l'université nantes angers le mans
PÔLE DE RECHERCHE ET D'ENSEIGNEMENT SUPÉRIEUR

UNIVERSITÉ DE NANTES

Thèse de Doctorat

Vivien DESVEAUX

Mémoire présenté en vue de l'obtention du
grade de Docteur de l'Université de Nantes
sous le label de l'Université de Nantes Angers Le Mans

École doctorale : Sciences et technologies de l'information et mathématiques

Discipline : Mathématiques et leurs interactions, section CNU 26
Unité de recherche : Laboratoire de Mathématiques Jean Leray (LMJL)

Soutenue le 26 novembre 2013

Contribution à l'approximation numérique des systèmes hyperboliques

JURY

Président :	**M. Stéphane CLAIN**, Professeur, Universidade do Minho
Rapporteurs :	**M. François JAMES**, Professeur, Université d'Orléans **M. Frédéric LAGOUTIÈRE**, Professeur, Université Paris-Sud
Examinateurs :	**M. Jean-François COULOMBEL**, Directeur de recherche CNRS, Université de Nantes **M. Christian KLINGENBERG**, Professeur, Universität Würzburg
Directeur de thèse :	**M. Christophe BERTHON**, Professeur, Université de Nantes
Co-directeur de thèse :	**M. Yves COUDIÈRE**, Professeur, Université Bordeaux I

Remerciements

Ces trois années de thèse furent pour moi une magnifique expérience scientifique, professionnelle et humaine. De nombreuses personnes ont contribué de près ou de loin à l'aboutissement de ce travail et je profite de ces quelques lignes pour leur exprimer toute ma gratitude.

En premier lieu, je tiens à remercier très vivement Christophe Berthon qui a su me transmettre sa passion pour les systèmes hyperboliques et les schémas volumes finis. Ses nombreux conseils et encouragements ont été un soutien précieux tout au long de ce travail. J'ai beaucoup apprécié sa disponibilité, son optimisme et sa persévérance. Ce fut pour moi un réel plaisir de travailler avec lui.

Je souhaite aussi adresser mes plus sincères remerciements à Yves Coudière pour tout ce qu'il m'a appris, que ce soit pendant mon année de M2 ou pendant ma thèse. Ses grandes qualités humaines ont également été très importantes pour moi.

Je remercie chaleureusement François James et Frédéric Lagoutière pour avoir accepté de rapporter mon travail et pour la pertinence de leurs remarques. Je suis également très reconnaissant envers Christian Klingenberg, Stéphane Clain et Jean-François Coulombel d'avoir accepté d'être membres de mon jury et pour l'intérêt qu'ils ont porté à mon travail.

J'aimerais adresser un remerciement particulier à Christian Klingenberg pour l'opportunité qu'il m'a offerte en m'invitant à l'université de Würzburg. Cela a été une expérience à la fois très agréable et fructueuse, puisqu'elle a débouché sur le quatrième chapitre de ce manuscrit. Je tiens également à le remercier pour ses nombreux conseils éclairés et pour sa gentillesse. J'en profite pour remercier toute son équipe et en particulier Markus, Ujjwal, Benedikt et Gero. Cela a été très stimulant de travailler avec eux.

Un grand merci à Guy Moebs pour ses compétences en optimisation et en parallélisation de codes. Son aide m'a été très précieuse : sans lui, certains de mes codes seraient probablement encore en train de tourner !

Je remercie Rodolphe Turpault pour les nombreuses discussions, mathématiques ou autres, que l'on a pu avoir, notamment pendant les dix-sept heures du voyage Padoue-Nantes. Merci aussi aux autres numériciens qui ont toujours été de bon conseil, notamment Hélène, Nicolas, Françoise et Anaïs.

Je souhaite remercier de manière générale tous les membres du laboratoire Jean Leray qui contribuent à en faire un environnement propice à la recherche, à la fois studieux et convivial. Je remercie en particulier tout le personnel administratif et technique qui sont toujours disponibles, souriants et efficaces (même dix minutes avant ma soutenance).

Et puis comment ne pas mentionner la formidable équipe des thésards ? Je les remercie pour l'excellente ambiance, pour leur soutien sans faille dans les moments difficiles, pour les pauses café, pour les soirées jeux, pour les tournois FIFA, pour les nombreuses discussions improbables, pour la taroinche... Je commence par ma « grande » sœur Céline avec qui j'ai partagé bien plus que la thématique hyperbolique. Je la remercie pour son sourire constant et sa bonne humeur permanente, pour ses invitations à des soirées jeux/expériences culinaires et pour ses bruits loufoques. Je n'oublierai pas non plus les quelques « blagues » amicales que l'on

a échangé (d'ailleurs j'ai gagné !). Je continue avec mon co-bureau permanent, Alex Q, qui m'a répété au moins 42 fois de regarder H2G2. J'ai beaucoup apprécié être avec lui au quotidien pendant ces trois années. Je le remercie notamment pour les discussions sans queue ni tête sur les virus, la redéfinition du temps et le scrabo. Je remercie également les autres qui à moment ou un autre ont partagé mon bureau : Julien pour son accueil chaleureux et pour ses encouragements, Vincent pour ses talents de quizzeur, Thomas pour ses connaissances encyclopédiques en matière de fantasy et de jeux et pour son goût prononcé pour les fichiers openoffice et Tristan pour ses compétences en LaTeX et notre passion commune pour ASOIAF. Un grand merci à Alex U avec qui j'ai partagé non seulement le module d'analyse numérique L3, mais aussi quelques voyages Angers-Nantes, un bon paquet d'énigmes et pas mal de jeux (online ou pas). Merci à Carlos pour son goût pour la langue française, à Hermann pour l'invention des GDTC, à Anne pour ses conseils concernant l'enseignement, à Salim pour m'avoir rappelé le théorème de Lusin, à Baptiste pour le pique-nique en bord de Loire, à Nicolas pour son calme à toute épreuve, à Carl pour ses soirées arrosées, à Gilberto pour le laser game, à Christophe pour le brin de folie qu'il a apporté au labo, à Ilaria pour son accent Italien, à Moudhaffar pour ses cours de géographie tunisienne, à Virgile pour ses réponses à mes questions sur les distributions, à Antoine pour les discussions sur la CLU pendant les pauses café. Et puis bonne chance aux derniers arrivés, Damien, Florian, Pierre, Valentin et Victor.

Je terminerai en remerciant ma famille qui m'a soutenu et encouragé pendant ces trois années, en particulier mes parents qui ont accepté de relire ma thèse sans rien y comprendre.

Table des matières

Introduction

Ce travail a pour objectif l'approximation numérique des solutions faibles des systèmes hyperboliques de lois de conservation par des méthodes de volumes finis. Dans différents contextes, on se concentre sur la construction de schémas numériques possédant des « bonnes propriétés » de précision, de robustesse et de stabilité. La propriété de précision correspond généralement à demander que les schémas développés soient d'ordre élevé. Pour que le schéma soit robuste, il doit préserver l'ensemble des états physiquement admissibles. Enfin, la stabilité est le plus souvent comprise en termes d'inégalités d'entropie discrètes. En effet, les solutions faibles des systèmes hyperboliques doivent vérifier des inégalités d'entropie pour être physiquement admissibles. Pour avoir un bon comportement, un schéma numérique doit donc vérifier un équivalent discret de ces inégalités.

Lorsque l'on rajoute un terme source à un système de lois de conservation, une autre propriété de précision est requise : il est souhaitable que le schéma numérique préserve le plus précisément possible les états d'équilibres du système.

On considère dans un premier temps un système hyperbolique de lois de conservation en une dimension d'espace, sans terme source, de la forme

$$\partial_t w + \partial_x f(w) = 0, \quad x \in \mathbb{R}, \quad t \in \mathbb{R}^+. \tag{1}$$

La fonction inconnue $w : \mathbb{R} \times \mathbb{R}^+ \to \Omega$ est à valeurs dans un ensemble convexe $\Omega \subset \mathbb{R}^d$ d'états physiquement admissibles et $f : \Omega \to \mathbb{R}^d$ est la fonction flux.

On suppose que le système (1) est hyperbolique, c'est-à-dire que la matrice jacobienne $\nabla_w f(w)$ est diagonalisable dans $\mathbb{R}$. Les solutions de (1) pouvant développer des discontinuités (voir [56, 80]), on considère les solutions au sens faible, c'est-à-dire au sens des distributions. Par ailleurs, afin d'écarter les solutions non physiques, on a besoin d'un critère supplémentaire : les inégalités d'entropie (voir [76, 77, 93]). On dit qu'une fonction convexe $\eta : \Omega \to \mathbb{R}$ est une entropie pour le système (1) s'il existe une fonction $\mathcal{G} : \Omega \to \mathbb{R}$, appelée flux d'entropie, telle que $\nabla_w f \nabla_w \eta = \nabla_w \mathcal{G}$. Une solution faible w du système (1) est alors dite entropique si pour tout couple d'entropie $(\eta, \mathcal{G})$, w vérifie au sens des distributions l'inégalité d'entropie

$$\partial_t \eta(w) + \partial_x \mathcal{G}(w) \leq 0. \tag{2}$$

On s'intéresse alors à l'approximation numérique des solutions de (1). De nombreuses méthodes de volumes finis d'ordre un existent dans la littérature. Le lecteur pourra trouver en détail les méthodes usuelles dans [56, 80, 20, 100, 59]. Considérons une discrétisation uniforme de $\mathbb{R}$ en cellules $K_i = [x_{i-1/2}, x_{i+1/2}[$, avec $\Delta x = x_{i+1/2} - x_{i-1/2}$ le pas d'espace supposé constant. En notant w_i^n une approximation de $w(x_i, t^n)$, la mise à jour au temps $t^{n+1} = t^n + \Delta t$ par un schéma volumes finis d'ordre un s'écrit

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x} \left(F\left(w_i^n, w_{i+1}^n\right) - F\left(w_{i-1}^n, w_i^n\right) \right), \tag{3}$$

où $F : \Omega \to \Omega \to \mathbb{R}^d$ est un flux numérique consistant ($F(w, w) = f(w)$, pour tout $w \in \Omega$).

Le schéma (3) est dit entropique s'il vérifie les inégalités d'entropie discrètes

$$\eta(w_i^{n+1}) \leq \eta(w_i^n) - \frac{\Delta t}{\Delta x}\left(G\left(w_i^n, w_{i+1}^n\right) - G\left(w_{i-1}^n, w_i^n\right)\right),$$

pour tout couple d'entropie $(\eta, \mathcal{G})$, où $G : \Omega \to \Omega \to \mathbb{R}$ est un flux numérique d'entropie consistant ($G(w,w) = \mathcal{G}(w)$).

La robustesse et la stabilité sont des notions relativement simples à obtenir pour des schémas d'ordre un. En effet, de nombreux schémas d'ordre un préservent l'ensemble Ω et/ou sont entropiques, par exemple, le schéma HLL [64], les extensions aux solveurs de Riemann simples [52, 53, 29], le schéma HLLC [101, 7, 100], le schéma de relaxation de Suliciu [20, 11], le schéma de Roe et son extension VFRoe [91, 64, 50, 26, 17]... Notons que pour obtenir ces propriétés, il est nécessaire de limiter le pas de temps selon une condition introduite par Courant, Friedrichs et Lewy (condition CFL).

Les schémas d'ordre un contiennent beaucoup de viscosité numérique et par conséquent, ils sont peu précis, notamment au voisinage des discontinuités. Pour réduire ce défaut, de nombreuses méthodes ont été introduites au cours des dernières décennies afin d'augmenter l'ordre de précision. On peut mentionner une stratégie s'appuyant sur la résolution exacte du problème de Riemann généralisé (GRP) [8, 22, 23, 19]. L'inconvénient de cette approche est que la résolution exacte du problème de Riemann n'est possible que pour des systèmes très simples. On lui préfère donc souvent une méthode moins précise, mais plus facile à mettre en œuvre, consistant à utiliser une approximation d'ordre élevé de la fonction flux de la forme $F(w_i^{n,+}, w_{i+1}^{n,-})$. Ici, les états $w_i^{n,\pm}$ sont des reconstructions d'ordre élevé de la solution dans la cellule K_i aux interfaces $x_{i\pm1/2}$. Le schéma d'ordre élevé ainsi obtenu s'écrit

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(F\left(w_i^{n,+}, w_{i+1}^{n,-}\right) - F\left(w_{i-1}^{n,+}, w_i^{n,-}\right)\right). \tag{4}$$

Plusieurs techniques existent pour reconstruire les états $w_i^{n,\pm}$, parmi lesquelles on mentionne la reconstruction MUSCL [102, 86, 73, 21, 37, 80, 72, 12, 33], les approches cinétiques d'ordre deux [86, 73], la reconstruction ENO/WENO [88, 108, 109], la reconstruction PPM [35], la reconstruction MOOD [34, 47]...

Il n'est généralement pas très difficile de rendre robuste ces méthodes en introduisant une limitation adaptée dans la technique de reconstruction. Par exemple, dans [80, 20], les auteurs montrent que les méthodes MUSCL basiques préservent l'ensemble Ω, dans [88], la robustesse est établie dans le cadre de la méthode WENO et dans [12], la robustesse de méthodes plus complexes est étudiée. Dans tous ces travaux, la robustesse est obtenue sous une condition plus restrictive que la condition CFL du schéma d'ordre un (au moins deux fois plus restrictive). D'autre part, pour toutes ces méthodes, la procédure de limitation permettant d'obtenir la robustesse est effectuée *a priori*. Ces limitations globales peuvent s'avérer trop fortes et engendrer une perte de précision. Pour corriger ce défaut, la méthode MOOD a été récemment introduite dans [34, 47]. En activant la limitation *a posteriori*, seulement localement sur les cellules où cela s'avère nécessaire, cette technique permet de diminuer la viscosité numérique du schéma. Par ailleurs, la méthode MOOD est naturellement robuste par construction sous la condition CFL d'ordre un, ce qui représente un gain en temps de calcul.

Il est beaucoup plus délicat de montrer la stabilité des méthodes d'ordre élevé. Plusieurs stratégies ont été proposées mais se révèlent peu concluantes. Une première approche a été proposée par [8, 22, 23] pour les méthodes GRP, mais reste limitée par la difficulté de la résolution exacte du problème de Riemann généralisé. De nouvelles techniques de projection ont été introduites dans [21, 37, 36], mais les méthodes numériques qui en résultent sont délicates à implémenter. Une autre méthode a été introduite dans [12] et des inégalités d'entropie discrètes

sont obtenues, mais pour un opérateur de dérivée discrète en temps non classique. Il n'est cependant pas clair que la solution convergée au sens du théorème de Lax-Wendroff, obtenue avec cette technique, vérifie les inégalités d'entropie (2) attendues.

On s'intéresse maintenant aux systèmes de lois de conservation en deux dimensions d'espace, qui s'écrivent sous la forme

$$\partial_t w + \partial_x f(w) + \partial_y g(w) = 0, \quad (x,y) \in \mathbb{R}^2, \quad t \geq 0, \tag{5}$$

où $w : \mathbb{R}^2 \times \mathbb{R}^+ \to \Omega$ est le vecteur des inconnues, $f, g : \Omega \to \mathbb{R}^d$ sont les fonctions flux respectivement dans la direction x et y et $\Omega \subset \mathbb{R}^d$ désigne l'ensemble convexe des états admissibles. L'hyperbolicité revient dans ce cadre à demander que pour tout vecteur unitaire $\nu = (\nu_x, \nu_y)^T \in \mathbb{S}^1$, la matrice

$$\nu_x \nabla_w f(w) + \nu_w \nabla_w g(w)$$

soit diagonalisable dans $\mathbb{R}$. On demande que les solutions faibles vérifient les inégalités d'entropie

$$\partial_t \eta(w) + \partial_x \mathcal{F}(w) + \partial_y \mathcal{G}(w) \leq 0, \tag{6}$$

pour tout triplet de fonctions $(\eta, \mathcal{F}, \mathcal{G})$, avec η convexe et

$$\nabla f \nabla \eta = \nabla \mathcal{F}, \quad \nabla g \nabla \eta = \nabla \mathcal{G}.$$

On présente brièvement le formalisme permettant de décrire les schémas volumes finis en deux dimensions d'espace (voir [56, 80, 100]). On considère une discrétisation $(K_i)_{i \in \mathbb{Z}}$ de $\mathbb{R}^2$. Pour chaque cellule K_i, on note $\gamma(i)$ l'ensemble des indices des cellules voisines de K_i. Pour tout $i \in \mathbb{Z}$ et $j \in \gamma(i)$, on introduit alors e_{ij} le côté commun entre K_i et K_j et ν_{ij} la normale extérieure à e_{ij}. Enfin, on note respectivement $|e_{ij}|$ et $|K_i|$ la longueur du côté e_{ij} et l'aire de la cellule K_i.

En notant w_i^n une approximation constante de la solution de (5) au temps t^n sur la cellule K_i, la mise à jour au temps $t^n + \Delta t$ par un schéma volumes finis d'ordre un est donnée par

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{|K_i|} \sum_{j \in \gamma(i)} |e_{ij}| \varphi(w_i^n, w_j^n, \nu_{ij}), \tag{7}$$

où $\varphi(\cdot, \cdot, \nu)$ est un flux numérique 1D dans la direction ν.

Pour montrer la robustesse et la stabilité du schéma (7), les techniques classiques consistent à réécrire se schéma comme une combinaison convexe de schémas 1D (voir [87, 88, 12]). Cependant, la condition CFL usuelle permettant d'obtenir ces propriétés n'est pas du tout optimale dès que le maillage considéré est non structuré. Cela peut entraîner des coûts de calcul excessifs.

On s'intéresse maintenant aux méthodes d'ordre élevé en deux dimensions d'espace. De la même manière qu'en une dimension d'espace, les méthodes usuelles consistent à utiliser des reconstructions d'ordre élevé dans l'évaluation des flux numériques. Ainsi, les schémas d'ordre élevé s'écrivent en 2D :

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{|K_i|} \sum_{j \in \gamma(i)} |e_{ij}| \varphi\left(w_{ij}, w_{ji}, \nu_{ij}\right), \tag{8}$$

où l'état w_{ij} est une reconstruction d'ordre élevé de la solution sur la cellule K_i au milieu du côté e_{ij}.

Plusieurs méthodes existent pour reconstruire les états w_{ij}. On se concentre ici sur la méthode MUSCL d'ordre deux qui consiste à reconstruire une approximation du gradient de la solution sur chaque cellule. Sur des maillages cartésiens, les reconstructions de gradients sont facilement définies par extension du cas 1D. Quand des maillages non structurés sont considérés,

l'évaluation précise des gradients est un problème plus délicat. Plusieurs techniques existent pour des cellules triangulaire (voir par exemple [83, 56, 82, 24, 27, 28]), mais ne s'étendent pas à des volumes de contrôle plus généraux.

L'autre difficulté concerne la limitation de la pente qui est nécessaire pour éviter l'apparition d'oscillations non physiques. Certains auteurs, comme dans [6, 103, 45], proposent des limiteurs globalement 2D afin de reconstruire des solutions plus précises. Pour diminuer la viscosité numérique dans les régions où la solution est régulière, il est possible d'utiliser un limiteur supplémentaire (voir [84, 74]). Une autre possibilité consiste à modifier les limiteurs usuels afin de satisfaire des conditions de stabilité du type principe du maximum (par exemple [69, 9, 81]). Enfin, plutôt que de reconstruire une pente unique sur chaque cellule, les auteurs de [24, 33] proposent d'introduire une pente pour chaque côté de la cellule.

La procédure de limitation permet aussi d'assurer que le schéma préserve Ω (voir [86, 87, 88, 43, 14]). La robustesse du schéma d'ordre élevé étant obtenue à partir de celle du schéma d'ordre un, les conditions CFL permettant d'assurer la robustesse de (8) sont souvent trop restrictives. Comme en 1D, la méthode MOOD, récemment introduite par Diot et al. [34, 47], préserve naturellement l'ensemble Ω sous la condition CFL d'ordre un.

On considère enfin des systèmes de lois de conservation avec un terme source de la forme

$$\partial_t w + \partial_x f(w) = s(w)\partial_x Z, \tag{9}$$

où $s : \Omega \to \mathbb{R}^d$ est un terme source et $Z : \mathbb{R} \to \mathbb{R}$ est une fonction régulière donnée. La particularité de ce type de système est l'existence de solution stationnaires non triviales, c'est-à-dire indépendantes du temps. Il est souhaitable que les schémas numériques construits pour approcher les solutions faibles de (9) capturent avec précision ces solutions stationnaires (ou certaines d'entre elles). On parle alors de schémas well-balanced (voir [10, 60, 61, 58]).

On s'intéresse à trois systèmes de la forme (9) : les équations de Saint-Venant avec topographie, les équations de Ripa avec topographie et les équations d'Euler avec gravité.

Le système de Saint-Venant est maintenant bien étudié. Il n'admet qu'un seul état d'équilibre au repos (avec une vitesse nulle), à constante près : le lac au repos. Cette solution est particulièrement simple, car elle est caractérisée par une relation algébrique linéaire. De nombreux schémas well-balanced existent pour le système de Saint-Venant (voir par exemple [25, 5, 20, 85, 15, 46]).

Pour le système de Ripa avec topographie, les états d'équilibre au repos sont gouvernés par une équation aux dérivées partielles non intégrable. Autrement dit, les états d'équilibre ne sont pas tous donnés par une relation algébrique explicite. Cela rend beaucoup plus compliquée la construction d'un schéma numérique préservant tous ces états stationnaires. Dans [32], les auteurs construisent un schéma numérique préservant deux états d'équilibre particuliers généralisant le lac au repos.

Enfin, la difficulté est encore plus grande pour les équations d'Euler avec gravité. Les états d'équilibre, comme ceux de Ripa sont décrits par une équation différentielle non intégrable. Par ailleurs, le système en lui-même est plus non linéaire que Saint-Venant ou Ripa, ce qui complique encore la dérivation de schéma well-balanced. On mentionne les travaux de Käppeli et Mishra [71] où un schéma préservant tous les états d'équilibre isentropiques est proposé.

Plan de la thèse

Premier chapitre : généralités sur les méthodes de volumes finis

Ce premier chapitre ne contient pas de nouveau résultat. Il présente un état de l'art des méthodes de volumes finis les plus classiques. On commence par introduire le formalisme général

des schémas volumes finis conservatifs et les propriétés usuelles de robustesse et de stabilité qu'ils doivent vérifier. On décrit ensuite le premier schéma volumes finis historique : le schéma de Godunov. Bien que l'on montre facilement que ce schéma est entropique, il présente plusieurs défauts qui le rendent de nos jours peu attractif. On s'intéresse alors à une classe de schémas qui généralise le schéma de Godunov. Il s'agit des schémas de type Godunov, introduits par Harten, Lax et van Leer [64]. Ces schémas sont basés sur l'utilisation de solveurs de Riemann approchés, contrairement au schéma de Godunov qui utilise la solution exacte du problème de Riemann. On exhibe aisément des conditions simples pour assurer qu'un schéma de type Godunov est robuste et entropique.

On s'intéresse ensuite aux schémas d'ordre élevé MUSCL introduits par van Leer [102]. Après avoir donné quelques exemples de limiteurs intervenant dans la procédure de reconstruction, on donne un lemme assurant la robustesse du schéma MUSCL dès que la reconstruction préserve Ω. On conclut en détaillant la méthode MOOD, présentée dans [34, 47]. Contrairement aux techniques MUSCL classiques, la procédure de limitation de la méthode MOOD est effectuée *a posteriori* et uniquement sur les cellules problématiques vis à vis de certains critères comme l'appartenance à Ω ou le respect d'un principe du maximum.

Deuxième chapitre : schémas MUSCL basés sur une reconstruction de gradients par maillages duaux (DMGR)

Ce chapitre est dédié à la construction de schémas d'ordre élevé performants sur des maillages 2D non structurés. Dans la première partie, on étudie la robustesse des schémas d'ordre élevé en deux dimensions d'espace. Des résultats similaires existent (voir [87, 88, 14]), mais sous des conditions CFL très restrictives qui peuvent être améliorées. Après avoir présenté de façon générale les schémas volumes finis en deux dimensions d'espace, on restreint notre étude à une famille de schémas particulièrement agréables à manipuler : les schémas de type Godunov 2D. Dans ce cadre, on établit une condition CFL optimale permettant d'assurer la robustesse des schémas 2D d'ordre un. On déduit alors par des techniques classiques que le schéma MUSCL d'ordre élevé est robuste sous une nouvelle condition CFL, qui, elle, n'est pas optimale, mais est moins restrictives que celles intervenant dans [87, 88, 14].

On présente ensuite la technique DMGR qui est inspirée des méthodes Discrete Duality Finite Volume (DDFV), utilisées dans le cadre des problèmes elliptiques et paraboliques (voir par exemple [65, 66, 49, 3]). L'idée de base consiste à écrire deux schémas MUSCL distincts sur deux maillages de $\mathbb{R}^2$ se recouvrant, un maillage primal et son maillage dual associé. On s'attend à ce que l'augmentation du nombre d'inconnues numériques qui en résulte permette de reconstruire des gradients très précis. Pour faciliter la compréhension, on décrit dans un premier temps la méthode DMGR en une dimension d'espace. On construit ensuite le maillage dual, avant de donner la procédure de reconstruction/limitation DMGR. À ce stade, l'intérêt de l'utilisation des deux maillages n'est pas flagrante. On précise donc comment les inconnues des deux schémas MUSCL nous permettent de reconstruire des états très précis aux sommets des cellules.

Dans la dernière partie, on montre plusieurs expériences numériques réalisées pour les équations d'Euler. Les résultats prouvent que le schéma DMGR est très précis et qu'il est capable de capturer des structures complexes très fines.

Troisième chapitre : Schémas MOOD entropiques d'ordre élevé pour les équations d'Euler

Le but de ce chapitre est de construire des schémas d'ordre élevé entropiques. Comme mentionné précédemment, plusieurs inégalités d'entropie discrètes ont été prouvées pour des sché-

mas d'ordre élevé (voir [21, 12]). Malheureusement, on ne sait pas si ces inégalités d'entropie sont suffisantes pour assurer que la solution convergée, au sens du théorème de Lax-Wendroff, soit entropique. Par conséquent, la notion de schéma d'ordre élevé entropique est ambigüe. La première partie est donc dédiée à l'étude du comportement des inégalités d'entropie discrètes dans un régime de convergence. On commence par rappeler le théorème de Lax-Wendroff pour des schémas d'ordre élevé en temps et en espace. On en profite pour exhiber des inégalités d'entropie discrètes « fortes » qui assurent que la solution convergée est entropique. On analyse ensuite le comportement des inégalités d'entropie discrètes « faibles » vérifiées par les schémas d'ordre élevé en espace et en temps. On montre que celles-ci coïncident dans le régime de convergence avec les inégalités fortes à une mesure positive près. Une conjecture énoncée dans [67], ainsi que de nombreuses expériences numériques, nous incitent à penser que cette mesure est concentrée (et strictement positive) sur les courbes de discontinuité de la solution convergée. Il en résulte que les inégalités d'entropie discrètes « faibles » ne sont pas l'outil approprié pour assurer le caractère entropique de la solution convergée.

Puisqu'il ne semble pas possible dans l'état actuel d'assurer la stabilité des schémas MUSCL classiques, on propose de rajouter une procédure de limitation *a posteriori* de type MOOD. Le principe est d'activer cette limitation uniquement sur les cellules qui ne vérifient pas les inégalités d'entropie discrètes « fortes ». Dans le cas des équations d'Euler, il y a une infinité d'entropies et bien entendu, il n'est pas question de toutes les tester. On montre alors qu'il est possible, pour certains schémas, d'obtenir toutes les inégalités d'entropie discrètes requises à partir d'une seule inégalité d'entropie discrète bien choisie.

En utilisant ce résultat essentiel, on peut maintenant donner l'algorithme de la procédure e-MOOD qui se base sur la technique MOOD, en utilisant l'inégalité d'entropie discrète particulière précédente comme critère d'activation de la limitation *a posteriori*. Plutôt qu'un schéma à part entière, la procédure e-MOOD est plutôt une méthode de stabilisation des schémas d'ordre élevé classiques. Pour conclure, la précision et la stabilité de cette procédure sont validées par plusieurs tests numériques.

Quatrième chapitre : Schémas well-balanced pour des systèmes de lois de conservation avec terme source

On s'intéresse ici à la dérivation de schémas well-balanced pour plusieurs systèmes avec terme source : les équations de Saint-Venant avec topographie, les équations de Ripa avec topographie et les équations d'Euler avec gravité. On commence par décrire ces systèmes en détails. On s'attache en particulier à décrire les solution stationnaires qui sont en général régies par une équation différentielle. Afin de définir de façon claire le caractère well-balanced d'un schéma, on choisit dans un premier temps de donner une interprétation locale des solutions stationnaires. On introduit ainsi, pour chaque système, une notion d'équilibre local entre deux états qui correspond à une discrétisation à l'ordre un de l'équation différentielle décrivant les états stationnaires.

Dans la deuxième partie, on présente de manière générale les schémas volumes finis en présence d'un terme source. On introduit ensuite les schémas de type Godunov dans ce cadre. Puis on définit une notion globale d'équilibre discret : une approximation discrète de la solution est une solution discrète stationnaire constante par morceaux si tous les couples d'états consécutifs vérifient l'équilibre local défini précédemment. La définition de schéma well-balanced suit alors naturellement : un schéma est dit well-balanced s'il préserve exactement toutes les solutions discrètes stationnaires constantes par morceaux.

On cherche ensuite à construire des schémas de type Godunov well-balanced en suivant les idées de [64, 52, 53, 30]. On considère des solveurs de Riemann approchés simples, c'est-à-dire constitués de N états intermédiaires constants. On commence par dériver un solveur de

Riemann approché pour le système de Ripa en faisant intervenir une linéarisation de l'équation décrivant l'équilibre local. Le premier schéma vérifie les propriétés requises, mais l'approche s'avère difficile à étendre aux équations d'Euler avec gravité. On introduit alors une modification du solveur pour Ripa en ne faisant pas intervenir la linéarisation de l'équilibre local. Cette modification rend possible l'extension aux équations d'Euler avec gravité. Les schémas construits par cette méthode sont well-balanced et préservent l'ensemble Ω. Cependant, la robustesse est obtenue au prix de vitesses d'ondes potentiellement très grandes, ce qui peut rendre le schéma très diffusif.

On développe une autre approche basée sur les méthodes de relaxation. Dans un premier temps, on présente le formalisme de ces méthodes, en suivant les travaux de [31, 16]. On se concentre ensuite sur le système de Saint-Venant en commençant par rappeler le modèle de relaxation de Suliciu (voir [20]). Ce modèle présente des difficultés liées à la résolution du problème de Riemann : d'une part, l'ordre des valeurs propres n'est pas fixé *a priori* et d'autre part, l'onde stationnaire liée au terme source contient de fortes non-linéarités. Pour contourner ce problème, on suggère d'introduire une modification de ce système en « transportant » artificiellement le terme source à la vitesse du fluide. Cette modification fait disparaitre les non linéarités précédentes mais rend le problème de Riemann sous-déterminé. Afin de fermer le système d'équations régissant le problème de Riemann, on rajoute une linéarisation de l'équation décrivant l'équilibre local. On justifie ce choix *a priori* arbitraire en écrivant la solution du « problème de Riemann » obtenue comme la solution d'un nouveau système de relaxation complètement déterminé. On montre finalement que le schéma obtenu est well-balanced et robuste. Cette approche est ensuite étendue au système de Ripa et aux équations d'Euler avec gravité.

On propose ensuite une extension MUSCL à l'ordre deux du schéma de relaxation développé pour les équations d'Euler avec gravité. Il est nécessaire, dans un premier temps, de modifier la définition de schéma well-balanced pour l'adapter aux schémas d'ordre deux. Pour cela, on introduit une notion de solution discrète stationnaire affine par morceaux qui prend en compte les pentes de la reconstruction. Le schéma MUSCL est alors dit well-balanced s'il préserve exactement toutes les solutions discrètes stationnaires affines par morceaux. Le schéma MUSCL est ensuite dérivé en suivant des techniques classiques. On prouve aisément qu'il est well-balanced et robuste. Pour conclure cette section, on donne quelques arguments expliquant pourquoi cette extension à l'ordre deux est beaucoup plus délicate à réaliser pour le système de Ripa.

On conclut ce chapitre en présentant plusieurs résultats numériques pour le système de Ripa et les équations d'Euler avec gravité.

Liste de publications

Plusieurs publications sont issues de ce travail : un article accepté pour publication dans une revue internationale :

- C. Berthon, Y. Coudière, V. Desveaux, *Second-order MUSCL schemes based on dual mesh gradient reconstruction (DMGR)*, accepté pour publication dans Math. Model. Numer. Anal. (disponible en ligne),

un article soumis :

- C. Berthon, V. Desveaux, *An entropic MOOD scheme for the Euler equations*

et un proceeding pour une conférence avec comité de lecture :

- C. Berthon, Y. Coudière, V. Desveaux, *Development of DDFV methods for the Euler equations*, Finite Volume for Complex Application VI, Springer Proceedings in Mathematics 4 (2011), pp. 117-124.

1

Généralités sur les méthodes de volumes finis

Ce chapitre présente un état de l'art des méthodes numériques de type volumes finis pour approcher les solutions des systèmes de lois de conservation. Les méthodes et les résultats présentés ici sont tous bien connus. On choisit de les rappeler brièvement par souci de complétude.

On considère un système de lois de conservation

$$\partial_t w + \partial_x f(w) = 0, \quad x \in \mathbb{R}, \quad t \in \mathbb{R}^+, \tag{1.1}$$

où $w : \mathbb{R} \times \mathbb{R}^+ \to \Omega$ est le vecteur des inconnues, $f : \Omega \to \mathbb{R}^d$ est une fonction flux supposée régulière et $\Omega \subset \mathbb{R}^d$ désigne un ensemble d'états physiquement admissibles, supposé convexe. Ce système est complété par une condition initiale

$$w(x, t = 0) = w_0(x), \quad x \in \mathbb{R}. \tag{1.2}$$

On suppose que le système (1.1) est hyperbolique, c'est-à-dire que la matrice jacobienne $\nabla_w f(w)$ est diagonalisable dans $\mathbb{R}$, pour tout vecteur $w \in \Omega$.

Le système (1.1) traduit la conservation des différentes quantités composant le vecteur w. En effet, en intégrant l'équation (1.1) sur le segment $[a, b]$ en espace, on obtient

$$\frac{d}{dt} \int_a^b w(x, t) dx = f(w(a, t)) - f(w(b, t)).$$

Autrement dit, la variation de la quantité $\int_a^b w(x, t) dx$ est égale au flux entrant au point a moins le flux sortant au point b.

Il est bien connu que les solutions des systèmes hyperboliques peuvent développer des discontinuités, même pour une donnée initiale très régulière (voir par exemple [56, 80]). Par conséquent, il est nécessaire de considérer des solutions faibles, c'est-à-dire des fonctions vérifiant (1.1)–(1.2) au sens des distributions. Les discontinuités pouvant apparaître dans les solutions faibles sont caractérisées par les relations de Rankine-Hugoniot :

$$f(w_R) - f(w_L) = \sigma(w_R - w_L),$$

où w_L et w_R sont les valeurs de l'inconnue de part et d'autre de la discontinuité et σ est la pente de la discontinuité dans le plan (t, x).

Les solutions faibles ne sont pas toujours physiquement admissibles et l'on a besoin d'un critère supplémentaire permettant d'écarter les solutions non physiques. C'est le rôle de l'entropie que nous introduisons maintenant. Une entropie pour le système (1.1) est une fonction convexe $\eta \in C^2(\Omega; \mathbb{R})$ telle qu'il existe une fonction $\mathcal{G} \in C^2(\Omega; \mathbb{R})$, appelée flux d'entropie, et vérifiant

$$\nabla_w f \nabla_w \eta = \nabla_w \mathcal{G}.$$

Une solution faible w est alors dite entropique si elle vérifie au sens des distributions l'inégalité d'entropie

$$\partial_t \eta(w) + \partial_x \mathcal{G}(w) \leq 0, \tag{1.3}$$

pour toute paire d'entropie $(\eta, \mathcal{G})$.

1.1 Présentation des méthodes de volumes finis

On s'intéresse ici à l'approximation numérique des solutions faibles du système (1.1). Pour cela, on discrétise l'espace $\mathbb{R}$ en une suite croissante de points $(x_{i+1/2})_{i\in\mathbb{Z}}$. Pour simplifier la présentation, on suppose que cette discrétisation est uniforme, c'est-à-dire $x_{i+1/2} - x_{i-1/2} = \Delta x$, où Δx est le pas d'espace supposé constant. On définit alors les volumes de contrôle $K_i = [x_{i-1/2}, x_{i+1/2}[$ et l'on note $x_i = (x_{i-1/2} + x_{i+1/2})/2$ le milieu de la cellule K_i. On discrétise également le temps de la façon suivante, $t^n = n\Delta t$, où Δt est le pas de temps. On note alors w_i^n une approximation de la solution exacte de (1.1)–(1.2) sur la cellule K_i et au temps t^n.

Dans les méthodes de volumes finis, on cherche à approcher la moyenne de la solution w de (1.1) sur chaque cellule K_i :

$$w_i^n \approx \frac{1}{\Delta x} \int_{K_i} w(x, t^n) dx.$$

En intégrant l'équation (1.1) sur le rectangle $K_i \times [t^n, t^{n+1}]$ et en divisant par Δx, on obtient

$$\frac{1}{\Delta x} \int_{K_i} w(x, t^{n+1}) dx = \frac{1}{\Delta x} \int_{K_i} w(x, t^n) dx - \frac{1}{\Delta x} \left(\int_{t^n}^{t^{n+1}} f(w(x_{i+1/2}, t)) dt - \int_{t^n}^{t^{n+1}} f(w(x_{i-1/2}, t) dt \right). \tag{1.4}$$

Cette relation nous donne la mise à jour exacte des moyennes au pas de temps suivant, mais les intégrales en temps des flux sont en général difficiles à évaluer exactement. L'équation (1.4) nous suggère de considérer des méthodes numériques de la forme

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x} \left(F_{i+1/2} - F_{i-1/2} \right),$$

où $F_{i+1/2}$ est une approximation de la moyenne du flux le long de l'interface $x = x_{i+1/2}$:

$$F_{i+1/2} \approx \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} f(w(x_{i+1/2}, t)) dt.$$

L'information se propageant à vitesse finie dans un système hyperbolique, il est raisonnable de considérer que l'approximation $F_{i+1/2}$ peut être obtenue à partir des deux valeurs w_i^n et w_{i+1}^n de l'inconnue de part et d'autre de l'interface, c'est-à-dire

$$F_{i+1/2} = F\left(w_i^n, w_{i+1}^n\right),$$

où la fonction $F : \Omega \times \Omega \to \mathbb{R}^d$ est appelée flux numérique. Cela nous amène à la forme générale des schémas volumes finis à trois points :

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x} \left(F\left(w_i^n, w_{i+1}^n\right) - F\left(w_{i-1}^n, w_i^n\right) \right). \tag{1.5}$$

Une méthode de volumes finis est donc entièrement déterminée par le choix d'un flux numérique.

Pour discrétiser la donnée initiale, on adopte une approximation constante par morceaux au sens L^2. Par conséquent, sur chaque cellule K_i, la donnée initiale est approchée par

$$w_i^0 = \frac{1}{\Delta x} \int_{K_i} w_0(x) dx.$$

Un schéma écrit sous la forme (1.5) est dit sous forme conservative. En effet, on voit facilement que pour toute suite $(w_i^n)_{i\in\mathbb{Z}} \in L^1(\mathbb{Z})$ telle que $(w_i^{n+1})_{i\in\mathbb{Z}} \in L^1(\mathbb{Z})$, on a

$$\sum_{i\in\mathbb{Z}} w_i^{n+1} = \sum_{i\in\mathbb{Z}} w_i^n$$

et ainsi la masse totale de w est conservée par le schéma.

Une condition fondamentale que doit vérifier le schéma numérique est la consistance. Le flux numérique F (et donc le schéma (1.5)) est consistant si et seulement si

$$\forall w \in \Omega, \quad F(w, w) = w.$$

Une autre propriété importante, aussi bien sur le plan numérique que sur le plan physique, est la robustesse. Un schéma est dit robuste (ou préservant Ω) si

$$\forall i \in \mathbb{Z}, \quad w_i^n \in \Omega \quad \Rightarrow \quad \forall i \in \mathbb{Z}, \quad w_i^{n+1} \in \Omega.$$

Enfin, on demande que le schéma vérifie un équivalent discret des inégalités d'entropie (1.3). Le schéma (1.5) est dit entropique, si pour toute paire d'entropie $(\eta, \mathcal{G})$, il existe un flux numérique d'entropie $G : \Omega \times \Omega \to \mathbb{R}$, consistant avec $\mathcal{G}$ (i.e. $G(w, w) = \mathcal{G}(w)$), tel que le schéma vérifie l'inégalité d'entropie discrète

$$\eta(w_i^{n+1}) \leq \eta(w_i^n) - \frac{\Delta t}{\Delta x} \left(G\left(w_i^n, w_{i+1}^n\right) - G\left(w_{i-1}^n, w_i^n\right) \right). \tag{1.6}$$

1.2 Schéma de Godunov

Le schéma de Godunov [57] est le schéma volumes finis le plus naturel. Il se base sur la résolution exacte du problème de Riemann qui est un problème de Cauchy dans lequel la donnée initiale est uniquement constituée de deux états constants séparés par une discontinuité :

$$\begin{cases} \partial_t w + \partial_x f(w) = 0, \\ w(x,0) = \begin{cases} w_L, & \text{si } x < 0, \\ w_R, & \text{si } x > 0. \end{cases} \end{cases} \tag{1.7}$$

On suppose connue la solution faible entropique auto-similaire de ce problème et on la note $W_{\mathcal{R}}\left(\frac{x}{t}, w_L, w_R\right)$. Il est bien connu que dans un système hyperbolique, les informations se propagent à vitesse finie. On note ainsi respectivement $\lambda^-(w_L, w_R)$ et $\lambda^+(w_L, w_R)$ la plus petite et la plus grande vitesse d'onde développée par le problème de Riemann $W_{\mathcal{R}}\left(\frac{x}{t}, w_L, w_R\right)$. Avant d'introduire le schéma de Godunov, on présente le résultat suivant qui sera utile dans la suite.

Lemme 1.1. *Supposons que Δt et Δx vérifient*

$$\frac{\Delta t}{\Delta x}\max|\lambda^{\pm}(w_L,w_R)|\leq\frac{1}{2}. \tag{1.8}$$

Alors la moyenne de la solution exacte du problème de Riemann (1.7) est donnée par

$$\frac{1}{\Delta x}\int_{-\Delta x/2}^{\Delta x/2}W_{\mathcal{R}}\left(\frac{x}{\Delta t},w_L,w_R\right)dx=\frac{w_L+w_R}{2}-\frac{\Delta t}{\Delta x}\left(f(w_R)-f(w_L)\right). \tag{1.9}$$

Démonstration. On intègre l'équation (1.7) sur le rectangle $[-\Delta x/2,\Delta x/2]\times[0,\Delta t]$ pour obtenir

$$\int_{-\Delta x/2}^{\Delta x/2}W_{\mathcal{R}}\left(\frac{x}{\Delta t},w_L,w_R\right)dx-\int_{-\Delta x/2}^{\Delta x/2}w(x,0)dx+\int_0^{\Delta t}f\left(W_{\mathcal{R}}\left(\frac{\Delta x}{2t},w_L,w_R\right)\right)dt$$
$$-\int_0^{\Delta t}f\left(W_{\mathcal{R}}\left(-\frac{\Delta x}{2t},w_L,w_R\right)\right)dt=0.$$

La condition (1.8) implique

$$W_{\mathcal{R}}\left(-\frac{\Delta x}{2t},w_L,w_R\right)=w_L\quad\text{et}\quad W_{\mathcal{R}}\left(\frac{\Delta x}{2t},w_L,w_R\right)=w_R,\quad\forall t\in[0,\Delta t].$$

On en déduit immédiatement l'équation (1.9). □

On considère maintenant une approximation de la solution à la date t^n, constante par morceaux,

$$W_{\Delta x}^n(x)=w_i^n,\quad\text{si } x\in K_i.$$

Remarquons qu'à chaque interface $x_{i+1/2}$, on a localement un problème de Riemann. On sait donc résoudre exactement le problème de Cauchy

$$\begin{cases}\partial_t w+\partial_x f(w)=0,\\ w(x,t^n)=W_{\Delta x}^n(x),\end{cases} \tag{1.10}$$

au moins pour des temps t^n+t, avec t petit. Plus précisément, la solution exacte de ce problème de Cauchy est constituée de la juxtaposition des problèmes de Riemann locaux $W_{\mathcal{R}}\left(\frac{x-x_{i+1/2}}{t},w_i^n,w_{i+1}^n\right)$, tant que ceux-ci n'interagissent pas. Une condition suffisante pour que les problèmes de Riemann locaux n'interagissent pas est la condition de Courant-Friedrichs-Lewy (CFL) introduite dans [42] :

$$\frac{\Delta t}{\Delta x}\max_{i\in\mathbb{Z}}\left|\lambda^{\pm}\left(w_i^n,w_{i+1}^n\right)\right|\leq\frac{1}{2}. \tag{1.11}$$

On note

$$W_{\Delta x}(x,t^n+t)=W_{\mathcal{R}}\left(\frac{x-x_{i+1/2}}{t},w_i^n,w_{i+1}^n\right),\quad\text{si } x\in[x_i,x_{i+1}[,$$

la solution du problème de Cauchy (1.10). Au temps $t^{n+1}=t^n+\Delta t$, la solution $W_{\Delta x}$ n'est pas constante sur chaque cellule K_i. On effectue donc une projection L^2 sur l'espace des fonctions constantes sur chaque cellule K_i pour définir

$$w_i^{n+1}=\frac{1}{\Delta x}\int_{K_i}W_{\Delta x}\left(x,t^n+\Delta t\right)dx. \tag{1.12}$$

On obtient ainsi une approximation de la solution au temps t^{n+1}, constante sur chaque cellule K_i, définie par

$$W_{\Delta x}^{n+1}(x)=w_i^{n+1},\quad\text{si } x\in K_i.$$

Pour résumer, le schéma de Godunov est constitué de deux étapes : une étape d'évolution exacte en temps, suivie d'une étape de projection en espace. On montre maintenant que le schéma de Godunov s'écrit sous forme conservative.

Lemme 1.2. *Supposons que la condition CFL (1.11) est vérifiée. Alors le schéma de Godunov (1.12) se réécrit sous forme d'un schéma conservatif (1.5), avec pour flux numérique*

$$F(w_L, w_R) = f\left(W_{\mathcal{R}}\left(0, w_L, w_R\right)\right). \tag{1.13}$$

De plus, le flux numérique F est consistant avec f.

Démonstration. Par définition, on a

$$\begin{aligned} w_i^{n+1} &= \frac{1}{\Delta x}\int_{K_i} W_{\Delta x}\left(x, t^n + \Delta t\right) dx \\ &= \frac{1}{\Delta x}\int_0^{\Delta x/2} W_{\mathcal{R}}\left(\frac{x}{\Delta t}, w_{i-1}^n, w_i^n\right) dx + \frac{1}{\Delta x}\int_{-\Delta x/2}^0 W_{\mathcal{R}}\left(\frac{x}{\Delta t}, w_i^n, w_{i+1}^n\right) dx. \end{aligned}$$

En intégrant l'équation (1.7) sur le rectangle $[0, \Delta x/2] \times [0, \Delta t]$, on obtient

$$\frac{1}{\Delta x}\int_0^{\Delta x/2} W_{\mathcal{R}}\left(\frac{x}{\Delta t}, w_L, w_R\right) dx = \frac{1}{2} w_R - \frac{\Delta t}{\Delta x}\left(f(w_R) - f\left(W_{\mathcal{R}}\left(0, w_L, w_R\right)\right)\right),$$

et en intégrant (1.7) sur le rectangle $[-\Delta x/2, 0] \times [0, \Delta t]$, on obtient

$$\frac{1}{\Delta x}\int_{-\Delta x/2}^0 W_{\mathcal{R}}\left(\frac{x}{\Delta t}, w_L, w_R\right) dx = \frac{1}{2} w_L - \frac{\Delta t}{\Delta x}\left(f\left(W_{\mathcal{R}}\left(0, w_L, w_R\right)\right) - f(W_L)\right).$$

On en déduit

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(f\left(W_{\mathcal{R}}\left(0, w_i^n, w_{i+1}^n\right)\right) - f\left(W_{\mathcal{R}}\left(0, w_{i-1}^n, w_i^n\right)\right)\right).$$

On a ainsi écrit le schéma de Godunov sous forme conservative, où le flux numérique F est défini par (1.13). Le flux numérique F est consistant avec f, puisque

$$\begin{aligned} F(w, w) &= f\left(W_{\mathcal{R}}\left(0, w, w\right)\right) \\ &= f(w). \end{aligned}$$

□

On montre également que le schéma de Godunov est entropique.

