)$$

et il est facile d'en déduire chaque valeur U_i^l, puisqu'il s'agit, pour chacune des p composantes de U, d'un problème scalaire découplé de toutes les autres inconnues U_j^l du domaine. Dans [76], il est également proposé, mais sans qu'il soit prétendu que cela soit une meilleure solution ou non (les résultats obtenus sont très proches), d'appliquer le *mass lumping* également à la matrice de masse de Galerkin qui apparait au second membre. Cette modification est désignée sous le nom de *Global lumping* dans la référence précitée, l'autre étant dénommée *Selective lumping*. Une alternative consistante pour évaluer m_{ij}^G serait d'approcher cette intégrale de volume par une formule de quadrature s'appuyant sur les degrés de liberté. De cette manière, on constate rapidement que la matrice de masse obtenue est diagonale (on utilise une propriété de l'interpolation de Lagrange), avec une erreur commise qui est de l'ordre de l'erreur de troncature du schéma. Pour tous les éléments autres que P_1, l'agrégation sur la diagonale n'est en fait plus suffisante et la diagonalisation de la matrice par une quadrature bien choisie devient la seule technique théoriquement admissible. Celle-ci est est d'ailleurs largement utilisée dans le domaine de la résolution des équations d'onde par la méthode de Galerkin (voir par exemple [29], [57], ou [28] pour une présentation détaillée).

4.3.3 Un traitement particulier pour le *divergence cleaning* ?

Interprétation du comportement de la correction hyperbolique (ou mixte)

Dans toute cette section sur les méthodes explicites, nous avons tacitement considéré que la correction de la divergence était incluse sous sa forme normale, c'est-à-dire une équation d'évolution hyperbolique sur le multiplicateur ψ (et éventuellement le terme de source linéaire parabolique au second membre). C'est la forme sous laquelle nous l'utilisons le plus fréquemment. Cependant, cela signifie que les erreurs de divergence sont transportées à la vitesse finie c_h. De plus, celle-ci est généralement prise inférieure

à la vitesse d'onde des ondes magnétosoniques rapides pour ne pas influer sur le critère de stabilité CFL (ni sur la définition du paramètre α du schéma de Lax-Friedrichs). Ces erreurs mettent donc un certain temps, généralement non négligeable au regard de la durée de la simulation, à sortir du domaine, à supposer que les conditions limites le permettent. Si les conditions aux bords sont périodiques ou constituées partout de parois, ce qui modéliserait une boîte, ou une chambre de confinement comme un tokamak, alors les erreurs de divergence ne sortiront pas. En revanche, elles peuvent être amorties (par la diffusion numérique propre au schéma ou grâce au terme de source parabolique), de manière à ce que ψ s'uniformise dans le domaine, ce qui traduit dans ce cas une disparition des erreurs sur la divergence pour une certaine définition numérique de celle-ci ! (Ou pour le dire différemment, les erreurs restantes sont rendues transparentes pour la méthode numérique : ceci est le but à atteindre de tout schéma, même hors du contexte $\mathcal{RD}$, car on ne peut pas imposer numériquement une divergence nulle quelle que soit sa définition discrète !) Pour s'en rendre compte, il faut remarquer que le gradient de ψ dans l'équation d'évolution du champ magnétique disparait. Pourtant, si les conditions aux limites isolent l'écoulement de l'extérieur, les erreurs sont toujours présentes puisqu'elles ne sont pas évacuées hors du domaine. Il semble donc que l'uniformisation de ψ résulte en un champ de vecteurs $\overrightarrow{B}$ à divergence nulle (pour une certaine définition numérique de cette divergence), qui diffère légèrement de celle qu'on aurait obtenue sans générer d'erreur. Lorsqu'il est présent, on peut se représenter ce comportement d'amortissement comme une reprojection de l'erreur (comme par la méthode de projection) progressive et délocalisée : les erreurs voyagent dans le domaine au fur et à mesure qu'elles se reprojettent sur une configuration à divergence nulle.

Or en implicite, on profite du fait que la résolution est itérative, à chaque pas de temps, pour annuler la dérivée en temps sur ψ et considérer que la correction des erreurs de divergence générées à chaque résolution temporelle est un problème stationnaire. Si la convergence est atteinte (ce qui est un autre problème), ceci rend compte d'une évacuation ou d'une uniformisation (i.e. un état d'équilibre) de ψ atteinte à chaque pas de temps. Il parait donc légitime de vouloir imiter la résolution stationnaire de l'implicite (obtenir une correction globale aboutie à chaque pas de temps), ou de vouloir s'en approcher. Si toutes choses étaient comparables par ailleurs, cette voie pourrait sembler meilleure que la résolution instationnaire classique. Il faut néanmoins relativiser ce sujet. Il est en fait tout à fait possible, voire probable, que cela n'améliore surtout que les résultats intermédiaires, et que si seule la solution finale nous intéresse, celle-ci obtenue par la voie explicite instationnaire classique soit tout à fait acceptable (à divergence discrète quasi nulle).

Pour résumer, la question que l'on peut se poser est : est-ce qu'une correction précise de la divergence à chaque pas de temps améliore la qualité de la solution **finale** par rapport à une résolution progressive (instationnaire) classique ? Le mot le plus important est bien "finale", car si on souhaite connaitre précisément diverses étapes intermédiaires, il est clairement préférable que la correction soit "propre" aux dates concernées. La réponse à la question précédente est : certainement, mais peut-être peu (on sous-entend que quoi qu'il arrive, cela ne peut pas dégrader la qualité de la solution finale). Ce qui amène naturellement à une autre question : dans quelle mesure y a-t-il amélioration, et à quel prix ? Cette fois-ci, la réponse est beaucoup moins évidente. Il se peut en effet qu'assez d'efforts soient déployés pour peu de gains sur la qualité de la solution. Ne sachant pas répondre plus précisément, nous avons voulu chercher, puis tester, des exemples de méthodes permettant de mettre en oeuvre la résolution stationnaire de la divergence de $\overrightarrow{B}$ en explicite.

Méthodes de découplage imaginées

Nous avons pensé qu'il fallait passer par une implicitation d'une partie des équations, relative au champ magnétique et à la correction de sa divergence. À partir de là, nous avons imaginé deux versions

(notées très originalement v_1 et v_2). Le point commun à celles-ci est que les $d+2$ premières équations, celles correspondant donc à $(\rho, \rho\overrightarrow{u}, E)$, sont résolues normalement par la méthode explicite. Ce sont les $d+1$ dernières, que nous allons rappeler ici, qui vont faire l'objet d'un traitement spécial :

$$\begin{aligned} \frac{\partial \overrightarrow{B}}{\partial t} + \overrightarrow{\nabla} \cdot \left(\overrightarrow{B}\overrightarrow{u}^t - \overrightarrow{u}\overrightarrow{B}^t \right) + \overrightarrow{\nabla}\psi &= \overrightarrow{0} \\ \frac{\partial \psi}{\partial t} + {c_h}^2 \overrightarrow{\nabla} \cdot \overrightarrow{B} &= -\frac{{c_h}^2}{{c_p}^2}\psi \end{aligned} \tag{4.3.17}$$

Comme nous l'avons souligné dans la section concernant l'implicite, et l'avons fait sans plus de justification dans cette section, nous commençons toujours par discrétiser en temps. C'est justement dans cette étape que les changements interviennent. Pour simplifier, nous allons omettre les étapes du schéma Runge-Kutta et présenter ces méthodes avec un schéma explicite quelconque du type :

$$\frac{\delta {U_h}^n}{\delta t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h^n) + g(U_h^n) = 0$$

Avec ce modèle, la première méthode (v_1) consiste à n'impliciter strictement que la partie *divergence cleaning* des équations et à résoudre un problème stationnaire sur la correction. Comme cette partie est linéaire, cela se fait en une itération. Pour être plus précis, on commence par résoudre les équations non corrigées, ce qui nous donne une solution $\left(\rho_h^{n+1}, (\rho\overrightarrow{u})_h^{n+1}, E_h^{n+1}, \overrightarrow{B}^*_h\right)$. Ensuite, on souhaite résoudre comme un problème stationnaire la partie *divergence cleaning*. Étant linéaire, cette partie se résoud simplement à l'aide d'un schéma de Galerkin en espace et s'écrit :

$$\forall M_i \in \Omega_h, \begin{cases} \displaystyle\int_{\mathcal{T}_i} \varphi_i \frac{\overrightarrow{B}_h^{n+1} - \overrightarrow{B}^*_h}{\Delta t^n} dV + \int_{\mathcal{T}_i} \varphi_i \overrightarrow{\nabla} \psi_h^{n+1} dV = \overrightarrow{0} \\ \displaystyle\int_{\mathcal{T}_i} \varphi_i \overrightarrow{\nabla} \cdot \overrightarrow{B}_h^{n+1} dV = 0 \end{cases}$$

et ce en une itération de la méthode Newton (que nous réutilisons dans le logiciel par commodité, même si le problème est linéaire). Nous ne détaillons pas davantage la construction de cette méthode, car on peut se rendre compte au final que cela revient à une méthode de projection, telle que celle utilisée par Brackbill et Barnes ([18]). Même si sa mise en oeuvre est un peu différente formellement, le principe est similaire.

La seconde méthode (v_2) s'appuie sur la connaissance de $\overrightarrow{u}^{n+1}$ pour s'attacher à résoudre en implicite un problème stationnaire en $\left(\overrightarrow{B}, \psi\right)$. Celui-ci s'écrit comme la limite de :

$$\begin{aligned} \frac{\partial \overrightarrow{B}_h}{\partial \tau} + \frac{\partial \overrightarrow{B}_h}{\partial t} + \overrightarrow{\nabla} \cdot \left(\overrightarrow{B}_h (\overrightarrow{u}_h^{n+1})^t - \overrightarrow{u}_h^{n+1} \overrightarrow{B}_h^t \right) + \overrightarrow{\nabla}\psi_h &= \overrightarrow{0} \\ \frac{\partial \psi_h}{\partial \tau} + {c_h}^2 \overrightarrow{\nabla} \cdot \overrightarrow{B}_h &= -\frac{{c_h}^2}{{c_p}^2}\psi_h \end{aligned}$$

lorsque $\dfrac{\partial}{\partial \tau} \to 0$. Cette méthode est une solution beaucoup trop forte, en ce sens qu'elle implique une hausse significative du temps de calcul pour résoudre un problème qui peut se résoudre par une méthode de projection v_1, même si elle est plus élégante (elle adapte toute l'équation de Faraday à la correction en temps réel). Elle nécessite en effet la résolution d'un problème non linéaire cette fois, même si $\overrightarrow{u}$ est connue, toujours avec la méthode de Newton-Raphson. Nous l'avons donc abandonnée au profit de la méthode de projection lorsque nous voulions un traitement particulier des erreurs de divergence.

En réalité, c'est justement tout l'intérêt de la correction hyperbolique que de réduire le coût de calcul d'une simulation MHD, par rapport à une méthode de projection, tout en essayant de faire en sorte que

les erreurs de divergence ne lui soient pas fatales. Dans ce but, on sacrifie la "propreté" de la solution, critère qu'on remplace par un transport contrôlé et amorti des erreurs de divergence. Par conséquent, ce qu'il faut déterminer est dans quel cas ceci est acceptable et ne nuit pas trop à la qualité de la solution, et cela dépend sans doute en grande partie des attentes ou des besoins des utilisateurs, en fonction des problèmes abordés.

4.4 Discrétisation des termes diffusifs des équations de la MHD résistive

Il ne nous reste plus qu'à discrétiser la partie diffusive des équations de la MHD dite résistive (2.1.12). Les phénomènes irréversibles que ces termes supplémentaires modélisent sont à l'origine de l'apparition de nombreuses instabilités, que nous souhaiterions parvenir à simuler.

4.4.1 Rappel des équations adimensionnées

Nous avons vu au chapitre 2 que les équations de la MHD résistive pouvaient écrites de manière synthétique :

$$\frac{\partial U}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U) - \overrightarrow{\nabla} \cdot \overrightarrow{F_D}(U) = -g(U) \tag{4.4.1}$$

où les flux $\overrightarrow{F}$ sont ceux de la MHD idéale traités jusqu'ici, et $\overrightarrow{F_D}$ les flux diffusifs que nous rajoutons seulement maintenant. Le terme de source g mis au second membre est soit nul, soit linéaire et égal à la partie parabolique de la correction mixte. Pour rappel, les flux diffusifs s'écrivent, dans la forme adimensionée donnée par le système (2.4.1) :

$$\overrightarrow{F_D}^t(U) = \begin{pmatrix} \frac{D}{Pe_\rho}\overrightarrow{\nabla}\rho \\ \\ \frac{\nu}{Re}\left(\overrightarrow{\nabla}\overrightarrow{u} + (\overrightarrow{\nabla}\overrightarrow{u})^t - \frac{2}{3}\overrightarrow{\nabla}\cdot\overrightarrow{u}\boldsymbol{I}\right) \\ \\ \frac{1}{Re}\boldsymbol{\tau}\overrightarrow{u} + \frac{\kappa}{Pe}\overrightarrow{\nabla}T + \frac{\eta}{Rm}\left(\overrightarrow{\nabla}\left(\frac{\overrightarrow{B}^2}{2}\right) - \left(\overrightarrow{B}\cdot\overrightarrow{\nabla}\right)\overrightarrow{B}\right) \\ \\ \frac{1}{Rm}\left(\eta\overrightarrow{\nabla}\overrightarrow{B} + (\overrightarrow{\nabla}\eta)\overrightarrow{B}^t - \overrightarrow{B}\cdot(\overrightarrow{\nabla}\eta)\boldsymbol{I}\right) \\ \\ 0 \end{pmatrix}$$

où les nombres sans dimension Pe_ρ, Re, Pe et Rm sont constants et peuvent être sortis de la divergence, $\boldsymbol{I}$ désigne la matrice identité de taille d, et où :

$$\boldsymbol{\tau}\overrightarrow{u} = \nu\left(\overrightarrow{\nabla}\cdot\left(\overrightarrow{u}\overrightarrow{u}^t\right) - \frac{2}{3}\left(\overrightarrow{\nabla}\cdot\overrightarrow{u}\right)\overrightarrow{u}\right)$$

L'écriture ci-dessus est valable pour des paramètres diffusifs (la diffusivité D, la viscosité ν, la conductivité thermique κ et la résistivité électrique η) variables. Néanmoins, nous les avons supposés constants dans nos travaux, ce qui suffit dans un premier temps pour générer les instabilités MHD qui nous intéressent. Dans [53], une résistivité constante par harmonique de Fourier est utilisée, les autres paramètres étant constants. Une approche semblable est utilisée dans [71]. C'est, bien entendu, un point qui pourrait être

appelé à être revu au fur et à mesure de l'évolution du code de calcul, en repartant de la forme continue que nous avons présentée. Avant de poursuivre, notons que le fait de supposer la résistivité constante nous permet de simplifier notre expression de $\overrightarrow{F_D}^t(U)$ sur le champ magnétique :

$$\left(\overrightarrow{F_D}^t\right)_{\overrightarrow{B}} = \frac{\eta}{Rm}\overrightarrow{\nabla}\overrightarrow{B}$$

Dès lors que les gradients de la résistivité n'apparaissent plus, il est plus commode de réécrire ces flux sous une forme quasi-linéaire pour l'usage que nous en faisons. On utilisera donc :

$$\forall 1 \leq k \leq d,\ \ (F_D)_k = \sum_{l=1}^{d} A_{kl}\frac{\partial U}{\partial x_l} \tag{4.4.2}$$

où les "matrices des flux" A_{kl} dépendent bien sûr elles-mêmes de la solution U. Leur détail est donné dans l'annexe D.

4.4.2 Discrétisation spatiale : la méthode de Galerkin

Nous avons vu que pour obtenir un schéma $\mathcal{RD}$ d'ordre élevé dans le cas général, celui-ci doit être formulé en espace de la manière suivante :

$$\forall M_i \in \Omega_h, \forall E \subset \Omega_h,\ \Phi_i^E(U_h) = \beta_i \Phi^E(U_h) = \beta_i \int_E r(U_h) dV$$

avec des matrices de ditribution β_i bornées. Cela reste vrai lorsqu'on étend le problème aux équations complètes. La seule différence est que le système à annuler n'est plus celui de la MHD idéale, mais (4.4.1). L'intégrale porte donc sur (en omettant le terme de source g déjà traité plus haut) :

$$r(U_h) = \frac{\partial U_h}{\partial t} + \overrightarrow{\nabla}\cdot\overrightarrow{F}(U_h) - \overrightarrow{\nabla}\cdot\overrightarrow{F_D}(U_h) \tag{4.4.3}$$

Dans le formalisme de Petrov-Galerkin, en conservant les notations vues précédemment, un schéma $\mathcal{LP}$ s'écrit naturellement :

$$\forall M_i \in \Omega_h, \int_{\Omega_h} (\varphi_i + \gamma_i)\, r(U_h) dV = \sum_{E \subset \mathcal{T}_i} \beta_i^E \left(\int_E \frac{\partial U_h}{\partial t} dV + \phi^E - \phi^{E,D} \right) = 0 \tag{4.4.4}$$

avec (4.4.3) et :

$$\phi^{E,D} = \int_E \overrightarrow{\nabla}\cdot\overrightarrow{F_D}(U_h) dV = \int_{\partial E} \overrightarrow{F_D}(U_h)\cdot\overrightarrow{n}\, d\partial E$$

Cette fluctuation diffusive fait intervenir des flux qui s'expriment en fonction des gradients de la solution U_h, et donc des fonctions de base φ_j. Ces gradients sont intégrés, comme on le déduit de l'expression précédente, sur les arêtes ou des faces des éléments considérés. Or ils ne sont pas continus d'un élément à un autre ! Si le calcul de $\phi^{E,D}$, pour un élément E donné, est rendu possible par le fait que les gradients à employer sur ∂E sont les restrictions de ceux définis dans E, cela pose malgré tout un problème sérieux de conservation. Pour être peut-être plus clair, prenons l'exemple simple de l'équation de continuité à intégrer sur deux éléments E_1 et E_2 qui possèdent un bord commun Γ_{12}. La partie advective donne pour chacune des deux fluctuations ($k = 1, 2$) :

$$\left(\phi^{E_k}\right)_\rho = \int_{E_k} \overrightarrow{\nabla}\cdot(\rho\overrightarrow{u})\, dV = \sum_{\Gamma \subset \partial E_k} \sum_{M_i \in \Gamma} (\rho\overrightarrow{u})_i \cdot \int_\Gamma \varphi_i \overrightarrow{n}\, d\Gamma$$

Si on s'intéresse à la conservation du schéma en sommant les résidus de chaque élément, comme le maillage est conforme, on voit bien qu'en ce qui concerne Γ_{12}, on va obtenir la contribution advective suivante :

$$C_{12} = \sum_{M_i \in \Gamma_{12}} (\rho\overrightarrow{u})_i \cdot \left(\int_{\Gamma_{12}} \varphi_i^{E_1}|_{\Gamma_{12}} \overrightarrow{n}_{1\to 2} d\Gamma_{12} + \int_{\Gamma_{12}} \varphi_i^{E_2}|_{\Gamma_{12}} \overrightarrow{n}_{2\to 1} d\Gamma_{12} \right)$$

Les polynômes de Lagrange sont définis par élément mais continus au passage d'un élément à un autre, donc :

$$\varphi_i^{E_1}|_{\Gamma_{12}} = \varphi_i^{E_2}|_{\Gamma_{12}}$$

Ce qui entraîne que la contribution C_{12} est nulle puisque la somme des normales $\overrightarrow{n}_{1\to 2}$ et $\overrightarrow{n}_{2\to 1}$ est nulle. Répétons à présent l'exercice pour un terme diffusif, comme celui de diffusion de matière qui se trouve dans la même équation. Les flucutations diffusives pour $k = 1, 2$ s'écrivent :

$$\left(\phi^{E_k,D}\right)_\rho = \frac{1}{Pe_\rho}\int_{E_k} \overrightarrow{\nabla}\cdot\left(D\overrightarrow{\nabla}\rho\right)dV = \frac{D}{Pe_\rho}\sum_{\Gamma\subset\partial E_k}\sum_{M_i\in\Gamma}\rho_i\int_\Gamma \overrightarrow{\nabla}\varphi_i\cdot\overrightarrow{n}d\Gamma$$

puisque D est une constante. De la même manière qu'au-dessus, l'étude de la conservation se basant sur la sommation des résidus, la contribution diffusive du bord Γ_{12} est :

$$C_{12}^D = \frac{D}{Pe_\rho}\sum_{M_i\in\Gamma_{12}}\rho_i\left(\int_{\Gamma_{12}}\overrightarrow{\nabla}\varphi_i^{E_1}|_{\Gamma_{12}}\overrightarrow{n}_{1\to 2}d\Gamma_{12} + \int_{\Gamma_{12}}\overrightarrow{\nabla}\varphi_i^{E_2}|_{\Gamma_{12}}\overrightarrow{n}_{2\to 1}d\Gamma_{12}\right)$$

Les normales sont toujours opposées, mais si les polynômes de Lagrange sont, par construction, continus aux interfaces Γ entre les éléments, ils n'y sont pas de classe $\mathcal{C}^1$:

$$\overrightarrow{\nabla}\varphi_i^{E_1}|_{\Gamma_{12}} \neq \overrightarrow{\nabla}\varphi_i^{E_2}|_{\Gamma_{12}}$$

et par conséquent, la contribution C_{12}^D n'est pas nulle. Le schéma (4.4.4) n'est donc pas conservatif, même si cela ne vient pas de la définition du schéma $\mathcal{RD}$ (i.e., dans le cas d'un schéma $\mathcal{LP}$, des matrices β_i dont la somme est bien égale à la matrice identité), mais de l'utilisation qui est faite de l'interpolation sur laquelle se base celui-ci. On peut en effet remarquer au passage que le même problème se pose avec le schéma de Galerkin (supposer $\gamma_i = 0$), méthode pour laquelle il est généralement résolu en intégrant par parties les termes diffusifs dans la formulation variationnelle (voir plus bas).

La présence des termes en dérivées secondes d'espace rend en outre l'obtention de l'ordre élevé plus délicate. En effet, à chaque dérivation, l'interpolation de la solution perd en efficacité comme nous avons pu le constater ne serait-ce qu'en dérivant une fois les flux : si on revient sur l'analyse de précision des problèmes instationnaires, on voit que le terme de divergence des flux est le seul en $O(h^{k+1})$ dans II, les autres étant en $O(h^{k+2})$. Ici, il faut donc s'attendre à la perte d'un ordre de précision supplémentaire. Ceci se traduit de façon particulièrement singulière sur les éléments P_1 et si les flux diffusifs sont linéaires en U, puisque la fluctuation diffusive est alors toujours nulle (il suffit de constater que pour le cas de l'équation d'advection-diffusion, $\Delta u_h = 0$). Ceci implique qu'avec une construction classique de schéma $\mathcal{LP}$, comme nous en avons vues auparavant, les termes linéaires de la partie diffusive des équations ne sont tout simplement pas résolus sur les éléments P_1 ! Pour notre système MHD, cela concerne par exemple le terme de diffusion de matière que nous avons vu plus haut.

Pour toutes ces raisons, depuis quelques années, il a fallu trouver une alternative pour traiter les problèmes diffusifs tels que les équations de Navier-Stokes ([90], [59], [97]). Or, dans le cas d'éléments P_1 et si les flux diffusifs sont linéaires en U (cas de l'équation d'advection-diffusion en scalaire), il est montré dans [90] que la distribution $\mathcal{RD}$ est équivalente, sur les termes diffusifs, à une discrétisation par le schéma de Galerkin. Cette démonstration rapide passe par une réécriture de chaque fonction de décentrement γ_i sous la forme d'une fonction bulle (dont la restriction aux bords de chaque élément est nulle), technique qui a été aussi reprise de façon plus pousée dans [97]. Au final, cela exprime une consistance entre les deux approches, qui est investiguée plus avant dans [59] notamment. Cela justifie que jusqu'à très récemment, et c'est le cas dans nos travaux également, il y ait eu systématiquement recours à un schéma de Galerkin pour discrétiser les termes diffusifs des équations.

La formulation variationnelle dans laquelle s'inscrit notre schéma est donc, concernant la partie diffusive, celle de Galerkin avec intégration par parties :

$$\forall M_i \in \Omega_h, \int_{\Omega_h} (\varphi_i + \gamma_i) \left(\frac{\partial U_h}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h) \right) dV + \int_{\Omega_h} \overrightarrow{\nabla} \varphi_i \cdot \overrightarrow{F_D}(U_h) dV - \int_{\partial \Omega_h} \varphi_i \overrightarrow{F_D} \cdot \overrightarrow{n} d\partial\Omega_h = 0 \tag{4.4.5}$$

On définit donc les parties diffusives des résidus partiels en conséquence, sur chaque élément E :

$$\phi_i^{E,D} = \int_E \overrightarrow{\nabla} \varphi_i \cdot \overrightarrow{F_D} dV \tag{4.4.6}$$

qu'on assemble ensuite avec la partie MHD idéale sur chaque élément contenant M_i pour résoudre (4.4.5), qui se réécrit pour chaque degré de liberté n'appartenant au bord du domaine :

$$\forall M_i \in \Omega_h, \sum_{E \subset \mathcal{T}_i} \left[\beta_i^E \left(\int_E \frac{\partial U_h}{\partial t} dV + \phi^E(U_h) \right) + \phi_i^{E,D} \right] = 0$$

Pour les degrés de liberté situés sur $\partial\Omega_h$, après avoir parcouru les éléments en assemblant les termes ci-dessus, il faut rajouter les termes de bord issus de l'intégration par parties dans (4.4.5). Pour calculer les contributions diffusives $\phi_i^{E,D}$, étant donné que les flux comportent là encore une majorité de termes non linéaires en U, et que ceux-ci ne sont pas polynomiaux, il est nécessaire d'employer une formule de quadrature (ou une interpolation des flux). La question de la précision nécessaire ou utile de cette formule trouve ici des réponses semblables à celles de la partie advective : une précision considérant des flux linéaires suffit théoriquement, mais il peut s'avérer préférable de consentir davantage d'efforts, surtout pour les ordres très élevés. Supposer les flux linéaires en U ne signifie pas pour autant que tout l'intégrande soit de degré k (le degré d'interpolation), car il faut prendre en compte le degré de $\overrightarrow{\nabla}\varphi_i$ qui s'y ajoute. Quoi qu'il en soit, lorsque nous employons une formule de quadrature à N_q points pour calculer (4.4.6), nous nous servons de la forme quasi-linéaire (4.4.2) des flux pour obtenir :

$$\begin{aligned} \phi_i^{E,D} &= \sum_{M_j \in E} \sum_{k,l=1}^{d} \left(\int_E \partial_{x_k} \varphi_i A_{kl}(U_h) \partial_{x_l} \varphi_j dV \right) U_j \\ &\approx |T| \sum_{M_j \in E} \sum_{k,l=1}^{d} \sum_{q=1}^{N_q} \omega_q \left(\partial_{x_k} \varphi_i(\overrightarrow{x}_q) A_{kl}(U_q) \partial_{x_l} \varphi_j(\overrightarrow{x}_q) \right) U_j \end{aligned}$$

Pour fixer les idées, nous pouvons illustrer la démarche en 2D, sur le cas simple des triangles P_1. La construction est quasiement identique pour les tétraèdres P_1 en 3D (il suffit d'étendre la définition des normales et de modifier l'expression des gradients des fonctions de base de façon adéquate). Les gradients des fonctions de base P_1 sont constants, ce qui fait que l'intégrande qui apparaît dans le calcul de $\phi_i^{E,D}$ ci-dessus est de degré égal à celui des matrices $A_{kl}(U_h)$. Dans cette configuration, supposer les flux diffusifs linéaires revient à faire la même hypothèse sur les matrices matrices A_{kl}, et c'est ce que nous allons faire dans cet exemple. Par conséquent, les signaux diffusifs sont évalués par une formule de quadrature exacte pour les fonctions linéaires, c'est-à-dire une simple moyenne arithmétique sur les valeurs aux sommets (c'est l'extension 2D de la formule des trapèzes en 1D). En reprenant la définition des normales de la

figure 3.5, et en particulier (3.1.16), on obtient :

$$\begin{aligned}
\forall M_i \in \Omega_h, \forall T \subset \mathcal{T}_i, \quad \phi_i^{T,D} &= \frac{|T|}{3} \sum_{M_q \in T} \sum_{k=1}^{d} \frac{(n_i)_k}{2|T|} \left(F_D\right)_k (U_q) \\
&= \frac{1}{6} \sum_{M_q \in T} \sum_{k=1}^{d} (n_i)_k \left[\sum_{l=1}^{d} A_{kl}(U_q) \left(\sum_{M_j \in T} \frac{(n_j)_l}{2|T|} U_j \right) \right] \\
&= \frac{1}{12|T|} \sum_{k=1}^{d} (n_i)_k \sum_{l=1}^{d} \left[\left(\sum_{M_q \in T} A_{kl}(U_q) \right) \left(\sum_{M_j \in T} (n_j)_l U_j \right) \right]
\end{aligned}$$

Dans cette configuration simple, et grâce à l'interpolation linéaire, on peut trouver la dernière expression qui est la mieux factorisée et donc la plus économe en calculs. Le cas général ne permet pas de manipulation semblable.

