

N° d'ordre : 2113

THESE

présentée à

L'UNIVERSITÉ BORDEAUX I

École Doctorale de Mathématiques

par **Luc MIEUSSENS**

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : MATHÉMATIQUES APPLIQUÉES

**MODÈLES À VITESSES DISCRÈTES ET MÉTHODES NUMÉRIQUES
POUR L'ÉQUATION DE BOLTZMANN-BGK**

Soutenue le mercredi 29 septembre 1999

Après avis de :

P. DEGOND, Directeur de Recherches
P. LE TALLEC, Professeur

Université Toulouse 3
École Polytechnique

Rapporteurs

Devant la commission d'examen formée de :

F. POUPAUD, Professeur
R. ABGRALL, Professeur
P. CHARRIER, Professeur
P. DEGOND, Directeur de Recherches
B. DUBROCA, Ingénieur de Recherches
G. DUFFA, Expert Senior
P. LE TALLEC, Professeur

Université de Nice Sophia-Antipolis
Université de Bordeaux I
Université de Bordeaux I
Université Toulouse 3
CEA/CESTA
CEA/CESTA
École Polytechnique

Président
Rapporteur
Examineurs

Remerciements

Je remercie très sincèrement Pierre Charrier pour m'avoir dirigé pendant cette thèse, toujours avec sérieux, rigueur, mais aussi avec humour et sympathie. J'ai beaucoup apprécié sa conception des mathématiques appliquées, grâce à laquelle j'ai pu aborder aussi bien des problèmes d'analyse que d'analyse numérique ou de calcul scientifique. Évidemment, un grand merci aussi à Bruno Dubroca pour tout l'aide qu'il m'a apportée au CESTA. Il m'a fait profiter de ses nombreuses connaissances, tant mathématiques que numériques ou informatiques, et m'aura décidément beaucoup appris. Merci enfin à tous les deux pour toutes les personnalités scientifiques qu'ils m'ont fait rencontrer, et pour les différentes conférences auxquelles ils m'ont fait participer. Grâce à eux, j'ai vécu une expérience de thèse particulièrement riche.

Je suis très reconnaissant à Pierre Degond et Patrick Le Tallec d'avoir accepté de rapporter cette thèse (pendant le mois de juillet). Merci aussi à Patrick Le Tallec pour m'avoir permis de participer à la rencontre Franco-Russe d'Akademgorodok.

J'ai beaucoup apprécié que Frédéric Poupaud et Rémi Abgrall acceptent de faire partie du jury, et je remercie l'expert senior Georges Duffa, autant pour avoir participé au jury que pour ses précieux conseils de physicien.

Il me faut aussi souligner que cette thèse a été cofinancée par le CEA-CESTA et la région Aquitaine. J'ai travaillé pendant trois ans au service informatique du CESTA, et je remercie particulièrement Gérard Gallice pour nos discussions scientifiques et footballistiques, Thierry Launay pour son aide précieuse sur le code DSMC, et Christine Chavas (ma secrétaire préférée). Dans ce service, j'ai croisé la route de nombreux thésards. En particulier, sous le regard de Michel Delpech et du Marsupilami, j'ai partagé le mythique bureau A115 avec Jean-Luc Feugeas, qui m'a beaucoup aidé au début de cette thèse, tant scientifiquement que par ses conseils sur la vie en général. Je regretterai longtemps les matches de lancer de boulettes avec Stéphane, ainsi que les duels avec les trois desperados Philippe, Frédéric et Christophe. Bonne chance à tous.

Par ailleurs, je remercie Jacques Schneider pour son éclairage précieux sur les méthodes à vitesses discrètes, ainsi que mon traducteur officiel Fabrice Véron.

Enfin, je n'oublie pas que j'ai fait cette thèse à l'issue d'un stage de DEA (avec Pierre Charrier et Bruno Dubroca). Le choix de ce stage m'a été conseillé par Bernard Hanouzet, et je l'en remercie très sincèrement.

*À mes parents, mes grands-parents, mes sœurs,
À mes amis,
À Karine.*

Table des matières

Introduction	11
1 Modèle à vitesses discrètes pour l'équation BGK	15
1.1 Equation BGK	16
1.1.1 Définitions	16
1.1.2 Nombre de Knudsen	16
1.1.3 Equation	17
1.1.4 Conditions aux limites	18
1.2 Construction du modèle BGK discret	19
1.2.1 Principe de discrétisation	19
1.2.2 Distribution d'équilibre discrète	19
1.3 Propriétés du modèle	22
1.3.1 Positivité, conservation, et entropie	22
1.3.2 Une grille de vitesse admissible	23
1.3.3 Bornes sur la vitesse et la température	25
1.3.4 Une remarque sur les écoulements plans	26
1.4 Conditions aux limites discrètes	27
1.4.1 Réflexion diffuse	27
1.4.2 Réflexion spéculaire	28
2 Etude théorique du modèle BGK discret	29
2.1 Notations et principaux résultats	30
2.2 Consistance de l'approximation de la maxwellienne	32
2.3 Un résultat d'existence et d'unicité pour le modèle discret	35
2.4 Un résultat de convergence pour le modèle BGK discret	38
3 Discrétisation de l'opérateur de transport axisymétrique	41
3.1 Equation de transport et lois de conservation	42
3.2 Discrétisation de f et définition des quantités macroscopiques	46
3.3 Discrétisation de la dérivée en vitesse	47
3.3.1 Discrétisation décentrée de l'équation non conservative	49
3.3.2 Modification "trigonométrique"	50
3.3.3 Discrétisation centrée "trigonométrique" de l'équation conservative	51
3.3.4 Discrétisation décentrée "trigonométrique" de l'équation conservative: conservation et entropie	52
3.4 Approximation de la vitesse radiale	53
3.4.1 Discrétisation centrée de l'équation conservative	54

3.4.2	Discrétisation décentrée de l'équation non conservative	55
3.5	Conclusion	56
4	Discrétisation de termes de collision BGK généralisés	57
4.1	Nombre de Prandtl pour l'équation BGK	57
4.2	Modèle ES	59
4.2.1	Modèle continu	59
4.2.2	Discrétisation	60
4.3	Modèles à deux temps de relaxation	62
4.3.1	Principe	62
4.3.2	Modèle de Levermore	62
4.3.3	Modèle aux 14 moments	63
4.4	Conclusion	66
5	Méthodes numériques pour la résolution du modèle BGK à vitesses discrètes	69
5.1	Schéma explicite	70
5.2	Résultat de convergence	72
5.2.1	Estimations	74
5.2.2	Convergence vers la solution	77
5.3	Schéma implicite linéarisé pour les écoulements stationnaires	80
5.3.1	Ecriture du schéma	80
5.3.2	Résolution du système linéaire	82
5.4	Calcul de $\alpha_{i,j}^n$	84
5.5	Prise en compte des conditions aux limites	84
6	Résultats numériques	87
6.1	Description du code et choix des paramètres	87
6.1.1	Implémentation	87
6.1.2	Détermination des bornes et du pas de la grille de vitesse	87
6.1.3	Paramètres de la méthode	89
6.2	Influence des paramètres	90
6.2.1	Influence de la grille des vitesses	90
6.2.2	Influence du maillage en espace	95
6.2.3	Comparaison ordre 1 - ordre 2 en espace	97
6.2.4	Comparaison des différents schémas en temps	98
6.3	Comparaison des schémas pour l'équation BGK axisymétrique	99
6.3.1	Etude sur un écoulement 1D	99
6.3.2	Evaluation du gain par rapport au 2D plan	106
6.3.3	Etude sur un écoulement 2D	106
6.4	Validation physique	108
6.4.1	Comparaison avec des résultats expérimentaux	108
6.4.2	Comparaison avec DSMC, NS, et autres méthodes	110
6.4.3	Influence du nombre de Prandtl	120
6.4.4	Domaine de validité de la méthode	124
6.5	Avantage de BGK pour certains écoulements	132
6.5.1	Écoulement en régime dense	132
6.5.2	Calcul d'une recirculation	137
6.6	Écoulements hypersoniques	141

Table des matières	9
Conclusion	149
Bibliographie	151

Introduction

L'étude aérodynamique de corps de rentrée, de navette spatiale ou de sondes planétaires conduit à considérer des régimes où le comportement du fluide ne peut pas être prédit correctement par les modèles classiques de la mécanique des milieux continus comme les équations de Navier-Stokes. C'est le cas à haute altitude où l'atmosphère est raréfiée mais aussi à plus basse altitude dans les régimes transitionnels où ces équations sont mises en défaut dans les chocs ou dans la couche limite. On doit alors utiliser un modèle cinétique reposant sur une description microscopique du fluide considéré comme un ensemble de particules. Cependant ces équations sont plus complexes et plus coûteuses à résoudre numériquement. L'objectif de ce travail est de proposer dans ce contexte de nouveaux modèles numériques pour le calcul d'écoulements en régime transitionnel, de les analyser et de les valider.

Le modèle de base pour les écoulements raréfiés est l'équation de Boltzmann qui s'écrit

$$\partial_t f + v \cdot \nabla_x f = Q(f, f),$$

où f est la fonction de distribution des particules (voir Cercignani [12]). Le premier membre décrit le vol libre de ces particules et le second membre décrit les collisions. Cette équation est complexe à traiter tant du point de vue théorique que du point de vue numérique. En particulier l'aspect binaire des collisions est numériquement coûteux et implique en général des algorithmes dont la complexité est quadratique. Aussi dans un souci de simplification, en particulier dans les régimes transitionnels, différents auteurs ont tenté de proposer des équations plus simples, conservant les propriétés principales de l'équation de Boltzmann. Le modèle le plus connu est sans doute celui de Bathnagar-Gross-Krook (BGK). Ce modèle consiste à remplacer le terme de collisions entre particules par un terme représentant le phénomène de relaxation de f vers l'équilibre maxwellien $M[f]$. Le modèle BGK s'écrit alors

$$\partial_t f + v \cdot \nabla_x f = \frac{1}{\tau}(M[f] - f).$$

On peut vérifier que cette équation possède beaucoup des propriétés de l'équation de Boltzmann telles que les lois de conservation macroscopiques et la dissipation d'entropie. En outre l'équation BGK possède les mêmes limites fluides que l'équation de Boltzmann, bien qu'avec des valeurs différentes des coefficients de transport. Enfin le traitement numérique du terme de relaxation conduit à des algorithmes plus simples dont la complexité est linéaire. Pour ces raisons, ce modèle est bien adapté aux écoulements en régime transitionnels et il nous servira de base pour le développement de modèles numériques. Il reste que l'équation BGK n'est pas à même de modéliser précisément les forts déséquilibres cinétiques et une de nos préoccupations sera de préciser son domaine de validité.

L'approche la plus classique pour le calcul numérique d'écoulements raréfiés repose sur les méthodes de simulation directe ou de type Monte-Carlo. La plus connue est la méthode DSMC de

Bird [5], mais il existe des variantes comme la méthode de Nanbu [47] ou celle de l'équipe de Kaiserslautern [26]. L'appellation "simulation directe" signifie que l'on utilise directement la dynamique moléculaire sans passer par la modélisation du gaz par une équation de type Boltzmann. On peut décrire grossièrement ces méthodes en disant qu'on simule un paquet de molécules que l'on fait se déplacer en ligne droite pendant un pas de temps (phase de vol libre), puis que l'on fait s'entrechoquer "au hasard" (phase de collision). Ce processus est itéré jusqu'à stationnarisation, puis on fait des moyennes pour calculer les quantités macroscopiques. De fait, cette méthode peut être vue comme une approximation de l'équation de Boltzmann au moyen d'un splitting, et on peut même résoudre l'équation BGK de cette façon en remplaçant la phase de collision par une phase de relaxation. Ces méthodes ont atteint une grande maturité depuis leur invention dans les années 50, et l'approche phénoménologique sous-jacente permet de traiter un grand nombre de phénomènes complexes comme les mélanges de gaz, la chimie, etc. Cependant, elles ont plusieurs défauts bien connus : leur coût devient élevé pour des écoulements denses, elles calculent difficilement des écoulements lents comme la recirculation en arrière d'un corps, enfin à cause de leur caractère stochastique, les résultats sont entachés d'un bruit qui nuit à la précision et que l'on ne peut diminuer qu'au prix d'une augmentation importante du coût de calcul. Ces défauts ont donc motivé une approche différente, basée sur la résolution déterministe de l'équation de Boltzmann. C'est cette voie que nous avons choisie dans ce travail.

Dans l'approche déterministe pour les équations de Boltzmann ou de BGK, la principale difficulté provient du traitement de la variable v qui décrit \mathbb{R}^3 tout entier. En pratique, on est condamné à travailler avec un ensemble tronqué et discrétisé de vitesses et donc à introduire des modèles à répartition discrète de vitesse. Les modèles cinétiques à vitesses discrètes peuvent être vus comme un première étape vers la discrétisation des équation cinétiques, mais ils peuvent être vus également comme une simplification de la description physique du gaz, où l'on suppose que les molécules ne se déplacent que selon un nombre de vitesses fini. Cette seconde approche, la théorie cinétique discrète, a été introduite initialement pour simplifier l'étude mathématique de l'équation de Boltzmann (voir l'ouvrage de Gatignol [24] pour des références sur le sujet et de nombreux développements). L'approche "numérique", plus récente, est motivée par la recherche d'une alternative à la méthode DSMC (voir Walus [63] pour un exposé complet de cette approche pour l'équation de Boltzmann). On peut citer par exemple Rogier et Schneider [55] ou Buet [11] pour l'équation de Boltzmann, Degond et Lucquin-Desreux pour l'équation de Fokker-Planck. Signalons également tous les travaux sur les méthodes de type Lattice-Boltzmann qui relèvent d'une approche comparable bien qu'elles attaquent plus particulièrement des régimes continus (voir [40, 52]). Mentionnons enfin l'approche par méthodes particulières déterministes (voir Mas-Gallic [39] ou Issautier [29]) qui comporte une phase de vol libre (comme DSMC) et une phase dans laquelle l'opérateur de collision est approché par une formule de quadrature, où les particules elles-mêmes jouent le rôle de points de quadrature.

Dans ce travail, nous proposons un modèle à vitesses discrètes original pour l'équation BGK. Notre objectif est "numérique", mais nous empruntons à la cinétique discrète la recherche d'un modèle cinétique discret conservant la positivité et vérifiant exactement les lois de conservation et de dissipation d'entropie. Ces propriétés ont l'avantage d'apporter beaucoup de robustesse aux méthodes numériques issues de ce modèle et d'autoriser un pas en vitesse beaucoup plus grand que d'autres approches moins rigoureuses. Pour l'équation BGK, le point essentiel est de fixer correctement la notion "d'équilibre discret" jouant dans notre contexte le rôle de maxwellienne. Nous y arrivons en utilisant une version discrète du principe de minimisation d'entropie (chapitre 1, théorème 1.1). Toutes les bonnes propriétés du modèle BGK à vitesses discrètes sont ensuite déduites (théorème 1.2). Cette approche peut sans doute être utilisée dans tous les modèles cinétiques com-

portant des termes de relaxation vers un équilibre (comme par exemple en transfert radiatif).

Le modèle BGK à vitesses discrètes étant obtenu, il est intéressant aussi bien par rigueur mathématique que pour justifier sa pertinence physique, d'étudier sa convergence vers l'équation BGK continue en vitesse quand la grille de vitesse tend vers \mathbb{R}^3 . Pour cela, on donne tout d'abord un théorème de consistance de l'approximation de la maxwellienne par la fonction d'équilibre discrète (chapitre 2, théorème 2.1). Ensuite, on donne un théorème d'existence et d'unicité du modèle discret en vitesse qui vérifie les lois de conservation et d'entropie (théorème 2.2). Remarquons qu'un tel résultat n'est pas connu pour les modèles à vitesses discrètes de Boltzmann. La dernière propriété est un résultat de convergence de cette solution discrète vers une solution de l'équation BGK continue (théorème 2.3). Nous utilisons pour cela la preuve de stabilité des solutions de l'équation BGK donnée par Perthame [49], et quelques techniques tirées du travail de Mischler sur un modèle discret de l'équation de Boltzmann [46].

La prise en compte d'écoulement axisymétriques est très importante en pratique et tout particulièrement pour les applications rencontrées au CEA-CESTA. Ceci nous a conduit à étudier ce problème.

Un aspect mathématiquement très intéressant des écoulements axisymétriques est que le passage des coordonnées cartésiennes aux coordonnées cylindriques fait apparaître un terme de gradient en vitesse de la fonction de distribution dans le terme de transport. Contrairement à l'équation en coordonnées cartésiennes, la vitesse ne joue alors plus seulement le rôle d'un paramètre. La façon dont on discrétise ce gradient de vitesse lors de la construction du modèle à vitesses discrètes peut faire perdre les lois de conservation, la positivité, ou la dissipation d'entropie.

Nous proposons alors plusieurs nouveaux schémas de discrétisation du terme de transport pour résoudre ce problème (chapitre 3). Notamment, nous introduisons deux discrétisations qui vérifient les propriétés de positivité, de conservation de la densité, de la quantité de mouvement axiale et de l'énergie, ainsi que de dissipation d'entropie. A notre connaissance, c'est la première fois que sont proposés des schémas possédant toutes ces propriétés à la fois, notamment la propriété d'entropie, qui pouvait sembler délicate à concilier avec les propriétés de conservation. De plus, nous insistons sur le fait que ces schémas sont des discrétisations du terme de transport, et peuvent donc s'appliquer à une large classe d'équations cinétiques (en particulier les équations de Boltzmann, de Fokker-Planck, ou celles du transfert radiatif).

Un autre problème important pour les applications physiques est la validité de la limite fluide de l'équation BGK. Il est bien connu que les équations de Navier-Stokes obtenues par un développement asymptotique de l'équation BGK n'ont pas des coefficients de transport physiquement corrects. Ceci traduit une relaxation vers l'équilibre des phénomènes thermiques et visqueux à la même vitesse, alors que physiquement l'une est plus lente que l'autre. Ce défaut peut se traduire par de mauvais résultats dans des cas où les phénomènes de transfert thermique sont importants.

Cependant, plusieurs modèles de type BGK ont été proposés dans la littérature pour introduire une deuxième vitesse de relaxation, notamment le modèle BGK-ES (pour "Ellipsoidal-Statistical") introduit par Holway [28]. Ce modèle vient d'être récemment repris par Andriès, Le Tallec et Perthame dans [2]. Ils ont montré une propriété de dissipation d'entropie dont la validité était jusqu'alors assez suspecte. Ce modèle possède donc toutes les propriétés de l'équation BGK, avec en plus une limite fluide correcte.

Nous avons alors appliqué notre approche de discrétisation en vitesse sur ce modèle BGK-ES. Nous montrons en particulier que pour préserver la relaxation vers l'équilibre, il est nécessaire de remplacer la température intervenant dans ce modèle par le tenseur d'énergie associé à l'équilibre

discret.

Cependant, le modèle BGK-ES est plus complexe que le modèle BGK, et nous n'obtenons pas de propriétés mathématiques aussi fortes que précédemment. Nous proposons alors un nouveau modèle de type BGK à deux vitesses de relaxation, basé sur les fonction d'équilibre généralisé de Levermore [36], qui possède une limite fluide correcte. Ce modèle présente néanmoins d'autres difficultés.

La dernière partie de ce travail concerne la résolution numérique des modèles proposés. On introduit d'abord un schéma explicite héritant des propriétés mathématiques fortes de l'équation BGK (positivité, conservation, dissipation d'entropie) et nous montrons qu'il converge vers la solution du modèle à vitesses discrètes. Puis nous construisons, pour le calcul des écoulements stationnaires, un schéma implicite linéarisé. Le schéma explicite est en effet extrêmement coûteux pour calculer de tels écoulements, du fait de la restriction sur le pas de temps nécessaire à la stabilité. Le schéma implicite que nous introduisons est semblable à celui de [65], mais nous apportons deux nouveautés essentielles :

- nous implicisons et linéarisons l'équilibre discret, ce qui permet d'accélérer la convergence vers l'état stationnaire,
- nous proposons une méthode itérative pour résoudre le très grand système linéaire qui apparaît dans le schéma. Nous mettons pour cela en évidence les structures creuses particulières des différentes matrices du système, liées aux rôles différents des variables d'espace et de vitesse. Cette méthode permet de faire converger rapidement le schéma implicite vers la solution stationnaire, en ne faisant que deux ou trois itérations du solveur linéaire.

Ces différentes méthodes numériques ont permis de développer un code de calcul (le code KISS, pour Kinetic Implicit Supersonic Simulation) autorisant le calcul d'écoulements 1D, 2D plan ou axisymétrique sur des maillages structurés curvilignes. Ce code sera utilisé au CESTA pour le dimensionnement de véhicules de rentrée et servira de maquette pour le développement d'un code industriel 3D.

Il a permis de valider les méthodes proposées, d'un point de vue numérique et physique. Nous montrons en particulier que le fait d'avoir un domaine de vitesse borné n'est pas forcément un frein à l'utilisation des méthodes déterministes. De plus, nous mettons en évidence un moyen simple de tester *a posteriori* si une grille de vitesse est adaptée à l'écoulement calculé, en visualisant la fonction de distribution. Nous montrons par ailleurs que le domaine de validité du modèle BGK comprend des écoulements denses et transitionnels en fort déséquilibre cinétique, allant jusqu'à des nombres de Knudsen locaux de 0.5. Nous comparons pour cela nos résultats avec des résultats obtenus par DSMC, ou avec des résultats expérimentaux. Finalement, nous montrons tout l'intérêt de notre approche pour des écoulements tels que la recirculation en arrière d'un corps. De tels écoulements sont très coûteux à calculer par les méthodes stochastiques de type DSMC, et sont même parfois inaccessibles. Les qualités de notre approche la rendent tout à fait envisageable pour ces problèmes.

Certains des résultats présentés dans ce travail ont déjà donné lieu à publication. L'existence d'un équilibre discret est parue dans une note aux Comptes Rendus de l'Académie des Sciences [16]. Un article accepté à Math. Models and Meth. in Applied Sciences [43] comporte l'essentiel des chapitres 1 et 5 et certains des résultats numériques du chapitre 6. Enfin, une note aux Comptes Rendus de l'Académie des Sciences sur la convergence du modèle à vitesses discrètes [44] est également acceptée pour publication.

Chapitre 1

Modèle à vitesses discrètes pour l'équation BGK

Pour résoudre numériquement l'équation BGK, la principale difficulté est la discrétisation de la variable de vitesse. La plupart de méthodes numériques existantes (comme celles d'Issautier [29], Aoki-Kanba-Takata [4], Yang-Huang [65]) ne vérifient pas les propriétés de conservation de la masse, de la quantité de mouvement, et de l'énergie, ainsi que la propriété d'entropie. Ceci aboutit à des algorithmes qui sont coûteux en raison de la grille de vitesse très fine nécessitée pour garantir la robustesse, en particulier avec les schémas implicites.

Une approche élégante qui permet d'avoir les propriétés de conservation et d'entropie avec des approximations de l'équation de Boltzmann est la théorie cinétique discrète. Elle a été beaucoup développée par Gatignol [24] qui a construit quelques modèles à vitesses discrètes de l'équation de Boltzmann. De plus, elle a démontré des propriétés de conservation et d'autres propriétés intéressantes. Ces modèles ont été initialement utilisés pour simplifier l'étude mathématique de l'équation de Boltzmann. Cependant, les méthodes numériques récentes développées par [55, 11] utilisent aussi des modèles à vitesses discrètes possédant les propriétés mentionnées ci-dessus.

Dans ce chapitre, nous présentons une approche similaire à celle de [55] et [11] pour le développement d'un modèle à vitesses discrètes de l'équation BGK. Nous proposons une discrétisation non-triviale de la maxwellienne, définie par un principe de minimisation d'entropie. Nous prouvons un résultat d'existence et d'unicité de cette approximation. Notre modèle à vitesses discrètes est donc conservatif et entropique. A notre connaissance, c'est la première fois qu'un tel modèle est présenté pour l'équation BGK (notons que la méthode conservative de Perthame-Coron [18] possède les propriétés mentionnées ci-dessus, mais ne peut pas être interprétée comme un modèle à vitesses discrètes). Par conséquent, contrairement aux méthodes non conservatives, notre modèle apparaît très économique.

Dans la section 1.2, nous donnons quelques définitions et notations utilisées constamment dans ce rapport. Nous rappelons les propriétés de l'équation BGK et de l'équilibre maxwellien, ainsi que quelques conditions aux limites. Nous présentons ensuite notre modèle à vitesses discrètes, basé sur la notion d'équilibre discret, et l'existence d'une fonction d'équilibre discrète est démontrée. Dans la section 1.3, nous étudions quelques propriétés du modèle discret. Nous donnons enfin dans la section 1.4 une interprétation discrète en vitesse des conditions aux limites.

1. La majeure partie de ce chapitre ainsi que le chapitre 5 ont fait l'objet d'un article intitulé *Discrete Velocity Model and Implicit Scheme for the BGK Equation of Rarefied Gas Dynamics*, à paraître dans *Math. Models and Meth. in Applied Sciences*.

1.1 Equation BGK

1.1.1 Définitions

Considérons un gaz constitué de molécules monoatomiques sans degré d'énergie interne, de même masse constante, et d'une même espèce chimique. Le principal outil permettant de décrire le gaz au niveau microscopique est la fonction $f(t, x, v)$ de distribution des vitesses des molécules. Elle représente la masse totale de molécules ayant la position x et la vitesse v par unité de volume de l'espace des phases. Les positions et les vitesses des molécules sont données dans \mathbb{R}^D . L'intégrale de toute fonction $g(v)$ est notée

$$\langle g \rangle = \int_{\mathbb{R}^D} g(v) dv.$$

Les quantités fluides usuelles sont obtenues par intégration de f sur l'espace des vitesses. Par exemple, la densité, la quantité de mouvement, et l'énergie totale du fluide sont données par

$$\rho = \langle f \rangle \in \mathbb{R}, \quad \rho u = \langle v f \rangle \in \mathbb{R}^D, \quad E = \langle \frac{1}{2} |v|^2 f \rangle \in \mathbb{R}. \quad (1.1)$$

Par conséquent, ces variables sont les $D + 2$ premiers moments de f par rapport à v . Les quantités microscopiques $1, v, \frac{1}{2}|v|^2$ (appelées invariants collisionnels) sont notées par le vecteur¹

$$\mathbf{m}(v) = (1, v, \frac{1}{2}|v|^2)^T \in \mathbb{R}^{D+2}. \quad (1.2)$$

Par analogie, les $D + 2$ premiers moments de f sont notés par le vecteur

$$\boldsymbol{\rho} = (\rho, \rho u, E)^T, \quad (1.3)$$

dont les composantes sont numérotées de 0 à $D + 1$: $\boldsymbol{\rho} = (\rho^{(i)})_{i=0..D+1}$. La définition (1.1) peut alors être réécrite sous une forme plus compacte

$$\boldsymbol{\rho} = \langle \mathbf{m} f \rangle, \quad (1.4)$$

et on dit alors que $\boldsymbol{\rho}$ est réalisable par f , ou que f réalise $\boldsymbol{\rho}$. Enfin, on définit l'entropie cinétique f par

$$H(f) = \langle f \log f \rangle, \quad (1.5)$$

et on rappelle que la température T et la pression p du gaz sont définies par

$$T = \frac{1}{DR} \langle |v - u|^2 f \rangle, \quad p = \rho RT, \quad (1.6)$$

où R est la constante du gaz.

1.1.2 Nombre de Knudsen

La description microscopique d'un gaz est nécessaire quand l'échelle de longueur associée aux gradients des quantités macroscopiques devient de l'ordre du libre parcours moyen. Pour préciser

1. Dans toute la suite de ce rapport, sauf mention contraire, les quantités écrites en gras sont des vecteurs de \mathbb{R}^{D+2} .

cette idée, on définit généralement le degré de raréfaction d'un gaz par le nombre de Knudsen, qui est le rapport entre le libre parcours moyen λ et une dimension caractéristique L de l'écoulement :

$$Kn = \frac{\lambda}{L}. \quad (1.7)$$

Cependant, la définition d'un nombre de Knudsen global pour tout l'écoulement, reposant sur une grandeur caractéristique souvent arbitraire, n'est pas toujours appropriée. En effet, généralement, les échelles de longueur de gradients sont supérieures au libre parcours moyen seulement dans certaines régions de l'écoulement. Il est plus pertinent de définir un nombre de Knudsen local, par exemple basé sur la longueur du gradient de densité

$$L = \frac{\rho}{\frac{d\rho}{dx}}.$$

On considère que l'erreur obtenue en utilisant les équations de Navier-Stokes devient significative quand le nombre de Knudsen local est plus grand que 0.1 (cf. Bird [5]).

1.1.3 Equation

L'équation classique décrivant l'évolution de f est l'équation de Boltzmann. Celle-ci traduit la variation de f au cours du temps due aux collisions binaires entre les molécules. Cette équation étant très complexe, des équations simplifiées, appelées équations modèles, ont été développées par différents auteurs. C'est le cas de l'équation BGK, introduite en 1954 par Bathnagar, Gross et Krook dans [51]. Les auteurs de ce modèle ont considéré que l'effet des collisions étant de ramener le gaz dans un état d'équilibre dit "maxwellien" (décrit par une fonction de distribution particulière donnée ci-dessous), on peut remplacer le terme de collision de l'équation de Boltzmann par un terme de relaxation vers l'équilibre maxwellien. L'équation BGK est alors la suivante

$$\partial_t f + v \cdot \nabla_x f = \frac{1}{\tau} (M[\boldsymbol{\rho}] - f), \quad (1.8)$$

où $\tau = \tau(\rho, T)$ est le temps de relaxation que l'on précisera par la suite (cf. chapitre 4). La distribution maxwellienne $M[\boldsymbol{\rho}]$ représentant l'équilibre local du gaz est définie par

$$M[\boldsymbol{\rho}] = \frac{\rho}{(2\pi RT)^{\frac{D}{2}}} \exp\left(-\frac{|v-u|^2}{2RT}\right). \quad (1.9)$$

Cette distribution est une gaussienne isotrope, qui a les mêmes $D+2$ premiers moments que f , *i.e.* $\langle \mathbf{m} M[\boldsymbol{\rho}] \rangle = \boldsymbol{\rho}$.

On peut facilement constater que $M[\boldsymbol{\rho}]$ est l'unique solution du problème de minimisation d'entropie suivant (voir par exemple [49])

$$(\mathcal{P}) \quad H(M[\boldsymbol{\rho}]) = \min \{H[g], g \geq 0 \text{ t.q. } \langle \mathbf{m} g \rangle = \boldsymbol{\rho}\}. \quad (1.10)$$

Cela signifie simplement que l'équilibre local minimise l'entropie de tous les états microscopiques possibles ayant les mêmes propriétés macroscopiques. Notons que ce problème de minimisation ne dépend que du vecteur de moments $\boldsymbol{\rho}$. Il est aisé de voir qu'il admet une unique solution pour tout vecteur de moments $\boldsymbol{\rho}$ si et seulement si $\rho, T > 0$. De plus, on obtient une expression de $M[\boldsymbol{\rho}]$ équivalente à (1.9) en résolvant (\mathcal{P}) par multiplicateurs de Lagrange

$$M[\boldsymbol{\rho}] = \exp(\boldsymbol{\alpha} \cdot \mathbf{m}(v)), \quad (1.11)$$

où les composantes du vecteur $\boldsymbol{\alpha}$ sont les multiplicateurs de Lagrange du problème (à une constante près). Ces multiplicateurs sont des fonctions de t et x , et peuvent être exprimés en terme de ρ, u, T par la relation inversible

$$\boldsymbol{\alpha} = \left(\log \left(\frac{\rho}{(2\pi RT)^{D/2}} \right) - \frac{|u|^2}{2RT}, \frac{u}{RT}, -\frac{1}{RT} \right)^T. \quad (1.12)$$

Avec cette caractérisation de l'équilibre local maxwellien, on obtient (au moins formellement) les propriétés suivantes de conservation de la masse, quantité de mouvement, énergie, et de dissipation d'entropie

$$\partial_t \langle \mathbf{m} f \rangle + \nabla_x \langle \mathbf{m} v f \rangle = 0, \quad (1.13)$$

$$\partial_t \langle f \log f \rangle + \nabla_x \langle v f \log f \rangle \leq 0. \quad (1.14)$$

De plus, il est possible de vérifier que les solutions de (1.8) sont positives.

Nous insistons sur le fait que dans un schéma numérique, la préservation de ces propriétés est essentielle pour obtenir une discrétisation robuste et économique. En fait, la plupart des méthodes couramment utilisées (particulaires, ordonnées discrètes ou différences finies) ne les satisfont pas. Nous présenterons dans les sections suivantes un modèle à vitesses discrètes pour lequel toutes ces propriétés sont effectivement vérifiées.

1.1.4 Conditions aux limites

Le problème de l'interaction d'un gaz avec une paroi est un sujet difficile, encore l'objet aujourd'hui de nombreuses recherches en mécanique. C'est pourquoi on ne considère souvent dans la pratique que des conditions aux limites relativement simples. La première est la condition dite spéculaire. Elle se traduit par la réflexion des particules incidentes sur une paroi à la façon de boules de billard; en d'autres termes, la distribution des molécules réfléchies (c.-à-d. de vitesse v telle que $v \cdot n(x) > 0$, où $n(x)$ est la normale à la paroi dirigée vers le fluide) est fonction de la distribution des molécules incidentes suivant la loi

$$f(t, x, v) = f(t, x, \mathcal{R}v), \quad v \cdot n(x) > 0, \quad (1.15)$$

où $\mathcal{R}v = v - 2(v \cdot n(x))n(x)$ est la vitesse symétrique de v par rapport à la paroi en x . Cette condition n'est généralement pas réaliste, et l'on considère souvent la condition dite diffuse. On suppose alors que les molécules incidentes sont absorbées par la paroi qui se comporte comme un gaz à l'équilibre. Ces molécules sont réémises à la température T_w de la paroi avec une vitesse aléatoire en moyenne égale à celle de la paroi u_w , selon une distribution maxwellienne

$$f(t, x, v) = \phi M[\boldsymbol{\rho}_w](v), \quad v \cdot n(x) > 0, \quad (1.16)$$

où $\boldsymbol{\rho}_w = (1, u_w, \frac{1}{2}|u_w|^2 + \frac{D}{2}RT_w)$ et ϕ est un paramètre déterminé pour que le flux de masse à travers la paroi soit nul, *i.e.*

$$\phi = - \left(\int_{v \cdot n(x) < 0} v \cdot n(x) f(t, x, v) dv \right) \left(\int_{v \cdot n(x) > 0} v \cdot n(x) M[\boldsymbol{\rho}_w](v) dv \right)^{-1}. \quad (1.17)$$

Enfin, on peut définir une troisième condition aux limites en couplant les deux précédentes : c'est la condition dite d'accommodation de Maxwell

$$f(t, x, v) = a f(t, x, \mathcal{R}v) + (1 - a) \phi M[\boldsymbol{\rho}_w](v), \quad v \cdot n(x) > 0, \quad (1.18)$$

où a est le paramètre d'accommodation compris entre 0 et 1.

1.2 Construction du modèle BGK discret

1.2.1 Principe de discrétisation

Soit \mathcal{K} un ensemble de N multi-indices de \mathbb{N}^D , défini par $\mathcal{K} = \{k = (k^{(i)})_{i=1..D}, k^{(i)} \leq K^{(i)}\}$, où $\{K^{(i)}\}$ sont des bornes données. Nous définissons un ensemble de vitesses discrètes \mathcal{V} de \mathbb{R}^D par

$$\mathcal{V} = \{v_k = k\Delta v + a, k \in \mathcal{K}\}, \quad (1.19)$$

où a est un vecteur arbitraire de \mathbb{R}^D et Δv est un scalaire². Les quantités microscopiques sur \mathcal{V} sont notées $\mathbf{m}_k = \mathbf{m}(v_k) = (1, v_k, \frac{1}{2}|v_k|^2)^T$. La fonction de distribution ‘‘continue’’ en vitesse f est alors remplacée par le vecteur à N composantes $f_{\mathcal{K}}(t, x) = (f_k(t, x))_{k \in \mathcal{K}}$, où chaque composante $f_k(t, x)$ est supposée être une approximation de $f(t, x, v_k)$. Les quantités fluides sont définies comme dans le cas continu, sauf que les intégrales de \mathbb{R}^D sont remplacées par des sommes discrètes sur \mathcal{V} . Ainsi, posant

$$\langle g \rangle_{\mathcal{K}} = \sum_{k \in \mathcal{K}} g_k \Delta v^D \quad (1.20)$$

pour tout vecteur $g \in \mathbb{R}^N$, nous définissons les moments discrets et l’entropie discrète de $f_{\mathcal{K}}$ par

$$\boldsymbol{\rho}_{\mathcal{K}} = \langle \mathbf{m} f_{\mathcal{K}} \rangle_{\mathcal{K}}, \quad (1.21)$$

$$H_{\mathcal{K}}(f_{\mathcal{K}}) = \langle f_{\mathcal{K}} \log f_{\mathcal{K}} \rangle_{\mathcal{K}}. \quad (1.22)$$

Notre modèle BGK discret en vitesse est alors le système de N équations

$$\partial_t f_k + v_k \cdot \nabla_x f_k = \frac{1}{\tau} (\mathcal{E}_k - f_k), \quad \forall k \in \mathcal{K}. \quad (1.23)$$

Le principal problème est de définir une approximation $\mathcal{E}_{\mathcal{K}}$ de l’équilibre maxwellien $M[\boldsymbol{\rho}]$, de sorte que les propriétés de conservation (1.13) et la propriété d’entropie (1.14) soient encore vérifiées.

1.2.2 Distribution d’équilibre discrète

Tout d’abord, il faut noter que l’approximation naturelle (utilisée par exemple par Yang and Huang dans [65])

$$\mathcal{E}_k = M[\boldsymbol{\rho}_{\mathcal{K}}](v_k), \quad \forall k \in \mathcal{K}, \quad (1.24)$$

ne peut satisfaire ces propriétés. Au lieu de cela, nous proposons d’utiliser la version discrète du problème de minimisation d’entropie (1.10). Soit $\mathcal{E}_{\mathcal{K}}$ définie par le minimum d’entropie discrète, avec les contraintes d’avoir les mêmes moments que $f_{\mathcal{K}}$, c.-à-d. $\mathcal{E}_{\mathcal{K}}$ est la solution du problème suivant

$$(\mathcal{P}_{\mathcal{K}}) \quad H_{\mathcal{K}}(\mathcal{E}_{\mathcal{K}}) = \min \{H_{\mathcal{K}}(g), g \geq 0 \in \mathbb{R}^N \text{ t.q. } \langle \mathbf{m} g \rangle_{\mathcal{K}} = \boldsymbol{\rho}_{\mathcal{K}}\}. \quad (1.25)$$

Evidemment, on doit vérifier que ce problème admet une solution unique et facilement calculable (résoudre directement $(\mathcal{P}_{\mathcal{K}})$ dans \mathbb{R}^N serait numériquement très coûteux).

Dans le cas continu, la condition $\rho, T > 0$ est suffisante pour caractériser la solution de (1.10) par la distribution maxwellienne. Cependant, ce n’est pas vrai dans le cas discret, car les calculs explicites ne sont pas possibles. Nous prouvons alors que sous une hypothèse naturelle sur \mathcal{V} , l’équilibre discret

2. Les résultats présentés dans cette section restent valides si Δv dépend de la direction (i) et de l’index k .

$\mathcal{E}_\mathcal{K}$ a une forme exponentielle *si, et seulement si*, une condition de “réalisabilité stricte” est vérifiée par $\rho_\mathcal{K}$. Ce résultat a été annoncé dans [41, 16, 42, 17].

Théorème 1.1. *Soit ρ un vecteur de moments de \mathbb{R}^{D+2} , tel que l'ensemble $E_\rho = \{g \geq 0 \in \mathbb{R}^N \text{ t.q. } \langle \mathbf{m}g \rangle_\mathcal{K} = \rho\}$ des distributions discrètes réalisant ρ est non vide. Alors le problème $(\mathcal{P}_\mathcal{K})$ admet une solution unique $\mathcal{E}_\mathcal{K}$ appelée équilibre discret. De plus, supposons que \mathcal{V} est tel que les quantités microscopiques satisfont*

$$\{\mathbf{m}_k, k \in \mathcal{K}\} \text{ est de rang } D + 2. \quad (1.26)$$

Alors il existe un unique vecteur α dans \mathbb{R}^{D+2} tel que $\mathcal{E}_\mathcal{K}$ s'écrit

$$\mathcal{E}_k = \exp(\alpha \cdot \mathbf{m}_k), \quad \forall k \in \mathcal{K}, \quad (1.27)$$

si et seulement si ρ est strictement réalisable, *i.e.*

$$\exists g \in E_\rho \text{ t.q. } g > 0. \quad (1.28)$$

Remarque 1.1. Grâce à ce résultat, le calcul de $\mathcal{E}_\mathcal{K}$ ne nécessite pas la solution d'un coûteux problème de minimisation dans \mathbb{R}^N : seul le calcul du vecteur α dans \mathbb{R}^{D+2} est nécessaire. Ce vecteur α est l'unique solution du système de $D + 2$ équations non-linéaires

$$\langle \mathbf{m} \exp(\alpha \cdot \mathbf{m}) \rangle_\mathcal{K} = \rho_\mathcal{K}, \quad (1.29)$$

puisque $\mathcal{E}_\mathcal{K}$ réalise $\rho_\mathcal{K}$. Ce système peut être résolu par un algorithme de Newton (voir chap. 5). Cette relation a déjà été utilisée pour la condition initiale de simulations numériques (voir [57, 10]). Cependant, à notre connaissance, c'est la première fois que le résultat d'existence et d'unicité est prouvé.

Remarque 1.2. Ce résultat est généralisé dans [16] pour des équilibres discrets associés à des moments d'ordre plus élevé. Cela permet de définir des versions discrètes des fermetures aux moments de Levermore (voir [36] pour les fermetures de Levermore et [15] pour la théorie discrète et des applications numériques).

Remarque 1.3. On montre dans la section suivante que la condition sur \mathcal{V} n'est pas très restrictive. Cependant, la condition (1.28) sur le vecteur de moments est plus restrictive que dans le cas continu. La raison de cette différence est principalement le nombre de degrés de liberté du problème : plus le nombre de vitesses discrètes est faible, moins il y a de degrés de liberté pour réaliser ρ , *i.e.* pour trouver g dans E_ρ , et donc plus il est difficile de trouver une distribution gaussienne discrète dans E_ρ . Par exemple, avec $D = 1$ et $N = 3$, si nous posons $\mathcal{V} = \{-1, 0, 1\}$ et $\rho = (2, 0, 2)$, alors E_ρ a un seul élément $g = (1, 0, 1)$. Par conséquent, l'équilibre discret $\mathcal{E}_\mathcal{K}$ est égal à g , et ne peut pas être une exponentielle (puisque $\mathcal{E}_2 = 0$), alors que $\rho, T > 0$ implique que l'équilibre continu est bien une maxwellienne. Dans la limite continue en vitesse, le nombre de degrés de liberté devient infini, et la condition (1.28) est toujours satisfaite pourvu que les conditions naturelles $\rho, T > 0$ soient vérifiées.

Preuve du théorème 1.1. Tout d'abord, notons que l'existence et l'unicité de $\mathcal{E}_\mathcal{K}$ est facilement obtenue puisque l'application $H_\mathcal{K}$ est clairement continue, coercive, et strictement coercive sur le sous-ensemble fermé E_ρ (nous n'écrivons pas la preuve de ces assertions).

La partie non triviale de ce théorème est la caractérisation de $\mathcal{E}_{\mathcal{K}}$. Pour cela, introduisons l'application $J(\boldsymbol{\beta}) = \langle \exp(\boldsymbol{\beta} \cdot \mathbf{m}) \rangle_{\mathcal{K}} - \boldsymbol{\beta} \cdot \boldsymbol{\rho}$ définie dans \mathbb{R}^{D+2} , et un second problème de minimisation

$$(\mathcal{P}'_{\mathcal{K}}) \quad \boldsymbol{\alpha} \text{ t.q. } J(\boldsymbol{\alpha}) = \min\{J(\boldsymbol{\beta}); \boldsymbol{\beta} \in \mathbb{R}^{D+2}\}. \quad (1.30)$$

Il est clair que ce problème implique le problème original $(\mathcal{P}_{\mathcal{K}})$ (voir (1.25)). En fait, une solution $\boldsymbol{\alpha}$ de $(\mathcal{P}'_{\mathcal{K}})$ satisfait nécessairement la relation d'extremum

$$J'(\boldsymbol{\alpha}) = \langle \mathbf{m} \exp(\boldsymbol{\alpha} \cdot \mathbf{m}) \rangle_{\mathcal{K}} - \boldsymbol{\rho} = 0. \quad (1.31)$$

Par conséquent, si l'on définit $\mathcal{E}_k = \exp(\boldsymbol{\alpha} \cdot \mathbf{m}_k)$ pour tout $k \in \mathcal{K}$, alors $\mathcal{E}_{\mathcal{K}}$ réalise $\boldsymbol{\rho}$. Finalement, cette forme exponentielle permet de conclure que $\mathcal{E}_{\mathcal{K}}$ est la solution de $(\mathcal{P}_{\mathcal{K}})$. Comme dans le cas continu, (cf. [49]), la convexité de $t \log t$ et la relation (1.31) impliquent que pour tout $g \in E_{\boldsymbol{\rho}}$

$$\begin{aligned} H_{\mathcal{K}}[\mathcal{E}_{\mathcal{K}}] &\leq H_{\mathcal{K}}[g] - \langle (1 + \boldsymbol{\alpha} \cdot \mathbf{m})(g - \mathcal{E}_{\mathcal{K}}) \rangle_{\mathcal{K}} \\ &= H_{\mathcal{K}}[g] + 0. \end{aligned}$$

Nous voulons à présent prouver l'existence et l'unicité d'une solution de $(\mathcal{P}'_{\mathcal{K}})$. L'avantage de ce second problème est que les conditions (1.26) et (1.28) apparaissent naturellement. Tout d'abord, nous prouvons que J est strictement convexe. Pour cela, il suffit de remarquer que la dérivée seconde de J satisfait

$$\begin{aligned} J''(\boldsymbol{\beta})(\boldsymbol{\xi}, \boldsymbol{\xi}) &= \boldsymbol{\xi}^T \langle \mathbf{m} \otimes \mathbf{m} \exp(\boldsymbol{\beta} \cdot \mathbf{m}) \rangle_{\mathcal{K}} \boldsymbol{\xi} \\ &= \sum_{k \in \mathcal{K}} \exp(\boldsymbol{\beta} \cdot \mathbf{m}_k) \left(\sum_i \xi^{(i)} \mathbf{m}_k^{(i)} \right)^2 \Delta v^D. \end{aligned}$$

Par conséquent, $J''(\boldsymbol{\beta})(\boldsymbol{\xi}, \boldsymbol{\xi}) \geq 0$ pour tout $\boldsymbol{\xi} \in \mathbb{R}^{D+2}$. De plus, l'hypothèse (1.26) s'écrit

$$\boldsymbol{\xi} \cdot \mathbf{m}_k = 0 \quad \forall k \in \mathcal{K} \Leftrightarrow \boldsymbol{\xi} = 0, \quad (1.32)$$

ce qui implique $J''(\boldsymbol{\beta})(\boldsymbol{\xi}, \boldsymbol{\xi}) > 0$ si $\boldsymbol{\xi} \neq 0$. Ceci prouve que $J''(\boldsymbol{\beta})$ est définie positive pour tout $\boldsymbol{\beta} \in \mathbb{R}^{D+2}$, et donc que J est strictement convexe. A présent, nous prouvons la coercivité de J dans les deux étapes suivantes.

étape 1: Dans cette première étape, nous montrons que J est coercive "dans chaque direction de \mathbb{R}^{D+2} ", c'est-à-dire

$$J(\lambda \boldsymbol{\omega}) = \sum_{k \in \mathcal{K}} \exp(\lambda \boldsymbol{\omega} \cdot \mathbf{m}_k) \Delta v^D - \lambda \boldsymbol{\omega} \cdot \boldsymbol{\rho} \xrightarrow{\lambda \rightarrow +\infty} +\infty$$

localement uniformément autour de tout $\boldsymbol{\omega}$ de la sphère unité S^{D+1} . Ce problème est divisé en deux cas. Premièrement, en supposant qu'il existe k_0 tel que $\boldsymbol{\omega} \cdot \mathbf{m}_{k_0} > 0$, nous obtenons $0 < c_1(\boldsymbol{\omega}) \leq \tilde{\boldsymbol{\omega}} \cdot \mathbf{m}_{k_0}$ pour tout $\tilde{\boldsymbol{\omega}}$ d'un petit voisinage $B(\boldsymbol{\omega}, \varepsilon(\boldsymbol{\omega}))$ de $\boldsymbol{\omega}$. Par conséquent

$$J(\lambda \tilde{\boldsymbol{\omega}}) \geq \exp(\lambda c_1(\boldsymbol{\omega})) \Delta v^D - \lambda c_2(\boldsymbol{\omega}, \boldsymbol{\rho}), \quad (1.33)$$

qui tend vers $+\infty$ uniformément sur $B(\boldsymbol{\omega}, \varepsilon(\boldsymbol{\omega}))$. Dans le second cas $\boldsymbol{\omega} \cdot \mathbf{m}_k \leq 0$ pour tout k , et la propriété (1.26) implique qu'il existe k_0 tel que $\boldsymbol{\omega} \cdot \mathbf{m}_{k_0} < 0$. Donc $\boldsymbol{\omega} \cdot \boldsymbol{\rho} < 0$, et il existe un voisinage $B(\boldsymbol{\omega}, \varepsilon(\boldsymbol{\omega}))$ de $\boldsymbol{\omega}$ dans lequel $\tilde{\boldsymbol{\omega}} \cdot \boldsymbol{\rho} \leq c_3(\boldsymbol{\omega}) < 0$. Par conséquent

$$J(\lambda \tilde{\boldsymbol{\omega}}) \geq -\lambda \tilde{\boldsymbol{\omega}} \cdot \boldsymbol{\rho} \geq -\lambda c_3(\boldsymbol{\omega}) \quad (1.34)$$

qui tend vers $+\infty$ uniformément sur $B(\boldsymbol{\omega}, \varepsilon(\boldsymbol{\omega}))$.

étape 2: Nous pouvons à présent prouver que J tend globalement uniformément vers $+\infty$ pour chaque direction $\boldsymbol{\omega}$. Considérons les voisinages $B(\boldsymbol{\omega}, \varepsilon(\boldsymbol{\omega}))$ définis ci-dessus pour tout $\boldsymbol{\omega}$ de S^{D+1} . Il est clair que la réunion de ces ensembles ouverts recouvre S^{D+1} . Sachant que S^{D+1} est compacte, cette réunion contient par conséquent un recouvrement fini, *i.e.* il existe $\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \dots, \boldsymbol{\omega}_L$ de S^{D+1} tel que

$$S^{D+1} \subset \bigcup_{i=1}^L B(\boldsymbol{\omega}_i, \varepsilon(\boldsymbol{\omega}_i)). \quad (1.35)$$

D'après l'étape précédente, nous savons que pour tout i et pour tout $M > 0$, il existe $R(\boldsymbol{\omega}_i, M) > 0$ tel que

$$\lambda \geq R(\boldsymbol{\omega}_i, M) \Rightarrow J(\lambda \tilde{\boldsymbol{\omega}}) \geq M \quad \forall \tilde{\boldsymbol{\omega}} \in B(\boldsymbol{\omega}_i, \varepsilon(\boldsymbol{\omega}_i)), \quad (1.36)$$

donc nous pouvons définir $R(M) = \max_{i=1..L} R(\boldsymbol{\omega}_i, M)$ qui est fini. D'après (1.36), il est donc clair que

$$\forall \tilde{\boldsymbol{\omega}} \in S^{D+1} : \quad \lambda \geq R(M) \Rightarrow J(\lambda \tilde{\boldsymbol{\omega}}) \geq M,$$

i.e. $J(\lambda \tilde{\boldsymbol{\omega}})$ tend vers l'infini uniformément en $\tilde{\boldsymbol{\omega}}$. La coercivité est alors évidente.

L'unicité dans \mathbb{R}^{D+2} d'une solution $\boldsymbol{\alpha}$ du problème $(\mathcal{P}'_{\mathcal{K}})$ est alors un classique résultat d'optimisation. Comme nous l'avons noté au début de la preuve, ceci implique que $\mathcal{E}_{\mathcal{K}} = (\exp(\boldsymbol{\alpha} \cdot \mathbf{m}_k))_{k \in \mathcal{K}}$ est l'unique solution de $(\mathcal{P}_{\mathcal{K}})$. Nous avons donc prouvé que sous l'hypothèse (1.26), la solution de $(\mathcal{P}_{\mathcal{K}})$ est caractérisée si la condition de réalisabilité stricte (1.28) est remplie. La réciproque est évidente, puisque $g = \mathcal{E}_{\mathcal{K}} = (\exp(\boldsymbol{\alpha} \cdot \mathbf{m}_k))_{k \in \mathcal{K}}$ est nécessairement strictement positive et réalise $\boldsymbol{\rho}$. \square

Le modèle est à présent complètement défini, excepté dans le cas où $E_{\boldsymbol{\rho}_{\mathcal{K}}}$ est vide. Dans ce cas, on peut poser $\mathcal{E}_{\mathcal{K}} = 0$, mais cela n'a pratiquement aucun intérêt, puisque le modèle contient implicitement le fait que $\boldsymbol{\rho}_{\mathcal{K}}$ est réalisé par $f_{\mathcal{K}} \geq 0$. On peut néanmoins remarquer que si $\boldsymbol{\rho}_{\mathcal{K}}$ n'est pas strictement réalisable, alors la définition de $\mathcal{E}_{\mathcal{K}}$ n'est pas entièrement satisfaisante, dans la mesure où l'on doit pratiquement résoudre un problème de minimisation dans \mathbb{R}^N . Cependant, on montre dans la section suivante que sous réserve d'une condition initiale raisonnable, alors $\boldsymbol{\rho}_{\mathcal{K}}$ reste strictement réalisable et donc que $\mathcal{E}_{\mathcal{K}}$ a la forme voulue.

1.3 Propriétés du modèle

1.3.1 Positivité, conservation, et entropie

Théorème 1.2. *Soit f_0 une fonction à valeurs dans \mathbb{R}^N strictement positives. Considérons le problème de Cauchy*

$$\partial_t f_k + v_k \cdot \nabla_x f_k = \frac{1}{\tau} (\mathcal{E}_k - f_k), \quad (1.37)$$

$$f_k(0, x) = f_k^0(x), \quad (1.38)$$

où $\mathcal{E}_\mathcal{K}$ est défini soit par $(\mathcal{P}_\mathcal{K})$, soit par $\mathcal{E}_\mathcal{K} = 0$ (si $E_{\rho_\mathcal{K}} = \emptyset$). Si ce problème admet une solution $f_\mathcal{K}$, alors on a (au moins formellement)

$$f_k(t, x) > 0 \quad \forall k, t, x, \quad (1.39)$$

$$\mathcal{E}_k = \exp(\boldsymbol{\alpha} \cdot \mathbf{m}_k) \quad \forall k, \quad (1.40)$$

$$\partial_t \langle \mathbf{m} f_\mathcal{K} \rangle_\mathcal{K} + \nabla_x \langle \mathbf{m} v f_\mathcal{K} \rangle_\mathcal{K} = 0, \quad (1.41)$$

$$\partial_t \langle f_\mathcal{K} \log f_\mathcal{K} \rangle_\mathcal{K} + \nabla_x \langle v f_\mathcal{K} \log f_\mathcal{K} \rangle_\mathcal{K} \leq 0. \quad (1.42)$$

Preuve. La preuve de ces assertions est semblable à celle du cas continu. Comme $\mathcal{E}_\mathcal{K}$ est toujours positive par définition, alors

$$\partial_t f_k + v_k \cdot \nabla_x f_k \geq -\frac{1}{\tau} f_k,$$

d'où

$$\frac{d}{dt} (f_k(t, x + tv_k)) \geq -\frac{1}{\tau} f_k(t, x + tv_k),$$

et l'on conclut à l'aide du lemme de Gronwall que

$$f_k(t, x) \geq f_k^0(x - tv_k) e^{-t/\tau} > 0.$$

On voit donc que dans ce cas, (1.28) est trivialement vérifiée (prendre $g = f_\mathcal{K}$) et donc que $\mathcal{E}_\mathcal{K}$ a bien une forme exponentielle. Pour les lois de conservation (1.41), il suffit de multiplier (1.37) par \mathbf{m}_k et de sommer sur k . Par définition de $\mathcal{E}_\mathcal{K}$, le terme source s'annule et on trouve (1.41). Pour la loi de dissipation d'entropie (1.42), on multiplie (1.23) par $1 + \log f_k$ et on somme sur k . On remarque alors que le second membre vérifie

$$\langle (\mathcal{E}_\mathcal{K} - f_\mathcal{K}) \log f_\mathcal{K} \rangle_\mathcal{K} = \langle (\mathcal{E}_\mathcal{K} - f_\mathcal{K}) (\log f_\mathcal{K} - \log \mathcal{E}_\mathcal{K}) \rangle_\mathcal{K} + \langle (\mathcal{E}_\mathcal{K} - f_\mathcal{K}) \log \mathcal{E}_\mathcal{K} \rangle_\mathcal{K}.$$

Le premier membre de droite est négatif ou nul, car la fonction \log est croissante, et le deuxième vaut $\boldsymbol{\alpha} \cdot \langle (\mathcal{E}_\mathcal{K} - f_\mathcal{K}) \mathbf{m} \rangle_\mathcal{K} = 0$. \square

Notons ici que ces propriétés nous permettent d'obtenir un résultat d'existence et d'unicité, ainsi qu'un résultat de convergence vers l'équation BGK continue (voir chapitre 2).

1.3.2 Une grille de vitesse admissible

Il nous semble nécessaire d'expliquer pourquoi la condition (1.26) sur \mathcal{V} n'est pas trop restrictive. Le résultat suivant montre qu'il est en fait suffisant de prendre une grille cartésienne ayant au moins deux points dans chaque direction (incluant 0), et au moins trois points dans une direction donnée (par exemple une grille avec au moins trois points dans chaque direction est suffisante).

Proposition 1.1. *Soit \mathcal{V} défini par (1.19) avec*

$$K^{(i)} \geq 1 \quad \forall i = 1..D. \quad (1.43)$$

S'il existe une direction i telle que

$$K^{(i)} \geq 2, \quad (1.44)$$

alors $\{\mathbf{m}_k, k \in \mathcal{K}\}$ satisfait (1.26).

Preuve. Supposons tout d'abord que $a = 0$ (voir (1.19)). De plus, quitte à renuméroter les axes, nous pouvons supposer qu'il existe au moins trois points dans la direction 1 (poser $i = 1$ dans (1.44)). Avec les hypothèses de la proposition, on peut trouver un sous-ensemble de $D + 2$ multi-indices $\{k_l\}_{l=1..D+2}$ de \mathcal{K} tel que (cf. figure 1.1) :

$$\begin{aligned} v_{k_1} &= 0, \\ v_{k_l} &= (v_{k_l}^{(j)} = \Delta v \text{ si } j < l, 0 \text{ sinon}) \quad \text{pour } l = 2..D + 1, \\ v_{k_{D+2}} &= (2\Delta v, 0, \dots, 0). \end{aligned}$$

Nous voulons montrer que $\{\mathbf{m}_{k_l}\}_{l=1..D+2}$ est une famille linéairement indépendante de \mathbb{R}^{D+2} , ce

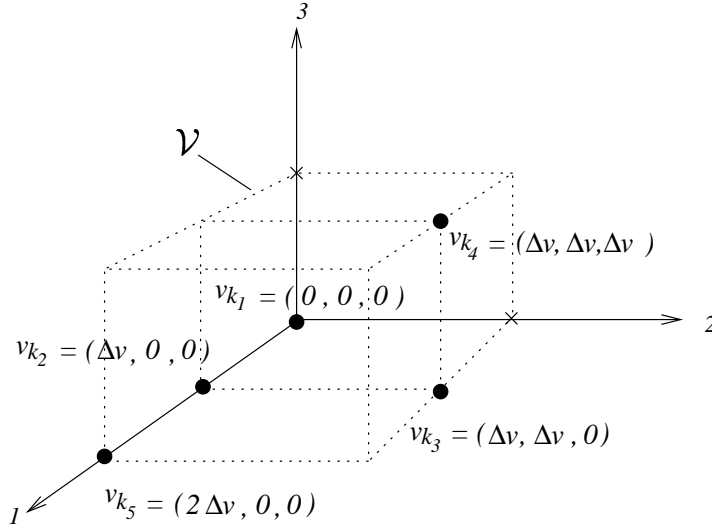


FIG. 1.1 – hypothèses de la proposition 1.1 pour $D = 3$: $K^{(1)} = 2$, $K^{(2)} = K^{(3)} = 1$, et donc $N = 12$.

qui prouvera la proposition. Considérons $\beta \in \mathbb{R}^{D+2}$ tel que $\beta \cdot \mathbf{m}_{k_l} = 0$ pour $l = 1..D + 2$, ce qui s'écrit $A\beta = 0$ sous forme matricielle, où

$$A = [\mathbf{m}_{k_1}, \dots, \mathbf{m}_{k_{D+2}}] = \begin{pmatrix} 1 & 1 & \dots & \dots & 1 & 1 \\ 0 & \Delta v & \dots & \dots & \Delta v & 2\Delta v \\ 0 & 0 & \dots & \dots & \Delta v & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & 0 & \Delta v & 0 \\ 0 & \frac{1}{2}\Delta v^2 & \dots & \dots & \frac{D}{2}|\Delta v|^2 & 2\Delta v^2 \end{pmatrix}.$$

On peut vérifier que le déterminant de A est $\det A = \Delta v^{D+2} \neq 0$. Ceci prouve que $\beta = 0$, et $\{\mathbf{m}_{k_l}\}_{l=1..D+2}$ est une famille linéairement indépendante, donc une base de \mathbb{R}^{D+2} .

Pour le cas $a \neq 0$, posons $\tilde{v}_k = v_k - a$ pour tout $k \in \mathcal{K}$. D'après le cas précédent, il est clair que $\{\tilde{\mathbf{m}}_k = \mathbf{m}(\tilde{v}_k)\}_{k=1..D+2}$ est une famille linéairement indépendante, ce qui signifie que $\tilde{A} = [\tilde{\mathbf{m}}_{k_1}, \dots, \tilde{\mathbf{m}}_{k_{D+2}}]$ est inversible. Posant $A = [\mathbf{m}_{k_1}, \dots, \mathbf{m}_{k_{D+2}}]$, nous avons le lemme suivant :

Lemme 1.1. *Il existe une matrice L triangulaire supérieure telle que $A^T = \tilde{A}^T L$ et $L_{ll} = 1$ pour $l = 0..D + 1$.*

On peut alors immédiatement en conclure que $\det A = \det \tilde{A} \times \det L \neq 0$. Il ne reste plus qu'à prouver le lemme.

Preuve du lemme. Soit $\lambda \in \mathbb{R}^{D+2}$, on a

$$\begin{aligned} (A^T \lambda)_i &= \sum_{j=0}^{D+1} \lambda^{(j)} m_{k_i}^{(j)} = \sum_{j=0}^{D+1} \lambda^{(j)} m^{(j)}(v_{k_i}) = \sum_{j=0}^{D+1} \lambda^{(j)} m^{(j)}(\tilde{v}_{k_i} + a) \\ &= \lambda^{(0)} + \sum_{j=1}^D \lambda^{(j)} (\tilde{v}_{k_i}^{(j)} + a^{(j)}) + \lambda^{(D+1)} \frac{1}{2} |\tilde{v}_{k_i} + a|^2 \\ &= \sum_{j=0}^{D+1} \gamma^{(j)} \tilde{m}_{k_i}^{(j)} = \tilde{A}^T \gamma \end{aligned}$$

où

$$\begin{aligned} \gamma^{(0)} &= \lambda^{(0)} + \sum_{j=1}^{D+1} \lambda^{(j)} a^{(j)} + \lambda^{(D+1)} \frac{1}{2} |a|^2, \\ \gamma^{(j)} &= \lambda^{(j)} + \lambda^{(D+1)} a^{(j)} \quad j = 1..D, \\ \gamma^{(D+1)} &= \lambda^{(D+1)}. \end{aligned}$$

Donc

$$\gamma = L\lambda \text{ où } L = \begin{pmatrix} 1 & a^{(1)} & \cdot & \cdot & \cdot & \cdot & a^{(D)} & \frac{1}{2}|a|^2 \\ & 1 & 0 & \cdot & \cdot & \cdot & 0 & a^{(1)} \\ & & \cdot & & & & & \cdot \\ & & & 1 & 0 & \cdot & 0 & a^{(i)} \\ & & & & \cdot & & & \cdot \\ & & & & & \cdot & & \cdot \\ & & & & & & 1 & a^{(D)} \\ & & & & & & & 1 \end{pmatrix}.$$

On a donc montré que

$$\forall \lambda \in \mathbb{R}^{D+2}, \quad A^T \lambda = \tilde{A}^T L \lambda$$

avec L triangulaire et $L_{l,l} = 1$ pour $l = 0..D + 1$, ce qui conclut la preuve du lemme et de la proposition. □

□

1.3.3 Bornes sur la vitesse et la température

Dans cette section, étant donné un ensemble de vitesses discrètes, nous étudions les propriétés des quantités macroscopiques, indépendamment du processus de collision (*i.e.* BGK ou Boltzmann). Le résultat suivant montre qu'une fois \mathcal{V} choisi, un modèle à vitesses discrètes ne peut pas décrire n'importe quel écoulement. Réciproquement, pour un écoulement donné, l'ensemble de vitesses discrètes doit être choisit avec attention pour permettre une représentation correcte.

Proposition 1.2. *Soit $f \in \mathbb{R}^N$ une fonction de distribution discrète positive. Alors les quantités macroscopiques vitesse et température associées à f dans \mathcal{V} par $u = \frac{1}{\rho} \langle v f \rangle_{\mathcal{K}}$ et $T = \frac{1}{DR\rho} \langle |v - u|^2 f \rangle_{\mathcal{K}}$ satisfont les estimations*

$$\min_{\mathcal{K}} v_k^{(i)} \leq u^{(i)} \leq \max_{\mathcal{K}} v_k^{(i)} \quad i = 1..D \quad (1.45)$$

$$\frac{1}{DR} \min_{\mathcal{K}} |v_k - u|^2 \leq T \leq \frac{1}{DR} \max_{\mathcal{K}} |v_k - u|^2. \quad (1.46)$$

Preuve. Par définition, on a

$$\rho u^{(i)} = \sum_{k \in \mathcal{K}} v_k^{(i)} f_k \Delta v^D, \quad \text{et} \quad \frac{DR}{2} \rho T = \sum_{k \in \mathcal{K}} \frac{1}{2} |v_k - u|^2 f_k \Delta v^D,$$

et sachant en outre que

$$\rho = \sum_{k \in \mathcal{K}} f_k \Delta v^D,$$

alors les relations (1.45) et (1.46) sont immédiates. \square

Ces inégalités montrent que la vitesse macroscopique et la température sont bornées supérieurement par des quantités respectivement proportionnelles à la vitesse maximum de la grille et à son carré. Cela implique que l'ensemble de vitesses discrètes doit être assez large pour prendre en compte les grandes vitesses et les grandes températures qui peuvent apparaître dans l'écoulement.