Lemme 1.3. *Supposons que la condition CFL (1.11) est vérifiée. Alors le schéma de Godunov (1.12) est entropique, c'est-à-dire que pour tout couple d'entropie $(\eta, \mathcal{G})$, le schéma vérifie l'inégalité d'entropie discrète (1.6), où le flux d'entropie G est défini par*

$$G(w_L, w_R) = \mathcal{G}\left(W_{\mathcal{R}}\left(0, w_L, w_R\right)\right).$$

Démonstration. Puisque η est une fonction convexe, l'inégalité de Jensen nous donne

$$\begin{aligned} \eta\left(w_i^{n+1}\right) &\leq \frac{1}{\Delta x}\int_{K_i} \eta\left(W_{\Delta x}\left(x, t^n + \Delta t\right)\right) dx \\ &\leq \frac{1}{\Delta x}\int_0^{\Delta x/2} \eta\left(W_{\mathcal{R}}\left(\frac{x}{\Delta t}, w_{i-1}^n, w_i^n\right)\right) dx + \frac{1}{\Delta x}\int_{-\Delta x/2}^0 \eta\left(W_{\mathcal{R}}\left(\frac{x}{\Delta t}, w_i^n, w_{i+1}^n\right)\right) dx. \end{aligned}$$

En intégrant l'inégalité d'entropie (1.3) sur le rectangle $[0, \Delta x/2] \times [0, \Delta t]$, on obtient

$$\frac{1}{\Delta x}\int_0^{\Delta x/2} \eta\left(W_{\mathcal{R}}\left(\frac{x}{\Delta t}, w_L, w_R\right)\right) dx \leq \frac{1}{2}\eta(w_R) - \frac{\Delta t}{\Delta x}\left(\mathcal{G}(w_R) - \mathcal{G}\left(W_{\mathcal{R}}\left(0, w_L, w_R\right)\right)\right),$$

et en intégrant (1.3) sur le rectangle $[-\Delta x/2, 0] \times [0, \Delta t]$, on obtient

$$\frac{1}{\Delta x}\int_{-\Delta x/2}^{0} \eta\left(W_{\mathcal{R}}\left(\frac{x}{\Delta t}, w_L, w_R\right)\right) dx \leq \frac{1}{2}\eta(w_L) - \frac{\Delta t}{\Delta x}\left(\mathcal{G}\left(W_{\mathcal{R}}\left(0, w_L, w_R\right)\right) - \mathcal{G}(W_L)\right).$$

On en déduit l'inégalité d'entropie discrète

$$\eta\left(w_i^{n+1}\right) \leq \eta\left(w_i^n\right) - \frac{\Delta t}{\Delta x}\left(\mathcal{G}\left(W_{\mathcal{R}}\left(0, w_i^n, w_{i+1}^n\right)\right) - \mathcal{G}\left(W_{\mathcal{R}}\left(0, w_{i-1}^n, w_i^n\right)\right)\right).$$

Il reste à montrer que le flux numérique d'entropie

$$G(w_L, w_R) = \mathcal{G}\left(W_{\mathcal{R}}\left(0, w_L, w_R\right)\right)$$

est consistant avec $\mathcal{G}$, ce qui est le cas car $W_{\mathcal{R}}(0, w, w) = w$. □

Bien que précis car basé sur la résolution exacte des problèmes de Riemann, le schéma de Godunov présente plusieurs défauts et souvent d'autres schémas lui sont préférés. En effet, la résolution exacte des problèmes de Riemann peut s'avérer pour certains systèmes très difficile, voire impossible, et souvent très coûteuse en temps de calcul. De plus, la précision obtenue par l'étape d'évolution exacte est gommée par l'étape de projection. Une alternative est apparue dans les années 80 avec les travaux de Roe [91] et de Harten, Lax et van Leer [64], consistant à utiliser une solution approchée du problème de Riemann plutôt que la solution exacte. Cette approche mène aux schémas de type Godunov que nous présentons maintenant.

1.3 Schémas de type Godunov

Le formalisme des schémas de type Godunov a été introduit par Harten, Lax et van Leer [64]. L'idée de base consiste à utiliser une approximation $\widetilde{W}\left(\frac{x}{t}, w_L, w_R\right)$ de la solution du problème de Riemann (1.7). Pour ne pas perdre d'information, on demande que le cône de dépendance de la solution exacte soit inclus dans le cône de dépendance de la solution approchée (voir Figure 1.1). On demande de plus que la moyenne de la solution exacte sur une cellule soit préservée par le solveur approché.

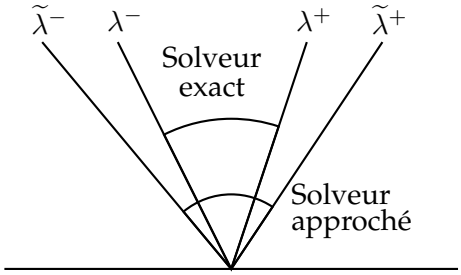


FIGURE 1.1 – Solveurs de Riemann exact et approché

Cela nous conduit à la définition suivante (voir [64]) :

Définition 1.4. On appelle solveur de Riemann approché une fonction $\widetilde{W} : \mathbb{R} \times \Omega \times \Omega \to \mathbb{R}^d$ telle que :

(i) il existe des vitesses $\widetilde{\lambda}^- \leq \lambda^-$ et $\widetilde{\lambda}^+ \geq \lambda^+$ telles que

$$\widetilde{W}(\xi, w_L, w_R) = \begin{cases} w_L, & \text{si } \xi < \widetilde{\lambda}^-, \\ \widetilde{W}(\xi, w_L, w_R), & \text{si } \widetilde{\lambda}^- < \xi < \widetilde{\lambda}^+, \\ w_R, & \text{si } \xi > \widetilde{\lambda}^+, \end{cases}$$

où $\lambda^\pm$ désignent la plus petite et la plus grande vitesse d'onde apparaissant dans la solution exacte du problème de Riemann (1.7) ;

(ii) si la condition CFL

$$\frac{\Delta t}{\Delta x} \max |\widetilde{\lambda}^\pm(w_L, w_R)| \leq \frac{1}{2} \tag{1.14}$$

est vérifiée, alors le solveur approché préserve la moyenne du solveur exact :

$$\frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_L, w_R\right) dx = \frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} W_{\mathcal{R}}\left(\frac{x}{\Delta t}, w_L, w_R\right) dx. \tag{1.15}$$

(iii) le solveur approché vérifie $\widetilde{W}(\xi, w, w) = w$ pour tout $\xi \in \mathbb{R}$ et pour tout $w \in \Omega$.

D'après le Lemme 1.1, on peut calculer la moyenne sur une cellule de la solution exacte du problème de Riemann. La propriété (1.15) se réécrit alors de la façon suivante :

Lemme 1.5. *Supposons que la condition CFL (1.14) est vérifiée. Alors l'équation (1.15) est équivalente à la consistance avec la forme intégrale :*

$$\frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_L, w_R\right) dx = \frac{1}{2}(w_L + w_R) - \frac{\Delta t}{\Delta x}\left(f(w_R) - f(w_L)\right). \tag{1.16}$$

Démonstration. Notons que la propriété (i) de la Définition 1.4 implique

$$\max |\lambda^\pm| \leq \max |\widetilde{\lambda}^\pm|,$$

donc la condition CFL (1.14) implique (1.8). La moyenne sur une cellule de la solution exacte du problème de Riemann est alors donnée par (1.9), ce qui permet de conclure. □

De la même façon que pour le schéma de Godunov, on peut construire un schéma numérique à partir de n'importe quel solveur de Riemann approché. Supposons que l'on connaît à la date t^n une approximation constante par morceaux de la solution,

$$W_{\Delta x}^n(x) = w_i^n, \quad \text{si } x \in K_i.$$

On note

$$W_{\Delta x}(x, t^n + t) = \widetilde{W}\left(\frac{x - x_{i+1/2}}{t}, w_i^n, w_{i+1}^n\right), \quad \text{si } x \in [x_i, x_{i+1}[,$$

la juxtaposition des solutions approchés des problèmes de Riemann. Pour éviter toute interaction entre les solveurs approchés, on impose la condition CFL

$$\frac{\Delta t}{\Delta x} \max_{i \in \mathbb{Z}} |\widetilde{\lambda}^\pm\left(w_i^n, w_{i+1}^n\right)| \leq \frac{1}{2}. \tag{1.17}$$

On définit alors la solution approchée au temps t^{n+1} en projetant $W_{\Delta x}$ sur l'espace des fonctions constantes sur chaque K_i :

$$w_i^{n+1} = \frac{1}{\Delta x} \int_{K_i} W_{\Delta x}\left(x, t^n + \Delta t\right) dx,$$

que l'on peut réécrire

$$w_i^{n+1} = \frac{1}{\Delta x}\int_0^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_{i-1}^n, w_i^n\right) dx + \frac{1}{\Delta x}\int_{-\Delta x/2}^0 \widetilde{W}\left(\frac{x}{\Delta t}, w_i^n, w_{i+1}^n\right) dx. \tag{1.18}$$

On dit que le schéma défini par (1.18) est un schéma de type Godunov associé au solveur de Riemann approché $\widetilde{W}$.

Remarquons que la solution exacte d'un problème de Riemann étant un choix particulier de solveur de Riemann approché, le schéma de Godunov est un schéma de type Godunov.

Il est important de vérifier qu'un schéma de type Godunov peut s'écrire sous forme conservative.

Proposition 1.6. *Sous la condition CFL (1.17), un schéma de type Godunov peut s'écrire sous forme conservative*

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(F\left(w_i^n, w_{i+1}^n\right) - F\left(w_{i-1}^n, w_i^n\right)\right), \tag{1.19}$$

où le flux numérique est défini par

$$F(w_L, w_R) = f(w_L) + \frac{\Delta x}{2\Delta t} w_L - \frac{1}{\Delta t}\int_{-\Delta x/2}^0 \widetilde{W}\left(\frac{x}{\Delta t}, w_L, w_R\right) dx, \tag{1.20}$$

ou de façon équivalente par

$$F(w_L, w_R) = f(w_R) - \frac{\Delta x}{2\Delta t} w_R + \frac{1}{\Delta t}\int_0^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_L, w_R\right) dx. \tag{1.21}$$

De plus le flux numérique F est consistant avec f.

Démonstration. Par définition, on a

$$\begin{aligned} w_i^{n+1} &= \frac{1}{\Delta x}\int_0^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_{i-1}^n, w_i^n\right) dx + \frac{1}{\Delta x}\int_{-\Delta x/2}^0 \widetilde{W}\left(\frac{x}{\Delta t}, w_i^n, w_{i+1}^n\right) dx \\ &= \frac{1}{\Delta x}\int_{-\Delta x/2}^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_{i-1}^n, w_i^n\right) dx - \frac{1}{\Delta x}\int_{-\Delta x/2}^0 \widetilde{W}\left(\frac{x}{\Delta t}, w_{i-1}^n, w_i^n\right) dx \\ &\quad + \frac{1}{\Delta x}\int_{-\Delta x/2}^0 \widetilde{W}\left(\frac{x}{\Delta t}, w_i^n, w_{i+1}^n\right) dx. \end{aligned}$$

La condition CFL (1.17) étant vérifiée, on peut utiliser la consistance avec la forme intégrale (1.16) et l'on trouve

$$\begin{aligned} w_i^{n+1} &= \frac{1}{2}(w_{i-1}^n + w_i^n) - \frac{\Delta t}{\Delta x}\left(f(w_i^n) - f(w_{i-1}^n)\right) - \frac{1}{\Delta x}\int_{-\Delta x/2}^0 \widetilde{W}\left(\frac{x}{\Delta t}, w_{i-1}^n, w_i^n\right) dx \\ &\quad + \frac{1}{\Delta x}\int_{-\Delta x/2}^0 \widetilde{W}\left(\frac{x}{\Delta t}, w_i^n, w_{i+1}^n\right) dx \\ &= w_i^n - \frac{\Delta t}{\Delta x}\left(f(w_i^n) + \frac{\Delta x}{2\Delta t} w_i^n - \frac{1}{\Delta t}\int_{-\Delta x/2}^0 \widetilde{W}\left(\frac{x}{\Delta t}, w_i^n, w_{i+1}^n\right) dx\right. \\ &\quad \left. - \left(f(w_{i-1}^n) + \frac{\Delta x}{2\Delta t} w_{i-1}^n - \frac{1}{\Delta t}\int_{-\Delta x/2}^0 \widetilde{W}\left(\frac{x}{\Delta t}, w_{i-1}^n, w_i^n\right) dx\right)\right). \end{aligned}$$

Le schéma s'écrit donc bien sous forme conservative (1.19) avec le flux numérique défini par (1.20). Pour obtenir la formulation équivalente (1.21) du flux numérique, on utilise à nouveau la consistance avec la forme intégrale (1.16)

$$\begin{aligned}
F(w_L, w_R) &= f(w_L) + \frac{\Delta x}{2\Delta t} w_L - \frac{1}{\Delta t} \int_{-\Delta x/2}^{0} \widetilde{W}\left(\frac{x}{\Delta t}, w_L, w_R\right) dx \\
&= f(w_L) + \frac{\Delta x}{2\Delta t} w_L - \frac{1}{\Delta t} \int_{-\Delta x/2}^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_L, w_R\right) dx \\
&\quad + \frac{1}{\Delta t} \int_{0}^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_L, w_R\right) dx \\
&= f(w_L) + \frac{\Delta x}{2\Delta t} w_L - \frac{\Delta x}{2\Delta t}(w_L + w_R) + (f(w_R) - f(w_L)) \\
&\quad + \frac{1}{\Delta t} \int_{0}^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_L, w_R\right) dx \\
&= f(w_R) - \frac{\Delta x}{2\Delta t} w_R + \frac{1}{\Delta t} \int_{0}^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_L, w_R\right) dx.
\end{aligned}$$

Enfin, en utilisant la propriété (iii) de la définition 1.4 dans l'expression (1.20) du flux numérique, on obtient $F(w, w) = f(w)$, ce qui prouve que le schéma est consistant. □

Remarque. Les formulations (1.20) et (1.21) du flux numérique créent une dissymétrie entre l'état gauche et l'état droit. On peut également expliciter le flux numérique sous une forme symétrique :

$$\begin{aligned}
F(w_L, w_R) = \frac{1}{2}\left(f(w_L) + f(w_R)\right) - \frac{\Delta x}{4\Delta t}\Bigg(& w_R - \frac{2}{\Delta x} \int_{0}^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_L, w_R\right) dx \\
& - w_L + \frac{2}{\Delta x} \int_{-\Delta x/2}^{0} \widetilde{W}\left(\frac{x}{\Delta t}, w_L, w_R\right) dx \Bigg). \quad (1.22)
\end{aligned}$$

Pour obtenir cette formulation, il suffit de faire la demi-somme entre (1.20) et (1.21).

Pour compléter cette présentation, il reste à exhiber des conditions pour qu'un schéma de type Godunov soit robuste et entropique. C'est l'objet des deux lemmes qui suivent.

Lemme 1.7. *Supposons que la condition CFL (1.17) est vérifiée. Si pour tous w_L et w_R dans Ω, le solveur de Riemann approché $\widetilde{W}(\xi, w_L, w_R)$ est à valeurs dans Ω, pour tout $\xi \in \mathbb{R}$, alors le schéma de type Godunov associé à $\widetilde{W}$ est robuste.*

Démonstration. Le schéma de type Godunov associé est défini par (1.18) qui est la moyenne d'une fonction à valeur dans Ω. L'ensemble Ω étant supposé convexe, w_i^{n+1} reste donc dans Ω. □

Lemme 1.8. *Supposons que la condition CFL (1.17) est vérifiée. Si pour tout couple d'entropie $(\eta, \mathcal{G})$ et pour tous w_L et w_R dans Ω, le solveur de Riemann approché $\widetilde{W}$ vérifie la consistance avec la forme intégrale de l'inégalité d'entropie*

$$\frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} \eta\left(\widetilde{W}\left(\frac{x}{t}, w_L, w_R\right)\right) dx \leq \frac{1}{2}\left(\eta(w_L) + \eta(w_R)\right) - \frac{\Delta t}{\Delta x}\left(\mathcal{G}(w_R) - \mathcal{G}(w_L)\right),$$

alors le schéma de type Godunov associé à $\widetilde{W}$ est entropique et vérifie l'inégalité (1.6).

Démonstration. L'entropie η étant convexe, on a par l'inégalité de Jensen

$$\begin{aligned}
\eta\left(w_i^{n+1}\right) &= \eta\left(\frac{1}{\Delta x}\int_{K_i} W_{\Delta x}(x,t^n+\Delta t)dx\right)\\
&\leq \frac{1}{\Delta x}\int_{K_i}\eta\left(W_{\Delta x}(x,t^n+\Delta t)\right)dx\\
&\leq \frac{1}{\Delta x}\int_0^{\Delta x/2}\eta\left(\widetilde{W}\left(\frac{x}{\Delta t},w_{i-1}^n,w_i^n\right)\right)dx + \frac{1}{\Delta x}\int_{-\Delta x/2}^0\eta\left(\widetilde{W}\left(\frac{x}{\Delta t},w_i^n,w_{i+1}^n\right)\right)dx\\
&\leq \frac{1}{\Delta x}\int_{-\Delta x/2}^{\Delta x/2}\eta\left(\widetilde{W}\left(\frac{x}{\Delta t},w_{i-1}^n,w_i^n\right)\right)dx - \frac{1}{\Delta x}\int_{-\Delta x/2}^0\eta\left(\widetilde{W}\left(\frac{x}{\Delta t},w_{i-1}^n,w_i^n\right)\right)dx\\
&\quad + \frac{1}{\Delta x}\int_{-\Delta x/2}^0\eta\left(\widetilde{W}\left(\frac{x}{\Delta t},w_i^n,w_{i+1}^n\right)\right)dx.
\end{aligned}$$

En utilisant l'hypothèse, on trouve

$$\begin{aligned}
\eta\left(w_i^{n+1}\right) &\leq \frac{1}{2}\left(\eta(w_{i-1}^n)+\eta(w_i^n)\right) - \frac{\Delta t}{\Delta x}\left(\mathcal{G}(w_i^n)-\mathcal{G}(w_{i-1}^n)\right)\\
&\quad - \frac{1}{\Delta x}\int_{-\Delta x/2}^0\eta\left(\widetilde{W}\left(\frac{x}{\Delta t},w_{i-1}^n,w_i^n\right)\right)dx + \frac{1}{\Delta x}\int_{-\Delta x/2}^0\eta\left(\widetilde{W}\left(\frac{x}{\Delta t},w_i^n,w_{i+1}^n\right)\right)dx\\
&\leq \eta\left(w_i^n\right) - \frac{\Delta t}{\Delta x}\left(\mathcal{G}(w_i^n) + \frac{\Delta x}{2\Delta t}\eta(w_i^n) - \frac{1}{\Delta t}\int_{-\Delta x/2}^0\eta\left(\widetilde{W}\left(\frac{x}{\Delta t},w_i^n,w_{i+1}^n\right)\right)dx\right.\\
&\quad \left. - \left(\mathcal{G}(w_{i-1}^n) + \frac{\Delta x}{2\Delta t}\eta(w_{i-1}^n) - \frac{1}{\Delta t}\int_{-\Delta x/2}^0\eta\left(\widetilde{W}\left(\frac{x}{\Delta t},w_{i-1}^n,w_i^n\right)\right)dx\right)\right).
\end{aligned}$$

Le schéma vérifie donc l'inégalité d'entropie discrète (1.6) attendue

$$\eta(w_i^{n+1}) \leq \eta(w_i^n) - \frac{\Delta t}{\Delta x}\left(G\left(w_i^n,w_{i+1}^n\right) - G\left(w_{i-1}^n,w_i^n\right)\right),$$

où le flux numérique d'entropie G est défini par

$$G(w_L,w_R) = \mathcal{G}(w_L) + \frac{\Delta x}{2\Delta t}\eta(w_L) - \frac{1}{\Delta t}\int_{-\Delta x/2}^0\eta\left(\widetilde{W}\left(\frac{x}{\Delta t},w_L,w_R\right)\right)dx.$$

Par ailleurs, G est consistant avec le flux d'entropie $\mathcal{G}$ grâce à la propriété (iii) de la définition 1.4. Le schéma est donc entropique. □

Remarque. En modifiant la démonstration, on aurait également pu aboutir à l'inégalité discrète d'entropie

$$\eta(w_i^{n+1}) \leq \eta(w_i^n) - \frac{\Delta t}{\Delta x}\left(\overline{G}(w_i^n,w_{i+1}^n) - \overline{G}(w_{i-1}^n,w_i^n)\right),$$

où le flux numérique d'entropie $\overline{G}$ est défini par

$$\overline{G}(w_L,w_R) = \mathcal{G}(w_R) - \frac{\Delta x}{\Delta t}\eta(w_R) + \frac{1}{\Delta t}\int_0^{\Delta x/2}\eta\left(\widetilde{W}\left(\frac{x}{\Delta t},w_L,w_R\right)\right)dx.$$

Les flux numériques d'entropie G et $\overline{G}$ sont en général différents. On constate qu'il n'y a donc pas une unique inégalité d'entropie discrète pour un couple d'entropie $(\eta,\mathcal{G})$ donné.

1.4 Schémas d'ordre élevé de type MUSCL

La méthode MUSCL a été introduite par van Leer [102] (voir aussi [87, 88, 21, 12]) dans le but de construire une approximation d'ordre élevé de la solution. La méthode MUSCL se décompose en deux étapes :

1. étape de reconstruction de l'inconnue aux interfaces (voir Figure 1.2) :

$$w_i^{n,\pm} = w_i^n \pm \Delta w_i^n;$$

2. étape d'évolution des états reconstruits en utilisant le flux numérique d'ordre un :

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(F\left(w_i^{n,+}, w_{i+1}^{n,-}\right) - F\left(w_{i-1}+, w_i^{n,-}\right)\right). \tag{1.23}$$

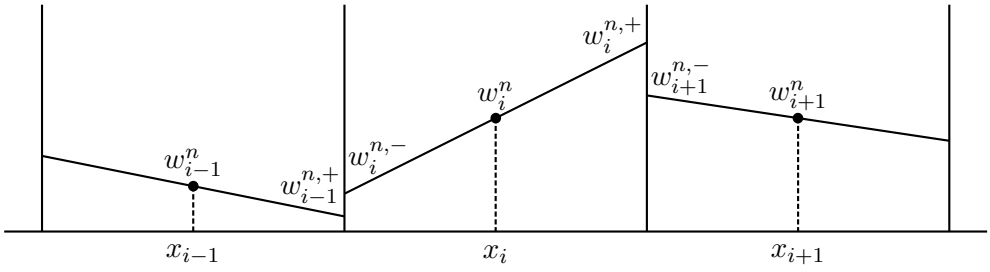


FIGURE 1.2 – Reconstruction affine par morceaux de la solution

Pour que le schéma (1.23) soit complet, il reste à préciser comment déterminer l'incrément Δw_i^n.

Dans le cas de la méthode MUSCL d'ordre deux, il est nécessaire de limiter la pente de la reconstruction afin d'assurer certaines propriétés comme la diminution de la variation totale dans le cas scalaire ou la robustesse (et la stabilité) dans le cas d'un système. Dans les méthodes MUSCL classiques, la limitation de la pente s'effectue en même temps que la reconstruction en utilisant un limiteur de pentes. Un limiteur est une fonction $L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ continue qui est consistante, au sens où $L(\sigma, \sigma) = \sigma$, pour tout σ dans $\mathbb{R}^d$, et bornée, c'est-à-dire qu'il existe une constante $M > 0$ telle que

$$\|L(\sigma_1, \sigma_2)\| \leq M \max\left(\|\sigma_1\|, \|\sigma_2\|\right).$$

Une fois le limiteur choisi, on définit l'incrément Δw_i^n sur chaque cellule par

$$\Delta w_i^n = \frac{1}{2} L(w_i^n - w_{i-1}^n, w_{i+1}^n - w_i^n).$$

On donne ici quelques exemples de limiteurs d'ordre deux qui seront utilisés dans ce manuscrit :

- limiteur minmod :

$$\text{minmod}(\sigma_L, \sigma_R) = \begin{cases} \min(\sigma_L, \sigma_R), & \text{si } \sigma_L > 0 \text{ et } \sigma_R > 0, \\ \max(\sigma_L, \sigma_R), & \text{si } \sigma_L < 0 \text{ et } \sigma_R < 0, \\ 0, & \text{sinon;} \end{cases}$$

- limiteur superbee :

$$\text{superbee}(\sigma_L, \sigma_R) = \text{maxmod}(\text{minmod}(2\sigma_L, \sigma_R), \text{minmod}(\sigma_L, 2\sigma_R)),$$

où la fonction maxmod est définie par

$$\text{maxmod}(\sigma_L, \sigma_R) = \begin{cases} \max(\sigma_L, \sigma_R), & \text{si } \sigma_L > 0 \text{ et } \sigma_R > 0, \\ \min(\sigma_L, \sigma_R), & \text{si } \sigma_L < 0 \text{ et } \sigma_R < 0, \\ 0, & \text{sinon;} \end{cases}$$

- limiteur monotonized central-difference (MC) :

$$\text{MC}(\sigma_L, \sigma_R) = \begin{cases} \min\left(2\sigma_L, 2\sigma_R, \frac{\sigma_L+\sigma_R}{2}\right), & \text{si } \sigma_L > 0 \text{ et } \sigma_R > 0, \\ \max\left(2\sigma_L, 2\sigma_R, \frac{\sigma_L+\sigma_R}{2}\right), & \text{si } \sigma_L < 0 \text{ et } \sigma_R < 0, \\ 0, & \text{sinon;} \end{cases}$$

- limiteur de van Leer :

$$\text{vanLeer}(\sigma_L, \sigma_R) = \begin{cases} \frac{\sigma_L|\sigma_R|+\sigma_R|\sigma_L|}{|\sigma_L|+|\sigma_R|}, & \text{si } \sigma_L\sigma_R \neq 0, \\ 0, & \text{sinon;} \end{cases}$$

- limiteur de van Albada 1 :

$$\text{vanAlbada1}(\sigma_L, \sigma_R) = \begin{cases} \frac{\sigma_L\sigma_R(\sigma_L+\sigma_R)}{\sigma_L^2+\sigma_R^2}, & \text{si } \sigma_L\sigma_R \neq 0, \\ 0, & \text{sinon.} \end{cases}$$

Notons qu'aucun de ces limiteurs ne permet d'assurer que les états reconstruits $w_i^{n,\pm}$ restent dans Ω. Il est donc parfois nécessaire d'effectuer une surlimitation de la pente pour assurer cette propriété.

En supposant que la reconstruction préserve l'ensemble Ω, on peut aisément montrer que le schéma MUSCL (1.23) est robuste.

Lemme 1.9. *Supposons que le schéma d'ordre un (1.18) est robuste sous la condition CFL (1.17). Supposons de plus que la reconstruction préserve l'ensemble Ω, c'est-à-dire*

$$\forall i \in \mathbb{Z}, \quad w_i^n \in \mathbb{Z} \quad \Rightarrow \quad \forall i \in \mathbb{Z}, \quad w_i^{n,\pm} \in \Omega.$$

Alors le schéma MUSCL (1.23) est robuste sous la condition CFL

$$\frac{\Delta t}{\Delta x} \max_{i\in\mathbb{Z}} \left\{ |\widetilde{\lambda}^{\pm}(w_i^{n,+}, w_{i+1}^{n,-})|, |\widetilde{\lambda}^{\pm}(w_i^{n,-}, w_i^{n,+})| \right\} \leq \frac{1}{4}. \tag{1.24}$$

Avant de prouver ce résultat, remarquons que la condition CFL (1.24) qui assure la robustesse du schéma MUSCL est deux fois plus restrictive que la condition CFL (1.17) qui assure la robustesse du schéma d'ordre un associé. En fait, la condition CFL (1.24) n'est pas optimale. En effet lorsque l'on fixe les incréments $\Delta w_i^n = 0$, on retrouve le schéma d'ordre un, mais la condition CFL (1.24) est clairement trop restrictive, puisque (1.17) suffit dans ce cas à assurer la robustesse. Il est malheureusement trop compliqué de trouver une condition CFL générale pour le schéma MUSCL qui serait « consistante » avec la condition CFL (1.17) associée au schéma d'ordre un. Il est possible de le faire pour certains systèmes de lois de conservation simples. Cela sort du cadre de cette brève présentation et l'on se contentera d'utiliser la condition CFL (1.24) qui a l'avantage d'être valable pour n'importe quel système de lois de conservation.

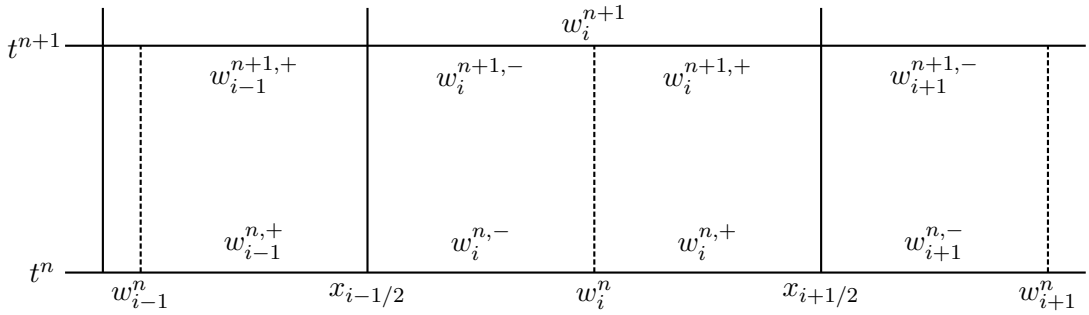


FIGURE 1.3 – Interprétation du schéma MUSCL comme moyenne de schémas d'ordre un

Démonstration. Au temps t^n, on suppose connue une approximation de la solution

$$W_{\Delta x}^n(x) = \begin{cases} w_i^{n,-} & \text{si } x \in [x_{i-1/2}, x_i], \\ w_i^{n,+} & \text{si } x \in [x_i, x_{i+1/2}]. \end{cases}$$

On fait évoluer cette solution approchée par le schéma d'ordre un (1.18) jusqu'au temps t^{n+1} et on obtient un état $w_i^{n+1,-}$ sur la cellule $[x_{i-1/2}, x_i]$ et un état $w_i^{n+1,+}$ sur la cellule $[x_i, x_{i+1/2}]$ (voir Figure 1.3). Ces états évolués s'écrivent

$$w_i^{n+1,-} = w_i^{n,-} - \frac{\Delta t}{\Delta x/2}\left(F\left(w_i^{n,-}, w_i^{n,+}\right) - F\left(w_{i-1}^{n,+}, w_i^{n,-}\right)\right), \tag{1.25}$$

$$w_i^{n+1,+} = w_i^{n,+} - \frac{\Delta t}{\Delta x/2}\left(F\left(w_i^{n,+}, w_{i+1}^{n,-}\right) - F\left(w_i^{n,-}, w_i^{n,+}\right)\right). \tag{1.26}$$

La condition CFL (1.24) permet d'assurer que les états $w_i^{n,\pm}$ donnés par les schémas d'ordre un (1.25) et (1.26) sont dans Ω. Enfin, on vérifie aisément que l'état w_i^{n+1} donné par le schéma MUSCL (1.23) s'écrit comme la moyenne des deux états $w_i^{n+1,-}$ et $w_i^{n+1,+}$. Puisque l'ensemble Ω est supposé convexe, on en déduit que l'état w_i^{n+1} reste dans Ω. □

1.5 La méthode MOOD

Dans les méthodes MUSCL classiques, la limitation de la pente est effectuée *a priori* et de manière globale sur toutes les cellules. L'inconvénient d'une telle méthode est que l'on limite la pente dans tout le domaine de calcul, y compris dans des zones où cela n'est pas nécessaire, ce qui peut engendrer une perte de précision.

Une autre approche a été développée récemment par Clain, Diot et Loubère [34, 47]. Il s'agit de la méthode MOOD (Multi-dimensional Optimal Order Detection) qui est basée sur une limitation *a posteriori* de la reconstruction. Le principe consiste à n'effectuer la procédure de limitation que sur les cellules « problématiques » où elle s'avère nécessaire.

Bien que la méthode MOOD ait été originellement conçue en deux dimensions d'espace, on présente ici, pour simplifier, une version 1D de cette méthode. On fixe un entier $d_{\max} > 0$ qui correspond au degré maximum des polynômes que l'on utilise dans la reconstruction. On suppose que pour tout entier j tel que $0 \leq j \leq d_{\max}$, on dispose d'une procédure de reconstruction qui fournit un polynôme $\mathcal{P}_i^j$, de degré j, approchant la solution de (1.1) sur la cellule K_i. Les polynômes $\mathcal{P}_i^j$ sont habituellement obtenus par des techniques d'interpolation ou de minimisation. On ne détaille pas ici le stencil nécessaire pour assurer la précision de la reconstruction. On suppose par contre que la reconstruction de degré zéro vérifie

$$\mathcal{P}_i^0(x) = w_i^n, \quad \forall x \in K_i.$$

La deuxième pierre angulaire de la méthode MOOD est l'existence d'un ensemble de critères de détection $\mathcal{A}$, dont le but est de tester si la solution numérique est acceptable ou pas. Un état évolué w_i^{n+1} vérifiant tous les critères de $\mathcal{A}$ est dit $\mathcal{A}$–éligible. Comme exemples de critères pouvant intervenir dans $\mathcal{A}$, on trouve par exemple dans [34, 47] la robustesse et des conditions de type principe du maximum discret, éventuellement tempérées par un processus de détection des extrema réguliers. L'ensemble $\mathcal{A}$ doit vérifier une propriété essentielle pour assurer que l'algorithme présenté plus bas s'arrête en un nombre fini d'itérations : sous la condition CFL (1.17), un état obtenu par le schéma d'ordre un doit être $\mathcal{A}$–éligible. Autrement dit, on doit avoir l'implication suivante :

$$\begin{cases} \forall i \in \mathbb{Z}, w_i^n \text{ est } \mathcal{A}\text{–éligible}, \\ w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(F\left(w_i^n, w_{i+1}^n\right) - F\left(w_{i-1}^n, w_i^n\right)\right) \end{cases} \quad \Rightarrow \quad w_i^{n+1} \text{ est } \mathcal{A}\text{–éligible.} \tag{1.27}$$

Avant de présenter l'algorithme de la méthode MOOD, on introduit sur chaque cellule, un degré effectif de reconstruction d_i. Tous ces degrés seront initialement fixés à $d_{\max}$ et ils seront décrémentés jusqu'à ce que la solution soit admissible selon l'ensemble de critères $\mathcal{A}$. La méthode MOOD est alors décrite par la procédure suivante.

1. **Initialisation des degrés :** tous les degrés effectifs de reconstruction d_i sont fixés à $d_{\max}$.
2. **Évaluation des états reconstruits :** sur chaque cellule K_i, on évalue des états reconstruits $w_i^{n,\pm}$ aux interfaces $x_{i\pm1/2}$ de la façon suivante :
$$w_i^{n,-} = \mathcal{P}_i^{d_{i-1,i}}\left(x_{i-1/2}\right) \quad \text{et} \quad w_i^{n,+} = \mathcal{P}_i^{d_{i,i+1}}\left(x_{i+1/2}\right),$$
où l'on a noté $d_{i,j} = \min(d_i, d_j)$.
3. **Évolution de la solution :** Les états sont mis à jour par
$$w_i^{n+1,\star} = w_i^n - \frac{\Delta t}{\Delta x}\left(F\left(w_i^{n,+}, w_{i+1}^{n,-}\right) - F\left(w_{i-1}^{n,+}, w_i^{n,-}\right)\right).$$
4. **Test d'$\mathcal{A}$–éligibilité :** pour chaque $i \in \mathbb{Z}$, si l'état $w_i^{n+1,\star}$ n'est pas $\mathcal{A}$–éligible, alors on décrémente le degré de reconstruction effectif d_i sur la cellule K_i.
   - Si tous les états $w_i^{n+1,\star}$ sont $\mathcal{A}$–éligibles, alors la solution est valide et l'on pose
   $$w_i^{n+1} = w_i^{n+1,\star}, \quad \forall i \in \mathbb{Z}.$$
   - Sinon, la solution est recalculée à partir de l'étape 2.

On montre aisément que cet algorithme se termine en un nombre fini d'itérations et que la solution obtenue vérifie toutes les conditions de l'ensemble $\mathcal{A}$.

Il peut paraître étrange d'utiliser les degrés $d_{i-1,i}$ et $d_{i,i+1}$ plutôt que le degré d_i pour la reconstruction des états $w_i^{n,\pm}$ sur la cellule K_i. Ce choix est pourtant nécessaire pour garantir que l'algorithme se termine. En effet, Si l'on utilise d_i dans l'étape 2 de l'algorithme, en supposant que celui-ci tombe à zéro, la solution évoluée sera donnée par

$$w_i^{n+1,\star} = w_i^n - \frac{\Delta t}{\Delta x}\left(F\left(w_i^n, w_{i+1}^{n,-}\right) - F\left(w_{i-1}^{n,+}, w_i^n\right)\right).$$

Cet état n'a *a priori* aucune raison d'être $\mathcal{A}$-éligible et on ne peut plus décrémenter le degré de reconstruction d_i. Par conséquent, l'algorithme risque de ne pas se terminer. L'utilisation de $d_{i-1,i}$ et $d_{i,i+1}$ oblige à calculer deux polynômes de degrés différents sur chaque cellule, ce qui alourdit le coût de calcul, mais est indispensable.

Remarquons par contre que seules les cellules dont le degré effectif de reconstruction a été décrémenté et les cellules qui leur sont adjacentes doivent être réactualisées. Par conséquent, seules ces cellules ont besoin d'être testées pour l'$\mathcal{A}$–éligibilité lors de la phase 4 de la prochaine itération. Cela permet de réduire considérablement le temps de calcul.

2

Schémas MUSCL basés sur une reconstruction de gradients par maillages duaux (DMGR)

Introduction

Ce travail est dédié à l'approximation numérique des solutions faibles des systèmes hyperboliques de lois de conservation en deux dimensions d'espace. Le système considéré s'écrit

$$\partial_t w + \partial_x f(w) + \partial_y g(w) = 0, \quad (x,y) \in \mathbb{R}^2, \quad t \geq 0, \tag{2.1}$$

où $w : \mathbb{R}^2 \times \mathbb{R}^+ \to \Omega$ est le vecteur des inconnues, $f, g : \Omega \to \mathbb{R}^d$ sont les fonctions flux respectivement dans la direction x et y et sont supposées régulières et $\Omega \subset \mathbb{R}^d$ désigne un ensemble d'états admissibles supposé convexe. Ce système est complété par la donnée initiale

$$w(x,y,t=0) = w_0(x,y), \quad (x,y) \in \mathbb{R}^2. \tag{2.2}$$

On suppose que le système (2.1) est hyperbolique, c'est-à-dire que pour tout vecteur unitaire $\nu = (\nu_x, \nu_y)^T \in \mathbb{S}^1$ et pour tout vecteur $w \in \Omega$, on demande que la matrice

$$\nu_x \nabla_w f(w) + \nu_w \nabla_w g(w)$$

soit diagonalisable dans $\mathbb{R}$. Les solutions des systèmes hyperboliques pouvant développer des discontinuités, on considère les solutions de (2.1)–(2.2) au sens faible.

Afin d'écarter les solutions non physiques, on suppose donc que les solutions faibles vérifient les inégalités d'entropie suivantes (voir [76, 77])

$$\partial_t \eta(w) + \partial_x \mathcal{F}(w) + \partial_y \mathcal{G}(w) \leq 0, \tag{2.3}$$

où $\eta : \Omega \to \mathbb{R}$ est convexe et $\mathcal{F}, \mathcal{G} : \Omega \to \mathbb{R}$ sont les flux d'entropie respectivement dans la direction x et y et vérifient les relation de compatibilité suivantes :

$$\nabla f \nabla \eta = \nabla \mathcal{F}, \quad \nabla g \nabla \eta = \nabla \mathcal{G}.$$

Durant les dernières décennies, la dérivation de schémas numériques pour approcher les solutions faibles de (2.1) était et reste un domaine de recherche très actif et de nombreuses méthodes sont proposées dans la littérature. Dans ce chapitre, on se concentre sur les schémas volumes finis d'ordre deux (voir [56, 80, 100] et les références incluses) et on s'intéresse à des extensions du célèbre schéma MUSCL introduit par van Leer [102]. La méthode MUSCL d'ordre deux est une des procédures de montée en ordre des méthodes de volumes finis les plus populaires grâce à la relative simplicité de sa dérivation. En effet, en s'appuyant sur un schéma volumes finis d'ordre un standard, van Leer [102] a suggéré d'introduire une reconstruction linéaire de la solution approchée sur chaque cellule pour évaluer le flux numérique de manière plus précise.

Les reconstructions de gradients sont facilement définies pour des problèmes en une dimension. On renvoie le lecteur au livre de LeVeque [80] (voir également [86, 21, 12, 72]) où plusieurs approches sont détaillées. Bien sûr, ces reconstructions MUSCL 1D s'étendent directement à des maillages 2D cartésiens. Cependant, l'évaluation correcte des gradients s'avère être une difficulté importante dans la dérivation du schéma MUSCL dès que des maillages non structurés sont considérés. Plusieurs techniques de reconstruction de gradients ont été introduites lorsque l'on utilise des triangles comme cellules de contrôle (voir [83, 56, 82, 24, 27, 28]) mais celles-ci ne s'étendent pas aisément à des volumes de contrôle plus généraux. Un des objectifs de ce travail est d'introduire une technique de reconstruction de gradients indépendante de la définition des cellules. Pour traiter cette difficulté, on propose une extension adaptée des méthodes « Discrete Duality Finite Volume » (DDFV) ([65, 66, 49, 3]) dans le cas des systèmes hyperboliques non linéaires.

En fait, la procédure de reconstruction linéaire s'avère être un point crucial pour empêcher l'apparition d'oscillations non physiques. Pour éviter de telles nuisances numériques, une procédure de limitation doit être envisagée. Une littérature importante est dédiée au développement de limiteurs pertinents. Pour un état de l'art non exhaustif, on renvoie le lecteur aux travaux suivants : [6, 103, 21, 69, 45, 20, 9, 43, 14, 84, 54, 24, 27, 33, 81, 74]. Dans [6, 103, 45], l'objectif principal est la dérivation de limiteurs globalement 2D pour obtenir des reconstructions pertinentes sur des maillages non structurés. Une telle approche peut engendrer trop de viscosité numérique dans les régions où les solutions sont régulières. Certains auteurs proposent un limiteur supplémentaire pour réduire artificiellement la viscosité numérique (voir [84, 74]). Afin d'assurer certaines propriétés de stabilité, certains articles sont dédiés à la modification des limiteurs usuels pour satisfaire des conditions de type principe du maximum (par exemple [69, 9, 81]). Dans ces travaux, le limiteur est basé sur une seule reconstruction de gradient par cellule. Une autre stratégie consiste à introduire une pente pour chaque côté à l'intérieur d'une cellule (voir [24, 33]).

La limitation est également un point-clé pour assurer des propriétés de robustesse. En effet, l'extension MUSCL n'est pas capable de rétablir l'invariance de l'ensemble Ω et une attention spécifique doit être portée à la procédure de limitation pour garantir cette propriété essentielle. Plusieurs stratégies ont été introduites pendant la dernière décennie. Par exemple, dans [86, 87, 88, 43, 14], les auteurs introduisent une limitation pertinente du gradient pour obtenir la robustesse requise sous une condition CFL restrictive. Plus récemment, pour éviter la restriction sur la CFL, dans [34, 47], une limitation a posteriori est suggérée en imposant une pente nulle du gradient quand la robustesse n'est pas vérifiée. Une telle technique rend le schéma robuste, mais au prix d'une perte locale de précision. Par conséquent, rendre le schéma robuste semble, en général, nécessiter une condition CFL plus restrictive et donc augmente le coût de calcul. Actuellement, il semble impossible d'éviter cette restriction additionnelle sur la CFL venant de la procédure MUSCL. Un des buts de ce travail est donc d'optimiser cette restriction. Mentionnons dès à présent que la condition CFL provient à la fois du nombre CFL associé au schéma d'ordre un et de la robustesse de la méthode MUSCL. Dans le cas d'un maillage 2D non structuré, on établit que la condition CFL usuelle d'ordre un n'est pas du tout optimale

(par exemple voir [88]) car elle introduit un facteur multiplicatif inapproprié et parfois important à la CFL, ce qui peut entraîner un coût de calcul excessif. On présentera une condition CFL d'ordre un « optimale ».

Le chapitre s'organise de la façon suivante. Dans la Partie 2.1, on rappelle brièvement les notations principales concernant les méthodes de volumes finis en 2D et l'on présente les schémas de type Godunov 2D qui vont être notre cadre de travail. On détaille ensuite l'obtention d'une CFL optimale permettant d'assurer la robustesse du schéma d'ordre un. On montre enfin comment on peut déduire de la robustesse du schéma d'ordre un que le schéma MUSCL préserve l'ensemble Ω. Dans la Partie 2.2, on présente le schéma DMGR. Cette approche est complétée par une procédure de limitation destinée à assurer la robustesse de la méthode. La Partie 2.3 est dédiée aux résultats numériques. On conclut finalement le chapitre avec quelques commentaires et perspectives.

2.1 Robustesse des schémas volumes finis en deux dimensions d'espace

2.1.1 Schémas volumes finis en deux dimensions d'espace

Afin de présenter le formalisme général des méthodes de volumes finis en deux dimensions d'espace, il est nécessaire d'introduire quelques notations pour décrire un maillage 2D non structuré. On considère un maillage de $\mathbb{R}^2$ constitué de volumes de contrôle polygonaux $(K_i)_{i\in\mathbb{Z}}$. Pour chaque cellule, on note $\gamma(i)$ l'ensemble des indices des cellules voisines de K_i. Pour tout $j \in \gamma(i)$, on appelle e_{ij} le côté commun qui sépare K_i et K_j et ν_{ij} la normale extérieure à e_{ij} (voir Figure 2.1). Enfin, on notera $|K_i|$ l'aire de la cellule K_i et $\mathcal{P}_i = \sum_{j\in\gamma(i)} |e_{ij}|$ son périmètre, où $|e_{ij}|$ désigne la longueur du côté e_{ij}.

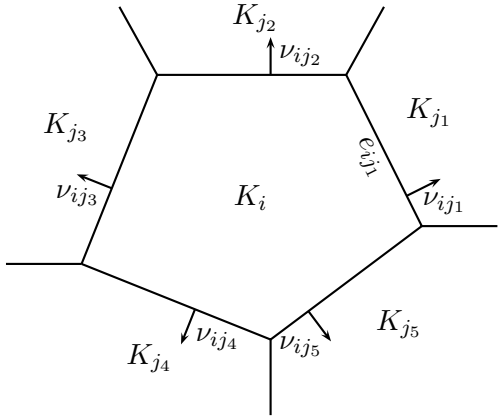


FIGURE 2.1 – Géométrie de la cellule K_i

On note w_i^n une approximation de la solution exacte de (2.1)–(2.2) sur la cellule K_i au temps t^n. La condition initiale (2.2) est discrétisée en prenant la moyenne sur chaque cellule

$$w_i^0 = \frac{1}{|K_i|}\int_{K_i} w_0(x,y)dxdy.$$

Pour un vecteur unitaire $\nu = (\nu_x, \nu_y)^T$ du cercle unité $\mathbb{S}^1$, on définit le flux dans la direction ν par

$$h_\nu(w) = \nu_x f(w) + \nu_y g(w). \tag{2.4}$$

De la même façon qu'en une dimension d'espace, on intègre alors l'équation (2.1) sur le prisme $K_i \times [t^n, t^{n+1}]$ pour obtenir par la formule de Green

$$\int_{K_i} w(x,y,t^{n+1})dxdy - \int_{K_i} w(x,y,t^n)dxdy + \sum_{j\in\gamma(i)} \int_{t^n}^{t^{n+1}} \int_{e_{ij}} h_{\nu_{ij}}(w(x,y))d\sigma dt = 0.$$

En considérant que sur l'arrête e_{ij}, on a localement un problème en une dimension, on est amenés à faire l'approximation

$$\int_{t^n}^{t^{n+1}} \int_{e_{ij}} h_{\nu_{ij}}(w(x,y))d\sigma dt \approx |e_{ij}|\Delta t\varphi(w_i^n, w_j^n, \nu_{ij}),$$

où $\varphi(\cdot,\cdot,\nu)$ est un flux numérique 1D pouvant éventuellement différer selon la direction ν. On en déduit la formulation générale d'un schéma volumes finis en deux dimensions d'espace :

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{|K_i|} \sum_{j\in\gamma(i)} |e_{ij}|\varphi(w_i^n, w_j^n, \nu_{ij}). \tag{2.5}$$

Une méthode de volumes finis sur un maillage non structuré est donc entièrement déterminée par la donnée dans chaque direction d'un flux numérique 1D.