Enfin, quoi qu'il soit fait en espace, les discrétisations temporelles mises en place dans les sections précédentes pour la MHD idéale restent valables pour la MHD complète. L'extension est triviale, les termes diffusifs étant traités exactement comme la divergence des flux advectifs et les termes de source.

4.4.3 Défauts de cette approche et alternatives

Il y a quelques remarques à faire sur la méthode que nous venons de présenter. Tout d'abord, on peut constater rapidement en consultant l'annexe D que les matrices A_{kl} sont très creuses. Leur intérêt réside dans l'exhibition du gradient de U_h par une expression quasi-linéaire des $\phi_i^{E,D}$, d'une façon claire et concise. Néanmoins, dans la pratique, une optimisation assez conséquente peut être réalisée en ne calculant que les termes utiles à chaque équation, c'est-à-dire en évitant le produit matrice-vecteur sur le système entier. Il en résulte un algorithme plus long et moins lisible, mais plus performant.

Plus important maintenant, nous devons nous poser la question de la précision réelle que procure notre méthode. Le fait est que si nous conduisons l'analyse de précision de la manière habituelle, que nous avons vue dans les chapitres 3 et 4, l'ordre élevé n'est démontré que lorsqu'on dispose d'une formulation $\mathcal{LP}$, qui distribue tous les termes des équations (y compris diffusifs donc) de la même façon. Ce n'est pas le cas dans notre méthode de Galerkin. On ne peut donc pas garantir une précision optimale. Ceci s'observe dans la pratique, comme il a été noté dans [59] : l'ordre optimal (celui d'un schéma $\mathcal{LP}$) est atteint lorsque l'écoulement est proche de la limite advective ou de la limite diffusive. Dans le premier cas, la précision est déterminée par le schéma $\mathcal{RD}$ advectif seulement, et dans le second uniquement par la méthode de Galerkin sur les termes diffusifs. Le problème concerne les régimes intermédiaires, où le couplage entre deux schémas différents détériore la précision. En mécanique des fluides, avec une conduction thermique à nombre de Prandtl constant, on dit qu'on observe une précision variable en fonction du nombre de Reynolds, optimale lorsque $Re << 1$ ou $Re \to \infty$ (on utilise en réalité un Reynods de maille, où L_0 est remplacé par h, car la taille des éléments joue un rôle important dans la capture des phénomènes visqueux), et dégradée dans les plages intermédiaires.

Partant de ce constant, il a depuis été convenu d'abandonner cette méthode. Des travaux en cours s'attachent, dans le cadre de la mécanique des fluides, à formuler des schémas $\mathcal{RD}$ pour tous les termes des équations, autrement dit en revenant à forme $\mathcal{LP}$ (4.4.4). Nous avons vu que les défauts de cette approche étaient la précision sur les termes diffusifs (l'interpolation étant dérivée deux fois) et la conservation qui est pourtant cruciale pour toute simulation numérique. Deux alternatives ont émergé pour améliorer les choses. La première est l'idée de reconstruire les gradients de la solution, en faisant appel aux données des éléments voisins, d'une façon continue aux interfaces et qui permette de gagner un ordre de précision sur

le terme diffusif (voir par exemple [22]). La seconde, plus récente, est celle du *First Order System* (FOS) développé par Nishikawa ([67], [68]) dont les travaux ont été motivés par les mêmes remarques que les précédentes, comme en témoigne [69]. L'idée est de traiter les gradients de la solution comme des variables à part entière, régies par des équations d'évolution hyperboliques. Ce faisant, les gradients sont interpolés avec la même précision que n'importe quelles autres variables (ce qui permet de recouvrer l'ordre élevé comme le ferait une reconstruction), et sont continus aux interfaces entre éléments comme le sont les fonctions d'interpolation (il s'agirait de polynômes de Lagrange dans notre contexte). En réalité, ceci est formulé dans un contexte stationnaire et on cherche donc la limite lorsque la dérivée en temps s'annule : il s'agit d'une méthode de relaxation sur les gradients, comme il en existe par exemple pour le calcul de la pression. En instationnaire, il faut donc s'attendre à devoir résoudre un problème stationnaire à chaque itération en temps physique. Néanmoins, comme ces équations d'évolution sont linéaires, la résolution de ces seules équations est atteinte en une seule itération de la méthode de Newton. Autrement dit, en instationnaire explicite, cela se traduirait par une inversion d'un système linéaire à chaque pas de temps, exactement comme la méthode de projection qui corrige la divergence du champ magnétique. Il y a d'ailleurs un parallèle manifeste entre les deux méthodes, la relaxation pouvant être vue comme une "correction" des gradients. On pourrait alors pousser la comparaison plus loin et essayer de transposer le *divergence cleaning* hyperbolique à la méthode de relaxation, c'est-à-dire en fin de compte simplement sacrifier la qualité des gradients calculés, en transportant les "erreurs" par rapport au gradient optimal au moyen d'une dérivée en temps physique non nulle sur les gradients. La réduction du temps de calcul, pour des méthodes explicites, serait sans doute importante. Cependant, on peut légitimement douter du fait que l'ordre élevé serait globalement préservé... Il faut noter que, comme le soulignent les auteurs dans [69], la méthode de relaxation est une solution onéreuse et potentiellement trop forte. Clairement, l'adoption de cette méthode multiplie par $d+1$ le coût en mémoire, d étant la dimension du problème, ce qui est beaucoup trop pour espérer simuler des problèmes 3D complexes à plusieurs millions de degrés de liberté de façon performante, sur les machines actuelles.

Pour résumer, la reconstruction des gradients et l'approche FOS sont deux alternatives qui permettent de calculer des gradients de U avec un ordre de précision supplémentaire et qui soient continus aux bords des éléments, ce qui garantit l'ordre élevé global d'un schéma $\mathcal{LP}$ et sa conservation. Ces méthodes et les analyses qui les accompagnent servent de point de départ à des travaux actuellement en cours et visant à produire des méthodes $\mathcal{RD}$ performantes pour résoudre les équations de Navier-Stokes. Pour éviter le surcoût en mémoire prohibitif du FOS, ceux-ci se sont plutôt orientés vers une formulation de type gradients reconstruits.

Chapitre 5

Tests numériques

Sommaire

Afin d'appuyer les raisonnements mis en oeuvre dans les précédents chapitres, il est maintenant temps de les confronter à l'expérience. Nous allons pour cela procéder progressivement, des cas en apparence les plus simples aux plus complexes. Mais avant toute chose, il nous semble approprié de présenter succintement la méthode de parallélisation mise en oeuvre dans le logiciel FluidBox.

5.1 Parallélisation

Un des attraits des schémas $\mathcal{RD}$ est leur compacité, qui en fait des méthodes tout à fait adaptées à une parallélilasation par décomposition de domaines. En effet, pour actualiser les valeurs de chaque degré de liberté du maillage, seuls les premiers voisins entrent en jeu. Dans notre contexte, puisque les degrés de liberté sont de type nodal et que l'assemblage des contributions (et de la matrice implicite le cas échéant) se fait par élément, deux degrés de liberté sont dits premiers voisins lorsqu'ils appartiennent à un même élément. Cette dénomination vient du cas P_1 (triangles ou tétraèdres), où les premiers voisins d'un noeud sont tous ceux qui sont reliés à lui par une arête. De façon classique, pour répartir la charge de travail entre plusieurs unités de calcul, l'idée est alors de découper le domaine en n partitions, où n désigne le nombre de processeurs que nous souhaitons utiliser, en équilibrant si possible les charges de travail des différents processeurs de manière équitable. Cette étape est très bien réalisée par le logiciel Scotch. Le découpage est effectué par arête (2D) ou par face (3D), ce qui signifie que les degrés de liberté appartenant à ces morceaux d'hyperplans sont dupliqués localement sur chaque sous-domaine généré.

Pour actualiser les valeurs de la solution en ces degrés de liberté, la méthode séquentielle ferait appel à des informations qui appartiennent désormais au(x) sous-domaine(s) voisin(s). En parallèle, les données

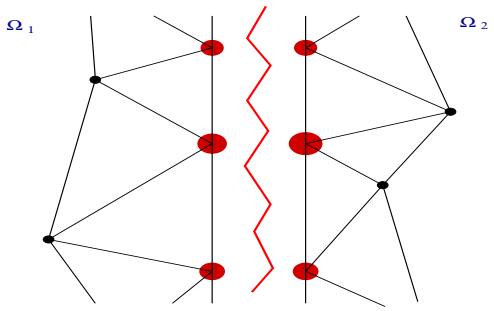


FIGURE 5.1 – Exemple de découpage simple illustrant localement un résultat que pourrait donner le partionnement par Scotch sur un maillage de triangles P_1. À ce stade, seuls les noeuds en rouge sont dupliqués.

disparues doivent être communiquées par les processeurs qui y ont accès, par le biais de fonctions MPI. Néanmoins, il n'est pas judicieux d'établir une communication à chaque fois qu'une donnée manquante devrait être utilisée, et de fonctionner exactement comme en séquentiel, car l'envoi et la réception des messages a un coût non négligeable. La méthode couramment utilisée consiste à dupliquer les premiers voisins manquants associés à chaque degré de liberté situé sur une frontière de partition. De cette manière, à chaque itération en temps (en instationnaire, ou pseudo-temps en stationnaire), les calculs peuvent être menés par chaque processeur de manière complètement autonome. Une fois la nouvelle solution calculée en chaque degré de liberté, il ne reste plus qu'à écraser les valeurs des degrés de liberté dupliqués qui ne sont par correctement mis à jour par chaque sous-domaine, la bonne valeur devant être communiquée par celui des processeurs voisins qui l'a obtenue. Ainsi, on minimise le nombre d'échanges au prix d'une augmentation généralement très légère de la consommation de mémoire, par rapport à la consommation totale. En effet, le surcoût en mémoire est proportionnel au nombre de degrés de liberté dupliqués, ou pour utiliser un autre vocabulaire, à la taille de la zone de chevauchement (*overlap*). Or, au moins pour les gros problèmes qui sont les plus limitants, cela ne représente habituellement qu'une partie négligeable du maillage.

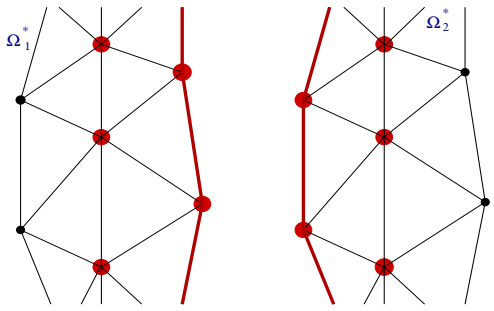


FIGURE 5.2 – Duplication de degrés de liberté (P_1 ici) supplémentaires (rouge) pour simplifier le schéma de communications. On travaille sur les nouveaux sous-domaines Ω_i^*. Si on fait l'expérience de pensée de fusionner les partitions de façon à revenir au maillage intial (dans cet exemple, en superposant les noeuds situés sur les arêtes verticales au centre), la zone d'*overlap* est la réunion de tous les éléments dont les sommets sont en rouge (délimitée par les lignes brisées rouge).

5.2 Études sur un cas simple 2D : une gaussienne MHD

La validation de nos méthodes va se faire sur des écoulements idéaux bidimensionnels, et sur des éléments P_1. En mécanique des fluides, une quantité importante de résultats ont été obtenus, même si beaucoup traitent de problèmes stationnaires (voir les travaux de thèse [62], [90] et [59], à l'exception notable de [75]). Nous voulons ici montrer comment se comportent les méthodes instationnaires que nous avons présentées, lorsqu'elles sont appliquées au système de la Magnétohydrodynamique (idéale). Ceci avait été abordé une première fois par [30] dans le cas 1D.

Pour commencer, nous avons jugé utile d'aborder un cas qui illustre le comportement des schémas $\mathcal{RD}$ instationnaires sur des problèmes réguliers, "lisses", où la capacité numérique à résoudre correctement les gradients forts n'est pas indispensable. Ce problème 2D considère une "boîte" carrée de côté 1, centrée en $(0,0)$ et ouverte de tous côtés, qui contient un plasma initialement statique. Celui-ci est davantage concentré au centre du domaine, selon une loi de densité de la forme d'une fonction gaussienne :

$$\rho(r) = (\rho_{\max} - \rho_0) e^{-\frac{4\ln(2)x^2}{H^2}}$$

Nous avons pris $\rho_{\max} = 10$, $\rho_0 = 1$ et la largeur à demi-hauteur $H = 0,1$ (voir la figure 5.3). Le rapport des capacités thermiques γ vaut $\frac{5}{3}$. De plus, comme on suppose que la température est uniforme, d'après la loi d'état des gaz parfaits, la pression suit l'évolution de la densité de la manière suivante :

$$p(r) = p_0 \frac{\rho(r)}{\rho_0}$$

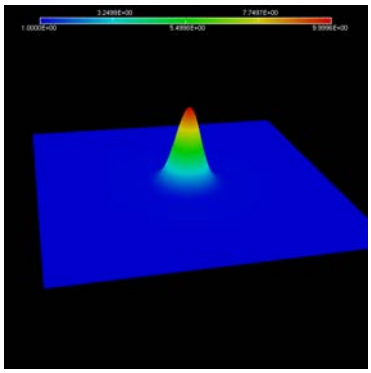


FIGURE 5.3 – Répartition de la densité $\rho(r)$ initiale

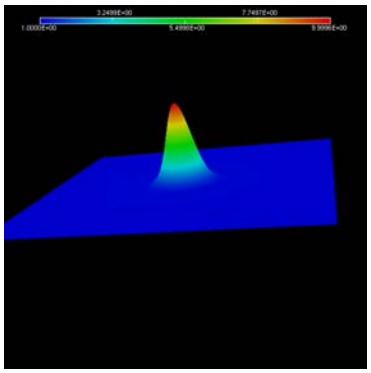
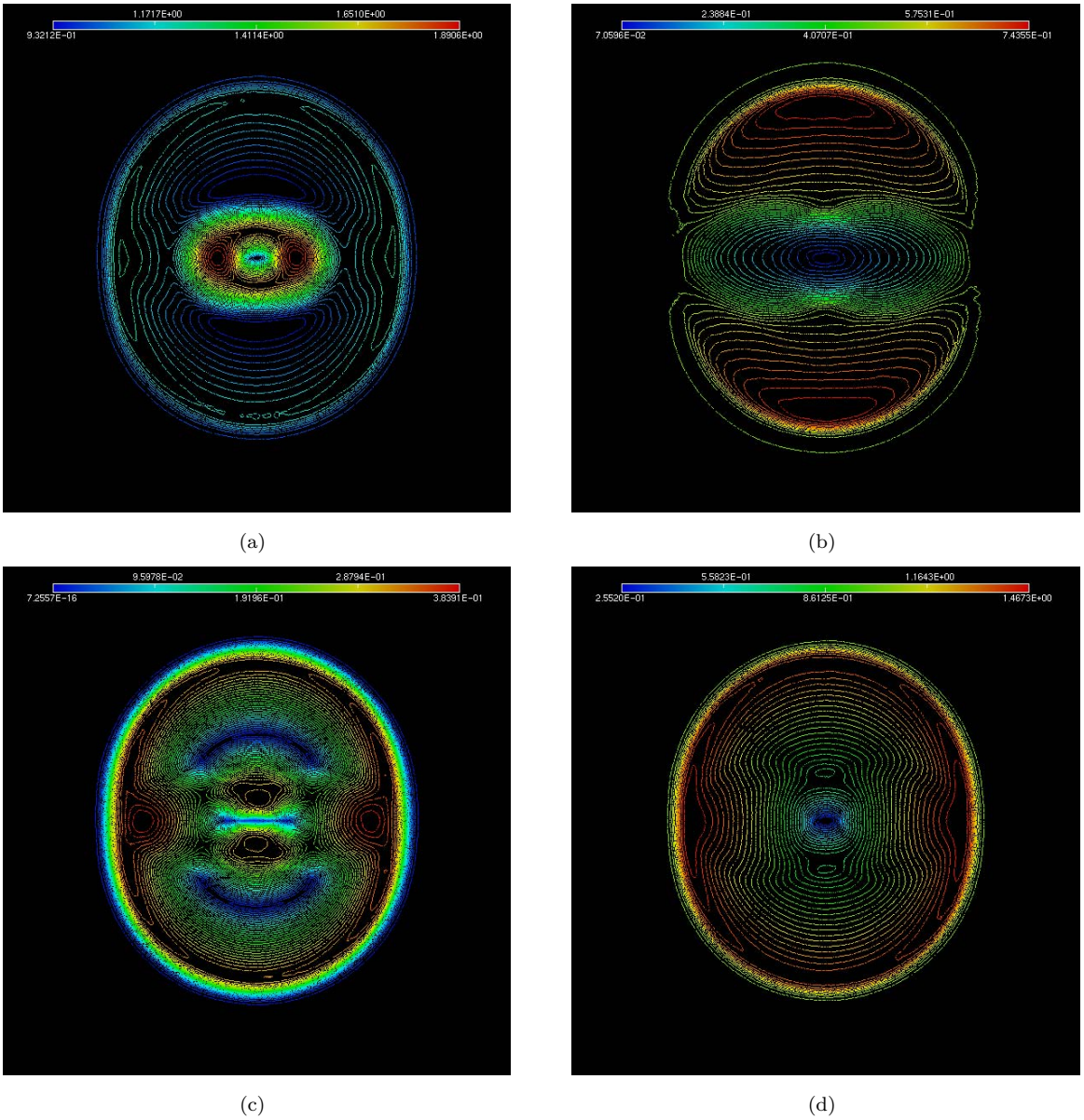


FIGURE 5.4 – Répartition de la pression $p(r)$ initiale

Ces gradients sont à peu près de l'amplitude limite au-delà de laquelle le schéma LDA, par exemple, sur un maillage relativement grossier (40×40), n'est plus assez robuste. Le saut est d'amplitude faible et s'étale sur plus d'un élément pour les maillages considérés, de telle sorte que le problème est bel et bien assez régulier. Si on s'en tient à cette description, en considérant que le fluide n'est pas un plasma et en résolvant les équations d'Euler, on s'attend intuitivement à voir le gradient de pression central donner naissance à une onde acoustiques se propageant radialement, visible sur les champs de la densité et de la vitesse car elle entraine la matière avec elle. Par symétrie, sa forme ne peut être que circulaire (un cercle qui s'étend progressivement), à l'image des vagues concentriques qu'on observerait en jetant une pierre dans un lac. Ce cas test simple peut être étendu en MHD simplement en ajoutant un champ magnétique ambiant non nul. On prendra ici un champ initialement uniforme et aligné selon l'axe horizontal, avec $B_x = 1$ et $B_y = 0$. Chaque particule est alors soumise à une force de Lorentz $q\overrightarrow{u} \wedge \overrightarrow{B}$ qui tend à lui imprimer une trajectoire hélicoïdale d'axe dirigé suivant $\overrightarrow{B}$, ce qui se traduit par un mouvement d'ensemble dans cette même direction : c'est l'idée de base du confinement magnétique.

Le cercle témoignant de la propagation de la matière (visible sur la densité), auquel on s'attendrait si le fluide était neutre, se verra donc aplati dans le sens horizontal.

Nous allons maintenant utiliser ce problème pour montrer les différences entre plusieurs techniques évoquées dans les chapitres précédents. Nous allons nous intéresser aux différents schémas d'ordre 2 que nous avons mis en oeuvre sur des maillages P_1. Pour chacune des simulations qui vont suivre, la méthode en temps sera celle de Runge-Kutta d'ordre 2 présentée au chapitre précédent. Le temps final de la simulation est $t_{\text{final}} = 0,18$, qui correspond au moment où le plasma atteint les bords et s'apprête à sortir du domaine. Pour fixer les idées, la solution référence que nous obtenons avec le schéma de Rusanov limité et stabilisé, sur un maillage très fin 400×400, est donnée par la figure 5.5.



(a) (b)

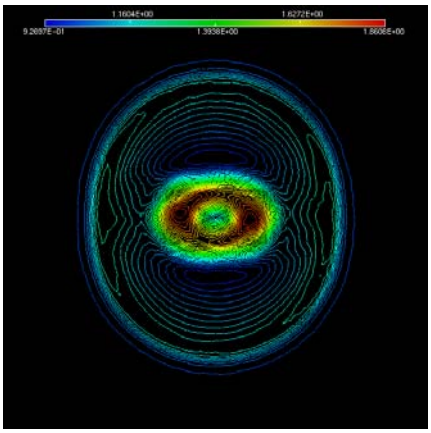
(c) (d)

FIGURE 5.5 – Profils de référence (à $t = 0,18$) : (a) densité ρ, (b) pression magnétique $\frac{\overrightarrow{B}^2}{2}$, (c) vitesse $\|\overrightarrow{u}\|$, (d) pression p

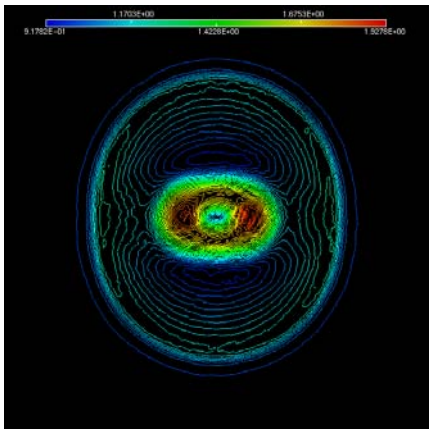
5.2.1 Comparaison des schémas $\mathcal{RD}$ d'ordre 2

Nous avons tendance à distinguer les schémas en deux catégories, non pas *upwind* ($\mathcal{MU}$) ou centrés, mais monotones ou naturellement $\mathcal{LP}$. Les schémas monotones à l'ordre 1 sont le schéma N ($\mathcal{MU}$) et celui de Rusanov (centré), qu'on limite pour obtenir l'ordre 2. Les schémas naturellement $\mathcal{LP}$ et donc d'ordre 2 sont le LDA et le SU. Pour cette comparaison, la limitation que nous utilisons est celle qui projette les résidus sur une base de vecteurs propres moyens, et qui projette ensuite les signaux obtenus sur le cercle circonscrit (revoir la section 3.4.1 pour plus de détails). De plus, nous ne faisons pas usage ici des outils que sont la stabilisation des schémas limités et le *divergence cleaning*. Sur un cas régulier comme celui-ci, nous allons voir que le fait de ne pas corriger le champ magnétique n'est pas fatal à la simulation. Les profils les plus sensibles sont ceux sur la densité et la pression magnétique, et afin d'alléger les pages suivantes nous ne comparerons que ceux-ci.

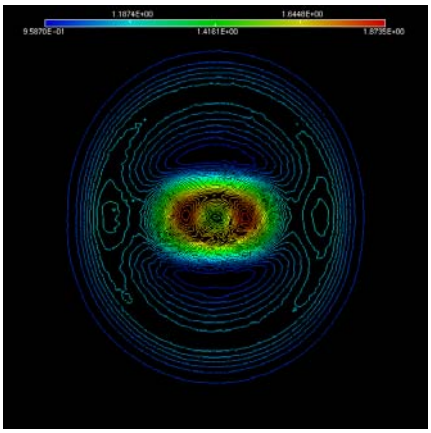
Les figures 5.6 et 5.7 montrent les résultats obtenus par les 4 schémas d'ordre 2 sur un maillage de taille raisonnable, 100×100. Tous les calculs ont été menés avec un nombre CFL égal à $0,9$, à l'exception du schéma SU pour lequel il est nécessaire de descendre à $0,8$. Ce dernier s'avère donc être le moins robuste, y compris vis-à-vis du schéma LDA.