Une autre propriété, plus inattendue, d'un écoulement à vitesses discrètes est que la température est bornée inférieurement. Cette borne est en fait fortement liée à Δv , comme on peut le voir dans l'exemple suivant. Considérons un écoulement de Couette 1D plan (*i.e.* un écoulement entre deux plaques de même température T_w , dont l'une est immobile alors que l'autre se déplace à la vitesse u_w). Dans un tel écoulement, la vitesse macroscopique croît continûment de 0 à u_w . Cela implique qu'il existe une position x pour laquelle $\min_{\mathcal{K}} |u(x) - v_k| = \frac{1}{2} \Delta v$. Donc en ce point $T(x)$ est bornée inférieurement par $\frac{\Delta v^2}{4DR}$, quelque soit la température de la plaque. Par conséquent, une petite température T_w pourrait causer un saut de température surestimé au voisinage des plaques. Cet exemple suggère que Δv doit être suffisamment petit pour prendre en compte les basses températures pouvant apparaître dans l'écoulement.

Nous proposons dans la section 6.1.2 une stratégie de construction de la grille de vitesses qui tient compte de ces contraintes.

Remarque 1.4. En mécanique statistique, il est préférable de définir la température d'un gaz à vitesses discrètes par le coefficient c tel que $\exp(\alpha \cdot \mathbf{m}_k) = a \exp(-\frac{1}{2c} |b - v_k|^2)$ (voir [13]). Le résultat de la proposition précédente ne donne pas de borne sur cette température. Cependant, Rogier et Schneider [55] ont montré que pour une maxwellienne continue dans un domaine de vitesses borné, ce coefficient peut être soit positif, si le domaine est assez large, soit négatif si le domaine est trop petit. Il semble donc raisonnable de penser que c'est encore vrai dans le cas discret. Remarquons cependant que dans la suite, nous utiliserons seulement la définition de la température donnée dans la proposition 1.2.

1.3.4 Une remarque sur les écoulements plans

Cette remarque est motivée par le fait que dans le cas continu en vitesse (avec par exemple $D = 3$), si $u^{(3)} = 0$ (*i.e.* un écoulement plan), alors (1.12) implique qu'on a aussi $\alpha^{(3)} = 0$. Il est intéressant de vérifier si c'est le cas dans notre cadre discret en vitesse, puisque cela réduirait le coût

du calcul de l'équilibre discret. En effet, on obtiendrait alors un problème dans \mathbb{R}^{D+1} plutôt que dans \mathbb{R}^{D+2} . Cependant, comme une relation explicitement inversible entre $\boldsymbol{\rho}$ et $\boldsymbol{\alpha}$ n'est pas connue, la situation n'est pas aussi évidente que dans le cas continu. Le résultat suivant montre pourtant que la propriété reste vraie.

Proposition 1.3. *Soit \mathcal{V} une grille de vitesses vérifiant (1.26) et symétrique par rapport au j^{eme} hyperplan $\{v^{(j)} = 0\}$. Soit $\boldsymbol{\rho} \in \mathbb{R}^{D+2}$ un vecteur de moments strictement réalisable dans \mathcal{V} (au sens de (1.28)), tel que $\rho^{(j)} = 0$ (i.e. $u^{(j)} = 0$). Alors le vecteur des multiplicateurs de Lagrange $\boldsymbol{\alpha}$ du théorème 1.1 vérifie de façon équivalente $\alpha^{(j)} = 0$.*

Preuve. Remarquons que l'hypothèse de symétrie de \mathcal{V} signifie que pour tout k dans \mathcal{K} , il existe k' dans \mathcal{K} tel que $v_k^{(i)} = v_{k'}^{(i)}$ si $i \neq j$ et $v_k^{(j)} = -v_{k'}^{(j)}$. Alors

$$\begin{aligned} \rho^{(j)} &= \sum_{k \in \mathcal{K}} v_k^{(j)} \exp(\boldsymbol{\alpha} \cdot \mathbf{m}_k) \Delta v^D \\ &= \sum_{\substack{k \text{ t.q.} \\ v_k^{(j)} > 0}} v_k^{(j)} \exp(\boldsymbol{\alpha} \cdot \mathbf{m}_k) \Delta v^D + \sum_{\substack{k \text{ t.q.} \\ v_k^{(j)} < 0}} v_k^{(j)} \exp(\boldsymbol{\alpha} \cdot \mathbf{m}_k) \Delta v^D, \end{aligned}$$

et l'on remarque que

$$\begin{aligned} \boldsymbol{\alpha} \cdot \mathbf{m}_k &= \alpha^{(0)} + \alpha^{(1)} v_k^{(1)} + \dots + \alpha^{(D)} v_k^{(D)} + \alpha^{(D+1)} \frac{1}{2} |v_k|^2 \\ &= r_k + \alpha^{(j)} v_k^{(j)}, \end{aligned}$$

où r_k ne dépend pas de $v_k^{(j)}$. Par conséquent, l'hypothèse de symétrie implique

$$\begin{aligned} \rho^{(j)} &= \sum_{\substack{k \text{ t.q.} \\ v_k^{(j)} > 0}} v_k^{(j)} \exp(r_k) [\exp(\alpha^{(j)} v_k^{(j)}) - \exp(-\alpha^{(j)} v_k^{(j)})] \Delta v^D \\ &= \sum_{\substack{k \text{ t.q.} \\ v_k^{(j)} > 0}} v_k^{(j)} \exp(r_k) \sinh(\alpha^{(j)} v_k^{(j)}) \Delta v^D. \end{aligned}$$

Tous les termes de cette somme sont de même signe, et sont nuls si et seulement si $\alpha^{(j)}$ est nul, ce qui conclut la preuve. \square

1.4 Conditions aux limites discrètes

1.4.1 Réflexion diffuse

Cette condition de réflexion se traduit naturellement dans le cadre discret. En effet, suivant l'interprétation donnée section 1.1.4, on considère que les molécules incidentes sont réémises par la paroi selon une distribution d'équilibre discrète associée à la température T_w et à la vitesse u_w . On pose donc

$$f_k(t, x) = \phi \mathcal{E}_k[\boldsymbol{\rho}_w], \quad v_k \cdot n(x) > 0, \quad (1.47)$$

où $\rho_w = (1, u_w, \frac{1}{2}|u_w|^2 + \frac{D}{2}RT_w)$ et ϕ est un paramètre déterminé pour annuler le flux de masse à travers la paroi :

$$\phi = - \left(\sum_{\substack{k \ t, q \\ v_k \cdot n(x) < 0}} v_k \cdot n(x) f_k(t, x) \Delta v^D \right) \left(\sum_{\substack{k \ t, q \\ v_k \cdot n(x) > 0}} v_k \cdot n(x) \mathcal{E}_k[\rho_w] \Delta v^D \right)^{-1} \quad (1.48)$$

1.4.2 Réflexion spéculaire

Cette condition est plus délicate à généraliser aux vitesses discrètes que la précédente. En effet, si on écrit la relation (1.15) en chaque point de la grille de vitesses, on obtient

$$f_k(t, x) = f_{k'}(t, x), \quad v_k \cdot n(x) > 0, \quad (1.49)$$

avec k' tel que $v_{k'} = \mathcal{R}v_k = v_k - 2(v_k \cdot n(x))n(x)$ soit la vitesse symétrique de v_k par rapport à la paroi en x . Il est clair que cette relation n'est justifiée que si la grille \mathcal{V} est confondue avec son image $\mathcal{R}\mathcal{V}$ dans la symétrie par rapport à la tangente n^\perp à la paroi (voir fig. 1.2).

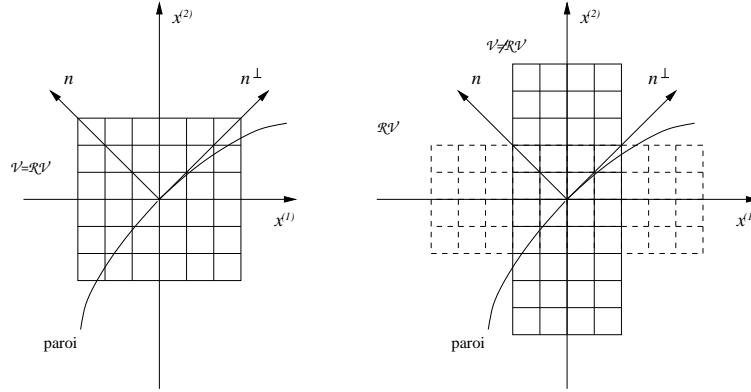


FIG. 1.2 – Grille de vitesses \mathcal{V} et symétrique $\mathcal{R}\mathcal{V}$ par rapport à la paroi (à gauche : $\mathcal{V} = \mathcal{R}\mathcal{V}$, à droite : $\mathcal{V} \neq \mathcal{R}\mathcal{V}$).

Cette propriété n'est en générale pas vérifiée et il convient donc de modifier la relation (1.49). Nous nous bornons ici à suggérer une méthode, bien que nous n'ayons pas effectivement utilisé cette condition aux limites dans nos calculs. Dans un premier temps, on peut considérer que si $v_{k'}$ n'appartient pas à \mathcal{V} , ceci signifie que $f_{k'}(t, x)$ est négligeable, et donc on pose $f_{k'} = 0$. Cependant, la condition aux limites obtenue n'est pas satisfaisante car elle crée artificiellement un flux de particules à travers la paroi. On peut alors introduire un coefficient correctif ϕ de la façon suivante

$$\text{pour } v_k \cdot n(x) > 0 : f_k(t, x) = \begin{cases} \phi f_{k'}(t, x) & \text{si } v_k \in \mathcal{V} \cap \mathcal{R}\mathcal{V}, \\ 0 & \text{sinon,} \end{cases} \quad (1.50)$$

déterminé pour que le flux normal à la paroi soit nul, c.-à-d.

$$\phi = \left(\sum_{v_k \cdot n(x) < 0} v_k \cdot n(x) f_k(t, x) \Delta v^D \right) \left(\sum_{\substack{v_k \cdot n(x) < 0 \\ v_k \in \mathcal{V} \cap \mathcal{R}\mathcal{V}}} v_k \cdot n(x) f_k(t, x) \Delta v^D \right)^{-1} \quad (1.51)$$

On peut néanmoins remarquer que cette correction ne garantit pas que les flux de quantités de mouvement, d'énergie et d'entropie soient nuls, contrairement à la condition continue. L'étude de cette condition de réflexion mériterait donc une étude plus fine que nous n'avons pas menée dans ce travail.

Chapitre 2

Etude théorique du modèle BGK discret

Dans les dix dernières années, quelques résultats mathématiques importants ont été obtenus sur l'équation BGK. Par exemple, Perthame a prouvé dans [49] un résultat global d'existence et de stabilité pour une solution distribution dans tout l'espace. Ce résultat a été étendu à des domaines bornés avec de nombreuses conditions aux limites par Ringeisen [54]. Plus récemment, Perthame et Pulvirenti ont montré dans [50] l'existence et l'unicité d'une solution intégrale ("mild solution") dans le cas de conditions aux limites périodiques, à l'aide d'estimations sur des normes L^∞ avec poids. Ce résultat a été généralisé à \mathbb{R}^D par Mischler dans [45]. Nous mentionnons aussi le résultat d'Issautier [29] qui a prouvé que la solution intégrale de Perthame et Pulvirenti est en fait une solution forte si l'on suppose une certaine régularité de la condition initiale. Cependant, il est important de remarquer que dans tous ces résultats, les différents auteurs supposent un temps de relaxation constant (ils posent $\tau = 1$). Ce n'est physiquement pas très réaliste car τ est plutôt une fonction de ρ et de T (cf. chap. 4). A notre connaissance, il n'existe pas pour l'instant de résultat d'existence global pour des temps de relaxation réalistes. Dans la suite de ce chapitre, nous ferons donc nous aussi l'hypothèse $\tau = 1$.

Il est naturel d'étudier la convergence de notre modèle discret vers l'équation BGK continue quand la grille de vitesses "tend" vers \mathbb{R}^D . Pour obtenir un tel résultat, la méthode est assez générale (voir par exemple la preuve de Mischler pour la convergence d'un modèle à vitesses discrètes de l'équation de Boltzmann [46]). Il y a essentiellement trois points distincts à prouver :

- la convergence d'une approximation du terme de collision, qui est local en t, x ,
- l'existence et l'unicité d'une solution du modèle discret,
- la convergence de cette solution vers une solution de l'équation continue.

Le premier point dépend fortement du problème. Schneider et al. [6] l'ont par exemple montré pour une quadrature de l'opérateur de collision de Boltzmann (voir aussi Heintz et Panferov [27]). Notre modèle est bien sûr complètement différent. La difficulté essentielle est que l'équilibre discret est implicitement défini par $f_{\mathcal{K}}$, et nous introduisons alors des problèmes de minimisations dérivés pour obtenir une propriété de coercivité uniforme.

Pour le second point, comme le note Mischler [46], un tel résultat n'est pas connu pour les modèles discrets de l'équation de Boltzmann. Cependant, grâce à la structure particulière du terme de collision de BGK et au fait que le domaine des vitesses est borné et discret, nous sommes capables de prouver l'existence et l'unicité d'une solution globale pour notre modèle. L'hypothèse sur la condition initiale continue est celle de Perthame [49].

1. Ce chapitre a fait l'objet d'une note aux *Comptes Rendus de l'Académie des Sciences*, parue en 1999 sous le titre *Convergence d'un modèle à vitesses discrètes pour l'équation de Boltzmann-BGK* (cf. [44])

Pour le dernier point, Mischler a proposé l'approche suivante. Il définit une formulation continue du modèle discret, et il prouve la convergence en utilisant le résultat de stabilité de DiPerna-Lions [21]. La principale difficulté dans cette méthode est la validité du lemme de moyenne en vitesse dans le contexte des vitesses discrètes, mais il a été prouvé par Mischler. Nous suivons donc une approche similaire en définissant une forme continue du modèle BGK discret, et nous utilisons la preuve de stabilité de Perthame [49] pour l'équation BGK.

Il est enfin important de noter le résultat récent d'Issautier [29] qui a prouvé un résultat de convergence pour une méthode particulière de l'équation BGK dans laquelle le temps, l'espace et la vitesse sont discrets. L'avantage de cette approche est l'approximation explicite de la maxwellienne (il prend $\mathcal{E}_k = M[\boldsymbol{\rho}_{\mathcal{K}}](v_k)$) qui permet de dériver facilement des estimations d'erreurs. Mais contrairement à notre modèle, la discrétisation d'Issautier n'est ni conservative ni entropique. Pour notre méthode, le prix à payer pour avoir ces propriétés est la définition implicite de $\mathcal{E}_{\mathcal{K}}$, ce qui rend difficile l'obtention d'estimations d'erreurs.

Le reste de ce chapitre s'articule de la façon suivante. Dans la section 2.1, nous donnons quelques notations qui diffèrent quelque peu du chapitre précédent, nous dérivons rigoureusement le modèle discret de l'équation BGK continue, et nous donnons les principaux résultats prouvés dans les sections suivantes. La section 2.2 est consacrée à la consistance de l'approximation de la maxwellienne. Dans la section 2.3, nous prouvons le résultat d'existence et d'unicité pour le modèle discret. Dans la section 2.4, nous prouvons le résultat final de convergence du modèle.

2.1 Notations et principaux résultats

Soient Δv_h et B_h deux suites de nombres réels tels que

$$\Delta v_h \xrightarrow{h \rightarrow 0} 0, \quad B_h \xrightarrow{h \rightarrow 0} +\infty. \quad (2.1)$$

Soit \mathcal{V}^h une grille de N_h vitesses définie par

$$\mathcal{V}^h = \{v_k^h = k\Delta v_h, \quad k \in \mathcal{K}^h\}, \quad (2.2)$$

où \mathcal{K}^h est l'ensemble de multi-indices $\mathcal{K}^h = \{k \in \mathbb{Z}^D, \quad |k| \leq B_h\}$. Nous définissons aussi les mailles de vitesse Λ_k^h par

$$\Lambda_k^h = [v_{k,1}^h - \frac{1}{2}\Delta v_h, v_{k,1}^h + \frac{1}{2}\Delta v_h] \times \dots \times [v_{k,D}^h - \frac{1}{2}\Delta v_h, v_{k,D}^h + \frac{1}{2}\Delta v_h]. \quad (2.3)$$

Une fonction de distribution discrète $g = (g_k)_{k \in \mathcal{K}^h}$ sur \mathcal{V}^h est un vecteur de \mathbb{R}^{N_h} dont les $D+2$ premiers moments sont définis par

$$\langle g \rangle_h = \sum_{k \in \mathcal{K}^h} g_k \Delta v_h^D, \quad \langle \mathbf{m}g \rangle_h = \sum_{k \in \mathcal{K}^h} \mathbf{m}(v_k^h) g_k \Delta v_h^D, \quad (2.4)$$

Dans tout ce chapitre, les sommes discrètes seront toujours notées $\langle \cdot \rangle_h$, et les intégrales des fonctions de v seront notées $\langle \cdot \rangle$.

La maxwellienne $M[\boldsymbol{\rho}_h]$ associée à un vecteur de moments donné $\boldsymbol{\rho}_h$ est approchée sur \mathcal{V}^h par $\mathcal{E}_{\mathcal{K}}^h = (\mathcal{E}_k^h)_{k \in \mathcal{K}^h}$. Cette approximation est définie par la version discrète du problème de minimisation d'entropie discrète

$$(\mathcal{P}_h) \quad H^h(\mathcal{E}_{\mathcal{K}}^h) = \min_{\mathcal{X}_{\boldsymbol{\rho}_h}} \left\{ H^h(g) = \langle g \log g \rangle_h \right\}, \quad (2.5)$$

avec $\mathcal{X}_{\boldsymbol{\rho}_h} = \{g \geq 0 \in \mathbb{R}^{N_h}, \langle \mathbf{m}g \rangle_h = \boldsymbol{\rho}_h\}$.

D'après le théorème 1.1, le problème (\mathcal{P}_h) a une solution unique si $\mathcal{X}_{\rho_h} \neq \emptyset$. De plus, si $B_h \geq 1$, alors il existe un unique vecteur $\alpha_h \in \mathbb{R}^{D+2}$ tel que

$$\mathcal{E}_k^h = \exp(\alpha_h \cdot \mathbf{m}(v_k^h)) \quad \forall k \in \mathcal{K}^h, \quad (2.6)$$

si, et seulement si, ρ_h est *strictement réalisable* dans \mathcal{V}^h , i.e.

$$\exists g > 0 \in \mathcal{X}_{\rho_h}. \quad (2.7)$$

Un résultat essentiel de ce chapitre, crucial pour la convergence du modèle discret, montre que l'approximation de $M[\rho]$ par $\mathcal{E}_{\mathcal{K}}^h$ est consistante si $\rho_h \xrightarrow{h \rightarrow 0} \rho$.

Théorème 2.1. *Soit $\{\rho_h\}_{h \leq 1}$ une suite de vecteurs de \mathbb{R}^{D+2} strictement réalisables sur \mathcal{V}^h pour tout h . Soit $\rho \in \mathbb{R}^{D+2}$ tel que $\rho, T > 0$. Alors, si $\rho_h \xrightarrow{h \rightarrow 0} \rho$, le vecteur α_h donné par (2.6), converge vers α défini par (1.12).*

Nous pouvons à présent définir notre approximation discrète en vitesse de l'équation BGK

$$\begin{aligned} \partial_t f + v \cdot \nabla_x f &= M[\rho] - f \\ f(0, x, v) &= f^0(x, v). \end{aligned} \quad (2.8)$$

Supposons que la condition initiale f^0 satisfait les estimations classiques

$$\int_{\mathbb{R}^D} \langle (1 + |x|^2 + |v|^2 + |\log f^0|) f^0 \rangle dx = \Gamma_0 < +\infty. \quad (2.9)$$

D'après le résultat de Perthame [49], nous savons que ces estimations assurent l'existence d'une solution distribution de (2.8). Nous définissons alors l'approximation suivante de f^0

$$f_k^{0,h}(x) = \min \left(\frac{1}{h}, \frac{1}{\Delta v_h^D} \int_{\Lambda_k^h} (f^0(x, v) + h \exp(-|x|^2 - |v|^2)) dv \right). \quad (2.10)$$

On peut vérifier que $f_{\mathcal{K}}^{0,h} = (f_k^{0,h})_{k \in \mathcal{K}}$ satisfait les estimations

$$\delta_0^h \phi(x) \leq f_k^{0,h}(x) \leq \frac{1}{h} \quad \text{p.p dans } \mathbb{R}_x^D, \quad \forall k \in \mathcal{K}^h, \quad (2.11)$$

$$\sup_{h \leq 1} \int_{\mathbb{R}^D} \langle (1 + |x|^2 + |v|^2 + |\log f_{\mathcal{K}}^{0,h}|) f_{\mathcal{K}}^{0,h} \rangle_h dx = \Gamma_1 < +\infty, \quad (2.12)$$

où $\delta_0^h = h \frac{1}{\Delta v_h^D} \min_k (\int_{\Lambda_k^h} \exp(-|v|^2) dv)$ et $\phi(x) = \exp(-|x|^2)$. L'approximation discrète en vitesse de (2.8) est alors

$$\begin{aligned} \partial_t f_k^h + v_k^h \cdot \nabla_x f_k^h &= \mathcal{E}_k^h - f_k^h \quad \text{dans } \mathcal{D}'([0, +\infty[\times \mathbb{R}_x^D) \quad \forall k \in \mathcal{K}^h, \\ f_k^h(0, x) &= f_k^{0,h}(x), \end{aligned} \quad (2.13)$$

où \mathcal{E}_k^h est naturellement définie par (\mathcal{P}_h) , avec

$$\rho_h = (\rho_h, \rho_h u_h, E_h)^T = \langle \mathbf{m} f_{\mathcal{K}}^h \rangle_h. \quad (2.14)$$

Le second résultat de ce chapitre montre que le modèle (2.13) admet une solution unique. De plus, grâce à la définition de l'équilibre discret, cette solution satisfait les lois de conservation et de dissipation d'entropie.

Théorème 2.2. *Le problème de Cauchy (2.13) admet une unique solution $f_{\mathcal{K}}^h = (f_k^h)_{k \in \mathcal{K}}$ dans $L^\infty([0, t_{max}[\times \mathbb{R}_x^D)^{N_h}$, pour tout $t_{max} > 0$. De plus, les relations de conservation et d'entropie suivantes sont vérifiées, au sens des distributions*

$$\partial_t \langle \mathbf{m} f_{\mathcal{K}}^h \rangle_h + \nabla_x \cdot \langle v \mathbf{m} f_{\mathcal{K}}^h \rangle_h = 0, \quad (2.15)$$

$$\partial_t \langle f_{\mathcal{K}}^h \log f_{\mathcal{K}}^h \rangle_h + \nabla_x \cdot \langle v f_{\mathcal{K}}^h \log f_{\mathcal{K}}^h \rangle_h \leq 0, \quad (2.16)$$

et $f_{\mathcal{K}}^h$ satisfait les estimations

$$\sup_h \sup_{[0, t_{max}[} \int_{\mathbb{R}^D} \langle (1 + |x|^2 + |v|^2 + |\log f_{\mathcal{K}}^h|) f_{\mathcal{K}}^h(t, x) \rangle_h dx \leq \Gamma_2(t_{max}), \quad (2.17)$$

$$\delta_0^h \phi(x - tv_k^h) e^{-t} \leq f_k^h(t, x) \leq \frac{1}{h} e^{N_h t} \quad p.p \ t, x. \quad (2.18)$$

Enfin, pour prouver la convergence de cette solution, nous définissons les fonctions constantes par maille de vitesse

$$f^h(t, x, v) = \sum_{k \in \mathcal{K}^h} f_k^h(t, x) \chi_k^h(v) \quad \text{et} \quad \mathcal{E}^h(t, x, v) = \sum_{k \in \mathcal{K}^h} \mathcal{E}_k^h(t, x) \chi_k^h(v), \quad (2.19)$$

où χ_k^h est la fonction indicatrice de Λ_k^h . On peut faire le lien entre l'équation continue (2.8) et le modèle discret (2.13) par l'équation

$$\begin{aligned} \partial_t f^h + C^h(v) \cdot \nabla_x f^h &= \mathcal{E}^h - f^h && \text{dans } \mathcal{D}'([0, +\infty[\times \mathbb{R}_x^D \times \mathbb{R}_v^D), \\ f^h(0, x, v) &= f^{0,h}(x, v) = \sum_{k \in \mathcal{K}^h} f_k^{0,h}(x) \chi_k^h(v), \end{aligned} \quad (2.20)$$

où

$$C^h(v) = \sum_{k \in \mathcal{K}^h} v_k^h \chi_k^h(v). \quad (2.21)$$

Nous donnons à présent le résultat de convergence :

Théorème 2.3. *Pour toutes suites $\Delta v_h, B_h$ satisfaisant (2.1), la suite $\{f^h\}_{h \leq 1}$ est faiblement convergente dans $L^1([0, t_{max}[\times \mathbb{R}_x^D \times \mathbb{R}_v^D) \forall t_{max} > 0$, à l'extraction d'une sous-suite près, vers une solution distribution de l'équation BGK (2.8).*

2.2 Consistance de l'approximation de la maxwellienne

De façon analogue à la preuve du théorème 1.1, on peut voir (cf. remarque 2.1) que α_h et α sont les solutions uniques des problèmes de minimisation suivants

$$J_h(\alpha_h) = \min_{\mathbb{R}^{D+2}} \left\{ J_h(\beta) = \langle \exp(\beta \cdot \mathbf{m}) \rangle_h - \beta \cdot \rho_h \right\}, \quad (2.22)$$

$$J(\alpha) = \min_{\mathcal{D}} \left\{ J(\beta) = \langle \exp(\beta \cdot \mathbf{m}) \rangle - \beta \cdot \rho \right\}, \quad (2.23)$$

avec $\mathcal{D} = \mathbb{R}^{D+2} \cap \{\beta, \beta^{(D+1)} < 0\}$. Rappelons que $(\beta^{(0)}, \dots, \beta^{(D+1)})$ sont les composantes de β . Tout d'abord, nous avons besoin de la proposition suivante

Proposition 2.1. *(P₁) J_h est strictement convexe et coercive dans \mathbb{R}^{D+2}*

(P₂) J est strictement convexe et coercive dans \mathcal{D}

(P₃) J_h est localement uniformément convergente vers J dans \mathcal{D}

(P₄) on a $\sup_h J_h(\alpha_h) < +\infty$

Remarque 2.1. Les propriétés (P_1) et (P_2) montrent bien que α_h et α sont les uniques solutions des problèmes (2.22) et (2.23)

Preuve de la proposition. La propriété (P_1) est prouvée dans la section 1.2.2 (preuve du théorème 1.1).

Pour (P_2) , il est clair que J est de classe C^2 , et que $J'' = \langle \mathbf{m} \otimes \mathbf{m} \exp(\boldsymbol{\beta} \cdot \mathbf{m}) \rangle$ est définie positive. Donc J'' est strictement convexe dans \mathcal{D} . La propriété de coercivité signifie que $J(\boldsymbol{\beta}_p) \xrightarrow{p \rightarrow +\infty} +\infty$ pour toute suite $\{\boldsymbol{\beta}_p\}_{p \geq 0} \subset \mathcal{D}$ s'approchant de la frontière de \mathcal{D} , c.-à-d. telle que

$$(i) \quad \beta_p^{(D+1)} \xrightarrow{p \rightarrow +\infty} 0^- \quad \text{ou} \quad (ii) \quad |\boldsymbol{\beta}_p| \xrightarrow{p \rightarrow +\infty} +\infty.$$

Pour le cas (i) , nous considérons une expression analytique de J

$$\begin{aligned} J(\boldsymbol{\beta}_p) &= \exp\left(\frac{1}{2} \sum_{i=1}^D |\beta_p^{(i)}|^2 / |\beta_p^{(D+1)}|\right) \exp(\beta_p^{(0)}) \left(\frac{2\pi}{|\beta_p^{(D+1)}|}\right)^{D/2} \\ &\quad - \beta_p^{(0)} \rho - \sum_{i=1}^D \beta_p^{(i)} \rho u^{(i)} - \beta_p^{(D+1)} E, \end{aligned} \quad (2.24)$$

où $u = (u^{(1)}, \dots, u^{(D)})$ est défini par (1.3). Nous étudions alors toutes les limites de $\beta_p^{(0)}, \beta_p^{(i=1..D)}$ et il apparaît que $J(\boldsymbol{\beta}_p)$ tend toujours vers $+\infty$. Nous résumons ici l'étude laborieuse de tous les cas possibles. Si $\beta_p^{(0)}$ est bornée inférieurement, alors on voit que les exponentielles de la puissance inverse de $\beta_p^{(D+1)}$ tendent plus vite vers $+\infty$ que la partie linéaire de J . Si $\beta_p^{(0)} \rightarrow -\infty$, alors on doit considérer deux cas différents. Si $-\beta_p^{(0)} \rho - \sum_{i=1}^D \beta_p^{(i)} \rho u^{(i)} \rightarrow +\infty$, alors $J(\boldsymbol{\beta}_p) \rightarrow +\infty$, quelle que soit la limite de $\beta_p^{(i=1..D)}$. Dans l'autre cas, $-\beta_p^{(0)} \rho - \sum_{i=1}^D \beta_p^{(i)} \rho u^{(i)}$ est borné supérieurement, ce qui implique que $\beta_p^{(i=1..D)}$ n'est pas borné; donc la première exponentielle tend assez vite vers $+\infty$ pour qu'on ait $J(\boldsymbol{\beta}_p) \rightarrow +\infty$.

Pour le cas (ii) , on peut supposer que $\beta_p^{(D+1)} \leq c < 0 \forall p$. Soit $R > 0$ tel que la boule $B(\boldsymbol{\alpha}, R) \subset \mathcal{D}$. Alors si p est assez grand (i.e. $p > p_0$), on a $\boldsymbol{\beta}_p \notin \bar{B}(\boldsymbol{\alpha}, R)$. Définissons $\boldsymbol{\gamma}_p = \frac{R}{|\boldsymbol{\beta}_p - \boldsymbol{\alpha}|} (\boldsymbol{\beta}_p - \boldsymbol{\alpha}) + \boldsymbol{\alpha}$, qui est sur la frontière S de $B(\boldsymbol{\alpha}, R)$, alors $\boldsymbol{\gamma}_p = \theta_p \boldsymbol{\alpha} + (1 - \theta_p) \boldsymbol{\beta}_p$ avec $0 < \theta_p = 1 - \frac{R}{|\boldsymbol{\beta}_p - \boldsymbol{\alpha}|} < 1$ si $p > p_0$. En raison de la stricte convexité de J , on a $J(\boldsymbol{\beta}_p) > \frac{1}{1-\theta_p} (J(\boldsymbol{\gamma}_p) - \theta_p J(\boldsymbol{\alpha}))$ pour $p > p_0$. De plus, $\boldsymbol{\alpha}$ est l'unique minimum de J qui est continu, et S est compacte, donc on a $J(\boldsymbol{\gamma}_p) - J(\boldsymbol{\alpha}) \geq m > 0$ pour tout $p > p_0$. Par conséquent $J(\boldsymbol{\beta}_p) \geq \frac{m}{1-\theta_p} + J(\boldsymbol{\alpha})$ qui tend vers $+\infty$ car $\theta_p \rightarrow 1^-$ quand $p \rightarrow \infty$.

Pour la propriété (P_3) , nous devons prouver que $e_h = |\langle \exp(\boldsymbol{\beta} \cdot \mathbf{m}) \rangle_h - \langle \exp(\boldsymbol{\beta} \cdot \mathbf{m}) \rangle| \rightarrow 0$ localement uniformément sur \mathcal{D} , ce qui est un problème de quadrature dans \mathbb{R}^D . Nous séparons e_h en deux termes

$$\begin{aligned} e_h &\leq \left| \int_{\mathbb{R}^D} \exp(\boldsymbol{\beta} \cdot \mathbf{m}(v)) dv - \Delta v_h^D \sum_{k \in \mathbb{Z}^D} \exp(\boldsymbol{\beta} \cdot \mathbf{m}(v_k^h)) \right| + \sum_{v_k^h \notin \mathcal{V}^h} \exp(\boldsymbol{\beta} \cdot \mathbf{m}(v_k^h)) \Delta v_h^D \\ &= E_1 + E_2. \end{aligned}$$

Nous utilisons le lemme suivant, donné dans [53]

Lemme 2.1. Il existe $c > 0$ indépendant de Δv tel que pour tout $g \in W^{m,1}(\mathbb{R}^D)$, $m > D$ on a

$$\left| \int_{\mathbb{R}^D} g(v) dv - \Delta v^D \sum_{k \in \mathbb{Z}^D} g(v_k) \right| \leq c \Delta v^m |g|_{m,1}, \quad (2.25)$$

où $|g|_{m,1} = \sum_{|l|=m} \|\partial^l g\|_{L^1}$.

Posons alors $g_\beta(v) = \exp(\beta \cdot \mathbf{m}(v))$, qui est dans $W^{m,1} \forall m \geq 0$. Nous déduisons de ce lemme que

$$E_1 \leq c \Delta v_h^m |g_\beta|_{m,1}.$$

On peut remarquer que puisque $|g_\beta|_{m,1} \leq \int |p(\beta, v)| \exp(\beta \cdot \mathbf{m}(v)) dv$ où p est un polynôme, alors $|g_\beta|_{m,1}$ est bornée sur tout sous-ensemble compact K de \mathcal{D} . Donc il existe une constante c_K dépendant uniquement de K , telle que $E_1 \leq c_K \Delta v_h^m$. Pour le terme E_2 , noter que $v_k^h \notin \mathcal{V}^h$ signifie $|v_k^h| \geq D_h$ (où D_h est le rayon \mathcal{V}^h). Donc

$$E_2 \leq \frac{1}{D_h^2} \sum_{k \in \mathbb{Z}^D} |v_k^h|^2 \exp(\beta \cdot \mathbf{m}(v_k^h)) \Delta v_h^D.$$

On applique le lemme 2.1 à $g_\beta(v) = |v|^2 \exp(\beta \cdot \mathbf{m}(v))$, ce qui donne $E_2 \leq \frac{1}{D_h^2} c_K$ sur tout compact K de \mathcal{D} . Par conséquent les bornes de E_1 et E_2 montrent que $e_h \rightarrow 0$, uniformément sur tout compact K de \mathcal{D} , ce qui termine la preuve de (P_3) .

Pour la dernière propriété (P_4) , remarquons que, par définition, $J_h(\alpha_h) \leq J_h(\beta)$ pour tout $\beta \in \mathcal{D}$. Grâce à (P_3) , on a $J_h(\beta) \rightarrow J(\beta)$, donc $J_h(\beta)$ est borné, et il existe $c > 0$ tel que $J_h(\alpha_h) \leq J_h(\beta) \leq c$ pour tout h . Ceci prouve (P_4) , et achève la preuve de la proposition 2.1. \square

L'idée de la démonstration du théorème 2.1 est alors de montrer que J_h est en fait coercive *uniformément en h* . Ainsi, (P_4) assure que α_h est borné, et (P_3) implique que $\alpha_h \rightarrow \alpha$.

Tout d'abord, soit un réel S tel que $S \geq J(\alpha)$. La coercivité de J implique qu'il existe un compact $K \subset \mathcal{D}$ (e.g. une boule fermée), telle que

$$J(\beta) \geq S + 1 \quad \forall \beta \in \mathcal{D} - \overset{\circ}{K}, \quad (2.26)$$

et on peut supposer que $\alpha \in \overset{\circ}{K}$ (qui dénote l'intérieur de K). La propriété (P_3) implique qu'il existe $h_0(K)$ dépendant uniquement du compact K tel que

$$|J_h(\beta) - J(\beta)| \leq \frac{1}{4} \quad \forall \beta \in K, \forall h \leq h_0(K). \quad (2.27)$$

Par conséquent, les relations (2.26) et (2.27) donnent

$$J_h(\beta) \geq S + \frac{3}{4} \quad \forall \beta \in \partial K, \forall h \leq h_0(K). \quad (2.28)$$

De plus, le fait que $\alpha \in \overset{\circ}{K}$ et la définition de S impliquent

$$J_h(\alpha) \leq |J_h(\alpha) - J(\alpha)| + J(\alpha) \leq S + \frac{1}{4} < S + \frac{3}{4} \quad \forall h \leq h_0(K). \quad (2.29)$$

Nous utilisons à présent le lemme suivant :

Lemme 2.2. *Soit ψ une fonction convexe d'un ouvert convexe Ω de \mathbb{R}^D . S'il existe un sous ensemble convexe compact $K \subset \Omega$ et une constante c tels que*

$$\psi(x) \geq c \text{ dans } \partial K \text{ et } \exists x_0 \in \overset{\circ}{K}, \psi(x_0) < c, \quad (2.30)$$

alors $\psi(x) > c$ dans $\Omega - K$.

Preuve du lemme 2.2. Soit $x \notin K$ et soit y le point d'intersection de ∂K avec $[x_0, x]$. Alors il existe $0 < \sigma < 1$ tel que $y = \sigma x_0 + (1 - \sigma)x$. La convexité de ψ implique

$$\psi(x) \geq \frac{1}{1 - \sigma}(\psi(y) - \sigma\psi(x_0)),$$

et donc l'hypothèse (2.30) donne $\psi(x) > c$, ce qui prouve le lemme. \square

Les relations (2.29) and (2.28) montrent que J_h satisfait les conditions du lemme 2.2 pour tout $h \leq h_0(K)$, et donc $J_h(\beta) \geq S + \frac{3}{4} \quad \forall h \leq h_0(K), \quad \forall \beta \in \mathbb{R}^{D+2} - K$. Par conséquent, pour tout $S \geq J(\alpha)$, il existe un compact K et un réel h_0 tels que

$$J_h(\beta) \geq S + \frac{3}{4} \quad \forall h \leq h_0(K), \quad \forall \beta \in \mathbb{R}^{D+2} - K.$$

Les fonctionnelles J_h sont donc coercives uniformément en h . Grâce à (P_4) , ceci prouve que $\alpha_h \in K$ pour tout $h \leq h_0(K)$. Puisque K est compact, alors α_h converge, à l'extraction d'une sous-suite près, vers $\bar{\alpha} \in \mathcal{D}$. La propriété (P_3) et l'unicité de α implique $\bar{\alpha} = \alpha$, et donc toute la suite α_h converge vers α , ce qui conclut la preuve du théorème.

2.3 Un résultat d'existence et d'unicité pour le modèle discret

Ce théorème étant indépendant de h , nous omettons l'indice h dans cette section quand il n'y a pas d'ambiguïté possible. Nous utilisons ici une méthode de point fixe pour l'opérateur Φ , défini par le problème non linéaire

$$F \stackrel{\text{def}}{=} \Phi(G) \iff \begin{cases} \partial_t F_k + v_k \cdot \nabla_x F_k + F_k = \mathcal{E}_k[\rho_G] \\ F_k(0, x) = f_k^0(x), \end{cases} \quad (2.31)$$

où $\rho_G = \langle \mathbf{m}G \rangle_h$, et $\mathcal{E}_k[\rho_G]$ est le minimum d'entropie sur \mathcal{X}_{ρ_G} (voir (2.5)). Cet opérateur est bien défini si G est strictement positif dans $L^\infty([0, t_{max}] \times \mathbb{R}_x^D)^N, \quad \forall t_{max} > 0$.

Dans un premier temps, nous donnons une zone invariante pour Φ . Pour travailler localement en espace et en temps, considérons des réels R et t_{max} positifs et le cône de dépendance de $\Phi(G)(t_{max}, x)$ sur la boule $\bar{B}(0, R)$

$$\Omega_R(t_{max}) = \left\{ (t, x); \quad t \leq t_{max} \text{ et } |x| \leq R + (t_{max} - t) \max_k |v_k| \right\}. \quad (2.32)$$

Notons que grâce au fait que les vitesses de propagation du modèle sont bornées par $\max_k |v_k|$, cet ensemble est compact. La zone invariante est donnée dans la proposition suivante :

Proposition 2.2. *L'ensemble*

$$\mathcal{F}_R = \left\{ G \in L^\infty(\Omega_R(t_{max}))^N, \quad \delta_0 e^{-t} \phi(x - tv_k) \leq G_k(t, x) \leq \frac{1}{h} e^{Nt}, \quad p.p (t, x) \right\} \quad (2.33)$$

est stable par Φ .

Preuve de la proposition 2.2. On vérifie tout d'abord que pour $G \in \mathcal{F}_R$, la fonction $F = \Phi(G)$ est bien définie sur $\Omega_R(t_{max})$ et ne dépend que des valeurs de G sur ce compact, ainsi que des valeurs de f^0 . En utilisant la représentation intégrale

$$F_k(t, x) = e^{-t} f_k^0(x - tv_k) + \int_0^t e^{s-t} \mathcal{E}_k[\rho_G(s, x - (t-s)v_k)] ds,$$

on obtient le résultat souhaité grâce au lemme suivant :

Lemme 2.3. *Soit $(t, x) \in \Omega_R(t_{max})$. Pour tout $0 \leq s \leq t$ et pour tout $k \in \mathcal{K}$, on a*

$$(s, x - (t - s)v_k) \in \Omega_R(t_{max}).$$

Preuve du lemme 2.3. Ce lemme repose sur le fait que les vitesses v_k étant bornées, on ne peut sortir de $\Omega_R(t_{max})$ en suivant les caractéristiques. D'après la définition (2.32), on a

$$t \leq t_{max} \quad \text{et} \quad |x| \leq R + (t_{max} - t) \max_k |v_k|,$$

et on veut montrer que

$$s \leq t_{max} \quad \text{et} \quad |x - (t - s)v_k| \leq R + (t_{max} - s) \max_k |v_k|.$$

Par définition de s , il est évident que $s \leq t_{max}$. De plus, par inégalité triangulaire, on a

$$\begin{aligned} |x - (t - s)v_k| &\leq |x| + (t - s)|v_k| \\ &\leq R + (t_{max} - t) \max_k |v_k| + (t - s) \max_k |v_k| \\ &= R + (t_{max} - s) \max_k |v_k|, \end{aligned}$$

ce qui conclut la démonstration du lemme. \square

Pour terminer la preuve de la proposition 2.2, on montre à présent que $F = \Phi(G)$ vérifie les mêmes estimations que G . Grâce au fait que \mathcal{V} est borné et discret, et grâce à la définition de $\mathcal{E}_{\mathcal{K}}[\rho_g]$, on a l'estimation suivante de l'équilibre discret

$$\mathcal{E}_{\mathcal{K}}[\rho_G(s, y)] \leq \sum_{k'} \mathcal{E}_{k'}[\rho_G(s, y)] = \sum_{k'} G_{k'}(s, y) \leq N \frac{1}{h} e^{Ns}. \quad (2.34)$$

Alors une représentation intégrale de $F = \Phi(G)$ donne la borne supérieure de (2.33) :

$$\begin{aligned} F_k(t, x) &= f_k^0(x - tv_k) + \int_0^t (\mathcal{E}_k[\rho_G(s, x - (t - s)v_k)] - F_k(s, x - (t - s)v_k)) ds \\ &\leq \frac{1}{h} + \int_0^t N \frac{1}{h} e^{Ns} ds = \frac{1}{h} e^{Nt}, \end{aligned}$$

et la borne inférieure est aisément obtenue en intégrant (2.31) le long des caractéristiques

$$\begin{aligned} F_k(t, x) &= e^{-t} f_k^0(x - tv_k) + \int_0^t e^{s-t} \mathcal{E}_k[\rho_G(s, x - (t - s)v_k)] ds \\ &\geq e^{-t} f_k^0(x - tv_k) \geq e^{-t} \delta_0 \phi(x - tv_k). \end{aligned}$$

\square

L'idée consiste à présent à appliquer un théorème de point fixe sur la zone invariante donnée par la proposition 2.2. Nous montrons tout d'abord que $G \mapsto \mathcal{E}_{\mathcal{K}}[\rho_G]$ est lipschitzien sur l'ensemble \mathcal{F}_R . Pour cela, remarquons que grâce au théorème 1.1, on a $\mathcal{E}_{\mathcal{K}}[\rho_G] = \exp(\alpha_G \cdot \mathbf{m}(v_k))$ pour tout $G \in \mathcal{F}_R$. On peut donc vérifier que l'application $\rho_G \mapsto \alpha_G \in \mathbb{R}^{D+2}$, définie dans l'ensemble $\{\rho_G \in \mathbb{R}^D \text{ strictement réalisable dans } \mathcal{V}\}$, est continûment différentiable : en fait, la matrice jacobienne de l'application réciproque est $\langle \mathbf{m} \otimes \mathbf{m} \exp(\alpha_G \cdot \mathbf{m}) \rangle_{\mathcal{V}}$, qui est définie positive en raison de la définition

de \mathcal{V} (voir la proposition 1.1), donc inversible. Or notons que les éléments de \mathcal{F}_R sont uniformément bornés et uniformément loin de 0, car

$$\delta_0 e^{-t} \phi(x - tv_k) \geq \delta_0 e^{-t_{max}} \min_{\Omega_R(t_{max})} \phi(x - tv_k) = C(t_{max}, R) > 0.$$

Donc l'opérateur $G \mapsto \mathcal{E}_K[\rho_G]$ est lipschitzien sur \mathcal{F}_R , *i.e.* il existe une constante $L(R, t_{max})$ positive qui ne dépend que de R et de t_{max} telle que

$$\left| \mathcal{E}_k[\rho_F] - \mathcal{E}_k[\rho_G] \right| (t, x) \leq L(R, t_{max}) \max_k \left| F_k - G_k \right| (t, x)$$

pour tout (t, x) de $\Omega_R(t_{max})$ et pour tout F et G de \mathcal{F}_R . Une technique classique en théorie des équations différentielles ordinaires permet alors de montrer qu'une certaine itérée de Φ est contractante sur \mathcal{F}_R . D'après le théorème du point fixe, il existe alors une unique fonction f_R dans $L^\infty(\Omega_R(t_{max}))^N$ telle que $\Phi(f_R) = f_R$ presque partout sur $\Omega_R(t_{max})$. En prenant une suite croissante de R , on peut donc construire une fonction f dans $L^\infty([0, t_{max}] \times \mathbb{R}_x^D)$ telle que $\Phi(f) = f$ presque partout. Par unicité de sa restriction à $\Omega_R(t_{max})$ pour tout $R > 0$, cette fonction est donc unique, ce qui conclut la partie existence et unicité du théorème.

Remarque 2.2. Nous avons choisi de travailler localement en espace sur $\Omega_R(t_{max})$ car la borne inférieure de (2.33) tend vers 0 quand $|x| \rightarrow +\infty$, ce qui ne permet pas d'obtenir une propriété de Lipschitz globale en espace de l'opérateur $G \mapsto \mathcal{E}_K[\rho_G]$. En effet, on s'approche alors de la frontière de l'ensemble des moments strictements réalisables, et il semble donc difficile de borner uniformément en x la dérivée de cet opérateur.

Pour obtenir les lois de conservation (2.15), il suffit de multiplier (2.13) par $\mathbf{m}(v_k)$, puis de sommer sur \mathcal{K} . Pour la loi de dissipation locale d'entropie (2.16), remarquons que si η est une fonction lipschitzienne, alors (voir [21])

$$\partial_t \eta(f_k) + v_k \cdot \nabla_x \eta(f_k) = (\mathcal{E}_k - f_k) \eta'(f_k) \quad \text{dans } \mathcal{D}'([0, t_{max}] \times \mathbb{R}_x^D)^N. \quad (2.35)$$

Mais $\eta(s) = s \log s$ n'est pas lipschitzienne près de 0, donc suivant Perthame [49], nous bornons sa dérivée en posant

$$\eta_R(0) = 0, \quad \eta'_R(s) = \max(-R, \min(R, 1 + \log s))$$

sur $[0, +\infty[$. Cette fonction est lipschitzienne et $\eta_R(s) \xrightarrow{R \rightarrow +\infty} \eta(s)$. D'après (2.35), on a

$$\eta_R(f_k^\#) - \eta_R(f_k^0) = \int_0^t \eta'_R(f_k^\#) (\mathcal{E}_k^\# - f_k^\#) ds \quad (2.36)$$

où $f_k^\#(t, x) = f_k(t, x + tv_k)$. Le membre de gauche de cette équation converge p.p vers $\eta(f_k^\#) - \eta(f_k^0)$ quand $R \rightarrow +\infty$. Il en va de même pour le membre de droite, puisque $|\eta'_R(f_k^\#)|$ est bornée supérieurement par $1 + |\log f_k^\#|$, qui est borné d'après la borne supérieure de (2.33). On peut alors passer à la limite dans (2.36), ce qui donne une formulation équivalente à (2.35). Finalement, on somme (2.35) sur $k \in \mathcal{K}$ et il est désormais classique de noter que

$$\langle (\mathcal{E}_K - f_K) \eta'(f_K) \rangle_h = \langle (\mathcal{E}_K - f_K) (\log f_K - \log \mathcal{E}_K) \rangle_h + \langle (1 + \log \mathcal{E}_K) (\mathcal{E}_K - f_K) \rangle_h.$$

D'après la définition de \mathcal{E}_K , le dernier terme est nul. Le second est négatif car $s \mapsto \log s$ est croissante. On obtient donc l'inégalité d'entropie (2.16).

L'estimation (2.17) est aisément déduite de (2.15) et (2.16), ce qui termine la preuve du théorème.

2.4 Un résultat de convergence pour le modèle BGK discret

Suivant Perthame [49], nous divisons la preuve en quatre étapes.

étape 1: convergence faible de f^h et $\langle \mathbf{m} f^h \rangle$

D'après (2.17), il est clair que f^h vérifie l'estimation

$$\sup_h \sup_{[0, t_{max}]} \int_{\mathbb{R}^D} \langle (1 + |x|^2 + |v|^2 + |\log f^h|) f^h \rangle dx \leq \Gamma_3(t_{max}) \quad (2.37)$$

pour tout $t_{max} > 0$. Nous en déduisons classiquement qu'il existe une sous-suite encore notée $\{f^h\}_h$ telle que

$$f^h \xrightarrow{h \rightarrow 0} f \quad \text{faiblement dans } L^1([0, t_{max}] \times \mathbb{R}_x^D \times \mathbb{R}_v^D), \quad \forall t_{max} > 0. \quad (2.38)$$

De plus, il est clair que $C^h(v)$ converge ponctuellement vers v et est localement uniformément bornée. Cela suffit avec (2.38) pour obtenir la convergence du membre de gauche de (2.20) vers $\partial_t f + v \cdot \nabla_x f$ dans $\mathcal{D}'([0, +\infty[\times \mathbb{R}_x^D \times \mathbb{R}_v^D)$.

Pour la convergence du membre de droite non linéaire, nous obtenons tout d'abord la convergence faible de $\langle \mathbf{m} f^h \rangle$. L'estimation (2.37) donne

$$\langle (1, v)^T f^h(t, x, v) \rangle \xrightarrow{h \rightarrow 0} (\rho, \rho v)^T = \langle (1, v)^T f(t, x, v) \rangle \quad \forall t_{max} > 0 \quad (2.39)$$

faiblement dans $L^1([0, t_{max}] \times \mathbb{R}_x^D)$.

Cependant, à cause du manque d'estimations sur le comportement de f^h pour les grandes vitesses, l'estimation (2.37) n'est pas suffisante pour obtenir la convergence faible de $\langle \frac{1}{2} |v|^2 f^h \rangle$. Nous utilisons alors le lemme de Perthame [49] pour estimer $|v|^3 f^h$. Ce lemme est basé sur l'effet dispersif de l'inversion de $\partial_t + v \cdot \nabla_x$, qui existe aussi dans le cas discret en vitesse.

Lemme 2.4. *Soit $F \in L^1(\mathbb{R}^+ \times \mathbb{R}_x^D)^{N_h}$ solution de*

$$\partial_t F_k + v_k \cdot \nabla_x F_k = g_k, \quad F_k(0, x) = 0 \quad \forall k \in \mathcal{K}^h,$$

où $g \geq 0$ satisfait

$$\int_0^{t_{max}} \int_{\mathbb{R}_x^D} \langle |v|^2 g(t, x) \rangle_h dx dt \leq c. \quad (2.40)$$

Alors, pour tout borné K de \mathbb{R}_x^D , on a

$$\int_0^{t_{max}} \int_K \langle |v|^3 F(t, x) \rangle_h dx dt \leq c \operatorname{diam}(K).$$

La preuve de ce lemme est exactement la même que celle donnée dans [49] et nous ne l'écrivons pas ici. Remarquons que l'équilibre discret $\mathcal{E}_{\mathcal{K}}^h$ a la même énergie que $f_{\mathcal{K}}^h$, et donc qu'il satisfait l'estimation (2.40) du lemme 2.4. Par conséquent, on déduit de ce lemme que pour tout compact K de \mathbb{R}_x^D , on a

$$\sup_h \int_0^{t_{max}} \int_K \sum_{|v_k^h| > R} |v_k^h|^2 f_k^h(t, x) \Delta v_h^D dt dx \leq \frac{c(K)}{R}. \quad (2.41)$$

On a alors naturellement

$$\sup_h \int_0^{t_{max}} \int_K \int_{|v| \geq 2R} |v|^2 f^h(t, x, v) dt dx dv \leq \frac{c(K)}{R}, \quad (2.42)$$

et donc

$$|v|^2 f^h \rightharpoonup |v|^2 f \quad \text{faiblement dans } L^1([0, t_{max}] \times K \times \mathbb{R}_v^D), \quad (2.43)$$

$$\langle \frac{1}{2} |v|^2 f^h \rangle \rightharpoonup E = \langle \frac{1}{2} |v|^2 f \rangle \quad \text{faiblement dans } L^1([0, t_{max}] \times K), \quad (2.44)$$

pour tout compact K de \mathbb{R}_x^D et pour tout $t_{max} > 0$. Nous avons donc montré que $\langle \mathbf{m} f^h \rangle \rightharpoonup \boldsymbol{\rho} = \langle \mathbf{m} f \rangle$ faiblement dans $L^1([0, t_{max}] \times K)$ pour tout compact K .

étape 2: Convergence faible de \mathcal{E}^h

Nous avons besoin du lemme suivant.

Lemme 2.5. *Pour tout $t_{max} > 0$, il existe une constante $c(t_{max})$ tel que*

$$\sup_h \sup_{[0, t_{max}]} \int_{\mathbb{R}^D} \langle (1 + |x|^2 + |v|^2 + |\log \mathcal{E}^h|) \mathcal{E}^h \rangle dx \leq c(t_{max}). \quad (2.45)$$

Preuve. Les bornes sur $(1 + |x|^2 + |v|^2) \mathcal{E}^h$ proviennent directement de la définition de \mathcal{E}^h et de l'estimation (2.37). Pour $\mathcal{E}^h |\log \mathcal{E}^h|$, notons qu'en raison de la définition de l'équilibre discret $\mathcal{E}_{\mathcal{K}}^h$, on a

$$\int_{\mathbb{R}^D} \mathcal{E}^h \log \mathcal{E}^h dv \leq \int_{\mathbb{R}^D} f^h \log f^h dv.$$

Alors une technique classique permet de passer de $\mathcal{E}^h \log \mathcal{E}^h$ à $\mathcal{E}^h |\log \mathcal{E}^h|$ (voir [49]) et donne

$$\int_{\mathbb{R}^{2D}} \mathcal{E}^h |\log \mathcal{E}^h| dx dv \leq c(t_{max}).$$

□

Ce lemme montre que \mathcal{E}^h est faiblement compact dans $L^1([0, t_{max}] \times \mathbb{R}^{2D})$, ainsi, il existe une fonction \mathcal{M} telle que $\mathcal{E}^h \rightharpoonup \mathcal{M}$ faiblement dans $L^1([0, t_{max}] \times \mathbb{R}_x^D \times \mathbb{R}_v^D)$. On peut alors déduire des étapes 1 et 2 que la limite faible f de f^h satisfait l'équation

$$\partial_t f + v \cdot \nabla_x f = \mathcal{M} - f \quad \text{dans } \mathcal{D}'. \quad (2.46)$$

Les étapes suivantes sont consacrées à la preuve de $\mathcal{M} = M[\boldsymbol{\rho}]$.

étape 3: Convergence forte de $\boldsymbol{\rho}_h$

L'extension au cadre discret des lemmes de moyenne obtenue par Mischler [46] implique que les moyennes en vitesse de f^h sur des ensembles bornés sont en fait fortement compactes. Quitte à extraire une sous-suite, on a alors pour tout $R > 0$

$$\int_{|v| < R} \mathbf{m}(v) f^h(t, x, v) dv \rightarrow \int_{|v| < R} \mathbf{m}(v) f(t, x, v) dv$$

fortement dans $L^1([0, t_{max}] \times \mathbb{R}_x^D)$. D'après les estimations uniformes (2.37) et (2.42), nous obtenons donc

$$\langle \mathbf{m} f^h \rangle \rightarrow \boldsymbol{\rho} = \langle \mathbf{m} f \rangle \quad \text{fortement dans } L^1([0, t_{max}] \times K)$$

pour tout compact K . Par ailleurs, comme $\langle (1, v)^T f^h \rangle = (\rho_h, \rho_h u_h)^T$ et $\langle \frac{1}{2} |v|^2 f^h \rangle = E_h + \rho_h \frac{D}{24} \Delta v_h^2$, alors il est clair que ρ_h et $\langle \mathbf{m} f^h \rangle$ sont asymptotiquement équivalents. Par conséquent, on conclut que $\rho_h \rightarrow \rho$ fortement dans $L^1([0, t_{max}] \times K)$, pour tout compact K , pour tout $t_{max} > 0$.

étape 4: Passage à la limite

En extrayant encore une sous-suite, on obtient $\rho_h(t, x) \rightarrow \rho(t, x)$ p.p dans $[0, t_{max}] \times \mathbb{R}_x^D$. Alors le théorème 2.1 implique que sur l'ensemble $\Omega = \{(t, x), \rho(t, x) \text{ et } T(t, x) > 0\}$, on a

$$\alpha_h(t, x) \rightarrow \alpha(t, x) \quad \text{p.p.}$$

Donc

$$\mathcal{E}^h \rightarrow M[\rho] \quad \text{p.p dans } \Omega \times \mathbb{R}_v^D,$$

et puisque

$$\rho_h(t, x) = \|\mathcal{E}^h(t, x, \cdot)\|_{L^1} \rightarrow \rho(t, x) = 0 \quad \text{p.p dans } {}^c\Omega,$$

alors

$$\mathcal{E}^h(t, x) \rightarrow 0 = M[\rho](t, x) \quad \text{p.p dans } {}^c\Omega \times \mathbb{R}_v^D.$$

Ceci prouve que \mathcal{E}^h converge ponctuellement vers $M[\rho]$. En combinant ce résultat avec celui de l'étape 2, on trouve que $\mathcal{M} = M[\rho]$. Par conséquent, le membre de droite de (2.13) converge vers $(M[\rho] - f)$ faiblement dans L^1 , et nous pouvons conclure que f est solution de l'équation BGK (2.8).

Remarque 2.3. Si la condition initiale f^0 est suffisamment régulière (par exemple $f^0 \in BV$), de sorte que $f^{0,h}$ converge vers f^0 dans L^1 , alors la convergence du théorème 2.3 est en fait forte dans L^1 . On peut en effet montrer par des arguments similaires à ceux de Lions [37] que $\log(1 + f^h)$ converge faiblement dans $\log(1 + f)$.

Remarque 2.4. Si l'on fait les hypothèses de Mischler [45] sur la condition initiale, alors la solution de BGK est unique. Dans ce cas, la suite entière f^h converge vers f .

Chapitre 3

Discrétisation de l'opérateur de transport axisymétrique

Dans la méthode discrète en vitesse que nous avons présentée dans les chapitres précédents, une grande attention a été portée aux propriétés que doit vérifier le terme source discret pour que les lois de conservation et de dissipation d'entropie soient satisfaites. De nombreux travaux ont d'ailleurs été menés ces dernières années sur d'autres équations cinétiques pour obtenir de telles propriétés. Dans ces travaux, la discrétisation du terme de transport ne pose en général pas de difficulté, car la variable de vitesse ne joue que le rôle d'un paramètre.

En revanche, ce n'est pas le cas lorsque l'on traite des écoulements en géométrie cylindrique. La transformation en coordonnées cylindrique permet de calculer un certain type d'écoulements tridimensionnels en deux dimensions d'espace seulement. De tels écoulements sont dits axisymétriques, comme par exemple un écoulement sans incidence autour d'une sphère ou d'un cône, ou un écoulement entre deux cylindres en rotation. Dans les équations cinétiques exprimées en coordonnées cylindriques, il apparaît des termes de gradient en vitesse dans l'opérateur de transport. La discrétisation en vitesse de ces équations introduit donc une difficulté supplémentaire par rapport aux équations cartésiennes. En effet, indépendamment de la discrétisation du terme source, les lois de conservation macroscopiques et la loi de dissipation d'entropie peuvent ne pas être vérifiées si la discrétisation du terme de transport n'est pas effectuée correctement.

Dans ce chapitre, nous mettons en évidence les propriétés que doit vérifier la discrétisation en vitesse de l'opérateur de transport en coordonnées cylindriques, pour obtenir les lois macroscopiques correctes. Il apparaît clairement que plusieurs schémas fréquemment utilisés dans la littérature ne vérifient pas ces propriétés. Nous proposons alors différents schémas pour corriger ce défaut. En particulier, nous proposons deux schémas qui garantissent la positivité de la fonction de distribution, et qui vérifient les lois de conservation ainsi que la loi de dissipation d'entropie. A notre connaissance, c'est la première fois que des méthodes garantissant toutes ces propriétés sont présentées. Nous insistons sur le fait que ces schémas peuvent s'utiliser pour une large classe d'équations cinétiques (Boltzmann, BGK, Fokker-Planck, etc.). Nous ne nous intéressons pas ici à la discrétisation des opérateurs de collision exprimés en coordonnées cylindriques (voir à ce sujet [34] pour Fokker-Planck).

Le reste de ce chapitre s'articule de la façon suivante. Dans la section 3.1, nous faisons des remarques générales sur les équations cinétiques et la façon dont on obtient les lois macroscopiques, en insistant sur la différence essentielle entre les représentations cartésiennes et cylindriques. Dans la section 3.2, nous donnons les définitions nécessaires à la discrétisation en vitesse. Dans la section 3.3, nous exposons les différentes propriétés que doit nécessairement vérifier une discrétisation différence

finies de la dérivée en vitesse pour obtenir les lois de conservation et d'entropie axisymétriques. Nous proposons différents modèles discrets qui vérifient un certain nombre de ces propriétés. Dans la section 3.4, nous étudions une autre approche, introduite récemment par Larina et Rykov, qui permet aussi d'obtenir des modèles ayant certaines propriétés de conservation. Nous rappelons leur idée ainsi que la discrétisation qu'ils proposent et les propriétés qu'elle vérifie. Cependant, leur modèle ne garantit pas la positivité de la fonction de distribution. Nous proposons donc une modification qui élimine ce problème tout en conservant les autres propriétés.

3.1 Equation de transport et lois de conservation

Considérons une équation cinétique de type Boltzmann sans force extérieure, en coordonnées cartésiennes

$$\partial_t f + v \cdot \nabla_X f = Q(f), \quad t > 0, \quad X = (x, y, z) \in \mathbb{R}^3, \quad v = (v_x, v_y, v_z) \in \mathbb{R}^3. \quad (3.1)$$

En multipliant cette équation par $\mathbf{m}(v) = (1, v, \frac{1}{2}|v|^2)^T$ et $(1 + \log f)$, on obtient les lois de conservation de la densité, quantité de mouvement, et énergie totale, et de dissipation d'entropie

$$\partial_t \int \mathbf{m} f dv + \nabla_X \cdot \int v \mathbf{m} f dv = 0, \quad (3.2a)$$

$$\partial_t \int f \log f dv + \nabla_X \cdot \int v f \log f dv = \int Q(f) \log f \leq 0. \quad (3.2b)$$

Pour résoudre numériquement cette équation cinétique, une première étape est généralement de la discrétiser par rapport à v . Pour garantir la robustesse des schémas numériques, on essaie d'assurer que les lois de conservation et de dissipation d'entropie sont encore vérifiées par l'équation discrétisée (ce qui a été fait pour BGK au chapitre 1). Pour cela, il est fondamental que la discrétisation de l'opérateur de collision respecte les équivalents discrets des deux propriétés suivantes

$$\int \mathbf{m} Q(f) dv = 0, \quad (3.3a)$$

$$\int Q(f) \log f dv \leq 0. \quad (3.3b)$$

Il existe plusieurs schémas garantissant ces propriétés au niveau discret pour différentes équations cinétiques : voir par exemple [11] ou [55] pour l'opérateur de Boltzmann, [43] pour l'opérateur de BGK, [19] pour l'opérateur de Fokker-Planck.

En revanche, lorsqu'on passe en coordonnées cylindriques, il apparaît que la discrétisation de l'opérateur de transport doit elle aussi vérifier des propriétés non triviales. La formulation axisymétrique de l'équation (3.1) s'obtient de la façon suivante. On exprime les variables d'espace dans un système de coordonnées cylindriques

$$(x, y, z) = (x, r \cos \varphi, r \sin \varphi),$$

et pour exploiter la symétrie axiale en espace, il faut en outre exprimer la vitesse v dans le repère local $(\vec{e}_x, \vec{e}_r, \vec{e}_\varphi)$

$$\begin{pmatrix} v_x \\ v_r \\ v_\varphi \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi & \sin \varphi \\ 0 & -\sin \varphi & \cos \varphi \end{pmatrix} \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}.$$

L'hypothèse d'écoulement axisymétrique se traduit alors par

$$\partial_\varphi f(t, x, r, \varphi, v_x, v_r, v_\varphi) = 0,$$

et l'équation cartésienne (3.1) devient

$$\partial_t f + v_x \partial_x f + v_r \partial_r f + \frac{v_\varphi^2}{r} \partial_{v_r} f - \frac{v_r v_\varphi}{r} \partial_{v_\varphi} f = Q(f), \quad (3.4)$$

où l'on voit apparaître des gradients en vitesse de f , qui sont en fait des termes d'inertie dus au repère local. Les caractéristiques de cette équation de transport sont plus compliquées que pour l'équation cartésienne, car ce sont maintenant les courbes de \mathbb{R}^4 définies par

$$\dot{x}(t) = v_x, \quad \dot{r}(t) = v_r, \quad \dot{v}_r(t) = \frac{v_\varphi^2}{r}, \quad \dot{v}_\varphi(t) = -\frac{v_r v_\varphi}{r}. \quad (3.5)$$

On peut cependant voir facilement qu'elles vérifient $v_r^2(t) + v_\varphi^2(t) = cte$, ce qui signifie que dans le plan (v_r, v_φ) , les caractéristiques sont des cercles. Par conséquent on peut remarquer que lorsqu'on remplacera l'espace des vitesses par un domaine borné en vue de la discrétisation de (3.4), il faudra que ce domaine soit de section circulaire dans le plan (v_r, v_φ) . En effet, dans le cas contraire, les caractéristiques vont rencontrer le bord du domaine, et on aura besoin de conditions aux limites en vitesse qui ne sont pas physiques, puisqu'elles n'existent pas dans l'équation cartésienne. Afin de discrétiser facilement ce domaine circulaire, il est plus judicieux d'exprimer à présent les composantes radiale et azimuthale de la vitesse (v_r, v_φ) dans un système de coordonnées polaires. On définit ζ and ω par

$$(v_r, v_\varphi) = (\zeta \cos \omega, \zeta \sin \omega),$$

et l'équation (3.4) devient

$$\partial_t f + v_x \partial_x f + \zeta \cos \omega \partial_r f - \frac{\zeta \sin \omega}{r} \partial_\omega f = Q(f). \quad (3.6)$$

Une expression conservative de cette équation est

$$\partial_t r f + v_x \partial_x r f + \zeta \cos \omega \partial_r r f - \zeta \partial_\omega (\sin \omega f) = r Q(f). \quad (3.7)$$

Par souci de simplification, nous ferons dans la suite l'hypothèse de parité

$$f(\omega) = f(-\omega), \quad (3.8)$$

qui est par exemple valable pour des écoulements extérieurs sans incidence autour de corps à géométrie axisymétrique. Notons cependant que tous les modèles et les propriétés donnés dans la suite sont identiques dans le cas où l'on ne fait pas cette hypothèse.