Contrairement au cas unidimensionnel, un schéma écrit sous la forme (2.5) n'est pas automatiquement conservatif. Pour qu'il le soit, on doit en plus vérifier que l'on a

$$\forall w_L, w_R \in \Omega, \quad \forall \nu \in \mathbb{S}^1, \quad \varphi(w_L, w_R, \nu) = -\varphi(w_R, w_L, -\nu). \tag{2.6}$$

Les autres propriétés du schéma (2.5) sont définies de la même façon qu'en une dimension. On dit que le schéma (2.5) est consistant (ou de façon équivalente que le flux numérique φ est consistant) si l'on a

$$\forall w \in \Omega, \quad \forall \nu \in \mathbb{S}^1, \quad \varphi(w,w,\nu) = h_\nu(w).$$

On dit que le schéma (2.5) est robuste si

$$\forall i \in \mathbb{Z}, \quad w_i^n \in \Omega \quad \Rightarrow \quad \forall i \in \mathbb{Z}, \quad w_i^{n+1} \in \Omega.$$

Pour un triplet d'entropie $(\eta, \mathcal{F}, \mathcal{G})$ fixé, on définit le flux d'entropie dans la direction ν par

$$\mathcal{H}_\nu(w) = \nu_x \mathcal{F}(w) + \nu_y \mathcal{G}(w), \quad \text{avec } \nu = (\nu_x, \nu_y)^T \in \mathbb{S}^1. \tag{2.7}$$

Le schéma (2.5) est dit entropique si pour tout triplet d'entropie $(\eta, \mathcal{F}, \mathcal{G})$, il existe un flux numérique d'entropie $\Phi : \Omega \times \Omega \times \mathbb{R}^2 \to \mathbb{R}$, consistant avec $\mathcal{H}_\nu$ (i.e. $\Phi(w,w,\nu) = \mathcal{H}_\nu(w)$), tel que le schéma vérifie l'inégalité d'entropie discrète

$$\eta(w_i^{n+1}) \leq \eta(w_i^n) - \frac{\Delta t}{|K_i|} \sum_{j\in\gamma(i)} |e_{ij}|\Phi(w_i^n, w_j^n, \nu_{ij}). \tag{2.8}$$

Comme en une dimension d'espace, il existe une classe de schémas pour laquelle il est relativement facile de montrer ces propriétés. Il s'agit des schémas de type Godunov que nous introduisons maintenant.

2.1.2 Schémas de type Godunov en deux dimensions d'espace

Dans cette étude, afin d'établir une condition CFL optimale, on restreint notre analyse aux flux numériques de type Godunov décrits par le travail de Harten, Lax et van Leer [64]. On introduit ainsi, dans chaque direction $\nu \in \mathbb{S}^1$, un solveur de Riemann approché unidimensionnel $\widetilde{W}_\nu\left(\frac{x}{t}, w_L, w_R\right)$ qui vérifie les propriétés de stabilité et de consistance de la Définition 1.4. On peut *a priori* choisir un solveur différent dans chaque direction ν. On note respectivement $\lambda_\nu^-(w_L, w_R)$ et $\lambda_\nu^+(w_L, w_R)$ la plus petite et la plus grande vitesse caractéristique développée par le solveur $\widetilde{W}_\nu\left(\frac{x}{t}, w_L, w_R\right)$. Signalons que par rapport aux notations de la Partie 1.3, on omet les tildes sur les vitesses, ce qui ne pose pas problème puisqu'il n'y a pas de confusion possible ici. Le solveur approché $\widetilde{W}_\nu$ doit en particulier vérifier sous la condition CFL

$$\frac{\Delta t}{\delta} \max\left|\lambda_\nu^\pm(w_L, w_R)\right| \leq \frac{1}{2}, \tag{2.9}$$

la consistance avec la forme intégrale

$$\frac{1}{\delta}\int_{-\frac{\delta}{2}}^{\frac{\delta}{2}} \widetilde{W}_\nu\left(\frac{x}{t}, w_L, w_R\right) dx = \frac{1}{2}(w_L + w_R) - \frac{\Delta t}{\delta}\left(h_\nu(w_R) - h_\nu(w_L)\right). \tag{2.10}$$

Soulignons que si le pas d'espace Δx admet une définition claire en une dimension, ce n'est pas le cas en 2D et on se contente donc pour l'instant de faire apparaitre un paramètre δ dans (2.10). La seule contrainte sur ce paramètre est de vérifier la condition CFL (2.9). Comme on s'y attend, δ va représenter une longueur de maillage caractéristique dont la définition précise sera donnée ultérieurement de façon à obtenir une condition CFL la moins restrictive possible.

La définition du flux numérique 2D suit alors la définition (1.20) du flux numérique pour un schéma de Godunov 1D :

$$\varphi(w_L, w_R, \nu) = h_\nu(w_L) + \frac{\delta}{2\Delta t} w_L - \frac{1}{\Delta t}\int_{-\frac{\delta}{2}}^{0} \widetilde{W}_\nu\left(\frac{x}{\Delta t}, w_L, w_R\right) dx. \tag{2.11}$$

On dit que le schéma volumes finis (2.5) est de type Godunov associé au solveur de Riemann approché $\widetilde{W}_\nu$ si son flux numérique est défini par (2.11). Dans toute la suite de ce chapitre, on suppose que le schéma (2.5) vérifie une telle définition. Par ailleurs, on suppose aussi que pour tout triplet d'entropie $(\eta, \mathcal{F}, \mathcal{G})$, le solveur de Riemann approché $\widetilde{W}_\nu$ vérifie la consistance avec la forme intégrale de l'inégalité d'entropie (2.3) :

$$\frac{1}{\delta}\int_{-\frac{\delta}{2}}^{\frac{\delta}{2}} \eta\left(\widetilde{W}_\nu\left(\frac{x}{t}, w_L, w_R\right)\right) dx \leq \frac{1}{2}\left(\eta(w_L) + \eta(w_R)\right) - \frac{\Delta t}{\delta}\left(\mathcal{H}_\nu(w_R) - \mathcal{H}_\nu(w_L)\right), \quad \forall \nu \in \mathbb{S}^1, \tag{2.12}$$

où $\mathcal{H}_\nu$ est défini par (2.7). Cela nous amène à définir le flux numérique d'entropie par

$$\Phi(w_L, w_R, \nu) = \mathcal{H}_\nu(w_L) + \frac{\delta}{2\Delta t}\eta(w_L) - \frac{1}{\Delta t}\int_{-\frac{\delta}{2}}^{0} \eta\left(\widetilde{W}_\nu\left(\frac{x}{\Delta t}, w_L, w_R\right)\right) dx. \tag{2.13}$$

La définition (2.11) du flux numérique permet d'obtenir facilement les propriétés de consistance et de conservation du schéma :

Lemme 2.1. *Le flux numérique (2.11) est consistant.*
De plus, si l'on impose la relation suivante :

$$\int_{-\frac{\delta}{2}}^{0} \widetilde{W}_\nu\left(\frac{x}{t}, w_L, w_R\right) dx = \int_{0}^{\frac{\delta}{2}} \widetilde{W}_{-\nu}\left(\frac{x}{t}, w_R, w_L\right) dx, \tag{2.14}$$

alors le schéma (2.5) est conservatif.

Démonstration. D'après la propriété (iii) de la Définition 1.4, on a

$$\begin{aligned}\varphi(w,w,\nu) &= h_\nu(w) + \frac{\delta}{2\Delta t}w - \frac{1}{\Delta t}\int_{-\frac{\delta}{2}}^{0} \widetilde{W}_\nu\left(\frac{x}{\Delta t}, w, w\right) dx \\ &= h_\nu(w),\end{aligned}$$

ce qui prouve la consistance.

En ce qui concerne la conservation, en utilisant l'hypothèse (2.14), on a

$$\varphi(w_L,w_R,\nu) = h_\nu(w_L) + \frac{\delta}{2\Delta t}w_L - \frac{1}{\Delta t}\int_{0}^{\frac{\delta}{2}} \widetilde{W}_{-\nu}\left(\frac{x}{t}, w_R, w_L\right) dx.$$

On utilise alors la consistance avec la forme intégrale (2.10) pour obtenir

$$\begin{aligned}\varphi(w_L,w_R,\nu) &= h_\nu(w_L) + \frac{\delta}{2\Delta t}w_L - \frac{\delta}{2\Delta t}(w_L+w_R) + h_\nu(w_R) - h_\nu(w_L) \\ &\quad + \frac{1}{\Delta t}\int_{-\frac{\delta}{2}}^{0} \widetilde{W}_{-\nu}\left(\frac{x}{t}, w_R, w_L\right) dx \\ &= h_\nu(w_R) - \frac{\delta}{2\Delta t}w_R + \frac{1}{\Delta t}\int_{-\frac{\delta}{2}}^{0} \widetilde{W}_{-\nu}\left(\frac{x}{t}, w_R, w_L\right) dx.\end{aligned}$$

Par la définition (2.4) du flux h_ν, on a $h_\nu(w) = -h_{-\nu}(w)$, ce qui nous permet de conclure que

$$\varphi(w_L,w_R,\nu) = -\varphi(w_R,w_L,-\nu),$$

et donc que le schéma est conservatif. □

Remarque. Soulignons que le flux numérique φ, donné par (2.11), est défini indépendamment du paramètre δ. En effet, on introduit

$$\Lambda^-(w_L,w_R,\nu) = \min\left(0, \lambda^-(w_L,w_R,\nu)\right) \text{ et } \Lambda^+(w_L,w_R,\nu) = \max\left(0, \lambda^+(w_L,w_R,\nu)\right).$$

Il s'agit en fait de la partie négative de $\lambda^-(w_L,w_R,\nu)$ et de la partie positive de $\lambda^+(w_L,w_R,\nu)$ pour lesquelles on a préféré ne pas utiliser les notations usuelles pour ne pas multiplier la présence de signes $-$ et $+$. Avec ces notations, on a

$$\int_{-\frac{\delta}{2}}^{0} \widetilde{W}_\nu\left(\frac{x}{\Delta t}, w_L, w_R\right) dx = \left(\frac{\delta}{2} + \Lambda^-(w_L,w_R,\nu)\Delta t\right) w_L + \int_{\Lambda^-(w_L,w_R,\nu)\Delta t}^{0} \widetilde{W}_\nu\left(\frac{x}{\Delta t}, w_L, w_R\right) dx.$$

Le flux numérique se réécrit alors

$$\varphi(w_L,w_R,\nu) = h_\nu(w_L) + \Lambda^-(w_L,w_R,\nu)w_L - \frac{1}{\Delta t}\int_{\Lambda^-(w_L,w_R,\nu)\Delta t}^{0} \widetilde{W}_\nu\left(\frac{x}{\Delta t}, w_L, w_R\right) dx.$$

Enfin, on peut prouver une extension du lemme 2.1 concernant l'entropie :

Lemme 2.2. *Le flux numérique d'entropie (2.13) est consistant pour tout triplet d'entropie $(\eta, \mathcal{F}, \mathcal{G})$. De plus, si l'on impose la relation suivante :*

$$\int_{-\frac{\delta}{2}}^{0} \eta\left(\widetilde{W}_\nu\left(\frac{x}{t}, w_L, w_R\right)\right) dx = \int_{0}^{\frac{\delta}{2}} \eta\left(\widetilde{W}_{-\nu}\left(\frac{x}{t}, w_R, w_L\right)\right) dx, \tag{2.15}$$

alors le flux numérique d'entropie vérifie la propriété suivante :

$$\Phi(w_L,w_R,\nu) + \Phi(w_R,w_L,-\nu) \geq 0. \tag{2.16}$$

Démonstration. La définition (2.13) du flux numérique d'entropie et la propriété (iii) de la définition 1.4 d'un solveur de Riemann approché impliquent facilement la consistance du flux numérique d'entropie.

Pour établir la propriété (2.16), les relations (2.13) et (2.15) nous donnent

$$\Phi(w_L, w_R, \nu) = \mathcal{H}_\nu(w_L) + \frac{\delta}{2\Delta t}\eta(w_L) - \frac{1}{\Delta t}\int_0^{\frac{\delta}{2}} \eta\left(\widetilde{w}_{-\nu}\left(\frac{x}{\Delta t}, w_R, w_L\right)\right) dx,$$

ce qui implique

$$\begin{aligned}\Phi(w_L, w_R, \nu) + \Phi(w_R, w_L, -\nu) = \mathcal{H}_\nu(w_L) + \mathcal{H}_{-\nu}(w_R) + \frac{\delta}{2\Delta t}\eta(w_L) + \frac{\delta}{2\Delta t}\eta(w_R) \\ - \frac{1}{\Delta t}\int_{-\frac{\delta}{2}}^{\frac{\delta}{2}} \eta\left(\widetilde{w}_{-\nu}\left(\frac{x}{\Delta t}, w_R, w_L\right)\right) dx.\end{aligned}$$

L'inégalité (2.12) nous donne alors

$$\Phi(w_L, w_R, \nu) + \Phi(w_R, w_L, -\nu) \geq \mathcal{H}_{-\nu}(w_R) + \mathcal{H}_\nu(w_R).$$

D'après la définition (2.7), le flux exact d'entropie vérifie $\mathcal{H}_{-\nu} = -\mathcal{H}_\nu$ et on en déduit (2.16). □

Remarque. Signalons que dans le lemme 2.2, la propriété (2.15) doit être vérifiée par toute entropie η, ce qui est a priori une condition forte. Cependant, la propriété de conservation ponctuelle

$$\widetilde{W}_\nu\left(\frac{x}{t}, w_L, w_R\right) = \widetilde{W}_{-\nu}\left(-\frac{x}{t}, w_R, w_L\right), \quad \forall x \in \mathbb{R}, \quad \forall t > 0, \tag{2.17}$$

implique à la fois les deux relations (2.14) et (2.15). Par ailleurs, la propriété (2.17) est vérifiée par tous les solveurs de Riemann approchés classiques (par exemple Godunov, HLL, Relaxation de Suliciu pour les équations d'Euler, voir [64, 56, 38, 20]).

2.1.3 Robustesse des schémas 2D d'ordre un

En considérant le flux numérique donné par (2.11) dans le résultat suivant, on présente une condition CFL d'ordre un optimale.

Théorème 2.3. *Pour tout w_L et w_R dans Ω, on suppose que $\widetilde{W}_\nu\left(\frac{x}{t}, w_L, w_R\right)$ reste dans Ω. Soit w_i^n appartenant à Ω pour tout $i \in \mathbb{Z}$. Supposons que la condition CFL suivante est vérifiée :*

$$\Delta t \frac{\mathcal{P}_i}{|K_i|} \max_{j\in\gamma(i)} \left|\lambda^\pm(w_i^n, w_j^n, \nu_{ij})\right| \leq 1, \quad \forall i \in \mathbb{Z}. \tag{2.18}$$

Alors les états w_i^{n+1}, donnés par (2.5) restent dans Ω.

De plus, le schéma (2.5) est entropique.

Démonstration. Tout d'abord, remarquons que les deux conditions CFL (2.9) et (2.18) coïncident dès que $w_L = w_i^n$, $w_R = w_j^n$ et $\delta = \frac{2|K_i|}{\mathcal{P}_i}$. Par conséquent, la formulation intégrale du flux numérique (2.11) peut être adoptée :

$$\varphi(w_i^n, w_j^n, \nu_{ij}) = h_{\nu_{ij}}(w_i^n) + \frac{|K_i|}{\mathcal{P}_i\Delta t}w_i^n - \frac{1}{\Delta t}\int_{-\frac{|K_i|}{\mathcal{P}_i}}^0 \widetilde{W}_{\nu_{ij}}\left(\frac{x}{\Delta t}, w_i^n, w_j^n\right) dx. \tag{2.19}$$

On rappelle que d'après la formule de Green, on a $\sum_{j\in\gamma(i)} |e_{ij}| h_{\nu_{ij}}(w_i^n) = 0$. En injectant (2.19) dans (2.5), on obtient

$$w_i^{n+1} = \frac{1}{|K_i|} \sum_{j\in\gamma(i)} |e_{ij}| \int_{-\frac{|K_i|}{\mathcal{P}_i}}^{0} \widetilde{W}_{\nu_{ij}} \left(\frac{x}{\Delta t}, w_i^n, w_j^n \right) dx. \tag{2.20}$$

Puisque $\widetilde{W}_{\nu_{ij}} \left(\frac{x}{\Delta t}, w_i^n, w_j^n \right)$ est à valeurs dans Ω qui est un ensemble convexe, on en déduit immédiatement que

$$\widehat{w}_{ij} := \frac{\mathcal{P}_i}{|K_i|} \int_{-\frac{|K_i|}{\mathcal{P}_i}}^{0} \widetilde{W}_{\nu_{ij}} \left(\frac{x}{\Delta t}, w_i^n, w_j^n \right) dx \in \Omega.$$

L'état w_i^{n+1} se réécrit comme une combinaison convexe des états $\widehat{w}_{ij}$:

$$w_i^{n+1} = \sum_{j\in\gamma(i)} \frac{|e_{ij}|}{\mathcal{P}_i} \widehat{w}_{ij}$$

et ainsi, la propriété de préservation de l'ensemble Ω est établie.

Pour prouver les inégalités d'entropie, on utilise la convexité de η et l'inégalité de Jensen dans l'équation (2.20) pour obtenir

$$\eta(w_i^{n+1}) \leq \frac{1}{|K_i|} \sum_{j\in\gamma(i)} |e_{ij}| \int_{-\frac{|K_i|}{\mathcal{P}_i}}^{0} \eta \left(\widetilde{W}_{\nu_{ij}} \left(\frac{x}{\Delta t}, w_i^n, w_j^n \right) \right) dx. \tag{2.21}$$

Par la formule de Green, on a

$$\eta(w_i^n) - \frac{\Delta t}{|K_i|} \sum_{j\in\gamma(i)} |e_{ij}| \left(\mathcal{H}_{\nu_{ij}}(w_i^n) + \frac{|K_i|}{\Delta t \mathcal{P}_i} \eta(w_i^n) \right) = 0. \tag{2.22}$$

En sommant les équations (2.21) et (2.22), on obtient

$$\begin{aligned} \eta(w_i^{n+1}) \leq\ & \eta(w_i^n) \\ & - \frac{\Delta t}{|K_i|} \sum_{j\in\gamma(i)} \left(\mathcal{H}_{\nu_{ij}}(w_i^n) + \frac{|K_i|}{\Delta t \mathcal{P}_i} \eta(w_i^n) - \frac{1}{\Delta t} \int_{-\frac{|K_i|}{\mathcal{P}_i}}^{0} \eta \left(\widetilde{W}_{\nu_{ij}} \left(\frac{x}{\Delta t}, w_i^n, w_j^n \right) \right) dx \right). \end{aligned}$$

La définition (2.13) du flux numérique d'entropie nous donne exactement les inégalités d'entropie discrètes (2.8). Le Lemme 2.2 assure enfin que le flux numérique d'entropie Φ est consistant. □

Pour conclure cette partie dédiée à la présentation du schéma d'ordre un adopté, on souligne que des techniques similaires existent dans la littérature (par exemple voir [88]). En général, une attention spécifique est portée à la condition CFL et plusieurs travaux [87, 88] montrent la robustesse en supposant une condition CFL deux fois plus restrictive que (2.18). Cependant, il est important de noter que cette restriction du pas de temps peut facilement être améliorée pour obtenir la condition CFL (2.18) dès que le flux numérique provient d'un solveur de Riemann approché. De plus, on remarque que la condition CFL (2.18) coïncide avec celle proposée dans [14] dès que les cellules sont des polygones réguliers. Un calcul simple montre par contre que la condition CFL (2.18) est moins restrictive que celle proposée dans [14] quand les cellules sont des polygones non réguliers. En effet, considérons une cellule rectangulaire de côtés Δx et Δy, avec $\Delta x < \Delta y$. La condition CFL proposée dans [14] s'écrit dans ce cas

$$\frac{\Delta t}{\Delta x} \max |\lambda^{\pm}| \leq \frac{1}{4}, \tag{2.23}$$

alors que (2.18) s'écrit

$$\Delta t \frac{\Delta x + \Delta y}{\Delta x \Delta y} \max |\lambda^{\pm}| \leq \frac{1}{2}. \tag{2.24}$$

On voit facilement que la condition CFL (2.23) est plus restrictive que (2.24). Si l'on suppose de plus que $\Delta x \ll \Delta y$, la CFL (2.24) devient

$$\frac{\Delta t}{\Delta x} \max |\lambda^{\pm}| \leq \frac{1}{2}.$$

On a alors gagné un facteur 2 par rapport à (2.23).

2.1.4 Le schéma MUSCL 2D

On considère w_i^n une approximation de la solution exacte de la solution du système (2.1) sur la cellule K_i au temps t^n. Pour obtenir une approximation au temps $t^{n+1} = t^n + \Delta t$, le schéma MUSCL consiste à faire évoluer la suite $(w_i^n)_{i \in \mathbb{Z}}$ de la façon suivante :

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{|K_i|} \sum_{j \in \gamma(i)} |e_{ij}| \varphi \left(w_{ij}, w_{ji}, \nu_{ij} \right), \tag{2.25}$$

où w_{ij} et w_{ji} sont des approximations d'ordre deux de la solution au milieu du côté e_{ij} de chaque côté de l'interface (voir Figure 2.2). La procédure de reconstruction qui permet de déterminer ces états sera l'objet de la Partie 2.2. On suppose à nouveau que le flux numérique φ est associé dans chaque direction $\nu \in \mathbb{S}^1$ à un solveur de Riemann approché $\widetilde{W}_\nu$. De plus, on suppose que $\widetilde{W}_\nu$ vérifie la consistance avec la forme intégrale de l'inégalité d'entropie (2.12), la propriété de conservation en moyenne (2.14) et la propriété de conservation de l'entropie en moyenne (2.15). On rappelle que les Lemmes 2.1 et 2.2 impliquent alors respectivement que le flux numérique φ vérifie la propriété de conservation (2.6) et que l'entropie vérifie la propriété de dissipation (2.16).

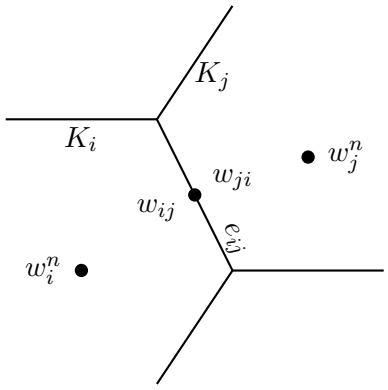


FIGURE 2.2 – Approximations d'ordre un w_i^n et w_j^n et reconstructions d'ordre deux w_{ij} et w_{ji}

L'objectif est maintenant d'étendre le Théorème 2.3 au schéma MUSCL d'ordre deux. Pour obtenir un tel résultat, on adopte la décomposition classique du schéma MUSCL en une combinaison convexe de schémas d'ordre un (par exemple, voir [87, 14]). Par conséquent, la restriction CFL du schéma MUSCL est fortement reliée à celle du schéma d'ordre un (2.18).

On doit tout d'abord introduire quelques notations supplémentaires. Soit G_i le centre de masse de la cellule K_i et pour tout $j \in \gamma(i)$, on appelle T_{ij} le triangle formé par le point G_i et le côté e_{ij}. Ces triangles sont essentiels puisque ce sont les sous-cellules sur lesquelles on va appliquer le Théorème 2.3. Soit $\gamma(i,j)$ l'ensemble des indices des deux sous-cellules voisines

de T_{ij} dans la cellule K_i. Pour tout $k \in \gamma(i,j)$, on note e^i_{jk} le côté commun qui sépare T_{ij} et T_{ik} et ν^i_{jk} la normale unitaire sortante à e^i_{jk} (voir Figure 2.3). Enfin, on désigne respectivement par $|T_{ij}|$ et $\mathcal{P}_{ij} = |e_{ij}| + \sum_{k\in\gamma(i,j)} |e^i_{jk}|$, l'aire de la sous-cellule T_{ij} et son périmètre, où l'on a noté $|e^i_{jk}|$ la longueur du côté e^i_{jk}.

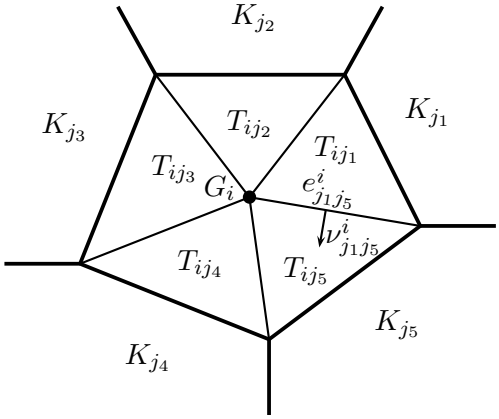


FIGURE 2.3 – Décomposition en sous-cellules de la cellule K_i

Avec ces notations, on peut établir la robustesse du schéma MUSCL d'ordre deux.

Théorème 2.4. *Pour tous w_L et w_R dans Ω, on suppose que $\widetilde{W}_\nu\left(\frac{x}{t}, w_L, w_R\right)$ reste dans Ω. On se donne des états moyens w_i^n et des états reconstruits w_{ij} qui appartiennent à Ω pour tout $i \in \mathbb{Z}$ et pour tout $j \in \gamma(i)$. Supposons que les états reconstruits vérifient la propriété de conservation suivante :*

$$\sum_{j\in\gamma(i)} \frac{|T_{ij}|}{|K_i|} w_{ij} = w_i^n. \tag{2.26}$$

Si la condition CFL

$$\Delta t \max_{j\in\gamma(i)} \frac{\mathcal{P}_{ij}}{|T_{ij}|} \max_{k\in\gamma(i,j)} \left\{|\lambda^\pm(w_{ij}, w_{ji}, \nu_{ij})|, |\lambda^\pm(w_{ij}, w_{ik}, \nu^i_{jk})|\right\} \leq 1, \quad \forall i \in \mathbb{Z}, \tag{2.27}$$

est vérifiée, alors les états évolués w_i^{n+1}, donnés par (2.25), restent dans Ω.

De plus, le schéma (2.25) vérifie les inégalités d'entropie suivantes :

$$\eta(w_i^{n+1}) \leq \overline{\eta}_i^n - \frac{\Delta t}{|K_i|} \sum_{j\in\gamma(i)} |e_{ij}| \Phi(w_{ij}, w_{ji}, \nu_{ij}), \tag{2.28}$$

où $\overline{\eta}_i^n$ est défini comme suit :

$$\overline{\eta}_i^n = \sum_{j\in\gamma(i)} \frac{|T_{ij}|}{|K_i|} \eta(w_{ij}). \tag{2.29}$$

Démonstration. On suppose connue une approximation de la solution au temps t^n, constante par morceaux, dont la valeur sur la sous-cellule T_{ij} est w_{ij}. On fait évoluer cette solution par le schéma d'ordre un (2.5) jusqu'au temps t^{n+1} et l'on obtient un état w_{ij}^{n+1} sur la cellule T_{ij}, donné par

$$w_{ij}^{n+1} = w_{ij} - \frac{\Delta t}{|T_{ij}|} \left(|e_{ij}| \phi(w_{ij}, w_{ji}, \nu_{ij}) + \sum_{k\in\gamma(i,j)} |e^i_{jk}| \phi(w_{ij}, w_{ik}, \nu^i_{jk}) \right).$$

La condition CFL (2.18) nécessaire pour pouvoir appliquer le Théorème 2.3 sur la sous-cellule T_{ij} est exactement la condition CFL (2.27). Cela implique donc que l'état w_{ij}^{n+1} est dans Ω. On calcule maintenant la combinaison convexe des w_{ij}^{n+1} pondérés par $|T_{ij}|/|K_i|$ et l'on trouve

$$\frac{1}{|K_i|}\sum_{j\in\gamma(i)}|T_{ij}|w_{ij}^{n+1} = \sum_{j\in\gamma(i)}\frac{|T_{ij}|}{|K_i|}w_{ij} - \frac{\Delta t}{|K_i|}\sum_{j\in\gamma(i)}|e_{ij}|\phi(w_{ij},w_{ji},\nu_{ij}) \\ - \frac{\Delta t}{|K_i|}\sum_{j\in\gamma(i)}\sum_{k\in\gamma(i,j)}|e_{jk}^i|\phi(w_{ij},w_{ik},\nu_{jk}^i).$$

Remarquons que dans le dernier terme, chacun des flux apparaît deux fois, mais dans des directions opposées. En utilisant la propriété de conservation (2.6) vérifiée par le flux numérique, on en déduit que

$$\sum_{j\in\gamma(i)}\sum_{k\in\gamma(i,j)}|e_{jk}^i|\phi(w_{ij},w_{ik},\nu_{jk}^i) = 0.$$

Grâce à la propriété de conservation de la reconstruction (2.26), on montre que les états intermédiaires w_{ij}^{n+1} vérifient la relation suivante :

$$\frac{1}{|K_i|}\sum_{j\in\gamma(i)}|T_{ij}|w_{ij}^{n+1} = w_i^n - \frac{\Delta t}{|K_i|}\sum_{j\in\gamma(i)}|e_{ij}|\phi(w_{ij},w_{ji},\nu_{ij}).$$

Enfin, la définition du schéma d'ordre deux (2.25) nous donne

$$w_i^{n+1} = \frac{1}{|K_i|}\sum_{j\in\gamma(i)}|T_{ij}|w_{ij}^{n+1}. \tag{2.30}$$

On a montré que l'état w_i^{n+1} est une combinaison convexe d'états dans Ω, il est donc lui-même dans Ω.

Pour établir les inégalités d'entropie (2.28)–(2.29), la seconde partie du Théorème 2.3 assure que l'on a

$$\eta(w_{ij}^{n+1}) \le \eta(w_{ij}) - \frac{\Delta t}{|T_{ij}|}\left(|e_{ij}|\Phi(w_{ij},w_{ji},\nu_{ij}) + \sum_{k\in\gamma(i,j)}|e_{jk}^i|\Phi(w_{ij},w_{ik},\nu_{jk}^i)\right).$$

L'inégalité de Jensen discrète, appliquée à l'équation (2.30), donne immédiatement

$$\eta(w_i^{n+1}) \le \frac{1}{|K_i|}\sum_{j\in\gamma(i)}|T_{ij}|\eta(w_{ij}^{n+1}),$$

et l'on en déduit

$$\eta(w_i^{n+1}) \le \eta_i^n - \frac{\Delta t}{|K_i|}\sum_{j\in\gamma(i)}|e_{ij}|\Phi(w_{ij},w_{ji},\nu_{ij}) - \frac{\Delta t}{|K_i|}\sum_{j\in\gamma(i)}\sum_{k\in\gamma(i,j)}|e_{jk}^i|\Phi(w_{ij},w_{ik},\nu_{jk}^i). \tag{2.31}$$

Enfin, la propriété de dissipation de l'entropie (2.16) assure que

$$\frac{\Delta t}{|K_i|}\sum_{j\in\gamma(i)}\sum_{k\in\gamma(i,j)}|e_{jk}^i|\Phi(w_{ij},w_{ik},\nu_{jk}^i) \ge 0$$

et le résultat est prouvé. □

Remarque. On a montré que le schéma d'ordre deux vérifie certaines inégalités d'entropie discrètes. On se gardera cependant de parler de schéma d'ordre deux entropique. En effet, la notion de schéma d'ordre deux entropique est délicate et sera l'un des objets de l'étude du chapitre 3.

Pour conclure cette partie, on montre que dès qu'on adopte une reconstruction affine, la propriété de conservation de la reconstruction (2.26) admet une réécriture équivalente très simple. En effet, considérons une fonction affine $\widetilde{w}_i : K_i \to \mathbb{R}^d$ définie par

$$\widetilde{w}_i(X) = w_{G_i} + \mu_i \cdot (X - G_i), \quad \text{avec } X \in K_i, \tag{2.32}$$

où $w_{G_i} \in \Omega$ est un vecteur constant et $\mu_i \in \mathbb{R}^d \times \mathbb{R}^d$ est une matrice constante donnée. On introduit Q_{ij} le milieu du côté e_{ij} pour définir

$$w_{ij} = \widetilde{w}_i(Q_{ij}). \tag{2.33}$$

On a alors la proposition suivante :

Proposition 2.5. *Supposons que les états reconstruits w_{ij} sont donnés par (2.32)–(2.33). Alors la condition (2.26) est vérifiée si et seulement si l'on a*

$$w_{G_i} = w_i^n. \tag{2.34}$$

Démonstration. D'une part, par définition du centre de masse G_i, on a

$$\int_{K_i} \widetilde{w}_i(X) dX = |K_i| w_{G_i}.$$

D'autre part, le centre de gravité d'un triangle étant situé aux deux tiers de la médiane en partant du sommet, l'intégrale de $\widetilde{w}_i$ sur les triangles T_{ij} est donnée par

$$\begin{aligned}
\int_{T_{ij}} \widetilde{w}_i(X) dX &= |T_{ij}| \widetilde{w}_i \left(\frac{2}{3} Q_{ij} + \frac{1}{3} G_i \right), \\
&= |T_{ij}| \left(w_{G_i} + \frac{2}{3} \mu_i \cdot (Q_{ij} - G_i) \right), \\
&= \frac{2}{3} |T_{ij}| w_{ij} + \frac{1}{3} |T_{ij}| w_{G_i}.
\end{aligned}$$

En sommant sur j, on obtient

$$\sum_{j \in \gamma(i)} |T_{ij}| w_{ij} = |K_i| w_{G_i},$$

et l'on en déduit l'équivalence entre les conditions (2.26) et (2.34). □

2.2 Le schéma DMGR

Dans cette partie, on décrit la technique DMGR. L'idée principale est de considérer deux maillages se recouvrant, à savoir un maillage primal et son maillage dual associé, puis d'écrire deux schémas volumes finis distincts sur ces maillages. Ce procédé augmente le nombre d'inconnues numériques, mais il permet de reconstruire des gradients très précis. Dans un premier temps, on dérive la technique DMGR en 1D, ce qui ne présente pas un grand intérêt sur le plan numérique. Cela va cependant permettre de mieux appréhender la dérivation 2D qui va suivre. On introduit ensuite quelques notations et on définit le maillage dual. Puis l'on décrit la procédure de reconstruction qui va être élaborée de manière à satisfaire les hypothèses du Théorème 2.4 pour assurer la robustesse requise. Enfin, on montre comment utiliser avantageusement les deux maillages afin d'obtenir des états à la fois aux sommets et aux centres de masse.

2.2.1 Le schéma DMGR en 1D

On considère, dans cette partie, un système de lois de conservation 1D

$$\partial_t w + \partial_x f(x) = 0. \tag{2.35}$$

On discrétise l'espace $\mathbb{R}$ en une suite croissante de points $(x_{i+1/2})_{i\in\mathbb{Z}}$ de manière uniforme, avec pour pas d'espace $\Delta x = x_{i+1/2} - x_{i-1/2}$ supposé constant. On définit les cellules primales $K_i = [x_{i-1/2}, x_{i+1/2}[$ et les cellules duales $K_{i+1/2} = [x_i, x_{i+1}[$, où l'on a noté $x_i = (x_{i-1/2} + x_{i+1/2})/2$ le milieu de la cellule K_i. Remarquons que l'ensemble des cellules $\{K_i\}_{i\in\mathbb{Z}}$ et l'ensemble $\{K_{i+1/2}\}_{i\in\mathbb{Z}}$ constituent deux maillages différents de la droite $\mathbb{R}$. On discrétise également le temps de la manière suivante, $t^n = n\Delta t$, où Δt est le pas de temps.

On suppose connues w_i^n et $w_{i+1/2}^n$ des approximations de la solution exacte de (2.35) au temps t^n respectivement sur la cellule K_i et sur la cellule $K_{i+1/2}$. On fait évoluer ces deux approximations par deux schémas MUSCL distincts :

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(F\left(w_i^{n,+}, w_{i+1}^{n,-}\right) - F\left(w_{i-1}^{n,+}, w_i^{n,-}\right)\right),$$

$$w_{i+1/2}^{n+1} = w_{i+1/2}^n - \frac{\Delta t}{\Delta x}\left(F\left(w_{i+1/2}^{n,+}, w_{i+3/2}^{n,-}\right) - F\left(w_{i-1/2}^{n,+}, w_{i+1/2}^{n,-}\right)\right),$$

où les états $w_i^{n,\pm}$ sont des reconstructions affines de la solution sur la cellule K_i à l'interface $x_{i\pm1/2}$ et les états $w_{i+1/2}^{n,\pm}$ sont des reconstructions affines de la solution sur la cellule $K_{i+1/2}$ à l'interface x_i et x_{i+1} (voir Figure 2.4).

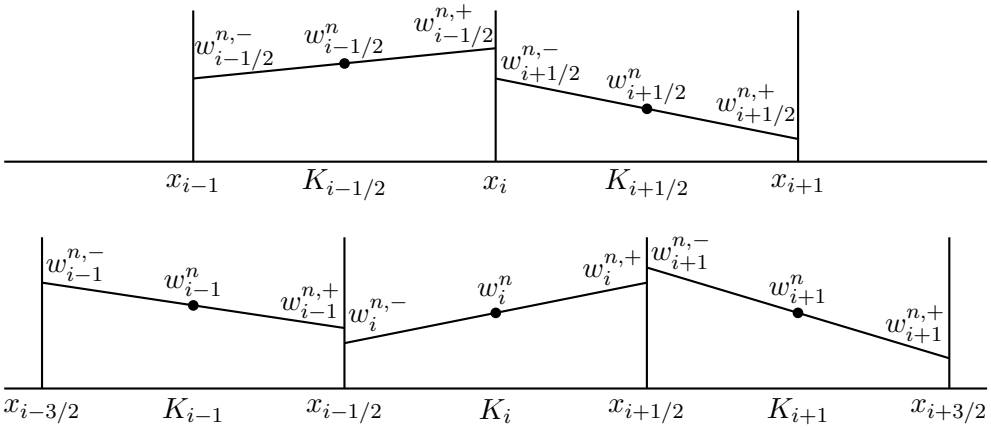


FIGURE 2.4 – Reconstruction affine par morceaux de la solution sur le maillage primal (bas) et le maillage dual (haut)

À ce stade, les deux schéma évoluent indépendamment. On va en fait les coupler pendant la phase de reconstruction. Considérons un limiteur 1D classique L. Pour tout $i \in \mathbb{Z}$, on définit alors les valeurs reconstruites aux interfaces par

$$w_i^{n,\pm} = w_i^n \pm L\left(w_i^n - w_{i-1/2}^n, w_{i+1/2}^n - w_i^n\right),$$

$$w_{i+1/2}^{n,\pm} = w_{i+1/2}^n \pm L\left(w_{i+1/2}^n - w_i^n, w_{i+1}^n - w_{i+1/2}^n\right).$$

L'idée de base du schéma DMGR est donc d'utiliser les approximations obtenues par les deux schémas pour reconstruire la solution aux différentes interfaces. En fait, l'approche est la même en 2D.

2.2.2 Maillage primal et maillage dual

On considère un maillage primal polygonal $\{K_i^p\}_{i\in\mathbb{Z}}$ de $\mathbb{R}^2$. On conserve toutes les notations introduites dans la partie 2.1 en leur ajoutant un exposant p pour préciser qu'il s'agit du maillage primal. On définit alors le maillage dual dont les cellules sont centrées autour des sommets du maillage primal. Pour cela, on note $\{S_i^p\}_{i\in\mathbb{Z}}$ l'ensemble de tous les sommets du maillage primal et pour tout $i \in \mathbb{Z}$, on définit $\delta(i)$ l'ensemble des indices j de toutes les cellules K_j^p qui admettent S_i^p pour sommet.

Pour chaque sommet primal S_i^p, on construit une cellule duale associée K_i^d obtenue en reliant les centres de masse $(G_j^p)_{j\in\delta(i)}$ des cellules voisines. Par construction, les sommets de la cellule K_i^d sont ainsi les points G_j^p avec $j \in \delta(i)$. Comme il est signalé par exemple dans [49], il est possible que les cellules K_i^d se chevauchent dans certains cas pathologiques. Par conséquent, on se limite à présent au cas où les cellules K_i^d forment une seconde partition de $\mathbb{R}^2$, appelée le maillage dual. On adopte à nouveau les notations introduites dans la partie 2.1 pour le maillage dual, mais en ajoutant un exposant d. On souligne que les sommets S_i^d du maillage dual coïncident par construction avec les centres de masse G_i^p du maillage primal. Par contre, le centre de masse G_i^d d'une cellule duale est en général distinct du sommet primal associé S_i^p (voir Figure 2.5).

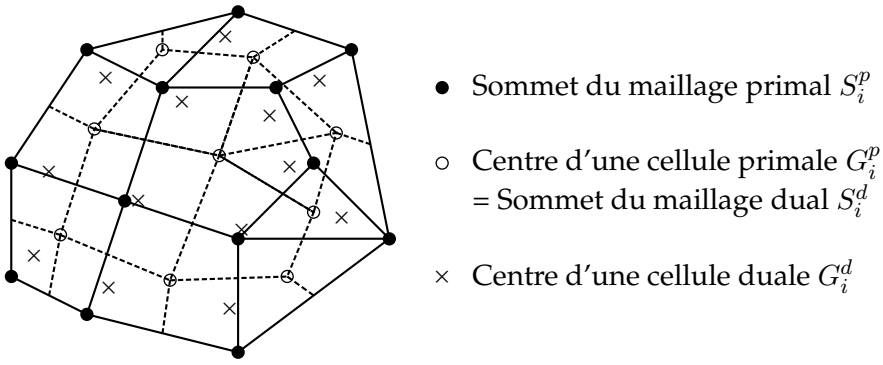


FIGURE 2.5 – Un exemple de maillage primal et son maillage dual associé

Maintenant que ces deux maillages sont construits, on écrit un schéma MUSCL sur chacun d'entre eux :

$$w_i^{p,n+1} = w_i^{p,n} - \frac{\Delta t}{|K_i^p|} \sum_{j\in\gamma^p(i)} |e_{ij}^p| \phi\left(w_{ij}^p, w_{ji}^p, \nu_{ij}^p\right), \tag{2.36}$$

$$w_i^{d,n+1} = w_i^{d,n} - \frac{\Delta t}{|K_i^d|} \sum_{j\in\gamma^d(i)} |e_{ij}^d| \phi\left(w_{ij}^d, w_{ji}^d, \nu_{ij}^d\right). \tag{2.37}$$

À ce stade, pour que le schéma soit complet, il ne reste plus qu'à préciser comment obtenir les états reconstruits w_{ij}^p et w_{ij}^d.

2.2.3 Procédure de reconstruction

On présente maintenant une procédure de reconstruction sur une cellule basée sur la connaissance du vecteur d'état aux sommets et au centre de masse de la cellule considérée. Puisque la procédure que l'on présente est la même sur les cellules primales et sur les cellules duales, on va détailler la reconstruction d'une façon unifiée et omettre l'exposant p ou d. De plus, dans cette partie, on suppose connue la valeur du vecteur des inconnues au sommets et au centre de masse de la cellule considérée. Les détails de l'évaluation de w aux sommets et au centre de masse seront l'objet de la partie suivante.

Pour alléger les notations, on considère une cellule K avec pour sommets $S_1, \ldots, S_k$. Afin de simplifier les calculs, on écrira abusivement $S_{k+1} = S_1$. Le centre de masse de la cellule K est noté G et pour $1 \leq i \leq k$, on introduit $T_{i+1/2}$ le triangle formé par les points S_i, S_{i+1} et G. Enfin, $Q_{i+1/2}$ désignera le milieu du côté S_iS_{i+1} (voir Figure 2.6).

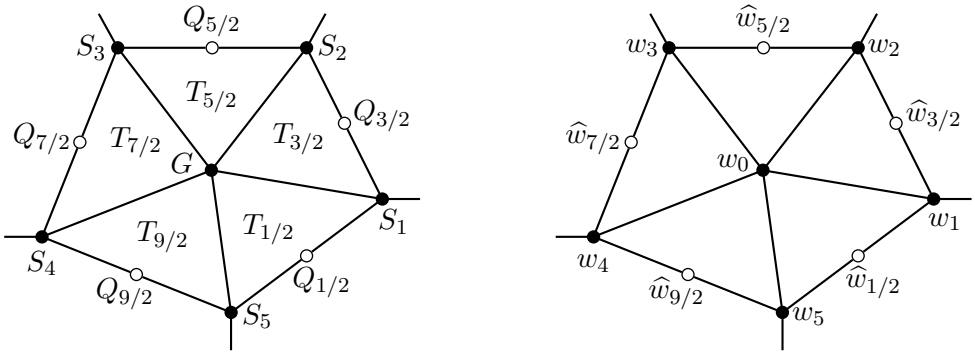


FIGURE 2.6 – Gauche : géométrie de la cellule K. Droite : états connus (points noirs) et états reconstruits inconnus (points blancs)

On suppose connus un état $w_i \in \Omega$ en chaque sommet S_i ainsi qu'un état $w_0 \in \Omega$ au centre de masse G. Notre but est de reconstruire un état $\widehat{w}_{i+1/2}$ en chaque point $Q_{i+1/2}$. On renvoie le lecteur à la Figure 2.6 pour une illustration des localisations des états connus et des états reconstruits.

Dans ce travail, on se limite aux reconstructions affines. D'après la Proposition 2.5, la condition (2.26) du Théorème 2.4 est satisfaite si et seulement si la reconstruction affine considérée $\widetilde{w}_\mu : K \to \mathbb{R}^d$ s'écrit

$$\widetilde{w}_\mu(X) = w_0 + \mu \cdot (X - G),$$

où μ représente une approximation précise du gradient spatial de la solution. Pour évaluer μ, on effectue une procédure en trois étapes.

Première étape : Reconstruction du gradient

Il existe une unique fonction affine $\overline{w}_{i+1/2} : T_{i+1/2} \mapsto \mathbb{R}^d$ telle que $\overline{w}_{i+1/2}(S_i) = w_i$, $\overline{w}_{i+1/2}(S_{i+1}) = w_{i+1}$ et $\overline{w}_{i+1/2}(G) = w_0$. On introduit alors $\overline{w} : K \mapsto \mathbb{R}^d$ la fonction continue affine par morceaux définie par

$$\overline{w}(X) = \overline{w}_{i+1/2}(X), \quad \forall X \in T_{i+1/2}.$$

Deuxième étape : Projection

On définit $\widetilde{w}_\mu : K \to \mathbb{R}^d$, la fonction affine qui résulte de la projection L^2 de $\overline{w}$ de la façon suivante :

$$\int_K \|\overline{w}(X) - \widetilde{w}_\mu(X)\|^2 \, dX = \min_{\alpha \in \mathbb{R}^d \times \mathbb{R}^2} \int_K \|\overline{w}(X) - \widetilde{w}_\alpha(X)\|^2 \, dX.$$

Par de simples arguments de convexité, on voit facilement qu'il existe un unique μ. On souligne que l'évaluation numérique de μ provient directement de la minimisation d'une fonction quadratique.

Troisième étape : Limitation des pentes

À ce stade, on ne peut pas utiliser directement les états reconstruits donnés par

$$\widehat{w}_{i+1/2} = \widetilde{w}_\mu(Q_{i+1/2}).$$

En effet, la reconstruction provenant de l'étape de projection décrite ci-dessus n'est pas nécessairement à valeurs dans Ω. Pour assurer que la reconstruction préserve Ω, on suggère de multiplier les pentes par un limiteur. On considère l'ensemble des limiteurs de pente admissibles défini comme suit :

$$F_{i+1/2} = \left\{\theta \in [0,1], \widetilde{w}_{s\mu}(Q_{i+1/2}) \in \Omega, \forall s \in [0,\theta]\right\}.$$

Puisque pour tout $0 \leq i \leq k-1$ on a $\widetilde{w}_0(Q_{i+1/2}) = w_0 \in \Omega$, aucun des ensembles $F_{i+1/2}$ n'est vide. On définit le limiteur de pente optimal de la façon suivante :

$$\beta = \min_{0\leq i\leq k-1} \sup(F_{i+1/2}) - \epsilon,$$

où $\epsilon > 0$ est un paramètre fixé tel que $\beta \in \bigcap_i F_{i+1/2}$. Par conséquent, on obtient que l'état $\widetilde{w}_{\beta\mu}(Q_{i+1/2})$ est dans Ω pour tout $0 \leq i \leq k-1$.

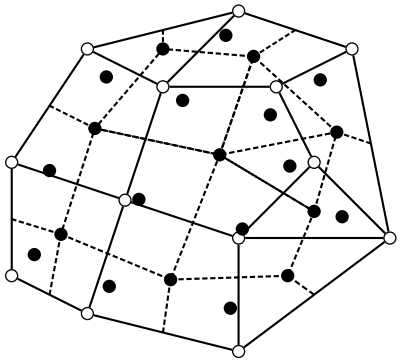
Les états reconstruits que l'on cherche à évaluer sont alors donnés par

$$\widehat{w}_{i+1/2} = \widetilde{w}_{\beta\mu}(Q_{i+1/2}).$$

Par construction, ils vérifient les hypothèses du Théorème 2.4 et ainsi le schéma DMGR préserve l'ensemble Ω sous la condition CFL (2.27).

2.2.4 Évaluation des états aux sommets et aux centres de masse

Pour conclure la présentation du schéma DMGR, nous devons préciser comment évaluer les états aux sommets et aux centres de masse à la fois sur les cellules primales et duales. Les états aux centres de masse sont donnés par les schémas numériques (2.36) et (2.37) respectivement pour le maillage primal et dual. D'autre part, les sommets du maillage dual coïncident exactement avec les centres de masses du maillage primal. Par conséquent, la valeur d'un état à un sommet dual est donnée par la valeur de l'état au centre de masse de la cellule primale associée (voir Figures 2.5 et 2.7).