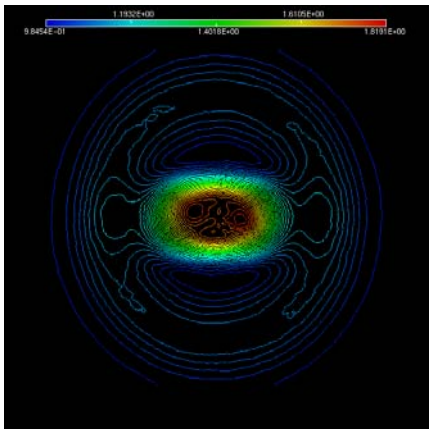
(a) LLF



(b) LN

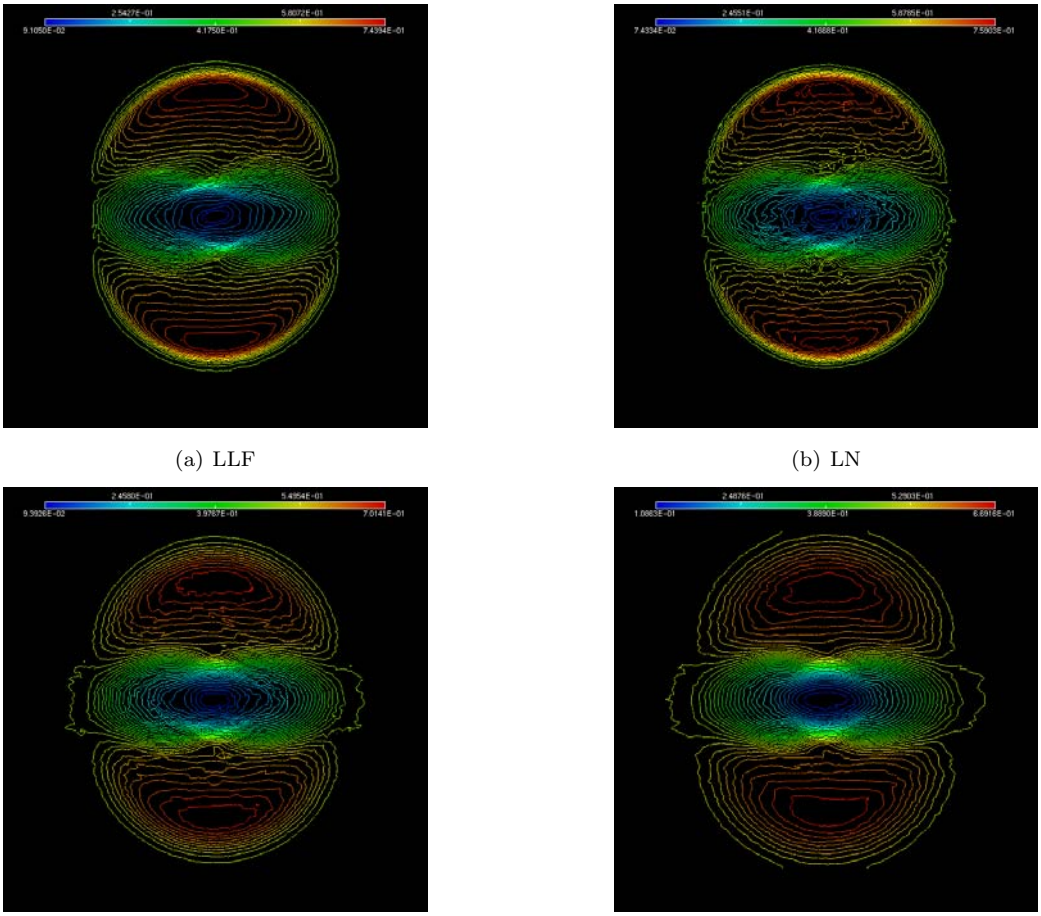


(c) LDA



(d) SU

FIGURE 5.6 – Profils de densité à $t = 0,18$ par les 4 schémas $\mathcal{RD}$: (a) Lax-Friedrichs (Rusanov) limité, (b) N limité, (c) LDA, (d) SU



(a) LLF (b) LN

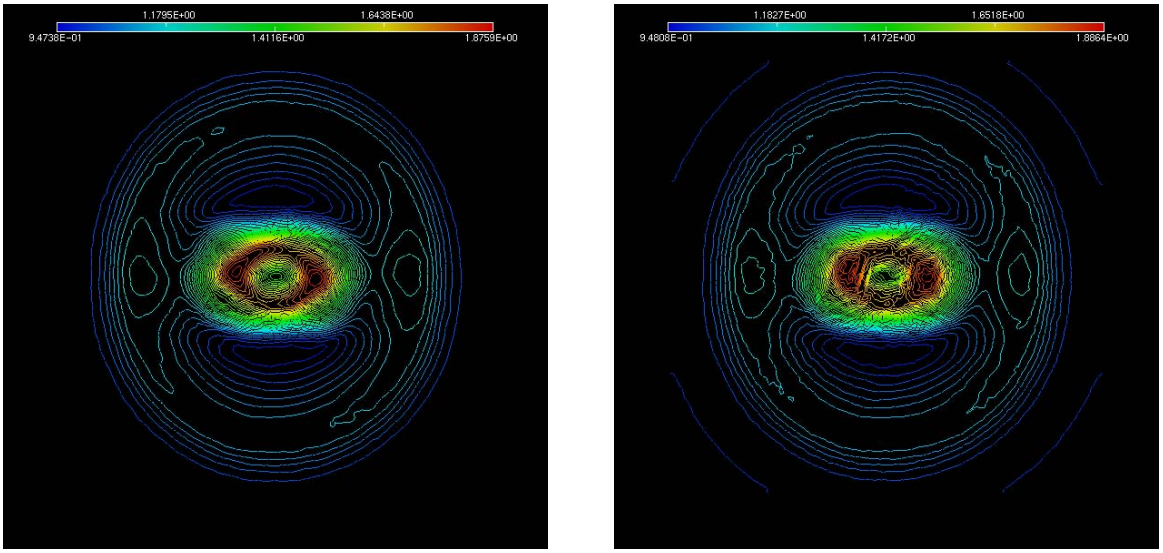
(c) LDA (d) SU

FIGURE 5.7 – Pression magnétique à $t = 0,18$ obtenue par les 4 schémas $\mathcal{RD}$: (a) Lax-Friedrichs (Rusanov) limité, (b) N limité, (c) LDA, (d) SU

Curieusement, même les schémas *upwind* montrent de plus fortes oscillations que le schéma de Lax-Friedrichs. Nous allons voir d'où cela vient.

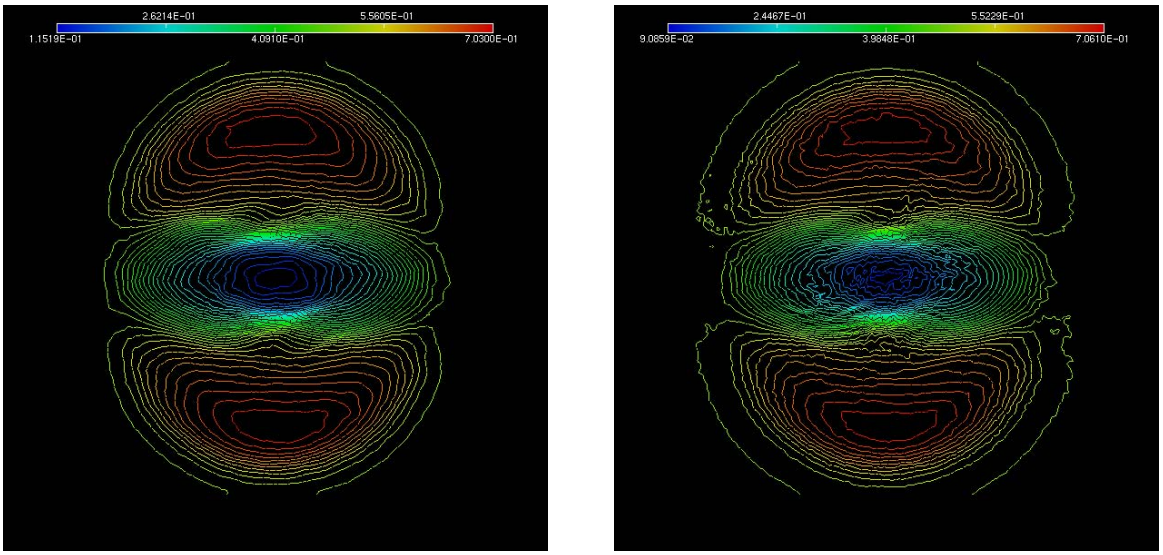
5.2.2 Remédier aux oscillations

Les oscillations peuvent avoir deux origines distinctes : un manque de stabilité de la distribution, ce qui n'est plus le cas dès qu'on dispose d'un caractère *upwind*, et le développement des erreurs commises sur la divergence du champ magnétique. Concernant la première, elle peut être compensée par l'ajout d'un terme de stabilisation, comme nous l'annoncions au chapitre 3. Les schémas qui en ont besoin sont ceux qui sont limités, le N et le Lax-Friedrichs, car les limitations que nous avons présentées ne tiennent absolument pas compte du sens de l'écoulement. On ne connaît pas à ce jour de façon de projeter les résidus dans un sens *upwind*. Les schémas LDA et SU ne sont pas atteints par ce problème et il serait inutile de leur adjoindre un terme de stabilisation. Les figures 5.8 et 5.9 montrent ce que donnent les deux schémas limités une fois stabilisés. Le maillage ne change pas et le nombre CFL reste à $0,9$.



(a) LLFS (b) LNS

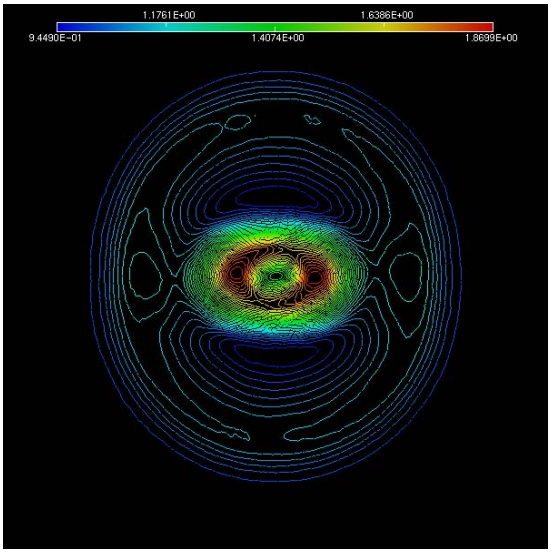
FIGURE 5.8 – Profils de densité à $t = 0.18$ obtenus par : (a) Lax-Friedrichs (Rusanov) limité stabilisé, (b) N limité stabilisé



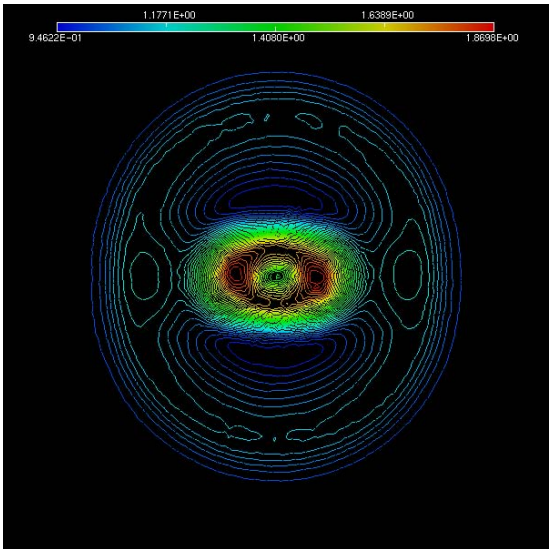
(a) LLFS (b) LNS

FIGURE 5.9 – Pression magnétique à $t = 0,18$ obtenue par : (a) Lax-Friedrichs (Rusanov) limité stabilisé, (b) N limité stabilisé

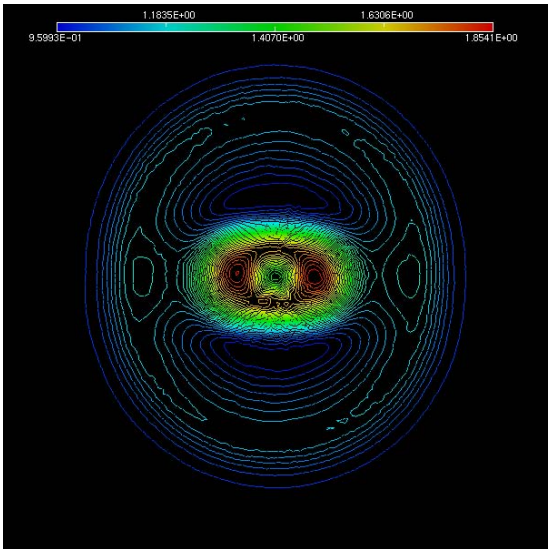
Enfin, nous rajoutons l'approche *divergence cleaning* hyperbolique et pouvons observer une nette diminution des oscillations sur les figures 5.10 et 5.11, avec tous les schémas.



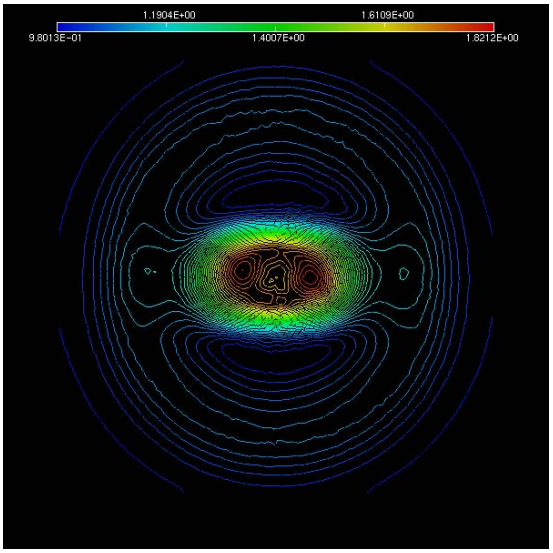
(a) LLFS



(b) LNS



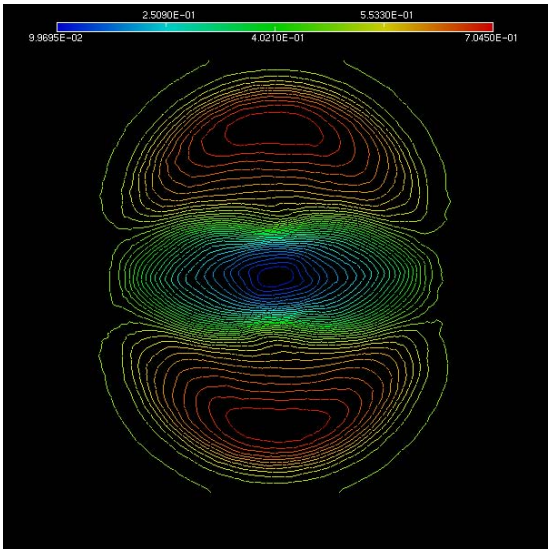
(c) LDA



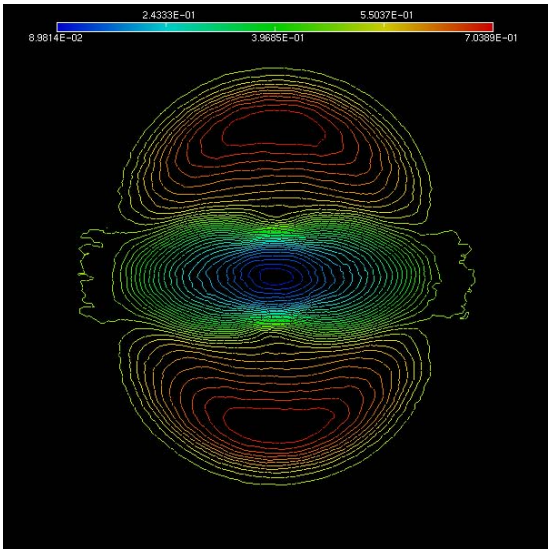
(d) SU

FIGURE 5.10 – Profils de densité à $t = 0,18$ obtenus, avec correction de la divergence, par : (a) Lax-Friedrichs (Rusanov) limité stabilisé, (b) N limité stabilisé, (c) LDA, (d) SU

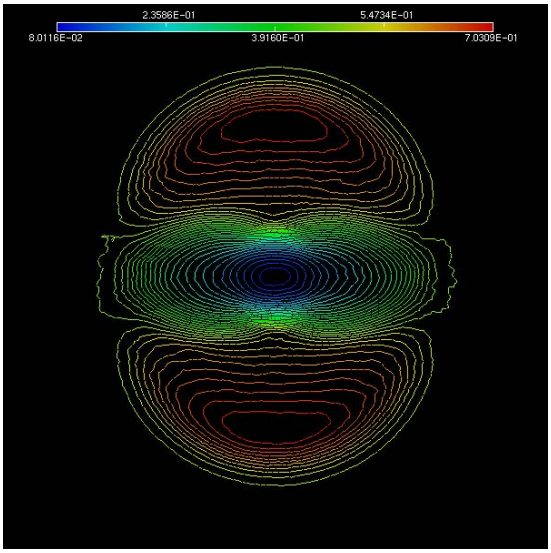
Ces résultats sont satisfaisants pour la résolution considérée, et le schéma SU fonctionne désormais avec un nombre CFL de $0,9$. Ceci confirme que les schémas *upwind* bruts ont tendance à davantage détériorer le caractère solénoïdal du champ magnétique, défaut qui est assez bien corrigé par l'approche hyperbolique du *divergence cleaning*. Cette comparaison nous incite à privilégier le schéma de Lax-Friedrichs limité et stabilisé comme schéma par défaut pour aborder tous les cas tests. C'est celui qui montre les résultats les plus satisfaisants. Son coût de calcul est moindre que celui des autres, et tient principalement (à hauteur de presque 50%) au calcul du terme de stabilisation.



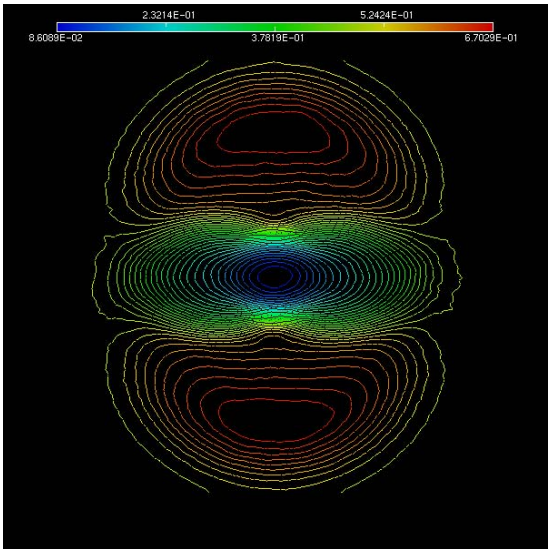
(a) LLFS



(b) LNS



(c) LDA



(d) SU

FIGURE 5.11 – Pression magnétique à $t = 0,18$ obtenue, avec correction de la divergence, par : (a) Lax-Friedrichs (Rusanov) limité stabilisé, (b) N limité stabilisé, (c) LDA, (d) SU

5.2.3 Sur le calcul du terme SUPG

Un détail important dans nos algorithmes est le calcul du terme de stabilisation, qui est le terme de décentrement du schéma SU. Sur des maillages P1, en accord avec les principes énoncés dans les chapitres précédents, nous allons comparer trois façons de le calculer. Rappelons-en tout d'abord l'expression, en notant r l'opérateur linéarisé qui décrit le système d'équations, soit $r = \partial_t + \overrightarrow{\lambda}(U) \cdot \overrightarrow{\nabla}$:

$$S_i(U_h) = \int_T r(\varphi_i) \tau r(U_h)$$

Cependant, nous n'interpolons la solution qu'en espace ce qui a pour effet d'annuler la partie temporelle du premier facteur :

$$r(\varphi_i) = \frac{\partial \varphi_i}{\partial t} + \overrightarrow{\lambda} \cdot \overrightarrow{\nabla} \varphi_i = \overrightarrow{\lambda} \cdot \overrightarrow{\nabla} \varphi_i$$

où $\overrightarrow{\lambda}$ est le vecteur des matrices jacobiennes. Compte tenu de la définition de τ avec la matrice N, on avait donc finalement établi :

$$S_i(U_h) = |T| \int_T \left(\overrightarrow{\lambda} \cdot \overrightarrow{\nabla} \varphi_i \right) N \left(\frac{\partial U_h}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U_h) \right) dV$$

Pour mettre en évidence l'importance du bon calcul de ce terme, nous allons tester trois façons de le calculer.

1. Simplifier en moyennant

Nous avions déjà évoqué cette simplification dans la section 3.4.2, qui consiste à moyenner le premier facteur sur chaque élément. Comme les gradients des fonctions de base sont constants en P_1, cela permettait d'écrire :

$$\begin{aligned} S_i^1(U_h) &= |T| \overrightarrow{\lambda}(\overline{U}) \cdot \frac{\overrightarrow{n}_i}{2|T|} N \int_T r(U_h) dV \\ &= \frac{1}{2} K_i N \Phi \end{aligned}$$

par définition des matrices K_i.

2. Une meilleure représentation des jacobiennes

La seconde idée est d'utiliser pour le premier une représentation du même ordre que pour les flux, à savoir linéaire si le résidu est calculé à l'ordre 2 (par exemple par une formule des trapèzes). Pour ce faire, une manière simple est d'ultiliser les degrés de liberté dont nous disposons pour interpoler les jacobiennes selon les fonctions de base :

$$\overrightarrow{\lambda}_h(U) = \sum_{M_i \in T} \overrightarrow{\lambda}(U_i) \varphi_i = \overrightarrow{\lambda}(U) + O(h^2)$$

On obtient alors :

$$\begin{aligned} S_i^2(U_h) &= |T| \left(\sum_{M_j \in T} \overrightarrow{\lambda}(U_j) \cdot \frac{\overrightarrow{n}_i}{2|T|} \right) N \int_T \varphi_i r(U_h) dV \\ &= \frac{1}{2} \left(\sum_{M_j \in T} \overrightarrow{\lambda}(U_j) \cdot \overrightarrow{n}_i \right) N \Phi_i^{T,G} \end{aligned}$$

où $\Phi_i^{T,G}$ désigne le résidu du schéma de Galerkin calculé en linéarisant les flux, soit :

$$\Phi_i^{T,G} = \sum_{M_j \in T} m_{ij}^G \frac{\partial U_j}{\partial t} + \frac{1}{6} \sum_{M_j \in T} \overrightarrow{F}(U_j) \cdot \overrightarrow{n}_j$$

avec la matrice de masse de Galerkin qui vaut sur les éléments P_1 :

$$m^G = |T| \begin{pmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{pmatrix}$$

3. Simplification pour être dissipatif

Il ne faut pas perdre de vue que l'intérêt de la stabilisation est d'être dissipative dans le cas continu, c'est-à-dire de respecter une inégalité sur l'énergie telle que présentée dans l'annexe B. Or nous avons vu que la partie temporelle du facteur de gauche s'annule, si bien que :

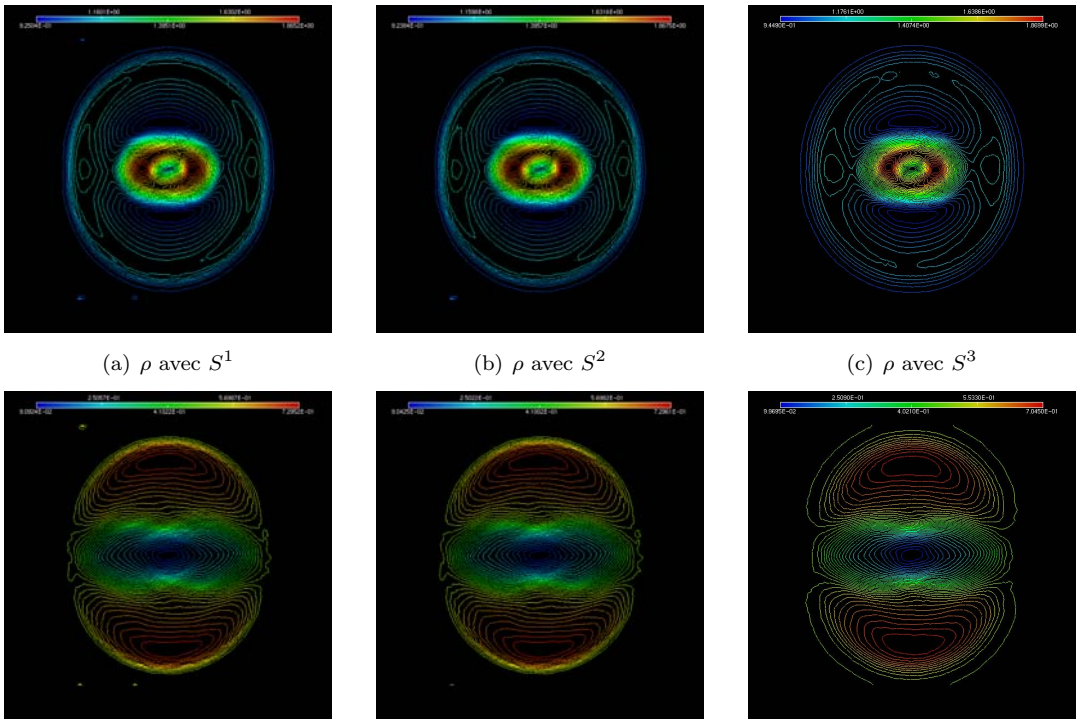
$$\sum_{M_i \in T} \langle U_i, S_i(U_h) \rangle = h^d \int_T \left(\overrightarrow{\lambda} \cdot (\overrightarrow{\nabla} U_h) \right) N \left(\frac{\partial U_h}{\partial t} + \overrightarrow{\lambda} \cdot (\overrightarrow{\nabla} U_h) \right) dV \not> 0$$

puisque même si la matrice N est positive, en décomposant, le terme contenant la dérivée en temps est de signe inconnu. Par conséquent, pour être dissipatif, il est formellement nécessaire de ne conserver que le terme stationnaire. Ainsi, la troisième version de la stabilisation que nous testerons est :

$$\begin{aligned} S_i^3(U_h) &= \frac{1}{2}\left(\sum_{M_j\in T}\overrightarrow{\lambda}(U_j)\cdot\overrightarrow{n}_i\right)N\int_T\varphi_i\sum_{M_j\in T}\overrightarrow{F}(U_j)\cdot\frac{\overrightarrow{n}_j}{2|T|}dV \\ &= \frac{1}{12}\left(\sum_{M_j\in T}\overrightarrow{\lambda}(U_j)\cdot\overrightarrow{n}_i\right)N\sum_{M_j\in T}\overrightarrow{F}(U_j)\cdot\overrightarrow{n}_j \end{aligned}$$

Comparaisons

Dans les comparaisons suivantes, la correction de la divergence est activée. On regarde les différences sur les profils finaux de densité et de pression magnétique. On commence par le Lax-Friedrichs, puis viennent le schéma N et enfin le schéma SU.



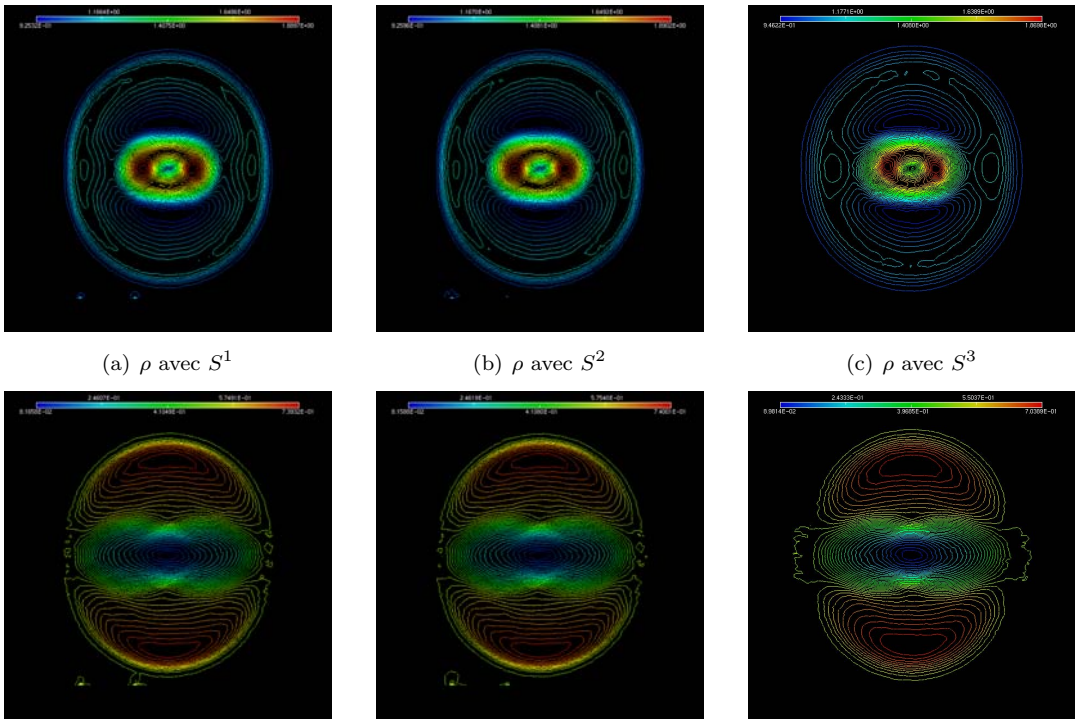
(a) ρ avec S^1 (b) ρ avec S^2 (c) ρ avec S^3

(d) Pression magnétique avec S^1 (e) Pression magnétique avec S^2 (f) Pression magnétique avec S^3

FIGURE 5.12 – Influence de la stabilisation sur le schéma LLFS. En haut : profils de densité avec, de gauche à droite S^1, S^2 et S^3. En bas : pression magnétique avec, de gauche à droite S^1, S^2 et S^3.