Pour obtenir les lois de conservation et de dissipation d'entropie axisymétriques, on peut faire les changements de variables précédents dans (3.2a), mais nous pensons qu'il est plus instructif de les obtenir à partir de l'une ou l'autre des équations (3.6) et (3.7). Cependant, les lois de conservation ne s'expriment pas de façon aussi simple qu'en coordonnées cartésiennes, et on doit introduire certaines quantités macroscopiques qui sont la densité ρ , les quantités de mouvement axiale et radiale ρu_x et

ρu_r , les contraintes normales $\rho\Theta_{xx}$, $\rho\Theta_{rr}$ et $\rho\Theta_{\varphi\varphi}$, la contrainte tangentielle $\rho\Theta_{xr}$, l'énergie totale E , et les flux de chaleur q_x et q_r :

$$\rho = \int f d\mu, \quad (3.9a)$$

$$\rho u_x = \int v_x f d\mu, \quad \rho u_r = \int \zeta \cos \omega f d\mu, \quad (3.9b)$$

$$\rho\Theta_{xx} = \int (v_x - u_x)^2 f d\mu, \quad \rho\Theta_{rr} = \int (\zeta \cos \omega - u_r)^2 f d\mu, \quad \rho\Theta_{\varphi\varphi} = \int \zeta^2 \sin^2 \omega f d\mu, \quad (3.9c)$$

$$\rho\Theta_{xr} = \int (v_x - u_x)(\zeta \cos \omega - u_r) f d\mu, \quad (3.9d)$$

$$E = \int \frac{1}{2}(v_x^2 + \zeta^2) f d\mu, \quad (3.9e)$$

$$q_x = \int \frac{1}{2}((v_x - u_x)^2 + (\zeta \cos \omega - u_r)^2 + \zeta^2 \sin^2 \omega)(v_x - u_x) f d\mu, \quad (3.9f)$$

$$q_r = \int \frac{1}{2}((v_x - u_x)^2 + (\zeta \cos \omega - u_r)^2 + \zeta^2 \sin^2 \omega)(\zeta \cos \omega - u_r) f d\mu, \quad (3.9g)$$

où l'on a noté $d\mu = \zeta dv_x d\zeta d\omega$ la mesure de Lebesgue dans \mathbb{R}^3 exprimée en coordonnées cylindriques. L'hypothèse de parité (3.8) implique que les autres quantités macroscopiques de même ordre en v sont nulles. On obtient alors les lois de conservation axisymétriques en multipliant (3.7) par $(1, v_x, \zeta \cos \omega, \frac{1}{2}(v_x^2 + \zeta^2))$ puis en sommant sur $\mathbb{R} \times \mathbb{R}^+ \times [0, 2\pi]$; en raison des propriétés de Q , on trouve

$$\begin{aligned} \partial_t r \int \begin{pmatrix} 1 \\ v_x \\ \zeta \cos \omega \\ \frac{1}{2}(v_x^2 + \zeta^2) \end{pmatrix} f d\mu + \partial_x r \int \begin{pmatrix} 1 \\ v_x \\ \zeta \cos \omega \\ \frac{1}{2}(v_x^2 + \zeta^2) \end{pmatrix} v_x f d\mu + \partial_r r \int \begin{pmatrix} 1 \\ v_x \\ \zeta \cos \omega \\ \frac{1}{2}(v_x^2 + \zeta^2) \end{pmatrix} \zeta \cos \omega f d\mu \\ - \int \begin{pmatrix} 1 \\ v_x \\ \zeta \cos \omega \\ \frac{1}{2}(v_x^2 + \zeta^2) \end{pmatrix} \zeta \partial_\omega (\sin \omega f) d\mu = 0. \end{aligned}$$

Par définition des quantités macroscopiques, on obtient

$$\begin{aligned} \partial_t r \begin{pmatrix} \rho \\ \rho u_x \\ \rho u_r \\ E \end{pmatrix} + \partial_x r \begin{pmatrix} \rho u_x \\ \rho u_x^2 + \rho\Theta_{xx} \\ \rho u_x u_r + \rho\Theta_{xr} \\ u_x E + \rho(\Theta_{xx} u_x + \Theta_{xr} u_r) + q_x \end{pmatrix} + \partial_r r \begin{pmatrix} \rho u_r \\ \rho u_x u_r + \rho\Theta_{xr} \\ \rho u_r^2 + \rho\Theta_{rr} \\ u_r E + \rho(\Theta_{xr} u_x + \Theta_{rr} u_r) + q_r \end{pmatrix} \\ = \int \begin{pmatrix} 1 \\ v_x \\ \zeta \cos \omega \\ \frac{1}{2}(v_x^2 + \zeta^2) \end{pmatrix} \zeta \partial_\omega (\sin \omega f) d\mu. \end{aligned} \quad (3.10)$$

De plus, en multipliant (3.7) par $(1 + \log f)$, on a

$$\begin{aligned} \partial_t r \int f \log f d\mu + \partial_x r \int v_x f \log f d\mu + \partial_r r \int \zeta \cos \omega f \log f d\mu + \int \zeta \cos \omega f d\mu \\ = \int (1 + \log f) \zeta \partial_\omega (\sin \omega f) d\mu + r \int Q(f) \log f d\mu. \end{aligned} \quad (3.11)$$

Le membre de gauche de (3.10) et (3.11) est uniquement dû à $\partial_t r f + v_x \partial_x r f + \zeta \cos \omega \partial_r r f$, où (v_x, ζ, ω) ne jouent que le rôle de paramètres. Cette partie ne pose donc pas de difficulté au niveau discret. Pour le second membre, un simple calcul montre que

$$\int \begin{pmatrix} 1 \\ v_x \\ \zeta \cos \omega \\ \frac{1}{2}(v_x^2 + \zeta^2) \end{pmatrix} \zeta \partial_\omega (\sin \omega f) d\mu = \begin{pmatrix} 0 \\ 0 \\ \rho \Theta_{\varphi\varphi} \\ 0 \end{pmatrix} \quad \text{et} \quad \int (1 + \log f) \zeta \partial_\omega (\sin \omega f) d\mu = \int \zeta \cos \omega f d\mu. \quad (3.12)$$

On obtient ainsi les lois de conservation et d'entropie suivantes

$$\partial_t r \rho + \partial_x r \rho u_x + \partial_r r \rho u_r = 0, \quad (3.13a)$$

$$\partial_t r \rho u_x + \partial_x r (\rho u_x^2 + \rho \Theta_{xx}) + \partial_r r (\rho u_x u_r + \rho \Theta_{xr}) = 0, \quad (3.13b)$$

$$\partial_t r \rho u_r + \partial_x r (\rho u_x u_r + \rho \Theta_{xr}) + \partial_r r (\rho u_r^2 + \rho \Theta_{rr}) = \rho \Theta_{\varphi\varphi}, \quad (3.13c)$$

$$\partial_t r E + \partial_x r (u_x E + \rho (\Theta_{xx} u_x + \Theta_{xr} u_r) + q_x) + \partial_r r (u_r E + \rho (\Theta_{xr} u_x + \Theta_{rr} u_r) + q_r) = 0. \quad (3.13d)$$

$$\partial_t r \int f \log f d\mu + \partial_x r \int v_x f \log f d\mu + \partial_r r \int \zeta \cos \omega f \log f d\mu = r \int Q(f) \log f d\mu \leq 0. \quad (3.14)$$

On remarque que seules les relations (3.13a), (3.13b), (3.13d) sont des lois de conservation à proprement parler, puisque la relation (3.13c) contient le terme source de pression $\rho \Theta_{\varphi\varphi}$.

D'après les relations de (3.12), il est clair que la discrétisation en vitesse, et plus particulièrement la discrétisation en ω va poser une difficulté. En effet, il faut que l'approximation discrète de $\partial_\omega (\sin \omega f)$ vérifie des analogues discrets de ces relations. Plus précisément, il faut que la discrétisation en ω vérifie des équivalents discrets de

$$\int_0^{2\pi} \partial_\omega (\sin \omega f) d\omega = 0, \quad (3.15a)$$

$$\int_0^{2\pi} \cos \omega \partial_\omega (\sin \omega f) d\omega = \int_0^{2\pi} \sin^2 \omega f d\omega, \quad (3.15b)$$

$$\int_0^{2\pi} \partial_\omega (\sin \omega f) \log f d\omega = \int_0^{2\pi} \cos \omega f d\omega. \quad (3.15c)$$

La relation (3.15a) sert à obtenir les équations de conservation (3.13a), (3.13b), (3.13d), la relation (3.15b) implique l'équation (3.13c), et (3.15c) suffit pour avoir l'équation de dissipation d'entropie (3.14). Remarquons qu'on peut affaiblir l'égalité (3.15c). En effet, pour obtenir la relation de dissipation d'entropie (3.14), il suffit de vérifier l'inégalité

$$\int_0^{2\pi} \partial_\omega (\sin \omega f) \log f d\omega \leq \int_0^{2\pi} \cos \omega f d\omega. \quad (3.15d)$$

Cependant, le taux de dissipation d'entropie est sans doute surestimé puisque la relation macroscopique obtenue est alors

$$\partial_t r \int f \log f d\mu + \partial_x r \int v_x f \log f d\mu + \partial_r r \int \zeta \cos \omega f \log f d\mu \leq r \int Q(f) \log f d\mu \leq 0,$$

mais l'important est qu'on dissipe effectivement l'entropie. Enfin, une autre propriété non triviale à respecter par la discrétisation est que les fonctions constantes en t, x, r, ω sont solutions de (3.7), ce qui équivaut à

$$\partial_\omega(\sin \omega) = \cos \omega. \quad (3.15e)$$

En effet, il est connu que la non préservation des écoulements constants donne généralement des schémas peu précis et non robustes.

Si l'on veut discrétiser l'équation non conservative (3.6), une procédure analogue montre que la discrétisation en ω doit vérifier des équivalents discrets de

$$\int_0^{2\pi} \sin \omega \partial_\omega f d\omega = - \int_0^{2\pi} \cos \omega f d\omega, \quad (3.16a)$$

$$\int_0^{2\pi} \cos \omega \sin \omega \partial_\omega f d\omega = \int_0^{2\pi} (-\cos^2 \omega + \sin^2 \omega) f d\omega, \quad (3.16b)$$

$$\int_0^{2\pi} \sin \omega \partial_\omega f (1 + \log f) d\omega = - \int_0^{2\pi} \cos \omega f \log f d\omega, \quad (3.16c)$$

ou

$$\int_0^{2\pi} \sin \omega \partial_\omega f (1 + \log f) d\omega \leq - \int_0^{2\pi} \cos \omega f \log f d\omega, \quad (3.16d)$$

où (3.16a) est nécessaire pour avoir les lois de conservation (3.13a), (3.13b), et (3.13d). La relation (3.16b) est nécessaire pour avoir (3.13c). Les relations (3.16c) et (3.16d) donnent (3.14). Pour respecter les solutions constantes en t, x, r, ω , la propriété à garantir est

$$\partial_\omega 1 = 0. \quad (3.16e)$$

Dans la suite, nous ne nous intéressons pas à la discrétisation de l'opérateur de collision (voir les chapitres précédents pour l'opérateur de BGK), nous étudions uniquement la discrétisation en ω de l'opérateur de transport. La discrétisation en v_x, ζ ne posant pas de difficulté (car ces variables jouent le rôle de paramètres), nous les laissons donc continues.

3.2 Discrétisation de f et définition des quantités macroscopiques

Dans cette section, nous introduisons quelques notations relatives à la discrétisation en ω , et nous définissons des approximations discrètes des quantités macroscopiques (3.9). La variable ω varie dans $[0, 2\pi]$, que l'on discrétise par $\{\omega_q\}_{q \geq 0}$. Nous notons f_q une approximation de $f(\omega_q)$. Les intégrales sur $[0, 2\pi]$ intervenant dans la définition des moments sont approchés par une formule de quadrature

$$\int_0^{2\pi} f(\omega) d\omega \approx \sum_{q \geq 0} f_q \Delta\omega.$$

Pour simplifier les expressions des quantités macroscopiques (3.9) après discrétisation, nous notons

$$\langle \sum_q f_q \rangle = \int_{\mathbb{R}} \int_0^{+\infty} \sum_{q \geq 0} f_q(v_x, \zeta) \Delta\omega dv_x \zeta d\zeta.$$

Ainsi, les quantités macroscopiques sont à présent définies par

$$\rho = \langle \sum_q f_q \rangle, \quad (3.17a)$$

$$\rho u_x = \langle \sum_q v_x f_q \rangle, \quad \rho u_r = \langle \sum_q \zeta \cos \omega_q f_q \rangle, \quad (3.17b)$$

$$\rho \Theta_{xx} = \langle \sum_q (v_x - u_x)^2 f_q \rangle, \quad \rho \Theta_{rr} = \langle \sum_q (\zeta \cos \omega_q - u_r)^2 f_q \rangle, \quad \rho \Theta_{\varphi\varphi} = \langle \sum_q \zeta^2 \sin^2 \omega_q f_q \rangle, \quad (3.17c)$$

$$\rho \Theta_{xr} = \langle \sum_q (v_x - u_x)(\zeta \cos \omega_q - u_r) f_q \rangle, \quad (3.17d)$$

$$E = \langle \sum_q \frac{1}{2}(v_x^2 + \zeta^2) f_q \rangle, \quad (3.17e)$$

$$q_x = \langle \sum_q \frac{1}{2}((v_x - u_x)^2 + (\zeta \cos \omega_q - u_r)^2 + \zeta^2 \sin^2 \omega_q)(v_x - u_x) f_q \rangle, \quad (3.17f)$$

$$q_r = \langle \sum_q \frac{1}{2}((v_x - u_x)^2 + (\zeta \cos \omega_q - u_r)^2 + \zeta^2 \sin^2 \omega_q)(\zeta \cos \omega_q - u_r) f_q \rangle, \quad (3.17g)$$

Les points de discrétisation ω_q et la formule de quadrature sont à définir pour chaque schéma, de façon à éliminer les termes de bord dans les intégrations par parties discrètes.

3.3 Discrétisation de la dérivée en vitesse

Dans cette section, nous explicitons les propriétés que doit vérifier la discrétisation des équations (3.6) et (3.7) sans terme source pour que les lois de conservation (3.13) et de dissipation d'entropie (3.14) soient satisfaites. Nous donnons en outre d'autres propriétés nécessaires pour que les schémas obtenus soient efficaces, comme la positivité et la conservation des écoulements uniformes. Nous étudions ensuite différents schémas de discrétisation.

Supposons que le terme $\partial_\omega f$ de l'équation non conservative (3.6) est approché par différences finies par $D(f)_q$ avec $q \geq 0$. L'approximation discrète de (3.6) sans terme source est alors

$$\partial_t f_q + v_x \partial_x f_q + \zeta \cos \omega_q \partial_r f_q - \frac{\zeta \sin \omega_q}{r} D(f)_q = 0. \quad (3.18)$$

Les équivalents discrets des relations (3.16) pour obtenir les lois de conservation et d'entropie sont

$$\sum_q \sin \omega_q D(f)_q = - \sum_q \cos \omega_q f_q, \quad (3.19a)$$

$$\sum_q \cos \omega_q \sin \omega_q D(f)_q = \sum_q (-\cos^2 \omega_q + \sin^2 \omega_q) f_q, \quad (3.19b)$$

$$\sum_q \sin \omega_q D(f)_q (1 + \log f_q) = - \sum_q \cos \omega_q f_q \log f_q, \quad (3.19c)$$

ou

$$\sum_q \sin \omega_q D(f)_q (1 + \log f_q) \leq - \sum_q \cos \omega_q f_q \log f_q, \quad (3.19d)$$

$$D(1)_q = 0. \quad (3.19e)$$

On a le résultat suivant

Proposition 3.1. – *Si les quantités macroscopiques sont définies par (3.17), alors elles vérifient les lois de conservation (3.13a), (3.13b), et (3.13d) si et seulement si la propriété (3.19a) est satisfaite;*

- *la propriété (3.19b) est équivalente à la loi de conservation de ρ_r (3.13c);*
- *la propriété (3.19c) (ou (3.19d)) implique la loi de dissipation d'entropie (3.14);*
- *les fonctions de distributions uniformes en t, x, r, q sont solutions de (3.18) si et seulement si la propriété (3.19e) est satisfaite.*

Preuve. Il suffit de multiplier l'équation (3.18) par $(1, v_x, \zeta \cos \omega_q, \frac{1}{2}(v_x^2 + \zeta^2))$ et de sommer sur q . On obtient les lois (3.13) avec des termes supplémentaires qui s'annulent si et seulement si les propriétés (3.19a) et (3.19b) sont satisfaites. Pour la relation d'entropie (3.14), il suffit de multiplier l'équation (3.18) par $(1 + \log f_q)$; il apparaît un terme qui est négatif ou nul si (3.19c) ou (3.19d) sont vérifiées. Enfin, l'équivalence entre (3.19e) et la préservation des écoulements uniformes est évidente. \square

Dans le cas de l'équation conservative (3.7), le terme $\partial_\omega(\sin \omega f)$ est approché par $D(\sin \omega f)_q$. L'approximation discrète de (3.7) sans terme source est alors

$$\partial_t r f_q + v_x \partial_x r f_q + \zeta \cos \omega_q \partial_r r f_q - \zeta D(\sin \omega f)_q = 0. \quad (3.20)$$

Les équivalents discrets des propriétés (3.15) pour obtenir les lois de conservation et d'entropie, ainsi que la préservation des écoulements uniformes, sont

$$\sum_q D(\sin \omega f)_q = 0, \quad (3.21a)$$

$$\sum_q \cos \omega_q D(\sin \omega f)_q = \sum_q \sin^2 \omega_q f_q, \quad (3.21b)$$

$$\sum_q D(\sin \omega f)_q \log f_q = \sum_q \cos \omega_q f_q, \quad (3.21c)$$

ou

$$\sum_q D(\sin \omega f)_q \log f_q \leq \sum_q \cos \omega_q f_q, \quad (3.21d)$$

$$D(\sin \omega)_q = \cos \omega_q. \quad (3.21e)$$

On peut alors montrer le même résultat que pour l'équation non conservative (proposition 3.1) si ces propriétés sont vérifiées.

Proposition 3.2. – *si les quantités macroscopiques sont définies par (3.17), alors elles vérifient les lois de conservation (3.13a), (3.13b), et (3.13d) si et seulement si la propriété (3.21a) est satisfaite;*

- *la propriété (3.21b) est équivalente à la loi de conservation de ρ_r (3.13c);*

- la propriété (3.21c) (ou (3.21d)) implique la loi de dissipation d'entropie (3.14);
- les fonctions de distributions uniformes en t, x, r, q sont solutions de (3.20) si et seulement si la propriété (3.21e) est satisfaite.

Dans les sections suivantes, nous étudions différents choix de discrétisation.

3.3.1 Discrétisation décentrée de l'équation non conservative

Cette méthode est notamment proposée par Shakhov dans [58]. En faisant l'hypothèse de parité (3.8), on considère l'équation (3.6) sur $[0, \pi]$, et D est défini par une approximation décentrée d'ordre 1

$$D(f)_q = \frac{f_{q+1} - f_q}{\Delta\omega}. \quad (3.22)$$

L'intervalle $[0, \pi]$ est discrétisé par $\omega_q = q\Delta\omega$ avec $q = 0..Q$ et $\omega_Q = \pi$. Les moments de f sont calculés par une formule de rectangles ouverte à gauche, *i.e.* en remplaçant \sum_q par $2 \sum_{q=1}^Q$ dans les formules (3.17). En fait, Shakhov [58] considère plutôt le domaine $[\pi, 2\pi]$, et prend donc $D(f)_q = \frac{f_q - f_{q-1}}{\Delta\omega}$, ce qui donne les mêmes propriétés. Les propriétés du modèle discret (3.18) avec cette définition (3.22) sont les suivantes.

Proposition 3.3. (i) la positivité de f est préservée :

$$f_q(0, x, r, v_x, \zeta) \geq 0 \quad \forall q, x, r, v_x, \zeta \Rightarrow f_q(t, x, r, v_x, \zeta) \geq 0 \quad \forall q, t, x, r, v_x, \zeta;$$

(ii) les écoulements uniformes en t, x, r, q sont solutions;

(iii) les équations de conservation (3.13) sont vérifiées à $O(\Delta\omega)$ près.

Preuve. Pour la positivité, on remarque que

$$\partial_t f_q + v_x \partial_x f_q + \zeta \cos \omega_q \partial_r f_q \geq -\frac{\zeta}{r} \sin \omega_q f_q,$$

car $\sin \omega_q \geq 0$ pour $q = 0..Q$. Par conséquent

$$\frac{d}{dt} f_q(t, x + tv_x, r + t\zeta \cos \omega_q, v_x, \zeta) \geq -\frac{\zeta}{r + t\zeta \cos \omega_q} f_q,$$

ce qui implique $f_q \geq 0$ si $f_q(t=0) \geq 0$. La propriété (ii) est évidente puisque (3.19e) est vérifiée. Pour montrer (iii), il suffit de remarquer que l'adjoint formel D^* de D vérifie

$$D^*(\sin \omega)_q = -\frac{\sin \omega_q - \sin \omega_{q-1}}{\Delta\omega} = -\cos \omega_q + O(\Delta\omega),$$

$$D^*(\cos \omega \sin \omega)_q = -\frac{\cos \omega_q \sin \omega_q - \cos \omega_{q-1} \sin \omega_{q-1}}{\Delta\omega} = -\cos^2 \omega_q + \sin^2 \omega_q + O(\Delta\omega).$$

Par conséquent, les propriétés (3.19a) et (3.19b) sont vérifiées à $O(\Delta\omega)$ près :

$$\sum_{q=1}^Q \sin \omega_q D(f)_q = \sum_{q=1}^Q D^*(\sin \omega)_q f_q = -\sum_{q=1}^Q \cos \omega_q f_q + O(\Delta\omega),$$

$$\sum_{q=1}^Q \cos \omega_q \sin \omega_q D(f)_q = \sum_{q=1}^Q D^*(\cos \omega \sin \omega)_q f_q = \sum_{q=1}^Q (-\cos^2 \omega_q + \sin^2 \omega_q) f_q + O(\Delta\omega),$$

et on conclut comme dans la preuve de la proposition 3.1. Remarquons que les termes de bord n'apparaissent pas dans les relations précédentes car $\sin \omega_0 = \sin \omega_Q = 0$. \square

Dans le cas où l'on ne fait pas l'hypothèse de parité (3.8), on discrétise l'intervalle $[0, 2\pi]$ par $\omega_q = q\Delta\omega$ avec $q = 0..Q$ et $\omega_Q = 2\pi - \Delta\omega$. On choisit $Q = 2m - 1$ de sorte que $\omega_m = \pi$. Ainsi l'opérateur D est naturellement généralisé par

$$D(f)_q = \begin{cases} \frac{f_{q+1} - f_q}{\Delta\omega} & \text{pour } q = 1..m, \\ \frac{f_q - f_{q-1}}{\Delta\omega} & \text{pour } q = m..Q, \end{cases} \quad (3.23)$$

afin de préserver la positivité de f . Pour éliminer les termes de bord dans les intégrations par parties, on définit les moments de f par les formules (3.17) en remplaçant \sum_q par $\sum_1^m + \sum_m^Q$, ce qui s'interprète comme une formule des rectangles ouverte à gauche sur $[0, \pi]$ et à droite sur $[\pi, 2\pi]$. On obtient alors les mêmes propriétés que sur $[0, \pi]$.

Remarque 3.1. Sone et al. [62] ont proposé une discrétisation décentrée d'ordre deux sur $[0, \pi]$ en introduisant

$$D(f)_q = \frac{1}{\Delta\omega} \left(-\frac{1}{2}f_{q+2} + 2f_{q+1} - \frac{3}{2}f_q \right).$$

Cependant, les termes de bord ne disparaissent pas dans les intégrations par parties discrètes, lorsqu'on utilise des formules de quadratures telles que celles des rectangles ou des trapèzes.

3.3.2 Modification "trigonométrique"

Nous modifions l'opérateur décentré d'ordre un précédent pour le rendre exact sur le cosinus et le sinus, *i.e.* on pose

$$D(f)_q = \frac{f_{q+1} - \cos \Delta\omega f_q}{\sin \Delta\omega}.$$

Les propriétés de l'équation discrète (3.18) sont données dans la proposition suivante.

Proposition 3.4. (i) la positivité de f est préservée;

(ii) les équations de conservation de ρ , ρu_x , E sont vérifiées exactement (Eq. (3.13a), (3.13b), et (3.13d));

(iii) les écoulements uniformes en t, x, r, q sont solutions à $O(\Delta\omega)$ près.

Preuve. La propriété de positivité est due au décentrage comme pour le schéma précédent. Pour les lois de conservation, il faut noter que D et son adjoint formel D^* ont les propriétés suivantes

$$\begin{aligned} D(\cos \omega)_q &= -\sin \omega_q & D^*(\cos \omega)_q &= \sin \omega_q \\ D(\sin \omega)_q &= \cos \omega_q & D^*(\sin \omega)_q &= -\cos \omega_q. \end{aligned}$$

Cela se voit facilement en remarquant que $\omega_{q\pm 1} = \omega_q \pm \Delta\omega$ et en développant $\cos \omega_{q\pm 1}$ et $\sin \omega_{q\pm 1}$. Ainsi, il est clair que la propriété (3.19a) qui garantit la conservation de ρ , ρu_x et E est vérifiée exactement, et la proposition 3.1 permet de conclure. En revanche, cette modification du schéma classique fait perdre la conservation des écoulements uniformes, car on a évidemment

$$D(1)_q = \frac{1 - \cos \Delta\omega}{\sin \Delta\omega} = 1 + O(\Delta\omega),$$

c'est-à-dire que (3.19e) est vérifiée à $O(\Delta\omega)$ près. □

Nous présentons dans la section suivante une discrétisation qui préserve les écoulements uniformes.

Remarque 3.2. On peut modifier l'opérateur décentré d'ordre deux de Sone et al. (cf. remarque 3.1) pour le rendre exact sur le cosinus et le sinus de la façon suivante

$$D(f)_q = \frac{1}{\sin \Delta\omega} \left(-\frac{1}{2}f_{q+2} + (1 + \cos \Delta\omega)f_{q+1} - \frac{1 + 2 \cos \Delta\omega}{2}f_q \right).$$

On vérifie facilement que $D(\sin \omega)_q = \cos \omega_q$ et $D^*(\cos \omega)_q = \sin \omega_q$. Mais comme on l'a noté dans la remarque 3.1, les termes de bord ne disparaissent pas dans les intégrations par parties discrètes, et cette modification ne suffit pas pour obtenir les propriétés (3.19).

3.3.3 Discrétisation centrée "trigonométrique" de l'équation conservative

Nous considérons ici une discrétisation de l'équation conservative sur $[0, 2\pi]$. Nous proposons un opérateur du même type que le précédent, exact pour le cosinus et le sinus, mais qui est aussi exact sur les fonctions constantes

$$D(f)_q = \frac{f_{q+1} - f_{q-1}}{2 \sin \Delta\omega}.$$

L'intervalle $[0, 2\pi]$ est discrétisé par $\omega_q = q\Delta\omega$ avec $q = 0..Q$ et $\omega_Q = 2\pi - \Delta\omega$. Les moments de f sont calculés par une formule des rectangles ouverte à droite, *i.e.* en remplaçant \sum_q par $\sum_{q=0}^Q$ dans les formules (3.17). Les propriétés de l'équation discrète (3.20) induites par cette définition sont les suivantes.

Proposition 3.5.

- (i) les écoulements uniformes sont préservés exactement;
- (ii) les équations de conservation de ρ , ρu_x , E sont vérifiées exactement (Eq. (3.13a), (3.13b), et (3.13d));
- (iii) l'équation d'évolution de ρu_r est vérifiée exactement (Eq. (3.13c)).

Preuve. Il suffit de noter que D vérifie

$$D(\cos \omega)_q = -\sin \omega_q, \quad D(\sin \omega)_q = \cos \omega_q,$$

et que $D^* = -D$. Ainsi la propriété (3.21e) est vérifiée, ce qui implique (i). En outre, on obtient la propriété (3.21b) car

$$\sum_{q=0}^Q \cos \omega_q D(\sin \omega f)_q = \sum_{q=0}^Q D^*(\cos \omega) \sin \omega_q f_q = -\sum_{q=0}^Q \sin \omega_q f_q,$$

ce qui implique (iii). Enfin, la forme symétrique de D et la périodicité de f impliquent la propriété (3.21a), donc (ii). \square

Cependant, cette méthode possède l'important défaut de ne pas garantir la positivité (ni a fortiori la propriété d'entropie), ce qui est dû à la discrétisation centrée. Remarquons enfin que le même opérateur utilisé sur l'équation non conservative donne les mêmes propriétés, sauf (iii), car la relation (3.19b) n'est vérifiée qu'à $O(\Delta\omega)$ près. En effet, un simple calcul donne $D(\cos \omega \sin \omega)_q = -\cos^2 \omega_q + \sin^2 \omega_q + O(\Delta\omega)$.

Remarque 3.3. Il est clair que la discrétisation centrée sans modification trigonométrique (*i.e.* $D(f)_q = \frac{f_{q+1} - f_{q-1}}{2\Delta\omega}$) ne vérifie aucune des propriétés de conservation. Elle vérifie seulement la propriété de préservation des écoulements uniformes.

3.3.4 Discrétisation décentrée "trigonométrique" de l'équation conservative : conservation et entropie

Nous proposons une modification décentrée de l'opérateur précédent pour obtenir la positivité :

$$D(\sin \omega f)_q = \frac{1}{2 \sin \frac{\Delta \omega}{2}} \left((\sin \omega_{q+\frac{1}{2}}^+ f_{q+1} + \sin \omega_{q+\frac{1}{2}}^- f_q) - (\sin \omega_{q-\frac{1}{2}}^+ f_q + \sin \omega_{q-\frac{1}{2}}^- f_{q-1}) \right),$$

où l'on a noté $a^+ = \frac{1}{2}(a + |a|)$, $a^- = \frac{1}{2}(a - |a|)$, et $\omega_{q\pm\frac{1}{2}} = \omega_q \pm \frac{\Delta \omega}{2}$. L'intervalle $[0, 2\pi]$ est discrétisé comme précédemment, ainsi que les moments de f , et on choisit $Q = 2m - 1$ de sorte que $\omega_m = \pi$. Les propriétés de l'équation discrète (3.20) sont alors les suivantes.

Proposition 3.6. (i) *les écoulements uniformes sont préservés exactement;*

(ii) *les équations de conservation de ρ , ρu_x , E sont satisfaites exactement (Eq. (3.13a), (3.13b), et (3.13d));*

(iii) *la positivité de f est préservée;*

(iv) *l'entropie totale est localement dissipée (Eq. (3.14)).*

Preuve. La propriété (i) est obtenue en remarquant que (3.21e) est vérifiée

$$D(\sin \omega)_q = \frac{1}{2 \sin \frac{\Delta \omega}{2}} (\sin \omega_{q+\frac{1}{2}} - \sin \omega_{q-\frac{1}{2}}) = \cos \omega_q.$$

Pour (ii), on remarque que $D(\sin \omega f)_q$ s'écrit sous forme d'un flux numérique $D(\sin \omega f)_q = h_{q+\frac{1}{2}} - h_{q-\frac{1}{2}}$, donc (3.21a) est évidente. La propriété (iii) est due au caractère décentré de la discrétisation. En effet

$$D(\sin \omega f)_q \geq \frac{1}{2 \sin \frac{\Delta \omega}{2}} (\sin \omega_{q+\frac{1}{2}}^- - \sin \omega_{q-\frac{1}{2}}^+) f_q,$$

donc en développant $\partial_r r f_q$ dans (3.20) et en divisant par r , on obtient l'inégalité

$$\partial_t f_q + v_x \partial_x f_q + \zeta \cos \omega_q \partial_r f_q \geq \frac{\zeta}{r} C_q f_q,$$

où $C_q = \frac{1}{2 \sin \frac{\Delta \omega}{2}} (\sin \omega_{q+\frac{1}{2}}^- - \sin \omega_{q-\frac{1}{2}}^+) - \cos \omega_q$. Il suffit donc de montrer que ce coefficient est toujours négatif ou nul pour conclure. Or on a

$$C_q = \begin{cases} -1 & \text{si } q = 0, \\ -\sin \omega_q \frac{\cos \frac{\Delta \omega}{2}}{2 \sin \frac{\Delta \omega}{2}} - \frac{1}{2} \cos \omega_q & \text{si } 0 < q < m, \\ 0 & \text{si } q = m, \\ \sin \omega_q \frac{\cos \frac{\Delta \omega}{2}}{2 \sin \frac{\Delta \omega}{2}} - \frac{1}{2} \cos \omega_q & \text{si } m < q < Q. \end{cases}$$

On voit donc qu'il suffit de montrer que $-\sin \omega_q \frac{\cos \frac{\Delta \omega}{2}}{2 \sin \frac{\Delta \omega}{2}} - \frac{1}{2} \cos \omega_q \leq 0$ pour $0 < q < m$, ce qui est évident si $0 < \omega_q < \frac{\pi}{2}$ car $\sin \omega_q$ et $\cos \omega_q$ sont alors tous deux positifs. Pour $\frac{\pi}{2} < \omega_q < \pi$, cette inégalité est équivalente à $-\tan \omega_q \geq \tan \frac{\Delta \omega}{2}$, ce qui est nécessairement vrai.

La propriété la plus remarquable de cette discrétisation est la dissipation d'entropie. On peut en effet montrer que (3.21d) est vérifiée : par un changement d'indice, il vient

$$\begin{aligned} \sum_{q=0}^Q D(\sin \omega f)_q \log f_q &= \sum_{q=0}^Q \frac{1}{2 \sin \frac{\Delta\omega}{2}} \left(\sin \omega_{q-\frac{1}{2}}^+ \log f_{q-1} + \sin \omega_{q+\frac{1}{2}}^- \log f_q - \right. \\ &\quad \left. \sin \omega_{q-\frac{1}{2}}^+ \log f_q - \sin \omega_{q+\frac{1}{2}}^- \log f_{q+1} \right) f_q. \end{aligned}$$

On utilise alors l'inégalité de convexité $t_2 \log t_1 \leq t_2 \log t_2 + t_1 - t_2$ pour les termes $f_q \log f_{q\pm 1}$; les logarithmes s'annulent et on obtient

$$\sum_{q=0}^Q D(\sin \omega f)_q \log f_q \leq \sum_{q=0}^Q \frac{1}{2 \sin \frac{\Delta\omega}{2}} \left(\sin \omega_{q-\frac{1}{2}}^+ (f_{q-1} - f_q) - \sin \omega_{q+\frac{1}{2}}^- (f_{q+1} - f_q) \right).$$

Par un nouveau changement d'indice, on trouve

$$\begin{aligned} \sum_{q=0}^Q D(\sin \omega f)_q \log f_q &\leq \sum_{q=0}^Q \frac{1}{2 \sin \frac{\Delta\omega}{2}} \left(\sin \omega_{q+\frac{1}{2}}^+ - \sin \omega_{q-\frac{1}{2}}^+ - \sin \omega_{q-\frac{1}{2}}^- + \sin \omega_{q+\frac{1}{2}}^- \right) f_q \\ &\leq \sum_{q=0}^Q \frac{1}{2 \sin \frac{\Delta\omega}{2}} \left(\sin \omega_{q+\frac{1}{2}} - \sin \omega_{q-\frac{1}{2}} \right) f_q \\ &= \sum_{q=0}^Q \cos \omega_q f_q, \end{aligned}$$

on a bien (3.21d), ce qui implique (iv) (cf. proposition 3.2). \square

Remarque 3.4. A notre connaissance, c'est la première fois qu'est présentée une discrétisation possédant simultanément les propriétés de positivité, de dissipation d'entropie, et de conservation de ρ , ρu_x et E . Notons de plus que si l'on prend la discrétisation décentrée d'ordre 1 classique sans modification trigonométrique (*i.e.* avec $\Delta\omega$ à la place de $2 \sin \frac{\Delta\omega}{2}$), alors on a seulement $D(\sin \omega)_q = \cos \omega_q + O(\Delta\omega)$, et seules les propriétés (ii) et (iii) sont vérifiées.

Remarque 3.5. L'équation d'évolution de ρu_r (3.13c) n'est obtenue qu'à $\Delta\omega$ près car on a

$$\begin{aligned} \partial_t r \rho u_r + \partial_x r (\rho u_x u_r + \rho \Theta_{xr}) + \partial_r r (\rho u_r^2 + \rho \Theta_{rr}) &= \rho \widetilde{\Theta}_{\varphi\varphi} \\ &= \rho \Theta_{\varphi\varphi} + O(\Delta\omega), \end{aligned}$$

où $\rho \widetilde{\Theta}_{\varphi\varphi} = \langle \sum_q \overline{\sin^2 \omega_q f_q} \rangle$ et $\overline{\sin^2 \omega_q} = \sin^2 \omega_q + O(\Delta\omega)$. On voit donc qu'on aurait pu obtenir l'équation d'évolution de ρu_r (3.13c) exactement en modifiant la définition de $\rho \Theta_{\varphi\varphi}$, *i.e.* en posant $\rho \Theta_{\varphi\varphi} = \rho \widetilde{\Theta}_{\varphi\varphi}$. Mais on constate alors que l'équation d'énergie (3.13d) n'est plus vérifiée qu'à $\Delta\omega$ près car E est à présent définie par $E = \langle \sum_q \frac{1}{2} (v_x^2 + \zeta^2 \cos^2 \omega_q + \zeta^2 \overline{\sin^2 \omega_q}) f_q \rangle = \langle \sum_q \frac{1}{2} (v_x^2 + \zeta^2) f_q \rangle + O(\Delta\omega)$. Cependant, il nous semble préférable de vérifier exactement l'équation d'énergie car c'est une loi de conservation, contrairement à l'équation d'évolution de ρu_r (3.13c).

3.4 Approximation de la vitesse radiale

Dans cette section, nous utilisons une idée de Larina et Rykov [33] pour obtenir une autre discrétisation ayant de bonnes propriétés. Leur idée peut être reformulée de la façon suivante : si

l'on utilise l'équation conservative, on voit que pour obtenir les propriétés (3.21), on peut tout autant jouer sur des approximations de $\cos \omega_q$ que sur $D(\sin \omega f)_q$. On prend alors un opérateur D classique, et on cherche une approximation $\overline{\cos} \omega_q$ de $\cos \omega_q$ telle que ces propriétés soient vérifiées. On peut en fait interpréter cette méthode comme un choix de vitesses radiales discrètes différent à $O(\Delta\omega^2)$ près des vitesses initiales, de façon à assurer certaines propriétés.

Cependant, il faut aussi définir les quantités macroscopiques de façon cohérente. Ainsi, considérons la discrétisation de l'équation conservative

$$\partial_t r f_q + v_x \partial_x r f_q + \zeta \overline{\cos} \omega_q \partial_r r f_q - \zeta D(\sin \omega f)_q = 0, \quad (3.24)$$

ou celle de l'équation non conservative

$$\partial_t f_q + v_x \partial_x f_q + \zeta \overline{\cos} \omega_q \partial_r f_q - \frac{\zeta \sin \omega_q}{r} D(f)_q = 0. \quad (3.25)$$

On voit que pour avoir les équations de conservation de ρ , ρu_x , et de E , il faut que l'on adopte les définitions suivantes

$$\rho u_r = \left\langle \sum_q \zeta \overline{\cos} \omega_q f_q \right\rangle, \quad (3.26a)$$

$$\rho \Theta_{rr} = \left\langle \sum_q (\zeta \overline{\cos} \omega_q - u_r)^2 f_q \right\rangle, \quad \rho \Theta_{xr} = \left\langle \sum_q (v_x - u_x) (\zeta \overline{\cos} \omega_q - u_r) f_q \right\rangle, \quad (3.26b)$$

$$q_x = \left\langle \sum_q \frac{1}{2} ((v_x - u_x)^2 + (\zeta \overline{\cos} \omega_q - u_r)^2 + \zeta^2 \sin^2 \omega_q) (v_x - u_x) f_q \right\rangle, \quad (3.26c)$$

$$q_r = \left\langle \sum_q \frac{1}{2} ((v_x - u_x)^2 + (\zeta \overline{\cos} \omega_q - u_r)^2 + \zeta^2 \sin^2 \omega_q) (\zeta \overline{\cos} \omega_q - u_r) f_q \right\rangle. \quad (3.26d)$$

On peut noter que le terme de pression $\rho \Theta_{\varphi\varphi}$ tel que $E = \frac{1}{2} \rho (u_x^2 + u_r^2) + \frac{1}{2} \rho (\Theta_{xx} + \Theta_{rr} + \Theta_{\varphi\varphi})$ vaut alors $\rho \Theta_{\varphi\varphi} = \left\langle \sum_q \zeta^2 (1 - \overline{\cos} \omega_q^2) f_q \right\rangle$. Voyons à présent le choix de discrétisation de Larina et Rykov, puis une modification que nous proposons.

3.4.1 Discrétisation centrée de l'équation conservative

Dans [33], Larina et Rykov ont proposé une discrétisation centrée d'ordre 2 de l'équation conservative (3.7). Leur méthode est relativement complexe car elle est basée sur une décomposition de f en parties paires et impaires par rapport à v_x et ω . Cependant, leur principe de discrétisation en ω peut être formulé de la façon suivante. L'opérateur D est défini par

$$D(f)_q = \frac{f_{q+1} - f_{q-1}}{2\Delta\omega}, \quad (3.27)$$

et on pose

$$\overline{\cos} \omega_q = D(\sin \omega)_q. \quad (3.28)$$

L'intervalle $[0, 2\pi]$ est discrétisé par $\omega_q = q\Delta\omega$ avec $q = 0..Q$ et $\omega_Q = 2\pi - \Delta\omega$. Les moments de f sont calculés par les formules (3.17) en remplaçant \sum_q par $\sum_{q=0}^Q$. L'équation obtenue est alors l'équation (3.24). Dans le schéma original de Larina et Rykov, l'opérateur D est en fait défini par $D(f)_q = \frac{1}{\Delta\omega} (f_{q+\frac{1}{2}} - f_{q-\frac{1}{2}})$, mais cela ne change rien aux propriétés de leur discrétisation, que nous rappelons ci-dessous.

Proposition 3.7. (i) *les écoulements uniformes en t, x, r, q sont préservés exactement;*

(ii) si les quantités macroscopiques sont définies par (3.26), alors les équations de conservation de ρ , ρu_x , E sont vérifiées exactement (Eq. (3.13a), (3.13b), (3.13d)).

Preuve. La propriété (i) est immédiate d'après la définition (3.28) de $\overline{\cos} \omega_q$. Quant à la propriété (ii), elle se prouve comme dans les sections 3.3.3 et 3.3.4 en notant que (3.21a) est vérifiée exactement. \square

On peut cependant noter que comme dans la section 3.3.3, cette méthode présente l'inconvénient de ne pas garantir la positivité de f (ni a fortiori une propriété d'entropie) en raison de la discrétisation centrée de ∂_ω . Pourtant ce défaut ne semble pas affecter les résultats numériques présentés par Larina et Rykov dans [33] et [32].

Remarque 3.6. En remplaçant $\Delta\omega$ par $\sin \Delta\omega$ dans la définition de D (3.27), on retrouve le schéma centré proposé dans la section 3.3.3, puisque alors $\overline{\cos} \omega_q = \cos \omega_q$.

Remarque 3.7. L'approximation $\overline{\cos} \omega_q$ de $\cos \omega_q$ peut s'interpréter comme un changement de vitesses discrètes radiales. En effet, après discrétisation de $[0, 2\pi]$ en $\{\omega_q\}_{q=0}^Q$, les vitesses discrètes radiales et azimuthales sont

$$v_r^q = \zeta \cos \omega_q, \quad v_\varphi^q = \zeta \sin \omega_q.$$

Avec la méthode de Larina et Rykov, la fonction de distribution $\{f_q\}$ est convectée par les vitesses radiales et azimuthales

$$\overline{v}_r^q = \zeta \overline{\cos} \omega_q, \quad v_\varphi^q = \zeta \sin \omega_q,$$

où $\overline{\cos} \omega_q = D(\sin \omega)_q = \cos \omega_q \frac{\sin \Delta\omega}{\Delta\omega}$, ce qui implique $\overline{v}_r^q = \frac{\sin \Delta\omega}{\Delta\omega} v_r^q$. En supposant que la variable ζ est bornée et discrète ($\zeta = (\zeta_l)_{l=0}^L$), on peut représenter les vitesses discrètes originelles et les vitesses discrètes modifiées sur la figure 3.1.

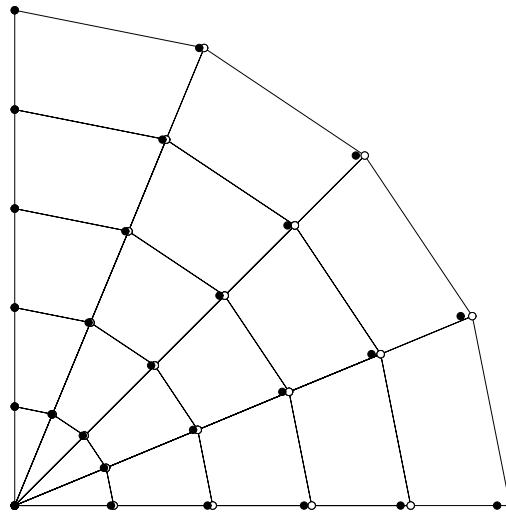


FIG. 3.1 – vitesses radiales et azimuthales pour $q = 0.4$ et $\Delta\omega = \frac{\pi}{8}$: $\circ (v_r^q, v_\varphi^q)$, $\bullet (\overline{v}_r^q, v_\varphi^q)$.

3.4.2 Discrétisation décentrée de l'équation non conservative

Nous proposons une méthode s'inspirant de l'idée précédente de Larina et Rykov. On peut l'interpréter comme une amélioration de la méthode classique exposée section 3.3.1. Elle permet

d'avoir une discrétisation possédant les mêmes propriétés que celle exposée section 3.3.4. L'opérateur D est défini par (3.22) sur $[0, \pi]$, et on pose

$$\overline{\cos} \omega_q = -D^*(\sin \omega)_q.$$

L'équation obtenue est alors l'équation (3.25), qui vérifie les propriétés suivantes.

Proposition 3.8. (i) *la positivité de f est préservée;*

(ii) *les écoulements uniformes en t, x, r, q sont solutions;*

(iii) *si les quantités macroscopiques sont définies par (3.26) (avec $2 \sum_{q=1}^Q$ au lieu de \sum_q), alors les équations de conservation de ρ , ρu_x , E sont vérifiées exactement (Eq. (3.13a), (3.13b), (3.13d));*

(iv) *l'entropie est localement dissipée (Eq. (3.14))*

Preuve. La propriété (i) est obtenue comme dans la section 3.3.1, ainsi que la propriété (ii). Pour la propriété (iii), il suffit de remarquer que la définition de $\overline{\cos} \omega_q$ implique

$$\sum_q \sin \omega_q D(f)_q = - \sum_q \overline{\cos} \omega_q f_q.$$

Cette relation est analogue à (3.19a) et permet de montrer la conservation de ρ , ρu_x , et E . Pour la propriété d'entropie (iv), une technique identique à celle utilisée section 3.3.4 donne le résultat. \square

3.5 Conclusion

Nous avons montré que la plupart des schémas utilisés actuellement pour calculer des écoulements axisymétriques ne vérifient pas certaines propriétés importantes. Les schémas décentrés de Shakhov et Sone (section 3.3.1) ne vérifient pas les propriétés de conservation, et le schéma centré d'ordre deux de Larina et Rykov (section 3.4.1) ne garantit pas la positivité de la fonction de distribution.

Parmi les différents schémas que nous avons introduits pour corriger ces défauts, trois paraissent prometteurs. Le schéma trigonométrique d'ordre deux (section 3.3.3) vérifie toutes les lois de conservation, ainsi que la loi d'évolution de ρu_r et la préservation des écoulements uniformes. Cependant, la positivité de f n'est pas préservée, et on peut se demander si l'approximation centrée ne pourrait pas être source d'instabilités. Le schéma décentré d'ordre un (section 3.3.4) vérifie toutes les lois de conservation, la préservation des écoulements uniformes, ainsi que la positivité de f . De plus, il vérifie une loi de dissipation d'entropie. Enfin le schéma décentré d'ordre un sur l'équation non conservative (section 3.4.2) vérifie lui aussi toutes ces propriétés, même s'il est moins élégant. En revanche, les autres schémas que nous avons introduits présentent tous le défaut de ne pas préserver les écoulements constants.

Il est néanmoins nécessaire de compléter cette étude par des essais numériques afin de mieux cerner les défauts et avantages de chaque schéma. La section 6.3 du chapitre 6 apporte quelques éléments de comparaison.

Chapitre 4

Discrétisation de termes de collision BGK généralisés

4.1 Nombre de Prandtl pour l'équation BGK

Lorsqu'on utilise l'équation de Boltzmann, on dispose d'un degré de liberté dans le modèle qui est le potentiel d'interaction. Suivant que l'on veut faire des calculs analytiques simples, ou qu'on veut approcher au mieux le comportement physique d'un gaz, on prend alors le potentiel maxwellien ou des sphères dures, ou un modèle plus complexe. De façon analogue, le paramètre de l'équation BGK est le temps de relaxation τ . Pour déterminer ce paramètre de façon à décrire au mieux la physique, la procédure communément utilisée est la méthode de Chapman-Enskog. On considère un développement asymptotique de f en puissance de τ autour de l'équilibre local $M[\boldsymbol{\rho}^\tau]$, en supposant que τ est petit :

$$f^\tau = M[\boldsymbol{\rho}^\tau] + \tau f_1 + \tau^2 f_2 + \dots$$

Ce développement n'est en fait pas un développement asymptotique à proprement parler car on ne développe pas les moments de f^τ . En particulier, le premier terme de ce développement est la maxwellienne construite sur $\boldsymbol{\rho}^\tau$. On tronque alors ce développement à l'ordre 2, et on l'injecte dans l'équation BGK (1.8). On obtient alors les équations de conservation

$$\partial_t \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix} + \nabla \cdot \begin{pmatrix} \rho u \\ \rho u \otimes u + pI \\ Eu + p \end{pmatrix} = \nabla \cdot \begin{pmatrix} 0 \\ -\Sigma \\ \Sigma u - q \end{pmatrix},$$

où $\Sigma = \langle (v - u) \otimes (v - u) f \rangle - pI$ est le tenseur de contraintes de cisaillement, dont l'expression à l'ordre 2 est

$$\Sigma = -\mu(\nabla u + \nabla u^T - \frac{2}{3}\nabla \cdot u) + O(\tau^2)$$

et $q = \langle \frac{1}{2}|v - u|^2(v - u) f \rangle$ est le flux de chaleur, qui vaut

$$q = -\kappa \nabla T + O(\tau^2).$$

Ces équations de conservation sont donc les équations de Navier-Stokes à $O(\tau^2)$ près. Les coefficients μ et κ sont respectivement les coefficients de viscosité et de transfert thermique, qui sont reliés à τ par les relations

$$\mu = p\tau \quad \text{et} \quad \kappa = \frac{5R}{2}p\tau. \quad (4.1)$$

Le choix d'un temps de relaxation τ revient donc à choisir une viscosité μ ou un coefficient de transfert thermique κ . Généralement, on impose $\mu = \mu_{Btz}$, où

$$\mu_{Btz} = \mu_{ref} \left(\frac{T}{T_{ref}} \right)^\omega \quad (4.2)$$

est le coefficient de viscosité obtenu par développement de Chapman-Enskog sur l'équation de Boltzmann après le choix d'un potentiel de collision, et μ_{ref} est la viscosité de référence à la température de référence T_{ref} . Ainsi, le modèle HS donne $\omega = \frac{1}{2}$, et le modèle Maxwellien donne $\omega = 1$. Pour un potentiel plus réaliste comme le potentiel VHS (introduit par Bird pour DSMC), ω est déterminé au mieux pour chaque gaz par méthode expérimentale ou par une simulation numérique (cf. [5]). Ainsi, on s'assure que proche de l'équilibre, l'équation BGK donne la bonne viscosité.

Cependant, on constate d'après (4.1) que le coefficient de transfert thermique κ est alors fixé à la valeur

$$\kappa = \frac{5}{2} R \mu,$$

ce qui implique que le modèle BGK ne laisse pas la possibilité d'adapter à la fois κ et μ au gaz considéré¹. On traduit habituellement ce phénomène par le fait que le nombre de Prandtl défini par

$$Pr = \frac{5}{2} \frac{R \mu}{\kappa} \quad (4.3)$$

est toujours égal à 1 avec le modèle BGK. Non seulement ce nombre de Prandtl est nécessairement fixé une fois pour toute dans le modèle, mais en plus il est très différent des valeurs que l'on obtient dans l'équation de Boltzmann avec certains potentiels de collision. Par exemple, avec le modèle HS, on trouve $Pr = \frac{2}{3}$ pour un gaz monoatomique, ce qui est assez proche des résultats expérimentaux. En général, pour la plupart des gaz, ces résultats donnent $Pr < 1$. Une conséquence physique de $Pr = 1$ au lieu de $Pr < 1$ est que les phénomènes de transfert thermiques (qui se font à la "vitesse" κ) sont trop lents par rapport aux phénomènes de transfert de quantité de mouvement (qui se font à la "vitesse" μ).

L'idée évidente pour pallier cette rigidité du modèle BGK est de rajouter un paramètre, de façon à pouvoir fixer indépendamment μ et κ . Beaucoup de modèles de ce type ont été proposés (voir les références dans les sections suivantes), mais peu respectent les trois propriétés fondamentales de donner des fonctions de distribution positives, d'avoir un nombre de Prandtl inférieur à 1, et de vérifier le H-théorème $\langle Q(f) \log f \rangle \leq 0$.

Dans la section 4.2, nous décrivons le modèle ES proposé par Holway [28] qui vérifie toutes ces propriétés (dont la dernière a été récemment prouvée par Andries-Le Tallec-Perthame [2]). Suivant l'approche exposée dans le chapitre 1, nous étudions sa discrétisation en vitesse. Cette discrétisation est possible, mais les bases mathématiques sont beaucoup plus difficile à obtenir que pour BGK. Dans la section 4.3, nous étudions deux modèles à deux temps de relaxation. Le premier a été proposé par Levermore, mais il a l'inconvénient de ne pas respecter la positivité si $Pr < 1$. Le deuxième modèle, que nous introduisons ici, vérifie toutes ces propriétés, mais semble pratiquement peu utilisable. Dans la section 4.4, nous présentons très succinctement les autres modèles existants.

1. Le choix de fixer $\mu = \mu_{Btz}$ est bien sûr arbitraire, et l'on pourrait tout autant fixer $\kappa = \kappa_{Btz}$, mais alors il n'y a plus de raison que $\mu = \mu_{Btz}$.

4.2 Modèle ES

4.2.1 Modèle continu

Ce modèle a été initialement proposé par Holway [28], puis étudié par Andries-Le Tallec-Perthame [2] qui ont prouvé certaines propriétés. L'idée consiste à remplacer la maxwellienne de BGK par une gaussienne non isotrope. Nous avons pour cela besoin des définitions suivantes

$$\begin{aligned}\rho\Theta &= \langle (v-u) \otimes (v-u)f \rangle \quad (\text{tenseur des contraintes}) \\ \theta &= \frac{1}{D}\text{Tr}(\Theta) = RT \quad (\text{température normalisée}).\end{aligned}$$

Introduisant alors un paramètre ν , on définit le tenseur

$$\mathcal{T} = (1-\nu)\theta I + \nu\Theta, \quad (4.4)$$

où I est le tenseur identité. Définissons aussi les dix premiers moments de la fonction de distribution f

$$\begin{aligned}\rho_{10} &= \langle \mathbf{m}_{10}f \rangle \\ &= (\rho, \rho u, \rho u \otimes u + \rho\Theta)^T,\end{aligned}$$

avec $\mathbf{m}_{10} = (1, v, v \otimes v)^T$ (nous considérons ici le cas particulier $D = 3$, ce qui donne 10 moments indépendants). De même, le vecteur de moments correspondant à \mathcal{T}

$$\rho_{10}^\nu = (\rho, \rho u, \rho u \otimes u + \rho\mathcal{T})^T.$$

La maxwellienne de l'opérateur de collision BGK est alors remplacée par la gaussienne

$$G[\rho_{10}^\nu] = \frac{\rho}{\sqrt{\det(2\pi\mathcal{T})}} \exp\left(-\frac{1}{2}(v-u)^T \mathcal{T}^{-1}(v-u)\right), \quad (4.5)$$

et le développement de Chapman-Enskog sur ce modèle donne alors

$$\mu = \frac{p\tau}{1-\nu} \quad \text{et} \quad \kappa = \frac{5}{2}Rp\tau,$$

ce qui donne la relation entre la paramètre ν et le nombre de Prandtl

$$Pr = \frac{1}{1-\nu}.$$

Le modèle permet donc d'obtenir n'importe quel nombre de Prandtl en ajustant ν . Remarquons que pour $\nu = 0$, on a $Pr = 1$, et on retrouve le modèle de BGK, car alors $\mathcal{T} = \theta I$ et donc $G[\rho_{10}^\nu] = M[\rho]$ (où M est la maxwellienne associée à f (cf. chap. 1)). De plus pour $\nu = -\frac{1}{2}$, on obtient $Pr = \frac{2}{3}$, ce qui est la valeur expérimentale pour les gaz monoatomiques.

Cependant, le modèle n'est en fait défini que pour un petit intervalle de ν , en dehors duquel \mathcal{T} n'est plus défini positif. Assez miraculeusement, la borne inférieure de cet intervalle est $\nu = -\frac{1}{2}$, qui est celle de la valeur de Pr que l'on veut obtenir ! Plus précisément, on donne à présent les propriétés prouvées dans [2] sur ce modèle. Introduisons le problème de minimisation d'entropie suivant

$$(\mathcal{P}^\nu) \quad \min\{H(g), \quad g \in \mathcal{X}_{\rho_{10}^\nu}\}, \quad \text{où } \mathcal{X}_{\rho_{10}^\nu} = \{g \geq 0, \langle \mathbf{m}_{10}g \rangle = \rho_{10}^\nu\}. \quad (4.6)$$

On a alors la proposition suivante

Proposition 4.1 (Andriès et al). *Pour tout tenseur défini positif Θ et $-\frac{1}{2} \leq \nu \leq 1$, on a*

(i) *le tenseur \mathcal{T} défini par (4.4) est défini positif et $\mathcal{X}_{\rho_{10}^\nu}$ est non vide ;*

- (ii) le problème (\mathcal{P}^ν) admet la gaussienne $G[\rho_{10}^\nu]$ comme unique solution ;
- (iii) l'entropie de la gaussienne satisfait $H(G) \leq H(f)$;
- (iv) l'égalité $H(G) = H(f)$ implique $f = M[\rho]$.

Cette proposition montre donc que le modèle ES vérifie toutes les propriétés nécessaires.

Remarque 4.1. Contrairement au cas du modèle BGK, les propriétés (iii) et (iv) ne sont pas évidentes ici car f ne vérifie pas les contraintes du problème de minimisation (*i.e.* f n'appartient pas à $\mathcal{X}_{\rho_{10}^\nu}$).

Remarque 4.2. Andriès et al. ont aussi proposé dans [2] une extension de ce modèle aux gaz polyatomiques qui possède les mêmes propriétés.

4.2.2 Discrétisation

Comme au chapitre 1, nous considérons une grille de vitesses discrètes \mathcal{V} indicées sur \mathcal{K} (cf. 1.19). On approche la fonction de distribution f sur \mathcal{V} par le vecteur $f_{\mathcal{K}} = (f_k)_{k \in \mathcal{K}}$, et ses dix premiers moments par

$$\begin{aligned} \rho_{10} &= \langle \mathbf{m}_{10} f_{\mathcal{K}} \rangle_{\mathcal{K}} \\ &= (\rho_{\mathcal{K}}, \rho_{\mathcal{K}} u_{\mathcal{K}}, \rho_{\mathcal{K}} u_{\mathcal{K}} \otimes u_{\mathcal{K}} + \rho_{\mathcal{K}} \Theta_{\mathcal{K}}). \end{aligned}$$

On définit enfin la température par $\theta_{\mathcal{K}} = \frac{1}{D} Tr(\Theta_{\mathcal{K}}) = \frac{1}{\rho_{\mathcal{K}}} \langle |v - u_{\mathcal{K}}|^2 \rangle_{\mathcal{K}}$ et l'entropie discrète $H_{\mathcal{K}} = \langle f_{\mathcal{K}} \log f_{\mathcal{K}} \rangle_{\mathcal{K}}$. Par la suite, nous omettrons l'indice \mathcal{K} chaque fois qu'il n'y a pas d'ambiguïté.

Suivant l'approche introduite au chapitre 1, nous définissons une approximation $\mathcal{G}_{\mathcal{K}}$ de la gaussienne G par le problème de minimisation d'entropie discrète

$$(\mathcal{P}_{\mathcal{K}}^\nu) \quad \min\{H_{\mathcal{K}}(g), \quad g \in \mathcal{X}_{\rho_{10}^\nu}\}, \quad \text{où } \mathcal{X}_{\rho_{10}^\nu} = \{g \geq 0, \langle \mathbf{m}_{10} g \rangle_{\mathcal{K}} = \rho_{10}^\nu\}, \quad (4.7)$$

avec $\rho_{10}^\nu = (\rho, \rho u, \rho u \otimes u + \rho \mathcal{T})$ et \mathcal{T} est le tenseur modifié défini par

$$\mathcal{T} = (1 - \nu)\theta I + \nu\Theta. \quad (4.8)$$

Cependant, on remarque que ce modèle ne peut être entropique car on n'a pas la relation

$$\mathcal{G}_{\mathcal{K}} = f_{\mathcal{K}} \Rightarrow f_{\mathcal{K}} = \mathcal{E}_{\mathcal{K}},$$

où $\mathcal{E}_k = \exp(\boldsymbol{\alpha} \cdot \mathbf{m}(v_k))$ est l'équilibre discret (*i.e.* l'approximation de la maxwellienne définie au chapitre 1). En effet, la relation $\mathcal{G}_{\mathcal{K}} = f_{\mathcal{K}}$ implique $\Theta = \mathcal{T}$, et donc $\mathcal{T} = \theta I$. Mais contrairement au cas continu (où $\mathcal{T} = \theta I$ implique $G = M$, ce qui se voit aisément dans l'expression (4.5)), cette relation n'implique pas $\mathcal{G}_{\mathcal{K}} = \mathcal{E}_{\mathcal{K}}$, pour la simple raison que $\langle (v - u) \otimes (v - u) \mathcal{E}_{\mathcal{K}} \rangle_{\mathcal{K}}$ est *a priori* différent de $\rho\theta I$. On définit alors le tenseur des contraintes à l'équilibre

$$\rho\Pi = \langle (v - u) \otimes (v - u) \mathcal{E}_{\mathcal{K}} \rangle_{\mathcal{K}},$$

et on modifie la définition du tenseur \mathcal{T} par

$$\mathcal{T} = (1 - \nu)\Pi + \nu\Theta. \quad (4.9)$$

On a alors la proposition suivante

Proposition 4.2. *La relation $\mathcal{G}_{\mathcal{K}} = f_{\mathcal{K}}$ implique $f_{\mathcal{K}} = \mathcal{E}_{\mathcal{K}}$.*

Preuve. la relation $\mathcal{G}_\mathcal{K} = f_\mathcal{K}$ implique $\mathcal{T} = \Pi$, et donc $\mathcal{E}_\mathcal{K} \in \mathcal{X}_{\rho_{10}^\nu}$. Or par convexité de $t \log t$, on a

$$\begin{aligned} H_\mathcal{K}(\mathcal{E}_\mathcal{K}) &\leq H_\mathcal{K}(\mathcal{G}_\mathcal{K}) + \langle (\mathcal{E}_\mathcal{K} - \mathcal{G}_\mathcal{K})(1 + \log \mathcal{E}_\mathcal{K}) \rangle_\mathcal{K} \\ &= H_\mathcal{K}(\mathcal{G}_\mathcal{K}) + \langle (\mathcal{E}_\mathcal{K} - \mathcal{G}_\mathcal{K})\boldsymbol{\alpha} \cdot \mathbf{m} \rangle_\mathcal{K} \\ &= H_\mathcal{K}(\mathcal{G}_\mathcal{K}) + 0, \end{aligned}$$

donc $\mathcal{E}_\mathcal{K}$ est solution de $(\mathcal{P}_\mathcal{K}^\nu)$ et par conséquent $\mathcal{E}_\mathcal{K} = \mathcal{G}_\mathcal{K}$, d'où le résultat. \square

Pour montrer que le modèle est bien défini, il faut à présent vérifier que si ρ_{10} est strictement réalisable sur \mathcal{V} (ce qui est l'analogue discret de $\rho > 0$ et Θ défini positif) et que $-\frac{1}{2} \leq \nu \leq 1$, alors

- (i) $\mathcal{X}_{\rho_{10}^\nu}$ est non vide, et ρ_{10}^ν est strictement réalisable,
- (ii) $(\mathcal{P}_\mathcal{K}^\nu)$ admet comme unique solution $\mathcal{G}_\mathcal{K} = (\exp(\boldsymbol{\alpha}_{10} \cdot \mathbf{m}_{10}(v_k)))_{k \in \mathcal{K}}$,
- (iii) $H_\mathcal{K}(\mathcal{G}_\mathcal{K}) \leq H_\mathcal{K}(f_\mathcal{K})$
- (iv) $H_\mathcal{K}(\mathcal{G}_\mathcal{K}) = H_\mathcal{K}(f_\mathcal{K}) \Rightarrow f_\mathcal{K} = \mathcal{G}_\mathcal{K}$

Ces propriétés sont bien les analogues discrets des propriétés continues de la proposition 4.1.

Remarquons tout d'abord que si (i) est vraie, alors la propriété (ii) est une simple conséquence du théorème 1.1 généralisé aux moments d'ordre supérieur (voir remarque 1.2). Cependant, nous n'avons pu prouver la propriété (i) ainsi que (iii) et (iv), et il semble difficile de dire si elles sont vraies ou pas. Par exemple, on élimine rapidement l'idée de considérer la distribution discrète g définie par $g_k = (1 - \nu)\mathcal{E}_k + f_k$. En effet, elle vérifie $\langle \mathbf{m}_{10}g \rangle_\mathcal{K} = \rho_{10}^\nu$; il suffirait donc de montrer que $g_k > 0 \forall k$ pour obtenir (i). Malheureusement, ceci n'est vrai que si $0 \leq \nu \leq 1$, ce qui ne donne pas la gamme de nombres de Prandtl corrects.

Par ailleurs, nous avons étudié quelques exemples dans lesquels on peut calculer explicitement $\mathcal{G}_\mathcal{K}$, $\mathcal{E}_\mathcal{K}$, etc. (il suffit de prendre un nombre de vitesses égal au nombre de moments). Alors qu'on trouve facilement des cas pour lesquels la propriété (i) est fautive si on prend la définition initiale (4.8) de \mathcal{T} , nous n'avons pu mettre cette propriété en échec si l'on prend la bonne définition de \mathcal{T} . Ceci laisse à penser que la propriété (i) est vraie. Cependant, on peut se demander comment peuvent apparaître les bornes $-\frac{1}{2} \leq \nu \leq 1$ dans une démonstration. Dans le cas continu, elles apparaissent dans des expressions algébriques, telles que la diagonalisation de θI et Θ dans une même base (cf. [2]). Or ici, les deux tenseurs qui interviennent dans la définition de \mathcal{T} sont Π et Θ , qui *a priori* n'ont pas de vecteurs propres communs. On ne peut donc espérer appliquer la méthode de preuve de [2].