- États connus par les schémas numériques
- ○ États inconnus que l'on doit déterminer

FIGURE 2.7 – États connus et inconnus sur le maillage primal et le maillage dual

À ce stade, on dispose de toutes les données pour pouvoir appliquer la procédure de reconstruction sur les cellules duales. Il en résulte une reconstruction affine $\widetilde{w}_i^d$ sur chaque cellule K_i^d.

Puis pour définir la reconstruction primale $\widetilde{w}_i^p$, on a besoin d'une évaluation du vecteur des inconnues aux sommets S_i^p. Comme on l'a signalé plus tôt, un sommet S_i^p du maillage primal ne coïncide pas nécessairement avec le centre de masse G_i^d de la cellule duale associée. Un premier choix basique pour l'état au point S_i^p serait de considérer la valeur au centre de masse

de la cellule duale associée $w_i^{d,n}$. Cependant il s'avère que cette approximation est trop grossière, particulièrement sur des maillages très déformés. Puisque l'on connaît la reconstruction duale $\widetilde{w}_i^d$, on suggère de considérer $\widetilde{w}_i^d(S_i^p)$ comme valeur de l'état au point S_i^p. Il s'agit d'une approximation bien plus précise comme vont l'illustrer les nombreux résultats numériques.

2.3 Résultats numériques

Le schéma DMGR a été implémenté dans un code non structuré pour approcher les solutions faibles des équations d'Euler bidimensionnelles :

$$\partial_t \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ E \end{pmatrix} + \partial_x \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ u(E+p) \end{pmatrix} + \partial_y \begin{pmatrix} \rho v \\ \rho uv \\ \rho v^2 + p \\ v(E+p) \end{pmatrix} = 0,$$

où ρ, (u, v), E et p désignent respectivement la densité, la vitesse, l'énergie totale et le pression du gaz. Le système est fermé par la loi des gaz parfaits

$$p = (\gamma - 1)\left(E - \rho\frac{u^2 + v^2}{2}\right),$$

où $\gamma \in]1, 3]$ est le coefficient adiabatique. Dans les expériences numériques qui vont suivre, γ sera fixé à 1.4. L'ensemble des états admissibles est défini par

$$\Omega = \left\{ w = (\rho, \rho u, \rho v, E)^T \in \mathbb{R}^4, \rho > 0, E - \frac{\rho}{2}(u^2 + v^2) > 0 \right\}.$$

Plusieurs cas-tests classiques ont été réalisés afin d'éprouver la précision et la stabilité de la méthode. Tout d'abord, on considère un tourbillon isentropique pour évaluer les erreurs et l'ordre numérique de convergence. En effet, la solution exacte du tourbillon isentropique est connue et suffisamment régulière. On s'intéresse ensuite à un cas-test de cisaillement afin de mettre l'accent sur les instabilités développées par ce phénomène. Celles-ci sont bien connues et ont été étudiées aussi bien théoriquement que numériquement (voir par exemple [105, 4, 41]). On poursuit en montrant plusieurs problèmes de Riemann 2D proposés dans [75]. On conclut la série de cas-tests avec deux expériences classiques, la double réflexion de Mach sur une rampe et la marche dans un écoulement Mach 3. Ces problèmes mettent l'accent sur la capacité du schéma DMGR à capturer précisément des chocs de forte amplitude et des discontinuités de contact. D'autre part, on utilise la double réflexion de Mach sur une rampe pour réaliser plusieurs comparaisons entre le schéma DMGR et une reconstruction MUSCL classique.

Pour tous ces tests, on choisit le flux numérique donné par le schéma de relaxation de Suliciu [38, 20]. On comparera les résultats obtenus par le schéma DMGR et par d'autres schémas. En ce qui concerne la restriction du pas de temps, on adopte les conditions CFL obtenues (2.18) ou (2.27) selon l'ordre de précision de la méthode utilisée. Afin d'illustrer l'efficacité de la discrétisation spatiale DMGR, on utilise seulement une approximation en temps d'ordre un. Pour pouvoir comparer de manière équitable, on doit réaliser les simulations avec un nombre de degrés de liberté (ddl) du même ordre. Le nombre de degrés de libertés est le nombre d'inconnues numériques du schéma : pour le schéma d'ordre un et le schéma MUSCL classique, il s'agit simplement du nombre de cellules du maillage, alors que pour le schéma DMGR, c'est le nombre de cellules du maillage primal plus le nombre de cellules du maillage dual.

Pour gérer les conditions de bord, la principale difficulté vient de la reconstruction sur les cellules touchant le bord. En effet, dès que la reconstruction au bord est connue, les flux au bord

sont évalués comme dans les méthodes de volumes finis usuelles. Dans un premier temps, on traite le cas d'une cellule duale touchant le bord, comme illustré par la Figure (2.8). Pour pouvoir appliquer la technique de reconstruction DMGR, il manque les valeurs des états aux points A' et B' et l'on doit les évaluer. Pour traiter ce problème, on peut envisager plusieurs techniques d'interpolation. On choisit la plus simple qui consiste à faire l'approximation $w(A') = w(A)$ et $w(B') = w(B)$. À ce stade, on peut effectuer la reconstruction sur les cellules duales et il reste à s'occuper des cellules primales. On remarque alors que l'on peut appliquer directement la procédure de reconstruction sur les cellules primales introduite dans la partie 2.2.4.

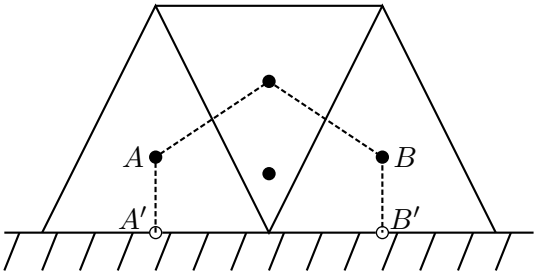


FIGURE 2.8 – Reconstruction sur une cellule touchant le bord. Points noirs : états connus. Points blancs : états inconnus

2.3.1 Tourbillon isentropique

Le problème du tourbillon isentropique est présenté dans [95, 47]. On considère un écoulement moyen caractérisé par $\rho_\infty = 1$, $(u_\infty, v_\infty) = (1, 1)$ et $p_\infty = 1$. On ajoute à cet écoulement un tourbillon isentropique, centré au point $(0, 0)$, qui est défini par des perturbations en vitesse et en température $T = p/\rho$, mais qui ne comprend pas de perturbation pour l'entropie $S = p/\rho^\gamma$. Le tourbillon est donné par les conditions

$$(\delta u, \delta v) = \frac{\beta}{2\pi} \exp\left(\frac{1 - r^2}{2}\right)(-y, x),$$

$$\delta T = -\frac{(\gamma - 1)\beta^2}{8\gamma\pi^2} \exp(1 - r^2), \quad \delta S = 0,$$

où $r^2 = x^2 + y^2$ et l'intensité du tourbillon est fixée à $\beta = 5$. La densité et la pression initiales sont ainsi données par

$$\rho = \rho_\infty \left(\frac{T}{T_\infty}\right)^{\frac{1}{\gamma-1}} = \left(1 - \frac{(\gamma - 1)\beta^2}{8\gamma\pi^2} \exp(1 - r^2)\right)^{\frac{1}{\gamma-1}},$$

$$p = \rho T = \left(1 - \frac{(\gamma - 1)\beta^2}{8\gamma\pi^2} \exp(1 - r^2)\right)^{\frac{\gamma}{\gamma-1}}.$$

Le domaine de calcul est donné par $[-5, 5] \times [-5, 5]$ et l'on impose des conditions de bord périodiques. On peut facilement vérifier que la solution exacte est simplement la convection du tourbillon initial à la vitesse moyenne (u_∞, v_∞).

Afin d'estimer l'ordre de convergence numérique du schéma DMGR, on calcule la solution numérique sur une série de maillages cartésiens raffinés. On arrête les simulations au temps $t = 10$ après une période, ce qui veut dire que la solution exacte au temps final et au temps

initial coïncident. On évalue les erreurs numériques L^1 et L^∞ relatives pour la densité de la façon suivante :

$$err_1 = \frac{\sum_i |\rho_i^N - \rho_i^0||K_i|}{\sum_i |\rho_i^0||K_i|} \quad \text{et} \quad err_\infty = \frac{\max_i |\rho_i^N - \rho_i^0|}{\max_i |\rho_i^0|},$$

où $(\rho_i^0)_i$ et $(\rho_i^N)_i$ désignent la valeur de la densité moyenne sur les cellules primales respectivement au temps initial et au temps final.

Dans le Tableau 2.1, on donne les erreurs L^1 et L^∞ et les taux de convergence pour le schéma d'ordre un et le schéma DMGR. On observe que l'ordre deux est effectivement atteint par le schéma DMGR. La Figure (2.9) présente les courbes de convergence correspondantes.

Schéma d'ordre un					**Schéma DMGR**				
Nb ddl	Erreur L^1		Erreur L^∞		Nb ddl	Erreur L^1		Erreur L^∞	
800	2.5718E-02	–	4.5570E-01	–	841	1.3296E-02	–	2.6987E-01	–
1600	2.4529E-02	0.15	4.3359E-01	0.15	3281	4.5387E-03	1.58	9.5782E-02	1.52
6400	2.0611E-02	0.25	3.9529E-01	0.13	12961	1.3228E-03	1.79	2.6453E-02	1.87
25600	1.5421E-02	0.42	3.0631E-01	0.37	52480	3.5605E-04	1.88	6.4104E-03	2.03
102400	1.0073E-02	0.61	1.9370E-01	0.66	205541	9.3363E-05	1.96	1.5533E-03	2.08
409600	5.9059E-03	0.77	1.0518E-01	0.88	409513	4.7932E-05	1.93	7.8221E-04	1.99

Tableau 2.1 – Erreurs L^1 et L^∞ et taux de convergence pour le problème du tourbillon isentropique en utilisant le schéma d'ordre un et le schéma DMGR

2.3.2 Cisaillement

On s'intéresse ici à un problème de cisaillement. Il s'agit en fait d'un problème de Riemann 1D simulé sur un maillage 2D non structuré. Le domaine de calcul est $[0,1] \times [0,1]$ et la donnée initiale est constituée d'un état gauche w_L pour $x < 0.5$ et d'un état droit w_R pour $x > 0.5$. Les deux pressions p_L et p_R sont choisies égales à 1 et les deux densités ρ_L et ρ_R sont fixées à 1.4 de manière à ce que les vitesses du son vaillent 1 des deux côtés de la discontinuité. Les deux vitesses normales u_L et u_R sont nulles alors que les vitesses tangentielles sont opposées :

$$v_L = -v_R = \overline{v}.$$

On va considérer plusieurs valeurs de $\overline{v} > 0$ afin de faire varier le nombre de Mach

$$M = \frac{\overline{v}}{\sqrt{\frac{\gamma p}{\rho}}} = \overline{v}.$$

De nombreuses études, aussi bien théoriques que numériques, ont été réalisées sur ce type de problème (voir [105, 4, 41]). Il en ressort que pour $M < \sqrt{2}$, l'écoulement développe des instabilités de Kelvin-Helmholtz qui grandissent exponentiellement, alors que les écoulements vérifiant $M > \sqrt{2}$ sont stables. Les expériences numériques que l'on présente maintenant montrent que le schéma DMGR a le comportement attendu. Ici, la perturbation nécessaire à la création d'instabilités ne sera pas ajoutée à la condition initiale, mais elle sera créée par la non-régularité du maillage. On va ainsi utiliser un maillage primal non structuré comme le montre la Figure 2.10. Afin de simplifier le traitement des conditions de bord, on utilise des cellules rectangulaires près du bord. Cela revient en fait à considérer un domaine fictif plus petit avec un maillage totalement non structuré. Pour pouvoir comparer les résultats avec ou sans perturbation, on effectue également les simulations sur un maillage cartésien. Les deux maillages contiennent environ 1.5×10^6 ddl. Pour tous les cas-tests considérés, on montre les solutions en densité et en vitesse tangentielle.

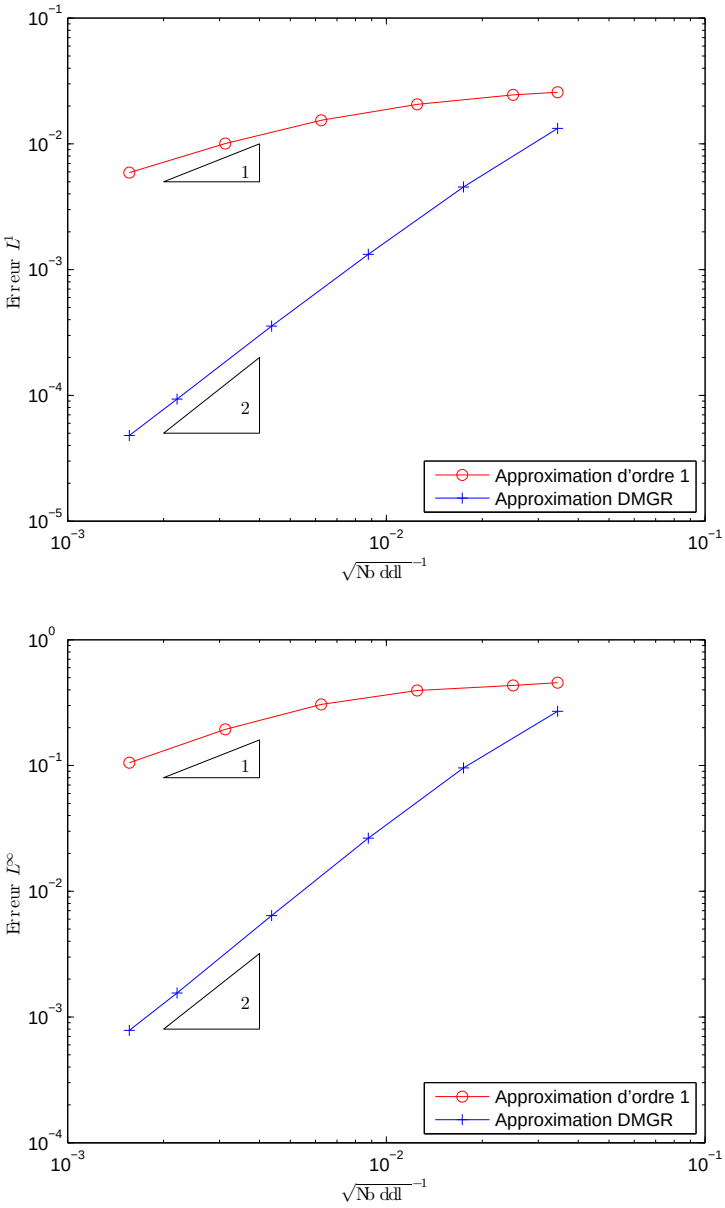


FIGURE 2.9 – Courbes de convergence pour le tourbillon isentropique approché par le schéma d'ordre un et le schéma DMGR. Haut : erreur L^1. Bas : erreur L^∞

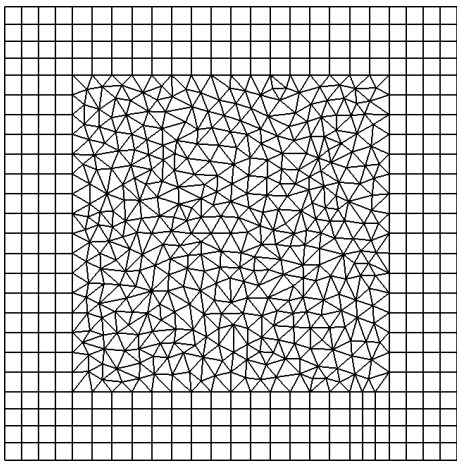


FIGURE 2.10 – Maillage primal non structuré utilisé pour le cisaillement et les problèmes de Riemann 2D

Dans un premier temps, on s'intéresse au cas instable, c'est-à-dire quand le nombre de Mach vérifie $M < \sqrt{2}$. On fixe ici $\overline{v} = 1$. La Figure 2.11 montre la solution sur un maillage cartésien au temps $t = 0.2$. Aucune instabilité de Kelvin-Helmholtz n'apparaît, ce qui est cohérent car il n'y a pas de perturbation et il s'agit en fait d'un cas-test purement 1D. On présente sur la Figure 2.12 les résultats sur un maillage non structuré aux temps $t = 0.1$, $t = 0.2$ et $t = 0.3$. On voit cette fois des instabilités de Kelvin-Helmholtz apparaître et grandir au cours du temps.

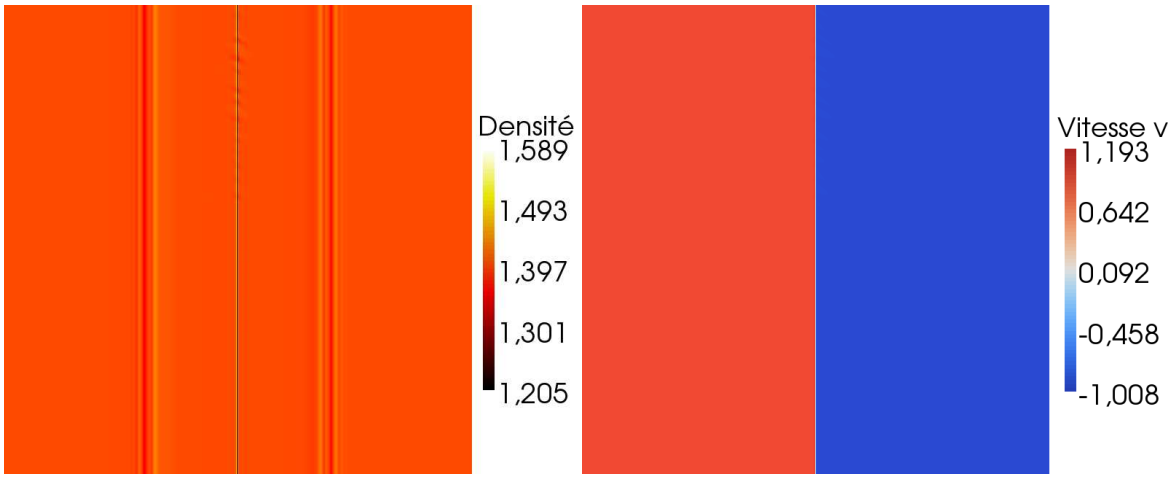


FIGURE 2.11 – Approximation DMGR en densité (gauche) et en vitesse (droite) du problème de cisaillement à Mach 1 sur un maillage cartésien contenant 1.5×10^6 ddl

On passe maintenant au cas stable en fixant la vitesse tangentielle à $\overline{v} = 3$. La simulation sur un maillage cartésien, que l'on peut voir Figure 2.13 au temps $t = 0.2$, montre à nouveau qu'il n'y a pas d'instabilité. Sur un maillage non structuré, on voit Figure 2.14 qu'au temps $t = 0.2$, aucune instabilité de Kelvin-Helmholtz n'apparaît.

2.3.3 Problèmes de Riemann 2D

On passe maintenant à la simulation de quelques problèmes de Riemann 2D proposés dans [75]. Le domaine de calcul est $[0, 1] \times [0, 1]$ et est divisé en quatre quadrants par les droites $x = 0.5$ et $y = 0.5$. Un problème de Riemann 2D est défini par un état constant initial sur chacun des

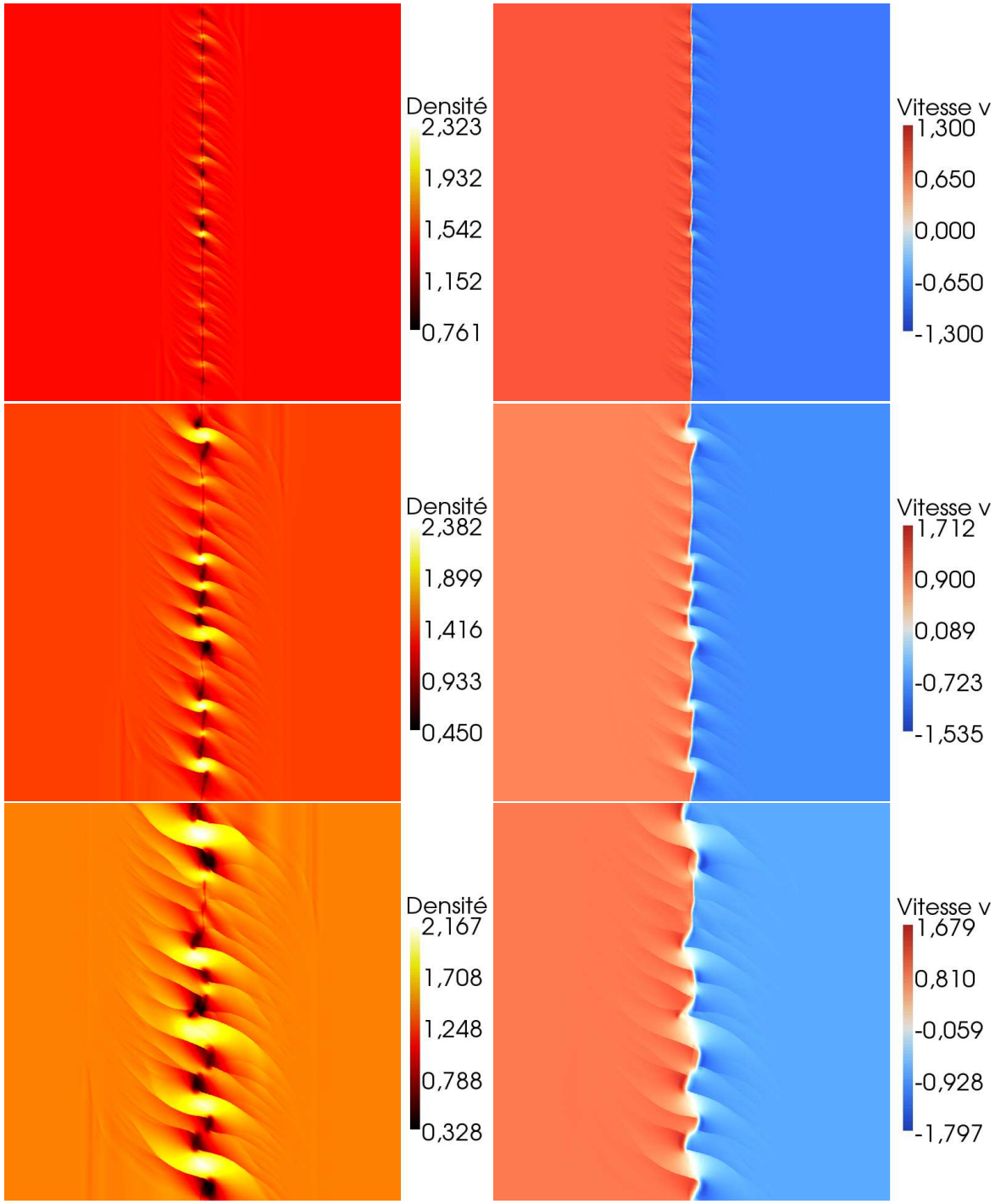


FIGURE 2.12 – Approximation DMGR en densité (gauche) et en vitesse (droite) du problème de cisaillement à Mach 1 sur un maillage non structuré contenant 1.5×10^6 ddl – Haut : $t = 0.1$. Milieu : $t = 0.2$. Bas : $t = 0.3$

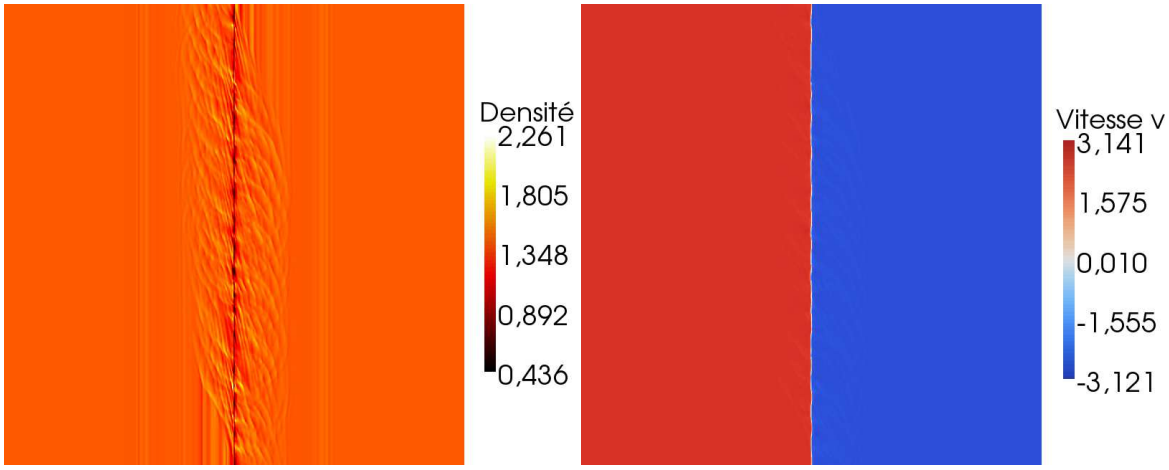


FIGURE 2.13 – Approximation DMGR en densité (gauche) et en vitesse (droite) du problème de cisaillement à Mach 3 sur un maillage cartésien contenant 1.5×10^6 ddl

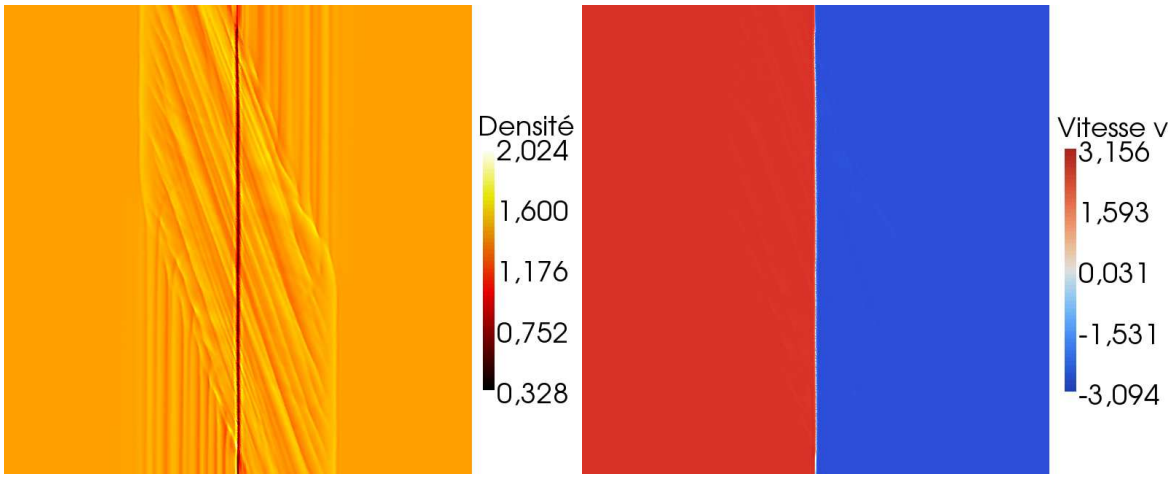


FIGURE 2.14 – Approximation DMGR en densité (gauche) et en vitesse (droite) du problème de cisaillement à Mach 3 sur un maillage non structuré contenant 1.5×10^6 ddl

quadrants. Dans le Tableau 2.2, on donne la valeur des quatre états initiaux pour chacun des trois problèmes de Riemann 2D considérés. Le premier problème correspond à la configuration 3 dans [75] (noté KT3), le deuxième à la configuration 5 dans [75] (noté KT5) et le dernier à la configuration 6 dans [75] (noté KT6). Les quatre problèmes de Riemann 1D entre les quadrants développent chacun exactement une onde : quatre chocs dans KT3 et quatre discontinuités de contact dans KT5 et KT6.

On utilise trois types de maillages primaux : un maillage purement cartésien, un maillage triangulaire régulier (voir Figure 2.15) et un maillage non structuré (voir Figure 2.10). Pour les deux derniers, une bande le long du bord est maillée de façon cartésienne afin de faciliter le traitement des conditions de bord.

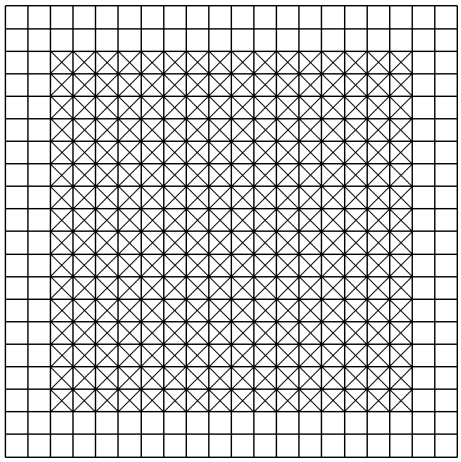


FIGURE 2.15 – Maillage primal triangulaire régulier utilisé pour les problèmes de Riemann 2D

		gauche				droite			
		ρ	u	v	p	ρ	u	v	p
KT3	haut	0.5323	1.206	0.0	0.3	1.5	0.0	0.0	1.5
	bas	0.138	1.206	1.206	0.029	0.5323	0.0	1.206	0.3
KT5	haut	2.0	-0.75	0.5	1.0	1.0	-0.75	-0.5	1.0
	bas	1.0	0.75	0.5	1.0	3.0	0.75	-0.5	1.0
KT6	haut	2.0	0.75	0.5	1.0	1.0	0.75	-0.5	1.0
	bas	1.0	-0.75	0.5	1.0	3.0	-0.75	-0.5	1.0

Tableau 2.2 – États initiaux des problèmes de Riemann 2D

On effectue les tests en utilisant environ 1.5×10^6 ddl. Le temps final des simulations est $t = 0.3$ pour les trois configurations. Tous les résultats présentés sont les solutions en densité.

On commence par montrer les résultats sur un maillage cartésien sur la Figure 2.16. On peut voir la précision du schéma DMGR et l'absence d'instabilités sur un maillage cartésien.

La Figure 2.17 montre les résultats obtenus sur le maillage triangulaire régulier. Sur KT3, les ondes de chocs restent stables comme on s'y attend. On voit cependant apparaître des petites instabilités dans la structure complexe centrale. Sur KT5 et KT6, des instabilités de Kelvin-Helmholtz se développent le long des discontinuités de contact.

On conclut en présentant Figure 2.18 les résultats sur le maillage non structuré. Les chocs présents sur le cas-test KT3 restent stables. Par contre, la structure centrale qui contient des contacts est complètement déstabilisée par la non-régularité du maillage. Sur les cas-tests KT5

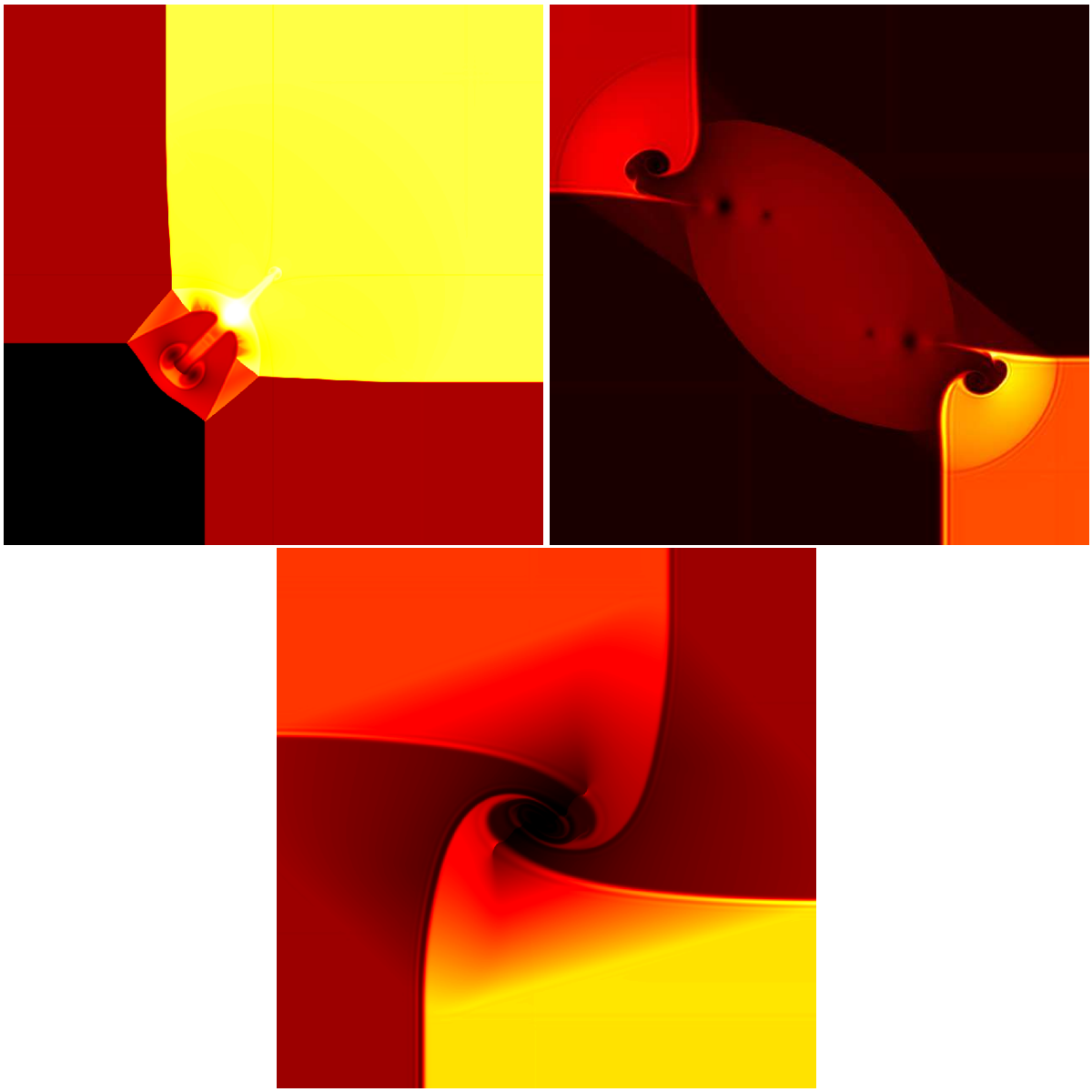


FIGURE 2.16 – Approximation DMGR en densité des problèmes de Riemann 2D sur un maillage cartésien contenant 1.5×10^6 ddl – En haut à gauche : KT3. En haut à droite : KT5. En bas : KT6

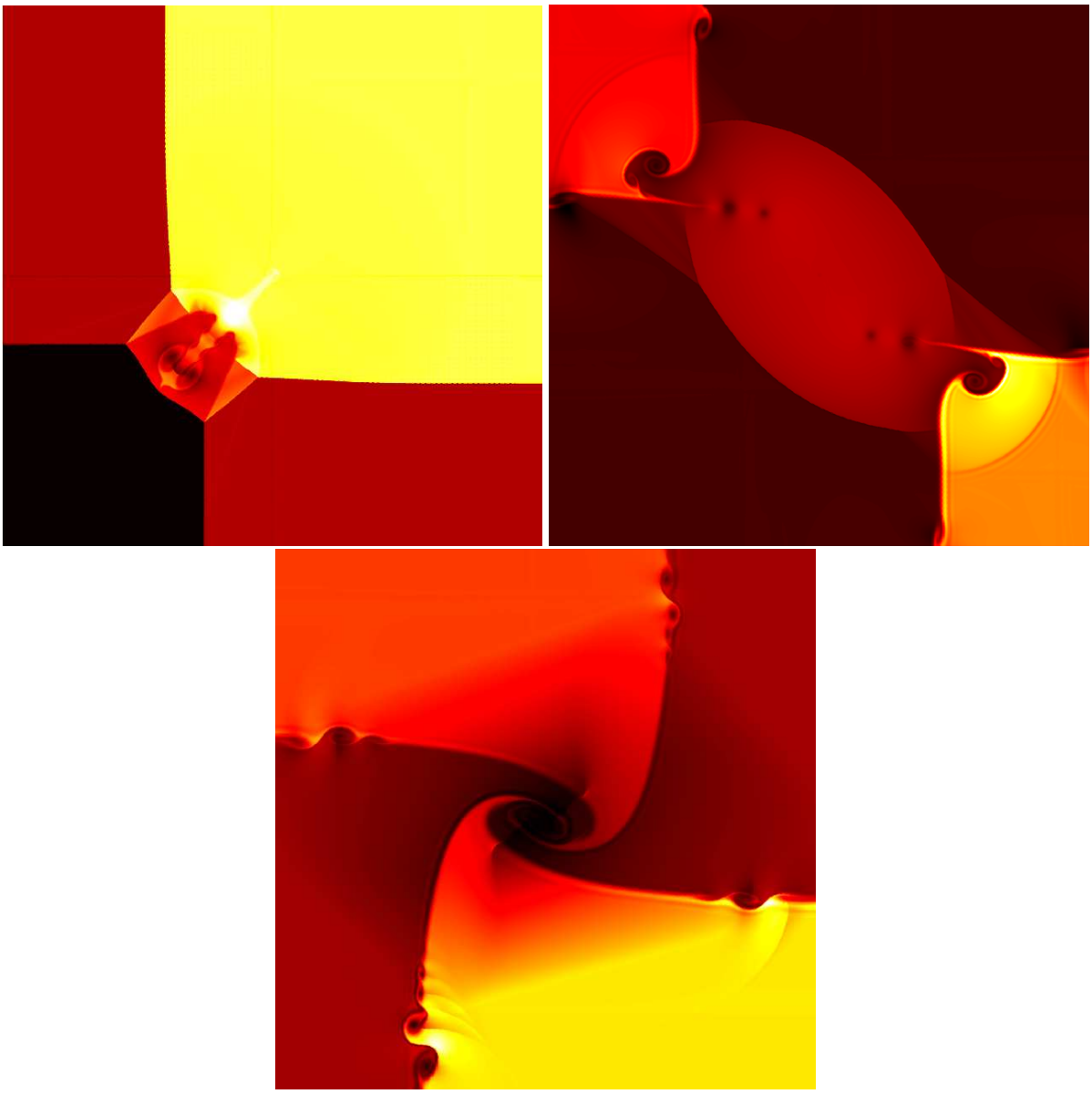


FIGURE 2.17 – Approximation DMGR en densité des problèmes de Riemann 2D sur un maillage triangulaire régulier contenant 1.5×10^6 ddl – En haut à gauche : KT3. En haut à droite : KT5. En bas : KT6

et KT6, de nombreuses instabilités de Kelvin-Helmholtz sont présentes le long des discontinuités de contact.

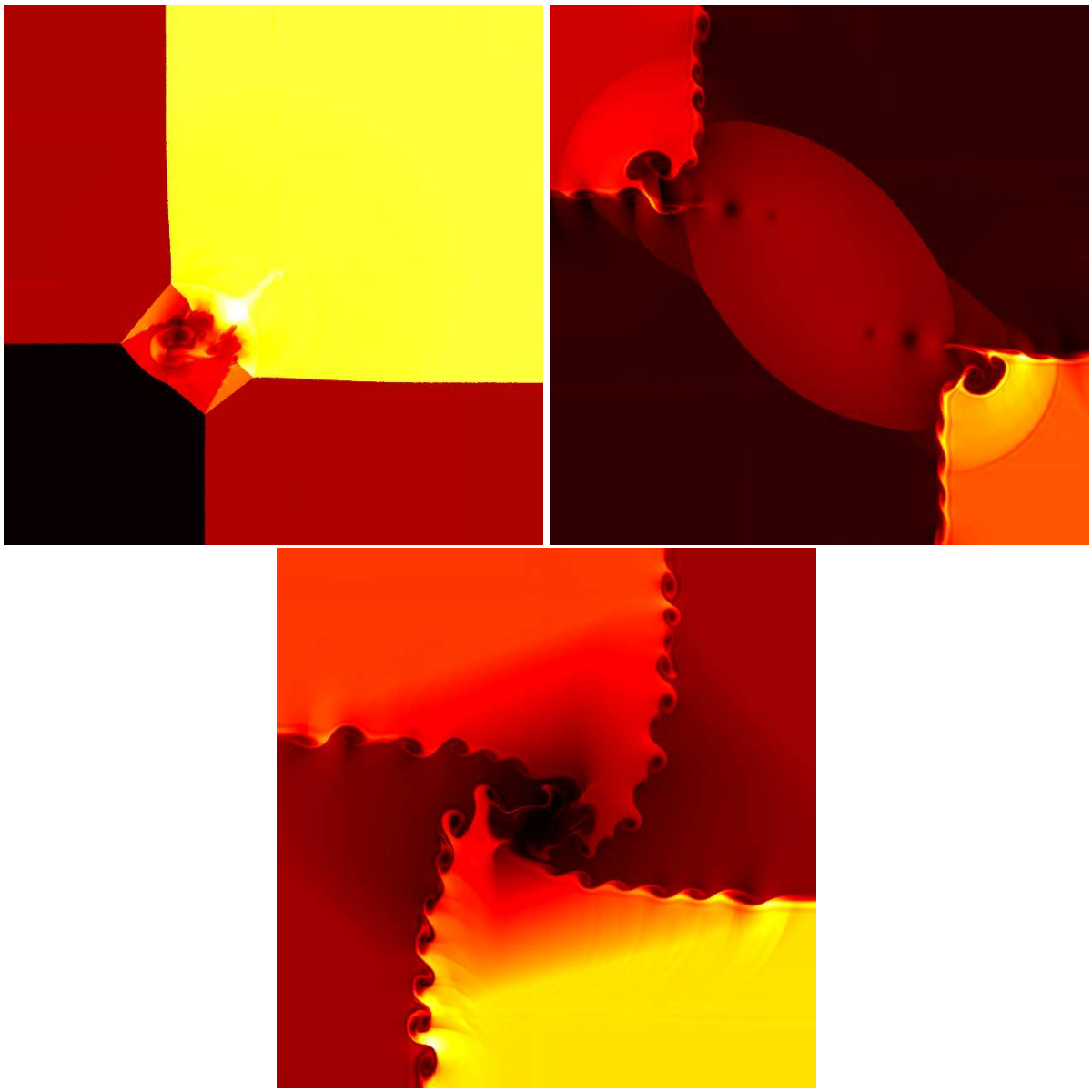


FIGURE 2.18 – Approximation DMGR en densité des problèmes de Riemann 2D sur un maillage non structuré contenant 1.5×10^6 ddl – En haut à gauche : KT3. En haut à droite : KT5. En bas : KT6

2.3.4 Double réflexion de Mach sur une rampe

L'expérience suivante est la double réflexion de Mach sur une rampe qui a été proposée par Woodward et Colella dans [106] (voir aussi [94, 14]). Elle met en jeu des chocs de forte amplitude et une structure très complexe due à des discontinuités de contact. Ce test consiste en l'interaction d'un choc planaire à Mach 10 avec une rampe faisant un angle de 30° avec l'axe des x.

On impose la condition initiale suivante au domaine de calcul entier : $\rho = 1.4$, $(u, v) = (0, 0)$, $p = 1$. On attribue des conditions de bord réfléchissantes aux frontières du bas et du haut. Sur

le bord de gauche, on impose une condition entrante donnée par $\rho = 8$, $(u, v) = (8.25, 0)$ et $p = 116.5$ pour créer le choc à Mach 10 et on impose une condition sortante sur le bord droit pour laisser le gaz s'échapper. On arrête la simulation au temps $t = 0.2$, avant que la structure n'atteigne la frontière du haut.

On effectue le test en utilisant un maillage non structuré contenant environ 3×10^6 ddl. Sur la Figure 2.19, on montre les résultats obtenus à la fois par le schéma d'ordre un, le schéma MUSCL usuel et le schéma DMGR. De plus, on affiche également un zoom sur la zone d'interactions. Afin de rendre la comparaison entre les schémas d'ordre deux pertinente, on utilise pour le schéma MUSCL usuel le limiteur introduit par Barth et Jespersen [6]. Ce limiteur est largement utilisé dans la littérature (par exemple, voir [45, 84, 74]). Le choc principal est connecté à un point de jonction triple par un choc incident et réfléchi. À partir du point triple, une ligne de glissement apparaît à l'intérieur de la structure venant du choc principal pour interagir avec la ligne de glissement.

Comme on s'y attend, le schéma DMGR capture très précisément cette structure d'ondes complexe. Par ailleurs, la ligne de glissement correspond à une discontinuité de contact et l'on peut voir que des instabilités de Kelvin-Helmholtz sont générées. On souligne que les schémas MUSCL classiques ne sont pas capables de capturer des structures aussi petites. Comme mentionné dans [94], l'apparition de si petites structures dans l'écoulement traduit le peu de viscosité numérique inhérente du schéma DMGR.

À propos de l'efficacité de la méthode DMGR, cette expérience numérique a besoin de $62\ 672$ itérations en temps, ce qui représente un pas de temps physique moyen de $\Delta t = 3.2 \times 10^{-6}$. La simulation a été effectuée en utilisant un code parallèle pour un temps CPU de 8h49. En comparaison, le schéma MUSCL standard avec le même nombre de ddl a besoin de 9h02 de temps CPU. Cette expérience numérique a été réalisée avec $78\ 453$ itérations en temps, ce qui correspond à un pas de temps physique moyen de $\Delta t = 2.5 \times 10^{-6}$. On remarque que la condition CFL du schéma DMGR est moins restrictive que celle du schéma MUSCL classique, puisqu'en considérant le même nombre de ddl, notre méthode implique des cellules plus grandes. En effet, pour obtenir 3×10^6 ddl, la méthode DMGR est basée sur un maillage primal composé de 2×10^6 cellules et un maillage dual composé de 1×10^6 cellules, tandis que la technique MUSCL nécessite un maillage de 3×10^6 cellules. Cela explique que les cellules impliquées dans le schéma MUSCL sont plus petites que les cellules impliquées dans le schéma DMGR. Cependant, il semble que le calcul sur chaque cellule soit un peu plus coûteux pour la reconstruction DMGR que pour la reconstruction MUSCL. Pour conclure, les deux méthodes ont des coûts de calcul similaires, mais avec une meilleure précision pour la technique DMGR.

2.3.5 Marche dans un écoulement Mach 3

Le dernier test numérique est dédié à la marche dans un écoulement Mach 3, introduit pour la première fois dans [106]. Le tunnel a pour largeur 1 et pour longueur 3. La marche est située à 0.6 unités de longueur du bord gauche du domaine et a une hauteur de 0.2. Initialement, le tunnel est rempli d'un gaz avec des conditions constantes de $\rho = 1.4$ pour la densité, $p = 1$ pour la pression et $(u, v) = (3, 0)$ pour la vitesse. Ces conditions décrivent un écoulement uniforme à Mach 3. Le long des murs et devant la marche, des conditions de bord réfléchissantes sont utilisées. Les mêmes conditions d'écoulement à Mach 3 sont appliquées en condition de bord entrante sur le bord gauche et une condition de bord sortante est utilisée sur le bord droit.

On utilise des maillages non structurés contenant environ 1.5×10^6 ddl pour réaliser ce cas-test. Les résultats numériques au temps final $t = 4$ sont présentés Figure 2.20, en utilisant le schéma d'ordre un et le schéma DMGR.