On constate que S^3 donne les résultats les moins parasités, au prix d'une diffusion légèrement supérieure aux autres choix, comme nous nous y attendions d'après les arguments précédents. En particulier, les oscillations que nous voyons surgir sur les bords avec S^1 et S^2 n'apparaissent pas avec S^3.

Voyons à présent les effets sur le schéma N limité et stabilisé.



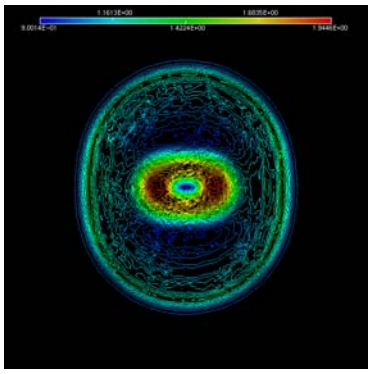
(a) ρ avec S^1 (b) ρ avec S^2 (c) ρ avec S^3

(d) Pression magnétique avec S^1 (e) Pression magnétique avec S^2 (f) Pression magnétique avec S^3

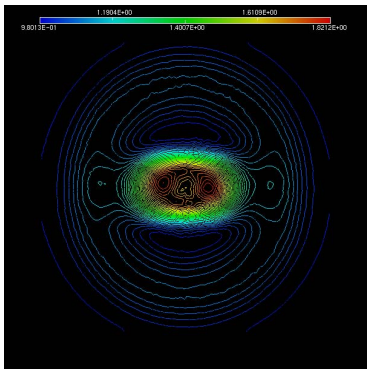
FIGURE 5.13 – Influence de la stabilisation sur le schéma LNS. En haut : profils de densité avec, de gauche à droite S^1, S^2 et S^3. En bas : pression magnétique avec, de gauche à droite S^1, S^2 et S^3.

Une fois encore, on peut tirer les mêmes conclusions, en notant que dans tous les cas il reste des oscillations. Une résolution plus fine permettrait de beaucoup les réduire, mais ce n'est pas ce que nous cherchons ici. La diffusion introduite par S^3 est assez faible, et diminue nettement en raffinant le maillage.

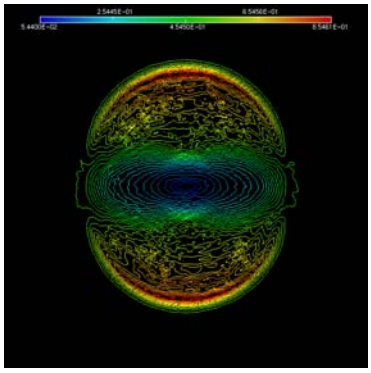
Et enfin, voyons ce qui se passe sur le schéma SU, qui lui n'est pas seulement stabilisé par ce terme : c'est tout le décentrement qu'il gouverne.



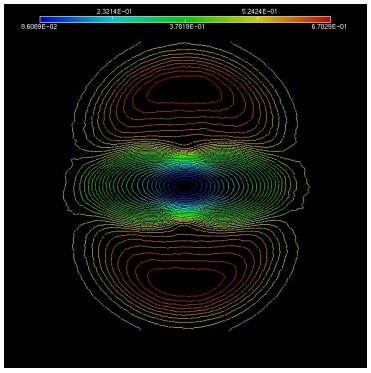
(a) ρ avec S^1



(b) ρ avec S^3



(c) Pression magnétique avec S^1



(d) Pression magnétique avec S^3

FIGURE 5.14 – Influence de la stabilisation sur le schéma SU. En haut : profils de densité avec, de gauche à droite S^1 et S^3. En bas : pression magnétique avec, de gauche à droite S^1 et S^3.

Ces résultats sont les plus surprenants. Le schéma de Galerkin n'est pas stable, et quand on lui adjoint un décentrement mal construit, cela se voit. La stabilisation classique par S^1 est beaucoup trop inefficace. La stabilisation par S^2, i.e. avec les termes instationnaires non dissipatifs (voire déstabilisants), ne passe tout simplement pas, même en baissant la CFL en deçà de $0,9$, d'où son absence dans la figure ci-dessus. D'un autre côté, la diffusion apportée par S^3 est vraiment très importante sur une telle résolution, que l'on qualifirait d'intermédiaire. L'usage du schéma SU avec S^3 devrait se faire sur des maillages plus fins.

En conclusion

Cette étude nous incite à considérer que la formulation S^3 est la plus robuste et la plus adaptée à ce que nous attendons du terme de stabilisation. C'est donc celle-ci que nous retiendrons par défaut, en veillant à rester vigilants sur la quantité de diffusion qu'elle introduit. Elle est plus adaptée à des maillages plus fins, comme nous allons le voir sur le problème suivant.

5.3 Autres problèmes académiques

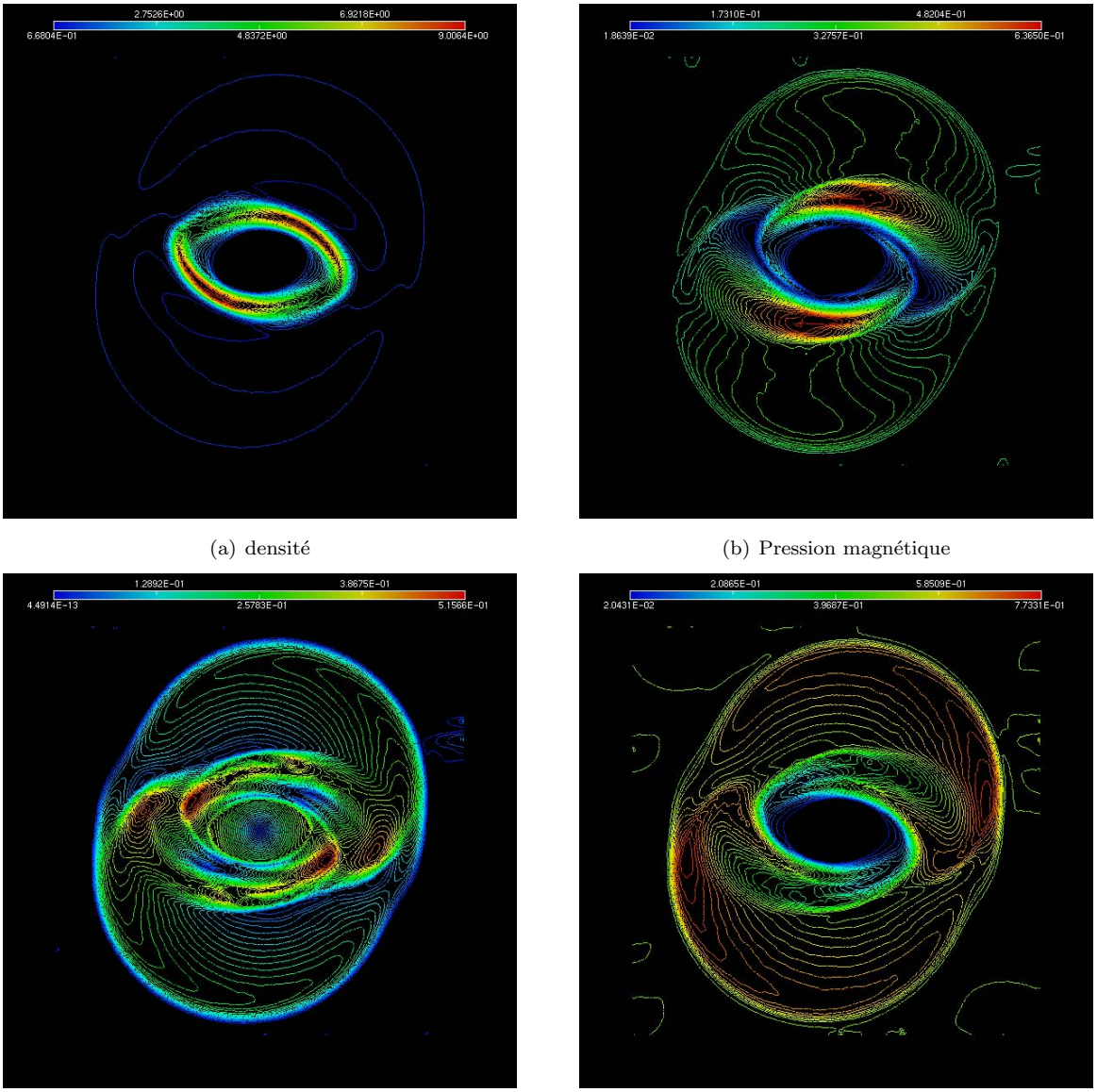
5.3.1 Le rotor

Ce problème est issu de [11], et a été repris par [95] en corrigeant une erreur de texte sur les temps finaux. Il y a donc dans ce dernier papier deux problèmes du rotor, que nous allons reproduire ici. Pour les deux, le plasma initial est contenu dans un carré de côté 1 centré en $(0,0)$ et divisé en deux zones, une au centre où le plasma tourne rapidement et l'autre où le plasma ambiant est statique. La zone centrale est délimitée par un cercle de rayon $r = 0,1$ centré en $(0,0)$. Le plasma y est dense, avec $\rho = 10$, et tourne avec une vitesse radiale constante $\omega = 20$ rad/s, de telle sorte que la vitesse en $r = 0,1$ vaut 2. Dans le reste du domaine, le plasma a une densité $\rho = 1$ et une vitesse nulle. Comme dans [11], une linéarisation est appliquée sur ρ et $\overrightarrow{u}$ de manière à éviter une discontinuité et un départ trop violent. Cette zone tampon est comprise entre $r = 0,1$ et $r = 0,115$: on a $\rho(0,1) = 10$, $\rho(0,115) = 1$ et ρ linéaire entre les deux, et de la même manière, $\overrightarrow{u}|_{r=0,1} = \begin{pmatrix} -y\omega \\ x\omega \end{pmatrix}$, $\overrightarrow{u}|_{r=0,115} = \overrightarrow{0}$ et $\overrightarrow{u}$ linéaire entre les deux. L'ensemble du plasma subit l'action d'un champ magnétique initialement uniforme et dirigé selon l'axe Ox, d'intensité $B_x = \frac{5}{\sqrt{4\pi}}$ pour le premier problème et $B_x = \frac{2,5}{\sqrt{4\pi}}$ pour le second. C'est la seule différence entre les deux. Enfin, pour les deux problèmes, la pression est uniforme à $p = 1$ et γ vaut $1,4$.

La force centrifuge du plasma central n'est compensée ni par un gradient de pression hydrodynamique ni par un gradient de pression magnétique. Si le fluide (idéal, sans viscosité) était neutre, il se propagerait en cercle comme pour le cas de la gaussienne vu plus haut, en conservant son moment angulaire. Ici, le champ magnétique horizontal tend à confiner, donc à freiner, les particules avec une intensité qui dépend de leur direction (la force de Lorentz est nulle lorsque la vitesse est selon x et maximale lorsqu'elle est suivant y). Il en résulte une propagation oblique, l'onde associée étant une onde d'Alfvén. Les figures 5.15 et 5.16 montrent les solutions aux deux problèmes.

Le second problème

On commence par le second problème, qui est le plus lent. Le résultat visible sur la figure 5.15 a été obtenu sur un maillage fin 200×200, avec un nombre CFL égal à $0,9$, et au bout de $0,3$ s.



(a) densité

(b) Pression magnétique

(c) vitesse

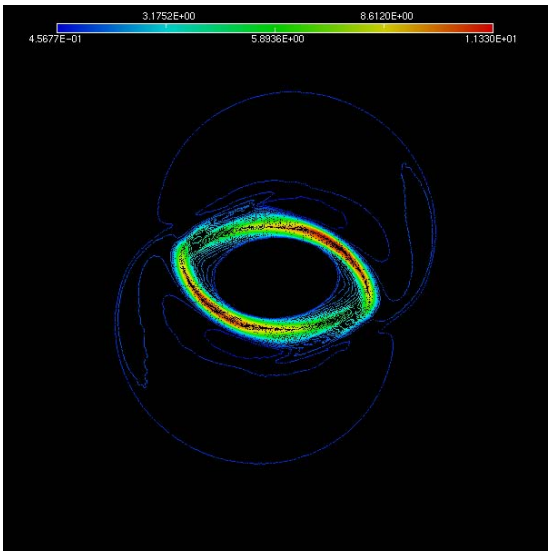
(d) pression

FIGURE 5.15 – Profils finaux du second problème, à $t = 0,3$ s, avec le schéma LLFS et la stabilisation S^3.

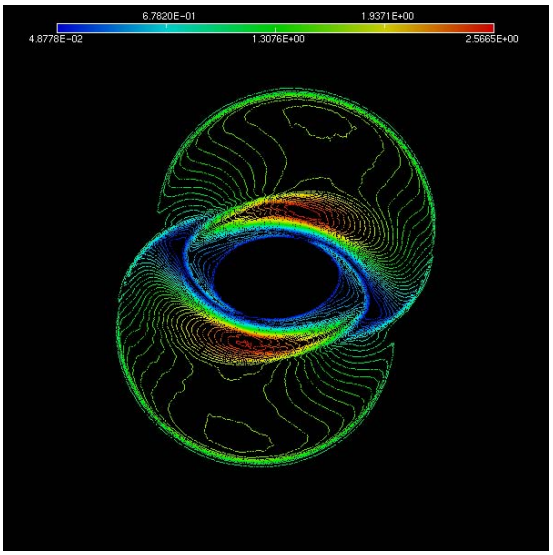
L'origine des oscillations observées sur les bords est incertaine. Leur apparition est très récente et visiblement sensible à certains paramètres encore non identifiés, car on ne les observe pas tout le temps. Il semblerait, mais ce n'est pas encore clair, que cela provienne de la stabilisation.

Le premier problème

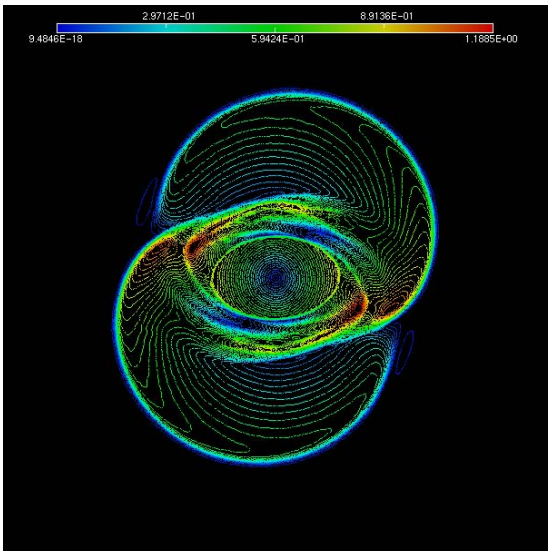
Cette simulation a été réalisée sur un maillage très fin 400×400, toujours avec une CFL égale à $0,9$, en faisant appel à 8 processeurs pendant une dizaine de minutes. Le champ magnétique est deux fois plus intense que sur le second problème, et comme l'écoulement est gouverné par des ondes d'Alfvén, de vitesse $\frac{\|\vec{B}\|}{\sqrt{\rho}}$ avec ρ initialisé à la même valeur dans les deux cas, le temps final de ce problème doit logiquement être à peu près divisé par 2 par rapport au cas précédent, soit un temps final à $0,15$ s.



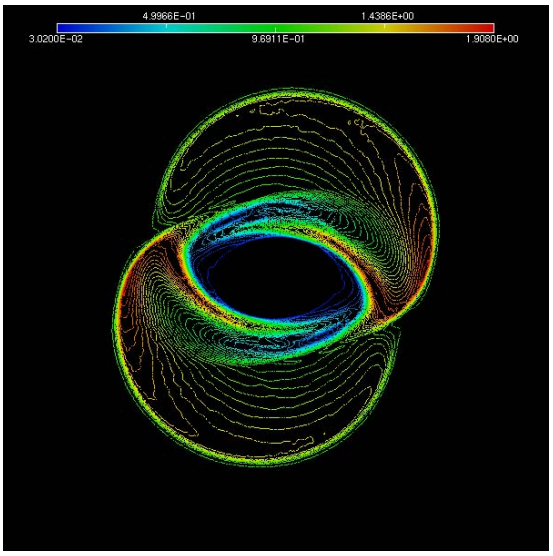
(a) densité



(b) Pression magnétique



(c) vitesse



(d) pression

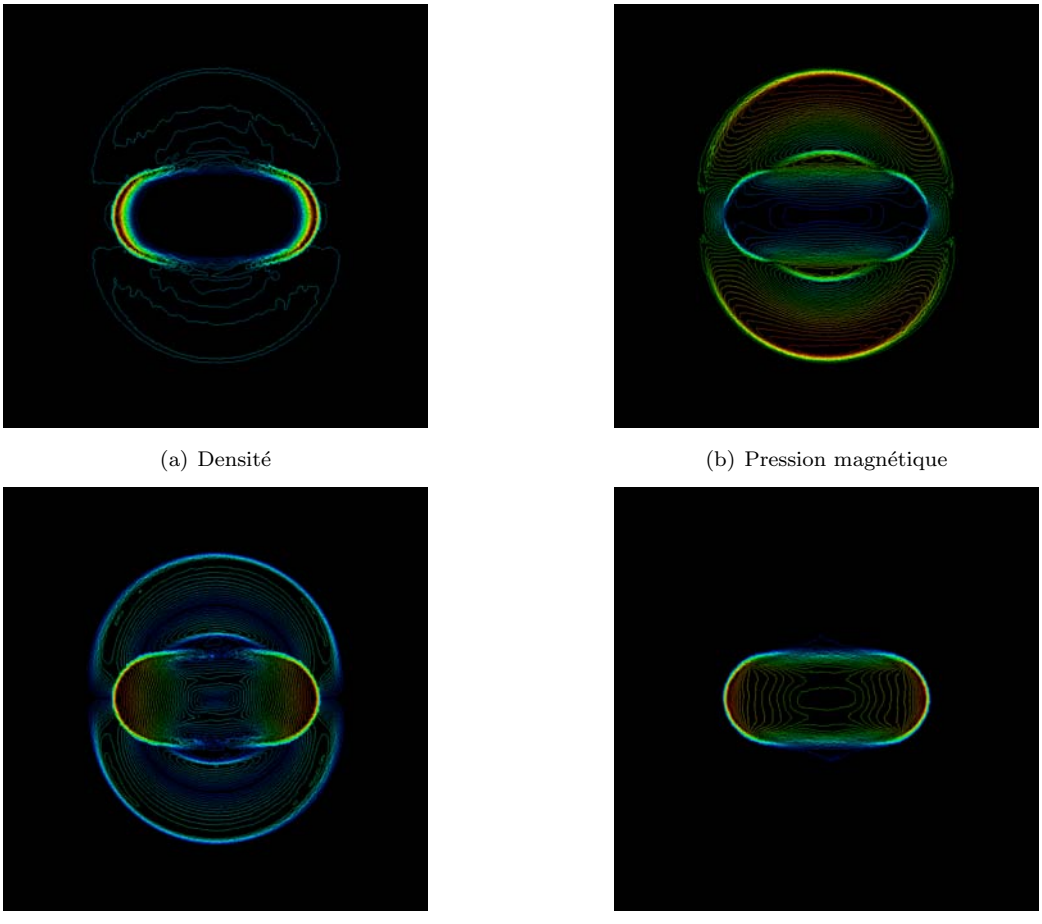
FIGURE 5.16 – Profils finaux du premier problème, à $t = 0,15$ s, avec le schéma LLFS et la stabilisation S^3.

On peut remarquer, au problème des oscillations près, que le passage à une résolution aussi fine n'apporte pas énormément, car les deux problèmes sont d'une difficulté très similaire.

5.3.2 Le *blast*

Ce cas test est fait pour tester la robustesse du schéma en ce qui concerne les chocs les plus violents. Il provient de [11]. Le plasma est à nouveau confiné dans un carré de côté 1 et centré en $(0, 0)$. Il est constitué de deux zones, l'une centrale délimitée par un cercle de rayon $r = 0,1$ et l'autre comprenant le reste du domaine. Le plasma est uniformément statique et de densité $\rho = 1$. En ce qui concernela répartition, c'est elle qui crée les conditions favorables à l'apparition de chocs rapides car on passe de $p = 1000$ dans la zone centrale à $p = 0,1$ en-dehors, et ce de manière discontinue, sans zone tampon. L'ensemble du domaine est initialement plongé dans un champ magnétique lui aussi uniforme, dirigé suivant l'axe Ox et d'intensité $B_x = \dfrac{100}{\sqrt{4\pi}}$. Le rapport des capacités calorifiques γ est égal à $1,4$.

L'écoulement atteint les bords du domaine aux alentours de $t = 0,01$, qui est la date finale considérée. Les schémas explicites présentés ont échoué à résoudre ce problème avec des CFL supérieures ou égales à $0,2$. Au mieux, le schéma LLFS avec la méthode Runge-Kutta d'ordre 2 arrive à la moitié du temps imposé. On est ici confronté au manque de positivité du schéma RK2-LLFS. Les résultats de la figure 5.17 n'ont donc pu être obtenus que par le schéma implicite, au prix d'efforts de calcul importants. En effet, comme nous l'avons précisé au chapitre précédent, il est très difficile de faire converger la méthode de Newton. Cette solution est donc très longue à obtenir, et la convergence est relativement pauvre. Cela suffit néanmoins à obtenir ces résultats, sur un maillage 200×200.



(a) Densité

(b) Pression magnétique

(c) Vitesse

(d) Pression hydrodynamique

FIGURE 5.17 – Profils finaux du *blast*, à $t = 0,01$ s, obtenus avec le schéma LLF en implicite avec le schéma de Gear en temps

Nous ne sommes pas encore en mesure de présenter de cas 3D, non pas a priori à cause des faiblesses des schémas mais du fait d'un problème informatique lié à la construction du graphe des maillages générés par Gmsh, qu'il nous faut revoir. En ce qui concerne la MHD résistive, nous manquons de cas tests permettant de la tester. Il était intialement prévu de simuler une instabilité de tokamak appelée "trou de courant", mais nous manquons à présent de références pour savoir comment mettre en place ce problème, et les quelques-unes que nous avons trouvées font appel à des modèles réduits difficilement transposables au notre, celui de la MHD complète.

Chapitre 6

Conclusion

Sommaire

6.1 Conclusions

Durant cette thèse, nous avons adapté la plupart des techniques utilisées dans le contexte $\mathcal{RD}$ pour résoudre les équations de la MHD, idéale et résistive (même si cette dernière n'a pas été validée). Nous avons décidé de corriger les erreurs de divergence sur le champ magnétique à l'aide d'une technique récente, le *divergence cleaning* hyperbolique ([35]). Pour l'inclure dans nos algorithmes et comprendre son fonctionnement, nous avons dû redériver le système propre des équations, allant jusqu'à chercher à construire un système propre entropique, chose qui s'est avérée difficile à cause de la non-existence physique du multiplicateur de Lagrange qui sert de variable de correction, et qui empêche donc la définition d'une entropie compatible avec la physique. Nous nous sommes donc ramenés à un système propre qui est une extension de celui obtenu par Roe et Balsara ([82]). Nous avons ensuite inclus la correction de la divergence dans les méthodes implicites de façon originale, en résolvant cette partie comme un problème stationnaire pour évacuer à chaque itération en temps les erreurs de divergence hors du domaine. Cela revient à supprimer la dérivée en temps sur la variable de correction ψ (multiplicateur de Lagrange) durant la méthode de Newton. Malheureusement, la convergence des schémas $\mathcal{RD}$ instationnaires s'est révélée être le principal obstacle au bon fonctionnement de nos méthodes. Dans certains cas, il est venu s'ajouter à cela le fait que pour des cas très raides (comme le *blast* présenté plus haut), le nombre CFL ne devait pas être pris trop grand, sous peine de dégrader encore plus, et de manière drastique, la convergence de la méthode de Newton, voire la faire diverger. Après plusieurs essais infructueux (différences finies, régularisation du schéma) sur lesquels nous aurions peut-être pu aller plus loin en prenant le risque de ne déboucher sur rien, nous avons décidé de nous assurer que les schémas en eux-mêmes étaient opérationnels pour des problèmes instationnaires en passant en explicite. Pour cela, nous avons bénéficié du travail de M. Ricchiuto ([76]) qui a permis l'application des schémas de Runge-Kutta avec des schémas $\mathcal{RD}$ en espace. Concernant la correction de la divergence, nous avons dans un premier temps conservé l'approche hyperbolique classique, avec une équation d'évolution sur ψ comme sur les autres variables. S'en est suivie une période de réflexion où nous nous sommes demandés si dans l'idéal, il ne serait pas préférable de corriger parfaitement le champ magnétique à chaque pas de temps. Nous avons voulu reproduire l'idée de l'implicite à l'explicite, sans nous préoccuper du coût de calcul engendré. Il s'est avéré que ce dernier

était bien plus important qu'une méthode de projection, et ces méthodes ont donc été abandonnées. En parallèle de tout ceci, il y a bien sûr eu épisodiquement d'autres sujets d'intérêt, pour améliorer la positivité des schémas limités par exemple (ce qui n'a rien donné) ou encore un traitement différent et plus adapté des conditions aux limites.

Pour ma part (je parle pour la première et unique fois à la première personne), beaucoup de mes connaissances ont été acquises ou consolidées lors de la rédaction du présent mémoire. Celles-ci m'auraient sans doute été fort utiles pour avancer de façon plus cohérente, mais nous avons tout de même pu déterrer de nombreuses pistes, à défaut d'avoir su les transformer en résultats montrables pour certaines. Certaines idées qui n'ont pas abouti seraient sans doute à reconsidérer, mais leur rentabilité à court terme n'est pas toujours évidente. Nous allons en évoquer quelques-unes dans ce qui suit.