La structure de Π est d'ailleurs entièrement dépendante de la grille de vitesses choisie. On peut par exemple montrer que Π est diagonal si \mathcal{V} est centré en u et symétrique, mais dans un écoulement où u varie, cela ne peut être vrai. On comprend donc que si la propriété (i) est vraie sur un intervalle de ν , celui-ci dépend fortement de la grille des vitesses \mathcal{V} , et cette dépendance est sans doute assez complexe. Ces arguments sont probablement reliés au fait que les modèles à vitesses discrètes ne respectent pas l'invariance galiléenne.

Pour les propriétés (iii) et (iv), la preuve de [2] en continue utilise fortement l'expression analytique de G en fonction de ρ_{10}^ν , ce qui n'est pas possible ici.

En conclusion, le modèle ES discret assure que la fonction de distribution reste positive et que l'on obtient l'équilibre discret si $f_\mathcal{K} = \mathcal{G}_\mathcal{K}$, sous l'hypothèse que ρ_{10}^ν reste strictement réalisable pour l'intervalle $\nu \in [-\frac{1}{2}, 1[$, ce qui n'a pas été prouvé. En pratique (cf. chapitre 6), ce modèle a quand même été implémenté et testé numériquement avec $\nu = -\frac{1}{2}$. Il n'apparaît pas de problème de réalisabilité, et des améliorations par rapport à BGK sont visibles.

4.3 Modèles à deux temps de relaxation

4.3.1 Principe

Dans le modèle BGK, où le terme source vaut

$$Q_{BGK}(f) = \frac{1}{\tau}(M[f] - f),$$

la relaxation de f vers $M[f]$ implique

$$\begin{aligned} \rho\Theta &= \langle (v-u) \otimes (v-u)f \rangle \xrightarrow{t \rightarrow +\infty} \rho\theta I && \text{(relaxation vers l'équilibre non visqueux),} \\ q &= \langle \frac{1}{2}|v-u|^2(v-u)f \rangle \xrightarrow{t \rightarrow +\infty} 0 && \text{(relaxation vers l'équilibre thermique).} \end{aligned}$$

Ces deux relaxations se font à la même vitesse $\frac{1}{\tau}$, ce qui implique un nombre de Prandtl $Pr = 1$. L'idée des modèles à deux temps de relaxation est d'introduire un deuxième équilibre dans le terme source pour découpler ces deux phénomènes.

4.3.2 Modèle de Levermore

Dans son article sur les modèles aux moments [36], Levermore a proposé le terme de collision suivant

$$Q_L(f) = \frac{1}{\tau_1}(M[f] - f) + \frac{1}{\tau_2}(G[f] - f),$$

où $G[f]$ est la gaussienne qui a les mêmes dix premiers moments que f :

$$G[f] = \frac{\rho}{\sqrt{\det(2\pi\Theta)}} \exp\left(-\frac{1}{2}(v-u)^T \Theta^{-1}(v-u)\right),$$

Les distributions $G[f]$ et f ont même tenseur des contraintes, mais on peut voir que le flux de chaleur associé à $G[f]$ est nul. Donc le deuxième terme de relaxation implique $q \rightarrow 0$, c'est-à-dire une relaxation vers l'équilibre thermique à la vitesse $\frac{1}{\tau_2}$. Le premier terme traduit alors $\Theta \rightarrow \theta I$, c'est-à-dire une relaxation vers l'équilibre non visqueux à la vitesse $\frac{1}{\tau_1}$.

Un développement de Chapman-Enskog de ce modèle donne

$$\tau_1 = \frac{\mu}{p} \quad \text{et} \quad \tau_2 = \left(\frac{5Rp}{2\kappa} - \frac{p}{\mu}\right)^{-1} = \tau_1 \frac{1}{(Pr - 1)}, \quad (4.10)$$

avec la définition (4.3) du nombre de Prandtl, et où $p = \rho\theta = \langle \frac{1}{2}|v-u|^2 f \rangle$ est la pression. Ainsi, comme on l'a exposé dans la section 1, il suffit de fixer une viscosité convenable μ , et les relations (4.10) permettent d'obtenir le nombre de Prandtl que l'on souhaite, et donc un coefficient de transfert thermique κ adapté.

Cependant, on voit que $\tau_2 > 0$ si et seulement si $Pr > 1$, ce qui n'est pas dans le domaine qui nous intéresse. Autrement dit, ce modèle ne permet d'avoir des bons nombres de Prandtl que pour $\tau_2 < 0$, ce qui ne garantit pas la positivité de f^2 . On peut même montrer avec le contre-exemple suivant que ce modèle ne préserve pas la positivité de f : soit f la fonction (avec $D=2$)

$$f(v_1, v_2) = a\chi_{[-b,b] \times [-c,c]},$$

2. Dans la référence [36], le nombre de Prandtl est défini par $Pr = \frac{2}{5} \frac{\kappa}{R\mu}$ (i.e. l'inverse de la définition correcte), ce qui donne $\tau_2 > 0 \Leftrightarrow Pr < 1$.

telle que

$$\rho = 1, \quad \Theta_{11} = 1, \quad \Theta_{22} = 2,$$

(i.e. $a = \frac{1}{4\sqrt{18}}, b = \sqrt{3}, c = \sqrt{6}$). Alors, on peut trouver un ensemble de mesure non nulle Ω (par exemple $\Omega = \{(v_1, v_2) \text{ t.q. } |v_2| > \max(c, \sqrt{12 \ln(2\sqrt{2}) + 2v_1^2}) \text{ et } |v_1| > b\}$) tel que

$$\frac{1}{\tau_1}(M[f] - f) + \frac{1}{\tau_2}(G[f] - f) < 0 \quad \text{dans } \Omega$$

avec $Pr = \frac{2}{3}$, ce qui implique que la solution du problème homogène peut devenir négative.

Afin de modifier ce modèle, nous donnons à présent une interprétation physique du fait que $Pr > 1$ si $\tau_2 > 0$. Les relations (4.10) peuvent encore s'écrire

$$\frac{1}{\tau_1} = \frac{p}{\mu} \quad \text{et} \quad \frac{1}{\tau_2} = (Pr - 1) \frac{1}{\tau_1},$$

ce qui implique $\frac{1}{\tau_2} < \frac{1}{\tau_1}$ pour des nombres de Prandtl compris entre 1 et 2. Ceci signifie que pour $1 < Pr < 2$, la relaxation thermique (vitesse $\frac{1}{\tau_2}$) est plus lente que la relaxation visqueuse (vitesse $\frac{1}{\tau_1}$). Physiquement, on veut $Pr < 1$, donc la relaxation thermique doit être *plus rapide* que la relaxation visqueuse. Nous présentons alors dans la section suivante un modèle qui donne les bonnes vitesses de relaxation, en utilisant cette interprétation.

4.3.3 Modèle aux 14 moments

Pour corriger le modèle précédent, on remplace la gaussienne par une fonction de distribution $E[f]$ à déterminer. Le terme de collision est alors

$$\frac{1}{\tau_1}(M[f] - f) + \frac{1}{\tau_2}(E[f] - f).$$

On veut imposer que le deuxième terme de relaxation traduise à présent une relaxation visqueuse, c'est-à-dire $\Theta \rightarrow \theta I$. Il faut donc que $\langle (v - u) \otimes (v - u) E[f] \rangle = \rho \theta I$ et que le flux de chaleur de $E[f]$ soit égal à celui de f , i.e. $\langle \frac{1}{2} |v - u|^2 (v - u) E[f] \rangle = \langle \frac{1}{2} |v - u|^2 (v - u) f \rangle$. Ainsi, le premier terme de relaxation traduit la relaxation thermique $q \rightarrow 0$. On utilise donc la fonction d'équilibre généralisée aux 14 moments de Levermore [36] $E[f] = \exp(\alpha_{14}^E \cdot \mathbf{m}_{14}(v))$, définie par la relation

$$\langle \mathbf{m}_{14} E[f] \rangle = (\rho, \rho u, \rho u \otimes u + \rho \theta I, Q, S),$$

avec $\mathbf{m}_{14}(v) = (1, v, v \otimes v, \frac{1}{2} |v|^2 v, |v|^4)$, $Q = \langle \frac{1}{2} |v|^2 v f \rangle = uE + \rho \theta u + q$, et $S = \langle |v|^4 f \rangle$. Le terme en $\frac{1}{2} |v|^2 v$ permet d'imposer le flux de chaleur, et le terme en $|v|^4$ assure la convergence des intégrales en vitesse. On a alors le résultat suivant

Proposition 4.3. *Le développement de Chapman-Enskog de l'équation*

$$\partial_t f + v \cdot \nabla_x f = \frac{1}{\tau_1}(M[f] - f) + \frac{1}{\tau_2}(E[f] - f) \quad (4.11)$$

donne

$$\tau_1 = \frac{\mu}{p} \frac{1}{Pr} \quad \text{et} \quad \tau_2 = \tau_1 \frac{Pr}{1 - Pr}.$$

On vérifie bien que le modèle se comporte comme prévu, puisque $\tau_2 > 0$ si $Pr < 1$. De plus, avec $Pr = \frac{2}{3}$, on a $\frac{1}{\tau_2} = \frac{1}{2} \frac{1}{\tau_1} < \frac{1}{\tau_1}$; la relaxation visqueuse est donc bien plus lente que la relaxation thermique.

Preuve de la proposition. Le développement de Chapman-Enskog est ici moins simple que pour l'équation BGK, car on doit linéariser la fonction $E[f]$, qui n'est pas calculable analytiquement. Introduisons tout d'abord quelques définitions. On note la maxwellienne $M[f]$ associée à f

$$M[f] = \exp(\boldsymbol{\alpha} \cdot \mathbf{m}(v)) = \exp(\boldsymbol{\alpha}_{14}^M \cdot \mathbf{m}_{14}(v)),$$

où $\mathbf{m}(v) = (1, v, \frac{1}{2}|v|^2)$ (cf. chapitre 1). On note $\boldsymbol{\rho}_{14}$ les quatorze premiers moments de f

$$\boldsymbol{\rho}_{14} = \boldsymbol{\rho}_{14}[f] = (\rho, \rho u, \rho u \otimes u + \rho \Theta, Q, S) = \langle \mathbf{m}_{14} f \rangle,$$

où $Q = \langle \frac{1}{2}|v|^2 v f \rangle = uE + \rho \Theta u + q$ et $S = \langle |v|^4 f \rangle$. On note de façon analogue les quatorze premiers moments de $E[f]$

$$\boldsymbol{\rho}_{14}^E = \boldsymbol{\rho}_{14}^E[E] = (\rho, \rho u, \rho u \otimes u + \rho \theta I, Q, S) = \langle \mathbf{m}_{14} E[f] \rangle, \quad (4.12)$$

et on remarque que la seule différence entre $\boldsymbol{\rho}_{14}$ et $\boldsymbol{\rho}_{14}^E$ est le tenseur des contraintes. Notons enfin $\boldsymbol{\rho}_{14}^M$ les quatorze premiers moments de la maxwellienne $M[f]$

$$\boldsymbol{\rho}_{14}^M = \boldsymbol{\rho}_{14}^M[f] = (\rho, \rho u, \rho u \otimes u + \rho \theta I, Q^M, S^M) = \langle \mathbf{m}_{14} M[f] \rangle,$$

où, *a priori*, $Q^M \neq Q$ et $S^M \neq S$.

Dans l'équation (4.11), on suppose que $\tau_1 = \varepsilon$ est petit, et du même ordre que τ_2 . On pose donc $\nu_2 = \frac{\varepsilon}{\tau_2} = O(1)$, et l'équation (4.11) s'écrit

$$\partial_t f^\varepsilon + v \cdot \nabla_x f^\varepsilon = \frac{1}{\varepsilon} \left[(M[f^\varepsilon] - f^\varepsilon) + \nu_2 (E[f^\varepsilon] - f^\varepsilon) \right].$$

Le développement de Chapman-Enskog revient à développer f^ε en puissance de ε autour de l'équilibre local maxwellien $M[f^\varepsilon]$:

$$f^\varepsilon = M[f^\varepsilon] + \varepsilon f^1 + O(\varepsilon), \quad (4.13)$$

pour trouver une approximation à l'ordre deux du tenseur des contraintes visqueuses et du flux de chaleur

$$\begin{aligned} \Sigma &= \langle (v - u) \otimes (v - u) f^\varepsilon \rangle - \rho \theta I, \\ q &= \langle \frac{1}{2}|v - u|^2 (v - u) f^\varepsilon \rangle. \end{aligned} \quad (4.14)$$

Dans un premier temps, l'introduction de (4.13) dans (4.14) donne

$$\begin{aligned} \Sigma &= \varepsilon \langle (v - u) \otimes (v - u) f^1 \rangle + O(\varepsilon^2) = \varepsilon \Sigma^1 + O(\varepsilon^2), \\ q &= \varepsilon \langle \frac{1}{2}|v - u|^2 (v - u) f^1 \rangle + O(\varepsilon^2) = \varepsilon q^1 + O(\varepsilon^2). \end{aligned} \quad (4.15)$$

Il faut donc déterminer la perturbation f^1 . On introduit le développement (4.13) dans l'équation (4.11), et on trouve

$$\partial_t M[f^\varepsilon] + v \cdot \nabla_x M[f^\varepsilon] = -(1 + \nu_2) f^1 + \nu_2 \boldsymbol{\beta} \cdot \mathbf{m}_{14} M[f^\varepsilon] + O(\varepsilon), \quad (4.16)$$

où on a linéarisé $E[f^\varepsilon]$ autour de $M[f^\varepsilon]$:

$$E[f^\varepsilon] = M[f^\varepsilon](1 + \varepsilon \boldsymbol{\beta} \cdot \mathbf{m}_{14}) + O(\varepsilon^2)$$

en posant $\boldsymbol{\beta} = \frac{1}{\varepsilon}(\boldsymbol{\alpha}_{14}^E - \boldsymbol{\alpha}_{14}^M)$ qui est en $O(1)$. En effet, définissons l'application $F(\boldsymbol{\gamma}) = \langle \mathbf{m}_{14} \exp(\boldsymbol{\gamma} \cdot \mathbf{m}_{14}) \rangle$, alors la relation (4.12) implique $\boldsymbol{\alpha}_{14}^E = F^{-1}(\boldsymbol{\rho}_{14}^E)$. Or

$$\begin{aligned} \boldsymbol{\rho}_{14}^E &= \boldsymbol{\rho}_{14} + (0, 0, \rho\theta I - \rho\Theta, 0, 0) \\ &= \boldsymbol{\rho}_{14} - \varepsilon(0, 0, \Sigma^1, 0, 0) + O(\varepsilon^2) \end{aligned}$$

d'après (4.15). Notons $\boldsymbol{\rho}_{\Sigma^1} = (0, 0, \Sigma^1, 0, 0)$, on a alors

$$\begin{aligned} \boldsymbol{\alpha}_{14}^E &= F^{-1}(\boldsymbol{\rho}_{14} - \varepsilon \boldsymbol{\rho}_{\Sigma^1} + O(\varepsilon^2)) \\ &= F^{-1}(\boldsymbol{\rho}_{14}) - \varepsilon (F'(\boldsymbol{\rho}_{14}))^{-1} \boldsymbol{\rho}_{\Sigma^1} + O(\varepsilon^2). \end{aligned} \quad (4.17)$$

Or le développement (4.13) implique

$$\boldsymbol{\rho}_{14} = \boldsymbol{\rho}_{14}^M + \varepsilon \boldsymbol{\rho}_{14}[f^1] + O(\varepsilon^2), \quad (4.18)$$

donc d'après (4.17) et (4.18), on obtient

$$\boldsymbol{\alpha}_{14}^E = F^{-1}(\boldsymbol{\rho}_{14}^M) + \varepsilon (F'(\boldsymbol{\rho}_{14}^M))^{-1} (\boldsymbol{\rho}_{14}[f^1] - \boldsymbol{\rho}_{\Sigma^1}) + O(\varepsilon^2).$$

Or sachant que, par définition, $\boldsymbol{\alpha}_{14}^M = F^{-1}(\boldsymbol{\rho}_{14}^M)$, et notant la matrice $A_{14}^M = F'(\boldsymbol{\rho}_{14}^M) \in \mathbb{R}^{14 \times 14}$, on obtient donc

$$\boldsymbol{\alpha}_{14}^E = \boldsymbol{\alpha}_{14}^M + \varepsilon (A_{14}^M)^{-1} (\boldsymbol{\rho}_{14}[f^1] - \boldsymbol{\rho}_{\Sigma^1}) + O(\varepsilon^2),$$

d'où, comme annoncé,

$$\boldsymbol{\beta} = (A_{14}^M)^{-1} (\boldsymbol{\rho}_{14}[f^1] - \boldsymbol{\rho}_{\Sigma^1}) + O(\varepsilon). \quad (4.19)$$

Les relations (4.16) et (4.19) donnent donc

$$f^1 = -\frac{1}{1 + \nu_2} \left[\partial_t M[f^\varepsilon] + v \cdot \nabla_x M[f^\varepsilon] - \nu_2 (A_{14}^M)^{-1} (\boldsymbol{\rho}_{14}[f^1] - \boldsymbol{\rho}_{\Sigma^1}) \cdot \mathbf{m}_{14} M[f^\varepsilon] \right] + O(\varepsilon), \quad (4.20)$$

et il reste à déterminer $\boldsymbol{\rho}_{14}[f^1]$. On multiplie (4.20) par \mathbf{m}_{14} et on intègre en vitesse, il vient

$$\boldsymbol{\rho}_{14}[f^1] = -\langle \mathbf{m}_{14} (\partial_t M[f^\varepsilon] + v \cdot \nabla_x M[f^\varepsilon]) \rangle - \nu_2 \boldsymbol{\rho}_{\Sigma^1} + O(\varepsilon).$$

Par conséquent

$$\begin{aligned} f^1 &= -\frac{1}{1 + \nu_2} \left[\partial_t M[f^\varepsilon] + v \cdot \nabla_x M[f^\varepsilon] + \nu_2 (A_{14}^M)^{-1} \left(\langle \mathbf{m}_{14} (\partial_t M[f^\varepsilon] + v \cdot \nabla_x M[f^\varepsilon]) \rangle \right. \right. \\ &\quad \left. \left. + (1 + \nu_2) \boldsymbol{\rho}_{\Sigma^1} \right) \cdot \mathbf{m}_{14} M[f^\varepsilon] \right] + O(\varepsilon). \end{aligned}$$

On injecte alors cette formule dans (4.15), et on trouve

$$\begin{aligned} \Sigma_{i,j} &= -\varepsilon \frac{1}{1 + \nu_2} \left[\langle (v_i - u_i)(v_j - u_j) (\partial_t M[f^\varepsilon] + v \cdot \nabla_x M[f^\varepsilon]) \rangle \right. \\ &\quad \left. + \nu_2 (A_{14}^M)^{-1} (\langle \mathbf{m}_{14} (\partial_t M[f^\varepsilon] + v \cdot \nabla_x M[f^\varepsilon]) \rangle + (1 + \nu_2) \boldsymbol{\rho}_{\Sigma^1}) \right. \\ &\quad \left. \cdot \langle (v_i - u_i)(v_j - u_j) \mathbf{m}_{14} M[f^\varepsilon] \rangle \right] + O(\varepsilon^2), \end{aligned} \quad (4.21)$$

$$\begin{aligned}
q_i = -\varepsilon \frac{1}{1 + \nu_2} & \left[\left\langle \frac{1}{2} |v - u|^2 (v_i - u_i) (\partial_t M[f^\varepsilon] + v \cdot \nabla_x M[f^\varepsilon]) \right. \right. \\
& + \nu_2 (A_{14}^M)^{-1} (\langle \mathbf{m}_{14} (\partial_t M[f^\varepsilon] + v \cdot \nabla_x M[f^\varepsilon]) \rangle + (1 + \nu_2) \boldsymbol{\rho}_{\Sigma^1}) \\
& \left. \left. \cdot \left\langle \frac{1}{2} |v - u|^2 (v_i - u_i) \mathbf{m}_{14} M[f^\varepsilon] \right\rangle \right] + O(\varepsilon^2). \tag{4.22}
\end{aligned}$$

On peut alors calculer les différents moments de $(\partial_t M[f^\varepsilon] + v \cdot \nabla_x M[f^\varepsilon])$ en fonction de ρ , ρu , et E et de leurs gradients, car ces quantités vérifient les équations d'Euler à $O(\varepsilon)$ près, d'après (4.13). Ensuite, dans la relation (4.21), Σ apparaît à gauche et à droite (car $\boldsymbol{\rho}_{\Sigma^1} = (0, 0, \Sigma^1, 0, 0)$ et $\Sigma^1 = \frac{1}{\varepsilon} \Sigma$ d'après (4.15)), et on peut donc résoudre l'équation. On obtient donc l'expression de Σ en fonction du gradient de u , que l'on injecte dans (4.22) pour trouver q . Ces calculs longs et laborieux peuvent être effectués avec MAPLE³ et on trouve

$$\begin{aligned}
\Sigma &= -\mu (\nabla u + \nabla u^T - \frac{2}{3} \operatorname{div} u I), \\
q &= -\kappa \nabla \theta,
\end{aligned}$$

avec

$$\tau_1 = \frac{\mu}{p} \frac{1}{Pr} \quad \text{et} \quad \tau_2 = \tau_1 \frac{Pr}{1 - Pr}.$$

□

Remarque 4.3. Ce modèle n'est pourtant pas complètement satisfaisant, pour plusieurs raisons. Tout d'abord, comme l'a montré Junk [31] en étudiant les modèles aux moments de Levermore, les fonctions quatorze moments du type $E[f]$ ne sont pas toujours bien définies, notamment quand on est proche de l'équilibre. De plus, si on veut discrétiser ce modèle en vitesse, on se heurte à la même difficulté que pour le modèle ES, c'est-à-dire que $E[f]$ n'a pas les mêmes moments que f . Il est donc difficile de montrer que la stricte réalisabilité de $\boldsymbol{\rho}_{14}[f]$ implique celle de $\boldsymbol{\rho}_{14}[E]$. Enfin, au niveau numérique, le calcul de $E[f]$ par un algorithme de Newton nécessite l'inversion de matrices 14×14 contenant le produit $\mathbf{m}_{14} \otimes \mathbf{m}_{14}$ qui est très mal conditionné. En effet, le premier élément de ce tenseur est 1, alors que le dernier est $|v|^8$. Cependant, il pourrait être utile dans les modèles aux quatorze moments pour lesquels de tels calculs sont effectués (cf. [48, 15]).

4.4 Conclusion

Il ressort de cette étude que, de tous les modèles étudiés, seul le modèle ES semble être utilisable. La discrétisation manque cependant de bases mathématiques solides, et ce modèle mériterait d'être étudié de façon plus approfondie.

Même si les modèles à deux temps de relaxation ne semblent pas utilisables pratiquement, cette étude apporte quelques réponses sur les phénomènes physiques mis en jeu, qui pourraient servir au développement d'autres modèles.

D'autres voies ont été explorées par de nombreux auteurs, en particulier le développement de modèles perturbatifs où la maxwellienne est perturbée par un polynôme (cf. le S-modèle de Shakhov [59], ou le modèle de Liu [38]). Mais par nature, ces modèles ne garantissent pas la positivité

3. Les calculs sont assez lourds, notamment l'inversion de la matrice 14×14 A_{14}^M . Cependant, comme le modèle est galiléen (car $E[f]$ et $M[f]$ sont invariants par translation et rotation), on peut poser $u = 0$, ce qui simplifie énormément toutes les expressions.

de la distribution. Une autre voie possible consiste peut-être à étudier les modèles à temps de relaxation dépendant de la vitesse (voir par exemple le modèle de Bouchut et Perthame [8] ou de Struchtrup [61]), car le terme de collision est alors plus proche de celui de Boltzmann. C'est une approche qui semble fructueuse dans la modélisation de fluides complexes par les méthodes Lattice Boltzmann (cf. [25]), même si les modèles existants ne sont pas encore utilisables tels quels dans notre approche de discrétisation.

Chapitre 5

Méthodes numériques pour la résolution du modèle BGK à vitesses discrètes

Nous proposons dans ce chapitre des schémas numériques pour discrétiser en temps et en espace le modèle BGK à vitesses discrètes présenté au chapitre 1, le but essentiel étant de construire un algorithme rapide et robuste pour calculer les écoulements stationnaires.

De nombreux schémas numériques ont été proposés pour l'équation BGK instationnaire (voir les références ci-après), mais la simulation d'écoulements stationnaires implique des difficultés différentes. Pour ces écoulements, on peut soit résoudre l'équation BGK stationnaire avec un schéma itératif (dû à la non linéarité du terme source), soit résoudre l'équation instationnaire en laissant le temps croître jusqu'à stationnarisation.

Dans la première méthode, la non linéarité peut être traitée par un algorithme de Newton, mais il est connu en CFD qu'une telle méthode n'est pas très robuste et peut ne pas converger. Elle peut aussi être traitée par une technique de point fixe, comme Aoki, Kanba, Takata dans [4] avec le schéma $v \cdot \nabla_x f^{n+1} = \frac{1}{\tau^n} (M[f^n] - f^{n+1})$, mais cette technique découple les termes de gain et de perte, et peut converger très lentement. On peut d'ailleurs interpréter cette méthode comme un schéma semi-implicite, où seul le terme de perte est implicite (car il donne des distributions négatives).

Dans la seconde méthode, le principal problème est que pour garantir la positivité des distributions, le pas de temps Δt doit satisfaire deux sévères conditions :

- Δt doit être plus petit que le temps de relaxation τ . Ceci est très restrictif dans les régimes transitionnels pour lesquels τ est petit,
- une molécule ne doit pas traverser plus d'une cellule du maillage pendant Δt . Cette condition est connue comme la condition CFL pour la convection. Par conséquent, Δt doit être très large dans les régimes à grande vitesse, comme pour les problèmes de rentrée atmosphérique.

Ces restrictions sur le pas de temps impliquent qu'un très grand nombre d'itérations est nécessaire pour atteindre la solution stationnaire.

La première contrainte a beaucoup été étudiée. Par exemple, Perthame et Coron dans [18] ont proposé un schéma de relaxation (où à la fois la forme exponentielle de l'équation BGK et la méthode de splitting sont utilisées) qui est stable pour n'importe quel τ petit. Une méthode similaire pour d'autres équations cinétiques a été proposée par Gabetta, Pareschi et Toscani dans [22]. La même idée a aussi été utilisée par Issautier dans [29] avec une discrétisation lagrangienne à la place de la méthode de splitting.

1. Les sections 5.1, 5.3, et 5.4 font l'objet d'une partie d'un article intitulé *Discrete Velocity Model and Implicit Scheme for the BGK Equation of Rarefied Gas Dynamics*, à paraître dans *Math. Models and Meth. in Applied Sciences*.

La seconde contrainte peut être plus restrictive que la première. Cependant, elle a reçu considérablement moins d'attention. Une méthode classique en CFD est d'utiliser une implicitation du terme de convection (voir par exemple [66] pour les équations d'Euler). Mais comme mentionné par Waluś in [63], les méthodes de splitting sont incompatibles avec les procédures implicites car les phases de transport et de collision sont alors trop découplées. En outre, la discrétisation lagrangienne de [29] n'est pas très adaptée aux écoulements stationnaires.

La méthode introduite par Yang et Huang dans [65] est complètement différente, et apparaît très intéressante pour les calculs stationnaires. Ils proposent un schéma implicite basé sur un schéma volumes finis explicite de l'équation BGK, et ils utilisent toutes les techniques des systèmes hyperboliques. Cependant, ils n'implicitent que le terme de perte, car la linéarisation de la maxwellienne est complexe. Leur méthode est donc une méthode semi-implicite, qui a les mêmes inconvénient de convergence lente que la méthode de [4] citée ci-dessus. Dans ce chapitre, nous présentons une approche similaire pour le modèle BGK à vitesses discrètes, mais nous utilisons un terme source totalement implicite.

Le plan de ce chapitre est le suivant. Nous proposons tout d'abord dans la section 5.1 un schéma explicite dont nous prouvons qu'il possède des propriétés de conservation et d'entropie, essentiellement dues aux propriétés du modèle à vitesses discrètes. Dans la section 5.2, nous prouvons un résultat de convergence de ce schéma vers le modèle à vitesses discrètes quand les pas de temps et d'espace tendent vers 0. Nous présentons dans la section 5.3 un schéma implicite linéarisé adapté aux calculs d'écoulements stationnaires, développé à partir du schéma explicite. Dans la section 5.4, nous indiquons comment calculer efficacement la fonction d'équilibre discrète. Enfin, nous présentons dans la section 5.5 comment on peut prendre en compte les conditions aux limites dans les schémas numériques.

5.1 Schéma explicite

Par soucis de simplicité, notre schéma est présenté en deux dimensions d'espace seulement, sur une grille cartésienne, mais toutes les propriétés établies ici sont vraies pour un espace à D dimensions et pour des maillages curvilignes. De plus, nous adoptons dans ce chapitre les notations du chapitre 1. L'équation à approcher est

$$\partial_t f_k + v_k^{(1)} \partial_x f_k + v_k^{(2)} \partial_y f_k = \frac{1}{\tau} (\mathcal{E}_k - f_k), \quad k \in \mathcal{K}. \quad (5.1)$$

Considérons une grille cartésienne en espace définie par les noeuds $(x_i, y_j) = (i\Delta x, j\Delta y)$ et par les mailles $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$. Considérons aussi une discrétisation en temps avec $t_n = n\Delta t$. Si $f_{i,j}^n = (f_{k,i,j}^n)_{k \in \mathcal{K}}$ est une approximation de $f_{\mathcal{K}}(t_n, x_i, y_j)$, les moments de $f_{i,j}^n$ sont naturellement $\rho_{i,j}^n = \langle \mathbf{m} f_{i,j}^n \rangle_{\mathcal{K}}$, et l'équilibre discret correspondant est désigné par $(\mathcal{E}_{k,i,j}^n)_{k \in \mathcal{K}}$. Si $\rho_{i,j}^n$ est strictement réalisable (dans le sens de (1.28)), l'équilibre discret est alors $\mathcal{E}_{k,i,j}^n = \exp(\alpha_{i,j}^n \cdot \mathbf{m}_k)$, où $\alpha_{i,j}^n$ est l'unique solution du système d'équations non-linéaires (voir remarque 1.1)

$$\langle \mathbf{m} \exp(\alpha_{i,j}^n \cdot \mathbf{m}) \rangle_{\mathcal{K}} = \rho_{i,j}^n. \quad (5.2)$$

Le terme de transport est un terme linéaire de convection, qui peut être approché par un schéma volumes finis standard. Pour le terme de relaxation non-linéaire, une technique classique d'approximation centrée est utilisée. Notre schéma s'écrit alors

$$\begin{aligned} f_{k,i,j}^{n+1} = & f_{k,i,j}^n - \frac{\Delta t}{\Delta x} (\mathcal{F}_{k,i+\frac{1}{2},j}^n - \mathcal{F}_{k,i-\frac{1}{2},j}^n) - \frac{\Delta t}{\Delta y} (\mathcal{F}_{k,i,j+\frac{1}{2}}^n - \mathcal{F}_{k,i,j-\frac{1}{2}}^n) \\ & + \frac{\Delta t}{\tau_{i,j}^n} (\mathcal{E}_{k,i,j}^n - f_{k,i,j}^n), \end{aligned} \quad (5.3)$$

où les flux numériques sont définis par

$$\mathcal{F}_{k,i+\frac{1}{2},j}^n = \frac{1}{2} \left(v_k^{(1)} f_{k,i+1,j}^n + v_k^{(1)} f_{k,i,j}^n - |v_k^{(1)}| (\Delta f_{k,i+\frac{1}{2},j}^n - \Phi_{k,i+\frac{1}{2},j}^n) \right) \quad (5.4)$$

avec la notation $\Delta f_{k,i+\frac{1}{2},j}^n = f_{k,i+1,j}^n - f_{k,i,j}^n$, et le limiteur de flux $\Phi_{k,i+\frac{1}{2},j}^n$ permet d'obtenir un schéma d'ordre deux. Par exemple

$$\Phi_{k,i+\frac{1}{2},j}^n = \begin{cases} 0 & \text{pour l'ordre un,} \\ \min\text{mod}(\Delta f_{k,i-\frac{1}{2},j}^n, \Delta f_{k,i+\frac{1}{2},j}^n, \Delta f_{k,i+\frac{3}{2},j}^n) & \text{pour l'ordre deux.} \end{cases} \quad (5.5)$$

Remarque 5.1. On peut obtenir un schéma analogue sur un maillage curviligne. Dans ce cas les vitesses apparaissant dans les flux numériques sont multipliées par des coefficients géométriques, variables sur le maillage. Les nouveaux flux peuvent s'écrire comme dans (5.4), à la différence que les vitesses sont alors indicées par i, j .

Grâce à la définition de notre modèle à vitesses discrètes, le schéma possède les propriétés désirées. Dans le cas d'un domaine non borné (*i.e.* $(i, j) \in \mathbb{Z}^2$), nous donnons ainsi la version numérique du théorème 1.2 :

Proposition 5.1. Soit $\{f_{k,i,j}^0\}_{k,i,j}$ une condition initiale vérifiant

$$f_{k,i,j}^0 > 0, \quad \forall k \in \mathcal{K}, \forall i, j \in \mathbb{Z}.$$

Alors sous la condition suivante sur le pas de temps

$$\Delta t \left(\max_{i,j} \left(\frac{1}{\tau_{i,j}^n} \right) + \max_{\mathcal{K}} \left(\frac{|v_k^{(1)}|}{\Delta x} + \frac{|v_k^{(2)}|}{\Delta y} \right) \right) < 1, \quad (5.6)$$

la suite $\{f^n\}_{n \geq 0}$ définie par le schéma d'ordre un (5.3) satisfait

$$f_{k,i,j}^n > 0 \quad \text{et} \quad \mathcal{E}_{k,i,j}^n = \exp(\alpha_{i,j}^n \cdot \mathbf{m}_k), \quad \forall n, \forall k \in \mathcal{K}, \forall i, j \in \mathbb{Z}. \quad (5.7)$$

De plus, on a pour tout n les relations de conservation et de dissipation d'entropie suivantes

$$\sum_{k,i,j} \mathbf{m}_k f_{k,i,j}^n \Delta v^D = \sum_{k,i,j} \mathbf{m}_k f_{k,i,j}^0 \Delta v^D \quad (5.8)$$

$$\sum_{k,i,j} f_{k,i,j}^{n+1} \log f_{k,i,j}^{n+1} \Delta v^D \leq \sum_{k,i,j} f_{k,i,j}^n \log f_{k,i,j}^n \Delta v^D. \quad (5.9)$$

Preuve. La relation (5.3) s'écrit

$$\begin{aligned} f_{k,i,j}^{n+1} &= \left(1 - \frac{\Delta t}{\tau_{i,j}^n} - |v_k^{(1)}| \frac{\Delta t}{\Delta x} - |v_k^{(2)}| \frac{\Delta t}{\Delta y} \right) f_{k,i,j}^n - \frac{\Delta t}{\Delta x} v_k^{(1)-} f_{k,i+1,j}^n \\ &\quad + \frac{\Delta t}{\Delta x} v_k^{(1)+} f_{k,i-1,j}^n - \frac{\Delta t}{\Delta y} v_k^{(2)-} f_{k,i,j+1}^n + \frac{\Delta t}{\Delta y} v_k^{(2)+} f_{k,i,j-1}^n \\ &\quad + \frac{\Delta t}{\tau_{i,j}^n} \mathcal{E}_{k,i,j}^n, \end{aligned} \quad (5.10)$$

avec les notations classiques $a^+ = \frac{a+|a|}{2}$ et $a^- = \frac{a-|a|}{2}$. Les cinq derniers termes du membre de droite sont tous strictement positifs; la positivité de $f_{k,i,j}^{n+1}$ est donc assurée si

$$1 - \frac{\Delta t}{\tau_{i,j}^n} - |v_k^{(1)}| \frac{\Delta t}{\Delta x} - |v_k^{(2)}| \frac{\Delta t}{\Delta y} \geq 0, \quad \forall k \in \mathcal{K}, \forall i, j \in \mathbb{Z},$$

ce qui équivaut à (5.6). La stricte positivité de $f_{k,i,j}^n$ pour tout n, i, j implique immédiatement que $\rho_{i,j}^n$ est strictement réalisable pour tout n, i, j . Par conséquent, d'après le théorème 1.1, l'équilibre discret s'écrit bien sous forme exponentielle $\mathcal{E}_{k,i,j}^n = \exp(\alpha_{i,j}^n \cdot \mathbf{m}_k)$.

Pour la propriété de conservation (5.8), il suffit de sommer la relation (5.10) sur $i, j \in \mathbb{Z}$ et $k \in \mathcal{K}$: les termes de flux s'annulent et on utilise la propriété $\langle \mathbf{m} \mathcal{E}_{\mathcal{K},i,j}^n \rangle_{\mathcal{K}} = \langle \mathbf{m} f_{\mathcal{K},i,j}^n \rangle_{\mathcal{K}}$.

Enfin, pour la relation d'entropie (5.9), remarquons que la relation (5.10) est une combinaison convexe de $f_{k,i,j}^n, f_{k,i+1,j}^n, f_{k,i,j+1}^n, f_{k,i-1,j}^n, f_{k,i,j-1}^n$ et $\mathcal{E}_{k,i,j}^n$. Puisque la fonction $t \log t$ est aussi convexe, on obtient

$$\begin{aligned} f_{k,i,j}^{n+1} \log f_{k,i,j}^{n+1} &\leq \left(1 - \frac{\Delta t}{\tau_{i,j}^n} - |v_k^{(1)}| \frac{\Delta t}{\Delta x} - |v_k^{(2)}| \frac{\Delta t}{\Delta y} \right) f_{k,i,j}^n \log f_{k,i,j}^n \\ &\quad - \frac{\Delta t}{\Delta x} v_k^{(1)-} f_{k,i+1,j}^n \log f_{k,i+1,j}^n + \frac{\Delta t}{\Delta x} v_k^{(1)+} f_{k,i-1,j}^n \log f_{k,i-1,j}^n \\ &\quad - \frac{\Delta t}{\Delta y} v_k^{(2)-} f_{k,i,j+1}^n \log f_{k,i,j+1}^n + \frac{\Delta t}{\Delta y} v_k^{(2)+} f_{k,i,j-1}^n \log f_{k,i,j-1}^n \\ &\quad + \frac{\Delta t}{\tau_{i,j}^n} \mathcal{E}_{k,i,j}^n \log \mathcal{E}_{k,i,j}^n. \end{aligned}$$

En sommant sur i, j , les flux numériques d'entropie s'annulent, puis une somme en k donne

$$\sum_{k,i,j} f_{k,i,j}^{n+1} \log f_{k,i,j}^{n+1} \leq \sum_{k,i,j} \left(1 - \frac{\Delta t}{\tau_{i,j}^n} \right) f_{k,i,j}^n \log f_{k,i,j}^n + \sum_{k,i,j} \frac{\Delta t}{\tau_{i,j}^n} \mathcal{E}_{k,i,j}^n \log \mathcal{E}_{k,i,j}^n.$$

Puisque l'équilibre discret minimise l'entropie, la relation (5.9) devient alors évidente. \square

5.2 Résultat de convergence

Le but de cette section est de montrer que le schéma explicite converge bien vers l'unique solution du modèle à vitesses discrètes quand les pas de temps et d'espace tendent vers 0. Cependant, il est plus simple de considérer ici un domaine borné en espace, avec des conditions aux limites périodiques. Le résultat d'existence du théorème 2.2 s'adapte sans difficulté à ce cadre. Plus précisément, sous l'hypothèse

$$\int_{\mathbb{T}_x \times \mathbb{T}_y} \langle (1 + |v|^2 + |\log f^0|) f^0 \rangle dx dy < +\infty, \quad (5.11)$$

où \mathbb{T}^2 est le tore plat de \mathbb{R}^2 , l'approximation f_k^0 définie par (2.10) vérifie

$$0 < C^0 \leq f_k^0(x, y) \leq A^0 \quad \text{p.p } x, y, \forall k \in \mathcal{K}, \quad (5.12)$$

où C^0 et A^0 sont deux constantes. On peut alors montrer, de la même façon que pour le théorème 2.2, que le modèle BGK à vitesses discrètes (5.1) admet une unique solution $f_{\mathcal{K}}$ dans $L^\infty([0, t_{max}] \times \mathbb{T}^2)^N$ pour tout $t_{max} > 0$.

Considérons le maillage de \mathbb{T}^2 défini par les mailles

$$\Lambda_i \times \Lambda_j = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}[\times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}[, \quad i, j = 1, \dots, M$$

où $x_{i\pm\frac{1}{2}} = (i \pm \frac{1}{2})\Delta x$ et $y_{j\pm\frac{1}{2}} = (j \pm \frac{1}{2})\Delta y$. Pour simplifier, nous supposons dans la suite que $\Delta x = \Delta y$. La condition initiale $f_{\mathcal{K}}^0$ est alors approchée sur ce maillage par

$$\begin{aligned} f_k^{0,\Delta x}(x, y) &= f_{k,i,j}^0 \quad \text{si } (x, y) \in \Lambda_i \times \Lambda_j \\ &= \frac{1}{\Delta x \Delta y} \int_{\Lambda_i \times \Lambda_j} f_k^0(x, y) dx dy. \end{aligned} \quad (5.13)$$

Considérons un temps fini $t_{max} > 0$ et la suite $\{f_{k,i,j}^n\}$ donnée par le schéma explicite d'ordre un de la section précédente

$$\begin{aligned} f_{k,i,j}^{n+1} &= f_{k,i,j}^n - \frac{\Delta t}{\Delta x} \left[v_k^{(1)+} (f_{k,i,j}^n - f_{k,i-1,j}^n) + v_k^{(1)-} (f_{k,i+1,j}^n - f_{k,i,j}^n) \right] \\ &\quad - \frac{\Delta t}{\Delta y} \left[v_k^{(2)+} (f_{k,i,j}^n - f_{k,i,j-1}^n) + v_k^{(2)-} (f_{k,i,j+1}^n - f_{k,i,j}^n) \right] \\ &\quad + \Delta t (\mathcal{E}_{k,i,j}^n - f_{k,i,j}^n) \end{aligned} \quad (5.14)$$

avec $i, j = 1, \dots, M$ et $n = 0, \dots, n_{max} - 1$, où n_{max} est tel que $n_{max}\Delta t \leq t_{max}$. Nous complétons la relation (5.14) par les conditions périodiques suivantes

$$f_{k,0,j}^n = f_{k,M,j}^n \quad f_{k,M+1,j}^n = f_{k,1,j}^n \quad (5.15)$$

$$f_{k,i,0}^n = f_{k,i,M}^n \quad f_{k,i,M+1}^n = f_{k,i,1}^n. \quad (5.16)$$

Soit $f_{\mathcal{K}}^{\Delta x}$ la fonction constante par maille définie par

$$f_k^{\Delta x}(t, x, y) = f_{k,i,j}^n \quad \text{si } (t, x, y) \in [t_n, t_{n+1}[\times \Lambda_i \times \Lambda_j, \quad (5.17)$$

avec $t_n = n\Delta t$, pour tout $k \in \mathcal{K}$. De façon analogue, on définit $\mathcal{E}_{\mathcal{K}}^{\Delta x}$ par

$$\begin{aligned} \mathcal{E}_k^{\Delta x}(t, x, y) &= \mathcal{E}_{k,i,j}^n \quad \text{si } (t, x, y) \in [t_n, t_{n+1}[\times \Lambda_i \times \Lambda_j, \\ &= \mathcal{E}_k[f_{\mathcal{K}}^{\Delta x}(t, x, y)]. \end{aligned}$$

Nous montrons alors le résultat de convergence suivant.

Théorème 5.1. *Si en plus de l'hypothèse (5.11), $f_{\mathcal{K}}^0$ vérifie l'hypothèse BV suivante*

$$\sup_k VT(f_k^0, \mathbb{T}^2) \leq B_0, \quad (5.18)$$

et si Δt et Δx sont liés par la condition CFL

$$\Delta t \left(1 + 2 \frac{\max_k |v_k|}{\Delta x} \right) = \sigma < 1, \quad (5.19)$$

alors la suite $\{f_{\mathcal{K}}^{\Delta x}\}_{\Delta x \leq 1}$ définie par (5.17) converge fortement dans $L^1([0, t_{max}[\times \mathbb{T}^2)^N$ quand $\Delta x \rightarrow 0$ vers l'unique solution $f_{\mathcal{K}}$ du modèle BGK discret (5.1).

Remarque 5.2. Nous adoptons ici la définition suivante de la variation totale sur \mathbb{T}^2

$$VT(g, \mathbb{T}^2) = \int_{\mathbb{T}_x} VT(g(x, \cdot), \mathbb{T}_y) dx + \int_{\mathbb{T}_y} VT(g(\cdot, y), \mathbb{T}_x) dy. \quad (5.20)$$

La preuve de ce théorème est exposée dans les deux sections suivantes (Sec. 5.2.1 et 5.2.2).

5.2.1 Estimations

Nous prouvons tout d'abord que l'estimation ponctuelle (5.12) de la condition initiale reste vraie en temps fini pour l'approximation numérique.

Lemme 5.1. *On a*

$$C^0 \leq f_{k,i,j}^0 \leq A^0, \quad (5.21)$$

$$C(t_{max}) \leq f_{k,i,j}^n \leq A(t_{max}), \quad (5.22)$$

pour tout $i, j = 1, \dots, M$, $k \in \mathcal{K}$, et $n = 0, \dots, n_{max}$. De plus, pour tout Δx vérifiant (5.19), on a

$$C(t_{max}) \leq f_k^{\Delta x}(t, x, y) \leq A(t_{max}), \quad (5.23)$$

dans $[0, t_{max}[\times \mathbb{T}^2$, où $C(t_{max})$ et $A(t_{max})$ sont des constantes calculées dans la démonstration ci-dessous, qui ne dépendent que de C^0 , A^0 et t_{max} .

Preuve. L'estimation (5.21) est évidente, et (5.23) est immédiatement déduite de (5.22). Prouvons alors cette estimation. D'après (5.14), on a

$$\begin{aligned} f_{k,i,j}^{n+1} \geq & \left(1 - \frac{\Delta t}{\Delta x} |v_k^{(1)}| - \frac{\Delta t}{\Delta y} |v_k^{(2)}| - \Delta t\right) f_{k,i,j}^n + \frac{\Delta t}{\Delta x} (v_k^{(1)+} f_{k,i-1,j}^n - v_k^{(1)-} f_{k,i+1,j}^n) \\ & + \frac{\Delta t}{\Delta y} (v_k^{(2)+} f_{k,i,j-1}^n - v_k^{(2)-} f_{k,i,j+1}^n). \end{aligned} \quad (5.24)$$

Les coefficients du membre de droite de l'inégalité sont tous positifs, on a donc

$$\min_{k,i,j} f_{k,i,j}^{n+1} \geq (1 - \Delta t) \min_{k,i,j} f_{k,i,j}^n, \quad (5.25)$$

et par conséquent

$$\min_{k,i,j} f_{k,i,j}^n \geq (1 - \Delta t)^n C_0 \quad \forall n \leq n_{max}. \quad (5.26)$$

On peut supposer que $\Delta t \leq \frac{1}{2}$, ce qui implique $1 - \Delta t \geq (1 + 2\Delta t)^{-1}$, d'où

$$\min_{k,i,j} f_{k,i,j}^n \geq (1 + 2\Delta t)^{-n} C_0 \geq e^{-2t_{max}} C_0 = C(t_{max}), \quad (5.27)$$

pour tout $\Delta t \leq \frac{1}{2}$ satisfaisant (5.19) et $n \leq n_{max}$. Pour prouver la borne supérieure de (5.22), nous remarquons que d'après (5.14), on a

$$f_{k,i,j}^{n+1} \leq (1 - \Delta t) \max_{k,i,j} f_{k,i,j}^n + \Delta t \mathcal{E}_{k,i,j}^n.$$

De plus, comme par définition $\mathcal{E}_{\mathcal{K},i,j}^n$ et $f_{\mathcal{K},i,j}^n$ ont les mêmes moments, alors on peut contrôler $\mathcal{E}_{\mathcal{K},i,j}^n$ de la façon suivante

$$\mathcal{E}_{k,i,j}^n \leq \sum_{k'} \mathcal{E}_{k',i,j} = \sum_{k'} f_{k',i,j} \leq N \max_{k,i,j} f_{k,i,j}^n. \quad (5.28)$$

On tire des deux inégalités précédentes que

$$\max_{k,i,j} f_{k,i,j}^{n+1} \leq (1 + (N-1)\Delta t) \max_{k,i,j} f_{k,i,j}^n,$$

d'où par récurrence

$$\max_{k,i,j} f_{k,i,j}^n \leq e^{(N-1)t_{max}} A_0 = A(t_{max}), \quad \text{pour } n = 0, \dots, n_{max},$$

ce qui est bien l'estimation cherchée. \square

Nous montrons à présent que la variation totale en temps et en espace de l'approximation numérique est contrôlée en temps fini.

Lemme 5.2. *Pour tout $k \in \mathcal{K}$, on a*

$$\sum_{j=1}^M \Delta y \sum_{i=2}^M |\Delta f_{k,i-\frac{1}{2},j}^0| + \sum_{i=1}^M \Delta x \sum_{j=2}^M |\Delta f_{k,i,j-\frac{1}{2}}^0| \leq B_0, \quad (5.29)$$

$$\sum_{j=1}^M \Delta y \sum_{i=2}^M |\Delta f_{k,i-\frac{1}{2},j}^n| + \sum_{i=1}^M \Delta x \sum_{j=2}^M |\Delta f_{k,i,j-\frac{1}{2}}^n| \leq B_1(t_{max}), \quad (5.30)$$

$$\sum_{i,j=1}^M \sum_{n=0}^{n_{max}-1} |\delta f_{k,i,j}^{n+\frac{1}{2}}| \Delta x \Delta y \leq B_2(t_{max}), \quad (5.31)$$

et donc

$$\begin{aligned} & \sup_{\Delta x \leq 1} \max_k \int_{[0,t_{max}] \times \mathbb{T}_x} VT(f_k^{\Delta x}(t, x, \cdot), \mathbb{T}_y) dt dx + \int_{[0,t_{max}] \times \mathbb{T}_y} VT(f_k^{\Delta x}(t, \cdot, y), \mathbb{T}_x) dt dy \\ & + \int_{\mathbb{T}^2} VT(f_k^{\Delta x}(\cdot, x, y), [0, t_{max}]) dx dy \leq B(t_{max}) \end{aligned} \quad (5.32)$$

où l'on note pour toute fonction de grille g

$$\Delta g_{i-\frac{1}{2},j}^n = g_{i,j}^n - g_{i-1,j}^n \quad \Delta g_{i,j-\frac{1}{2}}^n = g_{i,j}^n - g_{i,j-1}^n, \quad (5.33)$$

$$\delta g_{i,j}^{n+\frac{1}{2}} = g_{i,j}^{n+1} - g_{i,j}^n. \quad (5.34)$$

Preuve. D'après la définition (5.13), on a

$$\begin{aligned} |\Delta f_{k,i-\frac{1}{2},j}^0| & \leq \frac{1}{\Delta x \Delta y} \left| \int_{\Lambda_i \times \Lambda_j} f_k^0(x, y) dx dy - \int_{\Lambda_{i-1} \times \Lambda_j} f_k^0(x, y) dx dy \right| \\ & \leq \frac{1}{\Delta x \Delta y} \int_{\Lambda_i \times \Lambda_j} |f_k^0(x + \Delta x, y) - f_k^0(x, y)| dx dy, \end{aligned}$$

et donc

$$\begin{aligned} \sum_{j=1}^M \Delta y \sum_{i=2}^M |\Delta f_{k,i-\frac{1}{2},j}^0| & \leq \int_{\mathbb{T}_y} dy \frac{1}{\Delta x} \int_{\mathbb{T}_x} |f_k^0(x + \Delta x, y) - f_k^0(x, y)| dx \\ & \leq \int_{\mathbb{T}_y} VT(f_k^0(\cdot, y), \mathbb{T}_x) dy. \end{aligned}$$

De la même façon, on obtient

$$\sum_{i=1}^M \Delta x \sum_{j=2}^M |\Delta f_{k,i,j-\frac{1}{2}}^0| \leq \int_{\mathbb{T}_x} VT(f_k^0(x, \cdot), \mathbb{T}_y) dx,$$

d'où (5.29).

Pour la relation (5.30), on introduit deux définitions supplémentaires utiles à l'élimination des termes de bord :

$$f_{k,-1,j}^n := f_{k,M-1,j}^n \quad f_{k,i,-1}^n := f_{k,i,M-1}^n. \quad (5.35)$$

On déduit de (5.14) que

$$\begin{aligned} \Delta f_{k,i-\frac{1}{2},j}^{n+1} = & \Delta f_{k,i-\frac{1}{2},j}^n - \frac{\Delta t}{\Delta x} \left[v_k^{(1)+} (\Delta f_{k,i-\frac{1}{2},j}^n - \Delta f_{k,i-\frac{3}{2},j}^n) + v_k^{(1)-} (\Delta f_{k,i+\frac{1}{2},j}^n - \Delta f_{k,i-\frac{1}{2},j}^n) \right] \\ & - \frac{\Delta t}{\Delta y} \left[v_k^{(2)+} (\Delta f_{k,i-\frac{1}{2},j}^n - \Delta f_{k,i-\frac{1}{2},j-1}^n) + v_k^{(2)-} (\Delta f_{k,i-\frac{1}{2},j+1}^n - \Delta f_{k,i-\frac{1}{2},j}^n) \right] \\ & + \Delta t (\Delta \mathcal{E}_{k,i-\frac{1}{2},j}^n - \Delta f_{k,i-\frac{1}{2},j}^n), \end{aligned} \quad (5.36)$$

pour tout $i, j = 1, \dots, M$. On somme sur i, j et par un changement d'indice, on obtient

$$\sum_{i,j=1}^M |\Delta f_{k,i-\frac{1}{2},j}^{n+1}| \leq (1 - \Delta t) \sum_{i,j=1}^M |\Delta f_{k,i-\frac{1}{2},j}^n| + \Delta t \sum_{i,j=1}^M |\Delta \mathcal{E}_{k,i-\frac{1}{2},j}^n|.$$

Par ailleurs, on a

$$|\Delta \mathcal{E}_{k,i-\frac{1}{2},j}^n| \leq L \max_k |\Delta f_{k,i-\frac{1}{2},j}^n|,$$

où L est la constante de Lipschitz de $g \mapsto \mathcal{E}_{\mathcal{K}}[g]$ sur le sous espace fermé borné $\{g \in \mathbb{R}^{n_{max} \times M \times N}, C(t_{max}) \leq g_{k,i,j}^n \leq A(t_{max})\}$. Des deux inégalités précédentes, on déduit

$$\max_k \sum_{i,j=1}^M |\Delta f_{k,i-\frac{1}{2},j}^{n+1}| \Delta y \leq (1 - \Delta t + L \Delta t) \max_k \sum_{i,j=1}^M |\Delta f_{k,i-\frac{1}{2},j}^n| \Delta y.$$

On obtient la même relation pour la variation totale par rapport à j

$$\max_k \sum_{i,j=1}^M |\Delta f_{k,i,j-\frac{1}{2}}^{n+1}| \Delta x \leq (1 - \Delta t + L \Delta t) \max_k \sum_{i,j=1}^M |\Delta f_{k,i,j-\frac{1}{2}}^n| \Delta x,$$

et donc par récurrence, on a

$$\begin{aligned} & \max_k \sum_{i,j=1}^M |\Delta f_{k,i-\frac{1}{2},j}^n| \Delta y + |\Delta f_{k,i,j-\frac{1}{2}}^n| \Delta x \\ & \leq e^{(L-1)t_{max}} \max_k \sum_{i,j=1}^M |\Delta f_{k,i-\frac{1}{2},j}^0| \Delta y + |\Delta f_{k,i,j-\frac{1}{2}}^0| \Delta x. \end{aligned} \quad (5.37)$$

De plus, il est clair que

$$\sum_{i,j=1}^M |\Delta f_{k,i-\frac{1}{2},j}^0| \Delta y \leq 2 \sum_{j=1}^M \Delta y \sum_{i=2}^M |\Delta f_{k,i-\frac{1}{2},j}^0|,$$

et on a la même relation pour la variation en i . On obtient donc grâce à l'inégalité (5.37)

$$\begin{aligned} & \max_k \sum_{i,j=1}^M |\Delta f_{k,i-\frac{1}{2},j}^n| \Delta y + |\Delta f_{k,i,j-\frac{1}{2}}^n| \Delta x \\ & \leq 2 e^{(L-1)t_{max}} \max_k \sum_{j=1}^M \Delta y \sum_{i=2}^M |\Delta f_{k,i-\frac{1}{2},j}^0| + \sum_{i=1}^M \Delta x \sum_{j=2}^M |\Delta f_{k,i,j-\frac{1}{2}}^0|. \end{aligned}$$

Cette dernière estimation, associée à l'inégalité (5.29), implique

$$\max_k \sum_{i,j=1}^M |\Delta f_{k,i-\frac{1}{2},j}^n| \Delta y + |\Delta f_{k,i,j-\frac{1}{2}}^n| \Delta x \leq 2B_0 e^{(L-1)t_{max}} = B_1(t_{max}). \quad (5.38)$$

De façon évidente, on a $\sum_{j=1}^M \Delta y \sum_{i=2}^M |\Delta f_{k,i-\frac{1}{2},j}^n| \leq \sum_{i,j=1}^M |\Delta f_{k,i-\frac{1}{2},j}^n| \Delta y$, et on a la même relation pour la variation en i , ce qui implique finalement

$$\max_k \sum_{j=1}^M \Delta y \sum_{i=2}^M |\Delta f_{k,i-\frac{1}{2},j}^n| + \sum_{i=1}^M \Delta x \sum_{j=2}^M |\Delta f_{k,i,j-\frac{1}{2}}^n| \leq B_1(t_{max}),$$

qui est bien l'inégalité (5.30).

Pour prouver l'estimation (5.31) de la variation totale par rapport à n , on peut vérifier que (5.14) implique

$$\begin{aligned} \sum_{i,j=1}^M \sum_{n=0}^{n_{max}-1} |\delta f_{k,i,j}^{n+\frac{1}{2}}| &\leq \sum_{i,j=1}^M \sum_{n=0}^{n_{max}-1} \frac{\Delta t}{\Delta x} |v_k^{(1)}| |\Delta f_{k,i-\frac{1}{2},j}^n| + \frac{\Delta t}{\Delta y} |v_k^{(2)}| |\Delta f_{k,i,j-\frac{1}{2}}^n| \\ &\quad + \Delta t |\mathcal{E}_{k,i,j}^n - f_{k,i,j}^n|, \end{aligned}$$

donc avec les relations (5.38), (5.22) et (5.28), et en notant que $\sum_{i,j=1}^M \Delta x \Delta y = 1$, on obtient finalement

$$\begin{aligned} \max_k \sum_{i,j=1}^M \sum_{n=0}^{n_{max}-1} |\delta f_{k,i,j}^{n+\frac{1}{2}}| \Delta x \Delta y &\leq \max_k |v_k| \sum_{n=0}^{n_{max}-1} \Delta t B_1(t_{max}) + \sum_{n=0}^{n_{max}-1} \Delta t \sum_{i,j=1}^M (\mathcal{E}_{k,i,j}^n + f_{k,i,j}^n) \Delta x \Delta y \\ &\leq \max_k |v_k| B_1(t_{max}) t_{max} + (N+1) A(t_{max}) t_{max} \\ &= B_2(t_{max}). \end{aligned}$$

Ceci achève la démonstration puisque la dernière inégalité (5.32) est immédiatement déduite des trois autres. \square

5.2.2 Convergence vers la solution

Grâce à (5.32) qui est uniforme en Δx , on peut vérifier que les conditions du théorème de Riesz-Frechet-Kolmogorov sont satisfaites par la suite $\{f_{\mathcal{K}}^{\Delta x}\}$ (cf. [9]). Par conséquent, on peut extraire une sous-suite encore notée $\{f_{\mathcal{K}}^{\Delta x}\}_{\Delta x \leq 1}$ telle que

$$f_{\mathcal{K}}^{\Delta x} \xrightarrow{\Delta x \rightarrow 0} f_{\mathcal{K}} \quad \text{dans } L^1([0, t_{max}] \times \mathbb{T}^2) \text{ et p.p dans } [0, t_{max}] \times \mathbb{T}^2.$$

Il reste à prouver que $f_{\mathcal{K}}$ est une solution de (5.1); puisque cette solution est unique, cela prouvera que toute la suite converge vers $f_{\mathcal{K}}$. Dans ce but, nous établissons une forme faible discrète en temps et en espace de (5.14). Soit φ une fonction test de $C_0^\infty([0, t_{max}] \times \mathbb{T}^2)$. Définissons une approximation constante par maille $\varphi^{\Delta x}$ de φ par

$$\begin{aligned} \varphi^{\Delta x}(t, x, y) &= \varphi_{i,j}^n \quad \text{si } (t, x, y) \in [t_n, t_{n+1}] \times \Lambda_i \times \Lambda_j \\ &= \frac{1}{\Delta x \Delta y} \int_{\Lambda_i \times \Lambda_j} \varphi(t_n, x', y') dx' dy', \end{aligned} \quad (5.39)$$

et posons

$$\varphi_{0,j}^n = \varphi_{M,j}^n \qquad \varphi_{M+1,j}^n = \varphi_{1,j}^n \qquad (5.40)$$

$$\varphi_{i,0}^n = \varphi_{i,M}^n \qquad \varphi_{i,M+1}^n = \varphi_{i,1}^n. \qquad (5.41)$$

En multipliant le schéma explicite par $\varphi_{i,j}^n$, il vient

$$\begin{aligned} & \frac{f_{k,i,j}^{n+1} - f_{k,i,j}^n}{\Delta t} \varphi_{i,j}^n + v_k^{(1)+} \frac{f_{i,j}^n - f_{i-1,j}^n}{\Delta x} \varphi_{i,j}^n + v_k^{(1)-} \frac{f_{i+1,j}^n - f_{i,j}^n}{\Delta x} \varphi_{i,j}^n \\ & + v_k^{(2)+} \frac{f_{i,j}^n - f_{i,j-1}^n}{\Delta y} \varphi_{i,j}^n + v_k^{(2)-} \frac{f_{i,j+1}^n - f_{i,j}^n}{\Delta y} \varphi_{i,j}^n \\ & - (\mathcal{E}_{k,i,j}^n - f_{k,i,j}^n) \varphi_{i,j}^n = 0. \end{aligned} \qquad (5.42)$$

On somme alors sur $n = 0, \dots, n_{max} - 1$ et $i, j = 1, \dots, M$, et en utilisant un changement d'indice on obtient

$$\begin{aligned} & \sum_{n=1}^{n_{max}-1} \sum_{i,j=1}^M f_{k,i,j}^n \left(-\frac{\varphi_{i,j}^n - \varphi_{i,j}^{n-1}}{\Delta t} - v_k^{(1)+} \frac{\varphi_{i+1,j}^n - \varphi_{i,j}^n}{\Delta x} - v_k^{(1)-} \frac{\varphi_{i,j}^n - \varphi_{i-1,j}^n}{\Delta x} \right. \\ & \qquad \qquad \qquad \left. - v_k^{(2)+} \frac{\varphi_{i,j+1}^n - \varphi_{i,j}^n}{\Delta y} - v_k^{(2)-} \frac{\varphi_{i,j}^n - \varphi_{i,j-1}^n}{\Delta y} \right) \\ & - \sum_{n=1}^{n_{max}-1} \sum_{i,j=1}^M (\mathcal{E}_{k,i,j}^n - f_{k,i,j}^n) \varphi_{i,j}^n \\ & + \sum_{i,j=1}^M \frac{1}{\Delta t} (f_{k,i,j}^{n_{max}} \varphi_{i,j}^{n_{max}-1} - f_{k,i,j}^0 \varphi_{i,j}^0) \\ & + \sum_{i,j=1}^M \left\{ f_{k,i,j}^0 \left(-v_k^{(1)+} \frac{\varphi_{i+1,j}^0 - \varphi_{i,j}^0}{\Delta x} - v_k^{(1)-} \frac{\varphi_{i,j}^0 - \varphi_{i-1,j}^0}{\Delta x} \right. \right. \\ & \qquad \qquad \qquad \left. - v_k^{(2)+} \frac{\varphi_{i,j+1}^0 - \varphi_{i,j}^0}{\Delta y} - v_k^{(2)-} \frac{\varphi_{i,j}^0 - \varphi_{i,j-1}^0}{\Delta y} \right) \\ & \qquad \qquad \qquad \left. - (\mathcal{E}_{k,i,j}^0 - f_{k,i,j}^0) \varphi_{i,j}^0 \right\} = 0. \end{aligned}$$

On multiplie l'égalité précédente par $\Delta t \Delta x \Delta y$, et on la réécrit en termes de $f_k^{\Delta x}$:

$$\begin{aligned}
& \int_{[\Delta t, t_{max}]} \int_{\mathbb{T}^2} dt dx dy \left\{ f_k^{\Delta x}(t, x, y) \left(-\frac{\varphi^{\Delta x}(t, x, y) - \varphi^{\Delta x}(t - \Delta t, x, y)}{\Delta t} \right. \right. \\
& \quad - v_k^{(1)+} \frac{\varphi^{\Delta x}(t, x + \Delta x, y) - \varphi^{\Delta x}(t, x, y)}{\Delta x} - v_k^{(1)-} \frac{\varphi^{\Delta x}(t, x, y) - \varphi^{\Delta x}(t, x - \Delta x, y)}{\Delta x} \\
& \quad \left. \left. - v_k^{(2)+} \frac{\varphi^{\Delta x}(t, x, y + \Delta y) - \varphi^{\Delta x}(t, x, y)}{\Delta y} - v_k^{(2)-} \frac{\varphi^{\Delta x}(t, x, y) - \varphi^{\Delta x}(t, x, y - \Delta y)}{\Delta y} \right) \right\} \\
& - \int_{[\Delta t, t_{max}]} \int_{\mathbb{T}^2} (\mathcal{E}_k^{\Delta x}(t, x, y) - f_k^{\Delta x}(t, x, y)) \varphi^{\Delta x}(t, x, y) dt dx dy \\
& + \int_{\mathbb{T}^2} dx dy \left\{ (f_k^{\Delta x}(t_{max}, x, y) \varphi^{\Delta x}(t_{max} - \Delta t, x, y) - f_k^{\Delta x}(0, x, y) \varphi^{\Delta x}(0, x, y)) \right\} \\
& + \Delta t \int_{\mathbb{T}^2} dx dy \left\{ f_k^{\Delta x}(0, x, y) \left(\right. \right. \\
& \quad - v_k^{(1)+} \frac{\varphi^{\Delta x}(0, x + \Delta x, y) - \varphi^{\Delta x}(0, x, y)}{\Delta x} - v_k^{(1)-} \frac{\varphi^{\Delta x}(0, x, y) - \varphi^{\Delta x}(0, x - \Delta x, y)}{\Delta x} \\
& \quad \left. \left. - v_k^{(2)+} \frac{\varphi^{\Delta x}(0, x, y + \Delta y) - \varphi^{\Delta x}(0, x, y)}{\Delta y} - v_k^{(2)-} \frac{\varphi^{\Delta x}(0, x, y) - \varphi^{\Delta x}(0, x, y - \Delta y)}{\Delta y} \right) \right. \\
& \quad \left. - (\mathcal{E}_k^{\Delta x}(0, x, y) - f_k^{\Delta x}(0, x, y)) \varphi^{\Delta x}(0, x, y) \right\} = 0.
\end{aligned}$$

D'après (5.22) et (5.28), $f_k^{\Delta x}(0, x, y)$ et $\mathcal{E}_k^{\Delta x}(0, x, y)$ sont uniformément bornées par rapport à Δx , x et y , donc la quatrième intégrale de l'expression ci-dessus tend vers 0 quand $\Delta t \rightarrow 0$. Pour la troisième, notons que $\varphi^{\Delta x}(t_{max} - \Delta t, x, y) \rightarrow \varphi(t_{max}, x, y) = 0$ pour tout (x, y) quand $\Delta x \rightarrow 0$, et comme $f_k^{\Delta x}(t_{max}, x, y)$ est uniformément bornée, la première partie de la troisième intégrale disparaît, c.-à-d.

$$\int_{\mathbb{T}^2} f_k^{\Delta x}(t_{max}, x, y) \varphi^{\Delta x}(t_{max} - \Delta t, x, y) dx dy \xrightarrow{\Delta x \rightarrow 0} 0.$$

Pour la seconde partie de la troisième intégrale, on remarque que $f_k^{\Delta x}(0, x, y)$ est, par construction, la moyenne de f_k^0 sur une maille, donc elle converge pour presque tout (x, y) vers $f_k^0(x, y)$ (cf. [56]). Donc par le théorème de convergence de Lebesgue, on a

$$\int_{\mathbb{T}^2} f_k^{\Delta x}(0, x, y) \varphi^{\Delta x}(0, x, y) dx dy \xrightarrow{\Delta x \rightarrow 0} \int_{\mathbb{T}^2} f_k^0(x, y) \varphi(0, x, y) dx dy.$$

Pour la première intégrale, remarquons que puisque $f_k^{\Delta x}$ converge presque partout vers f_k , alors une autre application du théorème de Lebesgue permet de conclure que cette intégrale converge vers

$$\int_{[0, t_{max}]} \int_{\mathbb{T}^2} dt dx dy \left\{ f_k \left(-\partial_t \varphi - v_k^{(1)} \partial_x \varphi - v_k^{(2)} \partial_y \varphi \right) \right\}.$$

Il reste à passer à la limite dans le terme non-linéaire, c.-à-d. dans la seconde intégrale. Mais nous savons que l'application $g \mapsto \mathcal{E}_k[g]$ est continue, donc il est clair que

$$\mathcal{E}_k^{\Delta x}(t, x, y) = \mathcal{E}_k[f_k^{\Delta x}(t, x, y)] \xrightarrow{\Delta x \rightarrow 0} \mathcal{E}_k(t, x, y) = \mathcal{E}_k[f_k(t, x, y)],$$

pour presque tout (t, x, y) . Par conséquent cette seconde intégrale converge bien vers la bonne limite.