Une fois encore, le schéma DMGR donne de très bon résultats. Premièrement, on remarque que des instabilités de Kelvin-Helmholtz apparaissent à partir de la ligne de glissement du

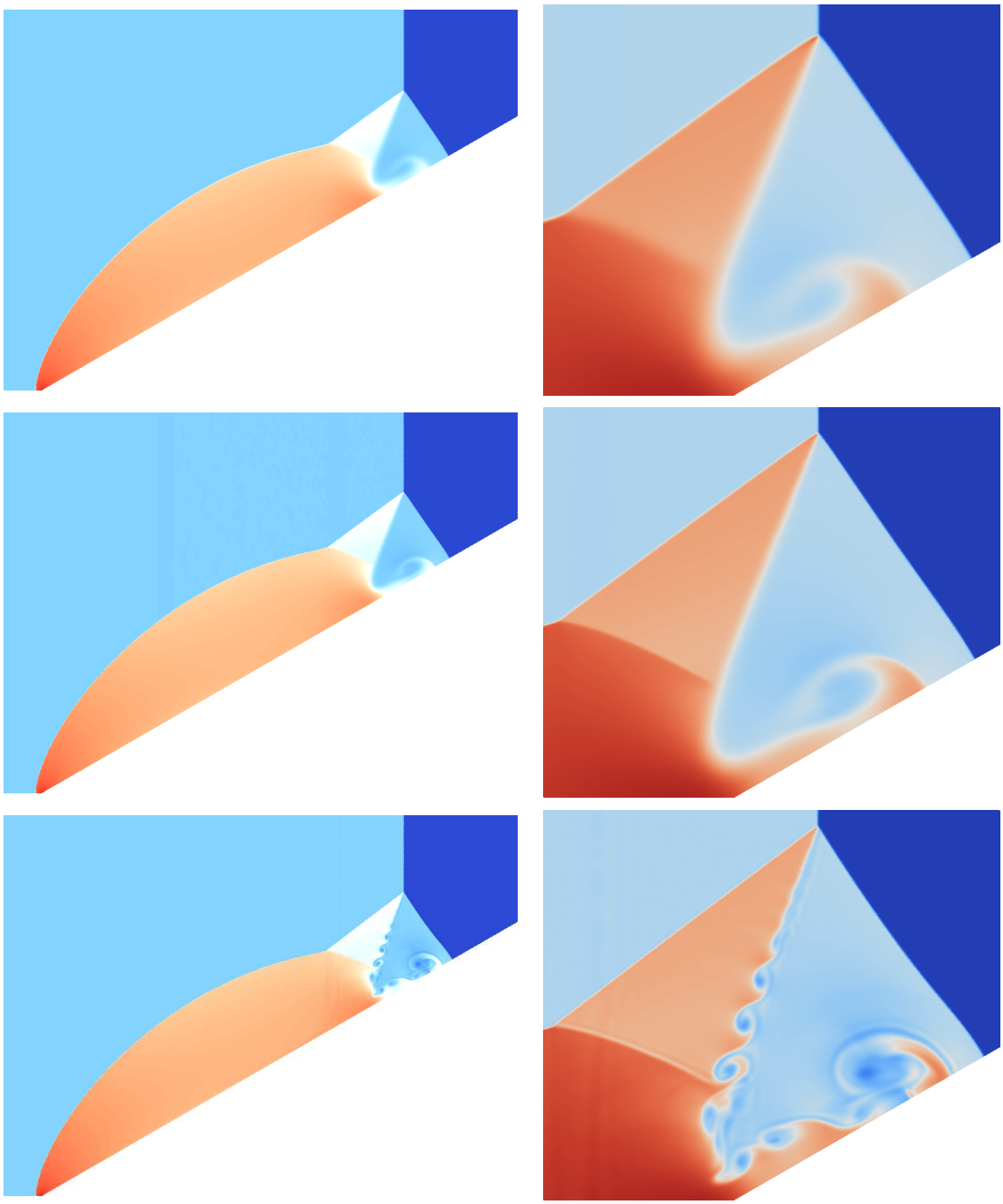


FIGURE 2.19 – Double réflexion de Mach sur une rampe avec 3×10^6 ddl – Gauche : domaine de calcul complet. Droite : zoom sur la zone d'interactions – Haut : approximation d'ordre un. Milieu : Approximation MUSCL classique. Bas : Approximation DMGR

haut. Ensuite, à propos du point triple, l'approximation d'ordre un le situe près de l'interface de la marche alors que, comme signalé dans [34], il devrait être situé exactement sur l'interface de la marche. Comme on peut le voir sur la Figure 2.21, ce problème de localisation est corrigé par le schéma DMGR. Cependant, on remarque que quelques oscillations se développent derrière le premier choc. Cela est probablement dû à l'absence d'un critère du genre TVD dans la technique de reconstruction DMGR présentée.

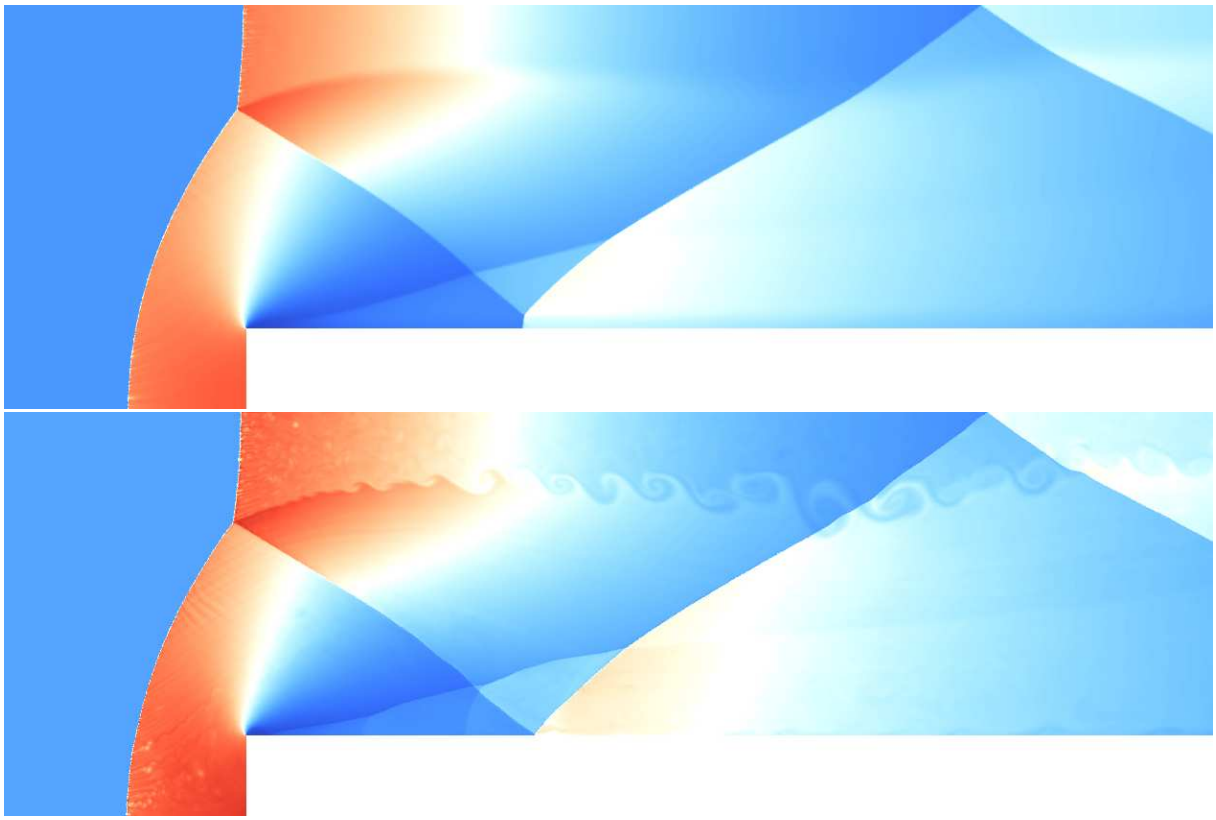


FIGURE 2.20 – Marche dans un écoulement Mach 3 avec 1.5×10^6 ddl – Haut : approximation d'ordre un. Bas : approximation DMGR

Conclusion

Dans ce travail, on a présenté une nouvelle stratégie pour approcher le gradient de la solution numérique. Cette procédure est indépendante de la définition du maillage (structuré ou non structuré) et elle permet d'introduire une amélioration des schémas MUSCL. Pour aborder cette question, on a construit un maillage dual. En considérant simultanément le maillage primal et le maillage dual, on peut reconstruire précisément le gradient de la solution. Ce procédé augmente le nombre de ddl comparé aux schémas MUSCL classiques mais ce nombre reste inférieur au nombre de ddl de la méthode Galerkin discontinue P^1. En effet, sur un maillage triangulaire avec N cellules, le nombre de ddl pour la méthode Galerkin discontinue est $3N$, contre environ $3N/2$ pour le schéma DMGR.

En ce qui concerne les extensions 3D, la difficulté majeure vient de la définition d'un maillage dual adapté. Plusieurs options ont été développées dans le contexte des méthodes DDFV pour approcher les solutions des équations elliptiques/paraboliques. Par exemple, dans [40, 2], la stratégie repose sur une localisation des inconnues aux centres et aux sommets des cellules. Des inconnues supplémentaires peuvent être considérées soit aux centres des faces [66], soit au

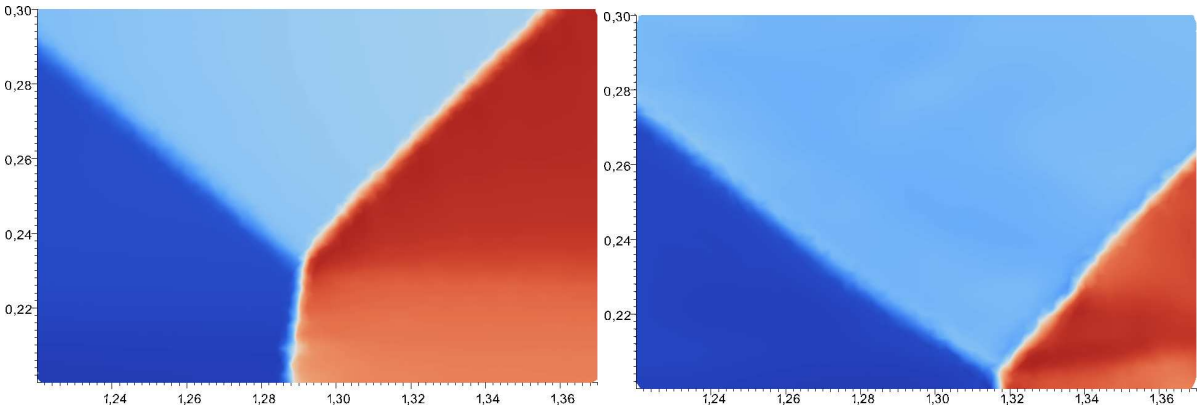


FIGURE 2.21 – Marche dans un écoulement Mach 3 avec 1.5×10^6 ddl – Zoom sur le point triple – Gauche : approximation d'ordre un. Droite : approximation DMGR

milieu des arêtes [39]. Il est possible d'adapter ces techniques DDFV 3D pour généraliser notre technique de reconstruction DMGR à des problèmes 3D.

Enfin, on a prouvé que la stratégie proposée est robuste en introduisant une condition de type CFL convenable. D'après plusieurs travaux récents, la CFL d'ordre deux obtenue n'est certainement pas optimale car elle ne redonne pas la CFL d'ordre un quand la reconstruction MUSCL redonne l'ordre un. Cependant, les simulations numériques réalisées ont montré un très bon comportement des approximations.

3

Schémas MOOD entropiques d'ordre élevé pour les équations d'Euler

Introduction

Le but de ce chapitre est de dériver des schémas d'ordre élevé entropiques pour approcher les solutions faibles des équations d'Euler données par

$$\begin{cases} \partial_t \rho + \partial_x \rho u = 0, \\ \partial_t \rho u + \partial_x (\rho u^2 + p) = 0, \\ \partial_t E + \partial_x (u(E+p)) = 0, \end{cases} \tag{3.1}$$

où la pression suit la loi des gaz parfaits

$$p = (\gamma - 1)\left(E - \frac{1}{2}\rho u^2\right),$$

pour un coefficient adiabatique $\gamma \in]1, 3]$ donné.

Pour alléger les notations, on introduit le vecteur des variables conservatives $w : \mathbb{R} \times \mathbb{R}^+ \to \Omega$ et la fonction flux $f : \Omega \to \mathbb{R}^3$ définis de la façon suivante :

$$w = (\rho, \rho u, E)^T \quad \text{et} \quad f(w) = \left(\rho u, \rho u^2 + p, u(E+p)\right)^T, \tag{3.2}$$

où Ω est l'ensemble convexe des états admissibles donné par

$$\Omega = \left\{ w \in \mathbb{R}^3, \rho > 0, e(w) = E - \frac{1}{2}\rho u^2 > 0 \right\}.$$

La fonction $e : \Omega \to \mathbb{R}^+$ désigne ici l'énergie interne.

Il est bien connu que le système (3.1) est hyperbolique et par conséquent, les solutions peuvent contenir des discontinuités (voir par exemple [55, 80, 93, 44] et les références incluses).

Afin d'exclure les solutions discontinues non physiques, on doit adjoindre au système des inégalités d'entropie (voir [76, 77, 93] pour plus de détails) :

$$\partial_t \rho \mathcal{F}(\ln(s)) + \partial_x \rho \mathcal{F}(\ln(s))u \leq 0, \quad \text{avec } s = \frac{p}{\rho^\gamma}, \tag{3.3}$$

où $\mathcal{F} : \mathbb{R} \to \mathbb{R}$ est une fonction régulière telle que

$$w \mapsto \eta(w) = \rho \mathcal{F}(\ln(s)) \tag{3.4}$$

soit convexe. Pour simplifier les notations, on pose

$$\mathcal{G}(w) = \rho \mathcal{F}(\ln(s))u. \tag{3.5}$$

Comme l'a montré Tadmor [99] (voir aussi [63, 93]), la fonction $\mathcal{F}$ doit vérifier

$$\mathcal{F}'(y) < 0 \quad \text{et} \quad \mathcal{F}'(y) < \gamma \mathcal{F}''(y), \quad \forall y \in \mathbb{R}. \tag{3.6}$$

On s'intéresse maintenant à l'approximation numérique des solutions faibles de (3.1). De nombreuses stratégies peuvent être trouvées dans la littérature en ce qui concerne les méthodes de volumes finis d'ordre un. On renvoie par exemple le lecteur à [56, 80, 20, 100, 59] où les techniques numériques usuelles sont détaillées. On considère une discrétisation uniforme de l'espace en cellules $K_i = [x_{i-1/2}, x_{i+1/2}[$, avec un pas de temps $\Delta x = x_{i+1/2} - x_{i-1/2}$ supposé constant. En notant Δt le pas de temps, une approximation d'ordre un de $w(x_i, t^n + \Delta t)$ est donnée par

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x} \left(F\left(w_i^n, w_{i+1}^n\right) - F\left(w_{i-1}^n, w_i^n\right) \right), \tag{3.7}$$

où le flux numérique $F : \Omega \times \Omega \to \mathbb{R}^3$ est lipschitzien et consistant :

$$F(w, w) = f(w).$$

Le pas de temps doit ici être restreint selon une condition CFL :

$$\frac{\Delta t}{\Delta x} \max_{i \in \mathbb{Z}} |\lambda^{\pm}\left(w_i^n, w_{i+1}^n\right)| \leq \frac{1}{2}, \tag{3.8}$$

où $\lambda^{\pm}\left(w_i^n, w_{i+1}^n\right)$ représentent des vitesses d'onde associées au flux numérique $F\left(w_i^n, w_{i+1}^n\right)$.

La définition du flux numérique peut être complétée par des propriétés additionnelles de robustesse et de stabilité. Concernant la robustesse, la méthode doit préserver à la fois la positivité de la densité et de l'énergie interne. Par conséquent, lorsque la suite $(w_i^n)_{i \in \mathbb{Z}}$ appartient à Ω, le schéma adopté doit vérifier $w_i^{n+1} \in \Omega$.

Dans ce travail, la stabilité du schéma est définie en termes d'entropie. On impose des inégalités d'entropie discrètes afin d'exclure, au niveau discret, les solutions non physiques indésirables. Les inégalités d'entropie discrètes que l'on souhaite obtenir s'écrivent

$$\frac{1}{\Delta t}\left(\eta\left(w_i^{n+1}\right) - \eta\left(w_i^n\right)\right) + \frac{1}{\Delta x}\left(G\left(w_i^n, w_{i+1}^n\right) - G\left(w_{i-1}^n, w_i^n\right)\right) \leq 0, \tag{3.9}$$

où $G : \Omega \times \Omega \to \mathbb{R}$ désigne le flux numérique d'entropie qui doit être consistant :

$$G(w, w) = \mathcal{G}(w).$$

Les méthodes de volumes finis d'ordre un (3.7) les plus classiques vérifient des propriétés de robustesse et/ou de stabilité. Par exemple, on se réfère à [64] pour le schéma HLL, à [52, 53, 29] pour l'extension aux solveurs de Riemann simples, à [101, 7, 20, 100] pour le schéma HLLC,

à [91, 64, 50, 26, 17] pour le schéma de Roe et son extension VFRoe. Cette liste n'est bien sûr pas exhaustive.

De nombreuses stratégies ont été proposées pour augmenter l'ordre de précision des schémas. L'une des plus populaires, qui est adoptée dans ce chapitre, est basée sur une reconstruction convenable de l'inconnue de chaque côté des interfaces situées en $x_{i+1/2}$. En effet, dans (3.7), $F\left(w_i^n, w_{i+1}^n\right)$ est une évaluation à l'ordre un de la fonction flux à l'interface $x_{i+1/2}$. L'extension à l'ordre deux (ou à l'ordre élevé) est obtenue en utilisant une évaluation à l'ordre deux (ou à l'ordre élevé) du flux, donnée par

$$F\left(w_i^{n,+}, w_{i+1}^{n,-}\right),$$

où $w_i^{n,\pm}$ désignent les états reconstruits sur la cellule K_i aux interfaces $x_{i\pm1/2}$. Les techniques pour obtenir $w_i^{n,\pm}$ sont largement étudiées dans la littérature et il est impossible de donner ici une liste complète de tous les articles dédiés à ce sujet. On mentionne simplement la reconstruction MUSCL [102, 86, 73, 21, 37, 80, 72, 12, 33], les approches cinétiques d'ordre deux [86, 73], la reconstruction ENO/WENO [88, 108, 109], la reconstruction PPM [35], la reconstruction MOOD [34, 47], et bien d'autres extensions...

Il s'avère difficile de prouver des propriétés de robustesse et de stabilité pour ces méthodes de volumes finis d'ordre élevé, qui s'écrivent maintenant

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(F\left(w_i^{n,+}, w_{i+1}^{n,-}\right) - F\left(w_{i-1}^{n,+}, w_i^{n,-}\right)\right). \tag{3.10}$$

La propriété de préservation de Ω que doit satisfaire le schéma (3.10) est maintenant bien étudiée. Elle est obtenue en introduisant une procédure de limitation adaptée dans la technique de reconstruction. On renvoie le lecteur à [80, 20] où les reconstructions MUSCL basiques sont présentées et à [12, 14] où la robustesse d'approches plus sophistiquées est étudiée. Dans [88], la robustesse requise est établie dans le cadre de la reconstruction WENO.

On souligne que ces procédures qui permettent d'obtenir la propriété de préservation de Ω impliquent une limitation *a priori*. En d'autres termes, ces limitations sont globales et s'avèrent parfois être trop fortes. Par conséquent, elles peuvent rendre le schéma trop diffusif. Pour corriger cette perte de précision, la méthode MOOD a été récemment présentée dans [34, 47]. Elle consiste à introduire une limitation *a posteriori*. Cette limitation est ainsi seulement locale en espace afin de réduire la viscosité numérique et d'augmenter la précision de la méthode.

Il est beaucoup plus délicat de montrer la stabilité des schémas d'ordre élevé donnés par (3.10). Plusieurs tentatives existent dans la littérature. Une stratégie proposée repose sur le problème de Riemann généralisé (GRP) [8, 22, 23]. Malheureusement, les solutions du GRP pour les équations d'Euler (3.1) sont très dures à obtenir et cela rend peu attrayant le schéma qui en résulte. Dans [37, 36], les auteurs suggèrent d'adopter des nouvelles techniques de projection mais les méthodes numériques obtenues sont, en général, sophistiquées et les extensions à des problèmes plus complexes semblent délicates. Dans le même esprit, on mentionne le travail de Bourdarias et al. [21] mais, comme les auteurs le précisent, le schéma obtenu ne peut pas être facilement implémenté. Plus récemment, dans [12], des inégalités d'entropie discrètes sont obtenues mais pour un opérateur de dérivation discrète en temps particulier (voir aussi [21]). De plus, pour obtenir ces résultats de stabilité , on doit malheureusement mettre en place des procédures de forte limitation ce qui engendre beaucoup de viscosité numérique. Par ailleurs, il n'est pas établi que les opérateurs de dérivation discrète en temps non classiques soient pertinents vis-à-vis du théorème de Lax-Wendroff. En d'autres termes, on ne sait pas montrer (à notre connaissance) que les inégalités d'entropie discrètes considérées convergent, en un sens à préciser, vers les inégalités d'entropie attendues (3.3). Dans [12] (voir aussi [73, 108, 109]), un critère supplémentaire de stabilité est obtenu en appliquant un principe du maximum sur l'entropie [99]. Cependant, cette condition de stabilité est plus faible que les inégalités d'entropie

discrètes usuelles et, par conséquent, on ne considérera pas ce principe du maximum dans ce travail.

Afin de développer des schémas d'ordre élevé robustes et entropiques, on adopte la technique MOOD (Multi-dimensional Optimal Order Detection) [34, 47] (voir aussi [68] pour une méthode proche). Dans la partie suivante, on donne nos principales motivations en étudiant brièvement la convergence des inégalités d'entropie discrètes utilisées dans [21, 12]. Ces motivations sont complétées par des expériences numériques réalisées avec des schémas MUSCL standards sur des maillages très raffinés. Il s'avère que ces approches numériques deviennent instables dès que la taille caractéristique du maillage est suffisamment petite. Par conséquent, on suggère de modifier les schémas MUSCL classiques ou de manière équivalente les reconstructions d'ordre élevé usuelles, en introduisant une limitation *a posteriori* selon une seule inégalité d'entropie. En effet, il n'est pas possible d'évaluer *a posteriori* les inégalités d'entropie en considérant l'espace entier des entropies convexes. Pour cela, dans la Partie 3.2, on prouve qu'à partir d'une seule inégalité d'entropie discrète bien choisie, on peut obtenir toutes les inégalités d'entropie discrètes requises. Munis de ce résultat, on dérive dans la Partie 3.3 le schéma e-MOOD en introduisant dans le schéma d'ordre élevé adopté initialement (ici, le schéma MUSCL pour simplifier) une limitation *a posteriori* basée sur la préservation de l'inégalité d'entropie discrète pertinente. Cette procédure est illustrée par plusieurs résultats numériques dans la Partie 3.4.

3.1 Principales motivations

L'objectif de ce travail est de dériver des schémas d'ordre élevé entropiques pour approcher les solutions faibles de (3.1). Un des principaux problèmes qui se pose lorsque l'on considère des schémas d'ordre élevé est de définir des inégalités d'entropie discrètes pertinentes. On rappelle que les inégalités d'entropie discrètes sont dérivées afin que la solution convergée préserve l'entropie. En d'autres termes, les inégalités d'entropie discrètes considérées doivent converger, au sens du théorème de Lax-Wendroff [78] (voir aussi [51, 80]), vers les inégalités d'entropie continues (3.9).

En fait, plusieurs inégalités d'entropie discrètes d'ordre élevé (MUSCL) ont été dérivées dans la littérature récente (par exemple, voir [21, 12]). Mais il n'est pas clair que ces inégalités discrètes vérifient le comportement de convergence attendu. Le but de cette partie est d'étudier brièvement le comportement des inégalités d'entropies discrètes d'ordre élevé usuelles dans le régime de convergence. À la fin de cette partie, on présente plusieurs expériences numériques qui démontrent l'incapacité des schémas MUSCL à restaurer (3.9), aussi bien avec une discrétisation en temps d'ordre un que d'ordre élevé. On ne justifie pas rigoureusement ces défaillances, mais on donnera quelques arguments.

Dans un premier temps, par soucis de complétude, on rappelle le théorème de Lax-Wendroff pour des schémas d'ordre élevé en espace et en temps. C'est également l'occasion de signaler les inégalités d'entropie discrètes d'ordre élevé que doit vérifier le schéma afin que la solution convergée préserve l'entropie selon (3.9). Ensuite, on passe brièvement en revue les inégalités d'entropie discrètes provenant des schémas d'ordre élevé en espace et en temps. On montrera que ces dernières coïncident avec les inégalités requises par le théorème de Lax-Wendroff à une mesure positive près. Autrement dit, les inégalités d'entropie discrètes usuelles semblent insuffisantes pour assurer que la solution convergée préserve l'entropie. Ce résultat négatif est illustré par plusieurs tests numériques.

3.1.1 Le théorème de Lax-Wendroff pour des schémas d'ordre élevé

On approche ici les solutions faibles d'un système hyperbolique de lois de conservation sous la forme condensée

$$\begin{cases} \partial_t w + \partial_x f(w) = 0, \\ w(x, t = 0) = w_0(x). \end{cases} \tag{3.11}$$

Ce système est complété par les inégalités d'entropie

$$\partial_t \eta(w) + \partial_x \mathcal{G}(w) \leq 0, \tag{3.12}$$

où η est une fonction convexe vérifiant

$$\nabla_w f \nabla_w \eta = \nabla_w \mathcal{G}.$$

Ces notations peuvent en particulier s'appliquer aux équations d'Euler en définissant le vecteur d'état et la fonction flux par (3.2) et en considérant (3.3) comme inégalités d'entropie.

On adopte un schéma en temps de Runge-Kutta à m étapes qui s'écrit

$$\begin{aligned} & w_i^{n,(0)} = w_i^n, \\ & w_i^{n,(\ell)} = w_i^n - \frac{\Delta t}{\Delta x} \sum_{j=0}^{\ell-1} c_{\ell,j} \left(F_{i+1/2}^{n,(j)} - F_{i-1/2}^{n,(j)} \right), \quad \ell = 1, \cdots, m, \\ & w_i^{n+1} = w_i^{n,(m)}. \end{aligned} \tag{3.13}$$

On suppose que les coefficients $c_{\ell,(j)}$ vérifient les propriétés de consistance suivantes

$$c_{\ell,j} \geq 0, \quad \sum_{j=1}^{m-1} c_{m,(j)} = 1. \tag{3.14}$$

Afin de permettre au schéma d'être d'ordre élevé en espace, on considère un flux numérique dépendant d'un large stencil :

$$F_{i+1/2}^{n,(j)} = F^s \left(w_{i-s+1}^{n,(j)}, \cdots, w_{i+s}^{n,(j)} \right), \tag{3.15}$$

où $F^s : \Omega^{2s} \to \mathbb{R}^3$ est continu et consistant :

$$F^s(w, \cdots, w) = f(w).$$

Comme d'habitude, la donnée initiale est approchée par

$$w_i^0 = \frac{1}{\Delta x} \int_{K_i} w_0(x) dx.$$

Pour simplifier ce qui va suivre, on introduit les fonctions constantes par morceaux suivantes :

$$\begin{aligned} w^{\Delta}(x,t) &= w_i^n, \quad \text{pour } (x,t) \in K_i \times [t^n, t^n + \Delta t[, \\ w^{\Delta,(\ell)}(x,t) &= w_i^{n,(\ell)}, \quad \text{pour } (x,t) \in K_i \times [t^n, t^n + \Delta t[. \end{aligned}$$

Théorème 3.1 (Lax-Wendroff)**.** *Supposons que la suite Δx tend vers 0 tout en préservant le ratio $\Delta t / \Delta x$ constant. On suppose que les hypothèses suivantes sont vérifiées :*

- *il existe un compact $K \subset \Omega$ tel que pour tout $0 \leq \ell \leq m$, la fonction $w^{\Delta,(\ell)}$ est à valeurs dans K ;*

- *la suite w^Δ converge dans $L^1_{loc}(\mathbb{R}\times\mathbb{R}^+;\Omega)$ vers une fonction w.*

Alors w est une solution faible de (3.11).

De plus, s'il existe un flux numérique d'entropie $G^s:\Omega^{2s}\to\mathbb{R}$ qui est lipschitzien et consistant :

$$G^s(w,\cdots,w)=\mathcal{G}(w),$$

et qui vérifie l'inégalité d'entropie discrète suivante :

$$\frac{1}{\Delta t}\left(\eta\left(w_i^{n+1}\right)-\eta\left(w_i^n\right)\right)+\frac{1}{\Delta x}\sum_{j=0}^{m-1}c_{m,j}\left(G^{n,(j)}_{i+1/2}-G^{n,(j)}_{i-1/2}\right)\leq 0, \tag{3.16}$$

où

$$G^{n,(j)}_{i+1/2}=G^s\left(w^{n,(j)}_{i-s+1},\cdots,w^{n,(j)}_{i+s}\right),$$

alors w est une solution entropique de (3.11).

Pour pouvoir appliquer ce théorème, le point délicat est l'obtention des inégalités d'entropie discrètes (3.16). De telles inégalités ont été prouvées pour plusieurs schémas d'ordre un en espace de la forme (3.7), comme le schéma de Godunov, les schémas HLL et HLLC, les schémas de relaxation et le schéma d'Osher [55, 21, 20, 11, 13]. Malheureusement, les extensions à l'ordre élevé de ces schémas d'ordre un entropiques ne vérifient pas les inégalités d'entropie discrètes d'ordre élevé données par (3.16). Notre but est maintenant de présenter les inégalités d'entropie discrètes d'ordre élevé vérifiées par les extensions d'ordre élevé en espace et en temps et d'étudier leur comportement dans le régime de convergence.

La preuve du Théorème de Lax-Wendroff (3.1) est classique et plusieurs versions peuvent être trouvées dans [78, 55, 51, 80]. Cependant, à cause de la présence des inégalités d'entropie d'ordre élevé (3.16), à notre connaissance, il n'existe pas de preuve complète dans la littérature. Bien que la preuve soit standard, on la détaille dans l'Appendice A.

3.1.2 Inégalités d'entropie discrètes d'ordre élevé en espace

Pour un schéma d'ordre un donné de la forme (3.7) qui vérifie les inégalités d'entropie discrètes d'ordre un (3.9), de nombreuses méthodes ont été introduites pour augmenter l'ordre de précision (par exemple voir [102, 8, 95, 34, 1]). Dans ce travail, on se restreint aux techniques de reconstruction MUSCL d'ordre deux en espace. On renvoie le lecteur à [107, 17]) où des reconstructions MUSCL d'ordre élevé sont considérées.

On rappelle que la méthode MUSCL est basée sur une reconstruction sur chaque cellule K_i d'états à chaque interface $x_{i\pm1/2}$ de la façon suivante :

$$w_i^{n,\pm}=w_i^n\pm\frac{1}{2}\mu_i^n. \tag{3.17}$$

L'incrément μ_i^n est défini par une fonction limiteur :

$$\mu_i^n=L\left(w_i^n-w_{i-1}^n,w_{i+1}^n-w_i^n\right), \tag{3.18}$$

où $L:\mathbb{R}^3\times\mathbb{R}^3\to\mathbb{R}^3$ est une fonction lipschitzienne qui vérifie

$$L(\mu,\mu)=\mu,\quad\forall\mu\in\mathbb{R}^3, \tag{3.19}$$

$$\exists M>0,\quad\|L(\mu_1,\mu_2)\|\leq M\max\left(\|\mu_1\|,\|\mu_2\|\right),\quad\forall\mu_1,\mu_2\in\mathbb{R}^3. \tag{3.20}$$

Des définitions précises de L sont nombreuses dans la littérature (par exemple, voir [80] et les références incluses). Remarquons dès à présent que les limiteurs usuels (minmod, superbee, MC,...) vérifient les conditions (3.19) et (3.20).

On obtient alors un schéma d'ordre deux de la forme (3.10) à partir d'un schéma d'ordre un (3.7). En ce qui concerne les inégalités d'entropie discrètes d'ordre deux associées à (3.10), plusieurs stratégies ont été récemment proposées. Par exemple, dans [12], on obtient, indépendamment du choix du limiteur L, les inégalités d'entropie discrètes suivantes

$$\frac{1}{\Delta t}\left(\eta\left(w_i^{n+1}\right)-\frac{1}{2}\left(\eta\left(w_i^{n,-}\right)+\eta\left(w_i^{n,+}\right)\right)\right)+\frac{1}{\Delta x}\left(G\left(w_i^{n,+},w_{i+1}^{n,-}\right)-G\left(w_{i-1}^{n,+},w_i^{n,-}\right)\right)\leq 0. \tag{3.21}$$

Un deuxième exemple peut être trouvé dans [21], où une procédure MUSCL spécifique est introduite pour obtenir

$$\frac{1}{\Delta t}\left(\eta\left(w_i^{n+1}\right)-\frac{1}{\Delta x}\int_{K_i}\eta\left(w_i^n+\frac{x}{\Delta x}\mu_i^n\right)dx\right)+\frac{1}{\Delta x}\left(G\left(w_i^{n,+},w_{i+1}^{n,-}\right)-G\left(w_{i-1}^{n,+},w_i^{n,-}\right)\right)\leq 0. \tag{3.22}$$

On remarque immédiatement que les dérivées discrètes en temps apparaissant dans (3.21) et dans (3.22) ne coïncident pas avec celles requises par (3.16). Notre but est maintenant d'illustrer le fait que ces variantes des inégalités d'entropie discrètes ne sont pas efficaces et pas pertinentes pour obtenir une solution convergée entropique.

Dans la suite, il va être utile d'unifier les notations en réécrivant (3.21) et (3.22) de la façon suivante :

$$\frac{1}{\Delta t}\left(\eta\left(w_i^{n+1}\right)-\eta\left(w_i^n\right)\right)+\frac{1}{\Delta x}\left(G_{i+1/2}^n-G_{i-1/2}^n\right)\leq\frac{1}{\Delta t}\left(P_i^n-\eta\left(w_i^n\right)\right), \tag{3.23}$$

où $P_i^n=P\left(w_i^n,\mu_i^n,\Delta x,\eta\right)$ est défini naturellement. En effet, on trouve pour (3.21) :

$$P(w,\mu,\Delta x,\eta)=\frac{\eta(w-\mu/2)+\eta(w+\mu/2)}{2}, \tag{3.24}$$

alors que l'on trouve pour (3.22) :

$$P(w,\mu,\Delta x,\eta)=\frac{1}{\Delta x}\int_{-\Delta x/2}^{\Delta x/2}\eta\left(w+\frac{x}{\Delta x}\mu\right)dx. \tag{3.25}$$

En fait, on peut interpréter P comme une projection pour approcher l'entropie évaluée en w_i^n. On impose l'existence d'une constante positive C telle que

$$0\leq P\left(w,\mu,\Delta x,\eta\right)-\eta(w)\leq C\left\|\nabla^2\eta\right\|\|\mu\|. \tag{3.26}$$

On vérifie aisément que cette propriété est vérifiée à la fois par (3.21) et par (3.22).

On va maintenant voir que $\frac{1}{\Delta t}(P_i^n-\eta(w_i^n))$ converge vers une mesure positive qui rend inadaptées les inégalités d'entropie discrètes (3.23). Afin d'illustrer de manière exhaustive l'échec de (3.23), on propose d'étendre ces inégalités d'entropie discrètes d'ordre élevé en espace en considérant des schémas d'ordre élevé en temps.

3.1.3 Inégalités d'entropie discrètes d'ordre élevé en temps

Pour augmenter l'ordre de précision en temps, on adopte ici le schéma classique de Runge-Kutta donné par (3.13). Afin d'écrire les inégalités d'entropie discrètes associées à (3.13), on considère une reformulation de (3.13) introduite par Shu et Osher ([96, 97]). Cela consiste à

écrire (3.13) comme une combinaison convexe de schémas d'ordre un en temps. On ne détaille pas ici tous les calculs présentés dans [96, 97], mais on rappelle simplement que pour toute famille de paramètres positifs $(\alpha_{\ell,j})_{\substack{1\leq\ell\leq m\\ 0\leq j\leq\ell-1}}$ telle que

$$\sum_{j=0}^{\ell-1}\alpha_{\ell,j}=1,\quad \forall 1\leq\ell\leq m, \tag{3.27}$$

le schéma de Runge-Kutta à m étapes (3.13) peut se réécrire de manière équivalente de la façon suivante :

$$w_i^{n,(0)}=w_i^n, \tag{3.28}$$

$$w_i^{n,(\ell)}=\sum_{j=0}^{\ell-1}\alpha_{\ell,j}\left(w_i^{n,(j)}-\frac{\beta_{\ell,j}}{\alpha_{\ell,j}}\frac{\Delta t}{\Delta x}\left(F_{i+1/2}^{n,(j)}-F_{i-1/2}^{n,(j)}\right)\right), \tag{3.29}$$

$$w_i^{n+1}=w_i^{n,(m)}, \tag{3.30}$$

où les coefficients $\beta_{\ell,j}$ sont donnés par

$$\beta_{\ell,j}=c_{\ell,j}-\sum_{k=j+1}^{\ell-1}\alpha_{\ell,k}c_{k,j}. \tag{3.31}$$

La famille $(\alpha_{\ell,j})_{\substack{1\leq\ell\leq m\\ 0\leq j\leq\ell-1}}$ est choisie de manière à assurer la positivité des coefficients $\beta_{\ell,j}$.

Puisque les paramètres $\alpha_{\ell,j}$ et $\beta_{\ell,j}$ sont supposés positifs, on remarque que les états intermédiaires $w_i^{n,(\ell)}$ sont obtenus par une combinaison convexe de schémas d'ordre un en temps, avec des pas de temps donnés par $\frac{\beta_{\ell,j}}{\alpha_{\ell,j}}\Delta t$.

On établit ensuite les inégalités d'entropie discrètes satisfaites par le schéma d'ordre élevé en temps (3.13) ou de manière équivalente par (3.28). On souligne que le résultat qui va suivre est indépendant de l'ordre en espace du schéma adopté.

Lemme 3.2. *On considère un schéma d'ordre un en temps donné par*

$$w_i^{n+1}=w_i^n-\frac{\Delta t}{\Delta x}\left(F_{i+1/2}^n-F_{i-1/2}^n\right),$$

$$F_{i+1/2}^n=F^s(w_{i-s+1}^n,\cdots,w_{i+s}^n),$$

vérifiant les inégalités d'entropie discrètes suivantes :

$$\frac{1}{\Delta t}\left(\eta\left(w_i^{n+1}\right)-\eta\left(w_i^n\right)\right)+\frac{1}{\Delta x}\left(G_{i+1/2}^n-G_{i-1/2}^n\right)\leq\delta\left(w_i^n\right), \tag{3.32}$$

où $G_{i+1/2}^n=G^s(w_{i-s+1}^n,\cdots,w_{i+s}^n)$ et $\delta\left(w_i^n\right)$ est une perturbation positive.

Supposons que les paramètres $\alpha_{\ell,j}>0$ sont choisis de manière à ce que les paramètres $\beta_{\ell,j}$ soient positifs. Alors le schéma (3.13) vérifie les inégalités d'entropie discrètes suivantes :

$$\frac{1}{\Delta t}\left(\eta\left(w_i^{n+1}\right)-\eta\left(w_i^n\right)\right)+\sum_{j=0}^{m-1}c_{m,j}\frac{1}{\Delta x}\left(G_{i+1/2}^{n,(j)}-G_{i-1/2}^{n,(j)}\right)\leq\sum_{j=0}^{m-1}\alpha_{m,j}\delta\left(w_i^{n,(j)}\right). \tag{3.33}$$

Avant de prouver ce résultat, on précise le rôle joué par la perturbation $\delta\left(w_i^n\right)$ qui est centrée en w_i^n mais qui peut dépendre d'autres états. Lorsqu'un schéma standard d'ordre un en temps et en espace de la forme (3.7) vérifie les inégalités d'entropie discrètes (3.9), la perturbation

disparaît, c'est-à-dire $\delta\left(w_i^n\right) = 0$. Dans ce cas, les inégalités (3.33) coïncident exactement avec les inégalités d'entropie discrètes requises (3.16). Plus généralement, cela signifie que dès que le schéma d'ordre un est entropique (avec $\delta\left(w_i^n\right) = 0$), alors le schéma d'ordre élevé en temps de Runge-Kutta associé est également entropique. La situation est différente lorsque $\delta\left(w_i^n\right) \neq 0$, et le second membre dans (3.33) doit être étudié avec attention.

Démonstration. On introduit les états intermédiaires suivants :

$$\widetilde{w}_i^{n,(j)} = w_i^{n,(j)} - \frac{\beta_{\ell,j}}{\alpha_{\ell,j}} \frac{\Delta t}{\Delta x} \left(F_{i+1/2}^{n,(j)} - F_{i-1/2}^{n,(j)} \right).$$

Puisque $\frac{\beta_{\ell,j}}{\alpha_{\ell,j}} \geq 0$, l'état $\widetilde{w}_i^{n,(j)}$ est simplement une mise à jour par un schéma d'ordre un en temps. D'après (3.32), les états intermédiaires $\widetilde{w}_i^{n,(j)}$ vérifient donc des inégalités d'entropie discrètes données par

$$\frac{1}{\Delta t}\left(\eta\left(\widetilde{w}_i^{n,(j)}\right) - \eta\left(w_i^{n,(j)}\right)\right) + \frac{\beta_{\ell,j}}{\alpha_{\ell,j}\Delta x}\left(G_{i+1/2}^{n,(j)} - G_{i-1/2}^{n,(j)}\right) \leq \delta\left(w_i^{n,(j)}\right).$$

En utilisant la formulation équivalente (3.28), on remarque que

$$w_i^{n,(\ell)} = \sum_{j=1}^{\ell-1} \alpha_{\ell,j} \widetilde{w}_i^{n,(j)}.$$

Ensuite, puisque η est une fonction convexe, on obtient

$$\eta\left(w_i^{n,(\ell)}\right) \leq \sum_{j=0}^{\ell-1} \alpha_{\ell,j} \eta\left(\widetilde{w}_i^{n,(j)}\right),$$

dont on déduit immédiatement

$$\eta\left(w_i^{n,(\ell)}\right) \leq \sum_{j=0}^{\ell-1} \left(\alpha_{\ell,j} \eta\left(w_i^{n,(j)}\right) - \beta_{\ell,j} \frac{\Delta t}{\Delta x}\left(G_{i+1/2}^{n,(j)} - G_{i-1/2}^{n,(j)}\right)\right) + \Delta t \sum_{j=0}^{\ell-1} \alpha_{\ell,j} \delta\left(w_i^{n,(j)}\right). \quad (3.34)$$

On établit maintenant par récurrence l'inégalité suivante :

$$\eta\left(w_i^{n,(\ell)}\right) \leq \eta\left(w_i^n\right) - \frac{\Delta t}{\Delta x} \sum_{j=0}^{\ell-1} c_{\ell,j}\left(G_{i+1/2}^{n,(j)} - G_{i-1/2}^{n,(j)}\right) + \Delta t \sum_{j=0}^{\ell-1} \alpha_{\ell,j} \delta\left(w_i^{n,(j)}\right), \quad 1 \leq \ell \leq m. \quad (3.35)$$

Supposons dans un premier temps que $\ell = 1$. En utilisant (3.27) et (3.31), on trouve alors aisément $\alpha_{1,0} = 1$ et $c_{1,0} = \beta_{1,0}$. On déduit alors (3.35) de (3.34).

On suppose ensuite que (3.35) est vérifiée pour tout j tel que $1 \leq j \leq \ell - 1$ et on établit l'inégalité pour ℓ. En remplaçant $\eta\left(w_i^{n,(j)}\right)$ par l'estimation (3.35) dans (3.34), on obtient

$$\begin{aligned}
\eta\left(w_i^{n,(\ell)}\right) \leq \sum_{j=0}^{\ell-1} \Bigg(& \alpha_{\ell,j} \left(\eta\left(w_i^n\right) - \frac{\Delta t}{\Delta x} \sum_{k=0}^{j-1} c_{j,k}\left(G_{i+1/2}^{n,(k)} - G_{i-1/2}^{n,(k)}\right)\right) \\
& - \beta_{\ell,j} \frac{\Delta t}{\Delta x}\left(G_{i+1/2}^{n,(j)} - G_{i-1/2}^{n,(j)}\right)\Bigg) + \Delta t \sum_{j=0}^{\ell-1} \alpha_{\ell,j} \delta\left(w_i^{n,(j)}\right), \\
\leq \eta\left(w_i^n\right) - \frac{\Delta t}{\Delta x} \sum_{j=0}^{\ell-1} & \left(\beta_{\ell,j} + \sum_{k=j+1}^{\ell-1} \alpha_{\ell,k} c_{k,j} \right)\left(G_{i+1/2}^{n,(j)} - G_{i-1/2}^{n,(j)}\right) + \Delta t \sum_{j=0}^{\ell-1} \alpha_{\ell,j} \delta\left(w_i^{n,(j)}\right)
\end{aligned}$$

et (3.35) est vérifiée. Puisque $w_i^{n+1} = w_i^{n,(m)}$ par (3.31), la preuve est achevée. □

Grâce à ce résultat, on peut maintenant présenter les inégalités d'entropie discrètes associées avec les schémas d'ordre élevé à la fois en temps et en espace (3.13)–(3.15). En effet, puisque le schéma d'ordre un en temps vérifie les inégalités d'entropie discrètes (3.23), on obtient une perturbation d'entropie discrète donnée par

$$\delta\left(w_i^n\right) = \frac{1}{\Delta t}\left(P_i^n - \eta\left(w_i^n\right)\right).$$

Par conséquent, on va examiner les inégalités d'entropie discrètes d'ordre élevé données par

$$\frac{1}{\Delta t}\left(\eta\left(w_i^{n+1}\right) - \eta\left(w_i^n\right)\right) + \sum_{j=0}^{m-1} \frac{c_{m,j}}{\Delta x}\left(G_{i+1/2}^{n,(j)} - G_{i-1/2}^{n,(j)}\right) \leq \sum_{j=0}^{m-1} \alpha_{m,j} \frac{1}{\Delta t}\left(P_i^{n,(j)} - \eta\left(w_i^{n,(j)}\right)\right). \tag{3.36}$$

Sous les hypothèses du Théorème 3.1, on obtient facilement la convergence faible du membre de gauche vers $\partial_t \eta(w) + \partial_x \mathcal{G}(w)$. Concernant le second membre, on pose

$$a^{\Delta}(x,t) = \sum_{j=0}^{m-1} \alpha_{m,j} \frac{1}{\Delta t}\left(P_i^{n,(j)} - \eta\left(w_i^{n,(j)}\right)\right), \quad (x,t) \in K_i \times [t^n, t^{n+1}[.$$

On introduit maintenant la mesure positive δ définie comme la limite faible-étoile de la suite a^{Δ}. On en déduit que dans la limite de Δx et Δt tendant vers zéro avec un ratio $\Delta t/\Delta x$ constant, l'inégalité (3.36) s'écrit

$$\partial_t \eta(w) + \partial_x \mathcal{G}(w) \leq \delta.$$

On suggère de comparer la mesure δ à la mesure de dissipation d'entropie β définie comme la limite faible-étoile de la suite suivante :

$$b^{\Delta}(x,t) = \sum_{j=0}^{m-1} \alpha_{m,j} \frac{1}{\Delta x}\left\|w_i^{n,(j)} - w_{i-1}^{n,(j)}\right\|^2, \quad (x,t) \in K_i \times [t^n, t^{n+1}[.$$

La mesure de dissipation d'entropie β a été étudiée par Hou et LeFloch [67] (voir ausi DiPerna [48]) dans le cas scalaire et avec un schéma d'ordre un en temps. Ils ont conjecturé que cette mesure est concentrée sur les courbes de discontinuité de w.

Dans le résultat qui suit, on établit que les mesures δ et β ont le même comportement.

Théorème 3.3. *La mesure δ est absolument continue par rapport à la mesure de dissipation d'entropie β.*

Démonstration. Soit ϕ une fonction test positive à support compact K et posons $\phi_i^n = \phi(x_i, t^n)$. Puisque P vérifie la propriété (3.26), on a

$$\sum_{i,n}\left(P_i^{n,(j)} - \eta\left(w_i^{n,(j)}\right)\right)\phi_i^n \Delta t \leq C\|\nabla^2\eta\| \sum_{i,n} \|\mu_i^{n,(j)}\|^2 \phi_i^n \Delta t,$$

où $\nabla^2\eta$ est bornée sur K et $\mu_i^{n,(j)}$ est l'incrément de la reconstruction défini par (3.18).

En utilisant (3.18) et (3.20), on obtient

$$\begin{aligned}\sum_{i,n}\left(P_i^{n,(j)} - \eta\left(w_i^{n,(j)}\right)\right)\phi_i^n \Delta t &\leq O(1)\sum_{i,n}\left(\|w_i^{n,(j)} - w_{i-1}^{n,(j)}\|^2 + \|w_{i+1}^{n,(j)} - w_i^{n,(j)}\|^2\right)\phi_i^n \Delta t, \\ &\leq O(1)\sum_{i,n}\|w_i^{n,(j)} - w_{i-1}^{n,(j)}\|^2\left(\phi_i^n + \phi_{i-1}^n\right)\Delta t.\end{aligned}$$

Puisque le ratio $\Delta t/\Delta x$ reste constant, on en déduit

$$\sum_{i,n}\sum_{j=0}^{m-1}\alpha_{m,j}\left(P_i^{n,(j)}-\eta\left(w_i^{n,(j)}\right)\right)\phi_i^n\Delta t\leq O(1)\sum_{i,n}\sum_{j=0}^{m-1}\alpha_{m,j}\|w_i^{n,(j)}-w_{i-1}^{n,(j)}\|^2\left(\phi_i^n+\phi_{i-1}^n\right)\Delta x.$$

En passant à la limite, on obtient

$$\int\phi d\delta\leq O(1)\int\phi d\beta,$$

ce qui conclut la preuve. □

Pour terminer cette partie, on souligne encore une fois que les inégalités d'entropie discrètes (3.36) ne sont pas suffisantes pour assurer la stabilité entropique requise.

3.1.4 Résultats numériques

On va maintenant illustrer numériquement les résultats précédents. Plus précisément, le but est d'évaluer numériquement la mesure δ introduite dans la partie précédente. D'après le travail de Hou et LeFloch [67], on s'attend à ce que cette mesure s'annule partout où la solution est continue. Inversement, quand la solution est discontinue, l'évaluation de δ doit donner $\delta>0$.