6.2 Perspectives

Les choses à faire ne manquent jamais. Mais parmi les problèmes non résolus, certains ne sont pas dus à l'introduction de la MHD et se posent aussi en mécanique des fluides. Parmi eux, on peut noter les problèmes de convergence des méthodes instationnaires implicites. Nous pensons que l'origine de cette difficulté vient de la mauvaise construction de la jacobienne du schéma, car celle-ci est très approximative et donc potentiellement en forte inadéquation avec le second membre. Pour le savoir, la voie directe serait une régularisation du schéma suivie d'une dérivation des jacobiennes exactes... ce qui s'annonce très lourd (dériver une matrice de vecteurs propres, par exemple, n'est pas une mince affaire !). Autrement, une alternative qui est en train de faire son chemin dans notre équipe est l'utilisation d'une méthode GMRES *matrix free*, dont l'épithète anglo-saxon permet de bien en saisir l'intérêt, puisque cela permettrait de passer outre le problème que nous évoquons. Si cette méthode s'avère efficace pour des problèmes de mécanique des fluides, il pourra sans doute être très intéressant de s'en servir en MHD. Elle repose en réalité sur une différenciation par différences finies, que nous avons eu l'occasion de tester pour construire nos jacobiennes (Adam Larat fut le premier à utiliser cette technique, cf. [59]). À cette époque, nous n'avions pas trouvé cette méthode assez robuste... même si cela est à nuancer quelque peu, car l'utilisation que nous en faisons repose sur des choix un peu arbitraires. À cet égard, le comportement des différences finies conserve encore une part d'oscurité.

Une autre amélioration possible est le passage aux ordres supérieurs, ne serait-ce que l'ordre 3, et à d'autres éléments finis. La méthodologie en mécanique des fluides stationnaire a été développée dans [59] (en ce qui concerne l'ordre très élevé), et nous disposons du schéma de Runge-Kutta d'ordre 3. Cela paraît donc faisable en "relativement" peu de temps. Sauf qu'il est très probable que le *mass lumping* utilisé sur les éléments P_1 soit une simplification trop grossière pour que la méthode fonctionne, et nous devrions dans ce cas avoir recours à des formules de quadrature judicieuses, à la manière de [29]. L'approche utilisée par ces auteurs est la seule alternative que nous ayons trouvée, et elle semble tout à fait appropriée pour nos schémas de type Runge-Kutta puisque la matrice de masse que nous devons inverser est strictement la même que celle de la méthode de Galerkin.

Une perspective encore plus proche est le passage aux géométries 3D. Aucun problème algorithmique ne se pose davantage en 3D qu'en 2D, du moins tant qu'on n'utilise que des tétraèdres. Cette étape pourrait être franchie en peu de temps. Il y a tout de même un risque, inconnu, de voir les problèmes de stabilité 2D prendre de l'ampleur en 3D. En particulier, il serait intéressant de voir dans quelle mesure la correction hyperbolique de la divergence suffit dans ces configurations. Il reste également des incertitudes quant à l'efficacité des conditions aux limites, qu'il serait bon d'élucider dans le même laps de temps.

Enfin, le principal problème auquel nous avons été confronté, est de concilier la résolution des chocs très violents et la limitation. Cette difficulté semble partagée par tous les schémas limités et n'est donc

pas l'apanage des schémas $\mathcal{RD}$. Si tout marche dans une bonne partie des cas, tant que les gradients ne sont que modérément importants voire un peu plus, il reste des situations extrêmes où aucun schéma limité ne fonctionne correctement. Le problème dit du *blast* que nous avons présenté en est un exemple, en cela qu'il contraint tellement la condition CFL du schéma explicite qu'il est nécessaire de le résoudre en implicite, malgré la lenteur de la convergence. Une première piste à laquelle nous avons pensé a été la construction d'un schéma sur le modèle du schéma de Lax-Friedrichs, mais avec des flux de Godounov plus complexes. Mais peu après avoir emprunté cette voie, nous avons réalisé qu'aucun flux n'était plus approprié que le Lax-Friedrichs pour assuer la positivité de la pression de la densité, et que notre quête avait donc de bonnes chances d'être vaine. D'autant que tout nouveau flux envisagé sera certainement limité pour atteindre l'ordre élevé, ce qui rend l'intérêt d'un tel effort discutable. Depuis, nous pensons à une autre alternative. La limitation ne sert qu'à atteindre l'ordre élevé en partant de schémas monotones mais d'ordre 1. Imaginons que nous arrivions à mettre en place le schéma sur des éléments d'ordre 6 ou plus. Une discontinuité (comme le front d'un choc par exemple) ne peut pas être représentée mieux qu'à l'ordre 2, ce qui signifie que toute représentation d'ordre supérieur n'apporterait rien. À défaut d'avoir de l'ordre élevé dans les chocs, on pourrait donc très bien imaginer que la limitation soit relachée, de sorte que dans les zones de choc, le schéma soit le schéma d'ordre 1 ou presque. Ainsi, la diffusion numérique opérant, le schéma pourrait être très robuste tout en restant asymptotiquement (en $h \to 0$) d'ordre 2. La limitation sur le cercle que nous avons développée semble toute indiquée pour ça, car il suffit, au lieu de projeter sur le cercle circonscrit de rayon $r = 1$, de projeter sur un cercle plus grand en augmentant simplement la valeur de r dans la formule. Cela signifie qu'on autorise par ce biais une plage plus large de coefficients β_i à rester inchangés. Nous avons commencé à tester cette option sans en avoir encore tiré de conclusions définitives... affaire à suivre.

Toutes ces questions font partie du cadre de la MHD idéale, mais il ne faut pas oublier la MHD résistive. Nous avons tout juste trouvé un cas de référence sur lequel valider notre approche, mais sans avoir ni toutes les données nécessaires ni encore les outils pour en critiquer les résultats (les problèmes sont généralement formulés dans le cadre de la MHD réduite, une simplification qui considère des variables telles que le courant ou les flux magnétiques, dont les relations avec nos variables ne sont pas triviales à exprimer numériquement). Il serait intéressant d'y consacrer davantage de temps et/ou de trouver des cas formulés avec nos variables (qui sont loin d'être exotiques) et bien documentés. À plus long terme, il serait bénéfique de modifier le schéma pour la MHD résistive en s'inspirant de [69] et des travaux de Guillaume Baurin, repris par Dante Desantis, qui sont en cours en mécanique des fluides, ceci afin d'obtenir une précision à peu près uniforme sur tous les cas tests.

Bibliographie

[1] R. Abgrall. Toward the ultimate conservative scheme : following the quest. *J. Comput. Phys.*, 167(2) :277–315, 2001.

[2] R. Abgrall. Essentially non-oscillatory Residual Distribution schemes for hyperbolic problems. *J. Comput. Phys.*, 214 :773–808, 2006.

[3] R. Abgrall and T. J. Barth. Residual distribution schemes for conservation laws via adaptive quadrature. *SIAM J. Sci. Comput.*, 24(3) :732–769, 2002.

[4] R. Abgrall and F. Marpeau. Residual distribution schemes on quadrilateral meshes. *J. Sci. Comput.*, 30(1) :131–175, 2007.

[5] R. Abgrall, K. Mer, and B. Nkonga. *A Lax-Wendroff type theorem for residual schemes*, pages 243–266. Innovative methods for numerical solutions of partial differential equations. World Scientific, 2002.

[6] R. Abgrall and M. Mezine. Construction of second order accurate monotone and stable residual distribution schemes for unsteady flow problems. *J. Comput. Phys.*, 188 :16–55, 2003.

[7] R. Abgrall and P. L. Roe. High order fluctuation schemes on triangular meshes. *J. Sci. Comput.*, 19(3) :3–36, 2003.

[8] R. Abgrall and J. Trefilick. An example of high order residual distribution scheme using non-Lagrange elements. *J. Sci. Comput.*, 45 :64–89, 2010.

[9] F. Assous, P. Degond, E. Heintze, P. A. Raviart, and J. Segre. On a finite-element method for solving the three-dimensional Maxwell equations. *J. Comput. Phys.*, 109 :222–237, 1993.

[10] D. S. Balsara. Linearized formulation of the Riemann problem for adiabatic and isothermal magnetohydrodynamics. *Astrophys. J. Suppl.*, 116 :119–131, 1998.

[11] D. S. Balsara and D. S. Spicer. A staggered mesh algorithm using high order Godunov fluxes to ensure solenoidal magnetic fields in magnetohydrodynamics simulations. *J. Comput. Phys.*, 149 :270–292, 1999.

[12] T. J. Barth. An energy look at the N scheme. Working notes, 1996.

[13] T. J. Barth. *Numerical Methods for Gasdynamic Systems on Unstructured Meshes*, volume 5 of *An Introduction to Recent Developments in Theory and Numerics for Conservation Laws, Lecture Notes in Computational Science and Engineering*, chapter 5, pages 195–. Springer-Verlag, 1999.

[14] J. Bazer and W. B. Ericson. Hydromagnetic shocks. *Astrophys. J.*, 129 :758–785, 1958.

[15] M. Bergot, G. Cohen, and M. Duruflé. Higher-order finite elements for hydrid meshes using new nodal pyramidal elements. *J. Sci. Comput.*, 42 :345–381, 2010.

[16] J. P. Boris. Relativistic plasma simulations - Optimization of a hybrid code. In *Proceedings of the 4th Conference on Numerical Simulation of Plasmas*, pages 3–67, Washington, D.C., November 1970. Naval Res. Lab.

[17] J. U. Brackbill. Fluid modeling of magnetized plasmas. *Space Sci. Rev.*, 42 :153–167, 1985.

[18] J. U. Brackbill and D. C. Barnes. Note : The effect of nonzero $\nabla \cdot B$ on the numerical solution of the magnetohydrodynamic equations. *J. Comput. Phys.*, 35 :426–430, 1980.

[19] M. Brio and C. C. Wu. An upwind differencing scheme for the equations of ideal magnetohydrodynamics. *J. Comput. Phys.*, 75 :400–422, 1988.

[20] A. N. Brooks and T. J. R. Hughes. Streamline upwind/petrov-galerkin formulations for convection dominated flows with particular emphasis on the incompressible navier-stokes equations. *Comput. Meth. Appl. Mech. Eng.*, 32 :199–259, 1982.

[21] J. C. Butcher. *Numerical Methods for Ordinary Diffrential Equations. Second edition.* John Wiley & Sons, Ltd., 2008.

[22] D. Caraeni and L. Fuchs. Compact third-order multidimensional upwind scheme for Navier-Stokes simulations. *Theoretical and Computational Fluid Dynamics*, 15(6) :373–401, 2002.

[23] J.C. Carette, H. Deconinck, H. Paillère, and P. L. Roe. Multidimensional upwinding - its relation to finite elements. *Int. J. Numer. Meth. Fl.*, 20(8) :935–955, 1995.

[24] P. Cargo and G. Gallice. Roe matrices for ideal MHD and systematic construction of Roe matrices for systems of conservation laws. *J. Comput. Phys.*, 136 :446–466, 1997.

[25] S. Chapman and T. G. Cowling. *The mathematical theory of non-uniform gases : an account of the kinetic theory of viscosity, thermal conduction and diffusion in gases.* Cambridge University Press, 1990.

[26] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems.* North Holland, Amsterdam, 1978.

[27] P. G. Ciarlet and P.-A. Raviart. Interpolation theory over curved elements, with applications to finite element methods. 1 :217–249, 1972.

[28] G. Cohen. *High Order Numerical Methods for Transient Wave Equations.* Springer-Verlag, Berlin, 2001.

[29] G. Cohen, P. Joly, J. E. Roberts, and N. Tordjman. Higher order triangular finite elements with mass lumping for the wave equation. *SIAM J. Numer. Anal.*, 38(6) :2047–2078, 2001.

[30] Á. Csík. *Upwind residual distribution schemes for general hyperbolic conservation laws and application to ideal magnetohydrodynamics.* PhD thesis, Katholieke Universiteit Leuven, Faculteit Wetenschappen Centrum voor Plasma-Astrofysica, Belgium, 2002.

[31] Á. Csík, M. Ricchiuto, and H. Deconinck. A conservative formulation of the multidimensional upwind Residual Distribution schemes for general conservation laws. *J. Comput. Phys.*, 179 :286–312, 2002.

[32] H. De Sterck, B. C. low, and S. Poedts. Complex magnetohydrodynamic bow shock topology in field-aligned low-β flow around a perfectly conducting cylinder. *Phys. Plasmas*, 5(11) :4015–4027, 1998.

[33] H. Deconinck and M. Ricchiuto. *Residual Distribution Schemes : Foundations and Analysis*, volume 3 of *Encyclopedia of Computational Mechanics*, chapter 19. John Wiley & Sons, Ltd, 2007.

[34] H. Deconinck, M. Ricchiuto, and K. Sermeus. Introduction to residual distribution schemes and comparison with stabilized finite elements. In H. Deconinck, editor, *VKI LS 2003-05, 33rd Computational Fluid dynamics Course.* von Karman Institute for Fluid Dynamics, 2003.

[35] A. Dedner, F. Kemm, D. Kröner, C. D. Munz, T. Schnitzer, and M. Wesenberg. Hyperbolic divergence cleaning for the MHD equations. *J. Comput. Phys.*, 175 :645–673, 2002.

[36] B. Després. *Lois de Conservation Eulériennes, Lagrangiennes et Méthodes Numériques.* Springer, 2010.

[37] M. Dumbser, D. S. Balsara, E. F. Toro, and C.-D. Munz. A unified framework for the construction of one-step finite volume and discontinuous Galerkin schemes on unstructured meshes. *J. Comput. Phys.*, 227 :8209–8253, 2008.

[38] C. R. Evans and J. F. Hawley. Simulation of magnetohydrodynamic flows : a constrained transport method. *Astrophys. J.*, 332 :659–677, 1988.

[39] L. Ferracina and M. N. Spijker. An extension and analysis of the Shu-Osher representation of Runge-Kutta methods. *Math. Comp.*, 74 :201–219, 2004.

[40] R. Feynman, R. Leighton, and M. Sands. *The Feynman Lectures on Physics*, volume 2. Addison-Wesley, 1964.

[41] L. Fezoui and B. Stoufflet. A class of implicit upwind schemes for Euler simulations with unstructured meshes. *J. Comput. Phys.*, 84 :174–, 1989.

[42] S. K. Godunov. A finite difference method for the numerical computation of discontinuous solutions of the equations of fluid dynamics. *Math. Sbornik*, 47 :271, 1959.

[43] S. K. Godunov. Symmetric form of the equations of magnetohydrodynamics. In *Numerical Methods for Mechanics of Continuum Medium*, volume 1, pages 26–31. 1972.

[44] H. Goedbloed and S. Poedts. *Principles of Magnetohydrodynamics : With Applications to Laboratory and Astrophysical Plasmas.* Cambridge University Press, 2004.

[45] S. Gottlieb and C.-W. Shu. Total-variation-diminishing Runge-Kutta schemes. *Math. Comp.*, 67 :73–85, 1998.

[46] S. Gottlieb, C.-W. Shu, and E. Tadmor. Strong-stability-preserving high-order time discretization methods. *SIAM Review*, 43 :89–112, 2001.

[47] A. Harten. On the symmetric form of systems of conservation laws with entropy. *J. Comput. Phys.*, 49 :151–164, 1983.

[48] J. S. Hesthaven. From electrostatics to almost optimal nodal sets for polynomial interpolation in a simplex. *SIAM J. Numer. Anal.*, 35(2) :655–676, 1998.

[49] I. Higueras. Representations of Runge-Kutta methods and strong stability preserving methods. *SIAM J. Numer. Anal.*, 43(3) :924–948, 2006.

[50] J. Hsu and A. Jameson. An implicit-explicit scheme for calculating complex unsteady flows. 40th AIAA Aerospace Sciences Meeting and Exhibit, 2002.

[51] T. J. R. Hughes, L. P. Franca, and G. M. Hulbert. A new finite element formulation for computational fluid dynamics : VIII. The Galerkin/least-squares method for advective-diffusive equations. *Comput. Meth. Appl. Mech. Eng.*, 73 :173–189, 1989.

[52] T. J. R. Hughes, L. P. Franca, and M. Mallet. A new finite element formulation for computational fluid dynamics : I. symmetric forms of the compressible Euler and Navier-Stokes equations and the second law of thermodynamics. *Comput. Meth. Appl. Mech. Engin.*, 54(2) :223–234, 1986.

[53] G. Huysmans, S. Pamela, E. Van Der Plas, and P. Ramet. Non-linear MHD simulations of Edge Localized Modes. In *36th EPS Plasma Physics Conference, Sofia, Bulgarie*, 2009.

[54] A. Jameson. Eigenvalues, eigenvectors and symmetrization of the magneto-hydrodynamics (mhd) equations. Presented at AFOSR Grantees and Contractors Meeting, August 7, 2006, Atlanta, GA.

[55] A. Jeffrey and T. Tanuiti. *Non-linear wave propagation : with applications to physics and magneto-hydrodynamics.* Academic Press, New York, 1964.

[56] S. Jund. *Méthodes d'éléments finis d'ordre élevé pour la simulation numérique de la propagation d'onde.* PhD thesis, Institut de Recherche Mathématique Avancée, Université Louis Pasteur, Strasbourg, 2007.

[57] S. Jund and S. Salmon. Arbitrary high-order finite element schemes and high-order mass lumping. *Int. J. Appl. Math. Comput. Sci.*, 17(3) :375–393, 2007.

[58] S. N. Kruzkov. Generalized solutions of the Cauchy problem in the large for nonlinear equations of first order. *Dokl. Akad. Nauk. SSSR*, 187 :29–32, 1969. (English transl.).

[59] A. Larat. *Conception and analysis of very high order distribution schemes. Application to fluid mechanics.* PhD thesis, INRIA Bordeaux Sud-Ouest et Université Bordeaux I, 2009.

[60] P. D. Lax. Hyperbolic systems of conservation laws II. *Comm. Pure and Appl. Math.*, 10 :536–566, 1957.

[61] B. Marder. A method for incorporating Gauss' law into electromagnetic PIC codes. *J. Comput. Phys.*, 68 :48–55, 1987.

[62] M. Mezine. *Conception de schémas distributifs pour l'aérodynamique stationnaire et instationnaire.* PhD thesis, Université Bordeaux I, 2002.

[63] C. D. Munz, P. Omnes, R. Scheider, E. Sonnendrücker, and U. Voss. Divergence correction techniques for Maxwell solvers based on a hyperbolic model. *J. Comput. Phys.*, 161(2) :484–511, 2000.

[64] C. D. Munz, R. Scheider, E. Sonnendrücker, and U. Voss. Maxwell's equations when the charge conservation is not satisfied. *C. R. Acad. Sci. Paris*, 328 :431–436, 1999.

[65] R. H. Ni. A multiple grid scheme for solving the Euler equations. *AIAA J.*, 20 :1565–1571, 1981.

[66] D. E. Nielsen and A. T. Drobot. An analysis and optimization of the pseudo-current method. *J. Comput. Phys.*, 89 :31–40, 1990.

[67] H. Nishikawa. A First-Order System Approach for Diffusion Equation. I : Second-Order Residual-Distribution Schemes. *J. Comput. Phys.*, 227 :315–352, 2007.

[68] H. Nishikawa. A First-Order System Approach for Diffusion Equation. II : Unification of Advection and Diffusion. *J. Comput. Phys.*, 229 :3989–4016, 2010.

[69] H. Nishikawa and P. L. Roe. On high-order fluctuation-splitting schemes for Navier-Stokes equations. In *Computational Fluid Dynamics 2004, Proceedings of the Third International Conference on Computational Fluid Dynamics, ICCFD3, Toronto, 12–16 July 2004*, pages 799–804. Springer Berlin Heidelberg, 2006.

[70] H. Paillère. *Multidimensional Upwind Residual Distribution Schemes for the Euler and Navier-Stokes Equations on Unstructured Grids.* PhD thesis, Université Libre de Bruxelles, 1995.

[71] S. Pamela. *Simulation Magnétohydrodynamique des Edge Localized Modes dans un tokamak.* PhD thesis, Université de Provence, 2010.

[72] B. Perthame and C.-W. Shu. On positivity preserving finite volume schemes for Euler equations. *Numer. Math.*, 73 :119–130, 1996.

[73] M. Polner. *Galerkin Least-Squares Stabilization Operators for the Navier-Stokes Equations - A Unified Approach.* PhD thesis, University of Twente, 2005.

[74] K. G. Powell. An approximate Riemann solver for magnetohydrodynamics (that works in more than one dimension). Technical report, ICASE Report No 94-24, NASA Langley Research Center, VA, 1994.

[75] M. Ricchiuto. *Construction and analysis of compact residual discretizations for conservation laws on unstructured meshes*. PhD thesis, von Karman Institue for Fluid Dynamics et Université Libre de Bruxelles, 2005.

[76] M. Ricchiuto and R. Abgrall. Explicit Runge-Kutta residual-distribution schemes for time dependent problems. *J. Comput. Phys.*, 229(16) :5653–5691, 2010.

[77] P. L. Roe. Approximate Riemann solvers, parameter vectors, and difference schemes. *J. Comput. Phys.*, 43 :357–372, 1981.

[78] P. L. Roe. *Flucuations and signals - a framework for numerical evolution problems*, pages 219–257. Numerical Methods for Fluid Dynamics. Academic Press, 1982.

[79] P. L. Roe. Fluctuations and signals, a framework for numerical evolution problems. In *Numerical Methods for Fluid Dynamics*. 1984.

[80] P. L. Roe. Characteristic-based schemes for the Euler equations. *Ann. Rev. Fluid Mech.*, 18 :337–365, 1986.

[81] P. L. Roe. Linear advection schemes on triangular meshes. Technical Report CoA 8720, Cranfield Institue of Technology, 1987.

[82] P. L. Roe and D. S. Balsara. Notes on the eigensystem of magnetohydrodynamics. *SIAM J. Appl. Math.*, 56(1) :57–67, 1996.

[83] P. L. Roe and D. Sidilkover. Optimum positive linear schemes for advection in two and three dimensions. *SIAM J. Numer. Anal.*, 29(6) :1542–1568, 1992.

[84] S. J. Ruuth and R. J. Spiteri. Two barriers on strong-stability-preserving time discretization methods. *J. Sci. Comput.*, 17 :211–220, 2002.

[85] D. Serre. *Systèmes de lois de conservation*, volume I. Diderot, 1996.

[86] C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *J. Comput. Phys.*, 77 :439–471, 1988.

[87] R. J. Spiteri and S. J. Ruuth. A new class of optimal high-order strong-stability-preserving time discretization methods. *SIAM J. Numer. Anal.*, 40 :469–491, 2002.

[88] R. Struijs. *A multi-dimensional upwind discretization method for the Euler equations on unstructured grids*. PhD thesis, Technische Universiteit Delft, 1994.

[89] E. Tadmor. A minimum entropy principle in the gas dynamics equations. *Appl. Numer. Math.*, 2(3-5) :211–219, 1986.

[90] C. Tavé. *Construction simple de schémas distribuant le résidu non-oscillants et d'ordre élevé pour la simulation d'écoulements stationnaires sur maillages triangulaires et hybrides*. PhD thesis, INRIA Futurs et Université Bordeaux I, 2007.

[91] T. E. Tezduyar and M. Senga. Stabilization and schock-capturing parameters in SUPG formulation of compressible flows. *Comput. Meth. Appl. Mech. Eng.*, 195 :1621–1632, 2006.

[92] E. F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics : A Practical Introduction*. Springer-Verlag, 3rd edition, 2009.

[93] M. Torrilhon. Exact solver and uniqueness conditions for Riemann problems of ideal magnetohydrodynamics. Technical report, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2002.

[94] M. Torrilhon. Uniqueness conditions for Riemann problems of ideal magnetohydrodynamics. *J. Plasma Physics*, 69 :253–276, 2003.

[95] G. Tóth. The $\nabla \cdot B = 0$ constraint in shock-capturing magnetohydrodynamics codes. *J. Comput. Phys.*, 161 :605–652, 2000.

[96] E. van der Weide and H. Deconinck. Positive matrix distribution schemes for hyperbolic systems, with application to the Euler equations. In *Computational Fluid Dynamics '96*, pages 747–753. ECCOMAS computational fluid dynamics conference, John Wiley & Sons, 1996.

[97] N. Villedieu. *High order discretization by Residual Distribution schemes.* PhD thesis, von Karman Institue for Fluid Dynamics et Université Libre de Bruxelles, 2009.

[98] C. Wervaecke. *Simulation d'écoulements turbulents compressibles par une méthode d'éléments finis stabilisés.* PhD thesis, INRIA Bordeaux Sud-Ouest et Université Bordeaux I, 2010.

Annexe A

Suppléments du problème continu

A.1 Formulaire d'analyse vectorielle

L'opérateur de dérivation vectorielle $\overrightarrow{\nabla}$ est une notation simplifiant l'écriture des trois opérateurs différentiels vectoriels classiques que sont le gradient, la divergence et le rotationnel. En coordonnées cartésiennes dans $\mathbb{R}^3$, on peut écrire :

$$\overrightarrow{\nabla} = \begin{pmatrix} \dfrac{\partial}{\partial x} \\ \dfrac{\partial}{\partial y} \\ \dfrac{\partial}{\partial z} \end{pmatrix}$$

Ceci permet de définir les trois opérateurs précédemment cités, en prenant un scalaire A et un vecteur $\overrightarrow{B} \in \mathbb{R}^3$ ayant les conditions de régularité suffisantes, et en utilisant les opérations vectorielles classiques dans l'espace euclidien $\mathbb{R}^3$:

$$\overrightarrow{\nabla} A = \overrightarrow{\mathrm{grad}}\,(A) = \begin{pmatrix} \dfrac{\partial A}{\partial x} \\ \dfrac{\partial A}{\partial y} \\ \dfrac{\partial A}{\partial z} \end{pmatrix}$$

$$\overrightarrow{\nabla} \cdot \overrightarrow{B} = \mathrm{div}\left(\overrightarrow{B}\right) = \frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z}$$

$$\overrightarrow{\nabla} \wedge \overrightarrow{B} = \overrightarrow{\mathrm{rot}}\left(\overrightarrow{B}\right) = \begin{pmatrix} \dfrac{\partial B_z}{\partial y} - \dfrac{\partial B_y}{\partial z} \\ \dfrac{\partial B_x}{\partial z} - \dfrac{\partial B_z}{\partial x} \\ \dfrac{\partial B_y}{\partial x} - \dfrac{\partial B_x}{\partial y} \end{pmatrix}$$

Ces règles ne sont valables que dans une base orthonormée fixe comme celle du repérage cartésien. Pour des bases non fixes comme un repérage cylindrique ou sphérique, les notations sont conservées mais le lien avec les opérations classiques ne se fait que par un changement de variables avec les coordonnées

cartésiennes. Grâce à la correspondance entre ces bases, les formules que nous allons maintenant présenter sont vraies dans chacune d'elles.