Nous avons donc prouvé que l'on peut passer à la limite dans chaque terme de la formulation faible discrète pour obtenir

$$\int_{[0,t_{max}]} \int_{\mathbb{T}^2} dt dx dy \left\{ f_k \left(-\partial_t \varphi - v_k^{(1)} \partial_x \varphi - v_k^{(2)} \partial_y \varphi \right) - (\mathcal{E}_k - f_k) \varphi \right\} - \int_{\mathbb{T}^2} f_k^0(x, y) \varphi(0, x, y) dx dy = 0,$$

pour tout fonction test φ dans $C_0^\infty([0, t_{max}] \times \mathbb{T}^2)$. Ceci est la formulation faible de (5.1) dont on peut montrer qu'elle est équivalente à la forme intégrale de (5.1) (cf. annexe 1 de [21]). Or on a montré au chapitre 2 l'unicité d'une solution de cette forme intégrale. Donc la limite f_K de $f_K^{\Delta x}$ est bien l'unique solution de (5.1) dans $L^\infty([0, t_{max}] \times \mathbb{T}^2)^N$, ce qui achève la démonstration.

5.3 Schéma implicite linéarisé pour les écoulements stationnaires

Pour résoudre une équation stationnaire numériquement, on peut utiliser le code explicite instationnaire existant en le menant à stationnarisation. Cette méthode se révèle cependant très coûteuse car nécessite un nombre de pas de temps très important. Une autre idée est de résoudre directement l'équation stationnaire par une méthode de Newton, améliorée par une technique de globalisation de type "line search" ou "trust region". Mais cette méthode échoue généralement car ne prend pas suffisamment en compte la physique du problème. On va donc utiliser un compromis de ces deux approches, qui peut s'interpréter à la fois comme un procédé de stationnarisation et comme une technique de globalisation : un schéma implicite linéarisé.

Dans cette section, nous partons du schéma explicite étudié dans la section 5.1. Nous implicitons les termes de transport ainsi que le terme source, qu'il est alors nécessaire de linéariser. Nous obtenons ainsi un schéma implicite linéarisé qui nécessite de résoudre à chaque itération un grand système linéaire, pour lequel nous donnons une méthode rapide de résolution.

5.3.1 Ecriture du schéma

Les schémas implicites sont habituellement obtenus à partir des schémas explicites en évaluant au temps t_{n+1} les termes qui créent des quantités négatives pour les grands pas de temps Δt (par exemple les flux numériques ou les termes sources). Si les termes non-linéaires ne sont pas différentiables, alors ils sont linéarisés, sinon ils restent évalués à t_n (comme les limiteurs de flux).

Le schéma présenté ici est obtenu de cette façon. Pour le terme de collision, il est clair que $-f_{k,i,j}^n$ (appelé le terme de perte) doit être implicité, puisqu'il produit des distributions négatives si Δt est grand. Le terme de gain, c'est-à-dire l'équilibre discret $\mathcal{E}_{k,i,j}^n$, est positif, et peut donc être laissé explicite (c'est la stratégie utilisée dans [65]). On obtient alors le schéma suivant, que nous appellerons *semi-implicite* :

$$f_{k,i,j}^{n+1} = f_{k,i,j}^n - \frac{\Delta t}{\Delta x} (\mathcal{F}_{k,i+\frac{1}{2},j}^{n+1} - \mathcal{F}_{k,i-\frac{1}{2},j}^{n+1}) - \frac{\Delta t}{\Delta y} (\mathcal{F}_{k,i,j+\frac{1}{2}}^{n+1} - \mathcal{F}_{k,i,j-\frac{1}{2}}^{n+1}) + \frac{\Delta t}{\tau_{i,j}^n} (\mathcal{E}_{k,i,j}^n - f_{k,i,j}^{n+1}). \quad (5.43)$$

Cependant les termes de gain et de perte sont alors évalués en des temps différents. Nous avons observé que cela ralentit considérablement la convergence du schéma (voir chap.6). Par conséquent, nous préférons implicitiser aussi le terme de gain. A l'opposé, l'implicitation du temps de relaxation

$\tau_{i,j}^n$ nécessiterait un calcul plus compliqué, et ne paraît pas très utile. Puisque l'équilibre discret est une fonction non-linéaire de f , on le linéarise de la façon suivante

$$\mathcal{E}_{k,i,j}^{n+1} \approx \mathcal{E}_{k,i,j}^n + [D_{i,j}^n (f_{i,j}^{n+1} - f_{i,j}^n)]_k, \quad (5.44)$$

où $D_{i,j}^n$ est la matrice jacobienne de l'application $g \in \mathbb{R}^N \mapsto \mathcal{E}[g]$ évaluée en $f_{i,j}^n$. Le schéma *implicite linéarisé* d'ordre un est alors le suivant

$$\begin{aligned} f_{k,i,j}^{n+1} + \frac{\Delta t}{\Delta x} (\mathcal{F}_{k,i+\frac{1}{2},j}^{n+1} - \mathcal{F}_{k,i-\frac{1}{2},j}^{n+1}) + \frac{\Delta t}{\Delta y} (\mathcal{F}_{k,i,j+\frac{1}{2}}^{n+1} - \mathcal{F}_{k,i,j-\frac{1}{2}}^{n+1}) + \frac{\Delta t}{\tau_{i,j}^n} (f_{k,i,j}^{n+1} - [D_{i,j}^n f_{i,j}^{n+1}]_k) \\ = f_{k,i,j}^n + \frac{\Delta t}{\tau_{i,j}^n} (\mathcal{E}_{k,i,j}^n - [D_{i,j}^n f_{i,j}^n]_k), \quad k \in \mathcal{K}, \quad i, j = 1, \dots, i_{max}, j_{max}. \end{aligned} \quad (5.45)$$

Pour le schéma d'ordre deux, les limiteurs de flux (non différentiables) ne sont pas implicites. Une écriture de ce schéma plus adaptée au calcul est la δ -forme matricielle suivante

$$\left(\frac{I}{\Delta t} + R^n + T \right) \delta f^n = RHS^n, \quad (5.46)$$

où $\delta f^n = f^{n+1} - f^n$, I est la matrice identité, T est la matrice telle que $(Tf^n)_{k,i,j} = \frac{1}{\Delta x} (\mathcal{F}_{k,i+\frac{1}{2},j}^n - \mathcal{F}_{k,i-\frac{1}{2},j}^n) + \frac{1}{\Delta y} (\mathcal{F}_{k,i,j+\frac{1}{2}}^n - \mathcal{F}_{k,i,j-\frac{1}{2}}^n)$ sans les limiteurs de flux, R^n est telle que $(R^n f^n)_{i,j} = \frac{1}{\tau_{i,j}^n} (f_{i,j}^n - D_{i,j}^n f_{i,j}^n)$, et

$$\begin{aligned} RHS_{i,j}^n = - \frac{1}{\Delta x} (\mathcal{F}_{k,i+\frac{1}{2},j}^n - \mathcal{F}_{k,i-\frac{1}{2},j}^n) - \frac{1}{\Delta y} (\mathcal{F}_{k,i,j+\frac{1}{2}}^n - \mathcal{F}_{k,i,j-\frac{1}{2}}^n) \\ + \frac{1}{\tau_{i,j}^n} (\mathcal{E}_{i,j}^n - f_{i,j}^n) \end{aligned} \quad (5.47)$$

qui contient les limiteurs pour le schéma d'ordre deux. Un calcul simple de dérivation d'applications composées donne la forme suivante de la jacobienne $D_{i,j}^n$

$$D_{i,j}^n[k, k'] = A^{-1}(\alpha_{i,j}^n) : \mathbf{m}_k \otimes \mathbf{m}_{k'} \mathcal{E}_{k,i,j}^n \Delta v^D, \quad (5.48)$$

où $A(\alpha_{i,j}^n) = \langle \mathbf{m} \otimes \mathbf{m} \exp(\alpha_{i,j}^n \cdot \mathbf{m}) \rangle_{\mathcal{K}}$.

En raison de la linéarisation de $\mathcal{E}_{\mathcal{K}}^{n+1}$, le schéma ne possède pas les propriétés de conservation et d'entropie prouvées pour le schéma explicite. Cependant, nous nous intéressons seulement ici aux solutions stationnaires. Plus précisément, après convergence nous avons $f^{n+1} = f^n$ et le schéma donne bien les lois de conservation et d'entropie stationnaires.

Remarquons la structure particulière des matrices T et R^n liée au rôle différent des variables d'espace et de vitesse. Si les quantités $f_{k,i,j}^n$ sont stockées ainsi: $f^n = (f_k^n)_{k \in \mathcal{K}}$ avec $f_k^n = (f_{k,i,j}^n)$, alors on peut voir (cf. figure 5.1) que T est une matrice $NM \times NM$, diagonale par blocs T_k , où T_k sont des matrices $M \times M$ pentadiagonales ($M = i_{max} \times j_{max}$ est le nombre de mailles), et que R^n est une matrice pleine de blocs diagonaux $R_{k,k'}^n$. On peut par exemple noter que si f_k^n est rangée par i puis par j , la $(i, j)^e$ ligne d'un bloc T_k est

$$[0, \dots, 0, T_{k,i-1,j}, 0, \dots, 0, T_{k,i,j-1}, T_{k,i,j}, T_{k,i,j+1}, 0, \dots, 0, T_{k,i+1,j}, 0, \dots, 0], \quad (5.49)$$

avec

$$\begin{aligned} T_{k,i-1,j} = -\frac{1}{\Delta x} v_k^{(1)+}, \quad T_{k,i,j-1} = -\frac{1}{\Delta y} v_k^{(2)+}, \quad T_{k,i,j} = \frac{1}{\Delta x} |v_k^{(1)}| + \frac{1}{\Delta y} |v_k^{(2)}|, \\ T_{k,i,j+1} = \frac{1}{\Delta y} v_k^{(2)-}, \quad T_{k,i+1,j} = \frac{1}{\Delta x} v_k^{(1)-}. \end{aligned} \quad (5.50)$$

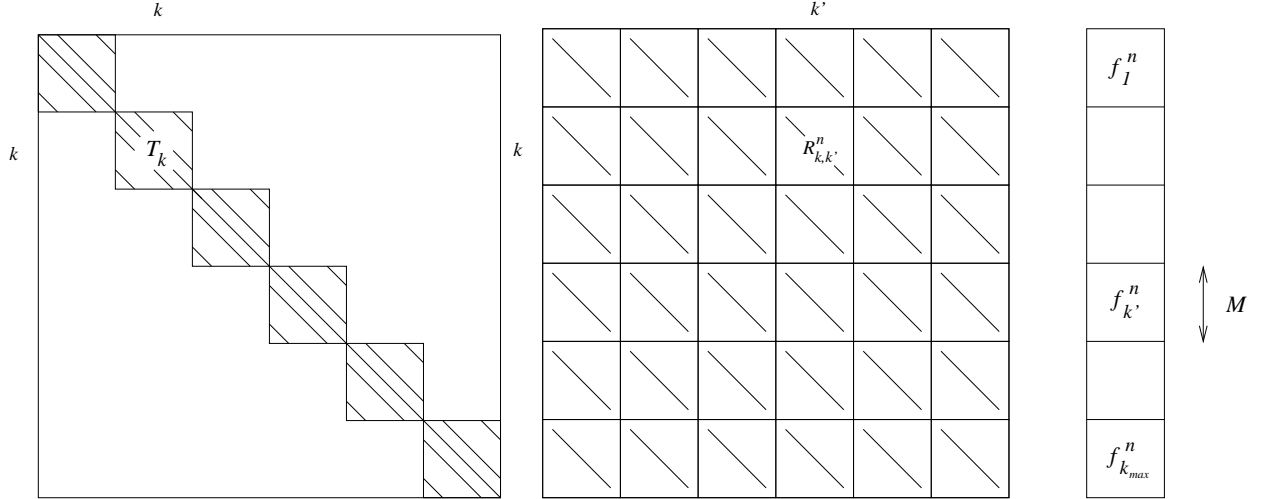


FIG. 5.1 – Matrices T et R^n , et stockage correspondant du vecteur f^n .

L'élément diagonal de la (i, j) ^e ligne d'un bloc $R_{k,k'}^n$ est

$$\frac{1}{\tau_{i,j}^n} (\delta_{k,k'} - D_{i,j}^n[k, k']), \quad (5.51)$$

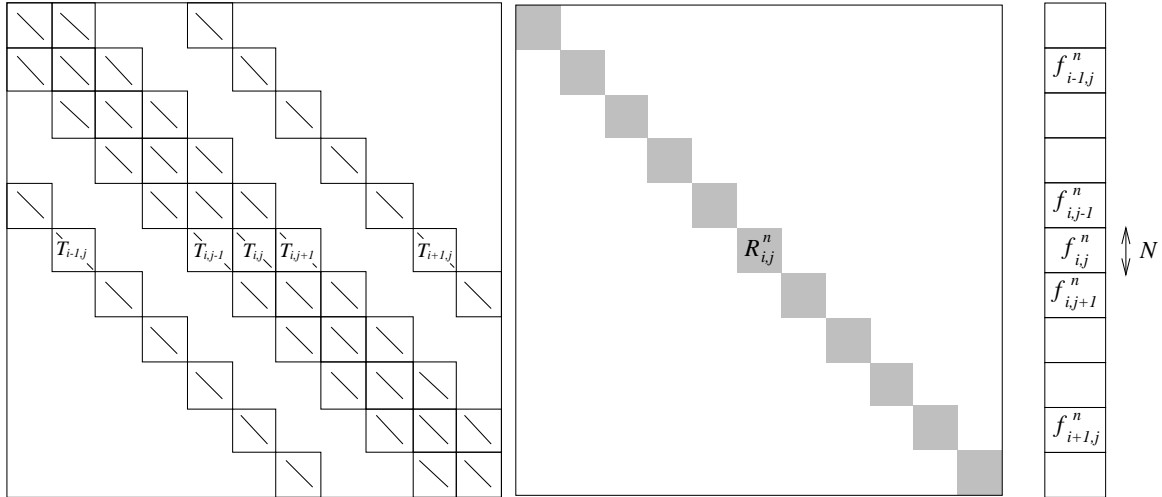
où $\delta_{k,k'}$ est le symbole de Kronecker, et $D_{i,j}^n[k, k']$ est définie dans (5.48). Ces structures creuses sont naturellement dues au fait que le phénomène de collision dans l'équation BGK est local en espace mais global en vitesse, alors que le phénomène de transport est numériquement "global" en espace, mais local en vitesse. On pourrait remarquer que les éléments de la diagonale d'un bloc T_k ne dépendent pas de (i, j) et sont donc tous égaux entre eux (ainsi que ceux des sous-diagonales), ce qui simplifierait le schéma. Mais dans la pratique, nous utilisons un maillage curviligne plutôt que cartésien, ce qui donne des coefficients géométriques dans les flux numériques qui dépendent de (i, j) . Par conséquent, les éléments de T_k dépendent pratiquement de (i, j) et c'est pourquoi nous n'utilisons pas ici toutes les simplifications permises par l'utilisation d'un maillage cartésien.

On peut envisager une autre forme de stockage de $f_{k,i,j}^n$ en rangeant f^n par maille d'espace puis par vitesse discrète. Si on arrange les mailles par i puis par j , i.e. $f^n = ((f_{i,j}^n)_{j=1,\dots,j_{max}})_{i=1,\dots,i_{max}}$, avec $f_{i,j}^n = (f_{k,i,j}^n)_{k \in \mathcal{K}}$, alors T est une matrice pentadiagonale par blocs, de blocs diagonaux de taille $N \times N$, et R^n est une matrice diagonale par blocs, de blocs pleins $R_{i,j}^n$ (cf. figure 5.2). Les éléments diagonaux des blocs $T_{i-1,j}$, $T_{i,j-1}$, $T_{i,j}$, $T_{i,j+1}$, $T_{i+1,j}$, représentés sur la figure (5.2) sont $T_{k,i-1,j}$, $T_{k,i,j-1}$, $T_{k,i,j}$, $T_{k,i,j+1}$, et $T_{k,i+1,j}$ donnés dans (5.50). Enfin les éléments $R_{i,j}^n[k, k']$ des blocs $R_{i,j}^n$ représentés figure (5.2) sont

$$R_{i,j}^n[k, k'] = \frac{1}{\tau_{i,j}^n} (\delta_{k,k'} - D_{i,j}^n[k, k']).$$

5.3.2 Résolution du système linéaire

Le système linéaire (5.46) à résoudre à chaque itération est très large ($NM \times NM$), et il paraît pertinent d'utiliser une méthode itérative adaptée aux différentes structures creuses des matrices. Nous présentons ici un algorithme basé sur un couplage entre les méthodes de Jacobi et de Gauss-Seidel en utilisant le stockage de la figure 5.1. Tout d'abord, R^n est séparée en sa partie diagonale

FIG. 5.2 – Matrices T et R^n pour un autre stockage du vecteur f^n .

par blocs Δ^n et sa partie extra diagonale par blocs E^n , *i.e.* $R^n = \Delta^n - E^n$ (c'est l'étape Jacobi). Le système (5.46) peut alors être résolu par l'algorithme suivant

1. $g^{(0)} = 0$,
2. pour $p = 0, \dots, P$, résoudre $\left(\frac{I}{\Delta t} + T + \Delta^n\right)g^{(p+1)} = RHS^n + E^n g^{(p)}$,
3. poser $\delta f^n = g^{(P+1)}$.

Puisque la matrice de ces systèmes linéaires est diagonale par blocs, de blocs pentadiagonaux $\frac{I}{\Delta t} + T_k + \Delta_k^n$, il est possible d'utiliser une méthode de Gauss-Seidel (par lignes ou par colonne, aller ou retour) en posant $T_k = M_k - N_k$. On obtient l'algorithme suivant

1. $g^{(0)} = 0$,
2. $\forall k \in \mathcal{K}$ et pour $p = 0, \dots, P$, résoudre

$$\left(\frac{I}{\Delta t} + M_k + \Delta_k^n\right)g_k^{(p+1)} = RHS_k^n + N_k g_k^{(p)} + [E^n g^{(p)}]_k, \quad (5.52)$$

3. poser $\delta f^n = g^{(P+1)}$.

Les systèmes linéaires (5.52) peuvent aisément être résolus par une succession de décompositions LU sur des systèmes tridiagonaux de $\mathbb{R}^{i_{max} \times i_{max}}$ ou $\mathbb{R}^{j_{max} \times j_{max}}$. Notons que le calcul du produit $E^n g$ n'est pas très coûteux, puisque les blocs de E^n sont diagonaux. En effet, nous avons

$$[E^n g]_{k,i,j} = \frac{1}{\tau_{i,j}^n} A^{-1}(\alpha_{i,j}^n) \mathbf{m}_k \mathcal{E}_{k,i,j}^n \cdot \left(\langle \mathbf{m} g_{i,j} \rangle_{\mathcal{K}} - \mathbf{m}_k g_{k,i,j} \Delta v^D\right). \quad (5.53)$$

Il suffit donc de calculer $A^{-1}(\alpha_{i,j}^n) \mathbf{m}_k \mathcal{E}_{k,i,j}^n$ au début de l'algorithme (un calcul local en k et i, j), puis de calculer $\langle \mathbf{m} g_{i,j} \rangle_{\mathcal{K}}$ dans chaque maille (ce qui est local en i, j), et finalement de former le produit scalaire. Ce calcul de $E^n g$ est donc local en i, j , et donc complètement parallélisable; son coût est en $O(NM)$.

Il est bien connu en CFD que seules quelques itérations de cet algorithme sont nécessaires pour faire converger le processus global (la boucle en n), il n'est donc pas utile de mener l'algorithme précédent à convergence. Le coût global d'une itération du schéma implicite est donc en $O(PNM)$, où $P = 2$ ou 3 .

5.4 Calcul de $\alpha_{i,j}^n$

Le système d'équations non-linéaires (5.2) peut-être résolu par l'algorithme de Newton suivant, où J est défini par $J(\beta) = \langle \exp(\beta \cdot \mathbf{m}) \rangle_{\mathcal{K}} - \beta \cdot \rho$ dans la preuve du théorème 1.1 :

1. soit $\alpha^{(0)} \in \mathbb{R}^{D+2}$,
2. résoudre le système linéaire $J''(\alpha^{(r)})\alpha^{(r+1)} = \alpha^{(r)} - J'(\alpha^{(r)})$ jusqu'à ce qu'un critère d'arrêt soit atteint,
3. poser $\alpha_{i,j}^n = \alpha^{(r)}$.

Un tel algorithme nécessite environ $O((D^2 + D)N)$ opérations par itération r et par maille (i, j) et est donc en $O(MN)$. Cependant, il peut apparaître trois difficultés.

Tout d'abord, bien que l'inversion de $J''(\alpha^{(r)}) = \langle \mathbf{m} \otimes \mathbf{m} \exp(\alpha^{(r)} \cdot \mathbf{m}) \rangle_{\mathcal{K}}$ dans l'étape 2 soit toujours théoriquement possible (voir la preuve du théorème 1.1), elle est numériquement difficile. En effet, puisque le premier élément de $\mathbf{m}_k \otimes \mathbf{m}_k$ est 1 alors que le dernier est $|v_k|^4$, le conditionnement de cette matrice peut être très grand. En fait, le conditionnement est aussi grand que les bornes de la grille de vitesses sont grandes, et dans les expérimentations numériques que nous avons pratiquées, il est souvent supérieur à 10^{20} . Ceci aboutit à un manque de robustesse que nous avons résolu par une technique de globalisation de type "back-tracking linesearch" (voir [20]).

La seconde difficulté est le choix de la donnée initiale $\alpha^{(0)}$. Un choix naturel est de prendre le paramètre α de l'équilibre continu maxwellien donné dans (1.12). Cependant, ce paramètre peut être très différent de la solution, en particulier lorsque N est petit. Un autre choix est de prendre $\alpha^{(0)} = \alpha_{i,j}^{n-1}$ calculé à l'itération globale précédente. L'avantage de cette méthode est que pour de petits pas de temps (comme dans le schéma explicite) ainsi que pour des écoulements stabilisés, la donnée initiale est proche de la solution. Pour une utilisation pratique du schéma implicite, une approche qui s'est révélée efficace est d'utiliser la première méthode au début du calcul, quand l'écoulement peut connaître de fortes variations, puis d'utiliser la seconde méthode quand l'écoulement est à peu-près stabilisé.

Ceci résout le troisième problème qui pourrait être un trop grand nombre d'itérations de l'algorithme de Newton. En fait, avec notre choix de $\alpha^{(0)}$, l'algorithme converge rapidement, puisqu'une seule itération est généralement nécessaire dans la plupart des cas que nous avons testés.

Enfin, l'utilisation de la proposition 1.3 permet la réduction de la taille du système (5.2) de $D + 2$ à $D + 1$ dans le cas d'écoulements plans (avec une grille de vitesses adaptée). Ceci diminue le coût de l'algorithme d'environ un facteur 1.3.

5.5 Prise en compte des conditions aux limites

Dans le cas d'un domaine borné Ω , le schéma doit prendre en compte les conditions aux limites de l'équation. Pour cela, on remarque que d'après la relation (5.46), le calcul de f^{n+1} au bord du domaine (*i.e.* aux mailles $(1, j)$, (i_{max}, j) , $(i, 1)$, et (i, j_{max})) nécessite de connaître les valeurs de f^n dans les mailles $(0, j)$, $(i_{max} + 1, j)$, $(i, 0)$, et $(i, j_{max} + 1)$ qui ne sont *a priori* pas définies. Ceci est classique en mécanique des fluides, en particulier dans les méthodes volumes finis : on définit alors ces mailles dites fictives de l'autre côté du bord de Ω , et f est calculée dans ces mailles de façon à rendre compte le mieux possible des conditions aux limites. Nous montrons ici que la condition de réflexion diffuse se traduit naturellement en termes de mailles fictives.

Plaçons nous dans le cas du bord $j = 1$. D'après la relation (5.3), on n'a besoin de f_k^n dans la maille $(i, 0)$ que pour les vitesses sortantes (*i.e.* $v_k \cdot n_i > 0$, cf. fig. 5.3). On va naturellement considérer que la distribution des molécules sortantes de la maille fictive $(i, 0)$ est égale à une fonction d'équilibre discrète centrée sur la vitesse de la paroi u_w , et de température égale à celle de

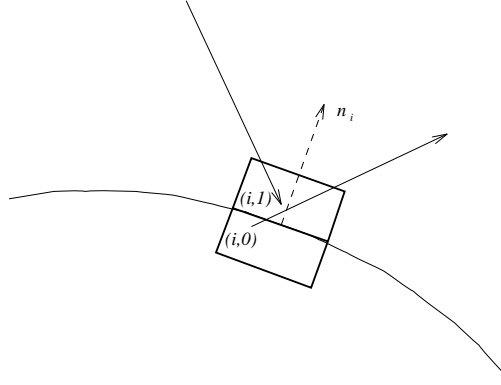


FIG. 5.3 – Vitesses sortantes et rentrantes à la paroi.

la paroi T_w , le paramètre φ de (1.48) étant déterminé pour que le flux de masse discret à travers la paroi soit nul. On a donc

$$f_{k,i,0}^n = \varphi \mathcal{E}_k[\boldsymbol{\rho}_w], \quad v_k \cdot n_i > 0, \quad (5.54)$$

où $\boldsymbol{\rho}_w = (1, u_w, \frac{1}{2}|u_w|^2 + \frac{D}{2}RT_w)$ et

$$\varphi = - \left(\sum_{\substack{k \text{ t. } q \\ v_k \cdot n_i < 0}} v_k \cdot n_i f_{k,i,1}^n \Delta v^D \right) \left(\sum_{\substack{k \text{ t. } q \\ v_k \cdot n_i > 0}} v_k \cdot n_i \mathcal{E}_k[\boldsymbol{\rho}_w] \Delta v^D \right)^{-1} \quad (5.55)$$

Pour les raisons que nous avons indiquées section 1.4.2, nous ne traitons pas ici le cas de la réflexion spéculaire.

Chapitre 6

Résultats numériques

Les méthodes précédentes ont servi de base à la réalisation d'un code de calcul qui nous a permis d'obtenir de nombreux résultats. Dans ce chapitre nous présentons quelques-uns de ces résultats qui illustrent toutes les capacités de notre approche. Nous évaluons l'influence des différents paramètres, la pertinence physique du modèle, le comportement de la méthode par rapport aux méthodes existantes, la faisabilité de certains calculs, et nous comparons différents schémas.

6.1 Description du code et choix des paramètres

6.1.1 Implémentation

Avec les différents schémas présentés précédemment, nous avons écrit le code de calcul KISS (Kinetic Implicit Supersonic Simulations), qui permet de calculer des écoulements 1D, 2D plan et 2D axi sur des maillages structurés curvilignes, en volumes finis "cell-centered". C'est un code séquentiel qui fonctionne sur l'IBM-SP2 du CESTA. Bien que nous n'utilisions qu'un processeur, le calculateur IBM-SP2 en possède plusieurs dont les caractéristiques sont de deux types différents :

- type 1 : processeur 66 MHz - 256 Mo,
- type 2 : processeur 120 MHz - 512 Mo,

Dans le code, nous avons privilégié la vitesse de calcul à la place mémoire. Par exemple, on calcule et on stocke à chaque itération en temps la fonction d'équilibre discrète $(\mathcal{E}_{k,i,j}^n)_{k,i,j}$ sur toutes les mailles, ce qui donne un tableau de dimension $N \times M$. L'autre approche, moins coûteuse en place mémoire mais plus coûteuse en temps, consisterait à ne stocker que les $\alpha_{i,j}^n$ (*i.e.* un tableau $5 \times N$) et de recalculer $\exp(\alpha_{i,j}^n \cdot \mathbf{m}_k)$ chaque fois que nécessaire. Le coût de calcul d'une exponentielle est cependant relativement élevé, c'est pourquoi nous avons opté pour la première solution. De même, pour la résolution du grand système linéaire du schéma implicite, nous stockons, à chaque pas de la méthode itérative, cinq grands tableaux de taille $N \times M$: la fonction de distribution $(f_{k,i,j}^n)_{k,i,j}$, l'équilibre discret $(\mathcal{E}_{k,i,j}^n)_{k,i,j}$, le tableau auxiliaire $(g_{k,i,j}^p)_{k,i,j}$ (qui correspond à l'incrément δf^n , voir section 5.3.2), le second membre explicite $(RHS_{k,i,j}^n)_{k,i,j}$, et la diagonale $(\Delta_{k,i,j}^n)_{k,i,j}$ de la matrice R^n . La place mémoire demandée par le code est donc supérieure à $5 \times N \times M$ (ce qui donne par exemple pour un maillage 50×50 et $10 \times 10 \times 10$ vitesses discrètes une taille de 95 Mo).

6.1.2 Détermination des bornes et du pas de la grille de vitesse

D'après la section 1.3.3, ces paramètres doivent être déterminés de façon à ce que la vitesse et la température maximales, ainsi que la température minimale de l'écoulement, puissent être

représentés sur la grille. Pour cela, supposons connues les quantités (ρ, u, T) de l'écoulement que l'on veut calculer dans un domaine Ω . On pose

$$v_{min}^{(i)} = \min_{\Omega}(u^{(i)} - cut\sqrt{RT}) \quad \text{et} \quad v_{max}^{(i)} = \max_{\Omega}(u^{(i)} + cut\sqrt{RT}) \quad (6.1)$$

pour $i = 1..D$, où cut est un paramètre de troncature. Si $cut = 1$, ces quantités sont les points d'inflexion dans les directions i des maxwelliennes $M(\rho, u, T)$, associées aux quantités u et T réalisant les minimum et maximum définis dans (6.1) (cf. figure 6.1). Si l'on suppose que le "support"

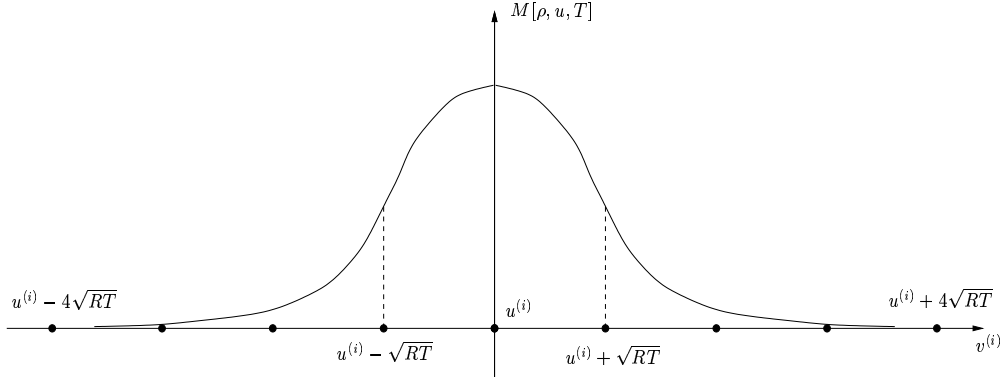


FIG. 6.1 – Points d'inflexion d'une maxwellienne dans la direction $v^{(i)}$ et points $u^{(i)} \pm cut(RT)^{\frac{1}{2}}$ avec $cut = 4$.

des fonctions de distribution de l'écoulement (*i.e.* l'intervalle contenant l'essentiel de la "masse") est le même que celui des maxwelliennes associées (ce qui semble vrai même dans le cas de fort déséquilibre), alors le domaine

$$\prod_{i=1}^D [v_{min}^{(i)}, v_{max}^{(i)}]$$

contient les supports de toutes les fonctions de distribution (cf. figure 6.2). En procédant de la sorte,

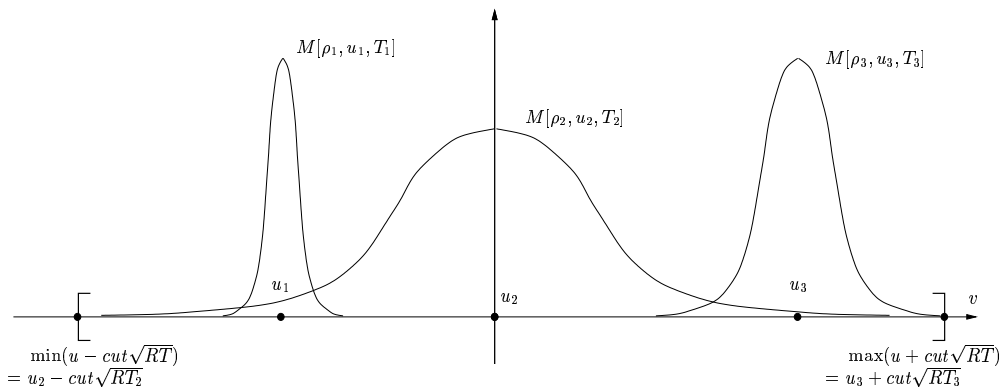


FIG. 6.2 – Domaine $[\min(u - cut(RT)^{\frac{1}{2}}), \max(u + cut(RT)^{\frac{1}{2}})]$ contenant tous les supports des fonctions de distribution de l'écoulement.

on s'assure que l'information essentielle de l'écoulement est capturée. On peut même montrer que

la vitesse $u(x)$ et la température $T(x)$ du gaz vérifient

$$\begin{aligned} v_{min}^{(i)} &\leq u^{(i)}(x) \leq v_{max}^{(i)} & \forall x \in \Omega, \\ T(x) &\leq \frac{1}{DR} \max_{v \in \Pi_{i=1}^D [v_{min}^{(i)}, v_{max}^{(i)}]} |v - u(x)|^2 & \forall x \in \Omega. \end{aligned}$$

Ainsi on peut décrire sur le domaine $\Pi_{i=1}^D [v_{min}^{(i)}, v_{max}^{(i)}]$ les vitesses et températures maximum de l'écoulement. Enfin, si on définit un pas de grille Δv tel que $\Delta v \leq \min_{\Omega} \sqrt{RT}$, et la grille $\mathcal{V} = \{v_k = k\Delta v, k^{(i)} = k_{min}^{(i)} \dots k_{max}^{(i)}\}$ avec $v_{k_{min}}^{(i)} = v_{min}^{(i)}$ et $v_{k_{max}}^{(i)} = v_{max}^{(i)}$, alors on assure en plus que

$$\frac{1}{DR} \min_k |v_k - u(x)|^2 \leq T(x) \quad \forall x \in \Omega,$$

ce qui implique que l'on peut décrire la température minimum de l'écoulement sur la grille \mathcal{V} (cf. section 1.3.3).

Il reste à déterminer les quantités $\min_{\Omega} |u - cut\sqrt{RT}|$ et $\max_{\Omega} |u + cut\sqrt{RT}|$. Dans certains cas, on peut les estimer si on connaît *a priori* la température minimum et les extrema de la vitesse. En effet, si la température de l'écoulement n'augmente pas trop, alors $u_{min} - cut\sqrt{RT_{min}}$ et $u_{max} + cut\sqrt{RT_{min}}$ suffisent à bien capturer l'information (car la racine carrée limite l'influence de l'augmentation de température). C'est par exemple le cas d'un écoulement de Couette où l'on a $u_{max} = u_{paroi}$ et $T_{min} = T_{paroi}$, avec une vitesse de paroi pas trop élevée. On peut encore citer le cas d'un choc stationnaire 1D où les valeurs de u et de T sont comprises entre les valeurs en amont et en aval du choc. Cependant, dans beaucoup d'écoulements, la température croît fortement et $\max_{\Omega} |u + cut\sqrt{RT}|$ est essentiellement donné par la température maximum, qui est difficile à estimer. Dans ce cas, nous effectuons un calcul Navier-Stokes préliminaire de l'écoulement (ce qui représente un faible surcoût). Nous estimons alors les quantités précédentes sur l'écoulement Navier-Stokes, ce qui donne des résultats tout à fait satisfaisants.

Remarque 6.1. Nous avons constaté que la valeur $cut = 4$ est un bon compromis entre le coût de la discrétisation et la précision des résultats (cf. section 6.2.1). C'est donc cette valeur que nous utilisons dans la plupart des résultats exposés dans ce chapitre.

6.1.3 Paramètres de la méthode

Le nombre d'itérations de la méthode itérative pour le système linéaire (cf. section 5.3.2) est fixé dans le code à deux (une aller et une retour). La méthode de Gauss-Seidel par lignes ou colonnes est choisie par l'utilisateur en début de calcul suivant la direction privilégiée de l'écoulement. Par exemple pour un écoulement de Couette avec des plaques parallèles à l'axe Oy , on choisira la méthode par colonnes pour mieux calculer les gradients en x qui sont les plus importants.

Le pas de temps Δt_{imp} du schéma implicite est calculé par la formule $\Delta t_{imp} = CFL\Delta t_{exp}$, où Δt_{exp} est le pas de temps du schéma explicite calculé avec la condition (5.6). Le nombre CFL vaut 1 en début de calcul et est incrémenté de 1 à chaque itération, jusqu'à un nombre d'itérations fixé par l'utilisateur, puis il prend la valeur maximum elle aussi fixée par l'utilisateur.

Le critère de convergence utilisé pour déterminer si l'écoulement obtenu est stationnaire est que le résidu quadratique global

$$\frac{1}{\Delta t} \sum_{k,i,j} |RHS_{k,i,j}^n|^2$$

diminue d'un facteur 10^5 . Ce facteur est bien sûr arbitraire, et il existe des cas où un facteur 10^3 suffit. Notons que ce critère est très différent du critère usuel de stationnarisation du nombre de molécules

(*i.e.* de la densité) utilisé dans les méthodes Monte-Carlo. Ceci rend délicat la comparaison du coût en temps CPU des deux approches. De plus, il faut mentionner que les processeurs de l'IBM-SP2 ayant des performances différentes les uns des autres, les comparaisons en temps CPU sont parfois des estimations, car nous n'avons pas la possibilité de choisir à chaque fois un processeur identique pour deux calculs consécutifs.

Pour les résultats numériques présentés dans les sections suivantes, nous utilisons, sauf mention contraire, le schéma implicite linéarisé d'ordre deux, donné section 5.3, avec des conditions de réflexion diffuse seulement. Le nombre CFL est fixé à 10000. Enfin, la viscosité permettant de définir le temps de relaxation τ est calculée par la formule (4.2), et les paramètres μ_{ref} et T_{ref} sont pris dans l'ouvrage de Bird ([5]-annexe A), ainsi que le paramètre ω (sauf pour les modèles HS ($\omega = \frac{1}{2}$) et maxwellien ($\omega = 1$)).

6.2 Influence des paramètres

6.2.1 Influence de la grille des vitesses

Choc stationnaire

Dans ce test, l'écoulement est initialisé avec deux maxwelliennes reliées par les relations de Rankine-Hugoniot. L'état stationnaire montre la transition entre les écoulements amont et aval. Les données utilisées sont $\rho_L = 6.63 \cdot 10^{-6} \text{ kg.m}^{-3}$, $T_L = 293 \text{ K}$, $u_L = 2551 \text{ m.s}^{-1}$ pour l'écoulement amont. Ces valeurs donnent un nombre de Mach de 8. le gaz considéré est de l'argon, et par conséquent $\omega = 0.81$ (avec le modèle VHS et les données de Bird [5]). La longueur du domaine de calcul est 0.5 m , et une grille 1D de 200 mailles est utilisée. La figure 6.3 montre les profils normalisés $\frac{q_L - q}{q_L - q_R}$ pour $q = \rho, T, u$, obtenus avec $11 \times 9 \times 9$ vitesses discrètes dans les directions (v_x, v_y, v_z) . Les bornes sont données par $[-3846, 5181] \times [-4513, 4513] \times [-4513, 4513]$, ce qui assure que les maxwelliennes amont et aval sont correctement représentées. Ceci nous permet de prendre en compte la majeure partie de l'information de l'écoulement. Ces profils sont légèrement moins étalés que ceux de Bird [5]. Les fonctions de distribution dans l'écoulement amont, dans le choc, et dans l'écoulement aval, sont tracées dans la figure 6.4. En dépit du faible nombre de vitesses discrètes, les résultats montrent clairement le phénomène de déséquilibre cinétique dans le choc (la distribution est assez éloignée de la distribution maxwellienne). Comme on s'y attend, les distributions amont et aval sont à peu près maxwelliennes (l'écoulement amont est moins bien représenté du fait de sa température plus basse).

Tout d'abord, nous montrons l'influence des bornes de la grille de vitesses sur les quantités microscopiques et macroscopiques. La figure 6.5 représente les distributions en amont, au choc, et en aval obtenues pour différentes bornes. Le nombre de vitesses discrètes étant adapté dans chaque cas pour donner la même précision. Les profils correspondants de ρ, u, T sont tracés dans la figure 6.6. Comme on s'y attend, les grandes bornes n'améliorent pas les résultats, puisque l'essentiel de l'information est déjà prise en compte avec les bornes normales (fig.6.5). Cependant, quand les bornes sont trop petites, les profils commencent à perdre leur précision (ils ne sont plus assez étalés). Une grande quantité d'information est perdue, en particulier dans la population des molécules de grande vitesse de l'écoulement amont, puisque les distributions au choc et en amont sont tronquées. Néanmoins, cette troncature donne une bonne indication pour vérifier *a posteriori* si la grille de vitesse est suffisamment large pour l'écoulement.

Nous présentons ensuite une étude des influences respectives du nombre de vitesses en x (direction de l'écoulement) et du nombre de vitesses en y et z . Pour cela, nous avons testé plusieurs grille de vitesse. Les résultats sont résumés dans la figure 6.7.

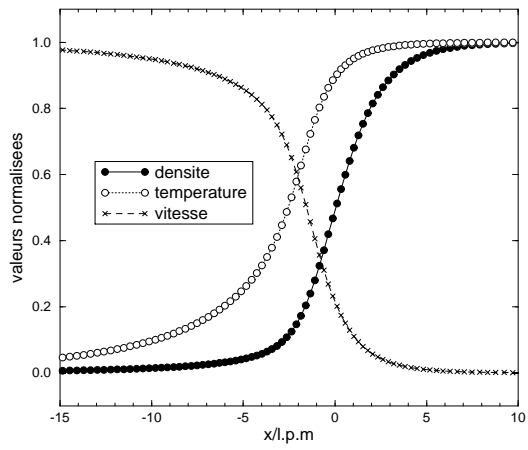


FIG. 6.3 – Profils normalisés pour un choc à Mach 8.

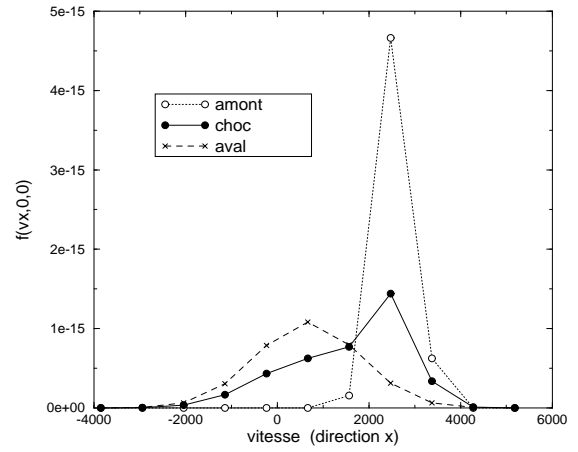


FIG. 6.4 – Fonctions de distribution dans le choc.

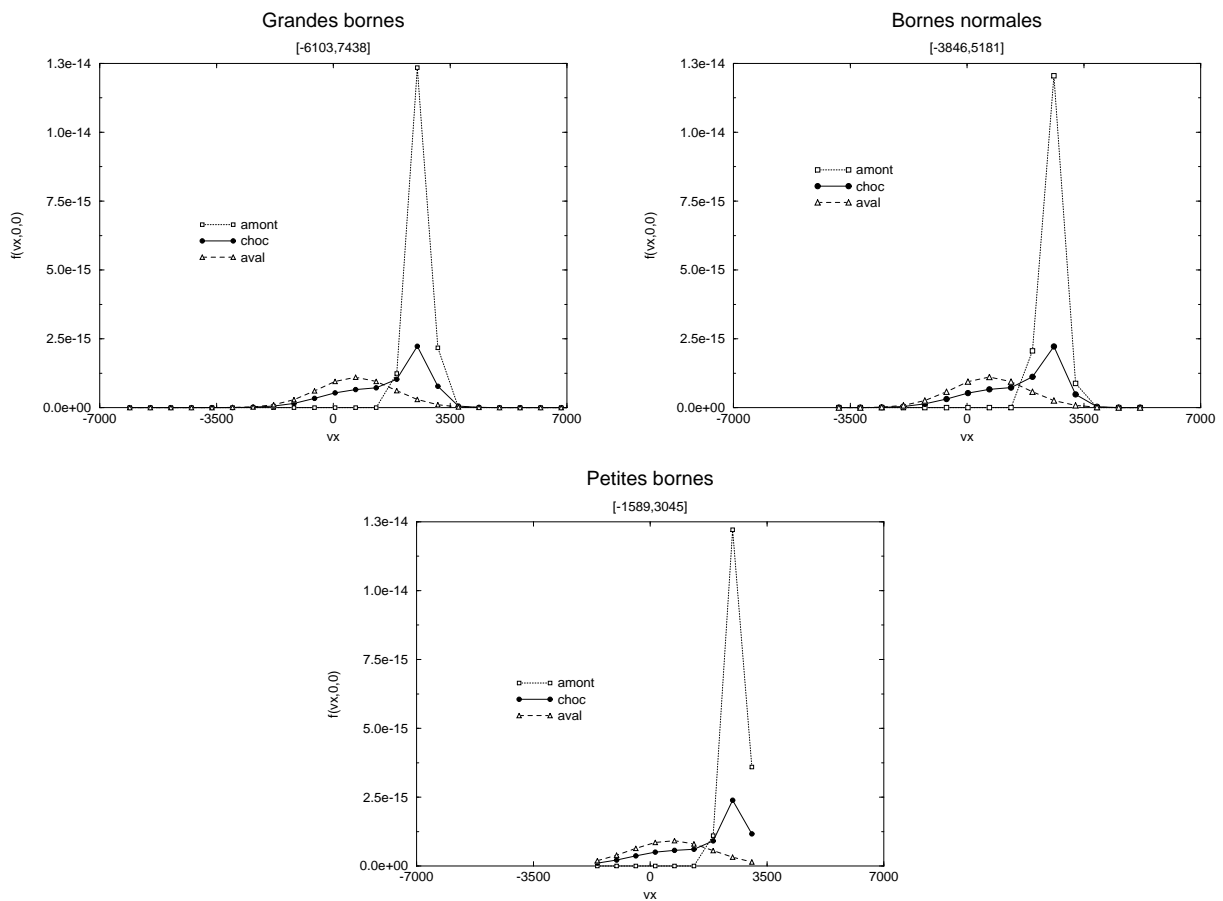


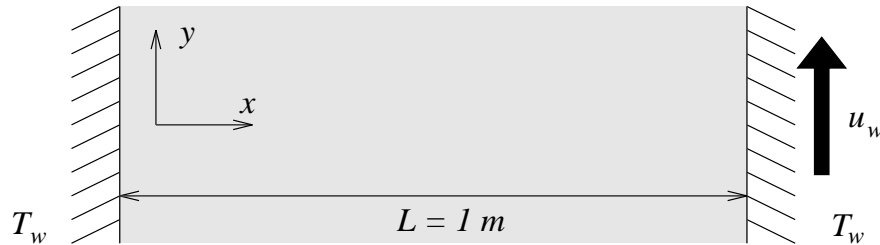
FIG. 6.5 – Influence des bornes de vitesse sur les fonctions de distribution.

Nous fixons dans un premier temps un petit nombre de vitesses en y et en z (7×7), et nous faisons varier le nombre de vitesses en x (courbes $11 \times 7 \times 7$, $17 \times 7 \times 7$, $21 \times 7 \times 7$, et $31 \times 7 \times 7$). On ne distingue alors quasiment aucune différence dans les profils. Nous n'avons pu tester un nombre de vitesses en x plus petit car le Newton pour calculer l'état droit ne converge plus.

En revanche, si on augmente le nombre de vitesses en y et z (13×13), on constate que les profils sont légèrement plus raides (courbes $15 \times 13 \times 13$, $17 \times 13 \times 13$, $21 \times 13 \times 13$). Mais là encore, l'augmentation du nombre de vitesses en x n'améliore rien. A contrario, si on prend un très petit nombre de vitesses en y et z (5×5), on s'aperçoit (courbe $11 \times 5 \times 5$) que l'on obtient des profils nettement plus étalés. Le nombre de vitesses $11 \times 9 \times 9$ semble finalement le nombre optimal pour ce cas (courbe $11 \times 9 \times 9$).

Écoulement de Couette plan 1D

Les données de l'écoulement sont celles de Bird dans [5]. Le gaz (argon), est contenu entre deux plaques maintenues à la température $T_w = 273 K$. La plaque de gauche est immobile, tandis que celle de droite se déplace avec la vitesse $u_w = 300 m.s^{-1}$ dans la direction y :



Le gaz est initialement à la température des plaques, et sa densité est $9.28 \cdot 10^{-6} kg.m^{-3}$. Le nombre de Knudsen basé sur la distance entre les plaques ($1 m$) est donc $9.25 \cdot 10^{-3}$. En utilisant une grille de 200 mailles dans la direction x et $13 \times 17 \times 13$ vitesses discrètes de bornes $[-913, 913] \times [-1253, 1253] \times [-913, 913]$, nous obtenons les profils tracés dans la figure 6.8. Nous trouvons un bon accord qualitatif avec les résultats de Bird, avec cependant des amplitudes plus grandes pour la densité et la température. Ceci est probablement dû au fait que le nombre de Prandtl donné par BGK est trop grand, *i.e.* que le transfert thermique est surestimé. Notons que la couche de Knudsen peut être observée sur le profil de pression : cette courbe est à peu près constante sauf près des plaques, où il y a une chute rapide sur une distance de l'ordre de quelques libre parcours moyens. Cette zone est aussi appelée couche limite cinétique.

Pour que l'influence de la grille de vitesse soit bien visible, nous étudions en fait un écoulement avec une vitesse u_w de $1000 m.s^{-1}$. Les différents nombre de vitesses utilisés sont $7 \times 9 \times 7$, $7 \times 17 \times 7$, et $7 \times 33 \times 7$ (le nombre de points en y est quasiment doublé à chaque fois), puis nous avons augmenté le nombre de points en x et en z avec $13 \times 17 \times 13$ et $17 \times 17 \times 17$. Enfin, nous avons testé un nombre de points égal à $7 \times 10 \times 7$, car on a ainsi un nombre pair de vitesses en y , ce qui implique l'absence de vitesse nulle parallèle à la paroi dans la grille.

Les résultats sont résumés fig. 6.9 et dans le tableau suivant :

	$7 \times 9 \times 7$	$7 \times 17 \times 7$	$7 \times 33 \times 7$	$13 \times 17 \times 13$	$17 \times 17 \times 17$	$7 \times 10 \times 7$
saut de T	18.7	17.9	17.8	14.3	13.9	17.9
saut de u	15.0	14.7	14.6	12.5	12.2	14.8
T_{max}	531.6	526.6	526.6	533.4	536.0	525.9
ρ_{min}	$7.5 \cdot 10^{-6}$	$7.6 \cdot 10^{-6}$	$7.7 \cdot 10^{-6}$	$7.7 \cdot 10^{-6}$	$7.7 \cdot 10^{-6}$	$7.5 \cdot 10^{-6}$

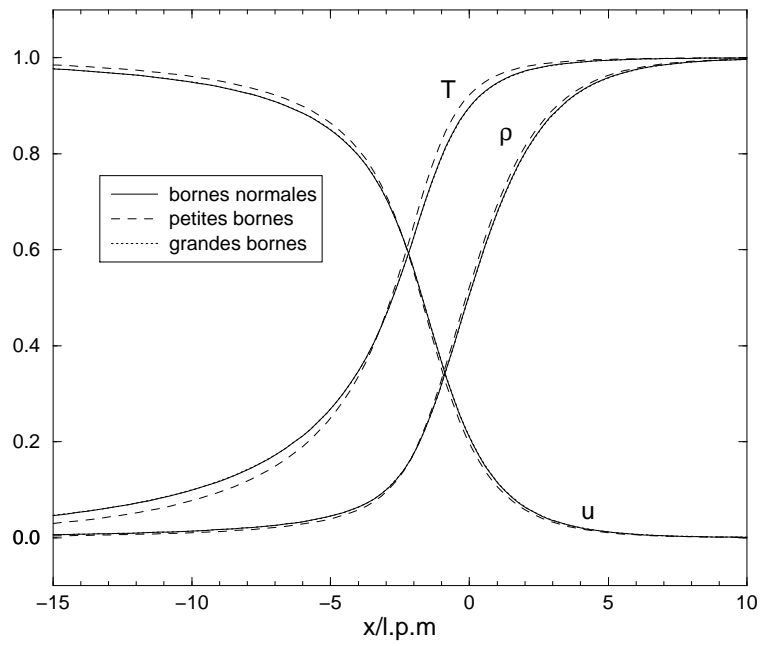


FIG. 6.6 – Influence des bornes de vitesse sur les quantités macroscopiques : densité ρ , température T , vitesse u .

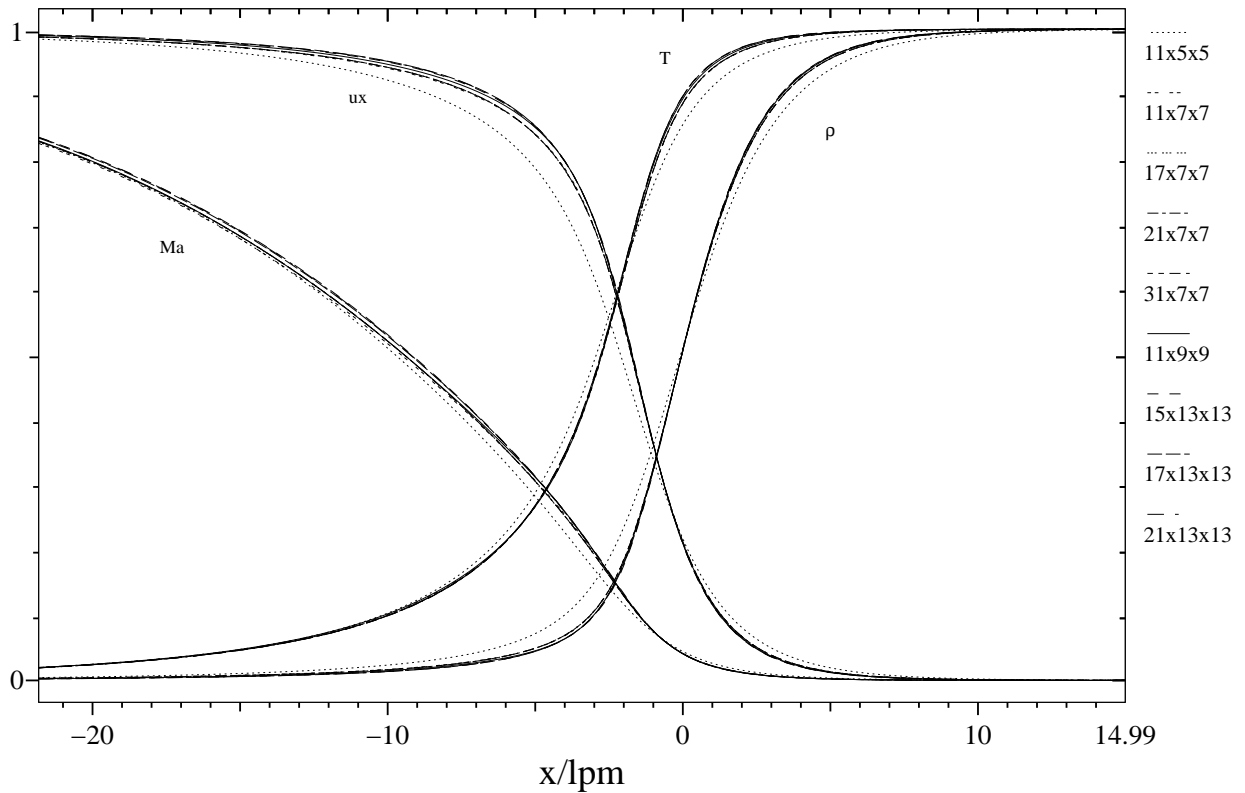


FIG. 6.7 – Influence du nombre de vitesses sur les profils (adimensionnés) d'un choc à Mach 8.

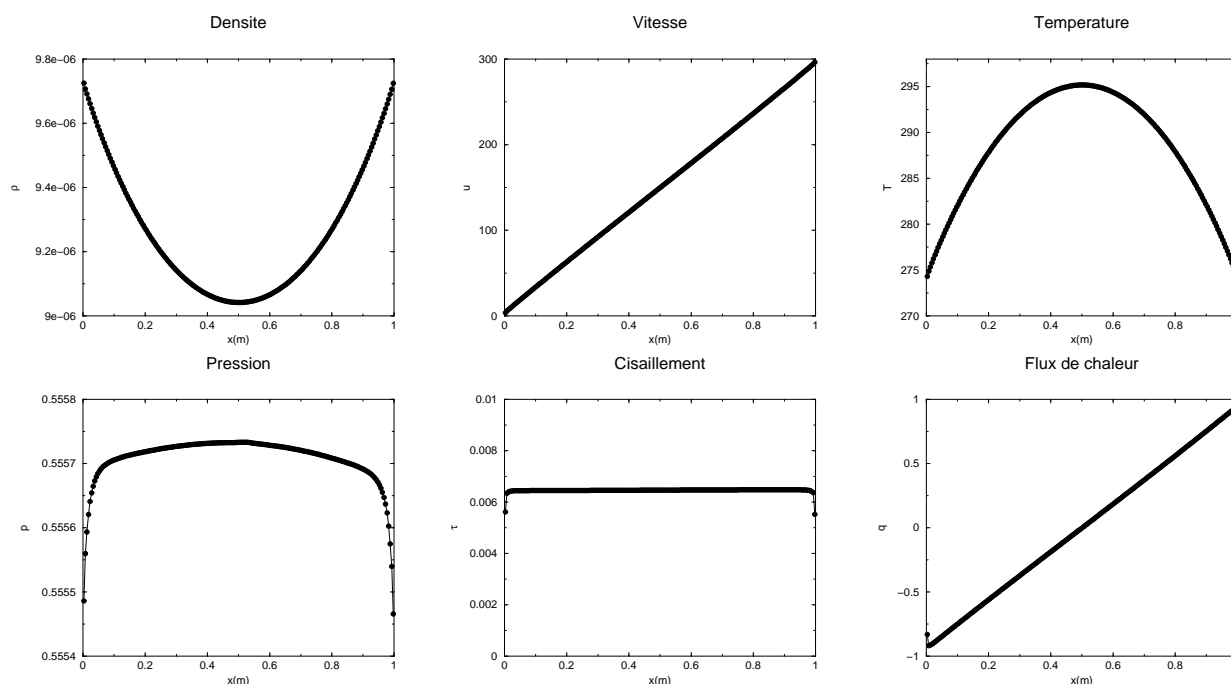


FIG. 6.8 – Profils dans l'écoulement de Couette 1D plan.

En considérant que la meilleure solution est celle obtenue avec $17 \times 17 \times 17$ vitesses discrètes (en gras dans le tableau ci-dessus), on peut faire les remarques suivantes. Tout d'abord, l'augmentation d'un nombre de vitesses en y (9 puis 17 puis 33) n'améliore ni les profils ni les sauts. En revanche, on constate sur les profils de pression que la couche de Knudsen est moins bien définie avec $7 \times 9 \times 7$ puisqu'elle s'étale plus. Si l'on augmente ensuite le nombre de vitesses en x et en y , (courbe $13 \times 17 \times 13$), on constate que les profils et les sauts s'améliorent, ce qui montre que la précision en x et en z est tout aussi importante que celle en y . En revanche, cela n'influe pas beaucoup sur l'épaisseur de la couche de Knudsen. Ensuite, on constate qu'avec $7 \times 10 \times 7$, les résultats sont légèrement meilleurs qu'avec $7 \times 9 \times 7$, sauf pour la température maximum. En outre le profil de pression montre une quasi disparition de la couche de Knudsen près de la paroi fixe. On peut interpréter ceci en notant qu'avec un nombre pair de vitesses en y , il n'y a pas de vitesse discrète égale à 0, ce qui fait qu'on perd de la précision dans les zones de faible vitesse, comme dans la couche de Knudsen.

Nous étudions alors l'influence des bornes du domaine, tout en gardant une densité de nombre de vitesses constante dans chaque direction. Pour cela, nous avons utilisé les résultats précédemment obtenus avec $13 \times 17 \times 13$ et un *cut* de 4. Ensuite, nous avons réduit les bornes du domaine en prenant un *cut* de 2, et en gardant la même densité de vitesses, ce qui impose de prendre $7 \times 13 \times 7$ points. Le résultat est présenté fig. 6.10 et dans le tableau ci-dessous :

	<i>cut</i> = 4	<i>cut</i> = 2
saut de température	14.3	17.0 (17.1 à droite)
saut de vitesse	12.5	11.6 (11.6 à droite)
température max.	533.4	670.4
densité min.	$7.7 \cdot 10^{-6}$	$8.3 \cdot 10^{-6}$

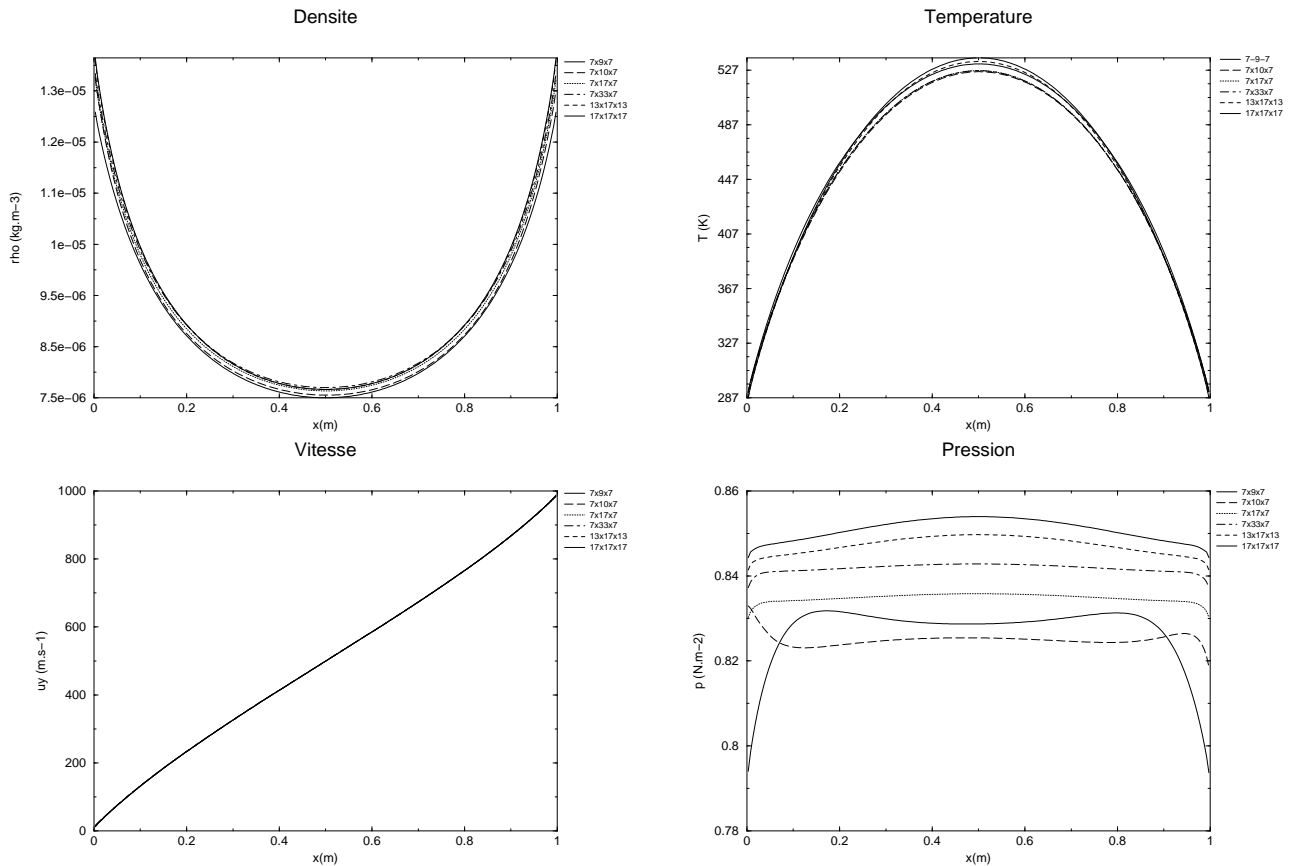


FIG. 6.9 – Influence du nombre de vitesses discrètes sur les profils .

On constate des profils de température et de densité très différents. Pour le *cut* de 2, le profil de densité est dissymétrique, et le minimum de densité surestimé. Le profil de température n'est pas tout à fait symétrique et le maximum est bien trop élevé. Ensuite, on note que les sauts sont sensiblement différents, mais surtout, on observe sur la courbe de pression que la couche de Knudsen n'est plus du tout discernable: elle est complètement mélangée à la zone centrale de l'écoulement. On comprend un peu ces phénomènes en observant les fonctions de distribution sur la figure 6.11. On voit que dans toute la moitié droite du gaz, les fonctions de distribution sont tronquées pour les grandes vitesses, à cause du *cut* trop petit, comme pour le choc stationnaire. Il semble clair que ceci conduit alors à la mauvaise représentation de l'écoulement.