Toutes les expériences numériques présentées sont basées sur la même stratégie. On adopte le flux numérique d'ordre un $F(w_L,w_R)$ donné par le schéma HLLC [101, 100]. L'avantage de ce choix est que le schéma HLLC est connu pour être robuste et vérifier les inégalités d'entropie discrètes (3.32) (voir [20, 11, 14, 30]). L'ordre deux en espace est obtenu par une reconstruction MUSCL (3.17)–(3.18). Les fonctions L que l'on utilise sont le limiteur minmod, le limiteur de van Albada 1, le limiteur de van Leer, le limiteur monotonized central-difference (MC) et le limiteur Superbee (voir Partie 1.4 ou [80] où tous ces limiteurs sont détaillés). Concernant la discrétisation en temps, on utilisera l'ordre un et l'ordre deux pour comparer les résultats.

D'après [80, 20, 12], on restreint le pas de temps Δt suivant la condition CFL

$$\frac{\Delta t}{\Delta x}\max_{i\in\mathbb{Z}}\left(\left|\lambda^{\pm}\left(w_{i+1/2}^{n,-},w_{i+1/2}^{n,+}\right)\right|,\left|\lambda^{\pm}\left(w_{i-1/2}^{n,+},w_{i+1/2}^{n,-}\right)\right|\right)\leq\frac{1}{4}.$$

Selon [12], cette restriction du pas de temps rend le schéma considéré robuste et lui permet de vérifier les inégalités d'entropie discrètes (3.23).

La pertinence des schémas considérés est évaluée en calculant l'erreur L^1 en densité au temps final :

$$E^{\Delta}=\Delta x\sum_{i\in\mathbb{Z}}\left|\rho_i^N-\rho^{ex}(x_i,T)\right|,$$

où $w^{ex}:\mathbb{R}\times\mathbb{R}^+\to\Omega$ est la solution exacte. De plus, pour évaluer la mesure δ, on calcule la masse totale de a^{Δ} :

$$I^{\Delta}=\Delta x\Delta t\sum_{n=0}^{N}\sum_{i\in\mathbb{Z}}a^{\Delta}(x_i,t^n).$$

On réalise deux expériences numériques. Les deux sont consacrées à l'approximation de la solution d'un problème de Riemann, c'est-à-dire que la donnée initiale est constituée de deux états constants séparés par une discontinuité située en $x=0$:

$$w_0(x)=\begin{cases}w_L & \text{si } x<0,\\ w_R & \text{si } x>0.\end{cases}\tag{3.37}$$

Dans le premier cas-test, les états gauche et droit sont donnés par

$$\begin{array}{ll} \rho_L = 1, & \rho_R = 0.1989, \\ u_L = -1, & u_R = 1, \\ p_L = 1.5, & p_R = 0.1564, \end{array} \tag{3.38}$$

pour que la solution exacte soit uniquement constituée d'une 1–détente continue (voir Figure 3.1).

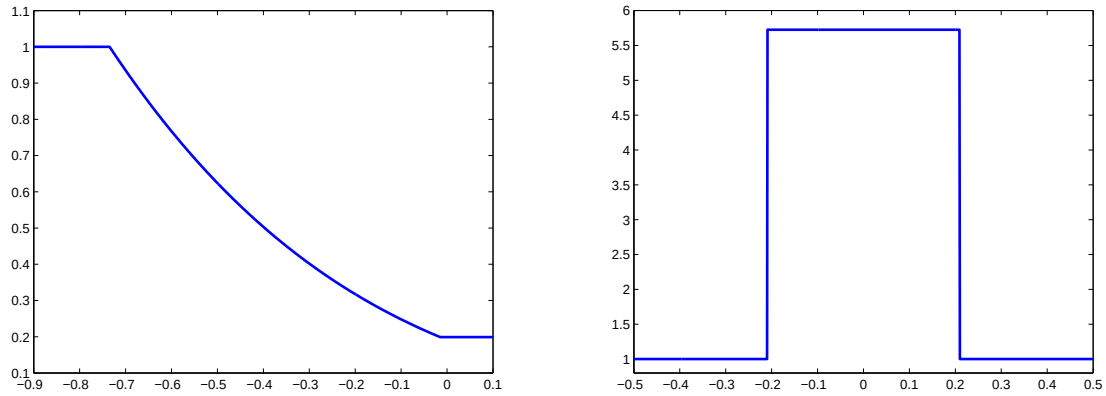


FIGURE 3.1 – Solution des problèmes de Riemann. Gauche : 1–détente. Droite : double choc

	minmod		van Albada 1		van Leer		MC		Superbee	
Δx	E^Δ	I^Δ	E^Δ	I^Δ	E^Δ	I^Δ	E^Δ	I^Δ	E^Δ	I^Δ
8.00E-3	7.54E-3	1.48E-2	9.98E-3	1.79E-2	1.09E-2	2.00E-2	1.34E-2	2.27E-2	3.42E-2	4.39E-2
4.00E-3	5.19E-3	9.36E-3	6.64E-3	1.12E-2	7.40E-3	1.25E-2	1.06E-2	1.52E-2	3.00E-2	3.22E-2
2.00E-3	3.12E-3	5.70E-3	4.17E-3	6.81E-3	5.00E-3	7.81E-3	9.38E-3	1.01E-2	2.35E-2	2.35E-2
1.00E-3	1.84E-3	3.38E-3	2.72E-3	4.10E-3	3.70E-3	4.96E-3	8.77E-3	8.53E-3	1.88E-2	1.84E-2
5.00E-4	1.07E-3	1.96E-3	2.01E-3	2.51E-3	2.82E-3	3.24E-3	8.58E-3	7.12E-3	2.48E-2	3.03E-2
2.50E-4	6.13E-4	1.12E-3	1.37E-3	1.55E-3	2.35E-3	2.26E-3	9.28E-3	6.38E-3	9.98E-2	7.50E-1
1.25E-4	3.63E-4	6.30E-4	1.19E-3	1.02E-3	2.44E-3	1.74E-3	3.84E-2	8.04E-2	1.21E-1	1.48E+1
6.25E-5	2.41E-4	3.53E-4	1.24E-3	7.54E-4	3.51E-3	2.18E-3	5.34E-2	1.42E-0		
3.12E-5	1.80E-4	1.99E-4	1.26E-3	6.18E-4						

Tableau 3.1 – Erreur L^1 et erreur entropique pour la 1–détente en utilisant un schéma d'ordre un en temps

Dans le Tableau 3.1, on donne l'évaluation de E^Δ et I^Δ obtenues avec un schéma d'ordre un en temps et différents limiteurs. Tout d'abord, on remarque que les limiteurs van Leer, MC et Superbee ne sont pas stables et des explosions numériques apparaissent avec un maillage très raffiné. Pour préciser le comportement de ces limiteurs, on montre Figure 3.3 la solution obtenue par le limiteur Superbee avec 1000 et 2000 cellules (respectivement $\Delta x = 10^{-3}$ et $\Delta x = 5 \times 10^{-4}$). Avec 1000 cellules, l'explosion numérique n'a pas encore eu lieu et on observe que le schéma fait apparaître un choc non entropique à l'intérieur de la détente. Cela confirme que le schéma MUSCL ne vérifie pas la stabilité requise. Avec 2000 cellules, le choc non entropique est toujours présent, mais l'explosion numérique commence à être importante. Si l'on augmente encore le nombre de cellules, le choc non entropique ne sera plus visible à cause de l'explosion numérique.

Le comportement des limiteurs minmod et van Albada 1 est meilleur car les deux schémas semblent converger, puisque E^Δ tend vers zéro quand Δx tend vers zéro. Pour ces deux limiteurs, la quantité I^Δ tend également vers zéro et donc la mesure δ également, ce qui est en accord avec la conjecture de Hou et LeFloch [67], puisque la solution convergée est continue. La Figure 3.2 illustre les résultats présentés dans le Tableau 3.1.

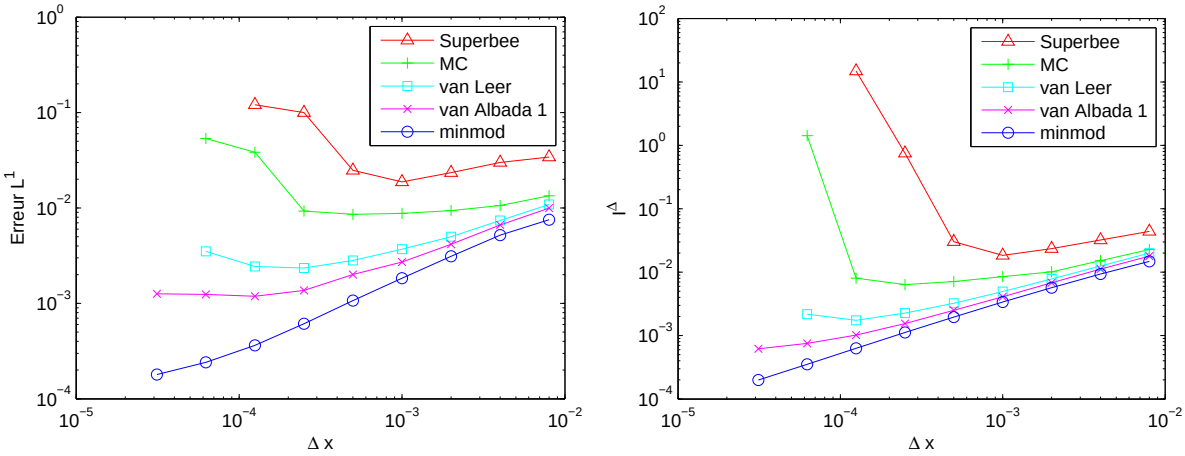


FIGURE 3.2 – 1–détente avec un schéma d'ordre un en temps. Gauche : erreur L^1. Droite : erreur entropique I^Δ

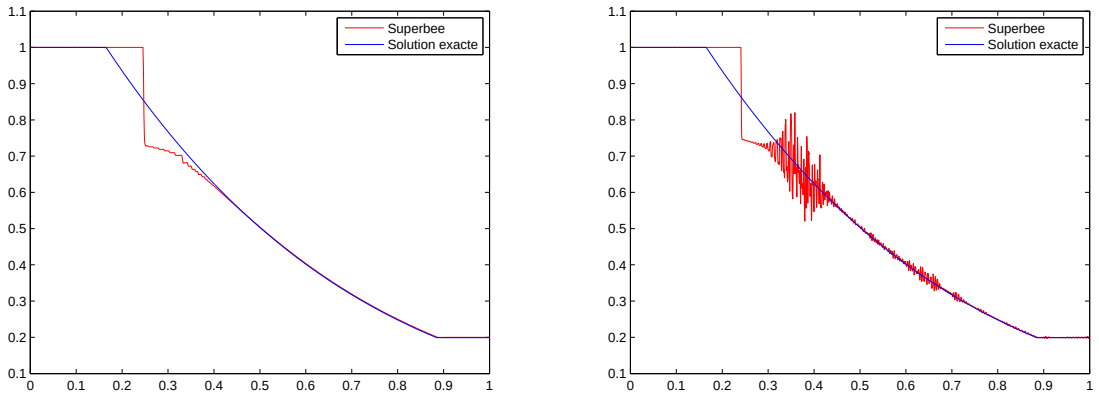


FIGURE 3.3 – 1–détente : résultat obtenus par le limiteur Superbee. Gauche : $\Delta x = 10^{-3}$. Droite : $\Delta x = 5 \times 10^{-4}$

Le Tableau 3.2 et le Figure 3.4 sont ensuite dédiés aux résultats obtenus avec un schéma d'ordre deux en temps. Mis à part le limiteur Superbee, tous les schémas semblent converger, tout comme la mesure δ qui tend vers zéro.

	minmod		van Albada 1		van Leer		MC		Superbee	
Δx	E^Δ	I^Δ	E^Δ	I^Δ	E^Δ	I^Δ	E^Δ	I^Δ	E^Δ	I^Δ
8.00E-3	6.82E-3	1.44E-2	5.41E-3	1.67E-2	4.55E-3	1.81E-2	3.87E-3	1.93E-2	4.93E-3	2.32E-2
4.00E-3	5.38E-3	9.08E-3	4.37E-3	1.03E-2	3.86E-3	1.10E-2	3.46E-3	1.16E-2	2.37E-3	1.38E-2
2.00E-3	2.72E-3	5.51E-3	2.19E-3	6.17E-3	1.94E-3	6.54E-3	1.74E-3	6.85E-3	1.16E-3	8.01E-3
1.00E-3	1.36E-3	3.25E-3	1.10E-3	3.60E-3	9.68E-4	3.79E-3	8.69E-4	3.96E-3	5.74E-4	4.61E-3
5.00E-4	6.82E-4	1.88E-3	5.49E-4	2.06E-3	4.84E-4	2.16E-3	4.35E-4	2.25E-3	2.89E-4	2.63E-3
2.50E-4	3.41E-4	1.07E-3	2.74E-4	1.16E-3	2.42E-4	1.16E-3	2.17E-4	1.27E-3	1.48E-4	1.51E-3
1.25E-4	1.71E-4	5.98E-4	1.37E-4	6.50E-4	1.21E-4	6.77E-4	1.09E-4	7.03E-4	7.68E-5	8.94E-4
6.25E-5	8.54E-5	3.31E-4	6.86E-5	3.59E-4	6.05E-5	3.73E-4	5.43E-5	3.88E-4	4.12E-5	5.66E-4
3.12E-5	4.27E-5	1.82E-4	3.43E-5	1.97E-4	3.03E-5	2.04E-4	2.72E-5	2.12E-4	4.92E-5	5.16E-4
1.56E-5							1.36E-5	1.15E-4	8.39E-3	1.03E-0

Tableau 3.2 – Erreur L^1 et erreur entropique pour la 1–détente en utilisant un schéma d'ordre deux en temps

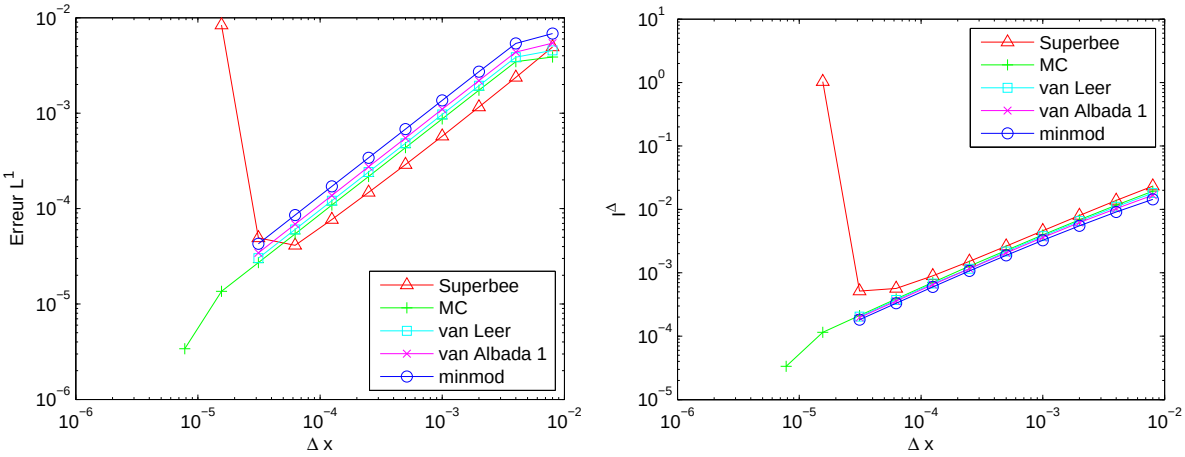


FIGURE 3.4 – 1–détente avec un schéma d'ordre deux en temps. Gauche : erreur L^1. Droite : erreur entropique I^Δ

La seconde expérience numérique que l'on propose est dédiée à l'approximation de solutions avec des chocs. On considère à nouveau un problème de Riemann où les états gauche et droits sont définis par

$$\begin{aligned} \rho_L &= 1, & \rho_R &= 1, \\ u_L &= 10, & u_R &= -10, \\ p_L &= 1, & p_R &= 1, \end{aligned} \tag{3.39}$$

pour obtenir une solution exacte composée de deux ondes de chocs se propageant avec des vitesses opposées (voir Figure 3.1).

Les résultats numériques obtenus en utilisant un schéma d'ordre un en temps sont présentés dans le Tableau 3.3 et la Figure 3.5. On remarque que les limiteurs van Leer, MC et Superbee entraînent une explosion numérique. En fait, il semble que les limiteurs minmod et van Albada 1 soient également instables mais l'explosion nécessite un maillage extrêmement raffiné. Par ailleurs, on souligne que la suite I^Δ semble converger vers une valeur positive avant l'explosion numérique. Cela confirme que la masse totale de la mesure δ est strictement positive.

Enfin, on présente dans le Tableau 3.4 et la Figure 3.6 le comportement de l'erreur L^1 et de la grandeur I^Δ en utilisant une discrétisation en temps Runge-Kutta d'ordre deux. Seul le limiteur Superbee engendre une explosion numérique alors que tous les autres schémas convergent (ou

	minmod		van Albada 1		van Leer		MC		Superbee	
Δx	E^Δ	I^Δ	E^Δ	I^Δ	E^Δ	I^Δ	E^Δ	I^Δ	E^Δ	I^Δ
8.00E-3	2.85E-2	1.19093	2.84E-2	1.50166	2.78E-2	3.69771	2.74E-2	6.60340	2.81E-2	7.39416
4.00E-3	8.79E-3	1.19348	8.42E-3	1.50362	7.99E-3	3.69027	8.36E-3	6.57974	9.96E-3	7.47961
2.00E-3	3.47E-3	1.19438	3.33E-3	1.50505	3.34E-3	3.67345	3.42E-3	6.63387	5.35E-2	7.59467
1.00E-3	1.18E-3	1.19488	1.14E-3	1.50596	1.36E-3	3.66646	1.59E-3	6.55043	2.06E-2	7.95907
5.00E-4	1.74E-3	1.19502	1.70E-3	1.50636	1.86E-3	3.66594	2.14E-3	6.61186	4.54E-2	10.0619
2.50E-4	1.07E-3	1.19521	1.05E-3	1.50650	1.29E-3	3.66420	3.97E-3	6.62561		
1.25E-4	7.62E-4	1.19529	7.53E-4	1.50648	1.04E-3	3.66326	1.89E-2	7.68269		
6.25E-5	1.56E-4	1.19531	1.69E-4	1.50625	7.78E-3	3.93815	2.18E-2	10.3855		
3.12E-5	9.59E-5	1.19545	4.23E-4	1.50600	1.31E-2	5.40179				
1.56E-5	1.46E-3	1.21136								

Tableau 3.3 – Erreur L^1 et erreur entropique pour le double choc en utilisant un schéma d'ordre un en temps

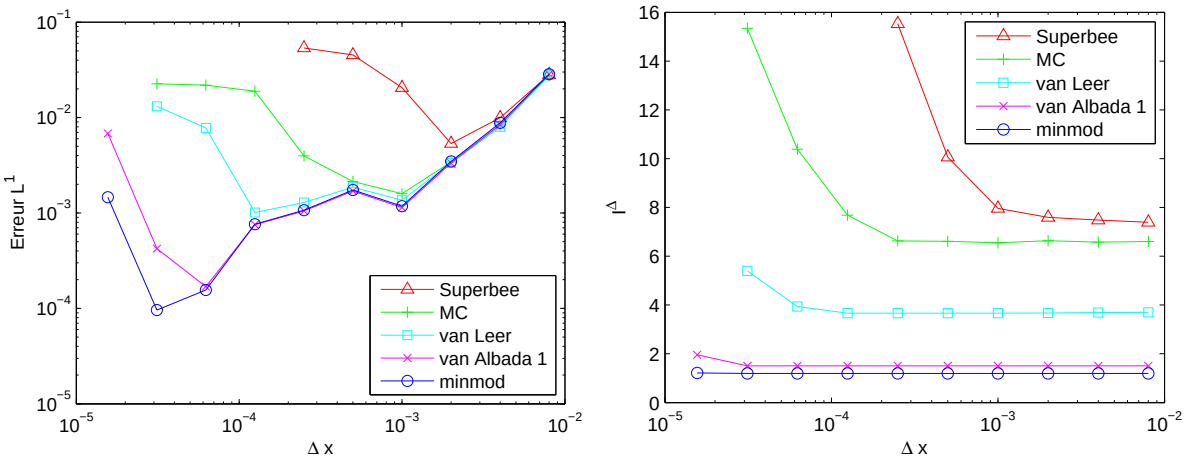


FIGURE 3.5 – Double choc avec un schéma d'ordre un en temps. Gauche : erreur L^1. Droite : erreur entropique I^Δ

semblent converger). Cependant, on remarque que I^Δ ne converge pas vers zéro, mais vers une valeur positive, ce qui est en accord avec la conjecture de Hou et LeFloch.

	minmod		van Albada 1		van Leer		MC		Superbee	
Δx	E^Δ	I^Δ	E^Δ	I^Δ	E^Δ	I^Δ	E^Δ	I^Δ	E^Δ	I^Δ
8.00E-3	2.87E-2	1.24970	2.85E-2	1.52251	2.80E-2	4.53453	2.86E-2	6.40651	2.96E-2	7.77453
4.00E-3	9.05E-3	1.25344	8.63E-3	1.52491	8.24E-3	4.56002	8.22E-3	6.57310	8.79E-3	7.95511
2.00E-3	3.60E-3	1.25475	3.43E-3	1.52644	3.55E-3	4.54521	3.39E-3	6.63387	4.20E-3	8.02181
1.00E-3	1.23E-3	1.25538	1.18E-3	1.52742	1.57E-3	4.53574	1.35E-3	6.68246	1.97E-3	8.08829
5.00E-4	1.77E-3	1.25556	1.72E-3	1.52786	2.06E-3	4.53815	1.71E-3	6.69711	2.85E-3	8.12208
2.50E-4	1.08E-3	1.25580	1.06E-3	1.52803	1.47E-3	4.53633	1.12E-3	6.75301	3.21E-3	8.15356
1.25E-4	7.59E-4	1.25594	7.46E-4	1.52813	1.21E-3	4.53547	8.41E-4	6.87392	1.92E-2	9.33707
6.25E-5	1.50E-4	1.25600	1.38E-4	1.52815	6.19E-4	4.53571	2.28E-4	6.93705	2.36E-2	13.8100
3.12E-5	6.52E-5	1.25602	5.64E-5	1.52817	5.45E-4	4.53550	1.49E-4	6.99102	3.22E-2	31.3636
1.56E-5	2.84E-5	1.25603			5.14E-4	4.53531	1.13E-4	7.02895		

Tableau 3.4 – Erreur L^1 et erreur entropique pour le double choc en utilisant un schéma d'ordre deux en temps

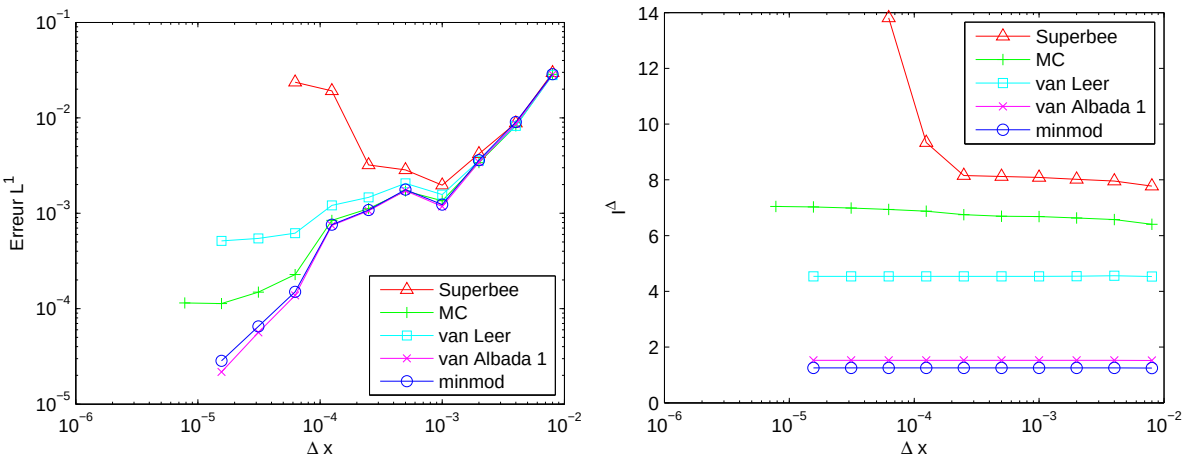


FIGURE 3.6 – Double choc avec un schéma d'ordre deux en temps. Gauche : erreur L^1. Droite : erreur entropique I^Δ

Ces expériences numériques semblent confirmer que la mesure δ est strictement positive, mais concentrée sur les courbes de discontinuité de la solution convergée w. Par conséquent, les inégalités d'entropie discrètes (3.23) (par exemple, celles données par [21, 12]) ne sont pas suffisantes pour assurer que la solution convergée soit entropique au sens du théorème de Lax-Wendroff (Théorème 3.1). Ces inégalités ne sont donc pas pertinentes, puisqu'elles n'empêchent pas les instabilités.

3.2 Obtention de toutes les inégalités d'entropie discrètes à partir d'une seule

D'après les résultats qui précèdent, un schéma d'ordre élevé doit satisfaire les inégalités d'entropie discrètes (3.16). En effet, les formulations non standards de la dérivée discrète en temps peuvent introduire des inégalités d'entropie inappropriées incluant une mesure positive. Afin de dériver un schéma d'ordre élevé capable de rétablir (3.16), on va adopter une technique *a posteriori* basée sur la vérification des inégalités d'entropie discrètes. Rappelons que les inégalités d'entropie discrètes (3.16) doivent être vérifiées pour toute paire d'entropie $(\rho\mathcal{F}(\ln(s)), \rho\mathcal{F}(\ln(s))u)$, où $\mathcal{F}$ est une fonction régulière satisfaisant (3.6). Les limitations *a posteriori* sont pertinentes

seulement lorsqu'un nombre fini d'estimations sont considérées, alors que l'on doit ici vérifier une infinité d'inégalités d'entropie discrètes.

Le but de cette partie est de donner des arguments pour obtenir toutes les inégalités d'entropie requises à partir d'une seule. Pour aborder ce problème, on commence par reformuler les paires d'entropie de la façon suivante :

Lemme 3.4. *Les paires d'entropie $(\eta, \mathcal{G})$, définies par (3.4)–(3.5), se réécrivent*

$$\eta(w) = \rho\psi(r(w)), \quad \mathcal{G}(w) = \rho\psi(r(w))u,$$

où l'on a posé

$$r(w) = -\frac{p^{1/\gamma}}{\rho}, \tag{3.40}$$

et ψ est une fonction régulière, croissante et convexe.

On souligne dès à présent que ce résultat n'est pas essentiel dans la suite, mais il facilite la présentation. En effet, on va voir que considérer des entropies $\eta(w)$ paramétrées par une fonction monotone convexe ψ sera plus pratique que de considérer des entropies paramétrées par une fonction $\mathcal{F}$ vérifiant la propriété (3.6). Cependant, on insiste sur le fait que toutes les dérivations de schémas qui vont suivre peuvent être réalisées en adoptant les paires d'entropie usuelles données par (3.4)–(3.5).

Démonstration. Dans un premier temps, remarquons que l'entropie spécifique, définie par (3.3), s'écrit $r(w) = -s(w)^{1/\gamma}$. On considère maintenant deux fonctions $\widetilde{\eta}$ et $\widetilde{\mathcal{G}}$ telles que l'on a

$$\widetilde{\eta}(w) = \rho\psi(r(w)) \quad \text{et} \quad \widetilde{\mathcal{G}}(w) = \rho\psi(r(w))u,$$

où ψ est une fonction régulière, croissante et convexe. En introduisant

$$\mathcal{F}(\ln(s)) := \psi\left(-s^{1/\gamma}\right),$$

on obtient

$$\mathcal{F}(y) = \psi\left(-\mathrm{e}^{y/\gamma}\right).$$

On en déduit

$$\mathcal{F}'(y) = -\frac{1}{\gamma}\psi'\left(-\mathrm{e}^{y/\gamma}\right) < 0$$

et

$$\mathcal{F}'(y) - \gamma\mathcal{F}''(y) = -\frac{1}{\gamma}\left(\psi'\left(-\mathrm{e}^{y/\gamma}\right) + \psi''\left(-\mathrm{e}^{y/\gamma}\right)\right) < 0.$$

Par conséquent, la fonction régulière $\mathcal{F}$ vérifie (3.6) et la paire $(\widetilde{\eta}, \widetilde{\mathcal{G}})$ est une paire d'entropie.

Réciproquement, considérons une paire d'entropie $(\eta, \mathcal{G}) = (\rho\mathcal{F}(\ln(s)), \rho\mathcal{F}(\ln(s))u)$, où $\mathcal{F}$ vérifie (3.6). Puisque l'on a

$$\mathcal{F}(\ln(s)) = \mathcal{F}\left(\gamma\ln(-r)\right),$$

on pose

$$\psi(r) := \mathcal{F}(\gamma\ln(-r)),$$

et l'on obtient les relations suivantes :

$$\eta(w) = \rho\psi(r) \quad \text{et} \quad \mathcal{G}(w) = \rho\psi(r)u.$$

Puisque (3.6) est vérifiée, on obtient aisément

$$\psi'(r) = \frac{\gamma}{r}\mathcal{F}'(\gamma\ln(-r)) > 0$$

et

$$\psi''((r) = -\frac{\gamma}{r^2}\left(\mathcal{F}'(\gamma\ln(-r)) - \gamma\mathcal{F}''(\gamma\ln(-r))\right) > 0.$$

Comme attendu, ψ est une fonction croissante et convexe et la preuve est achevée. □

À l'aide de ce résultat, on établit maintenant des conditions pour qu'une méthode de volumes finis préserve l'entropie dès qu'une seule inégalité d'entropie discrète bien choisie est vérifiée. Considérons un schéma conservatif donné par

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(F_{i+1/2}^n - F_{i+1/2}^n\right), \tag{3.41}$$

où $F_{i+1/2} = \left(F_{i+1/2}^{\rho}, F_{i+1/2}^{\rho u}, F_{i+1/2}^{E}\right)^T$ est un flux numérique consistant, de la forme (3.7) ou plus généralement de la forme (3.15).

Théorème 3.5. *Supposons que le schéma (3.41) soit robuste : si $w_i^n \in \Omega$ pour tout $i \in \mathbb{Z}$, alors $w_i^{n+1} \in \Omega$. Supposons que l'inégalité d'entropie discrète spécifique suivante soit vérifiée :*

$$\rho_i^{n+1} r_i^{n+1} \leq \rho_i^n r_i^n - \frac{\Delta t}{\Delta x}\left(F_{i+1/2}^{\rho} r_{i+1/2}^n - F_{i-1/2}^{\rho} r_{i-1/2}^n\right), \tag{3.42}$$

où l'on a posé

$$r_i^n = -\frac{(p_i^n)^{-1/\gamma}}{\rho_i^n} \quad \textit{et} \quad r_{i+1/2}^n = \begin{cases} r_{i+1}^n, & \textit{si } F_{i+1/2}^{\rho} < 0, \\ r_i^n, & \textit{si } F_{i+1/2}^{\rho} > 0. \end{cases} \tag{3.43}$$

Supposons enfin que la condition de type CFL supplémentaire suivante est vérifiée :

$$\frac{\Delta t}{\Delta x}\left(\max\left(0, F_{i+1/2}^{\rho}\right) - \min\left(0, F_{i-1/2}^{\rho}\right)\right) \leq \rho_i^n. \tag{3.44}$$

Alors le schéma (3.41) est entropique : pour toute fonction régulière, croissante et convexe ψ, on a

$$\rho_i^{n+1}\psi\left(r_i^{n+1}\right) \leq \rho_i^n\psi\left(r_i^n\right) - \frac{\Delta t}{\Delta x}\left(F_{i+1/2}^{\rho}\psi_{i+1/2}^n - F_{i-1/2}^{\rho}\psi_{i-1/2}^n\right),$$

avec $\psi_{i+1/2}^n$ défini par

$$\psi_{i+1/2}^n = \begin{cases} \psi\left(r_{i+1}^n\right) & \textit{si} F_{i+1/2}^{\rho} < 0, \\ \psi\left(r_i^n\right) & \textit{si} F_{i+1/2}^{\rho} > 0. \end{cases} \tag{3.45}$$

Avant de prouver ce résultat, on souligne le caractère particulier du flux numérique d'entropie qui intervient dans (3.42). En fait, on impose à l'entropie r de vérifier une propriété de type « transport ». Cette condition est clairement plus contraignante que les conditions habituelles, mais il existe des schémas numériques vérifiant ce type de propriété. Par exemple, le schéma de relaxation de Suliciu [20, 30] ou de manière équivalente le schéma HLLC [101, 100] sont des schémas d'ordre un entropiques qui impliquent un flux numérique d'entropie de la forme $F_{i+1/2}^{\rho}\mathcal{F}\left(\ln\left(s_{i+1/2}^n\right)\right)$ où

$$s_{i+1/2}^n = \begin{cases} p_{i+1}^n / \left(\rho_{i+1}^n\right)^{\gamma}, & \text{si } F_{i+1/2}^{\rho} < 0, \\ p_i^n / \left(\rho_i^n\right)^{\gamma}, & \text{si } F_{i+1/2}^{\rho} > 0. \end{cases}$$

Par conséquent, en introduisant $r_{i+1/2}^n = -\left(s_{i+1/2}^n\right)^{1/\gamma}$, ces schémas vérifient les inégalités (3.42)–(3.43).

Démonstration. D'après la définition de $r_{i+1/2}$ donnée par (3.43), on a la relation suivante :

$$F^{\rho}_{i+1/2} r_{i+1/2} = F^{\rho}_{i+1/2} \frac{r^n_i + r^n_{i+1}}{2} - \left| F^{\rho}_{i+1/2} \right| \frac{r^n_{i+1} - r^n_i}{2}.$$

En injectant cette relation dans (3.42), on obtient

$$r^{n+1}_i \leq \frac{a}{\rho^{n+1}_i} r^n_{i-1} + \frac{b}{\rho^{n+1}_i} r^n_i + \frac{c}{\rho^{n+1}_i} r^n_{i+1}, \tag{3.46}$$

où l'on a posé

$$a = \frac{\Delta t}{2\Delta x} \left(F^{\rho}_{i-1/2} + \left| F^{\rho}_{i-1/2} \right| \right), \tag{3.47}$$

$$b = \rho^n_i - \frac{\Delta t}{2\Delta x} \left(F^{\rho}_{i+1/2} + \left| F^{\rho}_{i+1/2} \right| - F^{\rho}_{i-1/2} + \left| F^{\rho}_{i-1/2} \right| \right), \tag{3.48}$$

$$c = \frac{\Delta t}{2\Delta x} \left(\left| F^{\rho}_{i+1/2} \right| - F^{\rho}_{i+1/2} \right). \tag{3.49}$$

Remarquons que

$$\begin{aligned} a + b + c &= \rho^n_i - \frac{\Delta t}{\Delta x} \left(F^{\rho}_{i+1/2} - F^{\rho}_{i-1/2} \right), \\ &= \rho^{n+1}_i > 0. \end{aligned}$$

On voit immédiatement que a et b sont positifs. Par ailleurs, la condition de type CFL supplémentaire (3.44) assure que le coefficient b est positif. Par conséquent, on a établi que le second membre de (3.46) est une combinaison convexe de r^n_{i-1}, r^n_i et r^n_{i+1}.

Considérons maintenant une paire d'entropie que l'on peut écrire, d'après le Lemme 3.4,

$$(\eta, \mathcal{G}) = (\rho\psi(r), \rho\psi(r)u),$$

où ψ est une fonction régulière, croissante et convexe. La fonction ψ étant croissante, on déduit de l'inégalité (3.46)

$$\psi\left(r^{n+1}_i\right) \leq \psi\left(\frac{a}{\rho^{n+1}_i} r^n_{i-1} + \frac{b}{\rho^{n+1}_i} r^n_i + \frac{c}{\rho^{n+1}_i} r^n_{i+1} \right).$$

En utilisant l'inégalité de Jensen, on obtient

$$\psi\left(r^{n+1}_i\right) \leq \frac{a}{\rho^{n+1}_i} \psi\left(r^n_{i-1}\right) + \frac{b}{\rho^{n+1}_i} \psi\left(r^n_i\right) + \frac{c}{\rho^{n+1}_i} \psi\left(r^n_{i+1}\right).$$

Ensuite, en remplaçant les coefficients a, b et c par leur valeur exacte donnée par (3.47), on trouve

$$\begin{aligned} \rho^{n+1}_i \psi(r^{n+1}_i) \leq \rho^n_i \psi(r^n_i) - \frac{\Delta t}{2\Delta x} \Big(& F^{\rho}_{i+1/2} (\psi(r^n_i) + \psi(r^n_{i+1})) - \left| F^{\rho}_{i+1/2} \right| (\psi(r^n_{i+1}) - \psi(r^n_i)) \\ & - F^{\rho}_{i-1/2} (\psi(r^n_{i-1}) + \psi(r^n_i)) + \left| F^{\rho}_{i-1/2} \right| (\psi(r^n_i) - \psi(r^n_{i-1})) \Big), \end{aligned}$$

que l'on peut réécrire

$$\rho^{n+1}_i \psi(r^{n+1}_i) \leq \rho^n_i \psi(r^n_i) - \frac{\Delta t}{\Delta x} \left(F^{\rho}_{i+1/2} \psi^n_{i+1/2} - F^{\rho}_{i-1/2} \psi^n_{i-1/2} \right),$$

où $\psi^n_{i+1/2}$ est défini par (3.45). La preuve est ainsi achevée. □

3.3 Le schéma e-MOOD pour les équations d'Euler

Dans cette partie, on dérive des schémas numériques d'ordre élevé en espace, donnés par (3.10), qui vérifient toutes les inégalités d'entropie discrètes requises (3.16). Pour simplifier la présentation, on se restreint ici aux méthodes numériques d'ordre un en temps. Cependant, d'après le Lemme 3.2, le schéma d'ordre élevé en espace, que l'on va maintenant détailler, s'étend très facilement en un schéma d'ordre élevé en temps entropique par une procédure de Runge-Kutta. Des extensions d'ordre élevé en temps seront utilisées pour réaliser les expériences numériques de la partie suivante.

Pour imposer les inégalités attendues (3.16), on suggère maintenant d'introduire une limitation *a posteriori* supplémentaire en reconstruisant les états $w_i^{n,-}$ et $w_i^{n,+}$ sur chaque cellule K_i. Cette technique de limitation *a posteriori* a été récemment introduite par Clain, Diot et Loubère [34, 47] dans le développement des schémas MOOD (Multi-dimensional Optimal Order Detection, voir Partie 1.5).

La technique MOOD permet d'étendre n'importe quel schéma d'ordre un, qui vérifie certaines propriétés, en un schéma d'ordre élevé en espace qui préserve ces propriétés. Elle est basée sur une procédure itérative pour déterminer, localement sur chaque cellule, la meilleure reconstruction préservant les propriétés imposées (ici, robustesse et stabilité).

On considère un schéma d'ordre un donné par (3.7). Sous la conditon CFL classique (3.8), on suppose que ce schéma d'ordre un vérifie les propriétés de robustesse et de stabilité suivantes :

Robustesse : Si tous les états initiaux w_i^n sont dans Ω, alors les états évolués w_i^{n+1} restent dans Ω.

Stabilité : Pour tout $i \in \mathbb{Z}$, l'inégalité d'entropie suivante est vérifiée :

$$\rho_i^{n+1} r_i^{n+1} \leq \rho_i^n r_i^n - \frac{\Delta t}{\Delta x}\left(F^\rho\left(w_i^n, w_{i+1}^n\right) r_{i+1/2}^n - F^\rho\left(w_{i-1}^n, w_i^n\right) r_{i-1/2}^n\right), \tag{3.50}$$

où $r_{i+1/2}^n$ est défini comme suit :

$$r_{i+1/2}^n = \begin{cases} r_{i+1}^n & \text{si } F^\rho\left(w_i^n, w_{i+1}^n\right) < 0, \\ r_i^n & \text{si } F^\rho\left(w_i^n, w_{i+1}^n\right) > 0. \end{cases} \tag{3.51}$$

On souligne à nouveau qu'un tel schéma d'ordre un existe. Par exemple, on renvoie le lecteur au schéma HLLC ou au schéma de relaxation de Suliciu [100, 30].

On adopte ensuite une procédure de reconstruction donnée par (3.17). Si l'incrément μ_i^n est défini par (3.18), on reste dans le cadre de la procédure MUSCL, mais μ_i^n peut être associé avec des techniques de reconstruction d'ordre plus élevé. Dans la suite, on suppose que la reconstruction préserve Ω :

$$w_i^{n,\pm} = w_i^n \pm \frac{1}{2}\mu_i^n \in \Omega, \quad \forall i \in \mathbb{Z}.$$

À ce stade, on remarque que la reconstruction vérifie la propriété de conservation suivante :

$$w_i^n = \frac{1}{2}\left(w_i^{n,-} + w_i^{n,+}\right). \tag{3.52}$$

Cette relation peut paraître restrictive. En effet, d'après des arguments présentés dans [12], il est possible d'utiliser une reconstruction telle que w_i^n ne soit pas une combinaison convexe de $w_i^{n,-}$ et $w_i^{n,+}$:

$$w_i^n \neq \alpha w_i^{n,-} + (1-\alpha) w_i^{n,+}, \quad \forall \alpha \in [0,1].$$

Cependant, la relation (3.52) rend la robustesse plus simple à obtenir. Par conséquent, par soucis de simplicité, on choisit ici de considérer une reconstruction vérifiant (3.52).

On peut maintenant présenter le schéma e-MOOD.

1. **Étape de reconstruction :** Pour chaque cellule K_i, on évalue des états reconstruits d'ordre élevé aux interfaces $x_{i-1/2}$ et $x_{i+1/2}$, donnés par

$$w_i^{n,\pm} = w_i^n \pm \frac{1}{2}\mu_i^n \in \Omega. \tag{3.53}$$

2. **Étape d'évolution :** On fait évoluer la solution reconstruite de la façon suivante :

$$w_i^{n+1,\star} = w_i^n - \frac{\Delta t}{\Delta x}\left(F\left(w_i^{n,+}, w_{i+1}^{n,-}\right) - F\left(w_{i-1}^{n,+}, w_i^{n,-}\right)\right), \quad \forall i \in \mathbb{Z}. \tag{3.54}$$

3. **Étape de limitation** *a posteriori* **:** On pose

$$\mathcal{E}_i^n = \rho_i^n r_i^n - \frac{\Delta t}{\Delta x}\left(F^\rho\left(w_i^{n,+}, w_{i+1}^{n,-}\right) r_{i+1/2}^n - F^\rho\left(w_{i-1}^{n,+}, w_i^{n,-}\right) r_{i-1/2}^n\right), \tag{3.55}$$

où $r_{i+1/2}^n$ est défini par (3.51). En notant $r_i^{n+1,\star} = r\left(w_i^{n+1,\star}\right)$, on a l'alternative suivante :

- Si pour tout $i \in \mathbb{Z}$, on a
$$\rho_i^{n+1,\star} r_i^{n+1,\star} \leq \mathcal{E}_i^n, \tag{3.56}$$
alors la solution est valide et l'on pose
$$w_i^{n+1} = w_i^{n+1,\star}, \quad \forall i \in \mathbb{Z}.$$
- Sinon, pour tout $i \in \mathbb{Z}$ tel que (3.56) n'est pas vérifiée, on pose $w_i^{n,\pm} = w_i^n$, puis on retourne à l'étape 2.

On précise que l'étape 3 est en fait une étape de préservation de l'entropie. En effet, la condition (3.56) coïncide exactement avec l'inégalité d'entropie requise.

Avant d'établir la robustesse et la stabilité vérifiées par le schéma e-MOOD, on souligne quelques différences entre cette procédure numérique et le schéma MOOD original introduit dans [34].

Tout d'abord, on rappelle que le schéma MOOD original utilise une procédure itérative sur le degré $0 \leq d_i \leq d_{\max}$ du polynôme intervenant dans la reconstruction. Il est important de noter que le degré d_i est défini localement sur chaque cellule K_i. Ensuite, durant l'étape de limitation *a posteriori*, si la propriété requise (ici, l'inégalité d'entropie discrète) n'est pas vérifiée, alors le degré d_i du polynôme est décrémenté et la méthode MOOD est à nouveau appliquée. Cette procédure itérative sur d_i s'arrête lorsque $d_i = 0$, puisque l'on a alors $\mu_i^n = 0$ et l'on retrouve un schéma d'ordre un qui, par hypothèse, vérifie la propriété attendue.

Pour simplifier la présentation de la méthode e-MOOD, on a choisi d'arrêter la procédure itérative à la fin de la première itération, en passant directement de la reconstruction de degré maximal à la reconstruction de degré zéro, correspondant au schéma d'ordre un. Bien sûr, il aurait été possible d'adopter une procédure constituée de plusieurs itération de $d_i = d_{\max}$ à $d_i = 0$. Les résultats de robustesse et de stabilité énoncés plus bas resteraient alors valables.

Ensuite, on met l'accent sur l'ordre effectif de précision. En effet, le schéma e-MOOD, mais aussi le schéma MOOD original, remplacent le schéma d'ordre élevé par une méthode d'ordre un dès que les propriétés requises ne sont pas vérifiées. Clairement, si la propriété requise est

« trop forte », la limitation sera activée sur le domaine de calcul entier et l'approximation obtenue aura une précision à l'ordre un. En pratique, on a considéré une étape de reconstruction donnée par un approche MUSCL usuelle et les améliorations dans les résultats numériques sont évidentes. De notre point de vue, le schéma e-MOOD doit être compris comme une technique de stabilisation et non seulement comme une procédure d'ordre élevé.

Le dernier point que l'on souhaite aborder concerne le degré effectif de la reconstruction. Dans la méthode MOOD originale, concernant la cellule K_i, on est obligé d'utiliser un polynôme de degré $\min(d_{i-1}, d_i)$ pour reconstruire l'état $w_i^{n,-}$ au point $x_{i-1/2}$ et un polynôme de degré $\min(d_i, d_{i+1})$ pour reconstruire l'état $w_i^{n,+}$ au point $x_{i+1/2}$. Cela nécessite de calculer deux reconstructions polynomiales de degrés différents sur chaque cellule, mais cela s'avère nécessaire pour assurer que l'algorithme MOOD se termine bien. En effet, la limitation *a posteriori* dans la méthode MOOD originale est basée sur la robustesse et sur un principe du maximum. Pour assurer que l'état w_i^{n+1} vérifie ces propriétés, il ne suffit pas d'imposer que la reconstruction soit de degré zéro sur la cellule K_i, il faut que tous les états intervenant dans le schéma donnant w_i^{n+1} soient des reconstructions de degré zéro, y compris les états $w_{i-1}^{n,+}$ et $w_{i+1}^{n,-}$.

Dans la méthode e-MOOD, ce problème ne se pose pas car on n'utilise pas de critère de type principe du maximum. Lorsque la procédure de limitation e-MOOD a été activée sur une cellule, alors le nouvel état évolué vérifie les propriétés requises de robustesse et de stabilité, quelque soit la reconstruction utilisée sur les cellules voisines. Cette remarque est essentielle car elle rend la méthode e-MOOD très simple à implémenter et d'un moindre coût que le schéma MOOD original.

On peut maintenant énoncer les propriétés de robustesse et de stabilité satisfaites par le schéma e-MOOD.

Théorème 3.6. *Supposons que le pas de temps Δt vérifie les deux conditions de type CFL suivantes :*

$$\frac{\Delta t}{\Delta x} \max_{i\in\mathbb{Z}} \left(\left| \lambda^{\pm} \left(w_i^{n,+}, w_{i+1}^{n,-} \right) \right|, \left| \lambda^{\pm} \left(w_i^{n,-}, w_i^{n,+} \right) \right| \right) \leq \frac{1}{4}, \tag{3.57}$$

$$\frac{\Delta t}{\Delta x} \left(\max \left(0, F_{i+1/2}^{\rho} \right) - \min \left(0, F_{i-1/2}^{\rho} \right) \right) \leq \rho_i^n. \tag{3.58}$$

Supposons que w_i^n et tous les états reconstruits $w_i^{n,\pm}$ définis par (3.53) sont dans Ω, pour tout $i \in \mathbb{Z}$.

Alors l'algorithme e-MOOD se termine en un nombre finis d'étapes. De plus, les états évolués w_i^{n+1} résultant de la procédure e-MOOD appartiennent à Ω pour tout $i \in \mathbb{Z}$. Enfin, pour toute fonction ψ régulière, croissante et convexe, le schéma e-MOOD vérifie l'inégalité d'entropie discrète

$$\begin{aligned} &\frac{1}{\Delta t} \left(\rho_i^{n+1} \psi \left(r_i^{n+1} \right) - \rho_i^n \psi \left(r_i^n \right) \right) \\ &\qquad + \frac{1}{\Delta x} \left(F^{\rho} \left(w_i^{n,+}, w_{i+1}^{n,-} \right) \psi \left(r_{i+1/2}^n \right) - F^{\rho} \left(w_{i-1}^{n,+}, w_i^{n,-} \right) \psi \left(r_{i-1/2}^n \right) \right) \leq 0, \end{aligned} \tag{3.59}$$

où $r_{i+1/2}^n$ est défini par (3.51). Le schéma e-MOOD est donc entropique.