Soient $\alpha \in \mathbb{R}$ et $\left(\overrightarrow{A}, \overrightarrow{B}\right) \in \mathbb{R}^3 \times \mathbb{R}^3$, alors :

$$\overrightarrow{\nabla} \cdot \left(\alpha \overrightarrow{A}\right) = \alpha \left(\overrightarrow{\nabla} \cdot \overrightarrow{A}\right) + \left(\overrightarrow{\nabla}\alpha\right) \cdot \overrightarrow{A} \tag{A.1}$$

$$\overrightarrow{\nabla} \wedge \left(\alpha \overrightarrow{A}\right) = \alpha \left(\overrightarrow{\nabla} \wedge \overrightarrow{A}\right) + \left(\overrightarrow{\nabla}\alpha\right) \wedge \overrightarrow{A} \tag{A.2}$$

$$\overrightarrow{\nabla} \left(\overrightarrow{A} \cdot \overrightarrow{B}\right) = \left(\overrightarrow{A} \cdot \overrightarrow{\nabla}\right) \overrightarrow{B} + \overrightarrow{A} \wedge \left(\overrightarrow{\nabla} \wedge \overrightarrow{B}\right) + \left(\overrightarrow{B} \cdot \overrightarrow{\nabla}\right) \overrightarrow{A} + \overrightarrow{B} \wedge \left(\overrightarrow{\nabla} \wedge \overrightarrow{A}\right) \tag{A.3}$$

$$\overrightarrow{\nabla} \cdot \left(\overrightarrow{A} \wedge \overrightarrow{B}\right) = -\overrightarrow{A} \cdot \left(\overrightarrow{\nabla} \wedge \overrightarrow{B}\right) + \overrightarrow{B} \cdot \left(\overrightarrow{\nabla} \wedge \overrightarrow{A}\right) \tag{A.4}$$

$$\overrightarrow{\nabla} \wedge \left(\overrightarrow{A} \wedge \overrightarrow{B}\right) = \overrightarrow{A} \left(\overrightarrow{\nabla} \cdot \overrightarrow{B}\right) - \left(\overrightarrow{A} \cdot \overrightarrow{\nabla}\right) \overrightarrow{B} - \overrightarrow{B} \left(\overrightarrow{\nabla} \cdot \overrightarrow{A}\right) + \left(\overrightarrow{B} \cdot \overrightarrow{\nabla}\right) \overrightarrow{A} \tag{A.5}$$

De plus, on peut composer les opérateurs et obtenir les trois propriétés remarquables suivantes :

$$\overrightarrow{\nabla} \wedge \left(\overrightarrow{\nabla}\alpha\right) = \overrightarrow{0} \tag{A.6}$$

$$\overrightarrow{\nabla} \cdot \left(\overrightarrow{\nabla} \wedge \overrightarrow{A}\right) = 0 \tag{A.7}$$

$$\overrightarrow{\nabla} \wedge \left(\overrightarrow{\nabla} \wedge \overrightarrow{A}\right) = \overrightarrow{\nabla} \left(\overrightarrow{\nabla} \cdot \overrightarrow{A}\right) - \left(\overrightarrow{\nabla} \cdot \overrightarrow{\nabla}\right) \overrightarrow{A} \tag{A.8}$$

L'opérateur $\left(\overrightarrow{\nabla} \cdot \overrightarrow{\nabla}\right)$ est également appelé le laplacien et est généralement noté Δ. En coordonnées cartésiennes, il s'écrit :

$$\Delta\alpha = \frac{\partial^2 \alpha}{\partial x^2} + \frac{\partial^2 \alpha}{\partial y^2} + \frac{\partial^2 \alpha}{\partial z^2}$$

Il peut également s'appliquer à un vecteur comme dans (A.8), auquel cas il suffit de reprendre la formule précedente et de l'appliquer à chaque composante du vecteur. Ceci peut se retrouver en appliquant successivement (2.1.9) et (2.1.8). Pour finir, on rappelle également quelques propriétés du produit vectoriel et du produit mixte. Soient trois vecteurs $\overrightarrow{A}$, $\overrightarrow{B}$ et $\overrightarrow{C}$ dans $\mathbb{R}^3$, alors :

$$\overrightarrow{A} \wedge \overrightarrow{B} = -\overrightarrow{B} \wedge \overrightarrow{A} \tag{A.9}$$

$$\overrightarrow{A} \wedge \left(\overrightarrow{B} \wedge \overrightarrow{C}\right) = \left(\overrightarrow{A} \cdot \overrightarrow{C}\right) \overrightarrow{B} - \left(\overrightarrow{A} \cdot \overrightarrow{B}\right) \overrightarrow{C} \tag{A.10}$$

$$\left(\overrightarrow{A} \wedge \overrightarrow{B}\right) \cdot \overrightarrow{C} = -\left(\overrightarrow{A} \wedge \overrightarrow{C}\right) \cdot \overrightarrow{B} = \left(\overrightarrow{B} \wedge \overrightarrow{C}\right) \cdot \overrightarrow{A} \tag{A.11}$$

Pour le produit mixte (A.11), toute permutation de deux vecteurs dans le membre de gauche change le signe de l'opération globale. Toutes ces propriétés sont retrouvables sur Wikipédia [1], ou dans [40] et [44].

1. http://fr.wikipedia.org/wiki/Analyse_vectorielle et http://fr.wikipedia.org/wiki/Produit_vectoriel

A.2 Dérivation du jacobien pour le changement de variables

Rappelons d'abord l'expression du jacobien :

$$\boldsymbol{J}\left(\overrightarrow{X},t\right)=\det\left(\frac{\partial\overrightarrow{f}}{\partial\overrightarrow{X}}\left(\overrightarrow{X},t\right)\right)$$

Il nous faut calculer la dérivée en temps de $\boldsymbol{J}$. Soit $h\in\mathbb{R}^+$, alors si $\forall\overrightarrow{X}\in\mathbb{R}^3$ fixé, nous notons la matrice $A(t)=\frac{\partial\overrightarrow{f}}{\partial\overrightarrow{X}}\left(\overrightarrow{X},t\right)$, nous pouvons écrire :

$$\begin{aligned} A(t+h) &= A(t)+A'(t)h+\mathrm{o}(h) \\ \Rightarrow\quad \det(A(t+h)) &= \det(A(t)+A'(t)h)+\mathrm{o}(h) \end{aligned}$$

Nous avons dit dans la section 2.1.2 que $\overrightarrow{f}$ était un $\mathcal{C}^\infty$-difféomorphisme, donc A est toujours inversible.

$$\Rightarrow\quad \det(A(t+h)) = \det(A(t))\left(\det(I+A^{-1}(t)A'(t)h)\right)+\mathrm{o}(h)$$

I est la matrice idéntité. La matrice $A^{-1}(t)A'(t)h$ est une matrice dont tous les coefficients sont proportionnels au "petit" paramètre h. Prenons une matrice H de petits paramètres quelconques et voyons l'accroissement du déterminant en l'identité, i.e. $\det(I+H)$. On va noter I_k et H_k respectivement la k-ième colonne de I et H, et h_{max} le plus grand coefficient de H.

$$\begin{aligned} \det(I+H) &= \det\left((I_k+H_k)_{1\leqslant k\leqslant n}\right) \\ \Rightarrow\quad \det(I+H) &= \det(I_1,I_2+H_2,\ldots,I_n+H_n)+\det(H_1,I_2+H_2,\ldots,I_n+H_n) \\ &\ldots \quad \text{(on répète le processus jusqu'au bout)} \\ \Rightarrow\quad \det(I+H) &= \det(I)+\sum_k \det(I_1,\ldots,I_{k-1},H_k,I_{k+1},\ldots,I_n)+\mathrm{o}(h_{max}) \\ &= 1+\sum_k D_k+\mathrm{o}(h_{max}) \end{aligned}$$

Les déterminants D_k sont formés par $n-1$ colonnes de I et une colonne de H. Si on développe chaque D_k suivant la k-ième ligne, on obtient immédiatement que $D_k=(-1)^{2k}H_{kk}det(I')=H_{kk}$ où I' est la matrice identité de taille $n-1$.

$$\begin{aligned} \det(I+H) &= 1+\sum_k H_{kk}+\mathrm{o}(h_{max}) \\ \Rightarrow\quad \det(I+H) &= 1+\mathrm{Tr}(H)+\mathrm{o}(h_{max}) \end{aligned}$$

Si nous reprenons maintenant $H=A^{-1}(t)A'(t)h$:

$$\begin{aligned} \det(A(t+h)) &= \det(A(t))\left(1+\mathrm{Tr}(A^{-1}(t)A'(t)h)\right)+\det(A(t))\mathrm{o}(h) \\ \Rightarrow\quad \det(A(t+h)) &= \det(A(t))+h\det(A(t))\,\mathrm{Tr}(A^{-1}(t)A'(t))+\mathrm{o}(h) \\ \Rightarrow\quad \frac{\det(A(t+h))-\det(A(t))}{h} &= \det(A(t))\,\mathrm{Tr}(A^{-1}(t)A'(t))+\mathrm{O}(h) \end{aligned}$$

On peut à présent passer à la limite $h \to 0$ et revenir au problème initial, car on vient de montrer que :

$$\frac{\partial \boldsymbol{J}}{\partial t} = \boldsymbol{J}\,\mathrm{Tr}\left(\left(\frac{\partial \overrightarrow{f}}{\partial \overrightarrow{X}}\right)^{-1}\frac{\partial}{\partial t}\left(\frac{\partial \overrightarrow{f}}{\partial \overrightarrow{X}}\right)\right)$$

$\overrightarrow{f}(\overrightarrow{X},t)$ peut être noté abusivement $\overrightarrow{x}(t)$, auquel cas on peut écrire de façon plus lisible :

$$\frac{\partial \boldsymbol{J}}{\partial t} = \boldsymbol{J}\,\mathrm{Tr}\left(\left(\frac{\partial \overrightarrow{X}}{\partial \overrightarrow{x}}\right)\frac{\partial}{\partial \overrightarrow{X}}\left(\frac{\partial \overrightarrow{f}}{\partial t}\right)\right)$$

où on a utilisé le théorème de Schwarz. Rappelons que la vitesse Eulérienne $\overrightarrow{u}(\overrightarrow{x},t)$ est assimilable en espace à la vitesse Lagrangienne $\partial_t \overrightarrow{f}(\overrightarrow{X},t)$, et nous obtenons finalement :

$$\begin{aligned}
\frac{\partial}{\partial t}\boldsymbol{J} &= \boldsymbol{J}\,\mathrm{Tr}\left(\overrightarrow{\nabla}_x \overrightarrow{X}\,\overrightarrow{\nabla}_X \overrightarrow{u}\right) \\
\Rightarrow \quad \frac{\partial}{\partial t}\boldsymbol{J} &= \boldsymbol{J}\,\mathrm{Tr}\left(\overrightarrow{\nabla}_x \overrightarrow{u}\right) \\
\Rightarrow \quad \frac{\partial}{\partial t}\boldsymbol{J} &= \boldsymbol{J}\overrightarrow{\nabla}\cdot\overrightarrow{u}
\end{aligned}$$

Annexe B

Commentaires sur la stabilité entropique

Sommaire

B.1 Formulation discrète

Nous rappelons un résultat que nous avons établi à la fin de la section 3.2.2 du chapitre 3. Le problème continu muni d'une entropie mathématique S vérifie, en tout point d'espace et de temps :

$$\left\langle W, \frac{\partial U}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(U) \right\rangle_{\mathbb{R}^p} \leq 0 \tag{B.1}$$

où $W = \nabla_U S$ désigne les variables entropiques et p le nombre de variables. Cette inégalité est équivalente à celle de Clausius-Duhem tant que le couple entropie-flux $(S, \overrightarrow{G}(S))$ est compatible avec les lois de conservation au sens de (2.3.7). Le but de cette section est de passer (B.1) au niveau discret pour y faire apparaître le schéma, et donc en déduire une contrainte qu'on puisse requérir à son encontre. Le premier impératif est donc de rester consistant avec la forme continue. Pour commencer, on ne travaille pas avec U mais avec une approximation consistante U_h, pour nous son interpolée de Lagrange. Comme W dépend de U, il faut également en définir une approximation consistante W_h. La précision détermine l'écart à l'inégalité de Clausius-Duhem. Si on consdère des projections de W et de U sur des espaces de polynômes de degrés différents, disons l et k respectivement, on obtient :

$$\left\langle \pi_h^l(W), \frac{\partial \pi_h^k(U)}{\partial t} + \overrightarrow{\nabla} \cdot \overrightarrow{F}(\pi_h^k(U)) \right\rangle_{\mathbb{R}^p} \leq 0 + O\left(h^{l+1+k+1}\right) \tag{B.2}$$

À partir de maintenant on utilisera la notation $\langle . \rangle$ sans ambiguïté pour désigner le produit scalaire $\langle . \rangle_{\mathbb{R}^p}$.

Par souci de clarté, nous ne procèderons pas ici à l'étude du schéma en temps, et donc nous ne discrétisons pas en temps à ce stade. Une fois l'interpolation faite, il faut inscrire l'expression précédente dans la formulation variationnelle que nous résolvons. À partir de maintenant, nous ferons appel aux notations et aux concepts du chapitre 3 de manière systématique. Avant de poursuivre, il est à noter que dans l'expression précédente, même avec une approximation de W constante par morceaux ($l = 0$), l'inégalité (B.1) est vérifiée avec une erreur d'ordre $k+2$ (soit une précision supérieure à celle du schéma !).

C'est appréciable, mais nous n'avons pas besoin d'une telle précision dans la mesure où nous ne cherchons pas à résoudre ce système, mais seulement à formuler une contrainte qui soit consistante avec (B.1). Si nous prenons $l = k$ et que la projection π_h se fait sur la base des polynômes de Lagrange φ_i, (B.2) devient :

$$\sum_{i\in\Omega_h} \left\langle W_i\varphi_i, \frac{\partial U_h}{\partial t} + \overrightarrow{\nabla}\cdot\overrightarrow{F}(U_h)\right\rangle \leq O\left(h^{2k+2}\right)$$

La contrainte en question consiste à dire que le schéma sera plus efficace et correspondra davantage à nos attentes (voir les sections 2.3.3 et 3.2.2 pour l'intérêt du respect de ce type d'inégalité) si on requiert la négativité de l'expression précédente quelle que soit l'erreur commise lors de la discrétisation, c'est-à-dire si on impose que le second membre soit strictement nul et non simplement "petit". Il faut de plus qu'elle soit formulée dans un sens faible, comme le problème que nous résolvons, ce qui est fait simplement en intégrant sur le domaine de travail car on obtient alors, φ_i étant une application scalaire réelle (donc trivialement autoadjointe) :

$$\sum_{i\in\Omega_h} \left\langle W_i, \int_{\Omega_h} \varphi_i \left(\frac{\partial U_h}{\partial t} + \overrightarrow{\nabla}\cdot\overrightarrow{F}(U_h)\right)\right\rangle dV \leq 0 \tag{B.3}$$

C'est à ce moment-là de l'analyse qu'il faut se demander dans quelle mesure la formulation variationnelle que nous résolvons vérifie cette inégalité. Il est intéressant de remarquer que le second membre du produit scalaire est exactement une méthode de Galerkin :

$$\forall i \in \Omega_h,\ \int_{\Omega_h} \varphi_i \left(\frac{\partial U_h}{\partial t} + \overrightarrow{\nabla}\cdot\overrightarrow{F}(U_h)\right) dV = 0$$

Comme noté dans [52], cela signifie que la méthode de Galerkin vérifie toujours une égalité d'entropie discrète :

$$\sum_{i\in\Omega_h} \left\langle W_i, \int_{\Omega_h} \varphi_i \left(\frac{\partial U_h}{\partial t} + \overrightarrow{\nabla}\cdot\overrightarrow{F}(U_h)\right)\right\rangle = 0$$

L'entropie est correctement transportée mais dès qu'une discontinuité est rencontrée, le schéma ne rend pas compte d'une augmentation d'entropie physique. C'est la raison pour laquelle, intrinsèquement, aucune méthode de Galerkin n'est capable d'appréhender les chocs.

Cependant, la formulation variationnelle que nous résolvons n'est pas de type Galerkin mais Petrov-Galerkin. Un des avantages de cette dernière est que l'égalité d'entropie n'est plus vraie, et on peut donc bel et bien espérer obtenir une inégalité. Nous résolvons en fait en chaque degré de liberté $i \in \Omega_h$ la formulation (3.6.5), i.e. :

$$\int_{\Omega_h} (\varphi_i I_p + \gamma_i) \left(\frac{\partial U_h}{\partial t} + \overrightarrow{\nabla}\cdot\overrightarrow{F}(U_h)\right) dV + \int_{\partial\Omega} (\varphi_i I_p + \gamma_i) \left(\overrightarrow{H}(U_h) - \overrightarrow{F}(U_h)\right)\cdot\overrightarrow{n}\, d\partial\Omega = 0 \tag{B.4}$$

où I_p est la matrice identité de taille p. Cette écriture rappelle au passage que chaque variable est interpolée via les mêmes polynômes, ceux de Lagrange, mais que les fonctions test (à travers les γ_i) sont des matrices, pleines a priori et continues par élément seulement (puisque tout schéma $\mathcal{RD}$ est défini par élément). Dans (B.4), on définit $\gamma_i = 0$ au-delà des premiers voisins de i, comme pour les polynômes de Lagrange φ_i. Tandis que pour tout élément $E \subset \mathcal{T}_i$, la matrice γ_i^E est définie à l'aide des matrices de distribution β_i^E comme $\gamma_i^E = \beta_i^E - \varphi_i^E$. L'égalité précédente peut donc être restreinte à la zone $\mathcal{T}_i$:

$$\int_{\mathcal{T}_i} (\varphi_i I_p + \gamma_i) \left(\frac{\partial U_h}{\partial t} + \overrightarrow{\nabla}\cdot\overrightarrow{F}(U_h)\right) dV + \int_{\partial\mathcal{T}_i\cap\partial\Omega} (\varphi_i I_p + \gamma_i) \left(\overrightarrow{H}(U_h) - \overrightarrow{F}(U_h)\right)\cdot\overrightarrow{n}\, d\partial\Omega = 0$$

ce qui revient à :

$$\sum_{E\subset\mathcal{T}_i} \int_E \varphi_i^E \left(\frac{\partial U_h}{\partial t} + \overrightarrow{\nabla}\cdot\overrightarrow{F}\right) dV = -\sum_{E\subset\mathcal{T}_i} \left[\int_E \gamma_i^E \left(\frac{\partial U_h}{\partial t} + \overrightarrow{\nabla}\cdot\overrightarrow{F}(U_h)\right) dV + \int_{\partial E\cap\partial\Omega} (\varphi_i I_p + \gamma_i)\,(H_n - F_n)\, d\partial\Omega\right]$$

qui se réécrit de manière synthétique, sachant que par définition, $\gamma_i^E = \beta_i^E - \varphi_i^E$:

$$\sum_{E\subset\mathcal{T}_i} \Phi_i^{E,G} = -\sum_{E\subset\mathcal{T}_i} \left(\Phi_i^E - \Phi_i^{E,G}\right) - \sum_{E\subset\mathcal{T}_i} \sum_{\Gamma\subset\partial E\cap\partial\Omega_h} \phi_i^\Gamma$$

avec $\Phi_i^{E,G}$ désignant une contribution du schéma de Galerkin. Dans le cas d'un problème stationnaire, il suffit de remplacer les 'Φ_i' par des 'ϕ_i' de part et d'autre. Or, la contrainte (B.3) correspond à :

$$\sum_{i\in\Omega_h} \left\langle W_i, \sum_{E\subset\mathcal{T}_i} \Phi_i^{E,G} \right\rangle \leq 0$$

et donc à :

$$\sum_{i\in\Omega_h}\sum_{E\subset\mathcal{T}_i} \left\langle W_i, \left(\Phi_i^E + \sum_{\Gamma\subset\partial E\cap\partial\Omega_h} \phi_i^\Gamma\right)\right\rangle \geq \sum_{i\in\Omega_h}\sum_{E\subset\mathcal{T}_i} \left\langle W_i, \Phi_i^{E,G}\right\rangle$$

Par commodité, on préfère réécrire les sommes sur les éléments plutôt que sur les degrés de liberté, soit :

$$\sum_{E\subset\Omega_h}\sum_{i\in E} \left\langle W_i, \left(\Phi_i^E + \sum_{\Gamma\subset\partial E\cap\partial\Omega_h} \phi_i^\Gamma\right)\right\rangle \geq \sum_{E\subset\Omega_h}\sum_{i\in E} \left\langle W_i, \Phi_i^{E,G}\right\rangle \tag{B.5}$$

On cherche à faire en sorte que le schéma et les conditions limites vérifient cette contrainte. Il est plus facile pour cela d'imposer qu'ils le fassent sur chaque élément, ce qui est une condition plus forte :

$$\forall E\subset\Omega_h,\ \sum_{i\in E} \left\langle W_i, \left(\Phi_i^E + \sum_{\Gamma\subset\partial E\cap\partial\Omega_h} \phi_i^\Gamma\right)\right\rangle \geq \sum_{i\in E} \left\langle W_i, \Phi_i^{E,G}\right\rangle \tag{B.6}$$

On appelle parfois le terme de gauche l'énergie par abus de langage, et analogie avec le cas de flux linéaires. La stabilité entropique demande donc que le schéma dissipe davantage cette énergie que le schéma de Galerkin, celui-ci ne faisant que la conserver (et ne créant donc pas d'entropie physique quand l'inégalité de Clausius-Duhem l'exige). Pour les éléments de bord seulement, les conditions limites participent à cette création. Mais il est plus difficile de concilier celles que nous présentées avec cette condition. En ce qui nous concerne, c'est un travail qui est toujours à l'état de perspectives de recherches, et si on devait étudier cette partie, on séparerait la somme en deux et on chercherait à ce que chacun des deux termes soit positif. C'est une condition plus forte, donc on resterait consistants. L'avantage serait de pouvoir travailler indépendamment sur les deux parties. La dissipation de l'énergie sur les bords est un problème d'autant plus complexe que nous n'avons pas spécifié les conditions limites portant sur l'entropie dans notre dérivation. La bonne façon de procéder serait de formuler l'impact la condition limite sur l'entropie (par exemple, pour une paroi glissante, $Su_n = 0$ sur les bords), de l'ajouter à la description continue (B.1) puis discrète (B.2), et d'intégrer le terme *ad hoc* dans (B.3). Ceci étant très éloigné de nos objectifs, nous ne nous sommes pas penchés sur ce problème. En ne considérant que le premier terme, ce qui est rigoureux pour tous les éléments non connectés à la frontière, on a :

$$\forall E\subset\Omega_h,\ \sum_{i\in E} \left\langle W_i, \Phi_i^E\right\rangle \geq \sum_{i\in E} \left\langle W_i, \Phi_i^{E,G}\right\rangle \tag{B.7}$$

Cette condition n'est déjà pas simple à imposer à un schéma, et ce plus encore au fur et à mesure qu'on se retrouve confrontés à des configurations plus complexes, telles que présentées dans la section 3.5. On serait donc déjà amplement satisfaits d'avoir (B.7), avant d'étudier les conditions limites. À notre connaissance, c'est toujours un thème de recherche ouvert.

B.2 Notes sur le schéma N

Clairement, il est possible d'estimer le signe de (B.7) si les contributions Φ_i s'écrivent linéairement en fonction des W_j, car cela revient alors à étudier une somme de formes bilinéaires définissant l'énergie, en manipulant les matrices associées.

$$\begin{aligned}\sum_{i\in T}\langle W_i, \beta_i^T \phi^T\rangle &= \sum_{i\in T}\left\langle W_i, \beta_i^T \sum_{j\in T} C_{ij} W_j\right\rangle \\ &= \sum_{i\in T}\sum_{j\in T}\langle W_i, M_{ij} W_j\rangle \\ &\geq 0 \quad ?\end{aligned}$$

C'est la procédure adoptée pour étudier les schémas N et LDA. Dans [12] et [3], il est montré que si on dispose d'une linéarisation en les variables entropiques W, sur des éléments linéaires P_1 (2D ou 3D), alors le schéma N vérifie (B.7), tandis que ce n'est pas le cas pour le schéma LDA (voir aussi [1] ou [75]). Malheureusement, dans le cas des équations d'Euler, on ne dispose d'une linéarisation que par rapport aux variables Z de Roe, pour lesquelles les flux sont quadratiques. En MHD, nous avons vu que nous ne disposions même pas de telles variables (section 2.3.4).

Remarque 11. *Concernant le schéma N, on peut noter qu'il est parfaitement approprié au cas de flux linéaires, où le principe du maximum formulable sur U en norme L^2 (cf. section 3.2.2) est vérifié d'un point de vue discret. La démonstration peut être trouvée dans [12] et [3]. La difficulté du passage au non linéaire vient du fait que les équations perdent ce principe.*

Pour avoir une chance de développer un schéma respectant l'inégalité d'entropie discrète (B.7), il faut disposer d'expressions faisant apparaître les quantités W_j. Tous les schémas $\mathcal{RD}$ s'exprimant en fonction de la fluctuation ϕ pour être conservatifs, il faut être capable de reformuler celle-ci en fonction des W_j. L'idée mise en avant dans [3] est de reformuler la divergence des flux à intégrer sous forme quasi-linéaire, non pas en dérivant selon les variables conservatives U qui sont résolues, mais selon les variables entropiques W, ce qui fait apparaître $\overrightarrow{\nabla} W$. Dans les cas où on ne dispose pas d'une linéarisation exacte (la très grande majorité des cas donc), on ne peut pas moyenner simplement (par exemple, traditionellement, via une moyenne arithmétique sur les sommets en P_1) la jacobienne et la sortir de l'intégrale. Au lieu de ça, les auteurs ont montré qu'il était possible, dans le cas P_1, de construire une jacobienne moyenne en variables entropiques sans perte de consistance, à l'aide d'une formule de quadrature. Dans [3], les auteurs parviennent alors à construire un schéma N basé sur ces jacobiennes qui soit consistant avec l'inégalité d'entropie, mais néanmoins sans qu'il la vérifie de manière discrète. Un inconvénient apparaît cependant en contrepartie : la perte formelle de la conservation. Celle-ci n'est en effet normalement vérifiée que grâce au calcul de ϕ sur le contour de l'élément, qui fait que par symétrie, toute erreur due au calcul numérique sur chaque bord est exactement compensée par le calcul sur les contours des éléments voisins. Or ici, on utilise une intégrale de volume. Toutefois, l'erreur de conservation peut être rendue aussi petite que souhaitée par la formule de quadrature, par exemple inférieure à l'erreur de troncature du schéma. En P_1, une quadrature à 3 points, voire 7 points dans certains cas, est suffisante d'après les tests effectués dans ce même papier. Ensuite, pour approcher l'intégrale, les auteurs interpolent logiquement W au lieu de U, ce qui reste une approximation consistante définie comme $U_h = U(W_h)$. C'est toutefois une méthode qui reste coûteuse et dont l'extension à des configurations plus complexes n'a pas encore démontré son efficacité. En particulier, il serait nécessaire de la tester de manière intensive afin de connaître le comportement de l'erreur de conservation et son impact réel sur la solution. Les schémas non parfaitement conservatif, basés sur des intégrales de volumes et qui sont utiles à la définition des schémas $\mathcal{MU}$, furent également étudiés par [31]. En particulier, le contrôle de l'erreur de conservation fut amélioré. Toutefois, dans tous les cas, on ne dispose pas encore d'inégalités d'entropie discrètes.