6.2.2 Influence du maillage en espace

Nous avons utilisé pour cette étude le cas de l'écoulement de Couette, avec $13 \times 17 \times 13$ vitesses discrètes et un *cut* de 4. Afin d'étudier l'influence de la taille des mailles sur la couche de Knudsen, nous avons testé trois maillages différents: un maillage grossier de 50 mailles uniformes, le maillage de 200 mailles uniformes utilisé précédemment, et un maillage de 200 mailles raffinées aux parois.

Pour les deux maillages fins, les mailles ont une longueur inférieure au libre parcours moyen initial ($\frac{1}{10}\lambda$ pour le maillage uniforme, de $\frac{1}{100}\lambda$ à λ pour le maillage raffiné), alors que pour le maillage grossier, les mailles sont de l'ordre du libre parcours moyen. On s'attend donc à une couche de Knudsen un peu plus étalée pour le maillage grossier, ce que l'on constate effectivement (fig. 6.12), bien que la différence soit faible. On constate aussi que le maillage raffiné à la paroi (maille

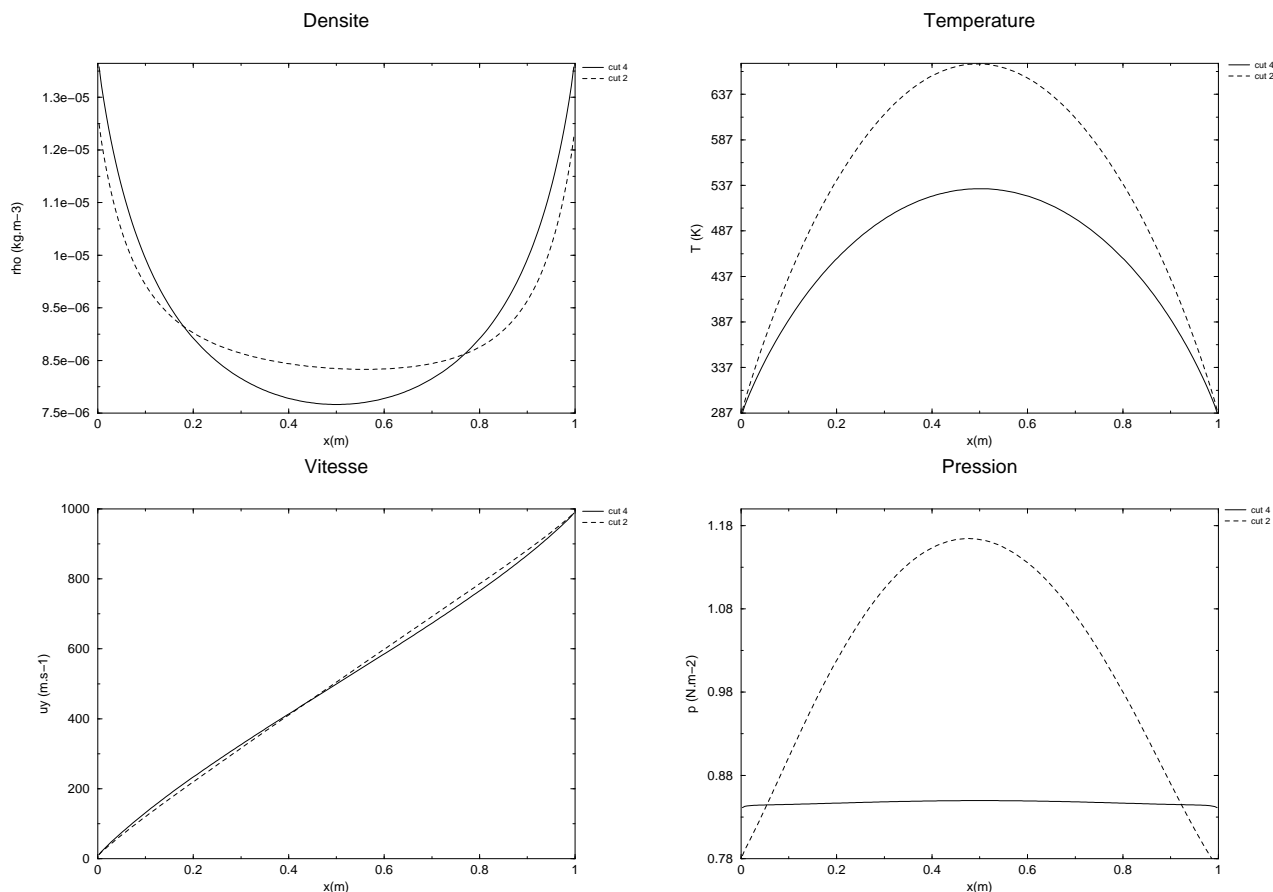


FIG. 6.10 – Influence des bornes sur les profils.

environ 12 fois plus petite que le maillage uniforme à 200 mailles) n'améliore pas sensiblement la représentation de la couche de Knudsen. En revanche, on voit dans le tableau ci-dessous que le raffinement du maillage améliore sensiblement les sauts de température et de vitesse :

	saut u	saut T
50 mailles	17.5	19.8
200 mailles	12.5	14.3
200 mailles (raffinées)	11.8	13.7

Enfin, comme le remarquent Perlat et Le Tallec [48], on a constaté l'existence d'une vitesse parasite u_x normale aux parois, qui semble due à une diffusion en espace. En effet, pour le maillage grossier, la vitesse normale à la paroi fixe atteint une valeur non négligeable par rapport à la vitesse parallèle. En raffinant les mailles à la paroi, ce phénomène s'atténue fortement comme on le voit dans le tableau ci-dessous, où l'on donne la vitesse normale u_x et la vitesse tangentielle u_y , ainsi que la différence relative en pourcentages :

	à gauche		à droite	
	u_y	u_x	u_y	u_x
50 mailles	17.5	-1.4 (8%)	982.4	2.9 (0.3%)
200 mailles	12.5	-0.5 (4%)	987.5	1.0 (0.1%)
200 mailles (raffinées)	11.8	-0.1 (1%)	988.1	0.2 (0.02%)

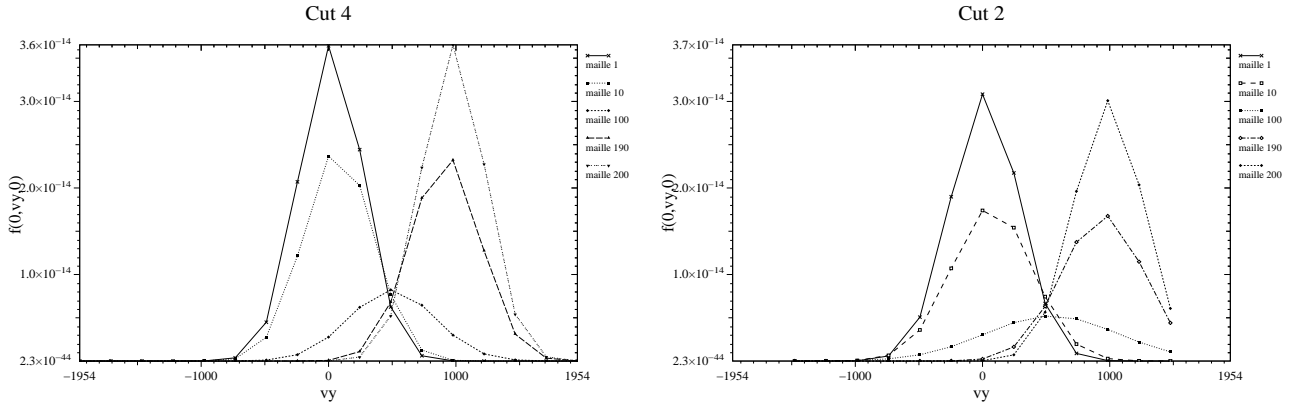


FIG. 6.11 – Influence des bornes sur les fonctions de distribution.

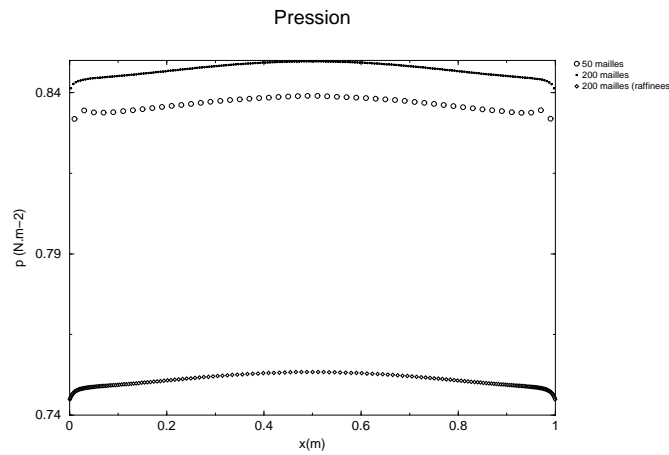


FIG. 6.12 – Influence du maillage sur la couche de Knudsen.

Il faut enfin noter que cette vitesse parasite est d'autant plus forte que la vitesse de la plaque u_w est élevée (à nombre de mailles constant), comme on le voit dans le tableau ci-dessous :

	u_y	u_x
$u_w = 300m.s^{-1}$	3.8	-0.04 (1%)
$u_w = 1000m.s^{-1}$	12.5	-0.5 (4%)

6.2.3 Comparaison ordre 1 - ordre 2 en espace

Nous montrons dans la figure (6.13) l'amélioration due aux schémas explicite et implicite d'ordre deux (en espace), pour un choc stationnaire à Mach 4. L'écoulement amont a les mêmes densité et température que dans le cas précédent. Le maillage est constitué de 200 mailles, mais nous utilisons à présent $9 \times 9 \times 9$ vitesses discrètes, et les bornes sont $[-2013, 2771] \times [-2392, 2392] \times [-2392, 2392]$. Ce test est motivé par le fait que les schémas d'ordre élevé qui ne résolvent pas le petit temps de relaxation (comme notre schéma implicite) peuvent dégénérer à un ordre inférieur. Ceci est par exemple étudié par Jin dans [30] pour les systèmes hyperboliques avec termes source raides. Cependant, cet inconvénient ne semble pas affecter nos résultats. Le schéma d'ordre un produit une dissipation numérique qui lisse trop les profils de densité, température et vitesse. Cette dissipation est clairement réduite par la discrétisation d'ordre deux en espace, ce qui donne des profils plus

raides. Notons que les résultats obtenus à la fois avec les schémas explicite et implicite d'ordre deux sont indifférentiables. Ceci prouve que le schéma implicite est réellement d'ordre deux en espace, et que la solution stationnaire est indépendante de la méthode de relaxation utilisée. Ces améliorations ont aussi été observées sur des calculs 2D.

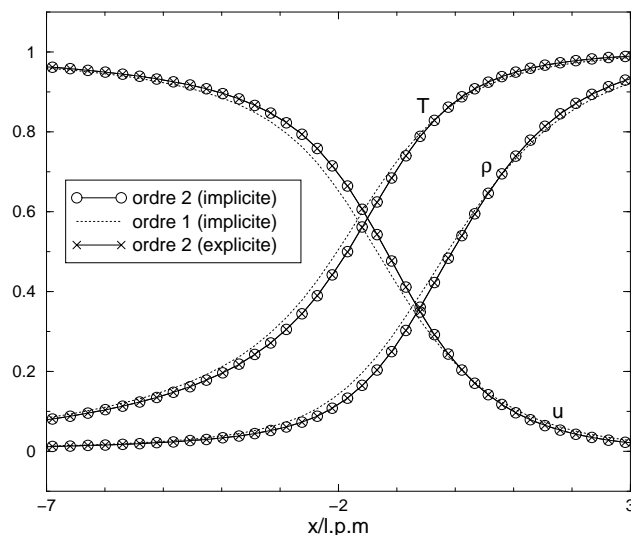


FIG. 6.13 – Comparaison entre les schémas d'ordre un et deux (pour un choc à Mach 4).

6.2.4 Comparaison des différents schémas en temps

Enfin, nous donnons quelques comparaisons entre les schémas explicite, implicite et semi-implicite, pour montrer que le schéma implicite est le mieux adapté aux calculs d'écoulements stationnaires. Le tableau suivant montre le temps CPU et le nombre d'itérations pour le schéma explicite et le schéma implicite linéarisé pour un choc stationnaire 1D à Mach 4 :

	iter.	CPU (s)
explicite	6000	16600
implicite	100	485

L'avantage du schéma implicite est indéniable. Bien que le schéma explicite évite de résoudre de grands systèmes linéaires (ce qui divise le coût d'une itération en temps par deux, comparé au schéma implicite), le nombre d'itérations nécessaires pour la convergence est si grand que le schéma implicite est plus efficace.

Dans la figure 6.14, nous donnons une comparaison de la convergence des schémas implicite et semi-implicite (où seul le terme de perte est implicite, voir section 5.3) pour un écoulement supersonique 2D plan autour d'un cylindre. Comme on l'a affirmé dans le chapitre précédent, le schéma semi-implicite est bien plus lent que le schéma implicite. En fait, le schéma implicite converge en environ 1000 itérations et 30 heures CPU, alors que le schéma semi-implicite nécessite 4000 itérations et 158 heures sur le même processeur. Notons qu'avec un calcul parallèle, le schéma implicite serait encore meilleur, puisque comme on l'a dit dans le chapitre 5, la linéarisation de $\mathcal{E}_{k,i,j}^n$ est complètement parallélisable.

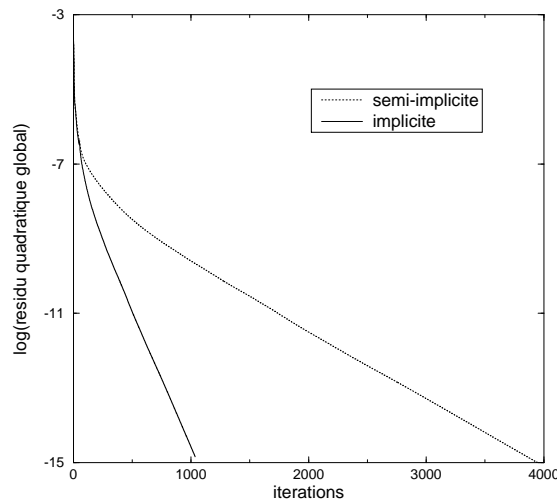


FIG. 6.14 – Courbe de convergence pour les schémas implicite et semi-implicite (écoulement 2D autour d'un cylindre).

6.3 Comparaison des schémas pour l'équation BGK axisymétrique

6.3.1 Etude sur un écoulement 1D

Nous utilisons ici un cas proposé et étudié numériquement par Sone et al. dans [60]. Il s'agit d'un écoulement de Couette cylindrique avec des conditions aux limites dites d'évaporation-condensation. On considère donc un gaz entre deux cylindres concentriques constitués d'une phase condensée du gaz. Chaque cylindre est maintenu à une température uniforme, le cylindre extérieur tourne à vitesse constante, et le cylindre intérieur est immobile. L'écoulement est alors uniforme en x et en φ et ne dépend que de r . L'intérêt de ce cas est qu'on peut le résoudre à la fois en 2D plan (dans le plan (y,z)) et en 1D axisymétrique (cf. figure 6.15). On peut donc valider les résultats axisymétriques en les comparant aux résultats de notre code 2D plan, ainsi qu'aux résultats de Sone et al. Les données de l'écoulement sont les suivantes :

gaz	argon
masse moléculaire (kg)	$0.663 \cdot 10^{-25}$
exposant ω de la loi de viscosité (modèle HS)	0.5
température de paroi (K)	1
vitesse cylindre intérieur (m/s)	0
vitesse cylindre extérieur (m/s)	6.1
pression cylindre intérieur (N/m^2)	1
pression cylindre extérieur (N/m^2)	1.1
rayon cylindre intérieur R_1 (m)	1
rayon cylindre extérieur R_2 (m)	2
nombre de Knudsen λ_1/R_1	$5 \cdot 10^{-2}$

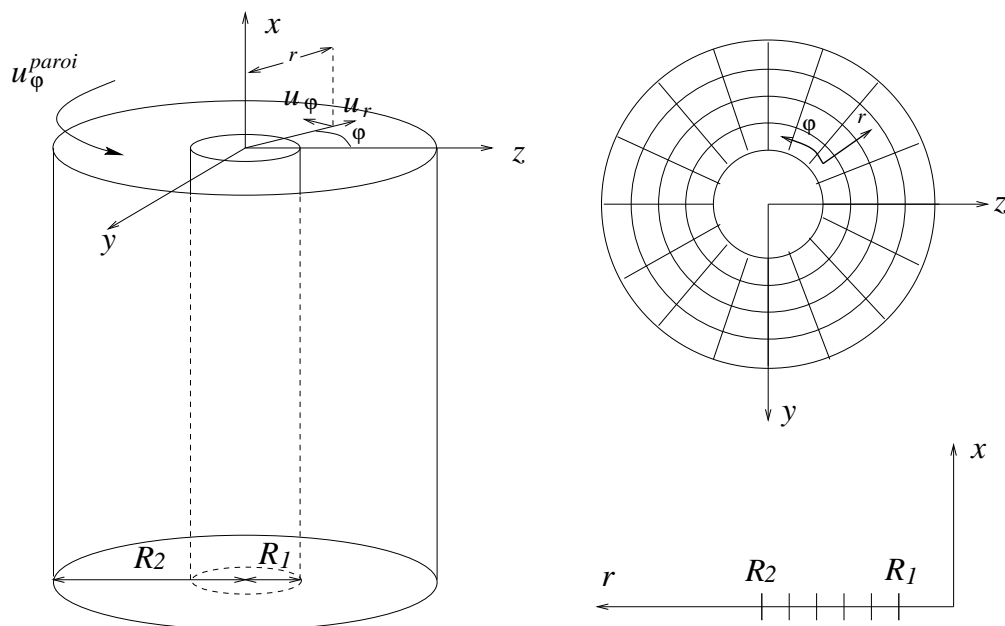


FIG. 6.15 – Géométrie de l'écoulement, et maillages en 2D plan et 1D axi.

Nous calculons d'abord l'écoulement avec le code 2D plan dans le plan (y, z) . Le maillage comporte 22 mailles dans la direction φ et 20 dans la direction r . Nous utilisons une grille de $9 \times 9 \times 9$ vitesses dans les direction (v_x, v_y, v_z) , de bornes $[-58, 58]^3$ données par $\pm 4\sqrt{RT_{paroi}}$, car la température ne varie pas beaucoup dans l'écoulement. Nous obtenons alors les profils de vitesse tangentielle u_φ , de vitesse radiale u_r , et de la température T en fonction du rayon, tracés dans la figure 6.16. On constate que ces profils sont identiques à ceux calculés par Sone et al. dans [60]. Nous considérons donc ces courbes comme des courbes de référence pour la suite de cette section. Notons que ces résultats ont été obtenus en 100 itérations et 19 min CPU sur un processeur de type 2.

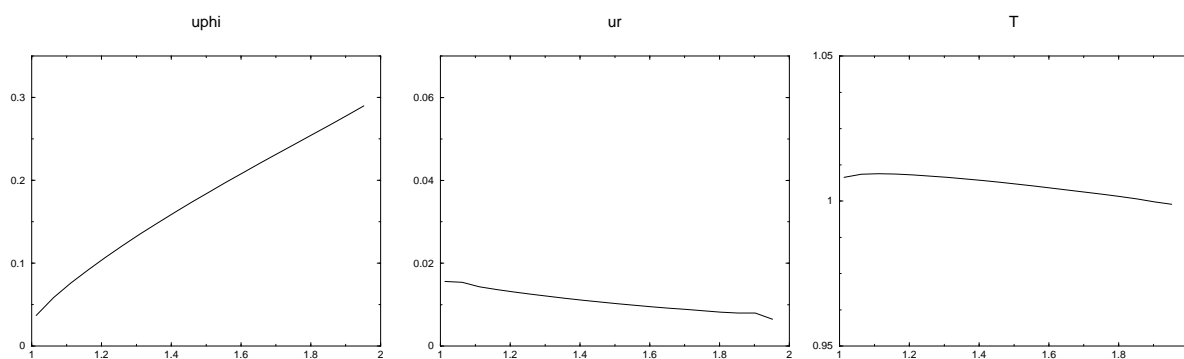


FIG. 6.16 – Profils pour l'écoulement de Couette cylindrique. Calcul plan 2D.

Nous comparons à présent les résultats obtenus pour la résolution de l'équation BGK axisymétrique avec les différents schémas présentés au chapitre 3. Nous utiliserons les abréviations suivantes

pour ces schémas :

- schémas pour l'équation de transport sous forme conservative
 - centré d'ordre 2 : CEC
 - centré d'ordre 2, modification trigonométrique : CEC-T
 - décentré d'ordre 1 : DEC
 - décentré d'ordre 1, modification trigonométrique : DEC-T
- schémas pour l'équation de transport sous forme non conservative
 - centré d'ordre 2 : CENC
 - centré d'ordre 2, modification trigonométrique : CENC-T
 - décentré d'ordre 1 : DENC
 - décentré d'ordre 1, modification trigonométrique : DENC-T
 - décentré d'ordre 2 : D2ENC
 - décentré d'ordre 2, modification trigonométrique : D2ENC-T

Rappelons dans le tableau ci-dessous les propriétés de tous ces schémas :

	écoulements const.	positivité	cons. $\rho, \rho u_x, E$	cons. ρu_r	entropie
CEC	OUI	<i>non</i>	<i>non</i>	<i>non</i>	<i>non</i>
CEC-T	OUI	<i>non</i>	OUI	OUI	<i>non</i>
DEC	<i>non</i>	OUI	OUI	<i>non</i>	<i>non</i>
DEC-T	OUI	OUI	OUI	<i>non</i>	OUI
CENC	OUI	<i>non</i>	<i>non</i>	<i>non</i>	<i>non</i>
CENC-T	OUI	<i>non</i>	OUI	<i>non</i>	<i>non</i>
DENC	OUI	OUI	<i>non</i>	<i>non</i>	<i>non</i>
DENC-T	<i>non</i>	OUI	OUI	<i>non</i>	<i>non</i>
D2ENC	OUI	<i>non</i>	<i>non</i>	<i>non</i>	<i>non</i>
D2ENC-T	OUI	<i>non</i>	<i>non</i>	<i>non</i>	<i>non</i>

Nous n'utilisons pas ici les schémas de type Larina et Rykov étudiés dans la section 3.4. Nous utilisons trois grilles de vitesse ayant les mêmes bornes, mais avec des nombres de points différents :

	v_x	ζ	ω
grille 1	9	6	18
grille 2	18	11	35
grille 3	18	11	60

Le nombre de points en ω est presque doublé à chaque fois, et la première grille a deux fois moins de points en v_x et en ζ que les autres. Nous cherchons à tester le comportement des différents schémas et à étudier la précision nécessaire pour approcher la dérivée de f en ω .

Pour la première grille, nous obtenons les trois profils tracés dans la figure 6.17. On constate tout d'abord que les schémas de discrétisation de l'équation conservative (CEC, CEC-T, DEC, DEC-T) sont meilleurs que les autres. Ensuite, il apparaît que les schémas CEC et CEC-T sont de loin les plus proches de la solution de référence. On observe le gain apporté par les modifications trigonométriques CEC-T et DEC-T des schémas CEC et DEC sur le profil de u_r . On constate aussi

un très mauvais profil de température obtenu avec la modification trigonométrique DENC-T du schéma DENC, sans doute dû au fait que les écoulements uniformes ne sont pas préservés par ce schéma. Hormis pour DENC-T, les profils de u_r pour les schémas de discrétisation de l'équation non conservative (CENC, CENC-T, DENC, D2ENC et D2ENC-T) sont surprenants, car u_r croît à l'approche du cylindre extérieur, à l'inverse de la solution de référence. Nous n'expliquons pas ce phénomène pour l'instant. Enfin, il est étonnant de constater le mauvais comportement du schéma D2ENC, puisque c'est pourtant celui utilisé par Sone et al. [60] pour obtenir la solution de référence. Il est possible que ces auteurs aient en fait utilisé un maillage en espace et en vitesse beaucoup plus fin, ce qui n'est pas indiqué dans leur article. En outre, ils utilisent une modification du schéma pour capturer la discontinuité en espace et en vitesse de la fonction de distribution, ce qui a peut être une influence sur le résultat.

Pour la deuxième grille, nous obtenons les profils tracés dans la figure 6.18. On constate que tous les profils se rapprochent de la solution de référence, et que la hiérarchie des schémas au niveau de la précision est la même que sur la première grille. On peut cependant noter que l'écart entre les schémas et leurs modifications trigonométriques a diminué. Cela était attendu, puisque la différence entre les deux version diminue avec $\Delta\omega$. On note enfin que les schémas de discrétisation de l'équation non conservative donnent toujours un très mauvais profil de u_r , en sens inverse de la solution de référence.

Pour la troisième grille, nous obtenons les profils tracés dans la figure 6.19. Le fait le plus important que l'on constate sur ce cas est l'incapacité du code à calculer l'écoulement avec les schémas centrés CEC, CEC-T, CENC et CENC-T. Au bout de quelques itérations, il y a trop de fonctions de distributions négatives, l'énergie interne devient négative, et le calcul s'arrête. On interprète ceci par le fait que ces schémas ne préservent pas la positivité, et que la discrétisation centrée d'un terme de convection génère des instabilités. Ceci devient sans doute apparent ici car on a beaucoup diminué le pas de discrétisation de la dérivée en ω . Les schémas d'ordre deux D2ENC et D2ENC-T fonctionnent encore car ils sont décentrés, et possèdent sans doute plus de stabilité. A part ce phénomène, la hiérarchie entre les autres schémas est la même que précédemment, et les meilleurs schémas semblent être DEC et DEC-T, qui sont très proches de la solution de référence. En outre, il semble que les schémas DENC, D2ENC et D2ENC-T ne convergent pas vers le bon profil de u_r , contrairement à la modification trigonométrique DENC-T.

En conclusion, on constate que la précision demandée en ω pour cet écoulement axisymétrique nécessite un grille beaucoup plus fine qu'en plan pour approcher les courbes de référence, excepté si l'on utilise les schémas centrés CEC et CEC-T qui sont d'ordre deux. Cependant, ces schémas précis ne semblent pas robustes. On constate aussi l'apport des modifications trigonométriques sur des grilles pas trop fines, même si ce gain est moins important que le gain de l'ordre deux par rapport à l'ordre un. En outre, la discrétisation de l'équation conservative semble bien plus précise que celle de l'équation non conservative. Finalement, il ressort de cette étude qu'en terme de précision et de robustesse, le meilleur schéma est le schéma décentré DEC-T de l'équation conservative avec modification trigonométrique, que nous avons introduit, ce qui n'est pas étonnant puisque c'est celui qui a le plus de propriétés. Le fait qu'il soit d'ordre un nécessite cependant une discrétisation fine en ω , qui risque d'être coûteuse.

Remarque 6.2. Nous n'avons pas constaté d'amélioration significative des résultats en testant une extension à l'ordre deux du schéma DEC-T (avec limiteur de pente de Yee). Cette étude mériterait cependant d'être approfondie.

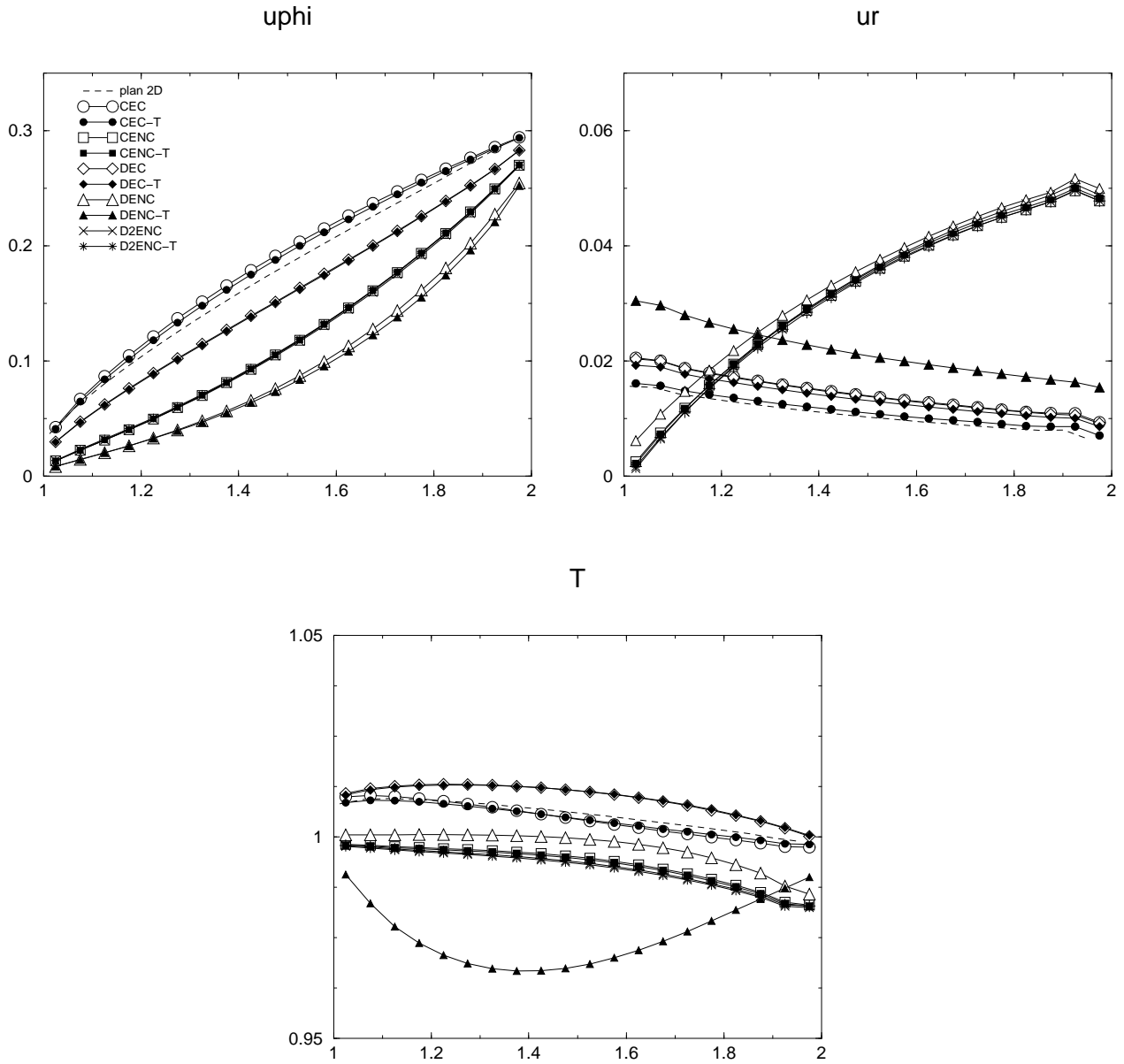


FIG. 6.17 – Comparaison des différents schémas axisymétriques pour l'écoulement de Couette cylindrique (grille 1). Noter que les courbes *CENC*, *CENC-T*, *D2ENC* et *D2ENC-T* sont quasiment confondues pour u_ϕ

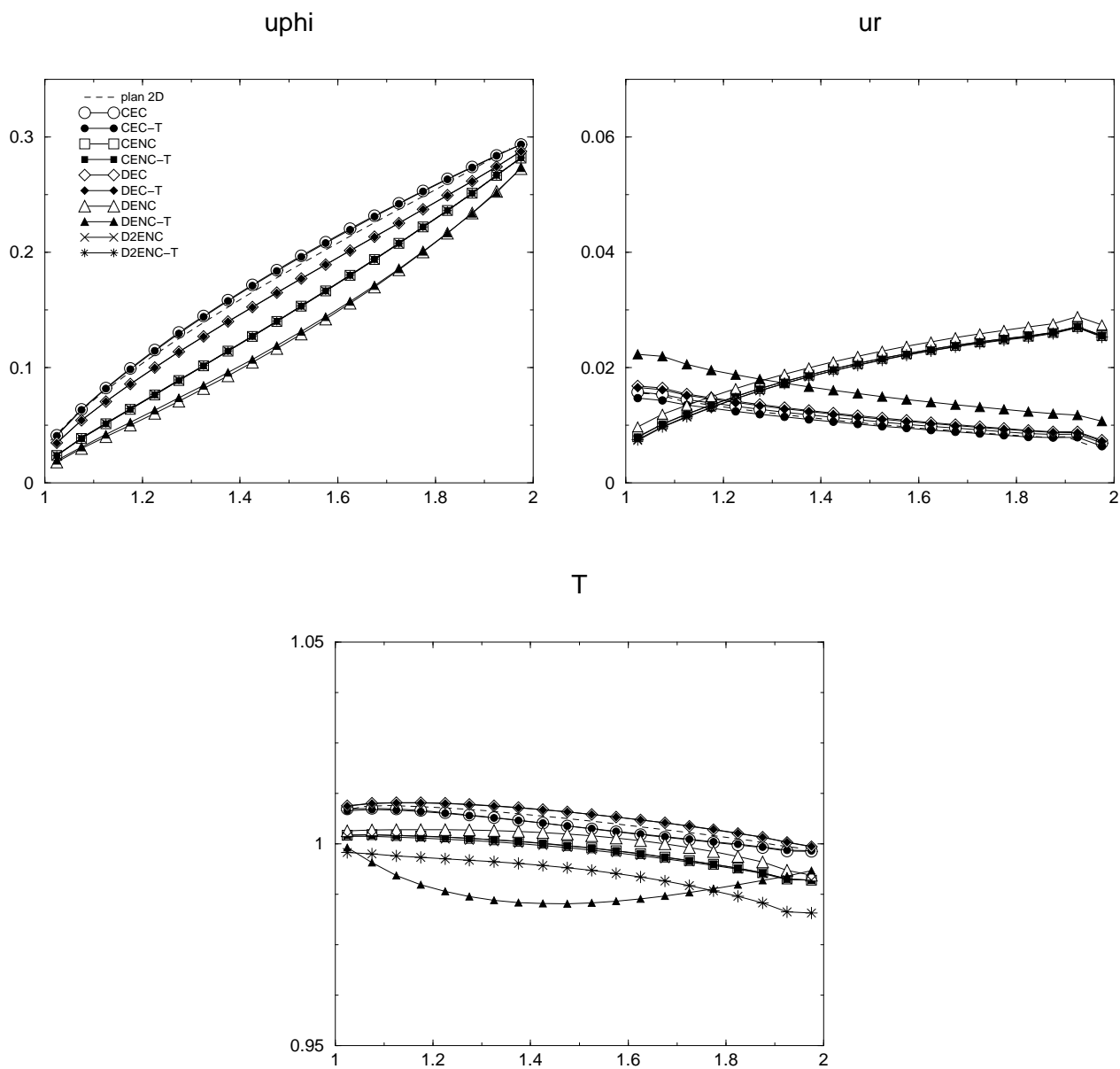


FIG. 6.18 – Comparaison des différents schémas axisymétriques pour l'écoulement de Couette cylindrique (grille 2).

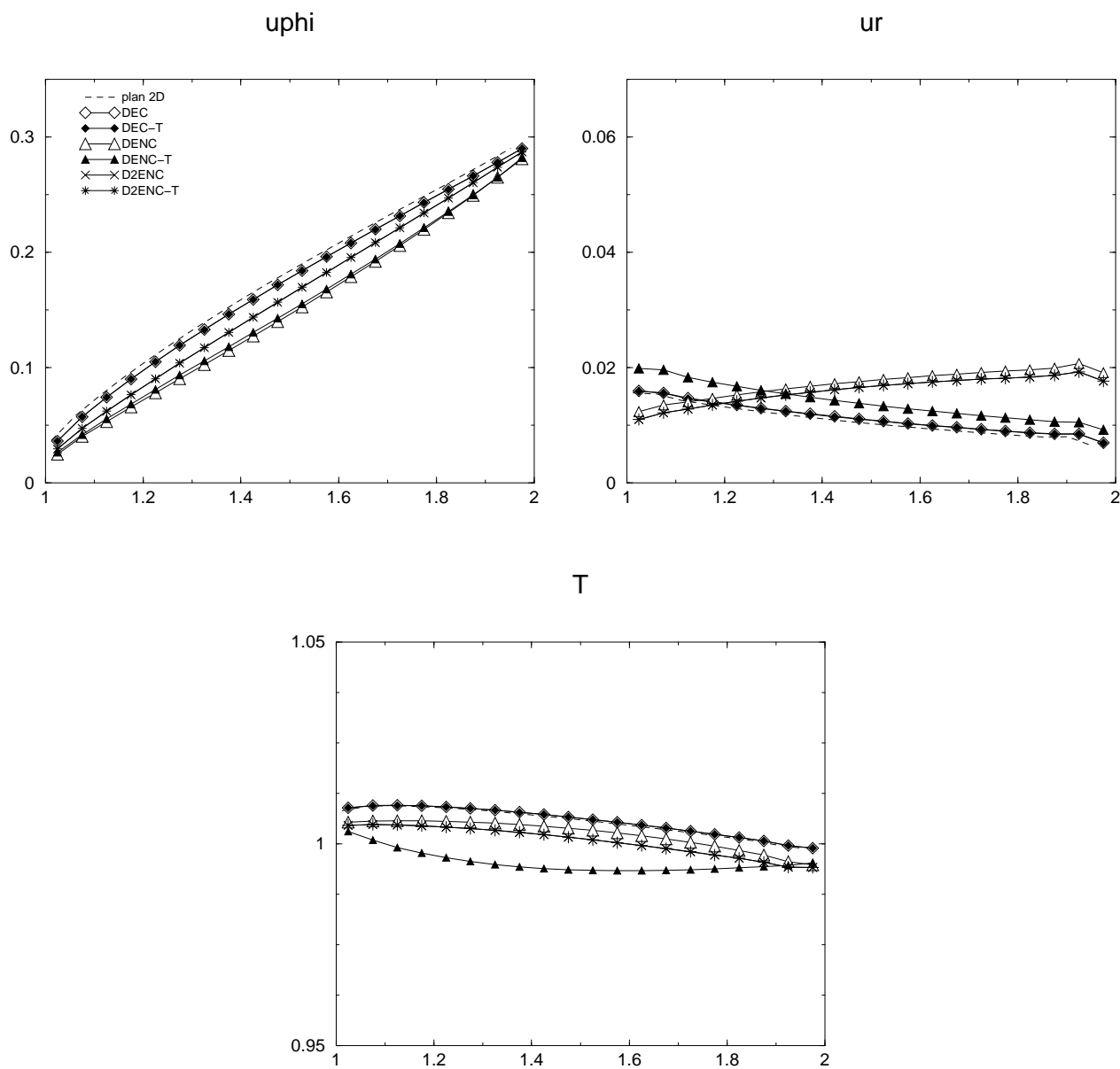


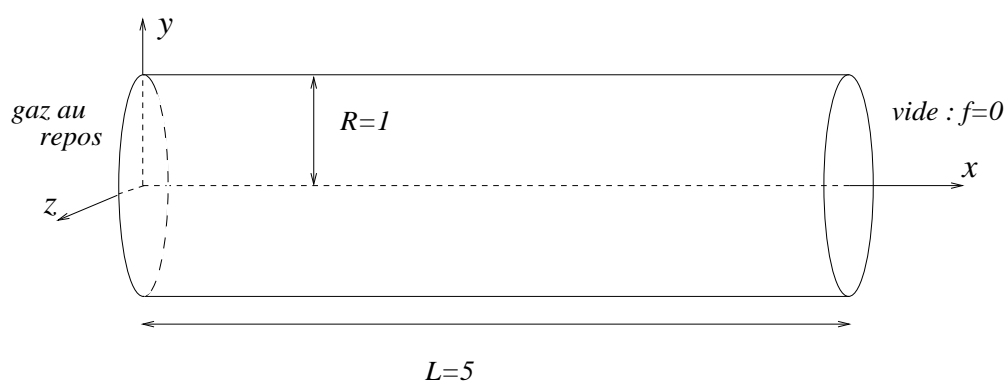
FIG. 6.19 – Comparaison des différents schémas axisymétriques pour l'écoulement de Couette cylindrique (grille 3).

6.3.2 Evaluation du gain par rapport au 2D plan

Les coûts de tous les calculs 1D axisymétriques sur le calcul précédent sont à peu près les mêmes. En moyenne on trouve 100 itérations et 150 s pour la grille 1, 110 itérations et 1400 s pour la grille 2, et 130 itérations et 3200 s pour la grille 3, avec un processeur de type 2. On constate que le gain en temps CPU par rapport au calcul 2D plan est d'un facteur 8 pour la grille 1, mais comme on l'a vu, seuls les schémas CEC et CEC-T donnent de bons résultats sur cette grille. Pour les grilles plus fines, le nombre de vitesses discrètes est tellement plus grand qu'en 2D plan que les calculs 1D axisymétriques sont plus chers. En effet, ces calculs sont de 1.2 fois (pour la grille 2) à 2.8 fois (pour la grille 3) plus lents que le 2D plan. L'intérêt d'un calcul axisymétrique pour cet écoulement n'apparaît pas évident.

6.3.3 Etude sur un écoulement 2D

Ce test a été proposé et étudié par Larina et Rykov dans [33]. On considère l'écoulement d'un gaz dans un tuyau de section circulaire défini dans la figure ci-dessous :



En entrée ($x = 0$), on introduit un gaz au repos (fonction de distribution maxwellienne). En sortie ($x = L$), le gaz est aspiré par le vide (fonction de distribution $f = 0$). Les caractéristiques de l'écoulement sont les suivantes :

gaz	argon
masse moléculaire (kg)	$0.663 \cdot 10^{-25}$
exposant de la loi de viscosité ω (modèle maxwellien)	1
température de paroi (K)	273.15
densité d'entrée (kg/m^3)	$0.350 \cdot 10^{-5}$
température d'entrée (K)	273.15
nombre de Knudsen λ_{entree}/R	$2 \cdot 10^{-2}$

Nous utilisons un maillage de 40×20 dans les directions x et r . Ce maillage est raffiné en sortie de tube pour tenir compte des forts gradients. La grille de vitesse comporte $11 \times 10 \times 11$ points dans les directions v_x , ζ et ω . Dans cet écoulement, la fonction de distribution est paire en ω , donc cette variable varie dans $[0, \pi]$. Les bornes en ζ sont $[0, 1400]$, calculées avec la température d'entrée (*i.e.* $\zeta_{max} = 4\sqrt{RT_{entree}}$). Les bornes en v_x sont plus grandes $[-1450, 1450]$ pour tenir compte de la vitesse élevée atteinte en sortie de tube.

Les schémas employés sont les mêmes que dans le paragraphe précédent. Seuls les schémas d'ordre un arrivent à converger. Pour les autres schémas (CEC, CEC-T, CENC, CENC-T, D2ENC,

D2ENC-T), le code s'arrête dès la dixième itération, car il apparaît des densités négatives. Avec le schéma décentré DENC-T, le code ne s'arrête pas, mais le calcul ne converge pas. Nous présentons dans la figure 6.20 les isovaleurs de densité, de température, et de nombre de Mach obtenues avec le schéma DEC-T. Dans la figure 6.21, nous présentons les profils de densité et de vitesse le long de l'axe de symétrie, obtenus avec les schémas DEC, DEC-T, DENC. Ces résultats sont très proches de ceux obtenus par Larina et Rykov, excepté en sortie de tube, où leurs profils sont plus raides, car leur maillage est sans doute plus raffiné que le nôtre. On constate qu'il n'y a quasiment aucune différence entre les schémas testés, excepté pour le schéma DEC sur la densité, qui est légèrement moins précis. On ne retrouve donc pas sur cet essai les différences importantes constatées sur l'écoulement 1D de Couette cylindrique. Notons enfin que pour tous ces schémas, le code converge en 430 itérations et 7h CPU sur un processeur de type 2.

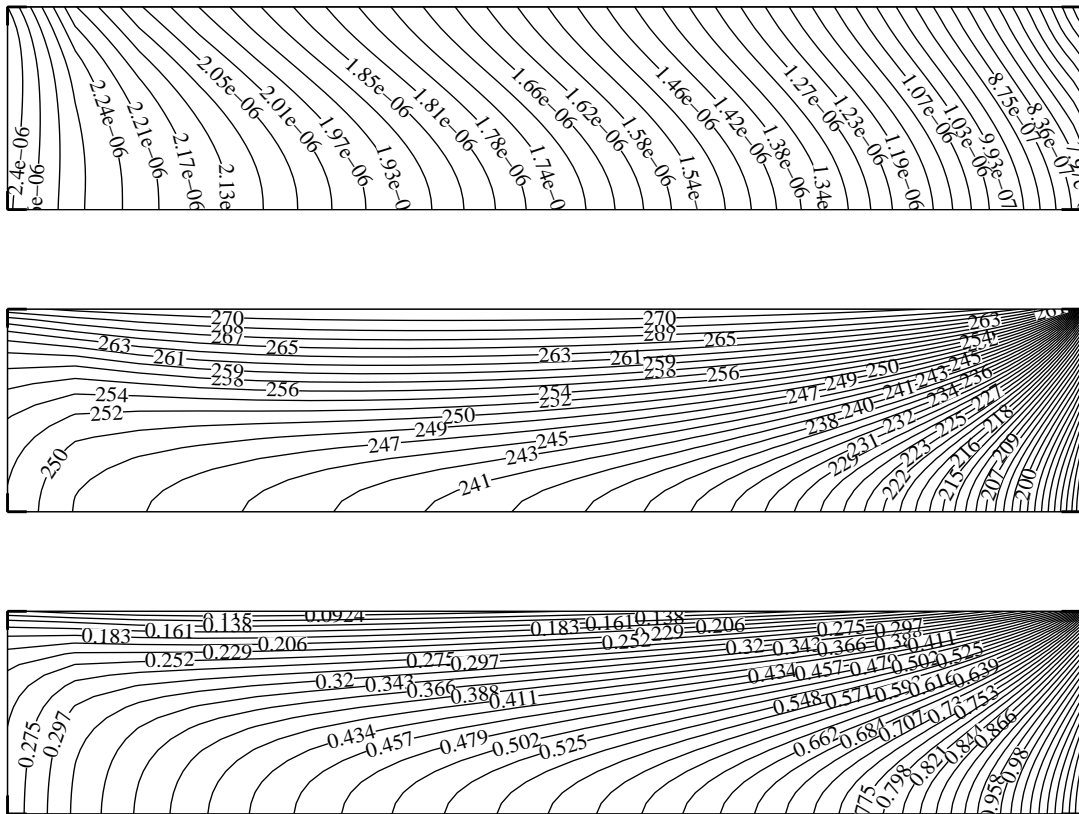


FIG. 6.20 – *Écoulement axisymétrique dans un tube avec sortie dans le vide pour le schéma DEC-T. Isovaleurs de densité (haut), température (milieu), et nombre de Mach (bas).*

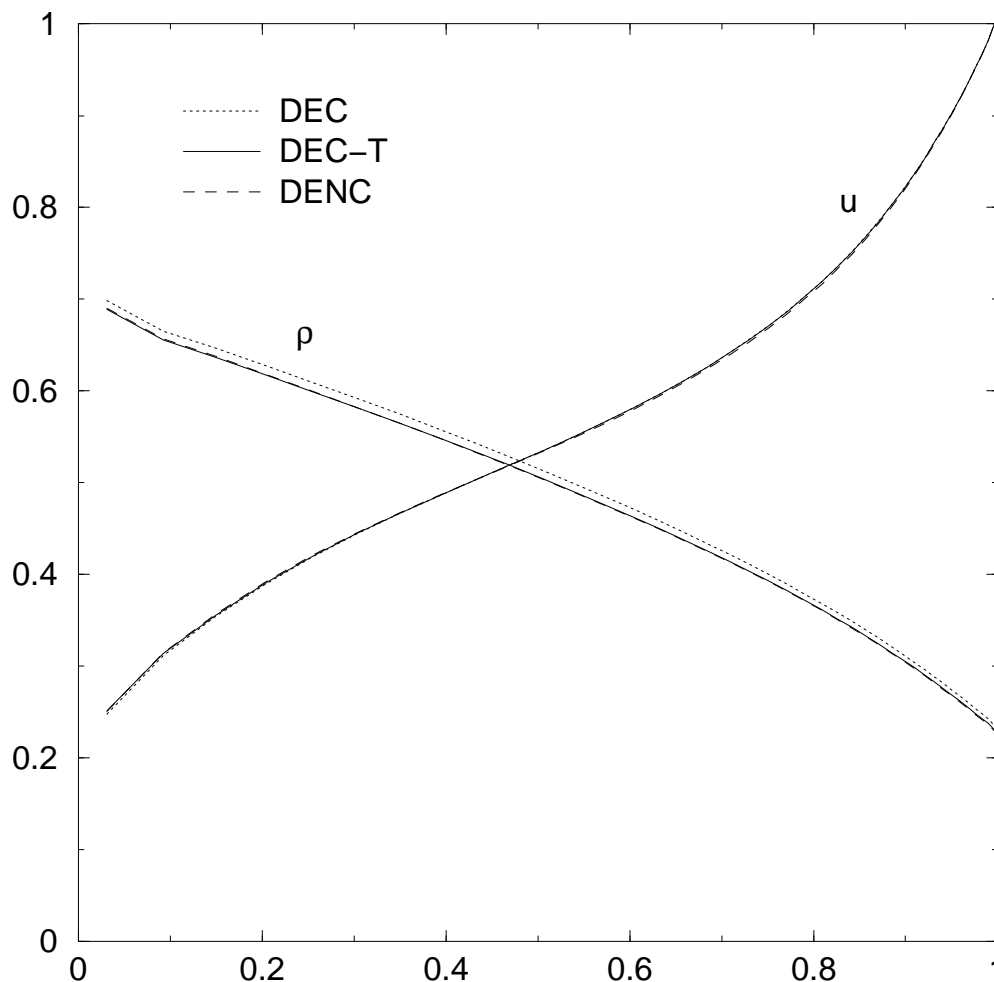


FIG. 6.21 – Profils de densité et de vitesse axiale pour différents schémas axisymétriques.

6.4 Validation physique

6.4.1 Comparaison avec des résultats expérimentaux

On peut définir l'épaisseur d'un choc, pour la densité par exemple, par la longueur

$$L_c = \frac{\rho_D - \rho_G}{\left(\frac{d\rho}{dx}\right)_{max}},$$

où ρ_D et ρ_G sont les densités en aval et en amont du choc (voir figure 6.22). Il est connu que l'épaisseur d'un choc dépend essentiellement de nombre de Mach amont. Pour un nombre de Mach croissant de 1 à 4, l'épaisseur du choc diminue. Elle atteint un minimum à Mach 4, puis elle croît quand le Mach augmente.

Nous montrons dans la figure 6.23 une comparaison de l'épaisseur de choc (ou plutôt l'inverse de l'épaisseur) donnée par notre méthode pour différents nombres de Mach, avec les données expérimentales données par Alsmeyer [1] et avec des résultats Navier-Stokes. On constate que notre méthode donne de très bons résultats, au moins jusqu'à Mach 7. Par comparaison, les équations de Navier-Stokes donnent de mauvais résultats dès Mach 2, avec une erreur de près de 50% à partir de Mach 4.

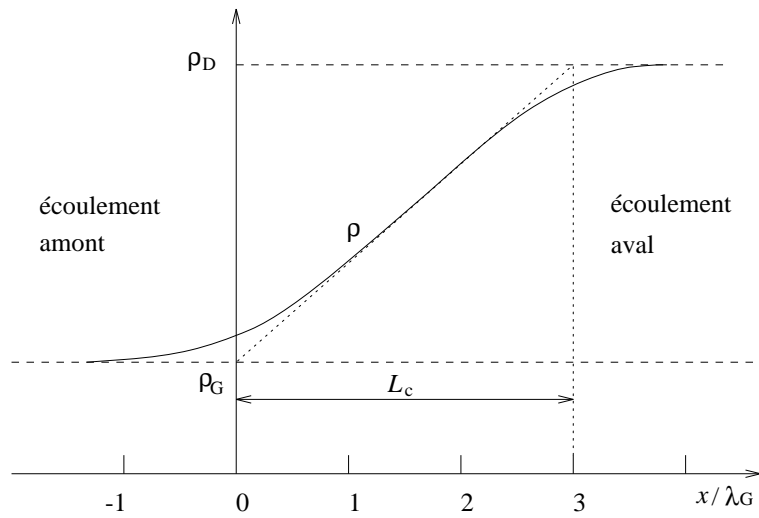


FIG. 6.22 – *Épaisseur d'un choc basée sur le profil de densité.*

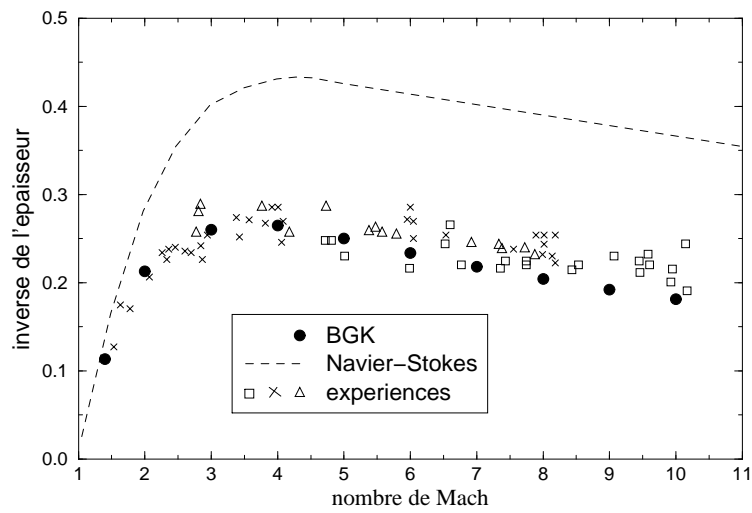


FIG. 6.23 – *Comparaison de BGK, Navier-Stokes, et des données expérimentales, pour l'inverse de l'épaisseur de choc dans de l'argon.*

6.4.2 Comparaison avec DSMC, NS, et autres méthodes

Rampe de compression

Dans ce test nous comparons notre méthode pour l'équation BGK avec la méthode DSMC (avec le code de J.-C. Lengrand [35]) et les équations de Navier-Stokes (résolues par le code ARES du CEA-CESTA sans condition de glissement). Nous étudions l'écoulement supersonique autour d'une plaque plane de 5 cm suivie d'une rampe de compression inclinée de 10° . Les données de l'écoulement sont les suivantes :

gaz	air
masse moléculaire (kg)	$4.815 \cdot 10^{-26}$
exposant ω de la loi de viscosité (modèle VHS)	0.77
densité amont ($kg.m^{-3}$)	$1.288 \cdot 10^{-4}$
température amont (K)	72.2
Mach amont	3.67
température de paroi (K)	290
Knudsen basé sur la longueur de la plaque	$6.7 \cdot 10^{-3}$

Pour les trois méthodes, nous utilisons le même maillage constitué de 70×70 mailles. La grille de vitesse pour le modèle BGK discret comporte $13 \times 11 \times 11$ vitesses avec les bornes $[-1500, 1500] \times [-1200, 1200] \times [-1200, 1200]$. Les paramètres utilisés pour DSMC sont les suivants :

nombre de molécules par maille	20
nombre de sondages	2300
temps entre deux collisions (s)	$5 \cdot 10^{-7}$
temps de stationnarisation (s)	0.0006 (soit 1200 it.)
temps maximum de simulation (s)	0.00407 (soit 8000 it.)

Pour le modèle BGK, le calcul a nécessité 260 itérations et 42 heures CPU. Pour DSMC le calcul converge en 46 heures CPU sur le même processeur de type 2. Le calcul Navier-Stokes est évidemment beaucoup moins coûteux, puisqu'il nécessite moins de 10 min CPU.

Les isovaleurs de densité et de température sont tracées dans la figure 6.24 pour les trois méthodes. On constate que les équations de Navier-Stokes donnent de très mauvais résultats en début de plaque, ainsi qu'au niveau du choc. Ceci se comprend en regardant le nombre de Knudsen local (voir Eq. (1.7)) : il vaut 0.13 en début de plaque, ce qui est au delà de la zone de validité pour les équations de Navier-Stokes. En revanche, les résultats obtenus avec BGK et DSMC sont très proches, comme on peut le voir plus clairement dans les figures 6.25 et 6.26. Ces figures représentent la distribution de densité et de température le long de trois lignes verticales. On peut seulement noter une différence près de la paroi pour la température (Navier-Stokes semble ici légèrement meilleur, sans doute à cause du mauvais nombre de Prandtl de BGK), ainsi qu'une légère différence au niveau du choc. Dans l'ensemble, les différences entre BGK et DSMC sont toutefois bien moindres que celles avec Navier-Stokes. On peut aussi remarquer le bruit dans les résultats de DSMC, induits par la nature stochastique de la méthode. De plus, on observe un défaut sur les résultats de DSMC en sortie de domaine près de la rampe. Ceci est sans doute dû aux conditions aux limites de cette méthode (on impose une condition de sortie dans le vide, ce qui n'est pas très réaliste). Comme on l'a noté section 6.1.3, bien que l'on donne ici les temps CPU des méthodes DSMC et BGK, une comparaison "honnête" n'est pas facile, car leurs critères de convergence sont très différents. Par

exemple, demander plus de sondages pour diminuer le bruit statistique peut faire accroître de façon considérable le temps CPU de DSMC.

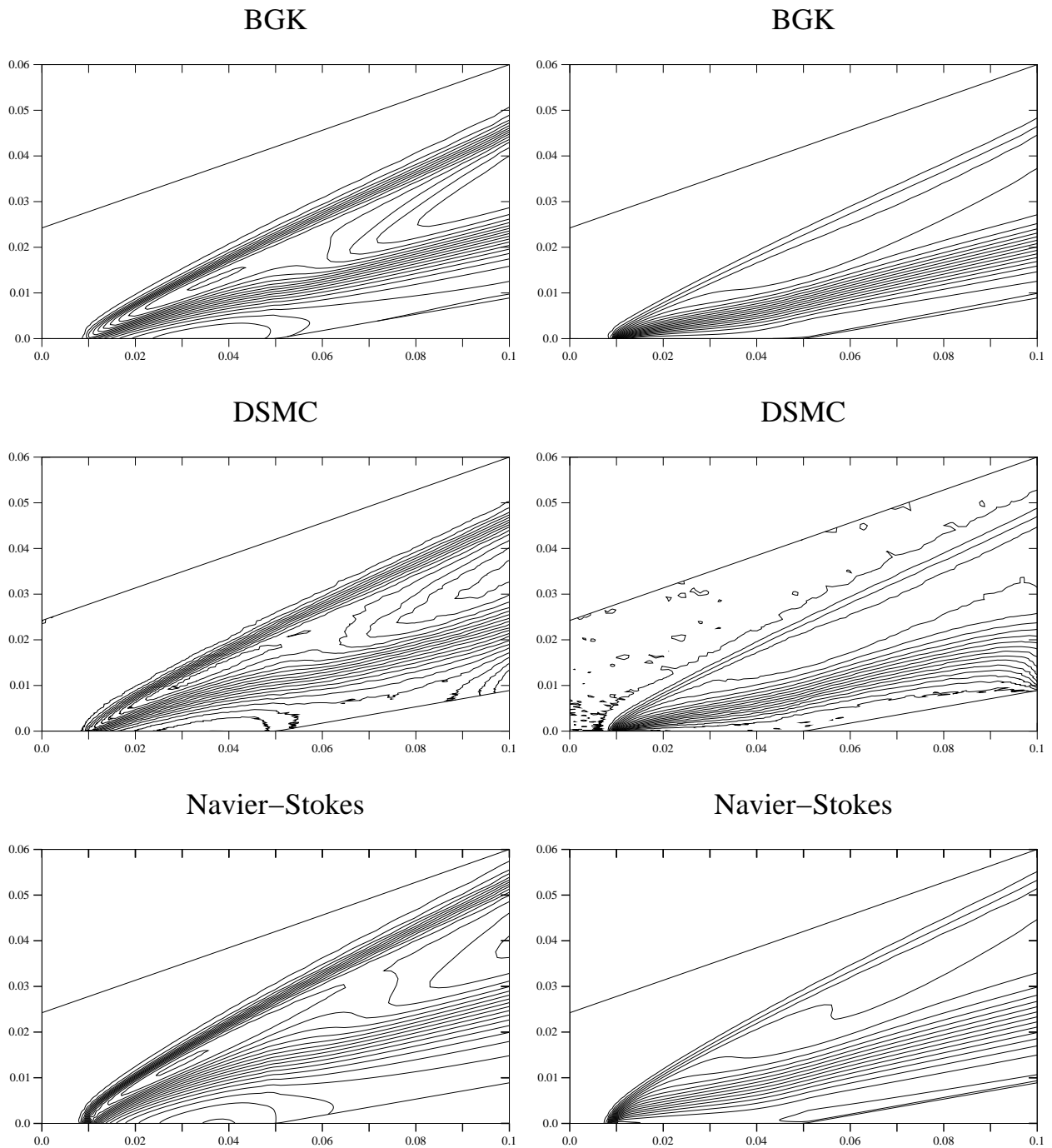


FIG. 6.24 – Rampe de compression : densité (gauche) et température (droite) pour BGK, DSMC, et Navier-Stokes.

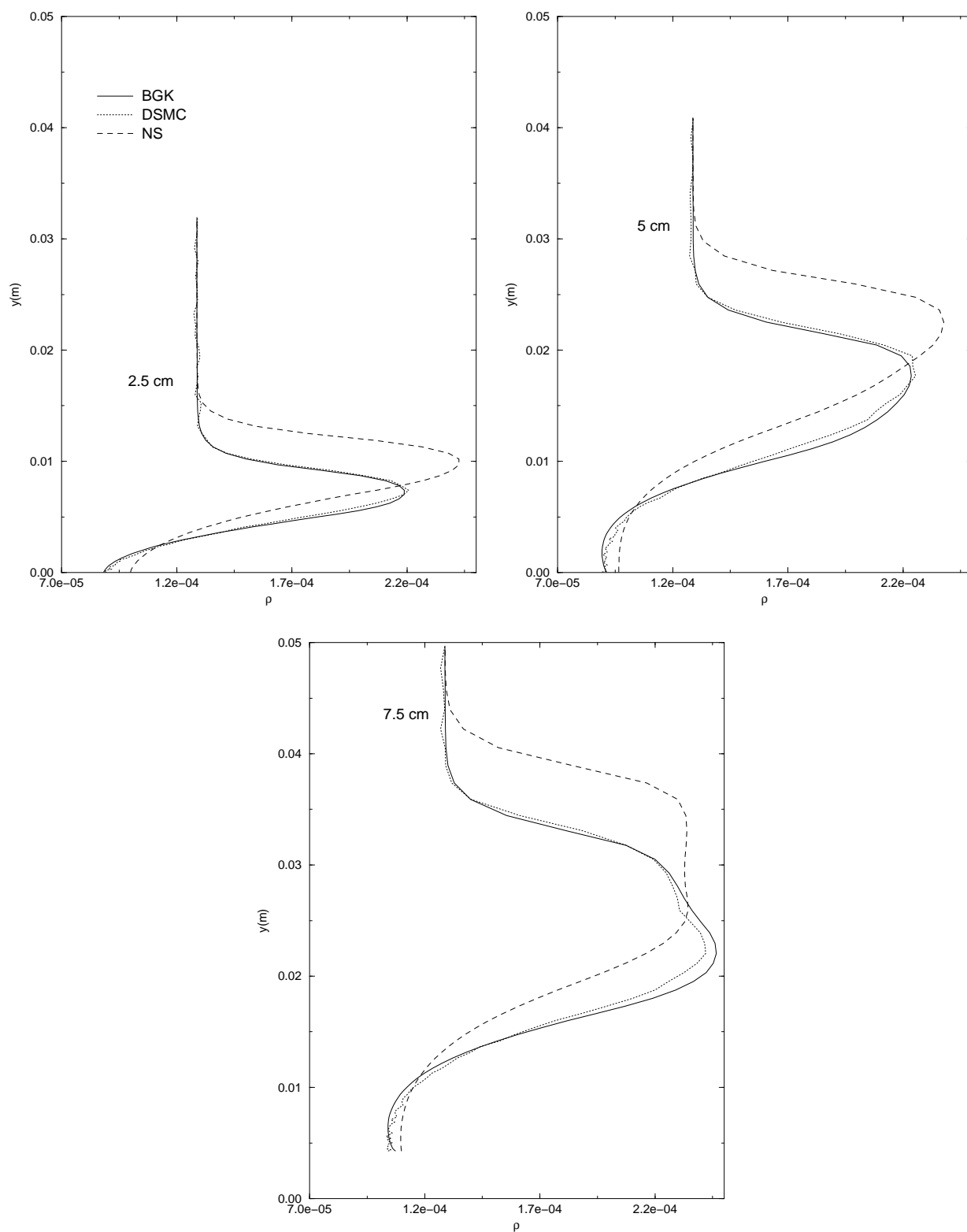


FIG. 6.25 – Profils de densité le long de trois lignes verticales ($x = 2.5, 5, 7.5$ cm), pour BGK, DSMC, et Navier-Stokes.

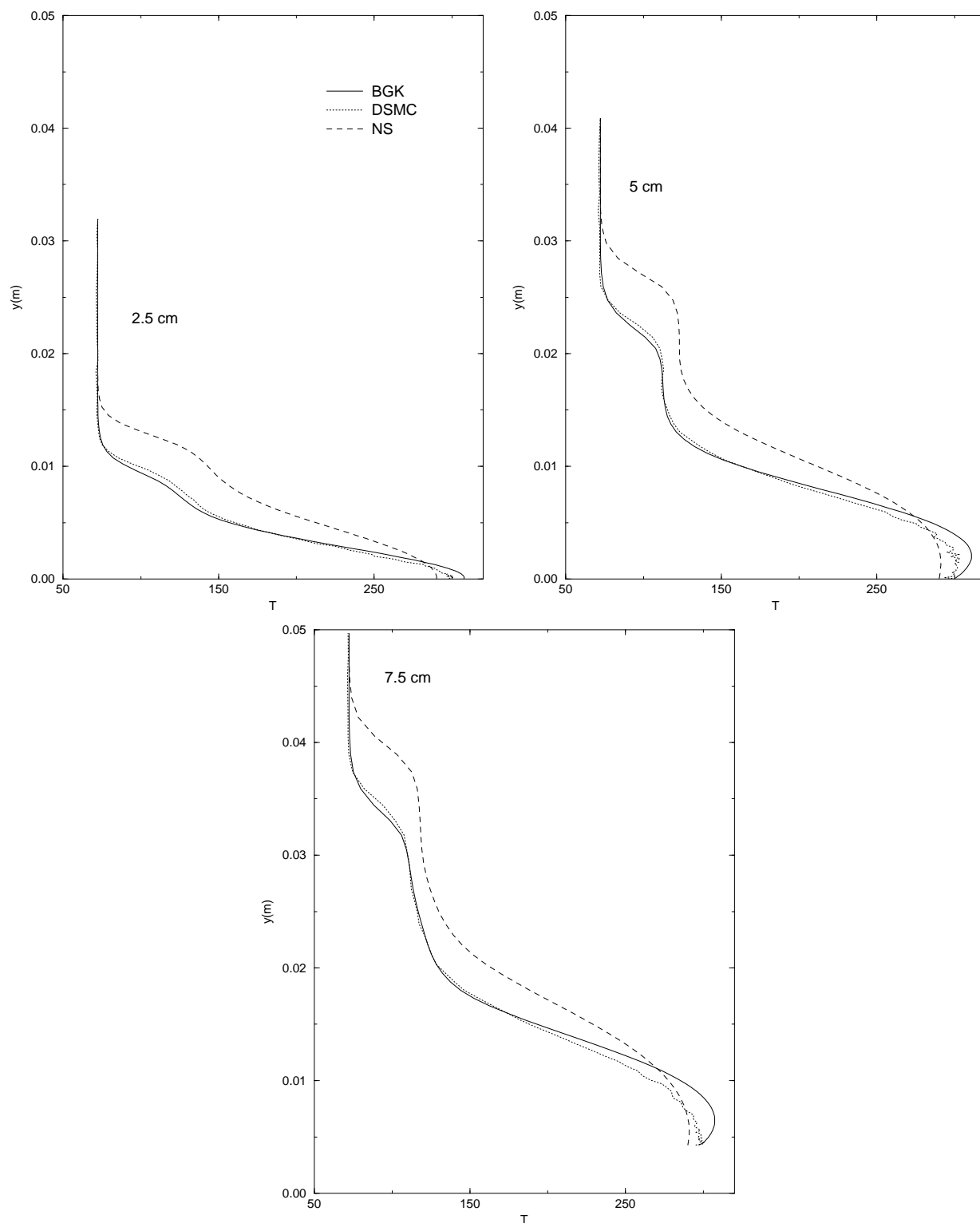


FIG. 6.26 – Profils de température le long de trois lignes verticales ($x = 2.5 - 5 - 7.5$ cm), pour BGK, DSMC, et Navier-Stokes.