Démonstration. Commençons par supposer que la procédure de limitation *a posteriori* a été activée sur la cellule K_i. On a alors $w_i^{n,\pm} = w_i^n$, ce qui implique que le schéma (3.54) est un schéma d'ordre un. On peut donc utiliser l'hypothèse (3.50) qui assure

$$\rho_i^{n+1,\star} r_i^{n+1,\star} \leq \mathcal{E}_i^n.$$

Par conséquent, le critère d'entropie est désormais vérifié sur la cellule K_i. La procédure de limitation *a posteriori* ne peut donc s'appliquer au plus qu'une fois par cellule et l'algorithme se termine en un nombre fini d'étapes.

On montre maintenant la robustesse du schéma e-MOOD. Puisque il n'y a pas de critère de robustesse dans la limitation *a posteriori*, il faut montrer que l'état $w_i^{n+1,\star}$, défini par (3.54), appartient à Ω pour tout $i \in \mathbb{Z}$. Sur chaque cellule K_i, on a alors deux alternatives :

- Soit la procédure de limitation *a posteriori* a été activée sur la cellule K_i et l'état $w_i^{n+1,\star}$ est alors obtenu par un schéma d'ordre un. Or celui-ci est robuste sous la condition CFL (3.8) par hypothèse. La condition CFL (3.57) étant plus restrictive que (3.8), on en déduit que l'état $w_i^{n+1,\star}$ est dans Ω.
- Soit la procédure de limitation *a posteriori* n'a pas été activée sur la cellule K_i et l'état $w_i^{n+1,\star}$ est alors obtenu par un schéma MUSCL. Grâce à l'hypothèse de conservation (3.52) vérifiée par la reconstruction et à la condition CFL (3.57), on peut alors appliquer le Lemme 1.9 qui assure que $w_i^{n+1,\star}$ est dans Ω.

Enfin, par définition du schéma e-MOOD, on a l'inégalité d'entropie discrète

$$\rho_i^{n+1} r_i^{n+1} \leq \rho_i^n r_i^n - \frac{\Delta t}{\Delta x}\left(F^{\rho}_{i+1/2} r^n_{i+1/2} - F^{\rho}_{i+1/2} r^n_{i-1/2}\right). \tag{3.60}$$

Sous la condition de type CFL (3.58), on peut appliquer le Théorème 3.5 et le schéma e-MOOD vérifie toutes les inégalités d'entropie discrètes requises. La preuve est ainsi achevée. □

3.4 Résultats numériques

Par soucis de consistance, les expériences numériques que l'on détaille maintenant suivent une stratégie similaire à celle présentée dans la Partie 3.1.4. Pour valider le schéma e-MOOD, on choisit comme flux numérique intervenant dans (3.54), le flux donné par le schéma HLLC [101, 100]. En ce qui concerne l'étape de reconstruction e-MOOD (3.53), on utilise des limiteurs MUSCL. On ne présente ici que les résultats obtenus avec le limiteur minmod et le limiteur Superbee. En effet, d'après les Tableaux 3.1–3.4, le limiteur minmod est le plus stable, alors que le limiteur Superbee est celui qui explose le plus rapidement. On compare systématiquement l'ordre un et l'ordre deux en temps. Afin de montrer l'amélioration apportée par la méthode e-MOOD, on compare également les résultats obtenus par le même limiteur, avec ou sans la procédure e-MOOD. Le pas de temps est restreint suivant les conditions CFL (3.57)–(3.58).

La première expérience numérique a pour but d'illustrer la précision du schéma e-MOOD en approchant une solution régulière périodique de (3.1). On considère une donnée initiale donnée sur le domaine de calcul $[0, 1]$ par

$$\begin{aligned}
\rho_0(x) &= \begin{cases} 1, & \text{si } x < 0.2 \text{ ou } x > 0.8, \\ 1 + \exp\left(\frac{(x-0.5)^2}{(x-0.2)(x-0.8)}\right) & \text{si } 0.2 < x < 0.8. \end{cases} \\
u_0(x) &= 1, \\
p_0(x) &= 1.
\end{aligned}$$

La solution exacte est donnée par

$$w(x,t) = w_0(x - at), \quad \text{avec } a = 1.$$

En utilisant des conditions de bord périodiques, la solution exacte en densité à $t = 1$ coïncide avec la donnée initiale (voir Figure 3.7).

Dans le Tableau 3.5 et la Figure 3.8, on donne l'erreur L^1 obtenue par le limiteur minmod et le limiteur Superbee en utilisant un discrétisation en temps de Runge-Kutta d'ordre deux.

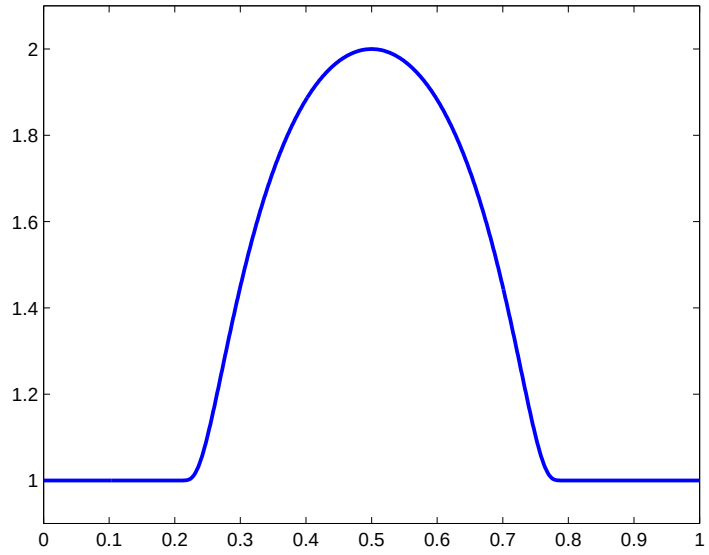


FIGURE 3.7 – Problème régulier : solution initiale et finale en densité

Les approximations MUSCL et e-MOOD donnent la même erreur avec le limiteur minmod. En fait, dans ce cas-test très régulier, le schéma MUSCL-minmod s'avère être déjà robuste et entropique. Par conséquent, la procédure e-MOOD reste inactive et on obtient les mêmes résultats. Par contre, en ce qui concerne le limiteur Superbee, la procédure e-MOOD impose une régularisation numérique et il n'y a pas d'explosion numérique, contrairement au schéma MUSCL-Superbee. Pour compléter cette expérience de validation, on donne également les résultats obtenus en utilisant un limiteur minmod d'ordre quatre présenté dans [107, 17]. On remarque que les schémas MUSCL et e-MOOD donnent des résultats similaires. Cela prouve que la méthode e-MOOD n'est pas trop diffusive et qu'elle préserve l'ordre de précision du schéma MUSCL tout en étant maintenant entropique d'après le Théorème 3.6.

	Schéma MUSCL			Schéma e-MOOD		
Δx	minmod	Superbee	minmod4	minmod	Superbee	minmod4
8.00E-3	3.75E-3	7.62E-4	1.22E-4	3.75E-3	7.62e-4	1.24e-4
4.00E-3	1.43E-3	4.98E-4	4.80E-5	1.43E-3	4.98e-4	4.81e-5
2.00E-3	4.94E-4	2.41E-4	7.16E-6	4.94E-4	2.41e-4	7.18e-6
1.00E-3	1.46E-4	7.73E-5	4.92E-7	1.49E-4	7.70e-5	4.93e-7
5.00E-4	3.91E-5	2.13E-5	1.48E-8	3.91E-5	2.11e-5	1.49e-8
2.50E-4	9.60E-6	5.44E-6	3.11E-10	9.60E-6	5.49e-5	3.17e-10
1.25E-4	2.40E-6	1.47E-6	1.89E-11	2.40E-6	1.67e-6	1.92e-11
6.25E-5	6.02E-7	2.55E-2	3.39E-12	6.02E-7	1.12e-6	3.40e-12
3.12E-5	1.51E-7			1.51E-7	1.55e-6	

Tableau 3.5 – Erreur L^1 pour le problème régulier en utilisant le schéma MUSCL et le schéma e-MOOD

On passe maintenant à l'approximation de la solution de problèmes de Riemann avec une donnée initiale donnée par (3.37). Dans le Tableau 3.6 et la Figure 3.9, on présente les résultats numériques obtenus en simulant une 1-détente avec une donnée initiale donnée par (3.38). On remarque immédiatement qu'il n'y a plus d'explosion numérique avec le schéma e-MOOD, mais l'ordre de précision attendu est conservé. Dans le Tableau 3.7 et la Figure 3.10, des résultats similaires sont obtenus pour le double choc avec une donnée initiale définie par (3.39).

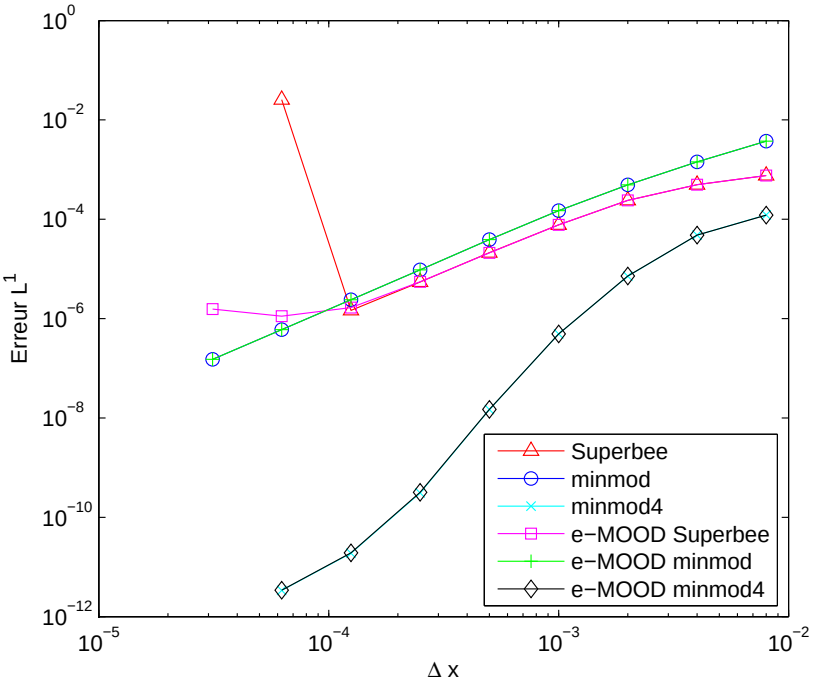


FIGURE 3.8 – Problème régulier : comparaison en norme L^1 entre le schéma MUSCL et le schéma e-MOOD

	ordre 1 en temps		ordre 2 en temps	
Δx	minmod	Superbee	minmod	Superbee
8.00E-3	2.33E-2	3.17E-2	1.53E-2	2.05E-2
4.00E-3	1.51E-2	1.99E-2	1.10E-2	1.20E-3
2.00E-3	9.26E-3	1.23E-2	6.30E-3	7.18E-3
1.00E-3	5.63E-3	7.45E-3	3.70E-3	4.23E-3
5.00E-4	3.33E-3	4.44E-3	2.13E-3	2.44E-3
2.50E-4	1.93E-3	2.66E-3	1.22E-3	1.40E-3
1.25E-4	1.11E-3	1.71E-3	6.88E-4	7.89E-4
6.25E-5	6.34E-4	1.14E-3	3.84E-4	4.39E-4
3.12E-5	3.62E-4	8.14E-4	2.16E-4	2.43E-4
1.56E-5			1.20E-4	1.34E-4

Tableau 3.6 – Erreur L^1 pour la 1-détente en utilisant le schéma e-MOOD

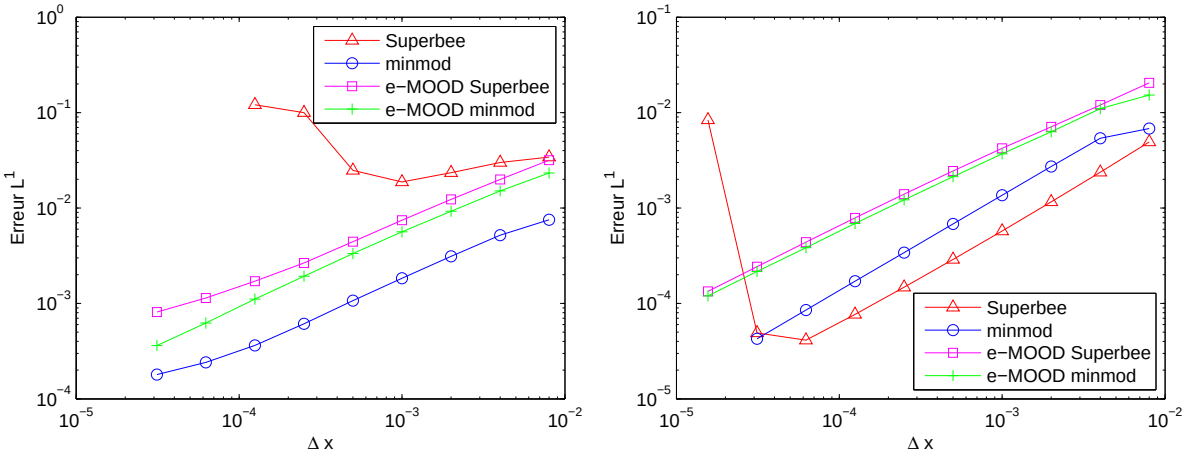


FIGURE 3.9 – 1–détente : comparaison en norme L^1 entre le schéma MUSCL et le schéma e-MOOD. Gauche : schéma d'ordre un en temps. Droite : schéma d'ordre deux en temps

	ordre 1 en temps		ordre 2 en temps	
Δx	minmod	Superbee	minmod	Superbee
8.00E-3	4.05E-2	3.99E-2	4.12E-2	4.12E-2
4.00E-3	1.63E-2	1.60E-2	1.66E-2	1.67E-3
2.00E-3	6.98E-3	7.17E-3	7.16E-3	7.31E-3
1.00E-3	2.74E-3	2.77E-3	2.82E-3	2.80E-3
5.00E-4	2.60E-3	2.57E-3	2.65E-3	2.63E-3
2.50E-4	1.51E-3	1.49E-3	1.53E-3	1.52E-3
1.25E-4	9.43E-4	9.45E-4	9.59E-4	9.58E-4
6.25E-5	2.60E-4	2.53E-4	2.68E-4	2.56E-4
3.12E-5	1.14E-4	1.17E-4	1.19E-4	1.16E-4
1.56E-5	4.92E-5	4.96E-5		

Tableau 3.7 – Erreur L^1 pour le double choc en utilisant le schéma e-MOOD

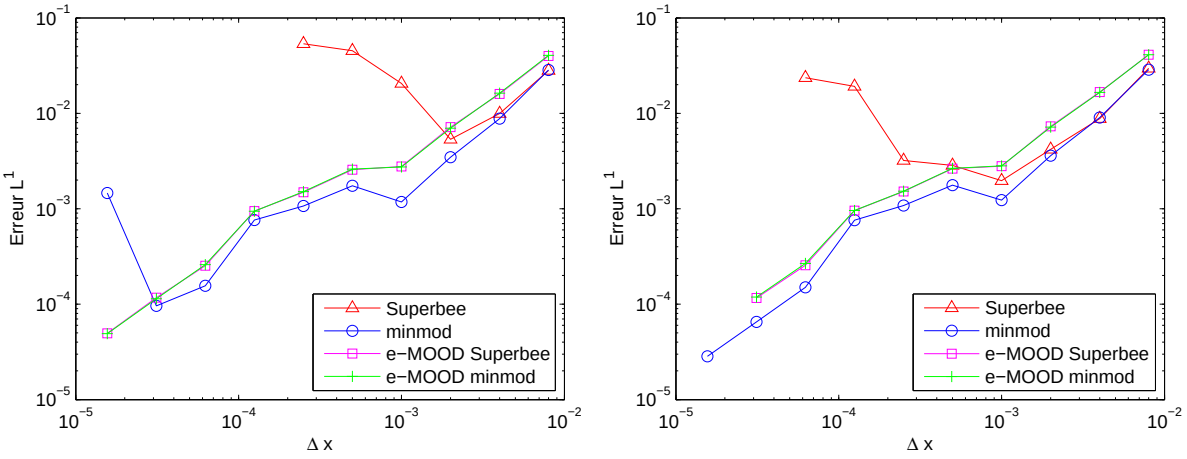


FIGURE 3.10 – Double choc : comparaison en norme L^1 entre le schéma MUSCL et le schéma e-MOOD. Gauche : schéma d'ordre un en temps. Droite : schéma d'ordre deux en temps

4

Schémas well-balanced pour des systèmes de lois de conservation avec terme source

Ce travail a été réalisé en collaboration avec l'équipe de Christian Klingenberg de l'université de Würzburg.

On s'intéresse ici à l'approximation numérique de certains systèmes hyperboliques avec termes sources de la forme

$$\partial_t w + \partial_x f(w) = s(w)\partial_x Z, \tag{4.1}$$

où $s : \Omega \to \mathbb{R}^d$ est un terme source régulier et $Z : \mathbb{R} \to \mathbb{R}$ est une fonction régulière donnée. La présence de termes sources pose une difficulté supplémentaire : les schémas numériques construits pour ces systèmes doivent préserver certains états d'équilibre (en plus des propriétés habituelles de robustesse et de stabilité) afin d'assurer leur précision dans certains régimes. Les états d'équilibre sont définis par le système d'équations aux dérivées partielles

$$\partial_x f(w) = s(w)\partial_x Z, \tag{4.2}$$

autrement dit, ce sont les solutions de (4.1) indépendantes du temps. Les schémas qui préservent les états d'équilibre sont dits well-balanced. Parmi les premiers à s'intéresser à des schémas ayant cette propriété, on peut citer Bermudez et Vasquez [10], Greenberg et al. [60, 61], puis Gosse [58].

Dans le cas du système de Saint-Venant avec un terme source de topographie, il n'y a qu'un seul état d'équilibre au repos (à constante près) : le lac au repos. Celui-ci est décrit par une relation algébrique linéaire, ce qui facilite la dérivation de schémas well-balanced. Une importante littérature est consacrée à la construction de schémas well-balanced pour le système de Saint-Venant. Le lecteur pourra se référer par exemple à [25, 5, 20, 85, 15, 46].

Le but de ce chapitre est de construire des schémas numériques well-balanced pour les équations d'Euler avec un terme source de gravité. Par rapport à Saint-Venant, ce système présente une difficulté majeure venant du terme source. En effet, contrairement au système de Saint-Venant, les états d'équilibre sont gouvernés par une équation aux dérivées partielles que

l'on ne peut pas intégrer. En d'autres termes, les états d'équilibre ne sont pas donnés par une relation algébrique, sauf dans quelques cas particuliers.

On s'intéresse également au système de Ripa [89, 90] qui représente une difficulté intermédiaire. Le système en lui-même est moins non linéaire que les équations d'Euler et sa structure semble plus proche de celle du système de Saint-Venant, mais les états d'équilibre sont gouvernés par une équation aux dérivées partielles que l'on ne peut pas intégrer et posent donc la même difficulté que ceux des équations d'Euler avec gravité.

Dans une première partie, nous présentons de manière détaillée les systèmes de Saint-Venant, de Ripa et d'Euler avec gravité. Pour ces trois systèmes, on s'attachera à donner une description des états d'équilibres au repos. Dans la deuxième partie, on introduit le formalisme des schémas volumes finis en présence d'un terme source. En effet, les techniques de volumes finis diffèrent légèrement par rapport à celles introduites dans le Chapitre 1. En particulier, on détaille la notion de schéma de type Godunov pour un système avec terme source. Par ailleurs, en raison de la forme non explicite des états d'équilibre qui devront ensuite être préservés par les schémas numériques, une extension de la définition d'un schéma well-balanced sera donnée. Dans la troisième partie, on construit des solveurs de Riemann simples en suivant le formalisme introduit par Harten, Lax et van Leer [64], puis appliqué aux systèmes avec terme source par Gallice [52, 53] et Chalons [29]. On se concentre d'abord sur le système de Ripa avant d'étendre la stratégie aux équations d'Euler avec gravité. Dans la quatrième partie, on choisit une approche différente basée sur les méthodes de relaxation. On développe ainsi un modèle de relaxation de type Suliciu (voir [38, 20]) duquel on déduit un schéma numérique well-balanced pour les trois modèles qui nous motivent dans cette étude. Dans la Partie 4.5, on présente une extension à l'ordre deux pour les équations d'Euler avec gravité. Cela nécessite une modification de la définition d'un schéma well-balanced permettant de prendre en compte les reconstructions. La Partie 4.6 est dédiée à une extension du schéma de relaxation en deux dimensions d'espace. Enfin la pertinence de ces schémas est illustrée dans la Partie 4.7 par plusieurs résultats numériques.

4.1 Les modèles

Dans cette partie, on présente les trois systèmes d'EDP étudiés ici : le système de Saint-Venant, le système de Ripa et les équations d'Euler avec gravité. Dans chaque cas, on spécifie les grandeurs physiques entrant en jeu, puis on détaille l'algèbre du système avant de décrire les états d'équilibre. Comme mentionné précédemment, les états d'équilibre pour le système de Ripa et pour les équations d'Euler avec gravité ne sont pas tous déterminés par des relations algébriques. La notion de schéma well-balanced nécessite donc une définition claire. Celle-ci repose sur le choix d'une discrétisation de l'équation différentielle régissant les états d'équilibre. On choisit de donner d'abord une interprétation locale de cette équation. Pour chaque système d'EDP considéré, on introduit une notion d'équilibre local entre deux états qui approche, en un certain sens, la solution de l'équation différentielle (4.2). Cette définition sera utilisée dans la Partie 4.2 pour généraliser la notion usuelle de schéma well-balanced.

4.1.1 Modèle de Saint-Venant

Le modèle de Saint-Venant [92] permet de décrire les écoulements en eaux peu profondes. La couche d'eau doit être suffisamment faible pour pouvoir considérer que la vitesse est constante sur l'épaisseur. Les grandeurs physiques intervenant dans ce système sont la hauteur d'eau

$h(x,t)$ et la vitesse $u(x,t)$. En une dimension d'espace, le système de Saint-Venant s'écrit

$$\begin{cases} \partial_t h + \partial_x hu = 0, \\ \partial_t hu + \partial_x \left(hu^2 + gh^2/2 \right) = -gh\partial_x Z, \end{cases} \tag{4.3}$$

où $Z(x)$ est une fonction régulière donnée représentant la topographie et g la constante de gravité. Dans cette étude, on ne s'intéresse pas aux transitions sec/mouillé (voir par exemple [5, 20, 18]) et par conséquent, on supposera que la hauteur d'eau est strictement positive. Le vecteur des variables conservatives $w = (h, hu)^T$ appartient donc à l'ensemble convexe des états admissibles

$$\Omega = \{w \in \mathbb{R}^2, h > 0\}.$$

Pour étudier les propriétés de ce système, il est utile de faire intervenir le vecteur des grandeurs physiques (incluant la topographie) $U = (h, u, Z)^T$. Avec ce jeu de variables, le système de Saint-Venant se réécrit sous la forme quasi-linéaire

$$\partial_t U + A(U)\partial_x U = 0,$$

où la matrice $A(U)$ est donnée par

$$A(U) = \begin{pmatrix} u & h & 0 \\ g & u & g \\ 0 & 0 & 0 \end{pmatrix}.$$

En introduisant la vitesse du son $c = \sqrt{gh}$, cette matrice $A(U)$ admet trois valeurs propres

$$0, \quad u \pm c,$$

avec pour vecteurs propres respectifs

$$\begin{pmatrix} h \\ -u \\ \frac{u^2}{g} - h \end{pmatrix}, \quad \begin{pmatrix} h \\ \pm c \\ 0 \end{pmatrix}.$$

Les vecteurs propres sont distincts en tout point de Ω sauf aux points critiques qui sont les points pour lesquels $u = \pm c$. En ces points où les champs propres collapsent, un phénomène de résonance se produit. On ne s'étendra pas plus sur ce phénomène et on renvoie le lecteur par exemple à [79]. Le système est hyperbolique en dehors des points critiques. Notons que les champs associés aux valeurs propres $u \pm c$ sont vraiment non linéaires, alors que celui associé à la valeur propre 0 est linéairement dégénéré, avec comme invariants de Riemann

$$hu, \quad \frac{u^2}{2} + g(h + Z).$$

On s'intéresse maintenant aux états d'équilibre du système (4.3) qui sont les solutions de

$$\begin{cases} \partial_x hu = 0, \\ \partial_x \left(hu^2 + gh^2/2 \right) = -gh\partial_x Z. \end{cases}$$

On peut intégrer ce système d'équations différentielles pour trouver la formulation explicite des états d'équilibre :

$$\begin{cases} hu = \text{ constante}, \\ \frac{u^2}{2} + g(h + Z) = \text{ constante}. \end{cases}$$

Dans cette étude, on se contente de préserver les états d'équilibre au repos, c'est-à-dire ceux dont la vitesse est identiquement nulle. Ces états sont gouvernés par

$$\begin{cases} u \equiv 0, \\ \partial_x \frac{h^2}{2} = -h\partial_x Z. \end{cases}$$

dont on déduit l'expression du lac au repos, qui est l'unique état d'équilibre au repos (à constante près) pour le système de Saint-Venant :

$$\begin{cases} u \equiv 0, \\ h + Z = \text{constante}, \end{cases} \tag{4.4}$$

Pour des raisons de simplicité dans le développement des méthodes numériques, il est utile d'introduire la notion d'équilibre local. On pourrait se passer de cette notion dans le cas des équations de Saint-Venant car les états d'équilibre admettent une formulation explicite. Elle sera cependant essentielle pour le modèle de Ripa et pour les équations d'Euler avec gravité et afin de pouvoir traiter les trois systèmes de la même façon, on choisit d'introduire également cette notion pour les équations de Saint-Venant.

Définition 4.1 (Équilibre local pour Saint-Venant)**.** Deux états (w_L, Z_L) et (w_R, Z_R) sont dits à l'équilibre local pour le système de Saint-Venant (4.3) si

$$\begin{cases} u_L = u_R = 0, \\ [h + Z] = 0, \end{cases} \tag{4.5}$$

où l'on a introduit la notation $[X] = X_R - X_L$ qui sera utilisée dans toute la suite.

Soulignons que la consistance de (4.5) avec (4.4) est immédiate.

4.1.2 Modèle de Ripa

Le modèle de Ripa est une extension du système de Saint-Venant qui permet de prendre en compte des gradients horizontaux de température. Comme pour le système de Saint-Venant, on retrouve comme inconnues la hauteur d'eau $h(x,t)$ et la vitesse $u(x,t)$, mais on a en plus la température $\theta(x,t)$. En une dimension d'espace, le système de Ripa s'écrit

$$\begin{cases} \partial_t h + \partial_x hu = 0, \\ \partial_t hu + \partial_x \left(hu^2 + gh^2\theta/2 \right) = -gh\theta\partial_x Z, \\ \partial_t h\theta + \partial_x hu\theta = 0, \end{cases} \tag{4.6}$$

où $Z(x)$ est une fonction régulière donnée représentant la topographie et g la constante de gravité. Le vecteur des variables conservatives $w = (h, hu, h\theta)^T$ appartient à l'ensemble convexe des états admissibles

$$\Omega = \{w \in \mathbb{R}^3, h > 0, \theta > 0\}.$$

En utilisant le vecteur des grandeurs physiques $U = (h, u, \theta, Z)^T$, le système de Ripa se réécrit sous la forme quasi-linéaire

$$\partial_t U + A(U)\partial_x U = 0,$$

où la matrice $A(U)$ est donnée par

$$A(U) = \begin{pmatrix} u & h & 0 & 0 \\ g\theta & u & gh/2 & g\theta \\ 0 & 0 & u & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

En notant la vitesse du son $c = \sqrt{gh\theta}$, la matrice $A(U)$ admet quatre valeurs propres :

$$0, \quad u, \quad u \pm c,$$

avec pour vecteurs propres respectifs

$$\begin{pmatrix} h \\ -u \\ 0 \\ \frac{u^2}{g\theta} - h \end{pmatrix}, \quad \begin{pmatrix} h \\ 0 \\ -2\theta \\ 0 \end{pmatrix}, \quad \begin{pmatrix} h \\ \pm c \\ 0 \\ 0 \end{pmatrix}.$$

Toutefois, pour les états vérifiant $u = \pm c$, la base de diagonalisation est perdue. Notons que quand $u = 0$, la valeur propre 0 devient double, mais la matrice $A(U)$ reste diagonalisable. Par conséquent, le système est hyperbolique en tout point de Ω, sauf aux points pour lesquels $u = \pm c$. Les champs associés aux valeurs propres $u \pm c$ sont vraiment non linéaires, alors que ceux associés aux valeurs propres 0 et u sont linéairement dégénérés. Les invariants de Riemann sont pour l'onde 0 :

$$hu, \quad \theta, \quad h + Z + \frac{u^2}{2g\theta}$$

et pour l'onde u :

$$u, \quad Z, \quad h^2\theta.$$

Les états d'équilibre pour le système de Ripa (4.6) sont solutions du système d'équations différentielles

$$\begin{cases} \partial_x hu = 0, \\ \partial_x \left(hu^2 + gh^2\theta/2 \right) = -gh\theta\partial_x Z, \\ \partial_x hu\theta = 0. \end{cases}$$

Il y a alors deux possibilités :

- soit u ne s'annule pas et l'on peut obtenir l'expression explicite suivante :

$$\begin{cases} hu = \text{ constante}, \\ \theta = \text{ constante}, \\ \frac{u^2}{2} + g\theta(h + Z) = \text{ constante}, \end{cases}$$

  qui décrit les états d'équilibre en mouvement ;

- soit u est identiquement nulle et le système devient

$$\begin{cases} u \equiv 0, \\ \partial_x \left(h^2\theta/2 \right) + h\theta\partial_x Z = 0. \end{cases} \tag{4.7}$$

  Cette équation différentielle régit les états d'équilibre au repos et contrairement au système de Saint-Venant, elle n'est pas intégrable. On ne peut donc pas obtenir d'expression analytique de tous les états d'équilibre au repos. Toutefois, on note que deux états d'équilibres sont remarquables car ils vérifient une équation algébrique. En effet, si l'on impose à θ d'être constante, alors on retrouve la solution dite du lac au repos :

$$\begin{cases} u \equiv 0, \\ \theta = \text{constante}, \\ h + Z = \text{constante}. \end{cases} \tag{4.8}$$

De même, si Z est constante, alors on a

$$\begin{cases} u \equiv 0, \\ Z = \text{constante}, \\ h^2\theta = \text{constante}. \end{cases} \tag{4.9}$$

On signale que dans leur article [32], Chertock, Kurganov et Liu construisent un schéma numérique pour le système de Ripa qui préserve les deux états d'équilibre (4.8) et (4.9). Dans cette étude, l'objectif est de construire des schémas qui préservent tous les états d'équilibres gouvernés par (4.7) et en particulier (4.8) et (4.9) de façon exacte.

Il reste à définir la notion d'équilibre local pour le système de Ripa. En fait, il s'agit d'une discrétisation « à l'ordre un » de l'équation (4.7).

Définition 4.2 (Équilibre local pour Ripa)**.** Deux états (w_L, Z_L) et (w_R, Z_R) sont dits à l'équilibre local pour le système de Ripa (4.6) si

$$\begin{cases} u_L = u_R = 0, \\ \left[h^2\theta/2\right] + \bar{h}\bar{\theta}[Z] = 0, \end{cases} \tag{4.10}$$

où l'on a utilisé la notation $\overline{X} = (X_L + X_R)/2$ qui sera utilisée dans toute la suite.

Pour justifier que (4.10) est bien consistant avec l'équation différentielle (4.7), supposons que $w_L = w(x)$ et $w_R = w(x + \Delta x)$, où w est une fonction régulière. De même, on suppose que $Z_L = Z(x)$ et $Z_R = Z(x + \Delta x)$, avec Z une fonction régulière. L'équation (4.10) devient alors

$$\frac{h(x+\Delta x)^2\theta(x+\Delta x)}{2} - \frac{h(x)^2\theta(x)}{2} + \frac{h(x)+h(x+\Delta x)}{2}\frac{\theta(x)+\theta(x+\Delta x)}{2}\left(Z(x+\Delta x) - Z(x)\right) = 0$$

Un développement limité nous donne

$$\partial_x(h^2\theta/2)(x) + h(x)\theta(x)\partial_x Z(x) = O(\Delta x)$$

et l'équation (4.10) est bien une approximation locale à l'ordre un de l'équation différentielle (4.7).

On remarque qu'avec la Définition 4.2, on retrouve l'état d'équilibre remarquable (4.8) si et seulement si la moyenne $\overline{h}$ est définie comme la moyenne arithmétique entre h_L et h_R. En effet, si $\theta_L = \theta_R$, la relation (4.10) se réécrit

$$\begin{cases} u_L = u_R = 0, \\ \theta_L = \theta_R, \\ \frac{h_L+h_R}{2}[h] + \overline{h}[Z] = 0. \end{cases}$$

On en déduit alors $[h] = -[Z]$, ce qui permet de retrouver (4.8). On retrouve également l'état d'équilibre (4.9) quelque soit le choix des moyennes $\overline{h}$ et $\overline{\theta}$. En effet, si $Z_L = Z_R$, l'équation (4.10) se réécrit

$$\begin{cases} u_L = u_R = 0, \\ Z_L = Z_R, \\ h_L^2\theta_L = h_R^2\theta_R. \end{cases}$$

On a choisit également de définir $\overline{\theta}$ par la moyenne arithmétique de θ_L et θ_R.

4.1.3 Équations d'Euler avec gravité

Les équations d'Euler avec gravité permettent de décrire la dynamique des fluides non visqueux soumis à un champ de gravité que l'on supposera constant. Les inconnues présentes dans le système sont la densité $\rho(x,t)$, la vitesse $u(x,t)$, l'énergie totale $E(x,t)$ et la pression $p(x,t)$. L'énergie totale du système s'écrit

$$E = \rho e + \rho u^2/2,$$

où e désigne l'énergie interne. En une dimension d'espace, les équations d'Euler avec gravité s'écrivent

$$\begin{cases} \partial_t \rho + \partial_x \rho u = 0, \\ \partial_t \rho u + \partial_x \left(\rho u^2 + p\right) = -g\rho, \\ \partial_t E + \partial_x (u(E+p)) = -g\rho u, \end{cases} \tag{4.11}$$

où g est la constante de gravité. Pour fermer le système, on considère une équation d'état générale

$$p = p(\rho, e). \tag{4.12}$$

Afin d'assurer l'hyperbolicité, on suppose que la loi de pression vérifie

$$\partial_\rho p + \frac{p}{\rho^2}\partial_e p > 0.$$

On définit alors la vitesse du son par

$$c^2 = \partial_\rho p + \frac{p}{\rho^2}\partial_e p.$$

Le vecteur des variables conservatives $w = (\rho, \rho u, E)^T$ appartient à l'ensemble convexe des états admissibles

$$\Omega = \left\{ w \in \mathbb{R}^3, \rho > 0, E - \frac{1}{2}\rho u^2 > 0 \right\}.$$

Pour pouvoir écrire le terme source sous la forme (4.1), on introduit la fonction $Z(x) = x$, ce qui nous donne le système

$$\begin{cases} \partial_t \rho + \partial_x \rho u = 0, \\ \partial_t \rho u + \partial_x \left(\rho u^2 + p\right) = -g\rho\partial_x Z, \\ \partial_t E + \partial_x (u(E+p)) = -g\rho u\partial_x Z. \end{cases}$$

Le vecteur des grandeurs physiques est alors $U = (\rho, u, p, Z)^T$ et l'on peut réécrire les équations d'Euler avec gravité sous la forme quasi-linéaire

$$\partial_t U + A(U)\partial_x U = 0,$$

où la matrice $A(U)$ est donnée par

$$A(U) = \begin{pmatrix} u & \rho & 0 & 0 \\ 0 & u & 1/\rho & g \\ 0 & \rho c^2 & u & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Cette matrice admet quatre valeurs propres :

$$0, \quad u, \quad u \pm c,$$

avec pour vecteurs propres respectifs

$$\begin{pmatrix} \rho \\ -u \\ \rho c^2 \\ \frac{u^2-c^2}{g} \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \rho \\ \pm c \\ \rho c^2 \\ 0 \end{pmatrix}.$$

Les vecteurs propres sont distincts en tout point de Ω sauf aux points vérifiant $u = \pm c$ pour lesquels la base de diagonalisation est perdue. Notons que quand $u = 0$, la valeur propre 0 devient double, mais la matrice $A(U)$ reste diagonalisable. Le système est donc hyperbolique en tout point de Ω, sauf aux points pour lesquels $u = \pm c$. Les champs associés aux valeurs propres $u \pm c$ sont vraiment non linéaires, alors que ceux associés aux valeurs propres 0 et u sont linéairement dégénérés.

Les états d'équilibre pour les équations d'Euler avec gravité sont décrits par le système d'équations différentielles

$$\begin{cases} \partial_x \rho u = 0, \\ \partial_x(\rho u^2 + p) = -g\rho\partial_x Z, \\ \partial_x u(E + p) = -g\rho u \partial_x Z. \end{cases}$$

Ce système n'est pas intégrable et l'on se limite aux états d'équilibre au repos gouvernés par l'équation

$$\begin{cases} u \equiv 0, \\ \partial_x p + g\rho\partial_x Z = 0. \end{cases} \tag{4.13}$$

Cette équation n'est pas non plus intégrable et l'on ne peut pas expliciter tous les états d'équilibre.

Il reste à définir la notion d'équilibre local qui correspond à une discrétisation « à l'ordre un » de l'équation (4.13).

Définition 4.3 (Équilibre local pour Euler avec gravité)**.** Deux états (w_L, Z_L) et (w_R, Z_R) sont dits à l'équilibre local pour le système d'Euler avec gravité si

$$\begin{cases} u_L = u_R = 0, \\ [p] + g\overline{\rho}[Z] = 0. \end{cases} \tag{4.14}$$

De même que pour Ripa, on montre aisément la consistance de (4.14) avec (4.13). En effet, supposons que $w_L = w(x)$, $w_R = w(x + \Delta x)$, $Z_L = Z(x)$ et $Z_R = Z(x + \Delta x)$, où w et Z sont des fonctions régulières. L'équation (4.14) devient alors

$$p(x + \Delta x) - p(x) + g\frac{\rho(x) + \rho(x + \Delta x)}{2}\left(Z(x + \Delta x) - Z(x)\right) = 0.$$

Un développement limité nous donne

$$\partial_x p(x) + g\rho(x)\partial_x Z(x) = O(\Delta x).$$

L'équation (4.14) est donc bien une approximation locale à l'ordre un de l'équation différentielle (4.13).

4.2 Schémas volumes finis pour des systèmes avec terme source

Les méthodes de volumes finis pour le système avec terme source (4.1) diffèrent légèrement de celles introduites dans le Chapitre 1. Par soucis de complétude, on présente brièvement ces méthodes qui sont des extensions des méthodes usuelles, sans terme source. On commence par introduire le formalisme général des méthodes de volumes finis. Puis on définit les solveurs de Riemann approchés et on étudie les schémas de type Godunov qui en résultent. Enfin, en s'appuyant sur l'équilibre local qui a été défini pour chaque modèle dans la Partie 4.1, on déduit naturellement une notion globale de solution discrète stationnaire : une approximation de la solution du système (4.1) est une solution discrète stationnaire constante par morceaux si tous les couples d'états successifs sont à l'équilibre local. Un schéma est alors dit well-balanced s'il préserve toutes les solutions discrètes stationnaires constantes par morceaux. Par ailleurs, on souligne que cette nouvelle définition implique la préservation exacte des solutions stationnaires analytiques usuelles pour Saint-Venant et Ripa.

4.2.1 Présentation des méthodes de volumes finis pour des systèmes avec terme source

On considère une discrétisation de l'espace $\mathbb{R}$ en une suite croissante de points $(x_{i+1/2})_{i\in\mathbb{Z}}$ que l'on suppose uniforme, c'est-à-dire $x_{i+1/2} - x_{i-1/2} = \Delta x$, où Δx est le pas d'espace. On définit alors les volumes de contrôle $K_i = [x_{i-1/2}, x_{i+1/2}[$ et on note $x_i = (x_{i-1/2} + x_{i+1/2})/2$ le milieu de la cellule K_i. On discrétise également le temps en introduisant $t^{n+1} = t^n + \Delta t$, où Δt est le pas de temps.

On introduit l'approximation suivante de la fonction Z :

$$Z_i = \frac{1}{\Delta x}\int_{K_i} Z(x)dx, \quad \forall i \in \mathbb{Z}.$$

Au temps t^n, on note w_i^n une approximation de la solution de (4.1) sur la cellule K_i. On adopte la mise à jour au temps t^{n+1} par un schéma volumes finis donnée par

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(F_{i+1/2}^n - F_{i-1/2}^n\right) + \frac{\Delta t}{2}\left(S(w_{i-1}^n, w_i^n)\frac{Z_i - Z_{i-1}}{\Delta x} + S(w_i^n, w_{i+1}^n)\frac{Z_{i+1} - Z_i}{\Delta x}\right), \quad (4.15)$$

où le flux numérique est défini par

$$F_{i+1/2}^n = F\left(w_i^n, Z_i, w_{i+1}^n, Z_{i+1}\right). \quad (4.16)$$

Ici, $F : \Omega \times \mathbb{R} \times \Omega \times \mathbb{R} \to \mathbb{R}^d$ désigne la fonction flux numérique et $S : \Omega \times \Omega \to \mathbb{R}^d$ désigne la fonction terme source numérique. La dépendance en Z dans le flux numérique F n'est pas naturelle et alourdit les notations. Elle est cependant indispensable afin d'éviter les confusions dans la suite. Formellement, Z participe à la viscosité relative au flux numérique.

Le flux numérique F est dit consistant avec f si

$$F(w, Z, w, Z) = f(w), \quad \forall w \in \Omega, \forall Z \in \mathbb{R}.$$

Le terme source numérique S est dit consistant avec s si

$$S(w, w) = s(w), \quad \forall w \in \Omega.$$

Le schéma numérique (4.15) est alors consistant avec (4.1) si le flux numérique F et le terme source numérique S sont tous les deux consistants.

La présence du terme source ne change pas la définition de la robustesse du schéma : le schéma (4.15) est dit robuste si

$$\forall i \in \mathbb{Z}, \quad w_i^n \in \Omega \quad \Rightarrow \quad \forall i \in \mathbb{Z}, \quad w_i^{n+1} \in \Omega.$$

On s'intéresse maintenant à une classe particulière de schémas volumes finis : les schémas de type Godunov.

4.2.2 Schémas de type Godunov en présence d'un terme source

On considère le problème de Riemann

$$\begin{cases} \partial_t w + \partial_x f(w) = s(w)\partial_x Z, \\ w(x,0) = \begin{cases} w_L & \text{si } x < 0, \\ w_R & \text{si } x > 0, \end{cases} \end{cases} \tag{4.17}$$

où la topographie Z est une fonction discontinue donnée de la forme suivante :

$$Z(x) = \begin{cases} Z_L & \text{si } x < 0, \\ Z_R & \text{si } x > 0. \end{cases} \tag{4.18}$$

On insiste sur le fait que le terme source $s(w)\partial_x Z$ est présent dans le problème de Riemann considéré, ce qui diffère de la présentation qui a été faite dans le Chapitre 1. On note $W_{\mathcal{R}}\left(\frac{x}{t}, w_L, Z_L, w_R, Z_R\right)$ la solution exacte du problème de Riemann (4.17)–(4.18) et on définit $\lambda^{\pm}(w_L, Z_L, w_R, Z_R)$ la plus petite et la plus grande vitesse d'onde apparaissant dans $W_{\mathcal{R}}$.

La présence du terme source ne change pas la définition d'un solveur de Riemann approché, mis à part qu'il dépend maintenant de Z_L et Z_R. Pour être complet, on redonne la définition en tenant compte de cette modification.

Définition 4.4. On appelle solveur de Riemann approché pour le système (4.1) une fonction $\widetilde{W} : \mathbb{R} \times \Omega \times \mathbb{R} \times \Omega \times \mathbb{R} \to \mathbb{R}^d$ telle que :

(i) il existe des vitesses $\widetilde{\lambda}^- < \lambda^-$ et $\widetilde{\lambda}^+ > \lambda^+$ telles que

$$\widetilde{W}(\xi, w_L, Z_L, w_R, Z_R) = \begin{cases} w_L & \text{si } \xi < \widetilde{\lambda}^-, \\ \widetilde{W}(\xi, w_L, Z_L, w_R, Z_R) & \text{si } \widetilde{\lambda}^- < \xi < \widetilde{\lambda}^+, \\ w_R & \text{si } \xi > \widetilde{\lambda}^+. \end{cases}$$

(ii) si la condition CFL

$$\frac{\Delta t}{\Delta x} \max |\widetilde{\lambda}^{\pm}(w_L, Z_L, w_R, Z_R)| \leq \frac{1}{2} \tag{4.19}$$

est vérifiée, alors

$$\frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx = \frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} W_{\mathcal{R}}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx. \tag{4.20}$$

(iii) le solveur approché vérifie $\widetilde{W}(\xi, w, Z, w, Z) = w$ pour tout $\xi \in \mathbb{R}$, pour tout $w \in \Omega$ et pour tout $Z \in \mathbb{R}$.

On peut calculer la moyenne de la solution exacte du problème de Riemann. En effet, en intégrant l'équation (4.17) sur le rectangle $[-\Delta x/2, \Delta x/2] \times [0 \times \Delta t]$, on trouve

$$\frac{1}{\Delta x}\int_{-\Delta x/2}^{\Delta x/2} W_{\mathcal{R}}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx = \frac{1}{2}(w_L + w_R) - \frac{\Delta x}{\Delta t}\left(f(w_R) - f(w_L)\right) + \frac{1}{\Delta x}\int_{-\Delta x/2}^{\Delta x/2}\int_0^{\Delta t} s\left(W_{\mathcal{R}}\left(\frac{x}{t}, w_L, Z_L, w_R, Z_R\right)\right)\partial_x Z(x) dt dx.$$

La propriété (4.20) se réécrit alors de la façon suivante :

Lemme 4.5. *Supposons que la condition CFL (4.19) est vérifiée. Alors l'équation (4.20) est équivalente à la consistance avec la forme intégrale :*

$$\frac{1}{\Delta x}\int_{-\Delta x/2}^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx = \frac{1}{2}(w_L + w_R) - \frac{\Delta t}{\Delta x}\left(f(w_R) - f(w_L)\right) + \frac{1}{\Delta x}\int_{-\Delta x/2}^{\Delta x/2}\int_0^{\Delta t} s\left(W_{\mathcal{R}}\left(\frac{x}{t}, w_L, Z_L, w_R, Z_R\right)\right)\partial_x Z(x) dt dx. \quad (4.21)$$

En général, il n'est pas possible de calculer de façon exacte l'intégrale du terme source. On introduit alors une approximation de cette intégrale :

$$\frac{1}{\Delta x \Delta t}\int_{-\Delta x/2}^{\Delta x/2}\int_0^{\Delta t} s\left(W_{\mathcal{R}}\left(\frac{x}{t}, w_L, Z_L, w_R, Z_R\right)\right)\partial_x Z(x) dt dx \simeq S(w_L, w_R)\frac{[Z]}{\Delta x}. \quad (4.22)$$

Cela nous donne une version approchée de (4.21) :

$$\frac{1}{\Delta x}\int_{-\Delta x/2}^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx = \frac{1}{2}(w_L + w_R) - \frac{\Delta t}{\Delta x}\left(f(w_R) - f(w_L)\right) + \Delta t S(w_L, w_R)\frac{[Z]}{\Delta x}. \quad (4.23)$$

Par abus de langage, on appelle encore solveur de Riemann approché pour (4.17) une fonction $\widetilde{W}$ qui vérifie la Définition 4.4, dans laquelle on a remplacé l'équation (4.20) par (4.23).

Maintenant que l'on a défini un solveur de Riemann approché, on construit un schéma de type Godunov en suivant la procédure habituelle. Considérons une approximation de (4.1) au temps t^n :

$$W^n_{\Delta x}(x) = w^n_i, \quad \text{si } x \in K_i = [x_{i-1/2}, x_{i+1/2}[.$$

On note

$$W_{\Delta x}(x, t^n + t) = \widetilde{W}\left(\frac{x - x_{i+1/2}}{t}, w^n_i, Z_i, w^n_{i+1}, Z_{i+1}\right), \quad \text{si } x \in [x_i, x_{i+1}[,$$

la juxtaposition des solutions approchés des problèmes de Riemann. Pour éviter toute interaction entre les solveurs approchés, on impose la condition CFL

$$\frac{\Delta t}{\Delta x}\max_{i\in\mathbb{Z}} |\widetilde{\lambda}^{\pm}\left(w^n_i, Z_i, w^n_{i+1}, Z_{i+1}\right)| \leq \frac{1}{2}. \quad (4.24)$$

On définit alors la solution approchée au temps t^{n+1} en projetant $W_{\Delta x}$ sur l'espace des fonctions constantes sur chaque K_i :

$$w_i^{n+1} = \frac{1}{\Delta x} \int_{K_i} W_{\Delta x}\left(x, t^n + \Delta t\right) dx,$$

que l'on peut réécrire

$$w_i^{n+1} = \frac{1}{\Delta x} \int_0^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_{i-1}^n, Z_{i-1}, w_i^n, Z_i\right) dx + \frac{1}{\Delta x} \int_{-\Delta x/2}^0 \widetilde{W}\left(\frac{x}{\Delta t}, w_i^n, Z_i, w_{i+1}^n, Z_{i+1}\right) dx. \tag{4.25}$$

On dit que le schéma défini par (4.25) est un schéma de type Godunov associé au solveur de Riemann approché $\widetilde{W}$.