Concernant la stabilité entropique des schémas centrés, il n'y a pas vraiment de résultats. Le schéma de Lax-Friedrichs d'ordre 1 en Volumes Finis vérifie une inégalité d'entropie discrète dans le cas des équations d'Euler (voir [89] dans le cas 1D). Les équations de la MHD idéale 1D ne présentent aucune particularité faisant obstacle à la généralisation de ce principe. Dans la version $\mathcal{RD}$, pour des configurations autres que P_1 notamment, la situation est moins claire du point de vue formel mais la présence d'une viscosité numérique importante assure en pratique un comportement similaire. Pour les schémas $\mathcal{RD}$ de Lax-Friedrichs et SU, le terme centré n'a clairement aucune propriété entropique particulière. En revanche, concernant le terme de stabilisation présent dans le schéma SU, il offre une dissipation, même si elle n'est pas entropique. Moyennant quelques modifications, il pourrait même offrir une véritable dissipation entropique discrète. C'est ce que nous allons montrer à présent.

B.3 Une voie de stabilisation entropique ?

Modification du terme SUPG

Avec les notions que nous avons présentées précédemment, on peut montrer que le terme de stabilisation qu'on emploie pour le schéma SU ou comme stabilisation des schémas limités, s'il est légèrement remanié avec un choix de τ proche de celui de la matrice N et une interpolation portant sur W, possède un mécanisme de dissipation d'entropie. Plus précisément, ce nouveau terme respecte l'inégalité d'entropie discrète (B.7) dans le cas stationnaire. Ceci est de plus vrai dans toute configuration. On note $A_k = \nabla_U F_k$ et $A_0 = \nabla_U W$ où W désigne les variables entropiques. On rappelle aussi que A_0^{-1} est symétrique définie positive et est un symétriseur à droite de chaque A_k. Pour simplifier, on définit également $\tau_N = \dfrac{\tau}{h^d}$. Sur un élément E quelconque, on peut alors écrire :

$$\begin{aligned}
\frac{1}{h^d}\sum_{i\in E}\langle W_i, S_i\rangle &= \sum_{i\in E} W_i^t \int_E \left(\sum_{k=1}^d A_k \partial_{x_k}\varphi_i\right)\tau_N\left(\sum_{k=1}^d A_k\partial_{x_k}U_h\right)dV \\
&= \int_E \sum_{k=1}^d \left(\sum_{i\in E} W_i\partial_{x_k}\varphi_i\right)^t A_k\tau_N\left(\sum_{k=1}^d A_k A_0^{-1}\partial_{x_k}W_h\right)dV \\
&= \int_E \left(\sum_{k=1}^d \left(\partial_{x_k}W_h\right)^t A_k\right)\tau_N\left(\sum_{k=1}^d A_k A_0^{-1}\partial_{x_k}W_h\right)dV
\end{aligned}$$

Cette première étape requiert de ne pas interpoler U mais bien W, ce qui signifie qu'on travaille avec l'approximation consistante $U_h = U(W_h)$. C'est une première différence avec ce qui est fait usuellement.

Au lieu de définir $\tau_N = N$ comme nous l'avons fait précédemment, on cherche une expression similaire mais qui permette à l'expression ci-dessus d'être positive. Pour cela, on veut faire apparaître A_0^{-1} dans le facteur de gauche. Ceci est rendu possible par une manipulation un peu particulière.

Théorème B.3.1. *Soit A une matrice symétrisable à droite par une matrice A_0^{-1} symétrique définie positive. On peut alors construire une matrice positive notée $(AA_0^{-1})^+$ telle que $(AA_0^{-1})^+ = A^+A_0^{-1}$.*

Démonstration. Cette preuve est inspirée des travaux de Barth [13]. On note $\widetilde{A} = AA_0^{-1}$. $\widetilde{A}$ étant une matrice symétrique, il existe C une matrice A_0-orthogonale, i.e. telle que $A_0^{-1} = CC^t$, qui permet de décomposer $\widetilde{A}$ comme suit :

$$\widetilde{A} = C\widetilde{\Lambda}C^t$$

où $\widetilde{\Lambda}$ est la matrice diagonale des valeurs propres de $\widetilde{A}$. D'autre part, on formule la décomposition de A ainsi :

$$A = R\Lambda L$$

Par définition, on peut donc écrire :

$$\begin{aligned} C\widetilde{\Lambda}C^t &= R\Lambda LCC^t \\ \Rightarrow \quad \widetilde{\Lambda} &= C^{-1}R\Lambda LC \end{aligned}$$

Les deux membres sont semblables, donc $\overline{\Lambda} = \Lambda : AA_0^{-1}$ et A ont les mêmes valeurs propres. De plus, C^{-1} et R ne diffèrent que d'un facteur sur chaque vecteur propre (on réutilisera parfois le terme de *scaling*). Il apparaît donc clairement possible de trouver une matrice de vecteur propres $\overline{R}$ dont le *scaling* permette de vérifier :

$$A = \overline{R}\Lambda\overline{R}^{-1} \text{ et } A_0^{-1} = \overline{R}\overline{R}^t$$

On en déduit alors que :

$$AA_0^{-1} = \overline{R}\Lambda\overline{R}^t$$

Comme noté dans [13], si on souhaite appliquer une fonction - par exemple la partie positive au sens des matrices - à AA_0^{-1}, il ne faut pas procéder directement sur cette matrice :

$$\left(AA_0^{-1}\right)^+ \neq \overline{R}\Lambda^+\overline{R}^{-1}$$

En effet, il faut prendre en compte le fait que cette matrice est celle d'un système modifié par la présence d'un tenseur de Riemann, qui en modifie la métrique. En guise d'illustration, sortons un instant du cadre strict de la preuve revenons au cas des lois de conservations symétrisables. Conformément aux notations précédentes, la matrice $A_kA_0^{-1}$ correspondrait en réalité au système d'équations :

$$A_0^{-1}\frac{\partial W_h}{\partial t} + \sum_{k=1}^{d} AA_0^{-1} \cdot \overrightarrow{\nabla} W_h = 0$$

C'est pour cette raison que nous avons pris soin de ne pas dire que nous prenions la partie positive de AA_0^{-1}, mais que nous voulions définir une matrice $\left(AA_0^{-1}\right)^+$ définie positive. La nuance est proportionnelle à A_0^{-1}. La définition de la matrice que nous cherchons est en fait liée à la partie positive de A, de la façon suivante :

$$\left(AA_0^{-1}\right)^+ = \overline{R}\Lambda^+\overline{LR}\overline{R}^t = \overline{R}\Lambda^+\overline{R}^t$$

On rappelle que AA_0^{-1} étant symétrique, sa décomposition est de la forme :

$$AA_0^{-1} = \widetilde{R}\Lambda\widetilde{R}^t$$

où $\widetilde{R}$ est cette fois réellement une matrice orthogonale, $\widetilde{R}^t = \widetilde{R}^{-1}$. On peut donc noter, là encore, que $\overline{R}$ ne diffère de $\widetilde{R}$ que par un *scaling* des vecteurs propres. La matrice $\left(AA_0^{-1}\right)^+$ ainsi construite est donc clairement positive, et la preuve est terminée. □

Dans le calcul de dissipation, on utilise maintenant un choix de τ_N qui, dans l'esprit, rappelle la matrice N et qui s'appuie sur les constructions de la preuve précédente. Définissons des matrices $\widetilde{K}_j$ comme :

$$\widetilde{K}_j(W_h) = \overrightarrow{A}(W_h) \cdot \overrightarrow{\nu}_j$$

où les $\overrightarrow{\nu}_j$ sont les directions d'*upwinding*, unitaires, introduites dans la section 3.5.2. En P_1, il s'agira des normales définies par 3.5 divisées par leur norme. Par linéarité, les matrices $\widetilde{K}_j$ restent symétrisables à droite par A_0^{-1}, et une construction proche de celle de la matrice N peut alors être :

$$\begin{aligned} \tau_N &= \left(\sum_{j\in E} \widetilde{K}_j{}^+ A_0^{-1}A_0\right)^{-1} \\ &= A_0^{-1}\left(\sum_{j\in E}\left(\widetilde{K}_jA_0^{-1}\right)^+\right)^{-1} \\ &= A_0^{-1}\widetilde{N} \end{aligned} \tag{B.1}$$

L'analyse de la matrice N effectuée dans [1] reste vraie en chaque point pour la construction de $\widetilde{N}$. Il est d'ailleurs intéressant de remarquer que dans cet articles, les matrices $\widetilde{K}_j$ ont été construites de la même manière, pour une bonne raison : l'inversibilité de la somme des K_j est équivalente à celle de la somme des $\widetilde{K}_j$. Les singularités n'apparaissent que lorsqu'on rencontre des points de stagnation. Dans la pratique, comme lors de la construction de la matrice N dans la section 3.3.2, on applique à $\left(AA_0^{-1}\right)^+$ une légère correction sur ses valeurs propres nulles pour garantir l'inversibilité. En outre, comme la matrice qui en résulte est définie positive, son inverse $\widetilde{N}$ l'est également. C'est cette propriété que nous allons mettre à profit.

Reprenons maintenant les calculs initiaux. Une différence à ne pas omettre entre N et $\widetilde{N}$ est la dimension : les normales, ou les directions d'upwinding, n'intervenant pas dans $\widetilde{N}$, cette matrice est de dimension seulement $[\lambda]^{-1}$. Il faut donc reconsidérer la définition de τ : $\tau = hA_0^{-1}\widetilde{N}$. Ceci donne :

$$\begin{aligned}\frac{1}{h}\sum_{i\in E}\langle W_i, S_i\rangle &= \int_E \left(\sum_{k=1}^{d}\left(\partial_{x_k}W_h\right)^t A_k\right) A_0^{-1}\widetilde{N}\left(\sum_{k=1}^{d} A_k A_0^{-1}\partial_{x_k}W_h\right) dV \\ &= \int_E \left(\sum_{k=1}^{d} A_k A_0^{-1}\partial_{x_k}W_h\right)^t \widetilde{N}\left(\sum_{k=1}^{d} A_k A_0^{-1}\partial_{x_k}W_h\right) dV \end{aligned} \tag{B.2}$$

En tout point de E, si la matrice $\widetilde{N}$ a bien été corrigée de manière à être positive, l'intégrande précédent est de la forme :

$$\left\langle V_h, \widetilde{N}V_h\right\rangle \geq 0$$

et on a donc bien :

$$\sum_{i\in E}\langle W_i, S_i\rangle \geq 0 \tag{B.3}$$

Conclusion

Pour résumer, par rapport à la définition de la stabilisation vue dans la section 3.4.2, on a modifié deux choses :

- l'interpolation de U se fait désormais à travers celle de W : $U_h = U(W_h)$,
- la définition de τ est modifiée en $\tau = h\tau_N = hA_0^{-1}\widetilde{N}$ et la nouvelle matrice $\widetilde{N}$ n'est plus évaluée en un état moyen.

Cela conduit à redéfinir le terme de stabilisation ainsi, sur tout élément E :

$$S_i^E = h\int_E \left((\overrightarrow{A}A_0^{-1})\cdot\overrightarrow{\nabla}\varphi_i\right)\widetilde{N}\left((\overrightarrow{A}A_0^{-1})\cdot W_h\right)dV$$

qui, avec $\widetilde{N}$ donnée par (B.1) et indépendante de i, permet de préserver la conservation du schéma global d'origine (la somme sur i des S_i reste nulle). Sous cette forme semi-continue, i.e. avec une solution interpolée mais une intégrale incalculable exactement, l'inégalité (B.3) est acquise. Mais dans la pratique, nous utiliserions une formule de quadrature à N_q points, du type :

$$\left(S_i^E\right)_h = \sum_{q=1}^{N_q}\omega_q\left((\overrightarrow{A}A_0^{-1})(\overrightarrow{x}_q)\cdot\overrightarrow{\nabla}\varphi_i(\overrightarrow{x}_q)\right)\widetilde{N}(\overrightarrow{x}_q)\left((\overrightarrow{A}A_0^{-1})(\overrightarrow{x}_q)\cdot\overrightarrow{\nabla}W_h(\overrightarrow{x}_q)\right)$$

où on a utilisé un raccourci de notation sur chaque matrice : $A(\overrightarrow{x}_q) = A\left(W_h(\overrightarrow{x}_q)\right)$. Or, il faut simplement voir que $W_h(\overrightarrow{x}_q) = \sum_{j\in E} W_j\varphi_j(\overrightarrow{x}_q)$ pour constater que :

$$\sum_{i\in E}\left\langle W_i, \left(S_i^E\right)_h\right\rangle = \sum_{q=1}^{N_q}\omega_q\left((\overrightarrow{A}A_0^{-1})(\overrightarrow{x}_q)\cdot\overrightarrow{\nabla}W_h(\overrightarrow{x}_q)\right)\widetilde{N}(\overrightarrow{x}_q)\left((\overrightarrow{A}A_0^{-1})(\overrightarrow{x}_q)\cdot\overrightarrow{\nabla}W_h(\overrightarrow{x}_q)\right) \geq 0$$

car il s'agit simplement de l'intégrande de (B.2) évalué en un point de quadrature et puisque $\omega_q > 0$. L'approximation $\left(S_i^E\right)_h$ transporterait donc l'inégalité d'entropie (B.3) au niveau complètement discret.

À noter en contrepartie que ce terme, de par la formulation du paramètre τ qui n'est plus constante par élément, s'éloigne de l'interprétation moindres carrés du terme SUPG original. Il offre néanmoins une stabilisation entropique au niveau discret qui peut être intéressante à exploiter. Nous n'avons pas encore testé ses capacités sous cette forme dans nos algorithmes, donc nous nous garderons d'anticiper ici les bénéfices qu'il pourrait offrir. De plus, à degré d'interpolation équivalent, le terme à intégrer est de degré plus élevé et peut nécessiter une formule de quadrature plus précise que pour la stabilisation à paramètre τ constant. Si des tests avéraient que cette précision est nécessaire pour rendre la dissipation effective, le coût supplémentaire pourrait être non négligeable. Cette question mériterait sans doute quelques investigations expérimentales.

De plus, le passage aux problèmes instationnaires n'est pas trivial du tout, car avec le terme $\widetilde{\partial_t U_h}$, si les fonctions de base ne comprennent pas une interpolation en temps (ce qui sera notre cas), on n'a aucune chance de pouvoir formuler une stabilisation entropique inspirée du terme SUPG. Pour nous, dans le cas instationnaire, le mieux est sans doute d'utiliser le terme sous la forme stationnaire que nous venons de donner, sans rajouter la dérivée en temps dans le dernier facteur.

Une simplification possible

Comme nous l'annoncions dans la section 3.4.2, quitte à s'éloigner de la formulation GLS stricte, en ne se souciant que de la préservation de la conservation, on pourrait simplifier cette stabilisation. En effet, on peut facilement remarquer dans l'analyse précédente que si toutes les matrices, tant A_k que A_0 et $\widetilde{N}$, sont évaluées en un même état moyen, l'inégalité entropique est préservée. Ceci signifie que la construction de $\widetilde{N}$ ne se fait plus, elle aussi, qu'en un état moyen. Au final, le gain en temps de calcul pourrait se révéler significatif. Supposons que nous définissions un état moyen $\overline{U}$ ou $\overline{W}$ par une simple moyenne arithmétique des valeurs de U ou de W dans l'élément. Dans chaque cas, on notera une matrice quelconque $\overline{C}$ en référence à $C(\overline{U})$ ou $C(\overline{W})$. Ce qui compte est de disposer d'une matrice moyenne, peu importe cette moyenne pour avoir l'inégalité. On définit alors une stabilisation simplifiée :

$$\overline{S_i}^E = h \int_E \left((\overrightarrow{\overline{A}}\,\overline{A_0}) \cdot \overrightarrow{\nabla} \varphi_i \right) \widetilde{\overline{N}} \left((\overrightarrow{\overline{A}}\,\overline{A_0}) \cdot \overrightarrow{\nabla} W_h \right) dV$$

où

$$\begin{aligned} \widetilde{\overline{N}} &= \left(\sum_{j \in E} \widetilde{\overline{K}}_j \overline{A_0} \right)^{-1} \\ \widetilde{\overline{K}}_j &= \widetilde{K}_j(\overline{W}) \end{aligned}$$

Le fait de moyenner les matrices est une approximation grossière, et lors d'une convergence en maillage, ce terme convergerait sans doute moins vite que le terme original S_i^E. Sa diffusion sera probablement plus importante. Il reste néanmoins consistant, dissipateur d'entropie, et peut être approché avec une précision arbitrairement élevée.

Annexe C

Construction de la matrice de Roe pour les conditions limites

Commençons par préciser nos notations :

$$\Delta x = x_2 - x_1$$

$$\overline{x} = \frac{\sqrt{\rho_1}x_1 + \sqrt{\rho_2}x_2}{\sqrt{\rho_1} + \sqrt{\rho_2}}$$

$$\underline{x} = \frac{\sqrt{\rho_2}x_1 + \sqrt{\rho_1}x_2}{\sqrt{\rho_1} + \sqrt{\rho_2}}$$

Notons quelques relations algébriques utiles pour les calculs :

$$\Delta xy = \underline{y}\Delta x + \overline{x}\Delta y$$

$$\underline{xy} = \underline{x}\,\underline{y}$$

$$\overline{\left(\frac{1}{x}\right)} = \frac{1}{\underline{x}}$$

Dans [19] et [24], la démarche a été effectuée dans le cas 1D où la contrainte sur la divergence du champ magnétique se résume à $B_x = B_x(t=0)$, autrement dit $\Delta B_x = 0$. Pour nous, par rotation, cela équivaudrait à $\Delta B_n = 0$. On suppose la normale unitaire. Si ce n'est pas le cas, il suffira de multiplier la matrice de Roe par sa norme une fois les calculs terminés (A_n étant linéaire en $\overrightarrow{n}$). Sans en tenir compte pour le moment, exprimons la variation des flux en fonction de ΔU selon le modèle de [24] :

$$\Delta F_n = \begin{pmatrix} \Delta(\rho u_n) \\ \Delta(\rho u_n \overrightarrow{u}) + \Delta\left(p + \frac{\overrightarrow{B}^2}{2}\right)\overrightarrow{n} - \Delta(B_n \overrightarrow{B}) \\ \Delta\left(\rho h u_n\right) - \Delta\left(B_n(\overrightarrow{u}\cdot\overrightarrow{B})\right) \\ \Delta(u_n \overrightarrow{B}) - \Delta(B_n \overrightarrow{u}) + \Delta\psi\overrightarrow{n} \\ {c_h}^2 \Delta B_n \end{pmatrix} \tag{C.1}$$

avec l'enthalpie massique h donnée par :

$$\rho h = E + p + \frac{\overrightarrow{B}^2}{2}$$

L'idée importante de [24] concernant ce calcul est de remarquer que le saut de pression magnétique doit faire intervenir le saut de densité, via la relation :

$$\Delta\left(\frac{\overrightarrow{B}^2}{2}\right) = X\Delta\rho + \underline{\overrightarrow{B}}\cdot\Delta\overrightarrow{B}$$

où

$$X = \frac{\left(\Delta\overrightarrow{B}\right)^2}{2\left(\sqrt{\rho_1}+\sqrt{\rho_2}\right)^2}$$

Sans cela, les valeurs propres de la matrice de Roe ne sont pas réelles. Tous calculs faits, on obtient finalement :

$$\Delta F_n = \left(\begin{array}{c} \Delta(\rho u_n) \\ \left[\left((\gamma-1)\frac{\overline{\overrightarrow{u}}^2}{2}+(2-\gamma)X\right)\overrightarrow{n} - \overline{u_n}\,\overline{\overrightarrow{u}}\right]\Delta\rho + \overline{u_n}\Delta(\rho\overrightarrow{u}) + (1-\gamma)\overline{\overrightarrow{u}}\cdot\Delta(\rho\overrightarrow{u})\overrightarrow{n} + \overline{\overrightarrow{u}}\Delta(\rho\overrightarrow{u})\cdot\overrightarrow{n} \\ +(\gamma-1)\Delta E + (2-\gamma)\underline{\overrightarrow{B}}\cdot\Delta\overrightarrow{B}\,\overrightarrow{n} - \underline{B_n}\Delta\overrightarrow{B} - \overline{\overrightarrow{B}}\Delta B_n \\ \left[\overline{u_n}\left((\gamma-1)\frac{\overline{\overrightarrow{u}}^2}{2}+(2-\gamma)X-\overline{h}\right)+\frac{B_n}{\underline{\rho}}\underline{\overrightarrow{B}}\cdot\overline{\overrightarrow{u}}\right]\Delta\rho + \left((1-\gamma)\overline{u_n}\,\overline{\overrightarrow{u}}+\overline{h}\,\overrightarrow{n}-\frac{B_n\overrightarrow{B}}{\underline{\rho}}\right)\cdot\Delta(\rho\overrightarrow{u}) \\ +\gamma\overline{u_n}\Delta E + \left((2-\gamma)\overline{u_n}\underline{\overrightarrow{B}} - \underline{B_n}\,\overline{\overrightarrow{u}}\right)\cdot\Delta\overrightarrow{B} - \overline{\left(\overrightarrow{u}\cdot\overrightarrow{B}\right)}\Delta B_n \\ \frac{B_n\,\overline{\overrightarrow{u}}-\overline{u_n}\overrightarrow{B}}{\underline{\rho}}\Delta\rho + \frac{\overrightarrow{B}}{\underline{\rho}}\Delta(\rho\overrightarrow{u})\cdot\overrightarrow{n} - \frac{B_n}{\underline{\rho}}\Delta(\rho\overrightarrow{u}) + \overline{u_n}\Delta\overrightarrow{B} - \overline{\overrightarrow{u}}\Delta B_n + \Delta\psi\overrightarrow{n} \\ {c_h}^2\Delta B_n \end{array}\right)$$

La matrice issue de ces expressions n'est pas symétrisable. De fait on aimerait récupérer, comme dans [24], une matrice proche des jacobiennes symétrisables que nous avions obtenues au chapitre 2, exprimée dans un jeu de variables consistant : $\underline{U} = (\underline{\rho}, \overline{\overrightarrow{u}}^t, \overline{h}, \underline{\overrightarrow{B}}^t, \underline{\psi})^t$. Pour cela, nous avions noté au chapitre précédent qu'il était nécessaire de rajouter certains termes dits de Powell (dûs à Godunov), proportionnels à $\overrightarrow{\nabla}\cdot\overrightarrow{B}$. De la même manière, on avait rajouté des termes en $\overrightarrow{\nabla}\psi$ pour que la correction ne brise pas l'invariance Galiléenne des équations. Tous ces termes devraient donc logiquement nous être nécessaire ici aussi. Si ce n'était pas le cas dans [24], c'est uniquement dû à la simplification qu'amène la résolution d'un problème 1D. Nous pourrions nous aussi considérer que nous avons affaire à un problème 1D suivant la direction $\overrightarrow{n}$ comme nous le faisions comprendre plus haut. Dans ce cas, il faudrait ôter la correction du flux de Roe, ne rien faire sur la composante ψ qui serait imposée fortement comme nulle (voir plus loin), et considérer que $\Delta B_n = 0$. Toutefois, nous préférons conserver la formulation complète pour compenser au mieux tous les termes distribués par le schéma sur le bord, et qui n'auraient pas dû l'être. Les termes source symétrisants n'étant pas conservatifs, on ne peut pas les incorporer dans ΔF_n. On prend donc le problème différemment. Dans le cas continu, les termes source de Powell compensent certains autres termes dans les équations et les font disparaître. Une manière consistante de procéder avec ΔF_n est donc d'identifier les termes correspondants et de les considérer nuls. Il s'agit clairement des termes proportionnels ΔB_n, qui s'annulent en 1D. Il faut également rajouter des termes proportionnels à $\Delta\psi$ pour rendre la matrice

symétrisable. Tout cela est globalement équivalent à rajouter à ΔF_n les termes suivants :

$$\Delta F_n := \Delta F_n + \begin{pmatrix} 0 \\ \overline{\underline{\vec{B}}}\Delta B_n \\ \overline{\left(\vec{u}\cdot\vec{B}\right)}\Delta B_n + \underline{B_n}\Delta\psi \\ \overline{\underline{\vec{u}}}\Delta B_n \\ \overline{u_n}\Delta\psi \end{pmatrix}$$

Le fait de rajouter ces termes pour définir la matrice de Roe, et en étudier le système propre, rejoint notre parti pris de considérer la matrice avec les termes symétrisants pour tout ce qui touche au système propre, mais de ne pas les prendre en compte dans les équations que nous résolvons. La matrice qui en résulte est :

$$\overline{A_n} = \begin{pmatrix} 0 & \vec{n}^t & 0 & \vec{0}^t & 0 \\ \overline{A_n}_{2:4,1} & \overline{u_n}I_3 + (1-\gamma)\vec{n}\overline{\vec{u}}^t + \overline{\vec{u}}\vec{n}^t & (\gamma-1)\vec{n} & (2-\gamma)\vec{n}\underline{\vec{B}}^t - \underline{B_n}I_3 & \vec{0} \\ \overline{A_n}_{5,1} & (1-\gamma)\overline{u_n}\overline{\vec{u}}^t + \overline{h}\vec{n}^t - \dfrac{\underline{B_n}\underline{\vec{B}}^t}{\underline{\rho}} & \gamma\overline{u_n} & (2-\gamma)\overline{u_n}\underline{\vec{B}}^t - \underline{B_n}\overline{\vec{u}}^t & \underline{B_n} \\ \dfrac{\underline{B_n}\overline{\vec{u}} - \overline{u_n}\underline{\vec{B}}}{\underline{\rho}} & \dfrac{\underline{\vec{B}}\vec{n}^t - \underline{B_n}I_3}{\underline{\rho}} & \vec{0} & \overline{u_n}I_3 & \vec{n} \\ 0 & \vec{0}^t & 0 & {c_h}^2\vec{n}^t & \overline{u_n} \end{pmatrix}$$

avec :

$$\overline{A_n}_{2:4,1} = \left((\gamma-1)\frac{\overline{\vec{u}}^2}{2} + (2-\gamma)X\right)\vec{n} - \overline{u_n}\overline{\vec{u}}$$

$$\overline{A_n}_{5,1} = \left((\gamma-1)\frac{\overline{\vec{u}}^2}{2} + (2-\gamma)X - \overline{h}\right)\overline{u_n} + \frac{\underline{B_n}}{\underline{\rho}}\underline{\vec{B}}\cdot\overline{\vec{u}}$$