Écoulement autour d'une sphère sans incidence

Ce test a pour but de compléter la validation de notre méthode de résolution de l'équation BGK axisymétrique. En effet la sphère a une géométrie de révolution, et un écoulement sans incidence autour d'un tel corps est nécessairement axisymétrique. Nous considérons un écoulement dont les caractéristiques suivantes correspondent à une altitude de 90 *km* :

gaz	argon
masse moléculaire (<i>kg</i>)	$0.663 \cdot 10^{-25}$
exposant ω de la loi de viscosité (modèle VHS)	0.81
température de paroi (<i>K</i>)	293
densité amont ($kg \cdot m^{-3}$)	$0.317 \cdot 10^{-5}$
température amont (<i>K</i>)	249
Mach amont	5
rayon <i>R</i> de la sphère (<i>m</i>)	0.1
Knudsen amont (basé sur <i>R</i>)	0.236
Knudsen au nez (basé sur <i>R</i>)	0.1

Pour tester notre méthode, nous utilisons un maillage de 60×50 mailles dans les directions tangentes et normales à la sphère. Le domaine de calcul est restreint à la partie avant de la sphère, car nous négligeons l'effet de l'écoulement en arrière du corps. La grille de vitesses comporte $11 \times 9 \times 21$ points dans les directions v_x, ζ, ω , la variable ω variant dans $[0, \pi]$ puisque la fonction de distribution est nécessairement paire en ω . Les bornes de la grille pour v_x et ζ sont $[-2300, 2300] \times [0, 2000]$, estimées par un calcul Navier-Stokes. La discrétisation du terme de transport axisymétrique utilise le schéma DEC-T (cf. page 101) qui est celui possédant les propriétés de positivité, de conservation et d'entropie. Les test précédents (section 6.3) ont en effet montré qu'il est le plus robuste et le plus précis. Le calcul converge en 137 itérations et 31 *h* CPU sur un processeur de type 2.

Nous avons effectué le même calcul avec le code DSMC, en utilisant le même maillage. Remarquons que ce maillage n'est pas suffisamment raffiné car la taille des mailles est supérieure au libre parcours moyen amont et au libre parcours moyen estimé à la paroi. Il ne faut donc pas s'attendre à des résultats très précis avec cette méthode. Les paramètres propres à DSMC que nous avons choisis sont les suivants :

nombre de molécules par maille	20
temps entre deux collisions (<i>s</i>)	$2 \cdot 10^{-6}$
temps de stationnarisation (<i>s</i>)	$2.72 \cdot 10^{-4}$ (soit 136 it.)
nombre de sondages	500
temps maximum de simulation (<i>s</i>)	$3 \cdot 10^{-3}$ (soit 1500 it.)

Le temps maximum de simulation est atteint en 1631 itérations et 18 *h* CPU sur le même processeur que pour la méthode BGK. Le temps CPU est donc ici inférieur au temps CPU nécessaire pour notre méthode, mais cela ne signifie pas grand chose, puisqu'il faudrait un temps bien supérieur pour avoir des résultats DSMC plus précis. Nous avons aussi effectué un calcul Navier-Stokes sans condition de glissement qui converge en quelques minutes.

Les résultats obtenus sont présentés dans les figures 6.27, 6.28, 6.29, et 6.30. Sur les isovaleurs de température et de pression (figure 6.27), on constate que les résultats DSMC sont très bruités, ce qui était prévisible. On voit aussi que les isovaleurs pour BGK oscillent un peu en fin de choc. Ce phénomène est aussi présent dans les résultats Navier-Stokes (bien que de moindre ampleur). C'est un problème classique en maillage structuré qui est dû à la viscosité numérique du schéma,

car les lignes de courant ne sont pas alignées avec les lignes de maillage. Cependant, les isovaleurs des méthodes BGK et DSMC sont assez semblables, ce qui n'est pas le cas des isovaleurs de Navier-Stokes.

Sur les figures 6.28 et 6.29, on a tracé des coupes de densité, de température et de pression en fonction du rayon r le long de deux ligne perpendiculaires à la paroi. L'une est la ligne d'arrêt (à $x = 0$), l'autre est inclinée de 45° par rapport à l'axe. Pour la ligne d'arrêt, les courbes DSMC et BGK sont assez proches, excepté au niveau de la paroi où il y a un écart de près de 20% sur la température et la densité. On constate cependant que l'écart entre DSMC et Navier-Stokes est beaucoup plus important, surtout pour la température. Pour la ligne à 45° , les courbes DSMC et BGK sont remarquablement proches. Au contraire, il y a une grande différence entre Navier-Stokes et DSMC, ce qui n'est pas étonnant puisqu'on trouve un nombre de Knudsen local de 0.6 dans cette zone (cf. section 1.1.2). On note par exemple un écart de près de 50% sur la température dans le choc ($r = 0.025 m$).

Enfin on compare sur la figure 6.30 la variation des coefficients aérodynamiques le long de la paroi en fonction de x . Ces quantités adimensionnées sont définies par

$$\begin{aligned} C_f &= \frac{(\sigma \cdot n)\tau}{0.5\rho_\infty u_\infty^2} && \text{coefficient de frottement,} \\ C_p &= \frac{(\sigma \cdot n)n}{0.5\rho_\infty u_\infty^2} && \text{coefficient de pression,} \\ C_h &= \frac{q \cdot n}{0.5\rho_\infty u_\infty^3} && \text{flux de chaleur normal,} \end{aligned}$$

où $\sigma = \langle (v - u) \otimes (v - u) f \rangle$ est le tenseur des contraintes, $q = \langle \frac{1}{2}(v - u)|v - u|^2 f \rangle$ est le flux de chaleur, et n et τ sont les vecteurs normal et tangent à la paroi. Ces courbes sont assez bruitées pour DSMC, ce qui est normal, mais aussi pour BGK et Navier-Stokes. Ceci est sans doute dû là encore au mauvais maillage.

Pour le coefficient de frottement, on constate un bon accord entre BGK et DSMC, excepté au dessus du point d'arrêt ($x = 0.02$). Dans cette zone, Navier-Stokes paraît meilleur, ce qui est étonnant au vu du nombre de Knudsen pourtant élevé. Après cette zone ($x \geq 0.03$), Navier-Stokes est bien moins proche de DSMC que BGK. Pour le coefficient de pression, les résultats BGK, DSMC et Navier-Stokes sont très proches, sauf en fin de corps ($x \geq 0.06$) où Navier-Stokes est moins bon. Pour le flux de chaleur, Navier-Stokes est meilleur que BGK près du point d'arrêt ($x \leq 0.03$), ce qui est sans doute dû au nombre de Prandtl, puis BGK est meilleur en fin de corps ($x \geq 0.03$).

Remarque 6.3. On peut là aussi avoir une idée du gain que l'on obtient en faisant un calcul 2D axisymétrique au lieu d'un calcul 3D. Nous avons calculé l'écoulement précédent en géométrie plane, avec le même nombre de mailles, et $11 \times 11 \times 11$ vitesses discrètes. On a donc à peu près deux fois moins de vitesses discrètes qu'en axisymétrique, ce qui est normal puisque le calcul axisymétrique nécessite une discrétisation plus fine en raison du gradient de vitesse. Nous avons cependant dû agrandir le domaine de calcul spatial, car en géométrie plane, le choc s'établit plus loin de la paroi. Le calcul converge alors en 194 itérations et 22 h CPU sur un processeur de type 2, ce qui est évidemment plus rapide que le calcul axisymétrique. Pour un calcul 3D, supposons que l'on ait 50 mailles dans la direction z : comme nos algorithmes sont de complexité linéaire, on peut supposer que le coût d'un calcul 3D serait donc 50 fois le coût du 2D plan, soit 1100 h CPU, contre 31 h CPU pour avoir le même résultat en axisymétrique. Il est clair que malgré le coût de la discrétisation fine en ω , le calcul axisymétrique est beaucoup moins coûteux que le calcul 3D.

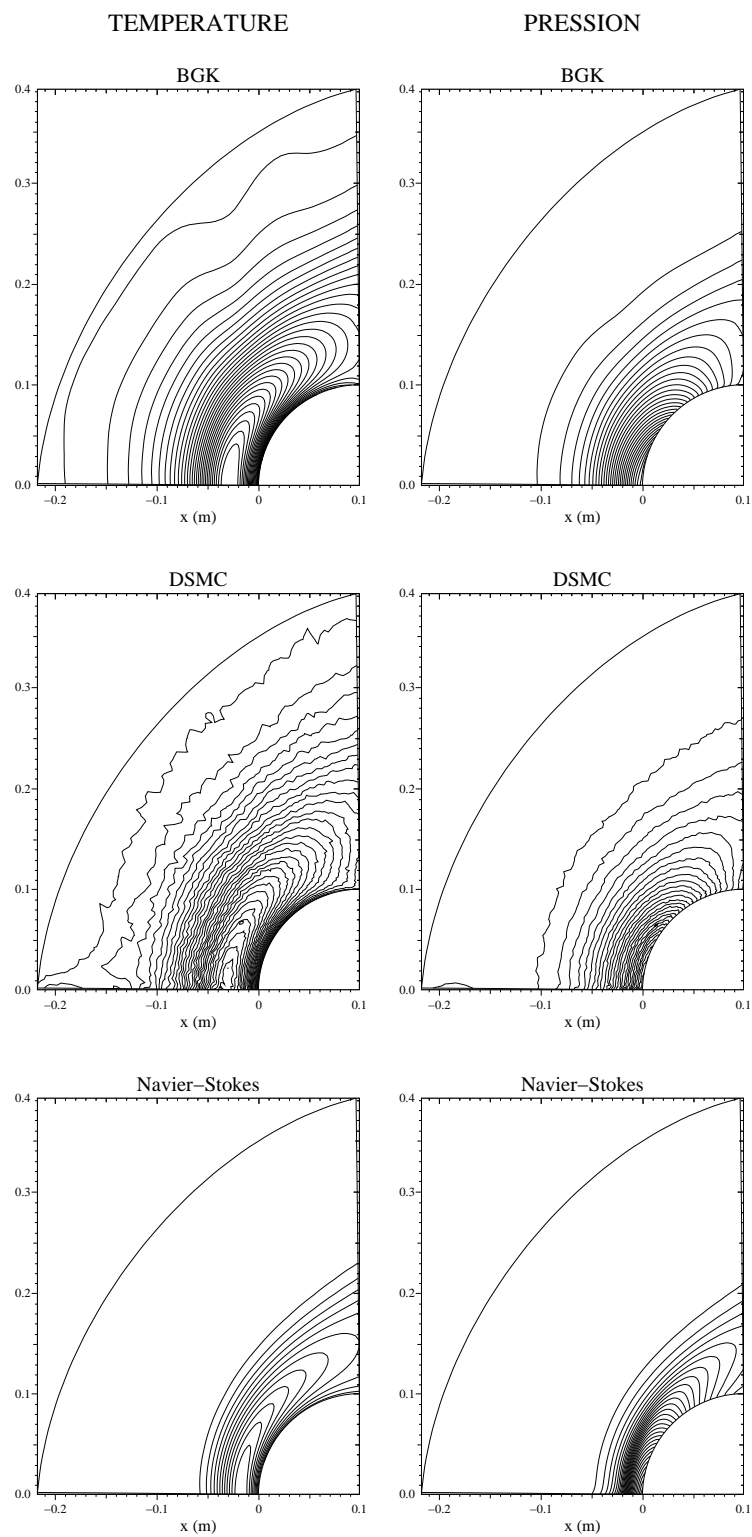


FIG. 6.27 – *Ecoulement axisymétrique autour d'une sphère. Isovaleurs de température et de pression pour BGK, DSMC, et Navier-Stokes.*

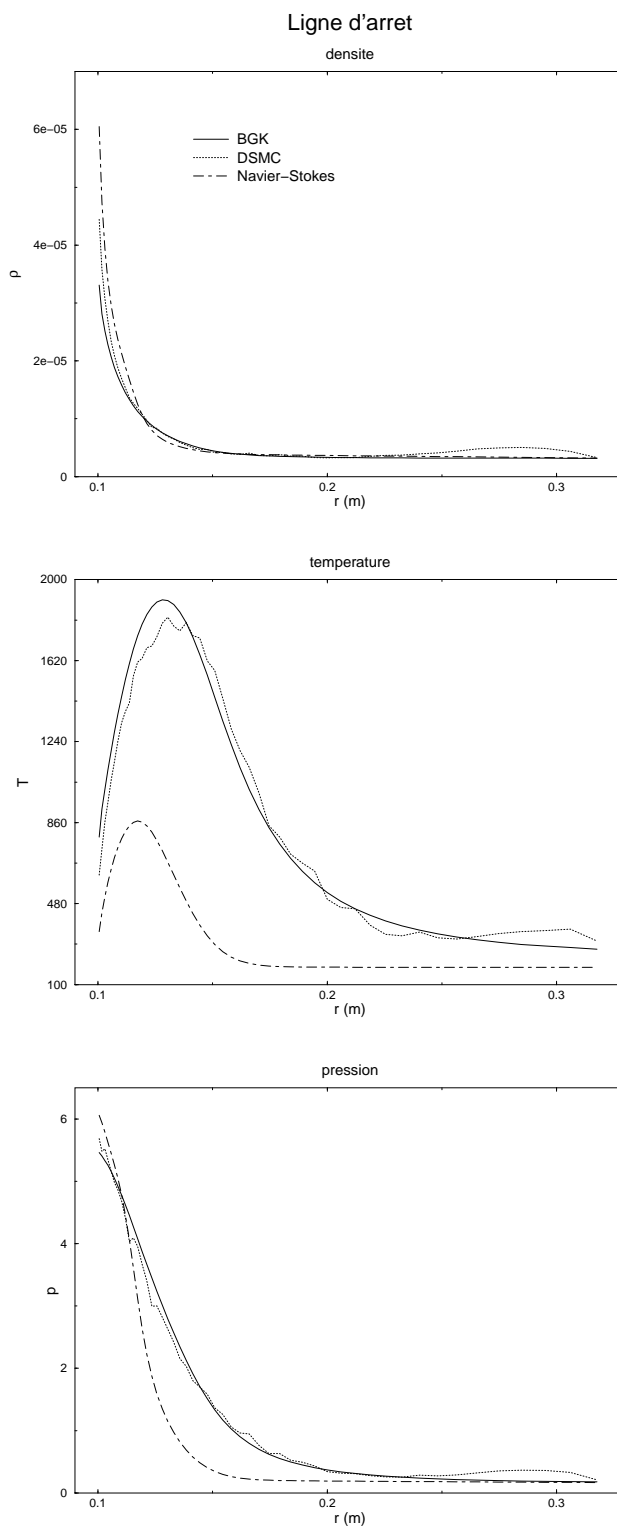


FIG. 6.28 – *Ecoulement axisymétrique autour d'une sphère. Coupes de densité, température et pression en fonction du rayon sur la ligne d'arrêt.*

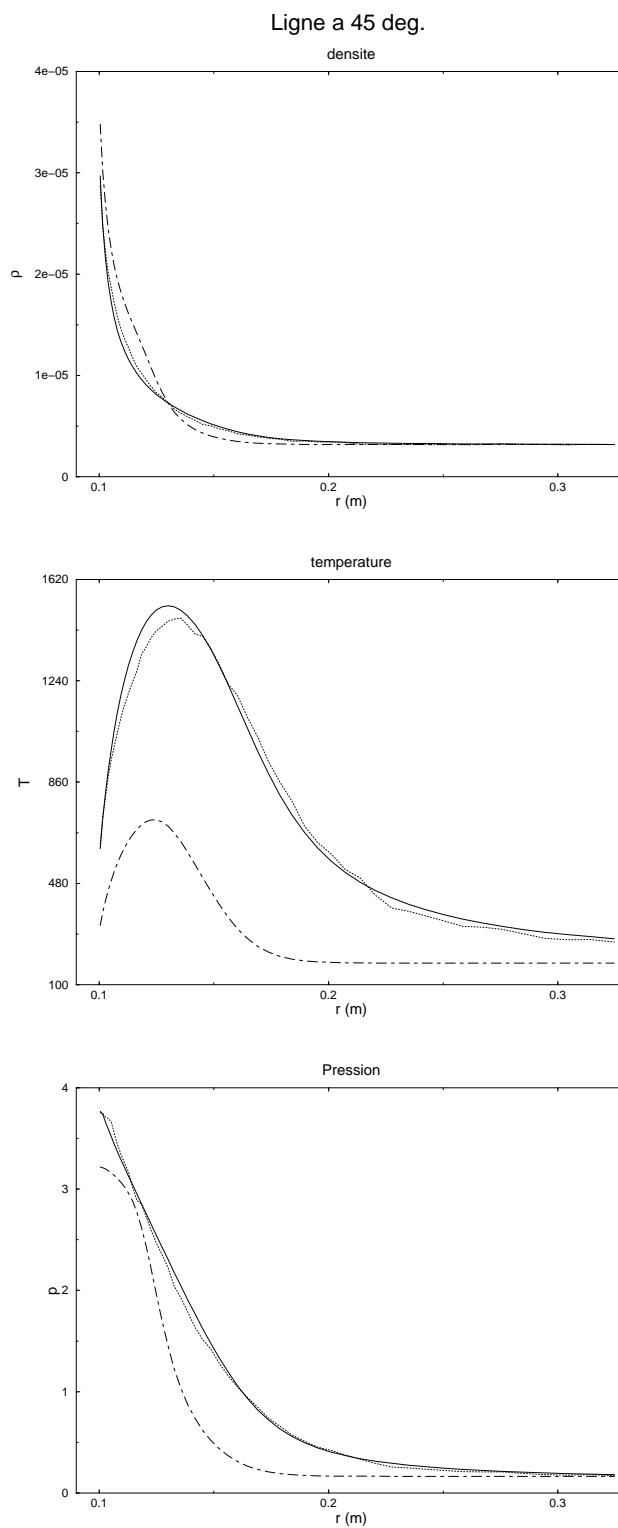


FIG. 6.29 – *Ecoulement axisymétrique autour d'une sphère. Coupes de densité, température et pression en fonction du rayon sur une ligne à 45°.*

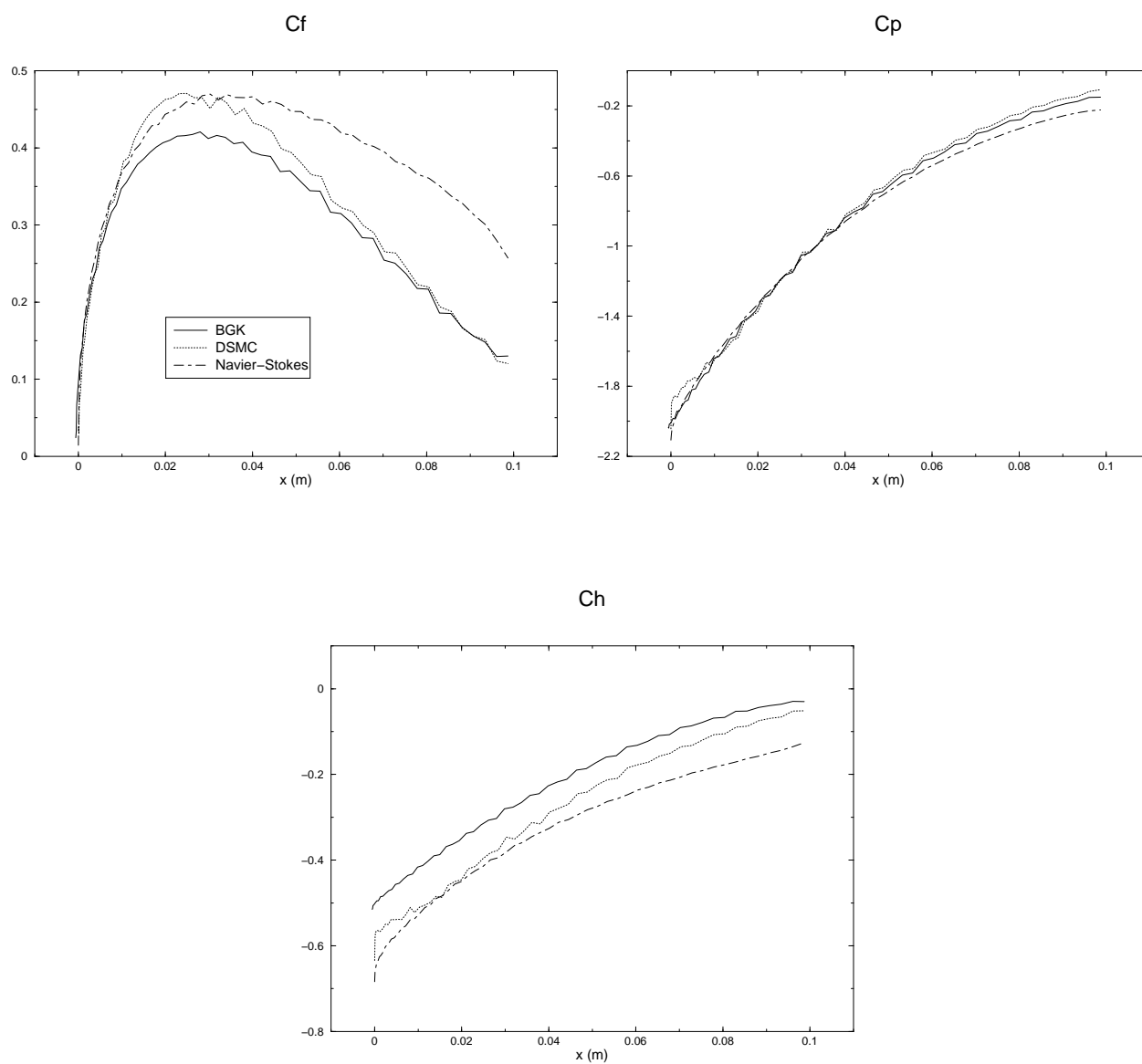


FIG. 6.30 – *Ecoulement axisymétrique autour d'une sphère. Coefficients aérodynamiques à la paroi.*

6.4.3 Influence du nombre de Prandtl

Écoulement de cisaillement 1D

Pour comparer les modèles BGK et BGK-ES, nous utilisons ici le cas d'un écoulement de Couette 1D plan étudié par Perlat dans [48]. La géométrie de cet écoulement est identique à celle de la section 6.2.1, sauf que l'on impose à droite un écoulement de densité égale à la densité initiale (on n'a donc plus de condition de réflexion diffuse à droite). Les caractéristiques suivantes sont déterminées pour redonner les grandeurs adimensionnées de [48], elles ne correspondent pas à un gaz existant :

masse moléculaire (kg)	$4.81 \cdot 10^{-26}$
exposant ω de la loi de viscosité (modèle HS)	0.5
viscosité de référence	21.93
température de référence (K)	273
température de paroi (K)	250
vitesse de paroi ($m.s^{-1}$)	1000
densité initiale ($kg.m^{-3}$)	1
température initiale (K)	250
distance L entre les plaques (m)	1
Knudsen initial (basé sur L)	0.1

Nous utilisons un maillage en espace constitué de 100 mailles dans la direction orthogonale aux plaques, et une grille de vitesses de bornes $[-1100, 1100] \times [-2100, 2100] \times [-1100, 1100]$ avec $17 \times 17 \times 17$ points, déterminée par la vitesse et la température de la paroi (cf. section 6.1.2).

Pour le modèle BGK, le calcul converge en 100 itérations et 36 min CPU. Pour le modèle BGK-ES, le calcul converge en 144 itérations et 1 h CPU. On compare dans le tableau suivant les sauts de vitesse et de température à la paroi ainsi que la température maximum et la densité minimum obtenues avec BGK et BGK-ES, aux résultats DSMC de [48] (les pourcentages à côté des valeurs sont l'écart relatif à la valeur DSMC) :

	saut u ($m.s^{-1}$)	saut T (K)	T_{max} (K)	ρ_{min} ($kg.m^{-3}$)
DSMC	72	59	420	0.7
BGK	75.8 (+5.3%)	64.2 (+8.8%)	460 (+9.5%)	0.638 (-8.9%)
BGK-ES	74.9 (+4%)	62.8 (+6.4%)	416 (-1%)	0.699 (-0.1 %)

Si l'on considère les résultats obtenus avec DSMC comme les résultats de référence, alors il apparaît clairement que le modèle BGK-ES améliore toutes les quantités étudiées, et en particulier les amplitudes de densité et de température. Pour cet écoulement, l'influence du nombre de Prandtl est donc importante.

Rampe de Compression

Nous étudions ici l'écoulement autour de la rampe de compression testé section 6.4.2, pour voir si le modèle BGK-ES améliore les profils de température à la paroi. Nous utilisons la même grille de vitesses et le même maillage en espace que dans cette section précédente. L'écoulement converge en 564 itérations et 60 h CPU sur un processeur de type 2

La figure 6.31 montre les isovaleurs de densité et de température pour les modèles BGK et BGK-ES. Les figures 6.32 et 6.33 montrent les coupes de densité et de température pour BGK, DSMC,

et BGK-ES, le long des trois lignes verticales $x = 2.5 - 5 - 7.5 \text{ cm}$. On constate que le modèle BGK-ES améliore effectivement les résultats de BGK sur les profils de température, au niveau de la paroi. Le calcul du choc semble cependant moins bon sur la densité pour le modèle BGK-ES. Mais globalement, les résultats donnés par BGK et BGK-ES ne sont pas très différents.

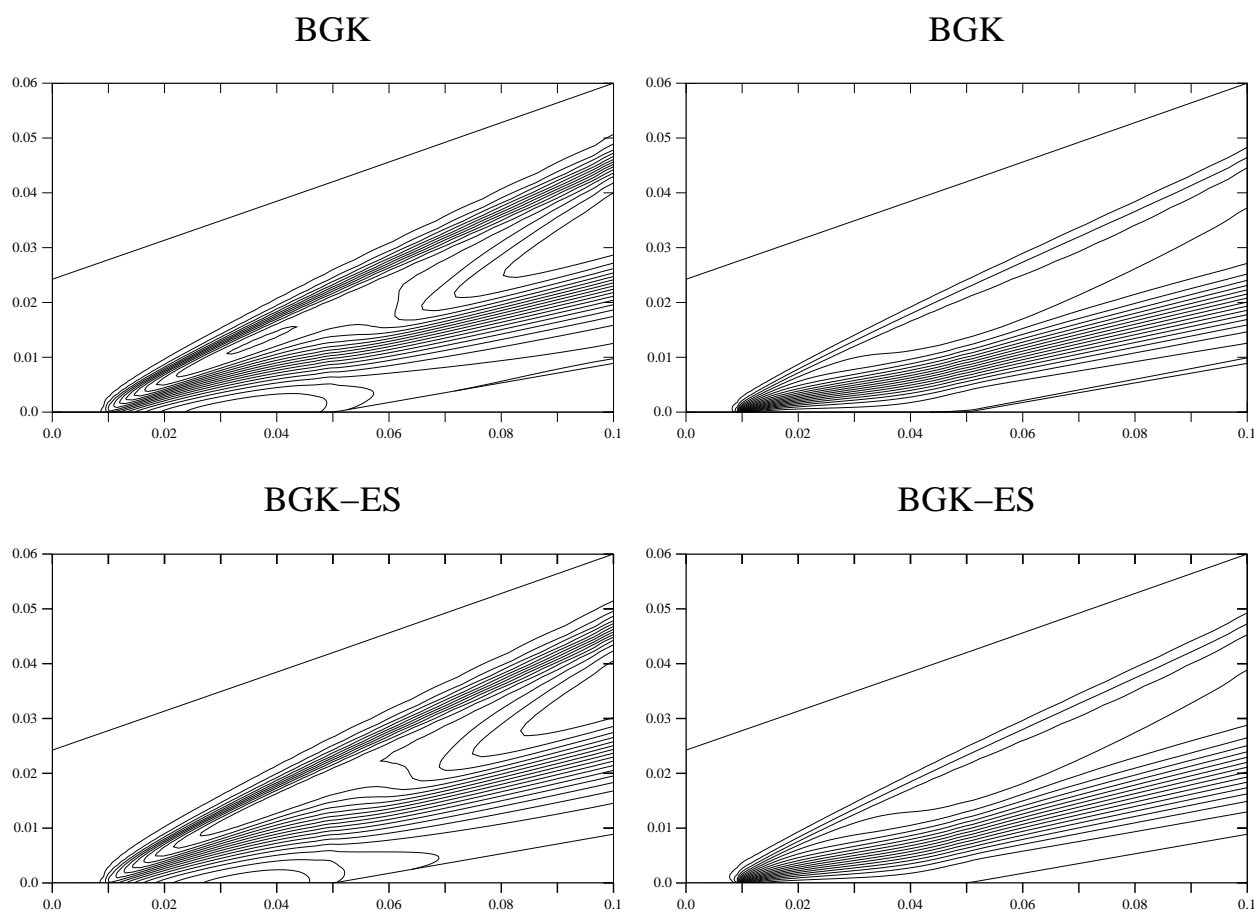


FIG. 6.31 – *Rampe de compression : densité (gauche) et température (droite) pour BGK et BGK-ES.*

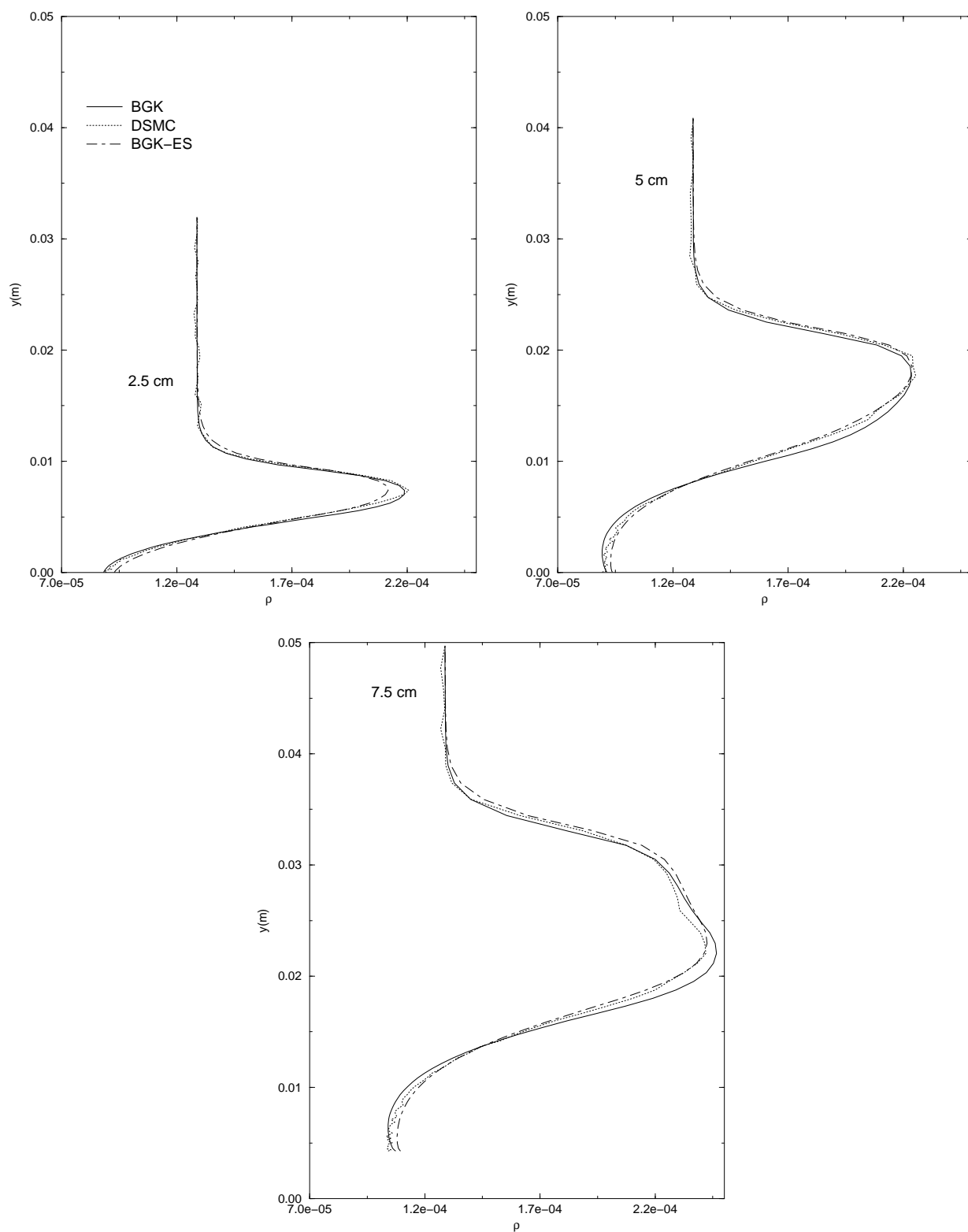


FIG. 6.32 – Profils de densité le long de trois lignes verticales ($x = 2.5, 5, 7.5$ cm), pour BGK, DSMC, et BGK-ES.

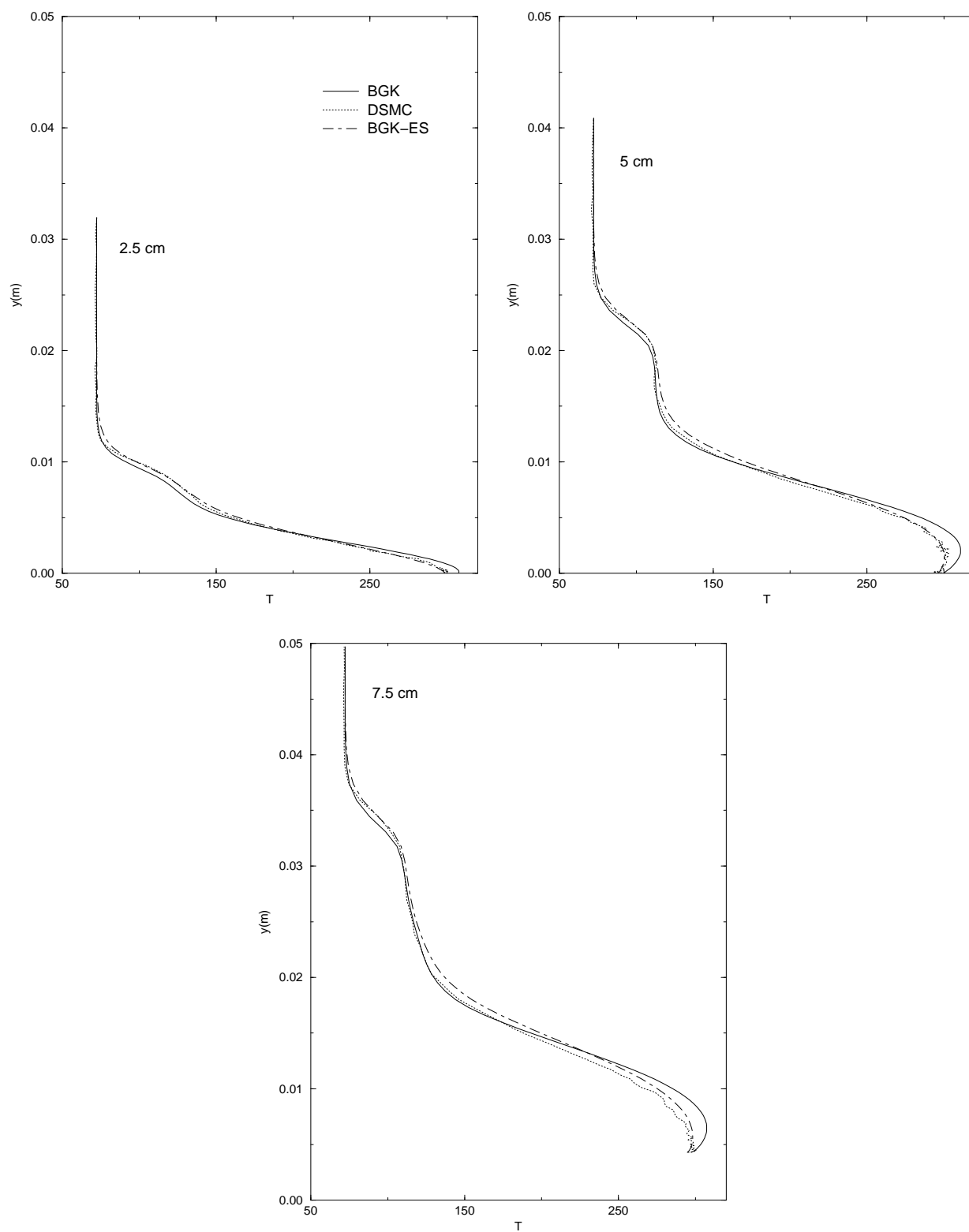


FIG. 6.33 – Profils de température le long de trois lignes verticales ($x = 2.5, 5, 7.5$ cm), pour BGK, DSMC, et BGK-ES.

6.4.4 Domaine de validité de la méthode

Nous étudions ici l'écoulement sur la rampe de compression de la section 6.4.2 pour des nombres de Knudsen cinq et dix fois plus grands. Les caractéristiques de l'écoulement ainsi que de la grille de vitesses sont exactement les mêmes que dans la section 6.4.2. Pour augmenter le nombre de Knudsen, nous avons simplement diminué la longueur de la rampe. Nous avons testé les modèles BGK et BGK-ES, en les comparant aux résultats obtenus avec le code DSMC. Le modèle Navier-Stokes n'a pas été testé ici, car on a vu section 6.4.2 qu'il est déjà mauvais pour un nombre de Knudsen plus petit.

Pour le premier essai, avec une rampe de longueur 1 cm , le nombre de Knudsen amont vaut 0.033 , soit cinq fois plus que le nombre de Knudsen de la section 6.4.2. Relativement à la longueur de la rampe, le domaine de calcul est un peu plus grand que précédemment, car le choc et la couche limite s'étalent plus.

Le calcul BGK converge en 140 itérations et 13 h CPU, contre 230 itérations et 15 h CPU pour le calcul BGK-ES. Pour le calcul DSMC, on divise le temps de stationnarisation par 5, ce qui donne $1.174 \cdot 10^{-4}\text{ s}$, et un temps maximum de simulation de $4 \cdot 10^{-3}\text{ s}$ pour 2000 sondages. Le calcul nécessite alors 14 h CPU. On remarque tout d'abord que les trois calculs ont convergé plus rapidement que pour le nombre de Knudsen 5 fois plus petit. Pour DSMC, la raison en est que lorsque le Knudsen augmente, alors le temps de collision augmente et le temps de stationnarisation diminue, ce qui implique que l'on fait moins d'itérations. Pour les modèles BGK et BGK-ES, l'accroissement du nombre de Knudsen et du temps de relaxation font diminuer le conditionnement du grand système linéaire, ce qui rend plus efficace le solveur linéaire, donc accélère la convergence.

Nous présentons à présent les isovaleurs obtenues avec les trois méthodes dans la figure 6.34, ainsi que les coupes de densité et de température effectuées selon quatre lignes verticales dans les figures 6.35 et 6.36. On voit aussi apparaître une discontinuité sur les isovaleurs de BGK et BGK-ES en début de plaque. Nous n'avons pas d'explication de ce phénomène pour l'instant (voir section 6.6 pour un autre exemple). En comparant les coupes de densité et de température, on voit que les résultats de BGK et BGK-ES sont bien moins bons que pour le nombre de Knudsen de la section 6.4.2. En particulier, le modèle BGK-ES donne de très mauvais résultats à la paroi et en début de plaque sur la densité. Pour la température, le modèle BGK-ES semble meilleur que BGK à la paroi, sauf pour la ligne $x = 0.4\text{ cm}$. Globalement, les modèles BGK et BGK-ES se sont donc dégradés avec l'augmentation du nombre de Knudsen. Notons enfin que le nombre de Knudsen local calculé au bord d'attaque vaut 0.3 .

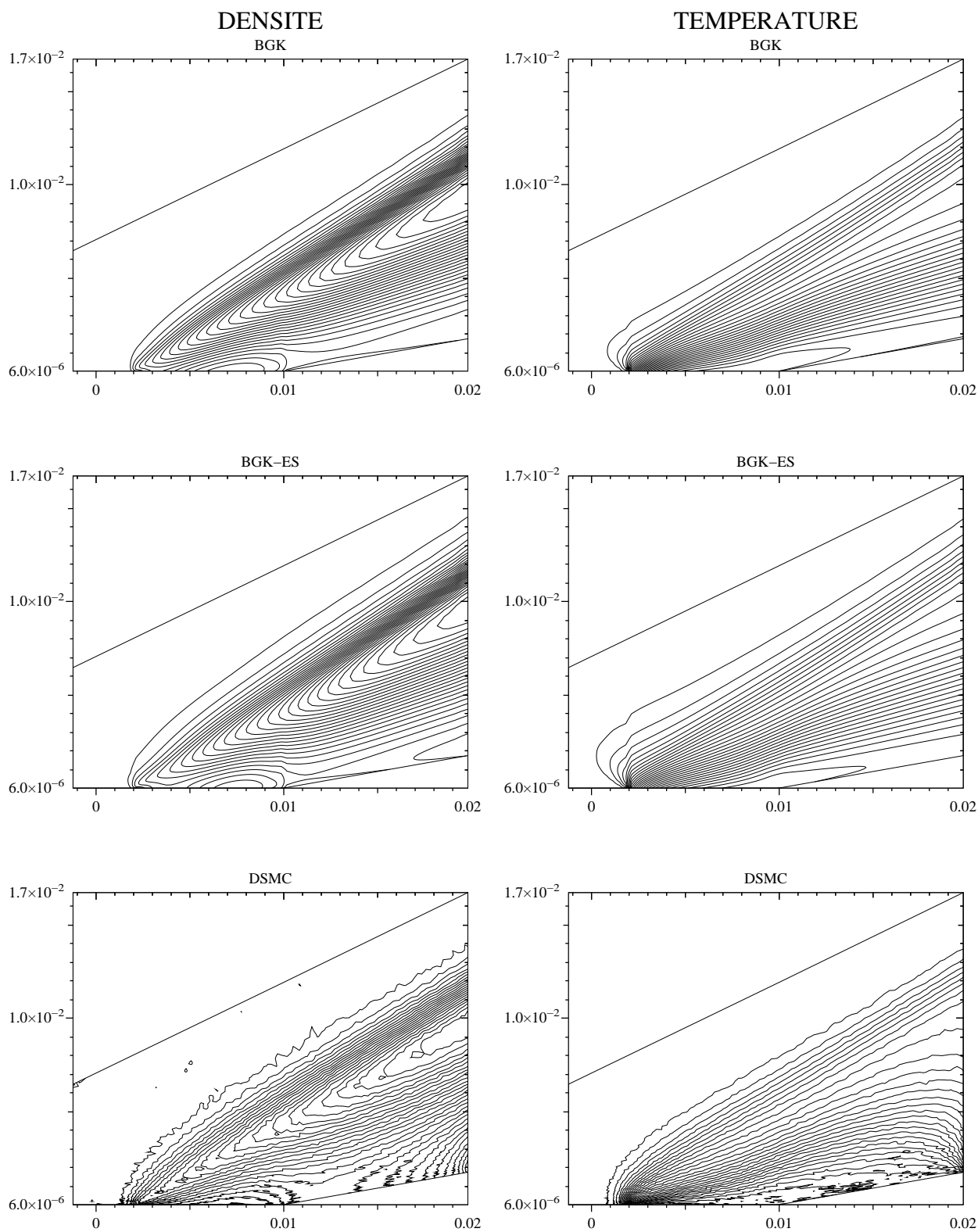


FIG. 6.34 – Isovaleurs pour la rampe de compression ($L = 1 \text{ cm}$, $Kn = 0.033$). Modèles BGK et BGK-ES, et méthode DSMC.

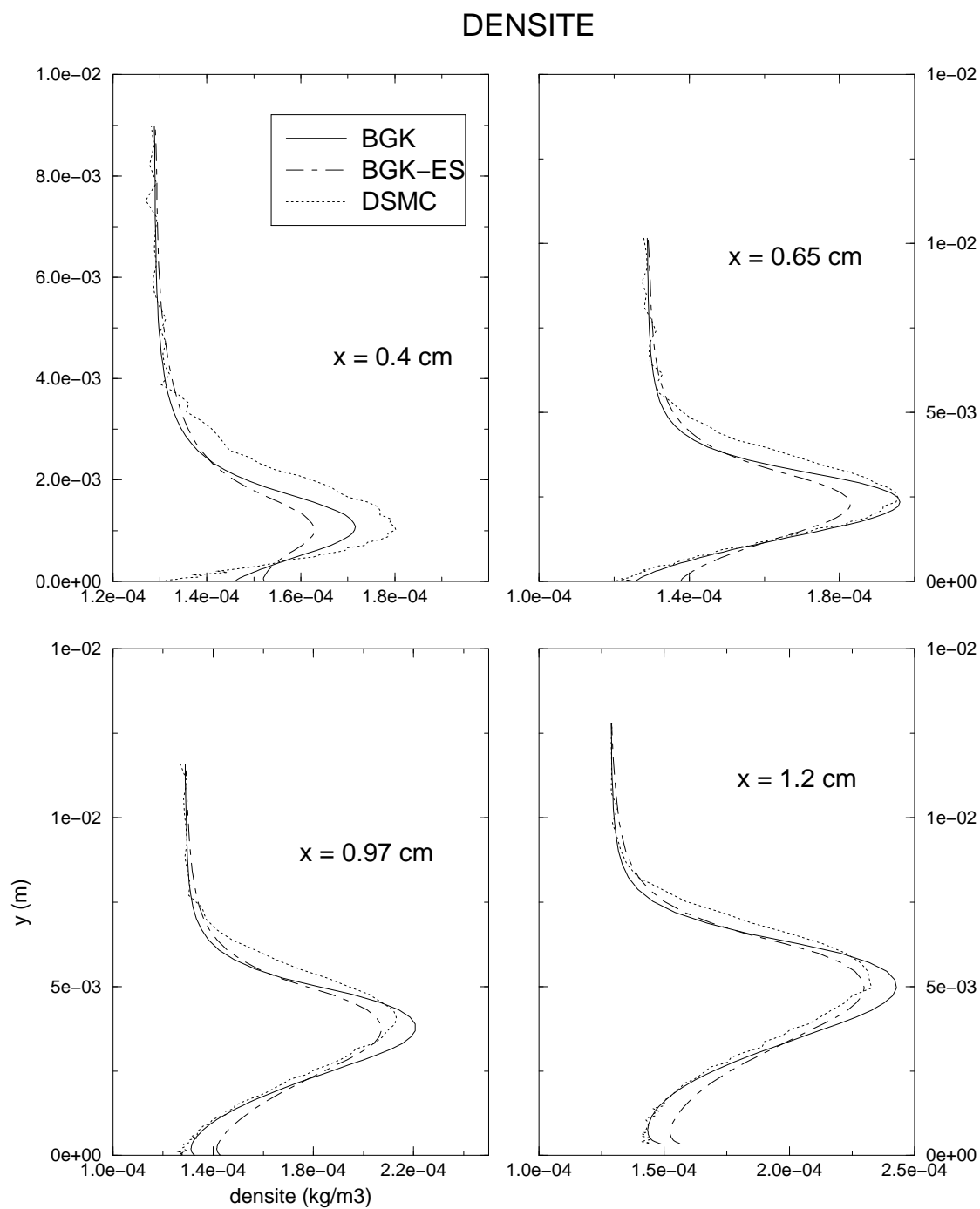


FIG. 6.35 – *Rampe de compression ($L = 1$ cm, $Kn = 0.033$) Coupes de densité suivant quatre lignes verticales pour BGK, BGK-ES, et DSMC.*

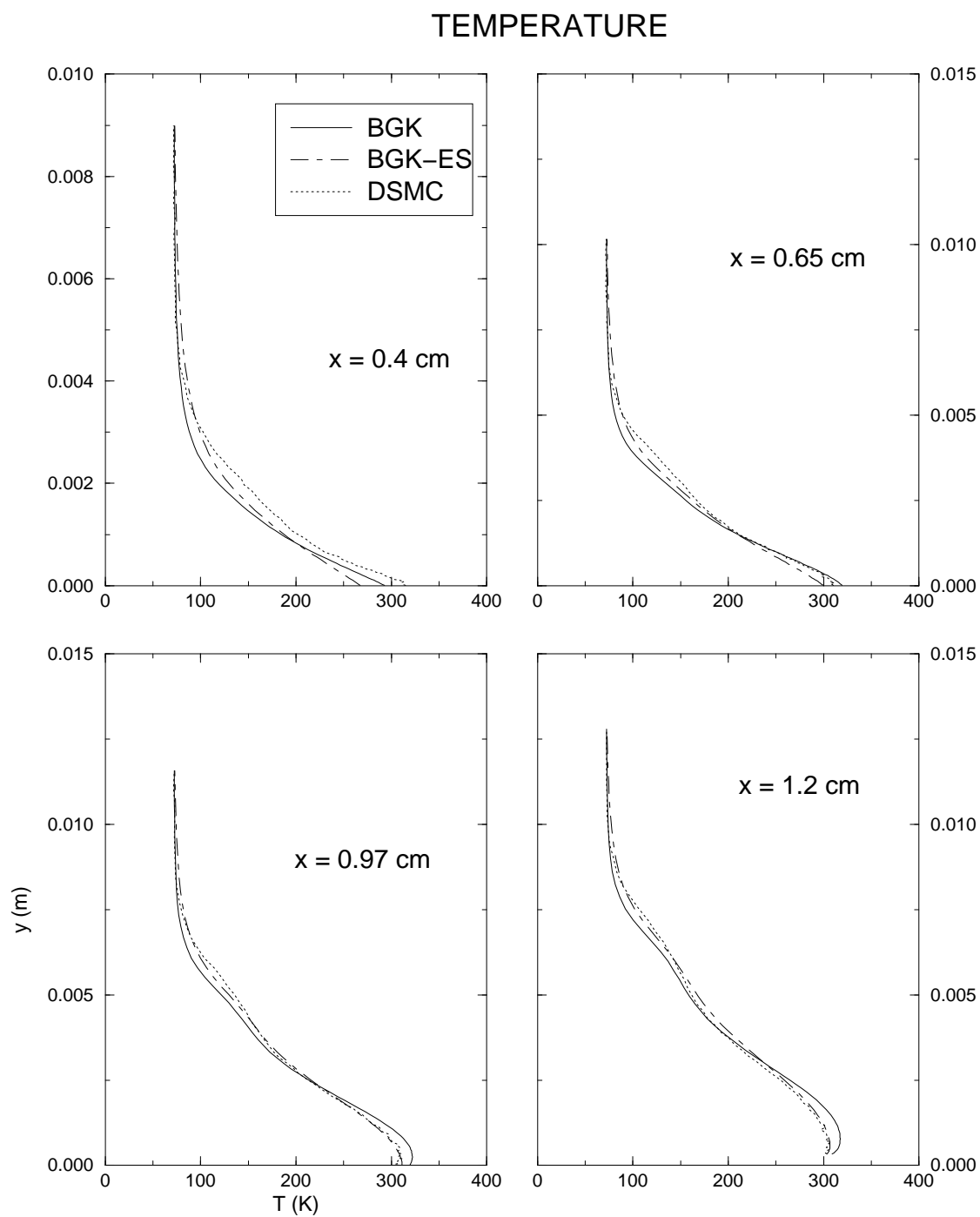


FIG. 6.36 – Rampe de compression ($L = 1$ cm, $Kn = 0.033$) Coupes de température suivant quatre lignes verticales pour BGK, BGK-ES, et DSMC.

Pour le deuxième essai, avec une rampe de longueur 0.5 cm , le nombre de Knudsen amont vaut 0.068 , soit dix fois plus que le nombre de Knudsen de la section 6.4.2, et deux fois plus que le nombre de Knudsen précédent. Relativement à la longueur de la rampe, le domaine de calcul est encore plus grand. Le calcul BGK converge en 140 itérations et 10 h CPU, contre 190 itérations et 12 h CPU pour le calcul BGK-ES. Nous présentons à présent les isovaleurs obtenues avec les trois méthodes dans la figure 6.37, ainsi que les coupes de densité et de température effectuées selon les mêmes quatre lignes verticales dans les figures 6.38 et 6.39. On constate sur les isovaleurs que la couche limite et le choc sont maintenant très mélangés. Sur les coupes de densité, BGK et BGK-ES semblent très différents de DSMC, et BGK-ES paraît légèrement plus mauvais que BGK au choc et à la paroi. Pour les coupes de température, DSMC paraît assez différent de BGK et BGK-ES sur la première coupe, mais les résultats sont proches sur les autres coupes. De plus, BGK-ES semble meilleur que BGK à la paroi sur les deux dernières coupes. Dans cet essai, le nombre de Knudsen local calculé au bord d'attaque est de 0.5 .

En conclusion, on voit que sur ce cas, le modèle BGK est beaucoup moins bon quand le nombre de Knudsen est cinq fois plus grand. Le modèle BGK-ES ne permet apparemment pas de beaucoup améliorer les résultats (les profils de température semblent meilleurs, mais au contraire les profils de densité sont moins bons que ceux de BGK). Il est cependant difficile de déduire une limite de validité de ces modèles sur la base de cette étude, en ne considérant que les nombres de Knudsen globaux ou locaux. En effet, les nombres de Knudsen globaux utilisés ici (0.033 et 0.066) sont plus petits que celui de l'écoulement autour de la sphère de la section 6.4.2, pour lequel les résultats de BGK sont très bons, et il en est de même pour les nombres de Knudsen locaux. Le problème de la validité de BGK est donc un problème qui semble dépendre fortement de chaque cas particulier que l'on étudie, notamment de la géométrie de l'obstacle.

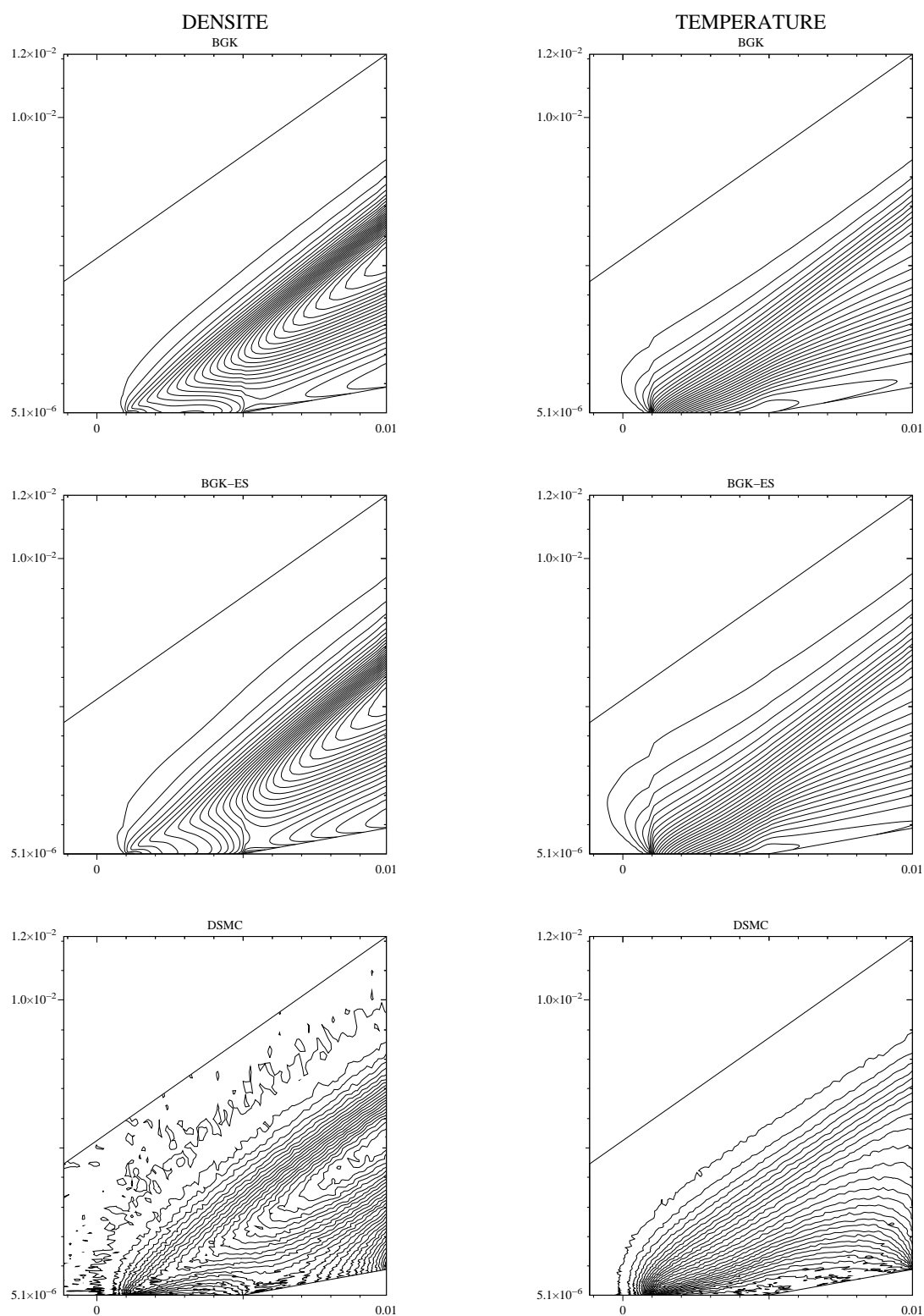


FIG. 6.37 – Isovaleurs pour la rampe de compression ($L = 0.5 \text{ cm}$, $Kn = 0.066$). Modèles BGK et BGK-ES, et méthode DSMC.

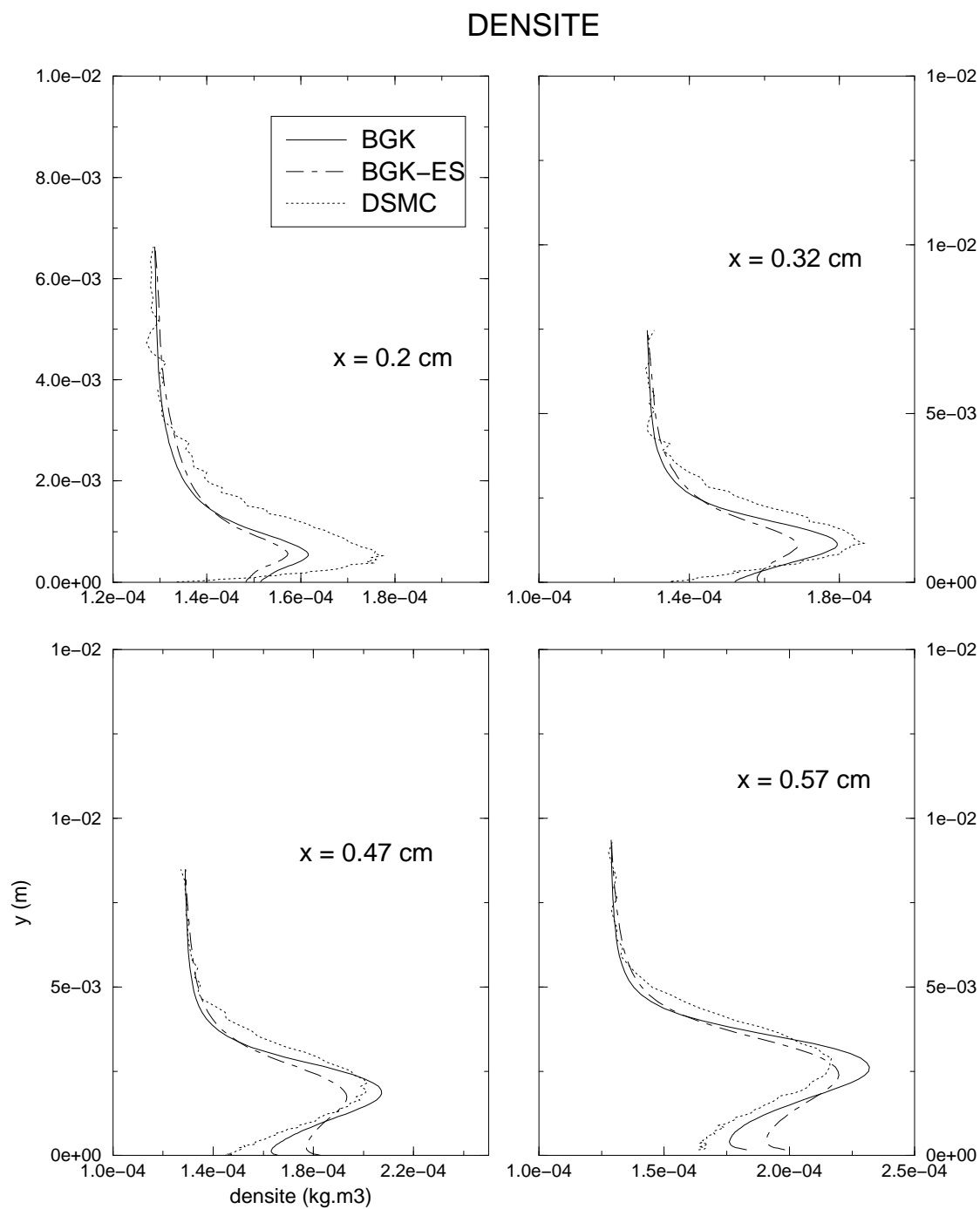


FIG. 6.38 – *Rampe de compression ($L = 0.5$ cm, $Kn = 0.066$) Coupes de densité suivant quatre lignes verticales pour BGK, BGK-ES, et DSMC.*

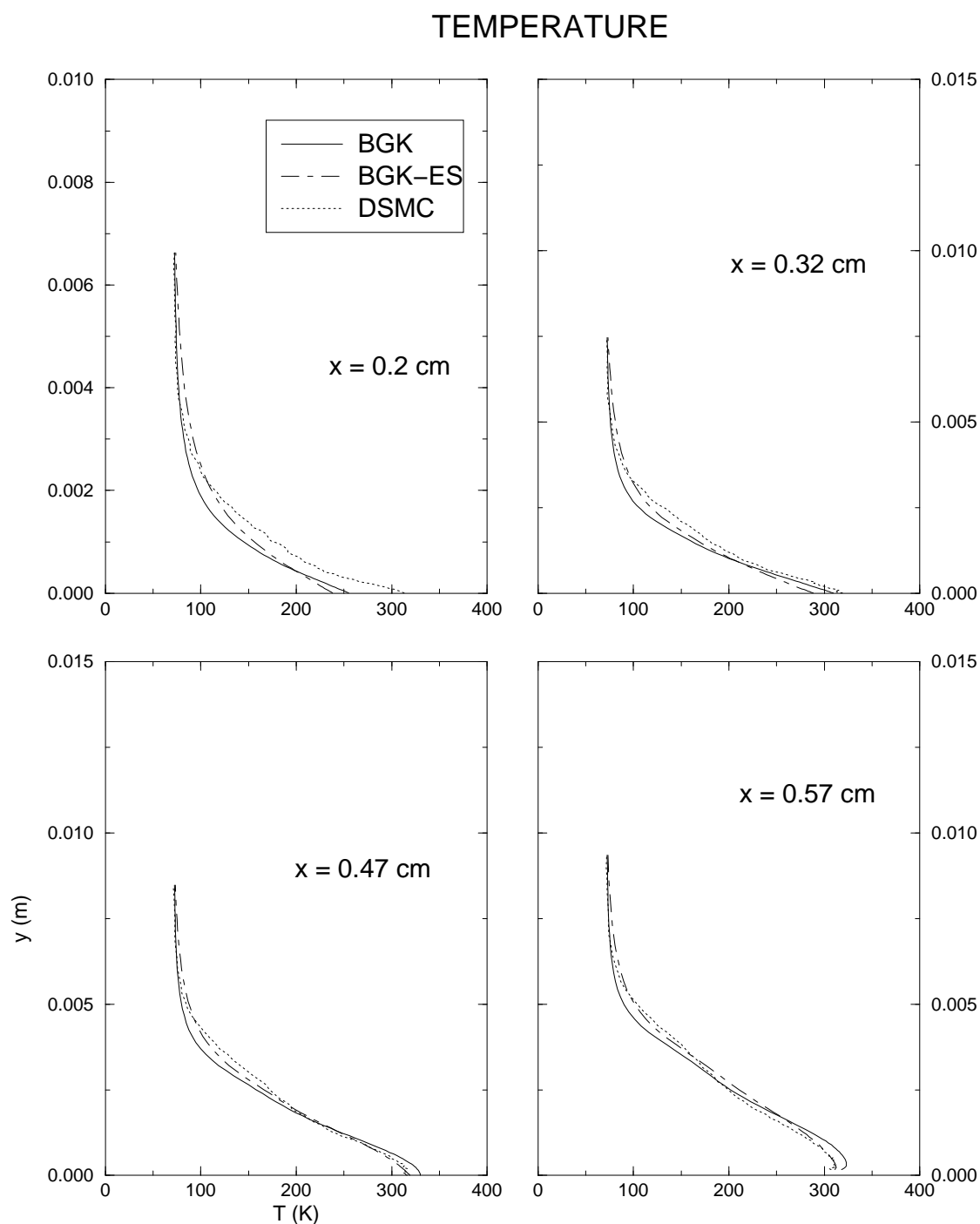


FIG. 6.39 – *Rampe de compression ($L = 0.5$ cm, $Kn = 0.066$) Coupes de température suivant quatre lignes verticales pour BGK, BGK-ES, et DSMC.*

6.5 Avantage de BGK pour certains écoulements

6.5.1 Écoulement en régime dense

Le but de ce test est de montrer que la résolution de l'équation BGK ne nécessite pas de maillage en espace aussi fin que la méthode DSMC. Plus précisément, nous voulons vérifier qu'il suffit d'utiliser un maillage assez fin pour capturer les gradients des quantités macroscopiques, c'est-à-dire adapté à la résolution des équations de Navier-Stokes. On pourra ainsi s'affranchir de la condition très restrictive de la méthode DSMC pour laquelle les mailles doivent être de taille inférieure au libre parcours moyen.

Nous considérons un écoulement plan sans incidence autour d'un cylindre infini, dont les caractéristiques suivantes sont celles d'un écoulement à 90 *km* d'altitude :

gaz	argon
masse moléculaire (<i>kg</i>)	$0.663 \cdot 10^{-25}$
exposant ω de la loi de viscosité (modèle HS)	0.5
température de paroi (<i>K</i>)	293
densité amont ($kg \cdot m^{-3}$)	$1.82 \cdot 10^{-5}$
température amont (<i>K</i>)	300
Mach amont	2.5
rayon <i>R</i> de la sphère (<i>m</i>)	1
Knudsen amont (basé sur <i>R</i>)	$6 \cdot 10^{-3}$
Knudsen au nez (basé sur <i>R</i>)	$4 \cdot 10^{-3}$

Cet écoulement est à faible Mach, on peut donc prendre une grille de vitesse assez grossière, ce qui permet d'utiliser un maillage en espace très fin. Les bornes de la grille, estimées par un calcul Navier-Stokes, sont $[-1600, 1600] \times [-1400, 1400] \times [-1300, 1300]$ avec $9 \times 7 \times 7$ points.