I_3 désigne la matrice idéntité de taille 3. Cette matrice est similaire aux jacobiennes du cas continu évaluées en $\underline{U}$, aux termes ci-dessus (proportionnels à X) près. La dérivation du système propre est donc similaire. Les valeurs propres de $\overline{A_n}$ sont, dans l'ordre croissant :

$$\overline{u_n} - \overline{c_f} \leq \overline{u_n} - \overline{c_a} \leq \overline{u_n} - \overline{c_s} \leq \overline{u_n} \leq \overline{u_n} + \overline{c_s} \leq \overline{u_n} + \overline{c_a} \leq \overline{u_n} + \overline{c_f}$$
$$\overline{u_n} - c_h \leq \overline{u_n} + c_h$$

où nous avons défini :

$$\overline{c_f} = \sqrt{\frac{1}{2}\left(\overline{a}^2 + \frac{\overline{\vec{B}}^2}{\underline{\rho}} + \sqrt{\left(\overline{a}^2 + \frac{\overline{\vec{B}}^2}{\underline{\rho}}\right)^2 - 4\overline{a}^2\frac{{\underline{B_n}}^2}{\underline{\rho}}}\right)}$$

$$\overline{c_f} = \sqrt{\frac{1}{2}\left(\overline{a}^2 + \frac{\overline{\vec{B}}^2}{\underline{\rho}} - \sqrt{\left(\overline{a}^2 + \frac{\overline{\vec{B}}^2}{\underline{\rho}}\right)^2 - 4\overline{a}^2\frac{{\underline{B_n}}^2}{\underline{\rho}}}\right)}$$

$$\overline{c_a} = \frac{|\underline{B_n}|}{\sqrt{\underline{\rho}}}$$

$$\overline{a}^2 = (2-\gamma)X + (\gamma-1)\left(\overline{h} - \frac{\overline{\vec{u}}^2}{2} - \frac{\overline{\vec{B}}^2}{\underline{\rho}}\right)$$

Les vecteurs propres et formes propres sont également proches de ceux présentés au second chapitre. On utilise ici encore, à l'instar de [24], la renormalisation de Roe et Balsara ([82]) qui fait intervenir les coefficients suivants :

$$\overline{\alpha_s} = \sqrt{\frac{\overline{c_f}^2 - \overline{a}^2}{\overline{c_f}^2 - \overline{c_s}^2}} \qquad \overline{\alpha_f} = \sqrt{\frac{\overline{a}^2 - \overline{c_s}^2}{\overline{c_f}^2 - \overline{c_s}^2}}$$

ainsi que, pour alléger les expressions :

$$\underline{b_n} = \frac{\underline{B_n}}{|\underline{B_n}|} = \mathrm{sgn}\left(\underline{B_n}\right) \quad , \quad \underline{\overrightarrow{B}}^{\perp} = \underline{\overrightarrow{B}} - \underline{B_n}\,\overrightarrow{n} \quad \text{et} \quad \underline{\overrightarrow{b}}^{\perp} = \frac{\underline{\overrightarrow{B}}^{\perp}}{\|\underline{\overrightarrow{B}}^{\perp}\|}$$

Les vecteurs propres sont donnés en 1D dans [24] (i.e. hormis ceux liés à la correction de la divergence). Pour être complets, on liste également les formes propres associées. Le lecteur peut vérifier facilement qu'on retrouve le système propre du problème continu symétrisable en prenant $\overline{x} = \underline{x} = x$ et $X = 0$.

$$\overline{r_0} = \begin{pmatrix} 1 \\ \\ \overline{\overrightarrow{u}} \\ \\ \dfrac{\overline{\overrightarrow{u}}^2}{2} + \dfrac{\gamma-2}{\gamma-1}X \\ \\ \overrightarrow{0} \\ \\ 0 \end{pmatrix} \qquad \overline{l_0}^t = \frac{1}{\overline{a}^2}\begin{pmatrix} \overline{a}^2 + (1-\gamma)\dfrac{\overline{\overrightarrow{u}}^2}{2} + (\gamma-2)X \\ \\ (\gamma-1)\overline{\overrightarrow{u}} \\ \\ (1-\gamma) \\ \\ (\gamma-1)\underline{\overrightarrow{B}} \\ \\ 0 \end{pmatrix}$$

$$\overline{r_{\pm a}} = \begin{pmatrix} 0 \\ \\ \overrightarrow{n} \wedge \underline{\overrightarrow{b}}^{\perp} \\ \\ \left(\overrightarrow{n} \wedge \underline{\overrightarrow{b}}^{\perp}\right) \cdot \overline{\overrightarrow{u}} \\ \\ \mp \dfrac{\underline{b_n}}{\sqrt{\underline{\rho}}}\,\overrightarrow{n} \wedge \underline{\overrightarrow{b}}^{\perp} \\ \\ 0 \end{pmatrix} \qquad \overline{l_{\pm a}}^t = \frac{1}{2}\begin{pmatrix} -\left(\overrightarrow{n} \wedge \underline{\overrightarrow{b}}^{\perp}\right) \cdot \overline{\overrightarrow{u}} \\ \\ \overrightarrow{n} \wedge \underline{\overrightarrow{b}}^{\perp} \\ \\ 0 \\ \\ \mp\sqrt{\underline{\rho}}\,\underline{b_n}\,\overrightarrow{n} \wedge \underline{\overrightarrow{b}}^{\perp} \\ \\ 0 \end{pmatrix}$$

$$\overline{r_{\pm f}} = \begin{pmatrix} \underline{\rho}\,\overline{\alpha_f} \\ \\ \underline{\rho}\,\overline{\alpha_f}\left(\overline{\overrightarrow{u}} \pm \overline{c_f}\,\overrightarrow{n}\right) \mp \underline{\rho}\,\overline{\alpha_s c_s}\,\underline{b_n}\,\underline{\overrightarrow{b}}^{\perp} \\ \\ \underline{\rho}\,\overline{\alpha_f}\left(\overline{h} - \dfrac{\underline{\overrightarrow{B}}^2}{\underline{\rho}} \pm \overline{u_n c_f}\right) \mp \underline{\rho}\,\overline{\alpha_s c_s}\,\underline{b_n}\left(\underline{\overrightarrow{b}}^{\perp} \cdot \overline{\overrightarrow{u}}\right) + \sqrt{\underline{\rho}\,\overline{\alpha_s a}}\|\underline{\overrightarrow{B}}^{\perp}\| \\ \\ \sqrt{\underline{\rho}}\,\overline{\alpha_s a}\,\underline{\overrightarrow{b}}^{\perp} \\ \\ 0 \end{pmatrix}$$

$$\overline{l_{\pm f}}^t = \frac{1}{2\underline{\rho}\overline{a}^2}\begin{pmatrix} (\gamma-1)\overline{\alpha_f}\dfrac{\overline{\overrightarrow{u}}^2}{2} \mp \overline{\alpha_f c_f u_n} \pm \overline{\alpha_s c_s}\underline{b_n}\left(\underline{\overrightarrow{b}}^{\perp}\cdot\overline{\overrightarrow{u}}\right) + (2-\gamma)\overline{\alpha_f}X \\ (1-\gamma)\overline{\alpha_f}\,\overline{\overrightarrow{u}} \pm \overline{\alpha_f c_f}\,\overline{\overrightarrow{n}} \mp \overline{\alpha_s c_s}\underline{b_n}\,\underline{\overrightarrow{b}}^{\perp} \\ (\gamma-1)\overline{\alpha_f} \\ (1-\gamma)\overline{\alpha_f}\underline{\overrightarrow{B}} + \sqrt{\underline{\rho}}\;\overline{a\alpha_s}\,\underline{\overrightarrow{b}}^{\perp} \\ 0 \end{pmatrix}$$

$$\overline{r_{\pm s}} = \begin{pmatrix} \underline{\rho}\overline{\alpha_s} \\ \underline{\rho}\overline{\alpha_s}\left(\overline{\overrightarrow{u}} \pm \overline{c_s}\,\overline{\overrightarrow{n}}\right) \pm \underline{\rho}\overline{\alpha_f c_f}\underline{b_n}\,\underline{\overrightarrow{b}}^{\perp} \\ \underline{\rho}\overline{\alpha_s}\left(\overline{h} - \dfrac{\overline{\overrightarrow{B}}^2}{\underline{\rho}} \pm \overline{u_n c_s}\right) \pm \underline{\rho}\overline{\alpha_f c_f}\underline{b_n}\left(\underline{\overrightarrow{b}}^{\perp}\cdot\overline{\overrightarrow{u}}\right) - \sqrt{\underline{\rho}\overline{\alpha_f a}}\|\underline{\overrightarrow{B}}^{\perp}\| \\ -\sqrt{\underline{\rho}}\;\overline{\alpha_f a}\,\underline{\overrightarrow{b}}^{\perp} \\ 0 \end{pmatrix}$$

$$\overline{l_{\pm s}}^t = \frac{1}{2\underline{\rho}\overline{a}^2}\begin{pmatrix} (\gamma-1)\overline{\alpha_s}\dfrac{\overline{\overrightarrow{u}}^2}{2} \mp \overline{\alpha_s c_s u_n} \mp \overline{\alpha_f c_f}\underline{b_n}\left(\underline{\overrightarrow{b}}^{\perp}\cdot\overline{\overrightarrow{u}}\right) + (2-\gamma)\overline{\alpha_s}X \\ (1-\gamma)\overline{\alpha_s}\,\overline{\overrightarrow{u}} \pm \overline{\alpha_s c_s}\,\overline{\overrightarrow{n}} \pm \overline{\alpha_f c_f}\underline{b_n}\,\underline{\overrightarrow{b}}^{\perp} \\ (\gamma-1)\overline{\alpha_s} \\ (1-\gamma)\overline{\alpha_s}\underline{\overrightarrow{B}} - \sqrt{\underline{\rho}}\;\overline{a\alpha_f}\,\underline{\overrightarrow{b}}^{\perp} \\ 0 \end{pmatrix}$$

$$\overline{r_{\pm h}} = \begin{pmatrix} 0 \\ \overrightarrow{0} \\ \underline{B_n} \\ \overrightarrow{n} \\ \pm c_h \end{pmatrix} \qquad \overline{l_{\pm h}}^t = \frac{1}{2}\begin{pmatrix} 0 \\ \overrightarrow{0} \\ 0 \\ \overrightarrow{n} \\ \pm\dfrac{1}{c_h} \end{pmatrix}$$

Annexe D

Matrices des flux diffusifs

Nous donnons ici les matrices permettant d'écrire les flux $\overrightarrow{F_D}$ sous la forme quasi-linéaire (4.4.2) dans une configuration 3D. Une manière encore plus simple que le calcul direct consiste à passer par la base des variables physiques $V^t = \left(\rho, \overrightarrow{u}, p, \overrightarrow{B}, \psi\right)$ pour déterminer les matrices B_{kl} :

$$\forall k \in \{x, y, z\}, \quad (F_D)_k = \sum_l A_{kl} \frac{\partial U}{\partial V} \frac{\partial V}{\partial x_l} = \sum_l B_{kl} \frac{\partial V}{\partial x_l} \tag{D.1}$$

et ensuite d'en déduire les matrices des flux A_{kl}. Il est bien évident, d'après l'expression des flux rappelée à la section 4.4.1, que ψ ne sera pas utile, et que les lignes et les colonnes correspondantes resteront vides. Nous reprenons bien sûr les notations vues dans les chapitres précédents, mais nous y ajoutons (pour simplifier les expressions) les quantités suivantes :

$$D^* = \frac{D}{Pe_\rho} \qquad \nu^* = \frac{\nu}{Re} \qquad \kappa^* = \frac{\kappa}{r_s Pe} \qquad \eta^* = \frac{\eta}{Rm}$$

où r_s désigne la constante massique du gaz parfait considéré, introduite dans la section 2.1.3, et dont l'apparition est due à la dérivation de la température qui vérifie :

$$T = \frac{p}{\rho r_s}$$

(Si on ne souhaite utiliser aucun adimensionnement des équations, cela revient à prendre les nombres sans dimension précédents égaux à 1, hormis le nombre de Reynolds magnétique Rm qui doit être pris égal à μ_0.) Les matrices B_{kl} sont alors données par :

$$B_{xx} = \begin{pmatrix} D^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{4}{3}\nu^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \nu^* & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \nu^* & 0 & 0 & 0 & 0 & 0 \\ -\kappa^* \frac{p}{\rho^2} & \frac{4}{3}\nu^* u & \nu^* v & \nu^* w & \frac{\kappa^*}{\rho} & -\eta^* B_x & \eta^* B_y & \eta^* B_z & 0 \\ 0 & 0 & 0 & 0 & 0 & \eta^* & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \eta^* & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \eta^* & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$B_{xy} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{2}{3}\nu^* & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \nu^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \nu^* v & -\frac{2}{3}\nu^* u & 0 & 0 & -\eta^* B_y & -\eta^* B_x & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$B_{xz} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{2}{3}\nu^* & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \nu^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \nu^* w & 0 & -\frac{2}{3}\nu^* u & 0 & -\eta^* B_z & 0 & -\eta^* B_x & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$B_{yx} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \nu^* & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{2}{3}\nu^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{2}{3}\nu^* v & \nu^* u & 0 & 0 & -\eta^* B_y & -\eta^* B_x & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$B_{yy} = \begin{pmatrix} D^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \nu^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{4}{3}\nu^* & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \nu^* & 0 & 0 & 0 & 0 & 0 \\ -\kappa^* \frac{p}{\rho^2} & \nu^* u & \frac{4}{3}\nu^* v & \nu^* w & \frac{\kappa^*}{\rho} & \eta^* B_x & -\eta^* B_y & \eta^* B_z & 0 \\ 0 & 0 & 0 & 0 & 0 & \eta^* & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \eta^* & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \eta^* & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$B_{yz} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{2}{3}\nu^* & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \nu^* & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \nu^* w & -\frac{2}{3}\nu^* v & 0 & 0 & -\eta^* B_z & -\eta^* B_y & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$B_{zx} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \nu^* & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{2}{3}\nu^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{2}{3}\nu^* w & 0 & \nu^* u & 0 & -\eta^* B_z & 0 & -\eta^* B_x & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$B_{zy} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \nu^* & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{2}{3}\nu^* & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{2}{3}\nu^* w & \nu^* v & 0 & 0 & -\eta^* B_z & -\eta^* B_y & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$B_{zz} = \begin{pmatrix} D^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \nu^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \nu^* & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{4}{3}\nu^* & 0 & 0 & 0 & 0 & 0 \\ -\kappa^* \frac{p}{\rho^2} & \nu^* u & \nu^* v & \frac{4}{3}\nu^* w & \frac{\kappa^*}{\rho} & \eta^* B_x & \eta^* B_y & -\eta^* B_z & 0 \\ 0 & 0 & 0 & 0 & 0 & \eta^* & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \eta^* & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \eta^* & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Pour calculer les matrices A_{kl}, puisque l'interpolation porte sur U et non sur V, il suffit de multiplier les matrices précédentes à droite :

$$\forall (k,l),\ A_{kl} = B_{kl} \frac{\partial V}{\partial U}$$

avec, en écriture compressée :

$$\frac{\partial U}{\partial V} = \begin{pmatrix} 1 & \overrightarrow{0}^t & 0 & \overrightarrow{0}^t & 0 \\ -\dfrac{\overrightarrow{u}}{\rho} & \dfrac{1}{\rho}\boldsymbol{I}_3 & \overrightarrow{0} & \overrightarrow{0}\,\overrightarrow{0}^t & \overrightarrow{0} \\ (\gamma-1)\dfrac{\overrightarrow{u}^2}{2} & (1-\gamma)\overrightarrow{u}^t & \gamma-1 & (1-\gamma)\overrightarrow{B}^t & 0 \\ \overrightarrow{0} & \overrightarrow{0}\,\overrightarrow{0}^t & \overrightarrow{0} & \boldsymbol{I}_3 & \overrightarrow{0} \\ 0 & \overrightarrow{0}^t & 0 & \overrightarrow{0}^t & 1 \end{pmatrix}$$

$\boldsymbol{I}_3$ étant toujours la matrice identité, de taille 3. On obtient donc au final les matrices suivantes :

$$A_{xx} = \begin{pmatrix} D^* & 0 & 0 & 0 & 0 \\ -\dfrac{4}{3}\dfrac{\nu^* u}{\rho} & \dfrac{4}{3}\dfrac{\nu^*}{\rho} & 0 & 0 & 0 \\ -\dfrac{\nu^* v}{\rho} & 0 & \dfrac{\nu^*}{\rho} & 0 & 0 \\ -\dfrac{\nu^* w}{\rho} & 0 & 0 & \dfrac{\nu^*}{\rho} & 0 \\ (A_{xx})_{5,1} & \dfrac{4}{3}\dfrac{\nu^* u}{\rho} + (1-\gamma)\dfrac{\kappa^* u}{\rho} & \dfrac{\nu^* v}{\rho} + (1-\gamma)\dfrac{\kappa^* v}{\rho} & \dfrac{\nu^* w}{\rho} + (1-\gamma)\dfrac{\kappa^* w}{\rho} & (\gamma-1)\dfrac{\kappa^*}{\rho} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \left((1-\gamma)\dfrac{\kappa^*}{\rho} - \eta^*\right) B_x & \left((1-\gamma)\dfrac{\kappa^*}{\rho} + \eta^*\right) B_y & \left((1-\gamma)\dfrac{\kappa^*}{\rho} + \eta^*\right) B_z & 0 \\ \eta^* & 0 & 0 & 0 \\ 0 & \eta^* & 0 & 0 \\ 0 & 0 & \eta^* & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{xy} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \dfrac{2}{3}\dfrac{\nu^* v}{\rho} & 0 & -\dfrac{2}{3}\dfrac{\nu^*}{\rho} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\dfrac{\nu^* u}{\rho} & \dfrac{\nu^*}{\rho} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\dfrac{1}{3}\dfrac{\nu^* uv}{\rho} & \dfrac{\nu^* v}{\rho} & -\dfrac{2}{3}\dfrac{\nu^* u}{\rho} & 0 & 0 & -\eta^* B_y & -\eta^* B_x & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{xz} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{2}{3}\frac{\nu^* w}{\rho} & 0 & 0 & -\frac{2}{3}\frac{\nu^*}{\rho} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\nu^* u}{\rho} & \frac{\nu^*}{\rho} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{3}\frac{\nu^* uw}{\rho} & \frac{\nu^* w}{\rho} & 0 & -\frac{2}{3}\frac{\nu^* u}{\rho} & 0 & -\eta^* B_z & 0 & -\eta^* B_x & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{yx} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\nu^* v}{\rho} & 0 & \frac{\nu^*}{\rho} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{2}{3}\frac{\nu^* u}{\rho} & -\frac{2}{3}\frac{\nu^*}{\rho} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{3}\frac{\nu^* uv}{\rho} & -\frac{2}{3}\frac{\nu^* v}{\rho} & \frac{\nu^* u}{\rho} & 0 & 0 & -\eta^* B_y & -\eta^* B_x & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{yy} = \begin{pmatrix} D^* & 0 & 0 & 0 & 0 \\ -\frac{\nu^* u}{\rho} & \frac{\nu^*}{\rho} & 0 & 0 & 0 \\ -\frac{4}{3}\frac{\nu^* v}{\rho} & 0 & \frac{4}{3}\frac{\nu^*}{\rho} & 0 & 0 \\ -\frac{\nu^* w}{\rho} & 0 & 0 & \frac{\nu^*}{\rho} & 0 \\ (A_{yy})_{5,1} & \frac{\nu^* u}{\rho} + (1-\gamma)\frac{\kappa^* u}{\rho} & \frac{4}{3}\frac{\nu^* v}{\rho} + (1-\gamma)\frac{\kappa^* v}{\rho} & \frac{\nu^* w}{\rho} + (1-\gamma)\frac{\kappa^* w}{\rho} & (\gamma-1)\frac{\kappa^*}{\rho} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \left((1-\gamma)\frac{\kappa^*}{\rho} + \eta^*\right) B_x & \left((1-\gamma)\frac{\kappa^*}{\rho} - \eta^*\right) B_y & \left((1-\gamma)\frac{\kappa^*}{\rho} + \eta^*\right) B_z & 0 \\ \eta^* & 0 & 0 & 0 \\ 0 & \eta^* & 0 & 0 \\ 0 & 0 & \eta^* & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{yz} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{2}{3}\frac{\nu^* w}{\rho} & 0 & 0 & -\frac{2}{3}\frac{\nu^*}{\rho} & 0 & 0 & 0 & 0 & 0 \\ -\frac{\nu^* v}{\rho} & 0 & \frac{\nu^*}{\rho} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{3}\frac{\nu^* vw}{\rho} & 0 & \frac{\nu^* w}{\rho} & -\frac{2}{3}\frac{\nu^* v}{\rho} & 0 & 0 & -\eta^* B_z & -\eta^* B_y & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{zx} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\nu^* w}{\rho} & 0 & 0 & \frac{\nu^*}{\rho} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{2}{3}\frac{\nu^* u}{\rho} & -\frac{2}{3}\frac{\nu^*}{\rho} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{3}\frac{\nu^* uw}{\rho} & -\frac{2}{3}\frac{\nu^* w}{\rho} & 0 & \frac{\nu^* u}{\rho} & 0 & -\eta^* B_z & 0 & -\eta^* B_x & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{zy} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\nu^* w}{\rho} & 0 & 0 & \frac{\nu^*}{\rho} & 0 & 0 & 0 & 0 & 0 \\ \frac{2}{3}\frac{\nu^* v}{\rho} & 0 & -\frac{2}{3}\frac{\nu^*}{\rho} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{3}\frac{\nu^* vw}{\rho} & 0 & -\frac{2}{3}\frac{\nu^* w}{\rho} & \frac{\nu^* v}{\rho} & 0 & 0 & -\eta^* B_z & -\eta^* B_y & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$
A_{zz} = \begin{pmatrix}
D^* & 0 & 0 & 0 & 0 \\
-\dfrac{\nu^* u}{\rho} & \dfrac{\nu^*}{\rho} & 0 & 0 & 0 \\
-\dfrac{\nu^* v}{\rho} & 0 & \dfrac{\nu^*}{\rho} & 0 & 0 \\
-\dfrac{4}{3}\dfrac{\nu^* w}{\rho} & 0 & 0 & \dfrac{4}{3}\dfrac{\nu^*}{\rho} & 0 \\
(A_{zz})_{5,1} & \dfrac{\nu^* u}{\rho} + (1-\gamma)\dfrac{\kappa^* u}{\rho} & \dfrac{\nu^* v}{\rho} + (1-\gamma)\dfrac{\kappa^* v}{\rho} & \dfrac{4}{3}\dfrac{\nu^* w}{\rho} + (1-\gamma)\dfrac{\kappa^* w}{\rho} & (\gamma-1)\dfrac{\kappa^*}{\rho} \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix}
$$

$$
\begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
\left((1-\gamma)\dfrac{\kappa^*}{\rho} + \eta^*\right) B_x & \left((1-\gamma)\dfrac{\kappa^*}{\rho} + \eta^*\right) B_y & \left((1-\gamma)\dfrac{\kappa^*}{\rho} - \eta^*\right) B_z & 0 \\
\eta^* & 0 & 0 & 0 \\
0 & \eta^* & 0 & 0 \\
0 & 0 & \eta^* & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}
$$

où les coefficients manquants sont :

$$
\begin{aligned}
(A_{xx})_{5,1} &= \frac{\kappa^*}{\rho}\left((\gamma-1)\frac{\overrightarrow{u}^2}{2} - \frac{p}{\rho}\right) - \frac{1}{3}\frac{\nu^* u^2}{\rho} - \frac{\nu^* \overrightarrow{u}^2}{\rho} \\
(A_{yy})_{5,1} &= \frac{\kappa^*}{\rho}\left((\gamma-1)\frac{\overrightarrow{u}^2}{2} - \frac{p}{\rho}\right) - \frac{1}{3}\frac{\nu^* v^2}{\rho} - \frac{\nu^* \overrightarrow{u}^2}{\rho} \\
(A_{zz})_{5,1} &= \frac{\kappa^*}{\rho}\left((\gamma-1)\frac{\overrightarrow{u}^2}{2} - \frac{p}{\rho}\right) - \frac{1}{3}\frac{\nu^* w^2}{\rho} - \frac{\nu^* \overrightarrow{u}^2}{\rho}
\end{aligned}
$$

Résumé

Au cours de ce travail, nous nous sommes attaché à la résolution numérique des équations de la Magnétohydrodynamique (MHD) auxquelles s'ajoute une loi hyperbolique de transport des erreurs de divergence. La première étape consista à symétriser le nouveau système de la MHD idéale afin d'en étudier le système propre, ce qui fut l'occasion de rappeler le rôle de l'entropie au niveau de ce calcul comme à celui de l'inégalité de Clausius-Duhem. La suite de cette thèse eut pour objectif la résolution de ces équations idéales à l'aide de schémas distribuant le résidu (notés $\mathcal{RD}$). Les quatre principaux schémas connus furent testés, et nous avons montré entre autres que le schéma N, qui a fait ses preuves sur les équations d'Euler en mécanique des fluides, n'était pas adapté aux équations de la MHD. Les stratégies classiques de limitation et de stabilisation purent être revisitées à ce moment. Les équations étant instationnaires, il fallut intégrer une discrétisation en temps et une distribution spatiale des termes d'évolution (et d'éventuelles sources). Nous avons d'emblée opté pour une approche implicite permettant d'être performant sur les simulations longues des expériences de tokamaks, et de traiter la correction de la divergence d'une manière originale et efficace. Les problèmes de convergence de la méthode de Newton-Raphson n'ayant pas été pleinement résolus, nous nous sommes tournés vers une alternative explicite de type Runge-Kutta. Enfin, nous avons réétabli les principes de la montée en ordre (en théorie, jusqu'à des ordres arbitraires, en prenant en compte le phénomène de Gibbs) à l'aide de tout type d'élément fini (bien construit) 2D ou 3D, sans avoir pu valider tous ces aspects. Nous avons également pris en compte les équations complètes de la MHD réelle classique (i.e. sans effet Hall) à l'aide d'un couplage $\mathcal{RD}$/Galerkin.

Mots-clés

Simulation, Magnétohydrodynamique, schémas distribuant le résidu, instationnaire, ordres élevés, correction de la divergence, MHD dissipative (complète)

Summary

During this thesis, we worked on the numerical resolution of the Magnetohydrodynamic (MHD) equations, to which we added a hyperbolic transport equation for the divergence errors of the magnetic field. The first step consisted in symmetrizing the new ideal MHD system in order to study its eigensystem, which was the opportunity to remind the role of the entropy in this calculation as well as in the Clausius-Duhem inequality. Next, we aimed at solving these ideal equations by the mean of Residual Distribution ($\mathcal{RD}$) schemes. The four main schemes were tested, and we showed among other things that the N scheme (although it has been proven very efficient with Euler equations in Fluid Mechanics) could not give satisfying results with the MHD equations. Classical strategies for the limitation and the stabilization were revisited then. Moreover, since we dealt with unsteady equations, we had to formulate a time discretization and a spatial distribution of the unsteady terms (as well as possible sources). We first choosed an implicit approach allowing us to be powerful on the long simulations needed for tokamak experiments, and to treat the divergence cleaning part in an original and efficient way. The convergence problems of our Newton-Raphson algorithm having not been fully resolved, we turned to an explicit alternative (Runge-Kutta type). Finally, we discussed about the principles of higher order schemes (theoretically, up to arbitrary orders, taking into account the Gibbs phenomenon) thanks to any type of 2D or 3D finite element (properly defined), without having been able to to validate all these aspects. We also implemented the dissipative part of the full MHD equations (in the classical sense, i.e. omitting the Hall effect) by the use of a $\mathcal{RD}$/Galerkin coupling.

Keywords

Simulation, Magnetohydrodyanmics, Residual Distribution schemes, unsteady flows, high orders, divergence cleaning, dissipative (full) MHD