Le premier maillage en espace utilisé comporte 50×50 points, nous l'appellerons maillage NS. Ce maillage est raffiné près de la paroi, par une technique utilisée pour la convergence en maillage des schémas numériques pour les équations de Navier-Stokes. Le raffinement est considéré comme suffisant quand l'approximation d'ordre un des flux calculés à la paroi est égale à l'approximation d'ordre deux.

Le deuxième maillage (maillage MC) comporte 80×350 mailles, de sorte que la longueur des mailles dans la direction normale à la paroi est partout de l'ordre du libre parcours moyen amont. Près de la paroi, cette longueur est inférieure à la moitié du libre parcours moyen à la paroi. Ce maillage respecte donc le critère des méthodes DSMC, au moins dans la direction normale.

Sur le maillage NS, notre schéma BGK converge en 3600 itérations et 70 *h* CPU. Sur le maillage MC, le résidu est divisé par 10^3 en 1630 itérations et 500 *h* CPU. On a supposé que l'écoulement est suffisamment convergé ainsi, car il était difficile de demander plus de temps CPU que les 20 jours déjà utilisés ! Nous avons aussi calculé l'écoulement sur les deux maillages avec le code Navier-Stokes.

Nous comparons les résultats obtenus avec BGK pour les deux maillages NS et MC sur les figures 6.40 et 6.41. On constate que les isovaleurs de température et de pression sont assez proches sur les deux maillages (fig. 6.40), même si évidemment le choc est plus fin sur le maillage MC. Pour les coefficients aérodynamiques (fig. 6.41), on constate là encore qu'il n'y a pas de grande différence entre les deux maillages, ce qui, sur cet essai, confirme notre hypothèse. En outre, on voit que le calcul par le code Navier-Stokes donne des résultats différents de modèle BGK, ce qui montre que le gaz n'est pas suffisamment dense pour être correctement décrit par un modèle fluide.

Pour la méthode DSMC, nous utilisons les paramètres suivants :

nombre de molécules par maille	20
temps entre deux collisions (s)	$5.37 \cdot 10^{-7}$
temps de stationnarisation (s)	$4.96 \cdot 10^{-3}$ (soit 9000 it.)
nombre de sondages	1000
temps maximum de simulation (s)	$5.37 \cdot 10^{-3}$ (soit 12000 it.)

Sur le maillage NS, le temps maximum de simulation est atteint en 100 h CPU sans que l'écoulement ait convergé (alors que BGK converge en 70 h CPU). En effet, la courbe d'évolution du nombre total de molécules au cours du temps (fig. 6.42) montre que ce nombre n'a pas stationnarisé au terme du calcul. Nous ne montrons pas les courbes isovaleurs et les coefficients aérodynamiques qui sont complètement bruités. Ces résultats étaient bien sûr attendus, puisque le maillage ne respecte pas du tout le critère DSMC. Sur le maillage MC, le seul calcul que nous ayons effectué a nécessité 83 h CPU pour seulement 631 itérations. Ceci permet d'estimer le temps nécessaire pour atteindre le temps maximum de simulation à deux mois et demi. Nous n'avons donc pas tenté de mener ce calcul à son terme.

En conclusion, nous avons vérifié que la résolution de l'équation BGK ne nécessite qu'un maillage de type Navier-Stokes, contrairement à la méthode DSMC qui ne donne aucun résultat exploitable sur un tel maillage. Nous avons aussi observé que les équations de Navier-Stokes ne suffisent pas pour décrire correctement l'écoulement. Ce test prouve donc que notre méthode est une approche tout à fait pertinente pour un écoulement sur lequel les méthodes classiques (DSMC et Navier-Stokes) ne donnent pas de bons résultats. Il serait toutefois intéressant de mener le calcul DSMC à son terme pour valider nos résultats obtenus avec le modèle BGK.

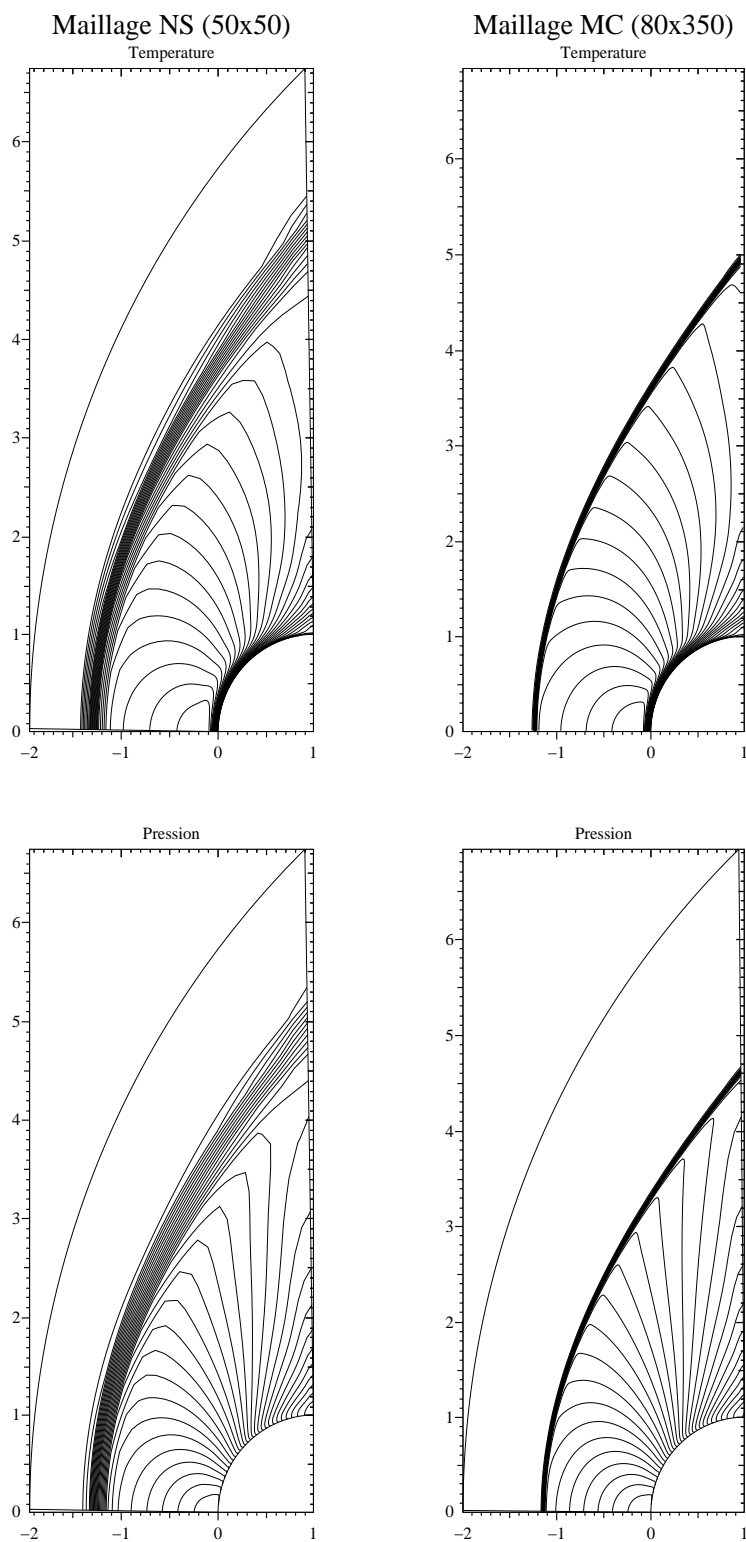


FIG. 6.40 – Isovaleurs pour un écoulement dense à Mach 2.5 autour d'un cylindre. Résultats BGK sur un maillage raffiné à la paroi (maillage NS) et sur un maillage respectant le critère $\Delta x \leq \lambda$ (maillage MC).

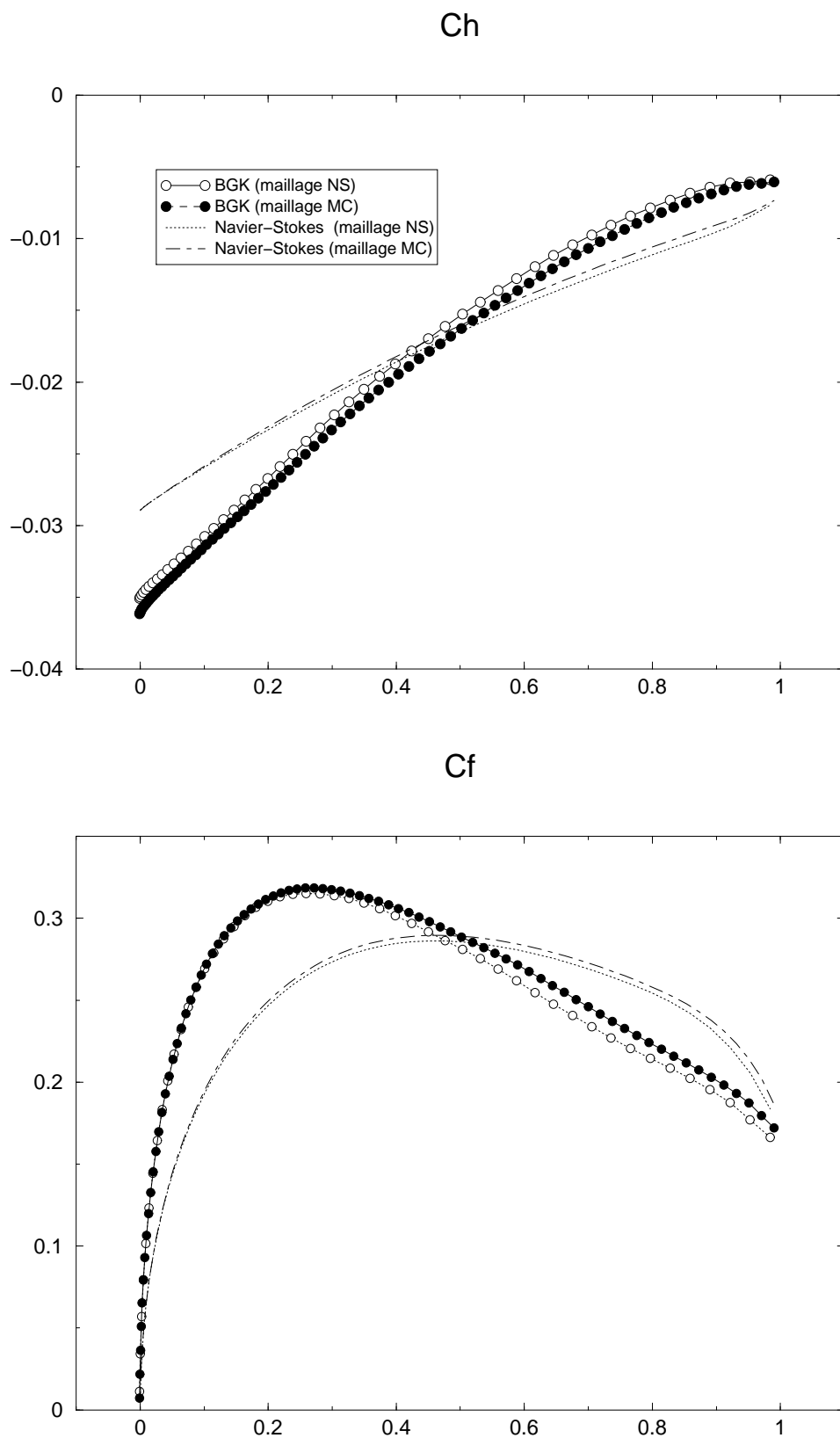


FIG. 6.41 – Coefficients aérodynamiques pour BGK et Navier-Stokes sur les deux maillages NS et MC.

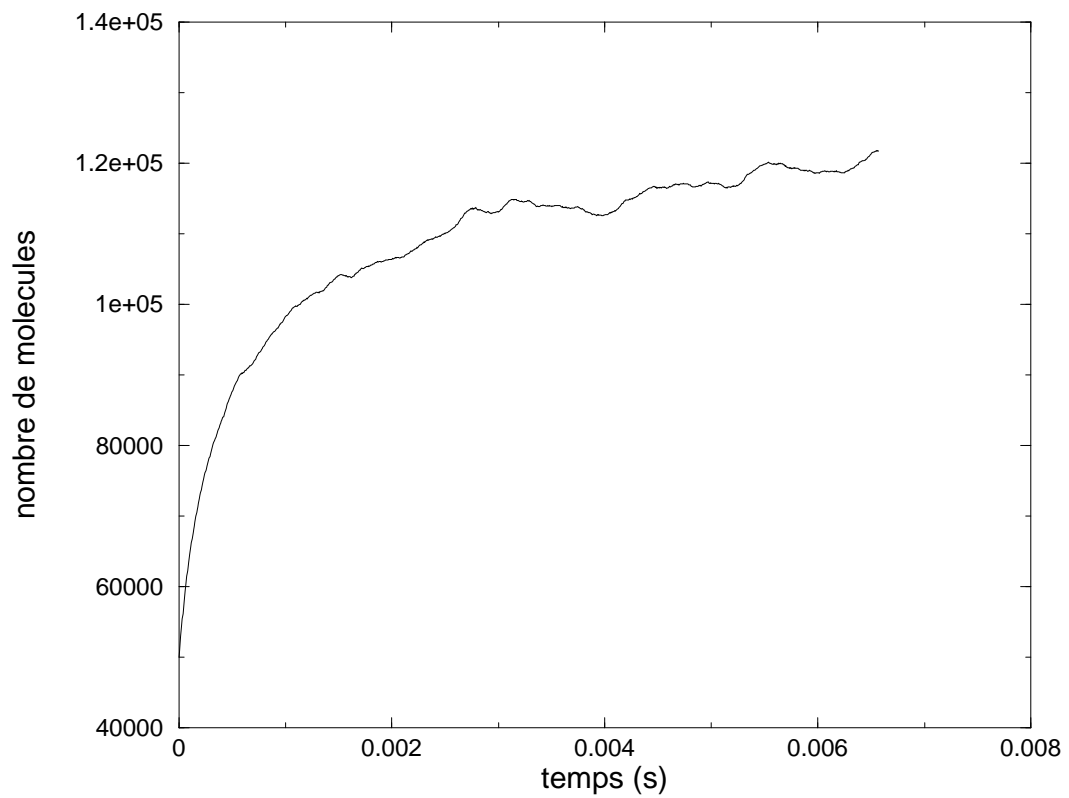


FIG. 6.42 – *Evolution du nombre total de molécules au cours du temps pour la simulation DSMC sur le maillage NS. Ce nombre n'a pas convergé.*

6.5.2 Calcul d'une recirculation

Il est connu que les méthodes Monte-Carlo sont coûteuses pour calculer des recirculations d'écoulements en arrière de corps. La première raison est que dans ces zones, la vitesse du gaz est en général très lente par rapport au reste de l'écoulement. Il faut donc un grand nombre d'itérations pour converger. Par ailleurs, la densité en arrière d'un corps est très faible, ce qui implique que si on ne donne pas assez de molécules en entrée, alors au bout d'un certain temps il n'y a plus assez de molécules numériques pour bien représenter le gaz dans cette zone, et les résultats sont donc très bruités. Comme notre méthode BGK à vitesses discrètes ne repose pas sur un nombre de molécules fixé, on doit pouvoir éviter le deuxième. De plus, le caractère implicite de notre schéma doit éliminer le problème de convergence lente.

Le cas que nous avons choisi pour vérifier ces hypothèses est un écoulement plan supersonique autour d'un cylindre. Les caractéristiques de l'écoulement sont les suivantes :

gaz	argon
masse moléculaire (kg)	$0.663 \cdot 10^{-25}$
exposant ω de la loi de viscosité (modèle HS)	0.5
température de paroi (K)	293
densité amont ($kg.m^{-3}$)	$0.31696 \cdot 10^{-5}$
température amont (K)	249
Mach amont	4
rayon du cylindre (m)	1
nombre de Knudsen amont	0.0358

Ces caractéristiques correspondent à celles d'un écoulement dans l'atmosphère à 90 km d'altitude (mais ici, le gaz est de l'argon, considéré comme monoatomique). Nous effectuons tout d'abord un calcul BGK sur une section entière du cylindre, pour en tester la faisabilité. Le maillage est constitué de 80×30 mailles dans les directions tangente et perpendiculaire à la section du cylindre. Les bornes de la grille de vitesse sont $[-2562, 2562] \times [-2462, 2462] \times [-2303, 2303]$, estimées par un calcul Navier-Stokes (voir section 6.1.2), et on prend 11 points dans chaque direction. Ce calcul converge en 1321 itérations et $54 h 18 min$ CPU sur un processeur de type 2. Les isovaleurs obtenues sont tracées dans la figure 6.43.

On réussit bien à capturer une petite recirculation à l'arrière du cylindre près de la paroi, mais on constate qu'il y a cependant trop peu de mailles à cet endroit. On raffine donc le maillage, et pour raison de symétrie, on peut ne calculer l'écoulement que sur la moitié du domaine. De plus, il est possible de rétrécir le domaine en négligeant la partie lointaine du sillage.

Nous prenons donc un maillage de 49×60 mailles, et une grille de vitesses identique au cas précédent. Nous testons alors notre méthode BGK et la méthode DSMC sur ce maillage. Les paramètres utilisés pour DSMC sont les suivants :

nombre de molécules par maille	20
nombre de sondages	800
temps entre deux collisions (s)	$1.75 \cdot 10^{-6}$
temps de stationnarisation (s)	0.02483 (soit 14000 it.)
temps maximum de simulation (s)	0.02903 (soit 16660 it.)

Notons que le maillage ne respecte le critère de taille de maille inférieur au libre parcours moyen que proche de la paroi. On peut donc prévoir que les résultats obtenus avec DSMC ne seront pas très bons.

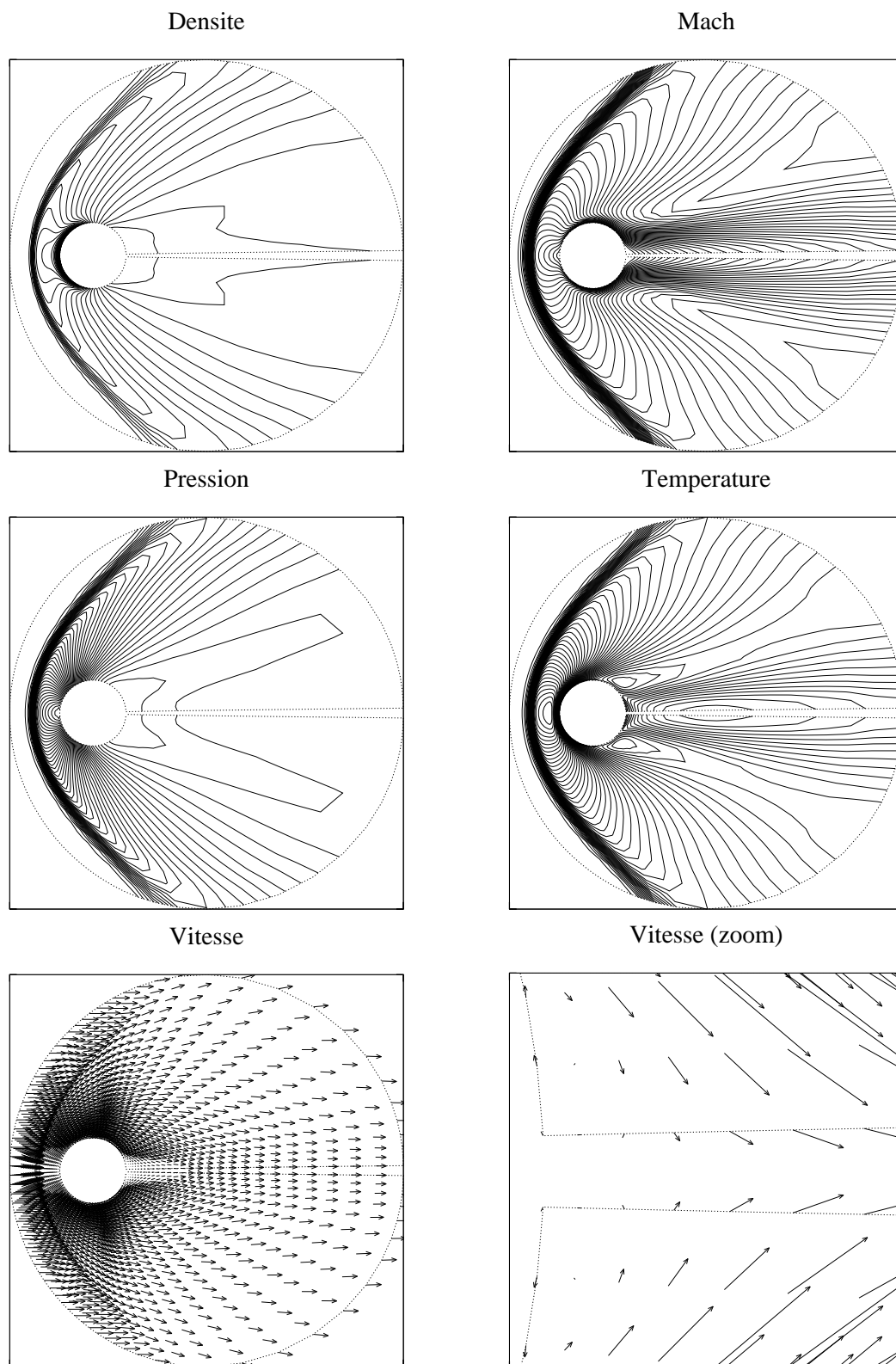


FIG. 6.43 – Isovaleurs pour l'écoulement supersonique autour d'un cylindre. Noter la recirculation en arrière du corps près de la paroi (zoom sur le champ de vitesses).

Pour BGK, le calcul converge en 1167 itérations et 90 h CPU. Sur le même processeur, le code DSMC atteint le temps maximum de simulation en 16000 itérations et 120 h CPU (ce qui est plus long que pour BGK), mais l'écoulement n'est pas stationnarisé. Le nombre total de molécules dans l'écoulement commence à être stabilisé en fin de calcul (voir figure 6.44), mais les mailles de la zone de recirculation ne contiennent que 3 molécules, ce qui est nettement insuffisant pour décrire le gaz de façon réaliste. On vérifie bien ce problème sur la figure 6.45, car la zone de recirculation (visualisée

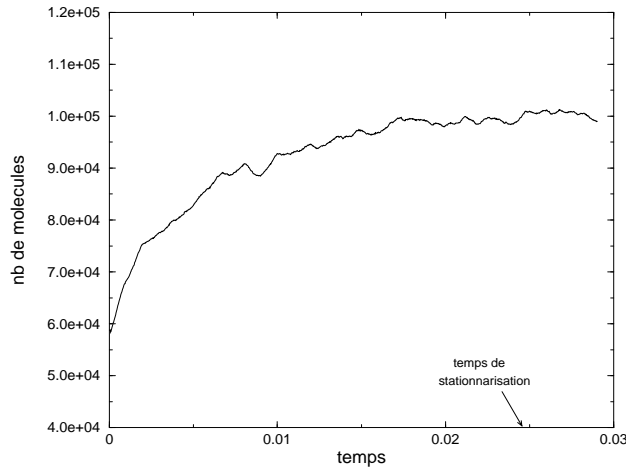


FIG. 6.44 – *Recirculation derrière un cylindre : évolution du nombre total de molécules dans l'écoulement (méthode DSMC). Noter que ce nombre n'a pas vraiment convergé.*

par un zoom sur le champs de vitesses) est très mal prise en compte par DSMC, contrairement à notre méthode. On constate aussi que les isochores sont extrêmement bruitées avec DSMC.

On peut donc conclure que notre méthode est avantageuse par rapport à DSMC sur ce genre de configuration. En effet, il faudrait un temps CPU beaucoup plus important pour que DSMC donne des résultats d'une qualité comparable à BGK. Nous n'avons pas tenté de mener le calcul DSMC à convergence pour comparer nos résultats, car le temps CPU nécessaire est prohibitif (il est déjà extrêmement important sur l'essai présenté ici : 120 h !).

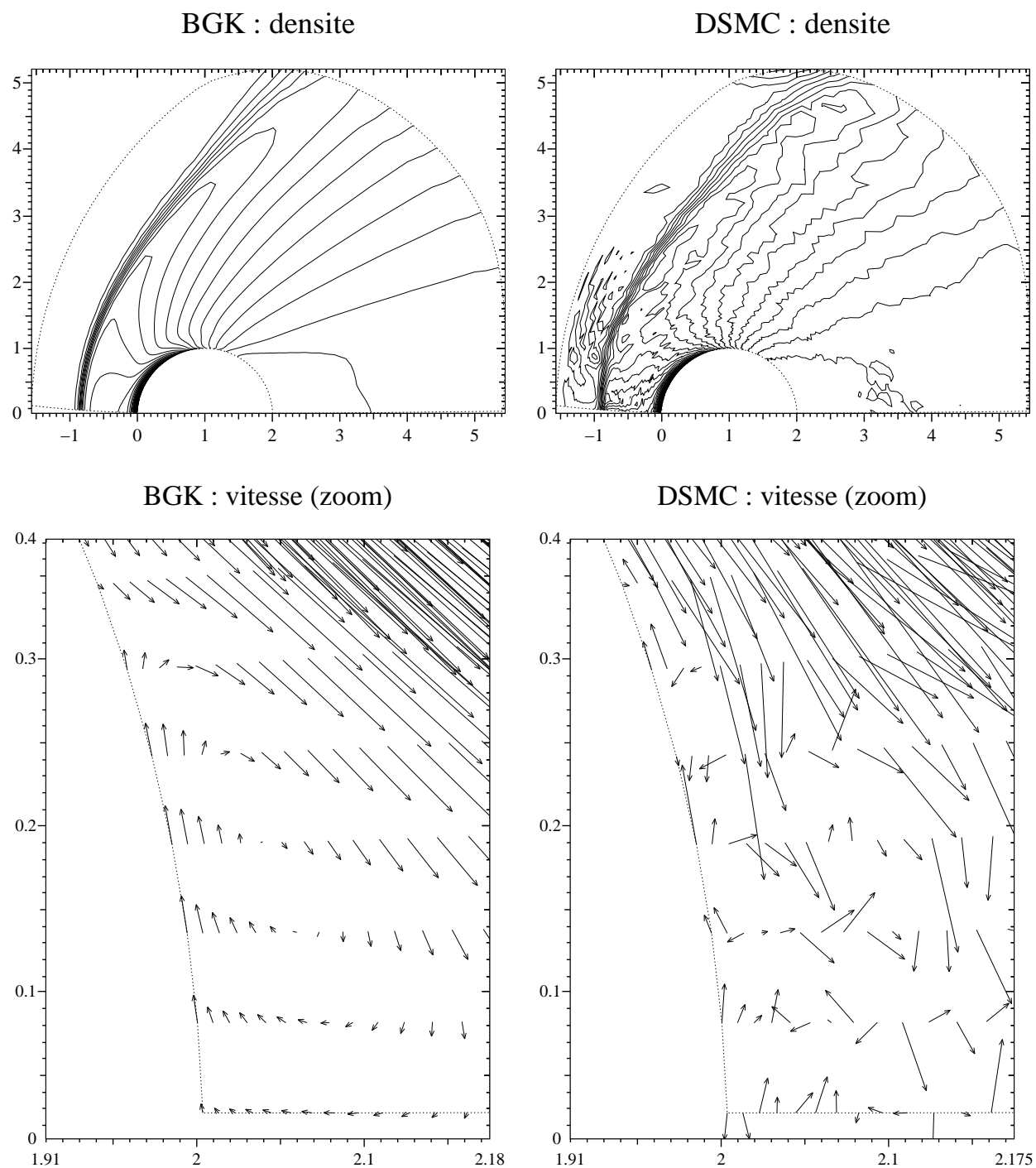
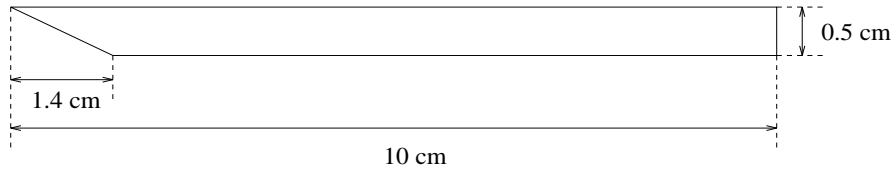


FIG. 6.45 – *Recirculation derrière un cylindre. Comparaison entre BGK et DSMC: densité (à gauche) et champ de vitesse dans la recirculation (à droite).*

6.6 Écoulements hypersoniques

Plaque biseautée

Ce cas a été étudié par Andriès-Bourgat-Le Tallec-Perthame dans [3] avec un code de type Monte-Carlo pouvant simuler les équations de Boltzmann, BGK, et BGK-ES. Il s'agit d'un écoulement hypersonique sans incidence à Mach 18.3 sur une plaque plane biseautée dont la géométrie est donnée dans la figure suivante :



Les caractéristiques de l'écoulement sont les suivantes :

gaz	azote
masse moléculaire (kg)	$4.65 \cdot 10^{-26}$
exposant ω de la loi de viscosité (modèle HS)	0.5
température de paroi (K)	286
densité amont ($kg.m^{-3}$)	$5.19 \cdot 10^{-5}$
température amont (K)	13.6
Mach amont	18.3
Knudsen amont (basé sur longueur de la plaque)	$1.4 \cdot 10^{-2}$
Knudsen au nez (basé sur longueur de la plaque)	$3.1 \cdot 10^{-3}$
Knudsen au nez (basé sur l'épaisseur de la plaque)	0.2

Pour un calcul BGK, nous utilisons un maillage de 75×29 mailles dans les directions tangentes et perpendiculaires à la plaque. Il est donc très grossier par rapport au maillage de [3] qui comporte environ 180×180 mailles. Nous évaluons les bornes de la grille de vitesses par un calcul Navier-Stokes qui donne les bornes $[-2300, 2300] \times [-1600, 1600] \times [-1300, 1300]$ et $31 \times 29 \times 27$ vitesses. Une telle grille donnerait cependant lieu à des calculs très longs, nous avons donc pris un nombre de vitesses plus faible avec $21 \times 21 \times 21$, ce qui correspond au nombre donné par le critère Navier-Stokes avec $cut = 3$ (cf. section 6.1.2).

Le calcul converge en 140 itérations et 36 h CPU. Il est déjà intéressant de voir que le code résiste à un écoulement aussi violent, ce qui confirme la robustesse du schéma implicite et de l'algorithme de Newton. En début de calcul, on constate qu'il apparaît des valeurs négatives de la fonction de distribution (jusqu'à 1 %), mais elles disparaissent rapidement après stabilisation de l'écoulement.

Les résultats sont présentés sous forme d'isovaleurs (figure 6.46 et 6.47) et de coefficients aérodynamiques à la paroi (figure 6.48). On constate tout d'abord que nos résultats semblent globalement assez proches de ceux de [3], alors que notre maillage est beaucoup plus grossier (ce qui confirme l'hypothèse de la section 6.5.1). De plus, on voit que les isovaleurs du nombre de Mach sont très étalées, comparé aux autres quantités. Nous avons déjà constaté ce phénomène sur certains des tests précédents, et il apparaît aussi sur les courbes de [3]. Comme le seul point commun entre notre

approche et celle de [3] est le modèle BGK sous-jacent, cet étalement est sans doute dû au modèle lui-même, et non pas à l'approche discrète ou au schéma numérique. Une interprétation possible est que dans le choc autour de la plaque, les molécules de très grande vitesse arrivant de l'amont ne collisionnent quasiment qu'avec les molécules de faible vitesse en aval. Cet aspect est gommé par le modèle BGK, puisque la fréquence de collision ne dépend pas de la vitesse des molécules. Une conséquence fâcheuse de cet étalement est qu'il oblige à prendre un domaine de calcul plus grand que ne le nécessiteraient les autres quantités macroscopiques. En effet, si la taille du domaine de calcul n'est pas suffisante, on constate que le calcul s'arrête quand les isovalues de Mach touchent le bord du domaine.

On observe aussi une légère discontinuité sur les isovalues de toutes les quantités macroscopiques au niveau du bord d'attaque de la plaque. Elle est encore plus visible sur le Mach. Ce phénomène ressemble à la discontinuité qui apparaît sur la rampe de compression (section 6.4.4), dont nous avons remarqué qu'elle augmente quand le gaz se raréfie. Nous avons aussi observé ce phénomène dans les résultats de [3], ce qui laisse penser qu'il est dû au modèle BGK lui-même, mais nous n'avons pas d'explication pour l'instant.

On peut avoir une idée de l'ampleur du déséquilibre cinétique près du bord d'attaque en notant que le nombre de Knudsen local vaut 0.5, et en visualisant la fonction de distribution réduite $F(v_x, v_y) = \int f(v_x, v_y, v_z) dv_z$. On distingue nettement sur la figure 6.47 la demi-maxwellienne de la paroi centrée en $v_x = 0$, et la maxwellienne de l'écoulement amont, centrée sur la vitesse amont $u_x = 1500$. Comme la température de paroi est plus forte que la température amont, la demi-maxwellienne est plus étalée.

Pour les coefficients aérodynamiques (figure 6.48), on voit que nos résultats sont très proches de ceux de [3], excepté pour le flux de chaleur C_h . Cela n'est pas dû au mauvais nombre de Prandtl du modèle BGK, puisque dans [3], il n'y a pas beaucoup de différence entre les résultats BGK et BGK-ES. Une cause possible est peut-être le faible nombre de vitesses discrètes de notre grille, qui conduit à une température de paroi peut-être surestimée (cf. section 1.3.3), et donc à un mauvais flux de chaleur.

Ce test mériterait donc d'être refait avec plus de vitesses discrètes et peut-être plus de mailles en espace, pour faire des comparaisons plus précises. Cependant, nous obtenons malgré tout des résultats globalement assez proches de ceux donnés par une simulation Monte-Carlo précise, en utilisant un maillage bien plus grossier. De plus, ce test prouve que notre schéma implicite linéarisé et notre algorithme de Newton sont assez robustes pour calculer des écoulements aussi violents.

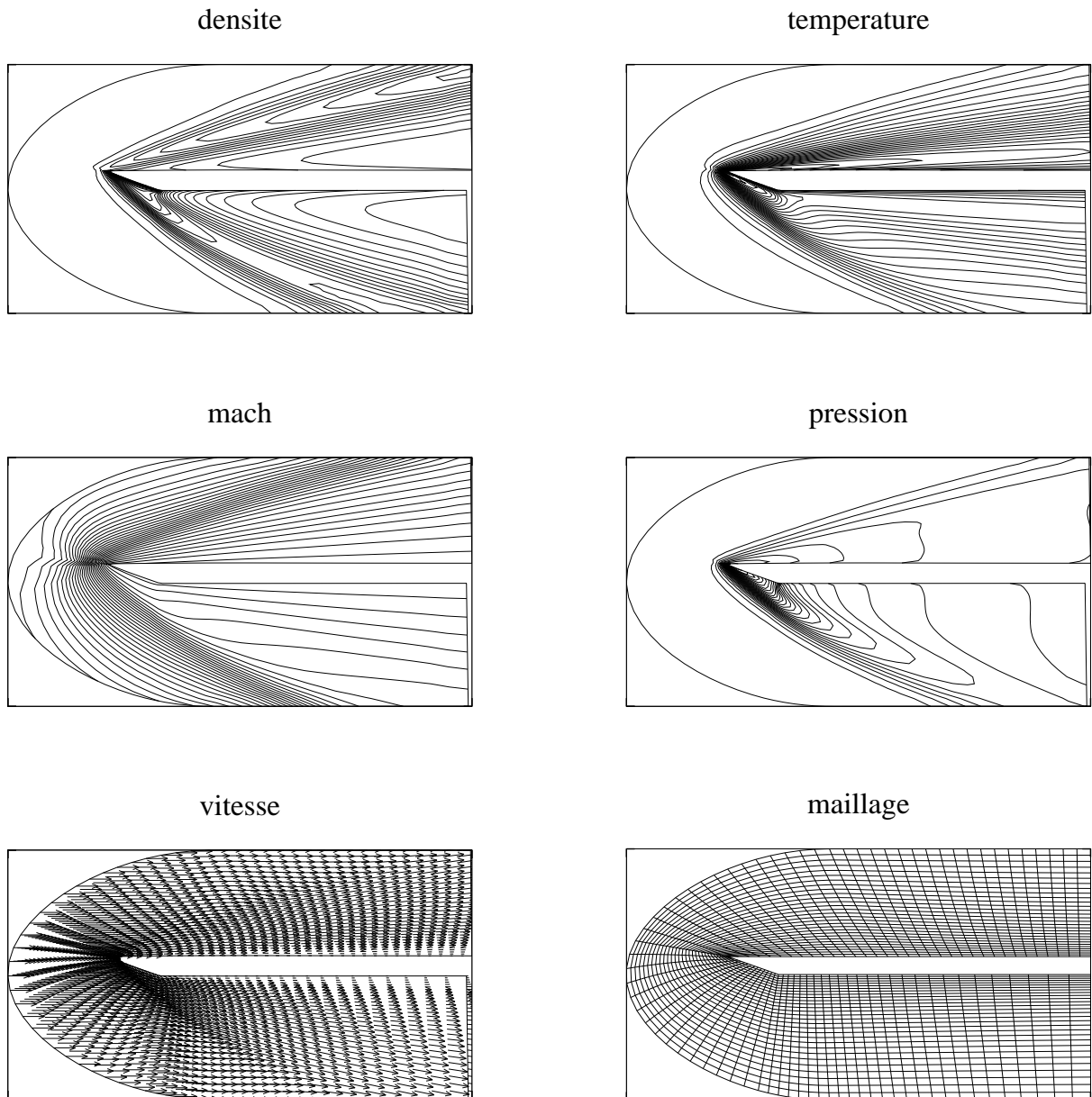


FIG. 6.46 – Isovaleurs et maillage pour l'écoulement hypersonique sur une plaque biseautée.

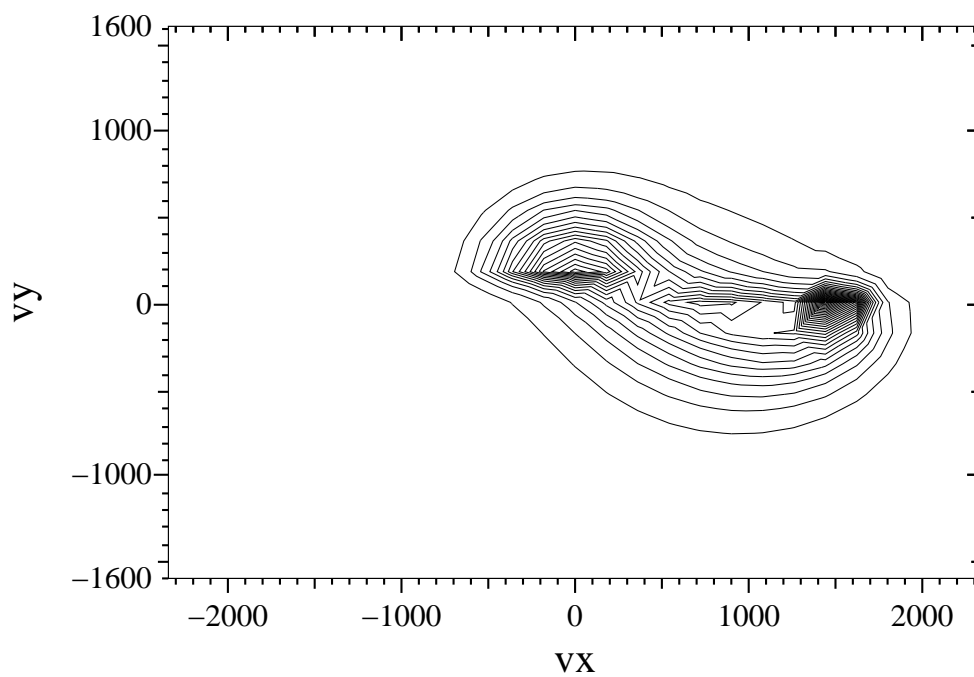


FIG. 6.47 – *Fonction de distribution réduite $F(v_x, v_y)$ au bord d'attaque de la plaque biseautée. On voit la demi-maxwellienne de la paroi centrée en $v_x = 0$ et la maxwellienne de l'écoulement amont centrée sur la vitesse amont ($u_x = 1500$).*

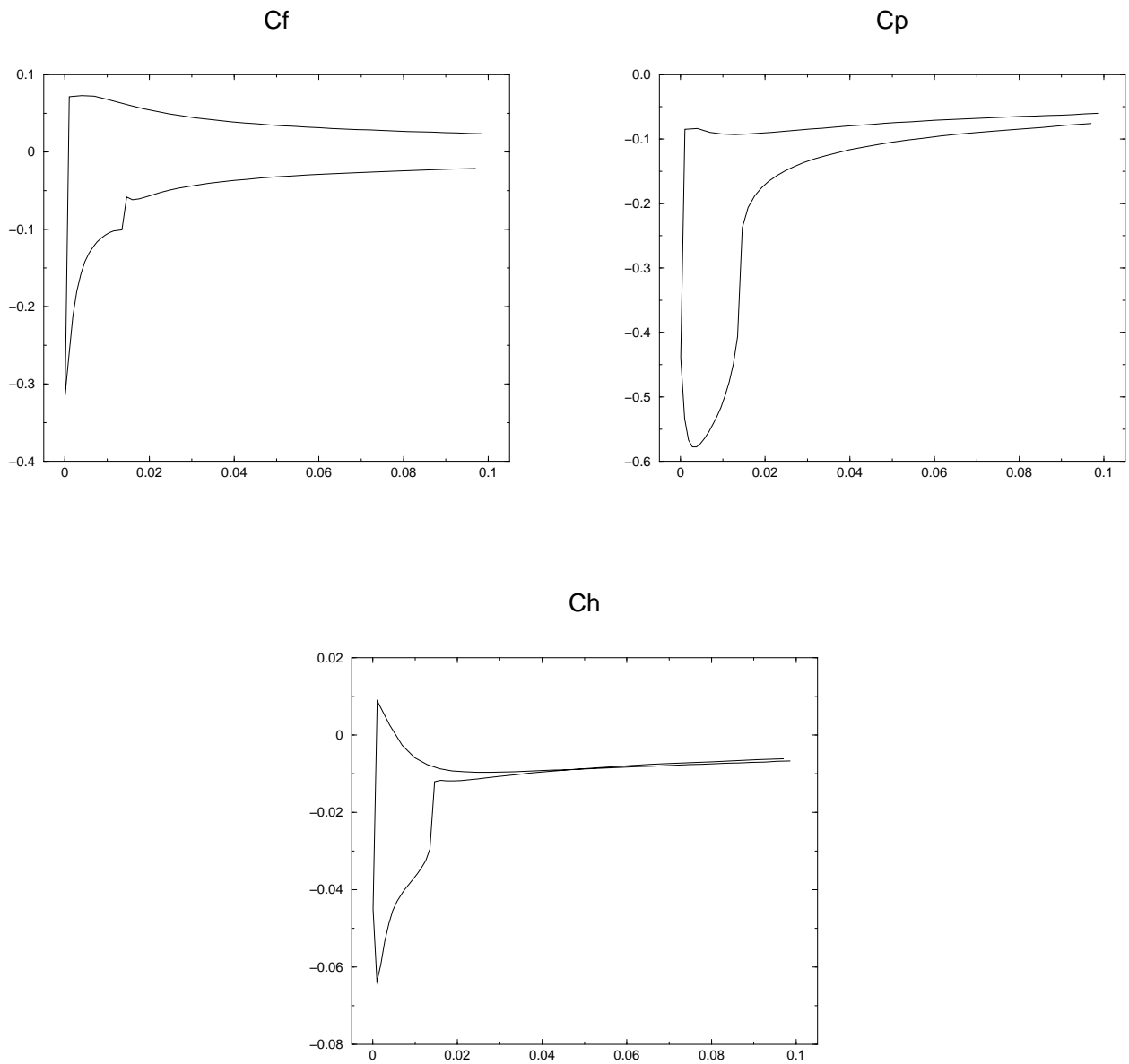


FIG. 6.48 – Coefficients aérodynamiques pour l'écoulement hypersonique sur une plaque biseautée.

Écoulement autour d'un cylindre

Ce dernier test est un écoulement plan à Mach 20 sur un cylindre infini que nous calculons avec le modèle BGK. Les caractéristiques suivantes correspondent à celle d'un écoulement à 90 km d'altitude :

gaz	argon
masse moléculaire (kg)	$0.663 \cdot 10^{-25}$
exposant ω de la loi de viscosité (modèle VHS)	0.5
température de paroi (K)	293
densité amont ($kg.m^{-3}$)	$0.317 \cdot 10^{-5}$
température amont (K)	249
Mach amont	20
rayon R du cylindre (m)	1
Knudsen amont (basé sur R)	$3.6 \cdot 10^{-2}$
Knudsen au nez (basé sur R)	$3.4 \cdot 10^{-3}$

Les bornes de la grille évaluées par un calcul Navier-Stokes sont $[-12000, 12000] \times [-11000, 11000] \times [-10000, 10000]$. Elles sont essentiellement données par la température maximale. Le même calcul Navier-Stokes donne un nombre de vitesses nécessaires de $47 \times 47 \times 45$, ce qui est énorme. Pour des raisons de temps, nous n'avons pas tenté de mener un tel calcul. Nous avons donc choisi de diminuer par deux le nombre de vitesses dans chaque direction, prenant alors $21 \times 21 \times 21$ vitesses. Nous avons par ailleurs utilisé un maillage assez grossier de 20×30 mailles seulement.

Il ne faut donc pas s'attendre ici à des résultats quantitativement corrects. Le but de ce calcul est seulement de voir si le code réussit à mener le calcul jusqu'à convergence. Il s'agit en effet d'un cas très sévère où il apparaît un choc frontal fort, avec des températures extrêmement grandes (supérieures à $30000 K$). Il constitue donc un bon test pour nos algorithmes. On peut aussi remarquer que comme dans le test précédent, on est obligé de prendre un domaine de calcul très grand, pour capturer les isovaleurs du nombre de Mach.

Le calcul converge en 1200 itérations et 66 h CPU. Les résultats obtenus sont donnés dans la figure 6.49 sous forme d'isovaleurs et de champ de vitesse. Comme prévu, les courbes isovaleurs de Mach sont très étalées, mais les autres courbes semblent bien calculées. On note une légère oscillation en fin de choc, probablement due au maillage grossier dont les lignes ne sont pas bien alignées avec l'écoulement. Comme cet écoulement est assez dense et très rapide, nous avons choisi de ne pas faire de calcul DSMC pour comparer nos résultats, car cela aurait été trop coûteux en temps.

Ce test prouve donc la robustesse de la méthode implicite et de l'algorithme de Newton. Il prouve aussi l'intérêt de notre modèle discret, qui est conservatif et qui dissipe l'entropie, puisque le code reste stable avec un faible nombre de vitesses, ce qui n'est probablement pas le cas avec d'autres modèles. En revanche, ce test montre aussi les limitations inhérentes à l'approche déterministe à vitesses discrètes. Le nombre de vitesses nécessaire dépend du nombre de Mach de l'écoulement et peut rapidement conduire à des calculs relativement coûteux.

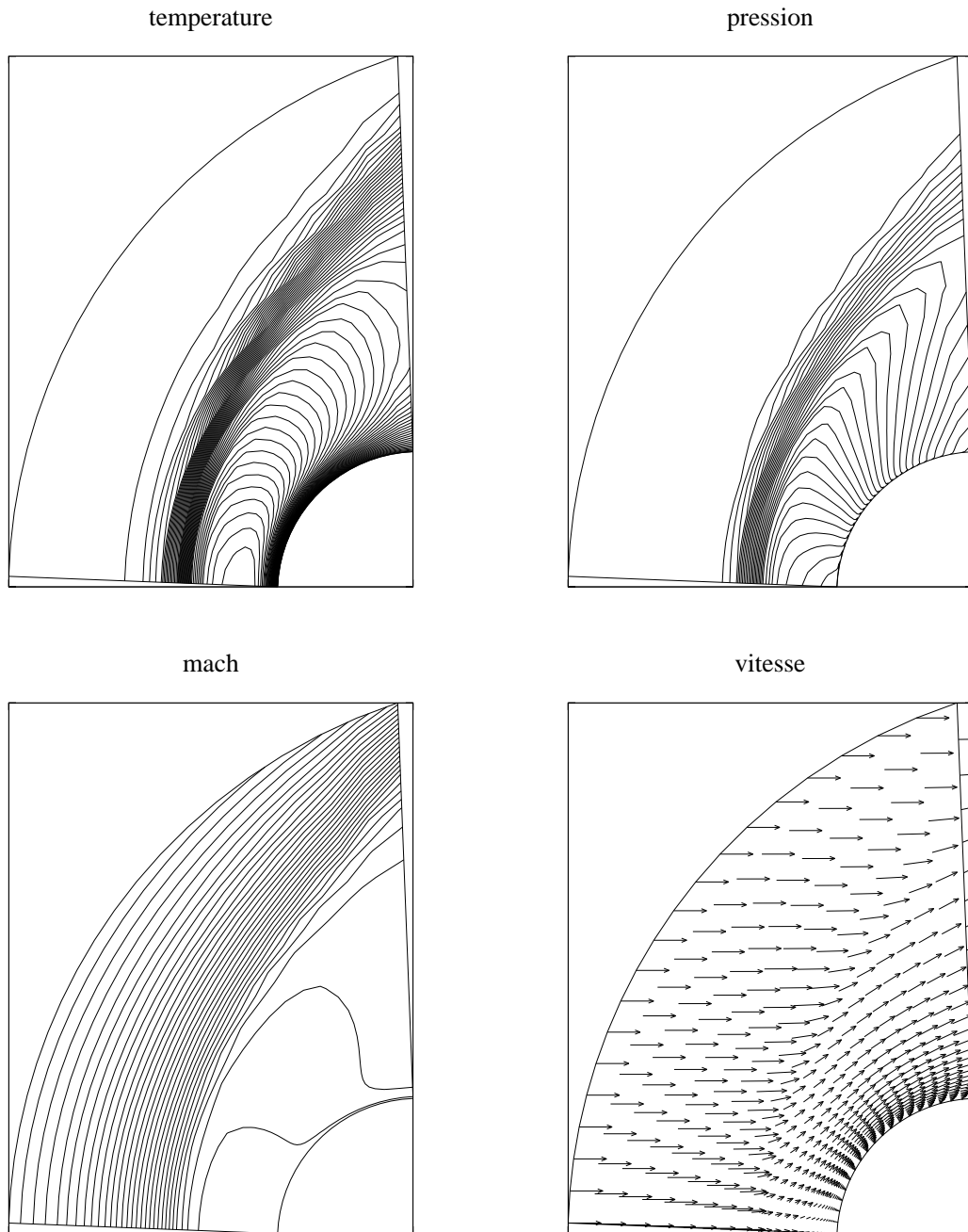


FIG. 6.49 – Isovaleurs pour un écoulement hypersonique à Mach 20 autour d'un cylindre.

Conclusion

Nous avons proposé une nouvelle méthode de résolution numérique de l'équation BGK, basée sur un modèle à vitesses discrètes et un schéma implicite linéarisé. Tout au long de ce travail, nous nous sommes attachés à obtenir des discrétisations possédant les mêmes propriétés importantes que les modèles continus : positivité, conservation des quantités macroscopiques, dissipation de l'entropie. Nous nous sommes assurés que nos modèles reposent sur des bases mathématiques solides, comme l'existence de la fonction d'équilibre discrète, ou la convergence du modèle à vitesses discrètes.

Nous avons validé ces méthodes sur des essais numériques qui ont mis en évidence le potentiel de notre approche pour calculer des écoulements stationnaires.

Nous avons écrit pour cela un code opérationnel 2D plan et 2D axisymétrique qui s'avère robuste et efficace. Ce code va servir de maquette pour le développement par le CEA d'un code 3D industriel.

Dans notre approche, nous avons choisi d'utiliser l'équation BGK, qui n'est qu'une équation modèle de l'équation de Boltzmann. Mais dans de nombreux cas, si le gaz n'est pas trop raréfié (encore que cette notion ne semble pas si claire), nos résultats sont proches de ceux obtenus avec l'équation de Boltzmann (simulée par DSMC), et les résultats obtenus avec les équations de Navier-Stokes ne sont en général pas bons. Ce sont précisément ces écoulements qui nous intéressaient à l'origine, et nous avons prouvé que notre approche est compétitive dans ces cas-là. De plus, notre méthode reste physiquement valide et d'un coût abordable pour certains régimes très denses, contrairement à DSMC, ce qui peut être important dans un contexte industriel.

Il est cependant clair que le domaine de validité de notre méthode ne couvre pas totalement celui de l'équation de Boltzmann, et il faudrait sans doute poursuivre cette étude pour préciser l'étendue de ce domaine.

Le problème du nombre de Prandtl de l'équation BGK n'est pas négligeable dans les situations où les transferts thermiques sont importants. Il semble être résolu par le modèle BGK-ES dans certains des cas étudiés, mais cela est loin d'être clair quand le nombre de Knudsen devient grand.

Nous envisageons à présent des développements de notre méthode dans plusieurs directions. D'un point de vue numérique, on peut remarquer que pour accélérer le calcul des écoulements plans 2D, on pourrait tenir compte de la parité de la fonction de distribution par rapport à v_z pour diviser par deux le nombre de vitesses discrètes (comme nous l'avons fait pour l'équation axisymétrique). De façon plus importante, on pourrait même éliminer la dépendance de f en v_z par l'introduction de deux fonctions de distributions réduites. Ceci a déjà été fait par de nombreux auteurs, mais nous ne l'avons pas réalisé dans ce travail car le but était de développer une approche générale, valable pour le 2D et le 3D. L'étude des schémas axisymétriques pour l'équation de transport mériterait aussi d'être approfondie, en particulier pour développer des extensions à l'ordre deux du schéma DEC-T. Nous avons en effet constaté qu'en dépit de ses bonnes propriétés (positivité, conservation, entropie), ce schéma manque de précision si la discrétisation n'est pas assez fine. Un autre développement possible est une amélioration de la grille de vitesses. Nous avons vu qu'une limitation de notre méthode est

le nombre important de vitesses discrètes nécessaire quand le nombre de Mach augmente. Ceci est dû au fait que notre grille de vitesse est fixe et de pas constant. On pourrait facilement envisager l'extension de la méthode à un pas non constant, mais le développement de notre modèle à grille variable en espace est lui un sujet plus difficile.

D'un point de vue physique, il serait intéressant de prendre en compte plus de complexité dans le modèle. Par exemple, nous envisageons dans un travail en cours de développer un modèle BGK à vitesses discrètes pour les gaz polyatomiques. Il paraît aussi nécessaire d'étudier de façon propre l'adaptation au cadre discret en vitesse des conditions aux limites spéculaires, même si ce type de réflexion n'est pas prépondérant dans les applications réelles. De plus, des défauts dans les résultats obtenus avec BGK suggèrent aussi d'étudier les modèles BGK avec temps de relaxation dépendant de la vitesse, pour mieux traiter les zones de chocs.

Enfin, nous envisageons d'utiliser notre approche de discrétisation en vitesse et de calcul de solutions stationnaires par schémas implicites pour d'autres problèmes de la physique faisant intervenir des termes de relaxation et une notion d'équilibre, comme par exemple l'équation du transfert radiatif.

Bibliographie

- [1] H. Alsmeyer. Density Profiles in Argon and Nitrogen Shock Waves Measured by the Absorption of an Electron beam. *J. Fluid Mech.*, 74(3):497–513, 1976.
- [2] P. Andries, P. Le Tallec, J.-P. Perlat, and B. Perthame. The Gaussian-BGK Model of Boltzmann Equation with Small Prandtl Number. preprint.
- [3] P. Andriès, J.-F. Bourgat, P. Le Tallec, and B. Perthame. Etude de modèles d'énergie interne bgk pour le calcul d'écoulements raréfiés. Technical report, INRIA, 1998.
- [4] K. Aoki, K. Kanba, and S. Takata. Numerical Analysis of a Supersonic Rarefied Gas Flow Past a Flat Plate. *Phys. Fluids*, 9(4), 1997.
- [5] G.A. Bird. *Molecular Gas Dynamics and the Direct Simulation of Gas Flows*. Oxford Science Publications, 1994.
- [6] A.V. Bobylev, A. Palczewski, and J. Schneider. A Consistency Result for a Discrete-Velocity Model of the Boltzmann Equation. *Siam J. Numer. Anal.*, 34(5):1865–1883, 1997.
- [7] A.V. Bobylev and J. Struckmeier. Implicit and iterative methods for the Boltzmann equation. *Transp. Th. Stat. Phys.*, 25(2):175–195, 1996.
- [8] F. Bouchut and B. Perthame. A BGK model for small prandtl number in the navier-stokes approximation. *Journal Stat. Physics*, 71(1-2):191–207, 1993.
- [9] H. Brezis. *Analyse fonctionnelle, théorie et applications*. Masson-Paris, 1983.
- [10] C. Buet. Résolution déterministe de l'équation de Boltzmann. Technical report, CEA, 1994.
- [11] C. Buet. A Discrete-Velocity Scheme for the Boltzmann Operator of Rarefied Gas Dynamics. *Transp. Th. Stat. Phys.*, 25(1):33–60, 1996.
- [12] C. Cercignani. *The Boltzmann Equation and Its Applications*, volume 68. Springer-Verlag, Lectures Series in Mathematics, 1988.
- [13] C. Cercignani. Temperature, Entropy, and Kinetic Theory. *J. Stat. Phys.*, 87(5-6), 1997.
- [14] S. Chapman and T.G. Cowling. *The Mathematical Theory of Non-Uniform Gases*. Cambridge University Press, 1970.
- [15] P. Charrier, B. Dubroca, and J.-L. Feugeas. Levermore's moment closure of discrete Boltzmann equations for non-equilibrium kinetic flows. In *21st RGD Symposium, Book of Abstracts*, Marseille, 1998.

- [16] P. Charrier, B. Dubroca, J.-L. Feugeas, and L. Mieussens. Discrete-velocity models for kinetic nonequilibrium flows. *C.R Acad. Sci.*, Serie I, 326(11):1347–1352, june 1998. Paris.
- [17] P. Charrier, B. Dubroca, and L. Mieussens. A numerical method for rarefied flow computation using a discrete velocity BGK model. In *21st RGD Symposium, Book of Abstracts*, Marseille, 1998.
- [18] F. Coron and B. Perthame. Numerical Passage from Kinetic to Fluid Equations. *SIAM J. Numer. Anal.*, 28(1):26–42, 1991.
- [19] P. Degond and B. Lucquin-Desreux. An entropy scheme for the Fokker-Planck collision operator of plasma kinetic theory. *Numer. Math.*, 68:239–262, 1994.
- [20] J.E. Dennis and R.B. Schnabel. *Numerical Methods for Unconstrained Optimization and Non-linear Equations*. Prentice-Hall Series in Computational Mathematics, 1983.
- [21] R.J. DiPerna and P.L. Lions. On the cauchy problem for Boltzmann equations: Global existence and weak stability. *Ann. Math*, **130**:321–366, 1989.
- [22] E. Gabetta, L. Pareschi, and G. Toscani. Relaxation Schemes for Non Linear Kinetic Equations. *SIAM J. Numer. Anal.*, 34(6):2168–2194, 1997.
- [23] V. Garzó and A. Santos. Comparison Between the Boltzmann and BGK Equations for Uniform Shear Flows. *Physica A*, 213:426–434, 1995.
- [24] R. Gatignol. *Théorie Cinétique des Gaz à Répartition Discrète de Vitesses*, volume 36 of *Lecture Notes in Physics*. Springer-Verlag, 1975.
- [25] L. Giraud, D. d’Humière, and P. Lallemand. A lattice-Boltzmann model for visco-elasticity. *Internat. Journal Modern Physics C*, 8(4):805–815, 1997.
- [26] F. Gropengiesser, H. Neunzert, J. Struckmeier, and B. Wiesen. Rarefied gas flow around a disc with different angles of attack. In A. E. Beylich, editor, *Rarefied gas dynamics*, pages 546–553. VCH, Weinheim, 1991.
- [27] A. Heintz and V. Panferov. A new consistent discrete-velocity model for the Boltzmann equation. preprint 1999:11, Chalmers, Göteborg University.
- [28] L. H. Holway. Kinetic theory of shock structure using an ellipsoidal distribution function. In New York Academic Press, editor, *Rarefied Gas Dynamics, Vol. 1 (Proc. Fourth Internat. Sympos. Univ. Toronto, 1964)*, pages 193–215, 1966.
- [29] D. Issautier. Convergence of a Weighted Particle Method for Solving The Boltzmann (B.G.K.) Equation. *SIAM J. Numer. Anal.*, 33(6):2099–2119, 1996.
- [30] S. Jin. Runge-Kutta method for hyperbolic conservation laws with stiff relaxation terms. *Journal of Comp. Phys.*, 122:51–67, 1995.
- [31] M. Junk. Domain of definition of Levermore’s five-moment system. *Journal Stat. Physics*, 93(5-6):1143–1168, 1998.
- [32] I. N. Larina and V. A. Rykov. A Method for Computing Axially Symmetric Rotating Flows of a Rarefied Gas at Small Knudsen Numbers. *Comp. Math. and Math. Phys*, 38(11):1936–1942, 1998.

- [33] I. N. Larina and V. A. Rykov. A Numerical Method for Calculating Axisymmetric Rarefied Gas Flows. *Comp. Maths. math. Phys.*, 38(8):1391–1403, 1998.
- [34] M. Lemou. Opérateur de Fokker-Planck-Landau axisymétrique. *C. R. Acad. Sci Paris, Série 1*, 326:179–184, 1998.
- [35] J.-C. Lengrand. Mise en oeuvre de la méthode de Monte Carlo pour la simulation numérique d'un écoulement de gaz raréfié. Technical report, CNRS, 1986. RC 86-4.
- [36] C.D. Levermore. Moment Closure Hierarchies for Kinetic Theories. *J. Stat. Phys.*, 83:1021–1065, 1996.
- [37] P.L. Lions. Compactness in Boltzmann's Equation via Fourier Integral Operators and Applications. II. *J. Math. Kyoto Univ*, 34(2):429–461, 1994.
- [38] G. Liu. A method for constructing a model for the Boltzmann equation. *Phys; Fluids A*, 2, 1990.
- [39] S. Mas-Gallic. A deterministic particle method for the linearized Boltzmann equation. *Transport Theory Statist. Phys.*, 16:855–887, 1987.
- [40] G. McNamara and G. Zanetti. Use of the Boltzmann equation to simulate lattice gas automata. *Phys. Rev. Lett.*, 61:2332, 1988.
- [41] L. Mieussens. An Implicit Discrete Velocity Scheme for the BGK Equation of Rarefied Gas Dynamics. In *First European Symposium on Applied Kinetic Theory, Book of Abstracts*, Toulouse, 1998.
- [42] L. Mieussens. An Implicit Discrete Velocity Scheme for the BGK Equation of Rarefied Gas Dynamics. In *16th ICNMF D, Book of Abstracts*, Arcachon, 1998.
- [43] L. Mieussens. Discrete Velocity Model and Implicit Scheme for the Bgk Equation of Rarefied Gas Dynamics. accepté pour publication à *Math. Models and Meth. in Applied Sciences*, 1998.
- [44] L. Mieussens. Convergence d'un modèle à vitesses discrètes pour l'équation de Boltzmann-BGK. *C.R Acad. Sci.*, Série I(328):1231–1236, 1999.
- [45] S. Mischler. Uniqueness for the BGK equation in \mathbb{R}^n and rate of convergence for a semi-discrete scheme. *Diff. and Int. Equations*, 9, 1996.
- [46] S. Mischler. Convergence of discrete velocity schemes for the Boltzmann equation. *Arch. Rat. Mech. Anal.*, 140:53–77, 1997.
- [47] K. Nanbu. Direct simulation scheme derived from the Boltzmann equation. *J. Phys. Soc. of Japan*, 45:2042–2049, 1980.
- [48] J-P. Perlat. Modélisation et calcul parallèle d'une couche limite cinétique. Thèse de l'université Paris 6, 1998.
- [49] B. Perthame. Global existence to the BGK model of Boltzmann Equation. *J. Diff. Eq.*, 82:191–205, 1989.
- [50] B. Perthame and M. Pulvirenti. Weighted L^∞ bounds and uniqueness for the Boltzmann BGK model. *Arch. Rat. Mech. Anal.*, 125:289–295, 1993.

- [51] E.P. Gross P.L. Bathnagar and M. Krook. A model for collision processes in gases. I. small amplitude processes in charged and neutral one-component systems. *Phys. Rev.*, 94:511–525, 1954.
- [52] Y.H. Qian, D. d’Humières, and P. Lallemand. Lattice bgk models for the navier-stokes equations. *Europhys. Lett.*, 17:479, 1992.
- [53] P.A. Raviart. An Analysis of Particle Methods. In F. Brezzi, editor, *Numerical Methods in Fluid Dynamics*, volume 1127 of *Lecture Notes in Mathematics*, pages 243–324. Springer-Verlag, Berlin, Heidelberg, New York, 1984.
- [54] E. Ringeisen. *Contributions à l’étude mathématique des équations cinétiques*. Thèse, Université Paris VII, 1991.
- [55] F. Rogier and J. Schneider. A Direct Method For Solving the Boltzmann Equation. *Transp. Th. Stat. Phys.*, 23(1-3):313–338, 1994.
- [56] W. Rudin. *Real and complex analysis*. Mcgraw-Hill, 1966.
- [57] J. Schneider. Une méthode déterministe pour la résolution de l’équation de Boltzmann. Thèse, Université Paris VI, 1993.
- [58] Y. M. Shakhov. The Axisymmetric Non-Linear Steady Flow of a Rarefied Gas in a Pipe of Circular Cross-Section. *Comp. Maths and Math. Phys.*, 36(8):1123–1131, 1996.
- [59] Ye. M. Shakhov. The Two-dimensional Non-linear Problem of the Motion of a Rarefied Gas Between Two Parallel Plates. *Zh. Vychisl. Mat. Mat. Fiz.*, 35:83–94, 1994.
- [60] Y. Sone, S. Takata, and H. Sugimoto. The behavior of a gas in the continuum limit in the light of kinetic theory: the case of cylindrical couette flows with evaporation and condensation. *Phys. Fluids*, 8(12), 1996.
- [61] H. Struchtrup. The BGK-Model with Velocity-Dependent Collision Frequency. *Continuum Mech. Thermodyn.*, 9:23–31, 1997.
- [62] H. Sugimoto and Y. Sone. Numerical analysis of steady flows of a gas evaporating from its cylindrical condensed phase on the basis of kinetic theory. *Phys. Fluids A*, 4(2), 1992.
- [63] W. Waluś. *Lecture Series in Mathematical Theory of the Boltzmann Equation*. Series on Advances in Mathematics for Applied Sciences. ed. N. Bellomo (World Scientific), 1994.
- [64] L.C. Woods. *An Introduction to the Kinetic Theory of Gases and Magnetoplasmas*. Oxford Science Publications, 1992.
- [65] J.Y. Yang and J.C. Huang. Rarefied Flow Computations Using Nonlinear Model Boltzmann Equations. *Journal of Computational Physics*, 120:323–339, 1995.
- [66] H. C. Yee. *A Class of High-Resolution Explicit and Implicit Shock-Capturing Methods*. von Karman Institute for Fluid Dynamics, Lectures Series, n°4. 1989.