Un schéma de type Godunov peut s'écrire sous la forme d'un schéma volumes finis (4.15)–(4.16).

Proposition 4.6. *Sous la condition CFL (4.24), le schéma de type Godunov (4.25) peut s'écrire sous la forme (4.15)–(4.16), où le flux numérique est défini par*

$$\begin{aligned} F(w_L, Z_L, w_R, Z_R) = \frac{1}{2}(f(w_L) + f(w_R)) \\ - \frac{\Delta x}{4\Delta t}\left(w_R - \frac{2}{\Delta x}\int_0^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx \right. \\ \left. -w_L + \frac{2}{\Delta x}\int_{-\Delta x/2}^0 \widetilde{W}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx \right) \end{aligned} \tag{4.26}$$

et le terme source numérique vérifie (4.22). De plus le flux numérique F est consistant avec f.

Notons qu'en l'absence de terme source, on retrouve la formulation symétrique du flux numérique (1.22) introduite dans le Chapitre 1.

La forme précise de la discrétisation du terme source sera donnée plus tard. Cependant, remarquons dès à présent que la consistance d'un schéma de type Godunov (4.25) est vérifiée dès que la propriété suivante est établie :

$$S(w, w) = s(w), \quad \forall w \in \Omega.$$

Démonstration. Pour simplifier les notations, on introduit

$$U_{i+1/2}^- = \frac{2}{\Delta x} \int_{-\Delta x/2}^0 \widetilde{W}\left(\frac{x}{\Delta t}, w_i^n, Z_i, w_{i+1}^n, Z_{i+1}\right) dx,$$

$$U_{i+1/2}^+ = \frac{2}{\Delta x} \int_0^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_i^n, Z_i, w_{i+1}^n, Z_{i+1}\right) dx.$$

On peut écrire (4.25) sous la forme

$$w_i^{n+1} = \frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_{i-1}^n, Z_{i-1}, w_i^n, Z_i\right) dx + \frac{1}{2}\left(U_{i+1/2}^- - U_{i-1/2}^-\right),$$

ce qui nous donne en utilisant la consistance avec la forme intégrale (4.23)

$$\begin{aligned} w_i^{n+1} = \frac{1}{2}\left(w_{i-1}^n + w_i^n\right) - \frac{\Delta t}{\Delta x}\left(f\left(w_i^n\right) - f\left(w_{i-1}^n\right)\right) + \frac{1}{2}\left(U_{i+1/2}^- - U_{i-1/2}^-\right) \\ + \Delta t S(w_{i-1}^n, w_i^n)\frac{Z_i - Z_{i-1}}{\Delta x}. \end{aligned} \tag{4.27}$$

De la même façon, on a

$$w_i^{n+1} = \frac{1}{\Delta x}\int_{-\Delta x/2}^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_i^n, Z_i, w_{i+1}^n, Z_{i+1}\right) dx + \frac{1}{2}\left(U_{i-1/2}^+ - U_{i+1/2}^+\right),$$

dont on déduit grâce à la consistance avec la forme intégrale (4.23)

$$w_i^{n+1} = \frac{1}{2}\left(w_i^n + w_{i+1}^n\right) - \frac{\Delta t}{\Delta x}\left(f\left(w_{i+1}^n\right) - f\left(w_i^n\right)\right) + \frac{1}{2}\left(U_{i-1/2}^+ - U_{i+1/2}^+\right) + \Delta t S(w_i^n, w_{i+1}^n)\frac{Z_{i+1} - Z_i}{\Delta x}. \quad (4.28)$$

En faisant la demi-somme de (4.27) et (4.28), on obtient

$$\begin{aligned} w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\Bigg(&\frac{1}{2}(f(w_i^n) + f(w_{i+1}^n) - \frac{\Delta x}{4\Delta t}\left(w_{i+1}^n - U_{i+1/2}^+ - w_i^n + U_{i+1/2}^-\right) \\ &-\frac{1}{2}(f(w_{i-1}^n) + f(w_i^n)) + \frac{\Delta x}{4\Delta t}\left(w_i^n - U_{i-1/2}^+ - w_{i-1}^n + U_{i-1/2}^-\right)\Bigg) \\ &+ \frac{\Delta t}{2}\left(S(w_{i-1}^n, w_i^n)\frac{Z_i - Z_{i-1}}{\Delta x} + S(w_i^n, w_{i+1}^n)\frac{Z_{i+1} - Z_i}{\Delta x}\right). \end{aligned}$$

En définissant le flux numérique F par (4.26), on obtient exactement le schéma (4.15).

Par ailleurs, en utilisant l'hypothèse (iii) de la définition 4.4, on voit aisément que le flux numérique F défini par (4.26) est consistant. □

Le lemme qui suit donne une condition simple pour qu'un schéma de type Godunov soit robuste.

Lemme 4.7. *Supposons que la condition CFL (4.24) est vérifiée. Si pour tous w_L et w_R dans Ω et pour tous Z_L et Z_R dans $\mathbb{R}$, le solveur de Riemann approché $\widetilde{W}(\xi, w_L, Z_L, w_R, Z_R)$ est à valeurs dans Ω, pour tout $\xi \in \mathbb{R}$, alors le schéma de type Godunov associé à $\widetilde{W}$ est robuste.*

Démonstration. Le schéma de type Godunov associé à $\widetilde{W}$ est défini par (4.25) qui est la moyenne d'une fonction à valeurs dans Ω. L'ensemble Ω étant supposé convexe, w_i^{n+1} reste donc dans Ω. □

4.2.3 Solutions discrètes stationnaires et schémas well-balanced

Pour les trois systèmes étudiés, on a défini une notion d'équilibre local entre deux états. On définit maintenant une notion globale de solution discrète stationnaire. Cela consiste à demander que tous les couples d'états consécutifs soient à l'équilibre local.

Définition 4.8 (Solution discrète stationnaire constante par morceaux)**.** Si $\forall i \in \mathbb{Z}$, les états (w_i^n, Z_i) et (w_{i+1}^n, Z_{i+1}) sont à l'équilibre local (Définitions 4.1, 4.2 ou 4.3 selon le système considéré), alors on dit que l'approximation $(w_i^n)_{i\in\mathbb{Z}}$ définit une solution discrète stationnaire constante par morceaux.

Dans le cas du système de Ripa (4.2) ou des équations d'Euler avec gravité (4.3), cette définition revient à demander que $(w_i^n)_{i\in\mathbb{Z}}$ soit une approximation consistante de l'équation différentielle (4.7) ou (4.13), respectivement. La définition de schéma d'ordre un well-balanced est alors naturelle.

Définition 4.9 (Schéma d'ordre un well-balanced)**.** Le schéma (4.15) est dit well-balanced si pour toute solution discrète stationnaire constante par morceaux $(w_i^n)_{i\in\mathbb{Z}}$, le schéma vérifie

$$\forall i \in \mathbb{Z}, \quad w_i^{n+1} = w_i^n.$$

Il est clair que la définition usuelle des schémas well-balanced pour Saint-Venant entre trivialement dans cette généralisation. En fait, la définition 4.9 permet d'étendre la notion well-balanced pour des systèmes dont les états d'équilibre recherchés ne sont pas naturellement donnés par des relations algébriques.

Il reste à trouver une condition simple pour qu'un schéma de type Godunov soit well-balanced. On commence par définir un solveur de Riemann approché well-balanced.

Définition 4.10 (Solveur de Riemann approché well-balanced)**.** Soit $\widetilde{W}(\xi, w_L, Z_L, w_R, Z_R)$ un solveur de Riemann approché. Celui-ci est dit well-balanced si pour tous états (w_L, Z_L) et (w_R, Z_R) à l'équilibre local (Définitions 4.1, 4.2 ou 4.3 selon le système considéré), le solveur de Riemann vérifie

$$\widetilde{W}(\xi, w_L, Z_L, w_R, Z_R) = \begin{cases} w_L & \text{si } \xi < 0, \\ w_R & \text{si } \xi > 0. \end{cases}$$

En d'autres termes, si la donnée initiale est à l'équilibre local, alors le solveur de Riemann approché doit exactement préserver cet équilibre local.

On a alors la proposition suivante :

Proposition 4.11. *Soit $\widetilde{W}$ un solveur de Riemann approché well-balanced. Alors le schéma de type Godunov associé à $\widetilde{W}$ est well-balanced.*

Démonstration. Le schéma de type Godunov associé au solveur $\widetilde{W}$ est défini par (4.25). Supposons que $(w_i^n)_{i\in\mathbb{Z}}$ est une solution discrète stationnaire constante par morceaux. Alors les états (w_{i-1}^n, Z_{i-1}) et (w_i^n, Z_i) d'une part et les états (w_i^n, Z_i) et (w_{i+1}^n, Z_{i+1}) d'autre part sont à l'équilibre local. Puisque le solveur de Riemann approché est well-balanced, on a

$$\widetilde{W}\left(\frac{x}{\Delta t}, w_{i-1}^n, Z_{i-1}, w_i^n, Z_i\right) = w_i^n, \quad \forall x \in [0, \Delta x/2]$$

et

$$\widetilde{W}\left(\frac{x}{\Delta t}, w_i^n, Z_i, w_{i+1}^n, Z_{i+1}\right) = w_i^n, \quad \forall x \in [-\Delta x/2, 0].$$

On en déduit immédiatement $w_i^{n+1} = w_i^n$. Le schéma de type Godunov associé à $\widetilde{W}$ est donc bien well-balanced. □

L'objectif est maintenant de construire des solveurs de Riemann approchés qui soient well-balanced.

4.3 Construction de solveurs simples de Riemann well-balanced

Dans cette partie, on construit des solveurs de Riemann approchés well-balanced en suivant le formalisme HLL [64]. On va de plus se concentrer sur des solveurs simples (voir Gallice [52, 53] et Chalons [29]), c'est-à-dire constitués de N états intermédiaires constants (voir Figure 4.1). Les états intermédiaires sont des inconnues et il faut déterminer un nombre égal d'équations. Ces équations doivent assurer que le solveur de Riemann vérifie la consistance avec la forme intégrale et qu'il est well-balanced. De plus, il est souhaitable que le système d'équations obtenu soit suffisamment simple pour que sa résolution reste à un coût de calcul raisonnable.

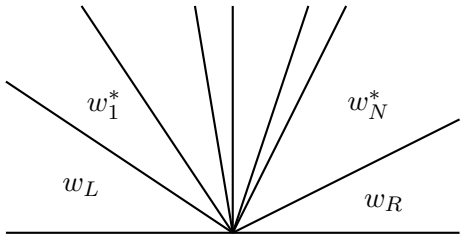


FIGURE 4.1 – Solveur de Riemann simple

Dans un premier temps, on développe un solveur de Riemann approché pour le système de Ripa pour obtenir un schéma well-balanced robuste. Cependant, malgré la simplicité de cette approche, le solveur obtenu pour Ripa ne pourra pas être directement étendu pour les équations d'Euler avec gravité. On introduit alors une modification du solveur pour Ripa qui rend possible l'extension aux équations d'Euler.

4.3.1 Un premier schéma well-balanced pour le système de Ripa

On considère ici le système de Ripa (4.6) pour lequel on développe un solveur de Riemann approché well-balanced suivant la Définition 4.10.

Le solveur de Riemann

Suivant l'étude du système de Ripa présentée dans la partie précédente, on considère un solveur de Riemann approché $\widetilde{W}(\xi, w_L, Z_L, w_R, Z_R)$ simple contenant quatre ondes de vitesse λ_L, 0, $\widehat{u}$ et λ_R séparant trois états intermédiaires constants w_L^*, w_0^* et w_R^*. On choisit les vitesses d'onde de manière à avoir $\lambda_L < 0 < \lambda_R$ et $\lambda_L < \widehat{u} < \lambda_R$. Par contre, la vitesse $\widehat{u}$ qui sera définie plus tard peut être positive ou négative (voir Figure 4.2).

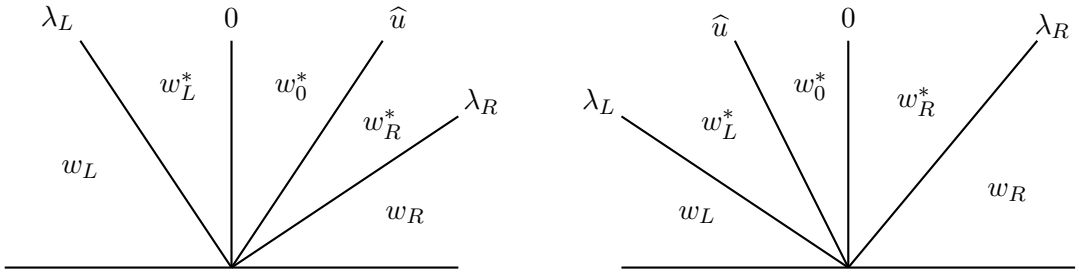


FIGURE 4.2 – Solveur de Riemann simple pour le système de Ripa. Gauche : cas $\widehat{u} > 0$. Droite : cas $\widehat{u} < 0$

Dans un premier temps, on remarque que le système de Ripa transporte simplement la température θ à la vitesse du fluide u. En effet, le système (4.6) nous donne aisément l'équation

$$\partial_t \theta + u \partial_x \theta = 0.$$

Par conséquent, il est naturel de fixer les inconnues en θ de la façon suivante :

$$\theta_L^* = \theta_L, \quad \theta_R^* = \theta_R, \quad \theta_0^* = \begin{cases} \theta_L, & \text{si } \widehat{u} > 0, \\ \theta_R, & \text{si } \widehat{u} < 0, \end{cases}$$

où $\widehat{u}$ est une approximation correcte de la vitesse u que l'on détaillera plus tard.

D'autre part, afin de simplifier le problème, on choisit de considérer h et u constants à travers l'onde de vitesse $\widehat{u}$. Il y a donc deux valeurs inconnues de h,

$$h_L^* \quad \text{pour } \lambda_L < \frac{x}{t} < 0,$$

$$h_R^* \quad \text{pour } 0 < \frac{x}{t} < \lambda_R,$$

et deux valeurs inconnues de u,

$$u_L^* \quad \text{pour } \lambda_L < \frac{x}{t} < 0,$$

$$u_R^* \quad \text{pour } 0 < \frac{x}{t} < \lambda_R.$$

Il nous faut donc quatre équations supplémentaires (voir Figure 4.3).

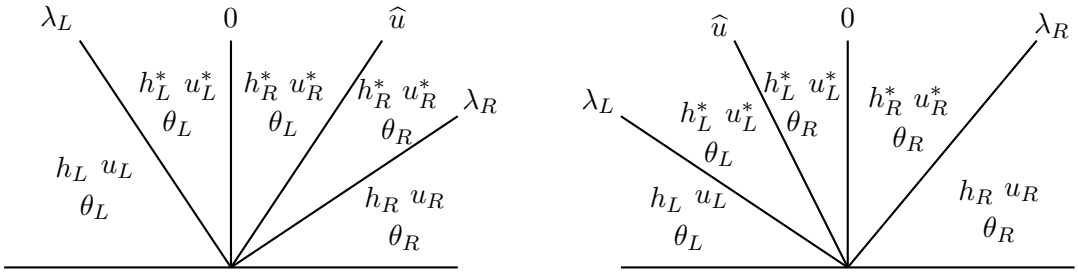


FIGURE 4.3 – Inconnues dans le solveur de Riemann simple pour le système de Ripa. Gauche : cas $\widehat{u} > 0$. Droite : cas $\widehat{u} < 0$

Les deux première équations nous sont fournies par la définition 4.4 d'un solveur de Riemann approché. En effet, sous la condition CFL

$$\frac{\Delta t}{\Delta x} \max(|\lambda_L|, |\lambda_R|) \leq \frac{1}{2}, \tag{4.29}$$

le solveur $\widetilde{W}$ doit vérifier

$$\frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx = \frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} W_{\mathcal{R}}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx, \tag{4.30}$$

où $W_{\mathcal{R}}$ est la solution exacte du problème de Riemann pour (4.6). D'après le Lemme 4.5, puisqu'il n'y a pas de terme source dans l'équation en h, on obtient directement

$$\frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} \widetilde{h}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx = \overline{h} - \frac{\Delta t}{\Delta x}[hu]. \tag{4.31}$$

Concernant la quantité de mouvement hu, le Lemme 4.5 nous donne

$$\begin{aligned}\frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} \widetilde{hu}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx = \overline{hu} - \frac{\Delta t}{\Delta x}\left[hu^2 + gh^2\theta/2\right] \\ - \frac{g}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} \int_0^{\Delta t} (h\theta)_{\mathcal{R}}\left(\frac{x}{t}, w_L, Z_L, w_R, Z_R\right) \partial_x Z dt dx.\end{aligned} \tag{4.32}$$

De plus, on a aisément

$$\begin{aligned}\frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} \widetilde{h}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx = \left(\frac{1}{2} + \lambda_L \Delta t\right) h_L - \lambda_L \Delta t h_L^* + \lambda_R \Delta t h_R^* \\ + \left(\frac{1}{2} - \lambda_R \Delta t\right) h_R,\end{aligned} \tag{4.33}$$

$$\frac{1}{\Delta x}\int_{-\Delta x/2}^{\Delta x/2}\widetilde{hu}\left(\frac{x}{\Delta t},w_L,Z_L,w_R,Z_R\right)dx=\left(\frac{1}{2}+\lambda_L\Delta t\right)h_Lu_L-\lambda_L\Delta t h_L^*u_L^*+\lambda_R\Delta t h_R^*u_R^* + \left(\frac{1}{2}-\lambda_R\Delta t\right)h_Ru_R. \quad (4.34)$$

De l'égalité entre (4.31) et (4.33) d'une part et entre (4.32) et (4.34) d'autre part, il résulte les deux relations suivantes :

$$\lambda_L(h_L-h_L^*)+\lambda_R(h_R^*-h_R)=-\left[hu\right], \quad (4.35)$$

$$\lambda_L\left(h_Lu_L-h_L^*u_L^*\right)+\lambda_R\left(h_R^*u_R^*-h_Ru_R\right)=-\left[hu^2+gh^2\theta/2\right] - \frac{g}{\Delta t}\int_{-\Delta x/2}^{\Delta x/2}\int_0^{\Delta t}(h\theta)_{\mathcal{R}}\left(\frac{x}{t},w_L,Z_L,w_R,Z_R\right)\partial_xZdtdx. \quad (4.36)$$

Comme mentionné dans la Partie 4.2.2, il est difficile de calculer exactement l'intégrale du terme source et l'on va utiliser une approximation. On montre qu'une fois la définition des états stationnaires choisie, il n'y a qu'une seule approximation de cette intégrale qui permette d'obtenir un solveur well-balanced. En effet, supposons momentanément que le solveur soit well-balanced et considérons deux états (w_L,Z_L) et (w_R,Z_R) à l'équilibre local. Nous avons donc $u_L=u_R=0$ et puisque le solveur est supposé well-balanced, $u_L^*=u_R^*=0$, $h_L^*=h_L$ et $h_R^*=h_R$. L'équation (4.36) nous donne alors

$$\frac{g}{\Delta t}\int_{-\Delta x/2}^{\Delta x/2}\int_0^{\Delta t}(h\theta)_{\mathcal{R}}\left(\frac{x}{t},w_L,Z_L,w_R,Z_R\right)\partial_xZdtdx=-\left[gh^2\theta/2\right].$$

De plus, puisque (w_L,Z_L) et (w_R,Z_R) vérifient l'équilibre local (4.10), on a

$$\left[h^2\theta/2\right]=-\bar{h}\bar{\theta}[Z].$$

On en déduit

$$\frac{1}{\Delta t}\int_{-\Delta x/2}^{\Delta x/2}\int_0^{\Delta t}(h\theta)_{\mathcal{R}}\left(\frac{x}{t},w_L,Z_L,w_R,Z_R\right)\partial_xZdtdx=\bar{h}\bar{\theta}[Z].$$

Cette relation est une égalité uniquement dans le cas où le solveur est supposé well-balanced et pour des états (w_L,Z_L) et (w_R,Z_R) à l'équilibre local. Dans le cas général, il est par conséquent naturel d'approcher l'intégrale du terme source de la façon suivante :

$$\frac{1}{\Delta t}\int_{-\Delta x/2}^{\Delta x/2}\int_0^{\Delta t}(h\theta)_{\mathcal{R}}\left(\frac{x}{t},w_L,Z_L,w_R,Z_R\right)\partial_xZdtdx\approx\bar{h}\bar{\theta}[Z]. \quad (4.37)$$

Notons que cela revient à définir le terme source numérique

$$S^{hu}(w_L,w_R)=-g\bar{h}\bar{\theta} \quad (4.38)$$

et à adopter la consistance approchée (4.23). On a immédiatement

$$\begin{aligned}S^{hu}(w,w)&=-gh\theta\\&=s^{hu}(w),\end{aligned}$$

ce qui montre la consistance du terme source numérique.

Afin de déterminer les quatre inconnues h_L^*, h_R^*, u_L^* et u_R^* du solveur de Riemann, il manque encore deux équations. On propose tout d'abord de préserver la continuité de l'invariant de Riemann hu à travers l'onde de vitesse 0, ce qui donne l'équation

$$h_L^* u_L^* = h_R^* u_R^*. \tag{4.39}$$

Enfin pour la dernière équation, on adopte une version « linéarisée » de la définition de l'équilibre local (4.10) :

$$\frac{h_R h_R^* \theta_R}{2} - \frac{h_L h_L^* \theta_L}{2} + \bar{h}\bar{\theta}[Z] = 0. \tag{4.40}$$

Cette dernière équation participe à assurer le caractère well-balanced du solveur. D'autres choix pourraient être faits, comme par exemple la version non linéarisée qui mènerait également à un solveur well-balanced, mais le système d'équations obtenu serait alors non linéaire et par conséquent beaucoup plus délicat à résoudre.

Pour simplifier les notations, on introduit les états intermédiaires du solveur HLL ([64])

$$h^{HLL} = \frac{\lambda_R h_R - \lambda_L h_L}{\lambda_R - \lambda_L} - \frac{1}{\lambda_R - \lambda_L}\left[hu\right] \tag{4.41}$$

et

$$q^{HLL} = \frac{\lambda_R h_R u_R - \lambda_L h_L u_L}{\lambda_R - \lambda_L} - \frac{1}{\lambda_R - \lambda_L}\left[hu^2 + gh^2\theta/2\right]. \tag{4.42}$$

On obtient alors le système composé des quatre équations (4.35), (4.36), (4.39), (4.40) qui se réécrit de la façon suivante :

$$\frac{\lambda_R h_R^* - \lambda_L h_L^*}{\lambda_R - \lambda_L} = h^{HLL}, \tag{4.43}$$

$$\frac{\lambda_R h_R^* u_R^* - \lambda_L h_L^* u_L^*}{\lambda_R - \lambda_L} = q^{HLL} - \frac{g}{\lambda_R - \lambda_L}\bar{h}\bar{\theta}[Z], \tag{4.44}$$

$$h_L^* u_L^* = h_R^* u_R^*, \tag{4.45}$$

$$\frac{h_R h_R^* \theta_R}{2} - \frac{h_L h_L^* \theta_L}{2} + \bar{h}\bar{\theta}[Z] = 0. \tag{4.46}$$

On introduit le moment intermédiaire

$$q^* = h_L^* u_L^* = h_R^* u_R^*.$$

On peut alors déterminer l'expression de q^* directement à partir de l'équation (4.44) :

$$q^* = q^{HLL} - \frac{g}{\lambda_R - \lambda_L}\bar{h}\bar{\theta}[Z]. \tag{4.47}$$

D'autre part, les équations (4.43) et (4.46) forment un système linéaire à deux inconnues qui nous donne

$$h_L^* = \frac{(\lambda_R - \lambda_L)\theta_R h_R h^{HLL} + 2\lambda_R \bar{h}\bar{\theta}[Z]}{\theta_L h_L \lambda_R - \theta_R h_R \lambda_L}, \tag{4.48}$$

$$h_R^* = \frac{(\lambda_R - \lambda_L)\theta_L h_L h^{HLL} + 2\lambda_L \bar{h}\bar{\theta}[Z]}{\theta_L h_L \lambda_R - \theta_R h_R \lambda_L}. \tag{4.49}$$

On en déduit alors les vitesses intermédiaires par

$$u_L^* = \frac{q^*}{h_L^*} \quad \text{et} \quad u_R^* = \frac{q^*}{h_R^*}.$$

Pour compléter le solveur de Riemann approché il reste à déterminer la vitesse de l'onde centrale $\widehat{u}$. Remarquons que l'on n'a pas encore assuré la consistance pour $h\theta$ et cette propriété

va permettre de déterminer $\widehat{u}$. Puisque l'équation en $h\theta$ ne contient pas de terme source, la consistance avec la forme intégrale s'écrit, d'après le Lemme 4.5,

$$\frac{1}{\Delta x}\int_{-\Delta x/2}^{\Delta x/2} \widetilde{h\theta}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx = \overline{h\theta} - \frac{\Delta t}{\Delta x}[hu\theta].$$

Supposons dans un premier temps que $\widehat{u} > 0$. Alors la moyenne en $h\theta$ du solveur approché est

$$\begin{aligned}\frac{1}{\Delta x}\int_{-\Delta x/2}^{\Delta x/2} \widetilde{W}^{h\theta}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx = &\left(\frac{1}{2} - \lambda_L\frac{\Delta t}{\Delta x}\right) h_L\theta_L - \lambda_L\frac{\Delta t}{\Delta x}h_L^*\theta_L + \widehat{u}\frac{\Delta t}{\Delta x}h_R^*\theta_L \\ &+ (\lambda_R - \widehat{u})\frac{\Delta t}{\Delta x}h_R^*\theta_R + \left(\frac{1}{2} - \lambda_R\frac{\Delta t}{\Delta x}\right) h_R\theta_R.\end{aligned}$$

Des deux dernières équations, on déduit la relation suivante :

$$h_Ru_R\theta_R - h_Lu_L\theta_L = \lambda_R\theta_R(h_R - h_R^*) + \lambda_L\theta_L(h_L^* - h_L) + \widehat{u}h_R^*(\theta_R - \theta_L). \tag{4.50}$$

En utilisant l'équation (4.35), on trouve

$$\widehat{u} = \frac{h_Lu_L + \lambda_L(h_L^* - h_L)}{h_R^*}.$$

Un calcul similaire montre que si $\widehat{u} < 0$, alors

$$\widehat{u} = \frac{h_Lu_L + \lambda_L(h_L^* - h_L)}{h_L^*}.$$

En supposant que les hauteurs intermédiaires h_L^* et h_R^* sont strictement positives (ce qui sera montré plus tard), alors $\widehat{u}$ est définie par

$$\widehat{u} = \begin{cases} \dfrac{h_Lu_L + \lambda_L(h_L^* - h_L)}{h_R^*}, & \text{si } h_Lu_L + \lambda_L(h_L^* - h_L) > 0, \\ \dfrac{h_Lu_L + \lambda_L(h_L^* - h_L)}{h_L^*}, & \text{sinon.} \end{cases} \tag{4.51}$$

Le solveur de Riemann approché $\widetilde{W}(\xi, w_L, Z_L, w_R, Z_R)$ est donc entièrement déterminé.

Le schéma complet

On suppose connue une approximation de la solution au temps t^n constante par morceaux :

$$W_{\Delta x}(x) = w_i^n, \quad \text{si } x \in K_i.$$

On rappelle que pour un schéma de type Godunov, la mise à jour est donnée par (4.25). Pour les composantes gouvernées par des équations sans terme source (h et $h\theta$), d'après la Proposition 1.6, le schéma se réécrit sous forme conservative

$$h_i^{n+1} = h_i^n - \frac{\Delta t}{\Delta x}\left(F^h\left(w_i^n, Z_i, w_{i+1}^n, Z_{i+1}\right) - F^h\left(w_{i-1}^n, Z_{i-1}, w_i^n, Z_i\right)\right),$$

$$(h\theta)_i^{n+1} = (h\theta)_i^n - \frac{\Delta t}{\Delta x}\left(F^{h\theta}\left(w_i^n, Z_i, w_{i+1}^n, Z_{i+1}\right) - F^{h\theta}\left(w_{i-1}^n, Z_{i-1}, w_i^n, Z_i\right)\right),$$

où les flux numériques F^h et $F^{h\theta}$ sont définis par

$$F^h(w_L, Z_L, w_R, Z_R) = f^h(w_L) + \frac{\Delta x}{2\Delta t}h_L - \frac{1}{\Delta t}\int_{-\Delta x/2}^{0} \widetilde{h}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx, \tag{4.52}$$

$$F^{h\theta}(w_L, Z_L, w_R, Z_R) = f^{h\theta}(w_L) + \frac{\Delta x}{2\Delta t} h_L\theta_L - \frac{1}{\Delta t}\int_{-\Delta x/2}^{0} \widetilde{h\theta}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx. \quad (4.53)$$

On déduit de l'équation (4.52) la formulation suivante du flux en h :

$$\begin{aligned} F^h(w_L, Z_L, w_R, Z_R) &= h_L u_L + \frac{\Delta x}{2\Delta t} h_L - \left(\frac{\Delta x}{2\Delta t} + \lambda_L\right) h_L + \lambda_L h_L^* \\ &= h_L u_L + \lambda_L (h_L^* - h_L). \end{aligned} \quad (4.54)$$

Pour calculer le flux en $h\theta$, remarquons que le signe de $\widehat{u}$ est le même que le signe de $F^h(w_L, Z_L, w_R, Z_R)$ (toujours en supposant que $h_L^* > 0$ et $h_R^* > 0$). Si $F^h(w_L, Z_L, w_R, Z_R) > 0$, on déduit de (4.53) que le flux en $h\theta$ s'écrit :

$$\begin{aligned} F^{h\theta}(w_L, Z_L, w_R, Z_R) &= h_L u_L\theta_L + \frac{\Delta x}{2\Delta t} h_L\theta_L - \left(\frac{\Delta x}{2\Delta t} + \lambda_L\right) h_L\theta_L + \lambda_L h_L^*\theta_L \\ &= \theta_L F^h(w_L, Z_L, w_R, Z_R). \end{aligned}$$

De même, si $F^h(w_L, Z_L, w_R, Z_R) < 0$, en utilisant la définition (4.51), on a

$$\begin{aligned} F^{h\theta}(w_L, Z_L, w_R, Z_R) &= h_L u_L\theta_L + \frac{\Delta x}{2\Delta t} h_L\theta_L - \left(\frac{\Delta x}{2\Delta t} + \lambda_L\right) h_L\theta_L + (\lambda_L - \widehat{u}) h_L^*\theta_L + \widehat{u} h_L^*\theta_R \\ &= \theta_L F^h(w_L, Z_L, w_R, Z_R) + \widehat{u} h_L^*(\theta_R - \theta_L) \\ &= \theta_R F^h(w_L, Z_L, w_R, Z_R). \end{aligned}$$

On introduit la notation

$$\theta_{LR} = \begin{cases} \theta_L & \text{si } F^h(w_L, Z_L, w_R, Z_R) > 0, \\ \theta_R & \text{si } F^h(w_L, Z_L, w_R, Z_R) < 0. \end{cases}$$

Le flux en $h\theta$ se réécrit alors sous la forme :

$$F^{h\theta}(w_L, Z_L, w_R, Z_R) = \theta_{LR} F^h(w_L, Z_L, w_R, Z_R). \quad (4.55)$$

Concernant hu, la présence du terme source nous oblige à utiliser la formulation (4.15)–(4.16). En utilisant la Proposition 4.6, on peut écrire le flux numérique sous la forme

$$\begin{aligned} F^{hu}(w_L, Z_L, w_R, Z_R) &= \overline{hu^2 + gh^2\theta/2} \\ &\quad - \frac{\Delta x}{4\Delta t}\left(h_R u_R - \frac{2}{\Delta x}\int_0^{\Delta x/2} \widetilde{hu}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx \right. \\ &\qquad \left. - h_L u_L + \frac{2}{\Delta x}\int_{-\Delta x/2}^{0} \widetilde{hu}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx\right) \end{aligned}$$

Un calcul immédiat donne

$$\frac{2}{\Delta x}\int_0^{\Delta x/2} \widetilde{hu}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx = 2\lambda_R \frac{\Delta t}{\Delta x} q^* + \left(1 - 2\lambda_R \frac{\Delta t}{\Delta x}\right) h_R u_R,$$

$$\frac{2}{\Delta x}\int_{-\Delta x/2}^{0} \widetilde{hu}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx = \left(1 + 2\lambda_L \frac{\Delta t}{\Delta x}\right) h_L u_L - 2\lambda_L \frac{\Delta t}{\Delta x} q^*.$$

On déduit des trois dernières équations

$$F^{hu}(w_L, Z_L, w_R, Z_R) = \overline{hu^2 + gh^2\theta/2} + \frac{\lambda_L}{2}(q^* - h_L u_L) + \frac{\lambda_R}{2}(q^* - h_R u_R). \quad (4.56)$$

On peut donc écrire le schéma complet :

$$\begin{cases} h_i^{n+1} = h_i^n - \dfrac{\Delta t}{\Delta x}\left(F_{i+1/2}^h - F_{i-1/2}^h\right), \\ h_i^{n+1}u_i^{n+1} = h_i^n u_i^n - \dfrac{\Delta t}{\Delta x}\left(F_{i+1/2}^{hu} - F_{i-1/2}^{hu}\right) \\ \qquad\qquad + \dfrac{\Delta t}{2}\left(S^{hu}(w_{i-1}^n, w_i^n)\dfrac{Z_i - Z_{i-1}}{\Delta x} + S^{hu}(w_i^n, w_{i+1}^n)\dfrac{Z_{i+1} - Z_i}{\Delta x}\right), \\ h_i^{n+1}\theta_i^{n+1} = h_i^n\theta_i^n - \dfrac{\Delta t}{\Delta x}\left(\theta_{i+1/2}^n F_{i+1/2}^h - \theta_{i+1/2}^n F_{i-1/2}^h\right), \end{cases} \tag{4.57}$$

où l'on a posé

$$F_{i+1/2} = F(w_i^n, Z_i, w_{i+1}^n, Z_{i+1}).$$

Les composantes du flux numérique sont définies par (4.54) pour h et par (4.56) pour hu. Le terme source numérique est défini par (4.38) et l'on a introduit

$$\theta_{i+1/2}^n = \begin{cases} \theta_i^n & \text{si } F^h(w_i^n, w_{i+1}^n) > 0, \\ \theta_{i+1}^n & \text{si } F^h(w_i^n, w_{i+1}^n) < 0. \end{cases}$$

Propriétés du schéma

On commence par établir que le schéma est well-balanced.

Proposition 4.12. *Le schéma numérique (4.57) est well-balanced, c'est-à-dire que si $\forall i \in \mathbb{Z}$ on a*

$$u_i^n = 0 \quad \textit{et} \quad \frac{(h_{i+1}^n)^2\theta_{i+1}^n}{2} - \frac{(h_i^n)^2\theta_i^n}{2} + \frac{h_i^n + h_{i+1}^n}{2}\frac{\theta_i^n + \theta_{i+1}^n}{2}(Z_{i+1} - Z_i) = 0,$$

alors $w_i^{n+1} = w_i^n$, $\forall i \in \mathbb{Z}$.

Démonstration. D'après le Lemme 4.11, il suffit de montrer que le solveur de Riemann approché $\widetilde{W}$ est well-balanced. On considère deux états (w_L, Z_L) et (w_R, Z_R) à l'équilibre local défini par (4.10). D'après (4.41), on a

$$h^{HLL} = \frac{\lambda_R h_R - \lambda_L h_L}{\lambda_R - \lambda_L}.$$

En utilisant (4.48), on en déduit

$$h_L^* = \frac{\theta_R h_R(\lambda_R h_R - \lambda_L h_L) + 2\lambda_R \bar{h}\bar{\theta}[Z]}{\lambda_R\theta_L h_L - \lambda_L\theta_R h_R}.$$

En utilisant la définition de l'équilibre local, on trouve

$$\begin{aligned} h_L^* &= \frac{\theta_R h_R(\lambda_R h_R - \lambda_L h_L) + \lambda_R\left(h_L^2\theta_L - h_R^2\theta_R\right)}{\lambda_R\theta_L h_L - \lambda_L\theta_R h_R} \\ &= h_L. \end{aligned}$$

De même, on trouve $h_R^* = h_R$. D'autre part, en utilisant la définition (4.42) de q^{HLL}, on a

$$q^{HLL} = \frac{g}{2(\lambda_R - \lambda_L)}\left(h_L^2\theta_L - h_R^2\theta_R\right).$$

La définition (4.47) et l'équilibre local permettent de conclure que $q^* = 0$. On en déduit que $u_L^* = u_R^* = 0$. Enfin, concernant θ, il suffit de montrer que $\widehat{u} = 0$, ce qui est le cas puisque $F^h(w_L, Z_L, w_R, Z_R) = 0$ d'après (4.54). □

Remarque. On souligne que le schéma (4.57) préserve en particulier les états d'équilibres remarquables (4.8) et (4.9) de façon exacte.

On s'intéresse maintenant à la robustesse du schéma.

Proposition 4.13. *Supposons que $\forall i \in \mathbb{Z}$, on choisit $\lambda^R_{i+1/2} - \lambda^L_{i-1/2}$ suffisamment grand pour que l'état intermédiaire du solveur HLL, défini par (4.41) soit dans Ω. On suppose également que $\forall i \in \mathbb{Z}$, on a :*

- *si $Z_{i+1} - Z_i > 0$, alors $\left|\frac{\lambda^R_{i+1/2}}{\lambda^L_{i+1/2}}\right|$ est suffisamment grand ;*
- *si $Z_{i+1} - Z_i < 0$, alors $\left|\frac{\lambda^L_{i+1/2}}{\lambda^R_{i+1/2}}\right|$ est suffisamment grand.*

Alors le schéma (4.57) est robuste, c'est-à-dire

$$\forall i \in \mathbb{Z}, \quad w_i^n \in \Omega \quad \Rightarrow \quad \forall i \in \mathbb{Z}, \quad w_i^{n+1} \in \Omega.$$

Démonstration. D'après le Lemme 4.7, la robustesse peut être montrée localement pour chaque solveur de Riemann approché. La température des états intermédiaires apparaissant dans le solveur de Riemann approché est soit θ_L, soit θ_R, qui sont tous les deux positifs par hypothèse. Il suffit donc de montrer que h_L^* et h_R^* sont strictement positifs. On peut réécrire ces états intermédiaires sous la forme suivante :

$$h_L^* = \frac{(\lambda_R - \lambda_L)\theta_R h_R}{\theta_L h_L \lambda_R - \theta_R h_R \lambda_L} h^{HLL} + \frac{2\bar{h}\bar{\theta}[Z]}{\theta_L h_L + \theta_R h_R \left|\frac{\lambda_L}{\lambda_R}\right|},$$

$$h_R^* = \frac{(\lambda_R - \lambda_L)\theta_L h_L}{\theta_L h_L \lambda_R - \theta_R h_R \lambda_L} h^{HLL} - \frac{2\bar{h}\bar{\theta}[Z]}{\theta_R h_R + \theta_L h_L \left|\frac{\lambda_R}{\lambda_L}\right|}.$$

Par hypothèse, on a $h^{HLL} > 0$, donc le premier terme des deux égalités précédentes est strictement positif.

Si $[Z] > 0$, alors on a immédiatement $h_L^* > 0$. De plus, on a

$$\frac{2\bar{h}\bar{\theta}[Z]}{\theta_R h_R + \theta_L h_L \left|\frac{\lambda_R}{\lambda_L}\right|} \to 0 \quad \text{quand} \quad \left|\frac{\lambda_R}{\lambda_L}\right| \to +\infty.$$

Donc pour $\left|\frac{\lambda_R}{\lambda_L}\right|$ suffisamment grand, h_R^* est également strictement positif.

De la même façon, si $[Z] < 0$, alors on a $h_R^* > 0$ et

$$\frac{2\bar{h}\bar{\theta}[Z]}{\theta_L h_L + \theta_R h_R \left|\frac{\lambda_L}{\lambda_R}\right|} \to 0 \quad \text{quand} \quad \left|\frac{\lambda_L}{\lambda_R}\right| \to +\infty.$$

Donc pour $\left|\frac{\lambda_L}{\lambda_R}\right|$ suffisamment grand, h_L^* est également strictement positif. □

Conclusion

On a construit un schéma numérique well-balanced et robuste pour le système de Ripa. Cependant, la condition sur les vitesses d'ondes permettant d'assurer la robustesse ne nous donne pas explicitement ces vitesses. On est donc amenés à choisir des vitesses éventuellement très grandes qui peuvent rendre le schéma trop diffusif. D'autre part, il n'est pas simple d'étendre directement cette approche au système d'Euler avec gravité. On va maintenant modifier le solveur que nous venons de construire pour permettre l'extension aux équations d'Euler avec gravité.

4.3.2 Un deuxième schéma well-balanced pour Ripa

L'idée ici est d'essayer de gagner un degré de liberté dans la dérivation du solveur de Riemann approché. Pour cela, on souhaite ne pas faire intervenir l'équation (4.40) obtenue par une linéarisation *ad hoc* de l'équilibre local. Celle-ci pourra alors être réintroduite pour l'extension aux équations d'Euler avec gravité dans la partie suivante.

Le solveur de Riemann

On considère un solveur de Riemann approché $\widetilde{W}(\xi, w_L, Z_L, w_R, Z_R)$ de la même forme que précédemment (voir Figure 4.2). Les inconnues en θ sont fixées de la manière suivante :

$$\theta_L^* = \theta_L, \quad \theta_R^* = \theta_R, \quad \theta_0^* = \begin{cases} \theta_L, & \text{si } \widehat{u} > 0, \\ \theta_R, & \text{si } \widehat{u} < 0, \end{cases}$$

où $\widehat{u}$ est une approximation correcte de la vitesse u que l'on détaillera plus tard. On considère h et u constants à travers l'onde de vitesse $\widehat{u}$. On a donc quatre inconnues h_L^*, h_R^*, u_L^* et u_R^* dans le solveur de Riemann qui sont les mêmes que pour le solveur précédent (voir Figure 4.3) et on cherche donc quatre équations.

À la place de l'équation (4.43), on utilise la relation de saut de Rankine-Hugoniot pour l'équation sur h à travers chacune des trois ondes λ_L, 0 et λ_R. Cela nous donne les trois équations suivantes :

$$h_L^* u_L^* - h_L u_L = \lambda_L \left(h_L^* - h_L\right), \tag{4.58}$$

$$h_R^* u_R^* = h_L^* u_L^*, \tag{4.59}$$

$$h_R u_R - h_R^* u_R^* = \lambda_R \left(h_R - h_R^*\right). \tag{4.60}$$

Pour l'équation manquante, on conserve l'équation provenant de la consistance avec la forme intégrale (4.23) pour hu :

$$\frac{\lambda_R h_R^* u_R^* - \lambda_L h_L^* u_L^*}{\lambda_R - \lambda_L} = q^{HLL} - \frac{g}{\lambda_R - \lambda_L} \bar{h}\bar{\theta}[Z]. \tag{4.61}$$

En considérant l'équation (4.61), on a implicitement défini le terme source numérique par

$$S^{hu}(w_L, w_R) = -g\bar{h}\bar{\theta}. \tag{4.62}$$

On constate immédiatement que le terme source numérique S est consistant avec s.

Le premier solveur que l'on a construit était naturellement consistant avec la forme intégrale par le choix des équations. Pour ce solveur, il nous faut vérifier cette propriété.

Proposition 4.14. *Si le solveur de Riemann approché $\widetilde{W}$ vérifie les équations (4.58), (4.59), (4.60) et (4.61) et que la vitesse $\widehat{u}$ est définie par*

$$\widehat{u} = \begin{cases} \dfrac{h_L u_L + \lambda_L(h_L^* - h_L)}{h_R^*}, & \text{si } h_L u_L + \lambda_L(h_L^* - h_L) > 0, \\ \dfrac{h_L u_L + \lambda_L(h_L^* - h_L)}{h_L^*}, & \text{si } h_L u_L + \lambda_L(h_L^* - h_L) < 0, \end{cases} \tag{4.63}$$

alors $\widetilde{W}$ vérifie la consistance avec la forme intégrale (4.23).

Démonstration. La consistance en hu coïncide exactement avec l'équation (4.61). De plus, le choix de $\widehat{u}$ permet exactement d'obtenir la relation de consistance en $h\theta$ donnée par (4.50).

Il reste à montrer la consistance en h. On a

$$\begin{aligned}\frac{1}{\Delta x}\int_{-\Delta x/2}^{\Delta x/2}\widetilde{h}\left(\frac{x}{\Delta t},w_L,w_R\right)dx &= \left(\frac{1}{2}+\lambda_L\frac{\Delta t}{\Delta x}\right)h_L-\lambda_L\frac{\Delta t}{\Delta x}h_L^*+\lambda_R\frac{\Delta t}{\Delta x}h_R^*\\ &\quad+\left(\frac{1}{2}-\lambda_R\frac{\Delta t}{\Delta x}\right)h_R\\ &= \overline{h}-\frac{\Delta t}{\Delta x}\left(\lambda_R\left(h_R-h_R^*\right)+\lambda_L\left(h_L^*-h_L\right)\right).\end{aligned}$$

Les équations (4.58), (4.59) et (4.60) impliquent la séquence d'égalités suivante :

$$\begin{aligned}\lambda_R\left(h_R-h_R^*\right)+\lambda_L\left(h_L^*-h_L\right) &= h_Ru_R-h_R^*u_R^*+h_L^*u_L^*-h_Lu_L\\ &= [hu].\end{aligned}$$

Il en résulte immédiatement la consistance en h attendue. □

Résolution du système et schéma complet

Dans un premier temps, on résout le système donné par les équations (4.58), (4.59), (4.60) et (4.61) afin de déterminer complètement les états intermédiaires du solveur de Riemann approché. La résolution de $q^*=h_L^*u_L^*=h_R^*u_R^*$ est la même que pour le schéma précédent et, en utilisant (4.61), on trouve

$$q^*=q^{HLL}-\frac{g}{\lambda_R-\lambda_L}\bar{h}\bar{\theta}[Z].$$

Les équations (4.58) et (4.60) nous donnent alors

$$h_L^*=h_L+\frac{1}{\lambda_L}\left(q^*-h_Lu_L\right), \tag{4.64}$$

$$h_R^*=h_R+\frac{1}{\lambda_R}\left(q^*-h_Ru_R\right). \tag{4.65}$$

On en déduit simplement les vitesses

$$u_L^*=\frac{q^*}{h_L^*}\quad\text{et}\quad u_R^*=\frac{q^*}{h_R^*}.$$

En suivant la même procédure que pour le schéma précédent, on peut écrire le schéma complet sous la forme

$$\begin{cases}h_i^{n+1}=h_i^n-\dfrac{\Delta t}{\Delta x}\left(F_{i+1/2}^h-F_{i-1/2}^h\right),\\ h_i^{n+1}u_i^{n+1}=h_i^nu_i^n-\dfrac{\Delta t}{\Delta x}\left(F_{i+1/2}^{hu}-F_{i-1/2}^{hu}\right)\\ \qquad\qquad+\Delta t\left(S^{hu}\left(w_{i-1}^n,w_i^n\right)\dfrac{Z_i-Z_{i-1}}{\Delta x}+S^{hu}\left(w_i^n,w_{i+1}^n\right)\dfrac{Z_{i+1}-Z_i}{\Delta x}\right),\\ h_i^{n+1}\theta_i^{n+1}=h_i^n\theta_i^n-\dfrac{\Delta t}{\Delta x}\left(\theta_{i+1/2}^nF_{i+1/2}^h-\theta_{i+1/2}^nF_{i-1/2}^h\right),\end{cases} \tag{4.66}$$

où l'on a posé

$$F_{i+1/2}=F(w_L,Z_L,w_R,Z_L).$$

Les composantes du flux numérique sont définies par

$$F^h(w_L, Z_L, w_R, Z_L) = h_L u_L + \lambda_L(h_L^* - h_L),$$

$$F^{hu}(w_L, Z_L, w_R, Z_R) = \overline{hu^2 + gh^2\theta/2} + \frac{\lambda_L}{2}(q^* - h_L u_L) + \frac{\lambda_R}{2}(q^* - h_R u_R)$$

et le terme source numérique est défini par (4.62). Avec cette écriture, les schémas (4.57) et (4.66) paraissent identiques. Ils diffèrent en fait simplement sur la définition des hauteurs intermédiaires h_L^* et h_R^*, données à présent par (4.64) et (4.65).

Propriétés du schéma

L'intérêt principal de ce schéma est qu'il est well-balanced sans qu'on ait eu besoin d'utiliser l'équation définissant l'équilibre local.

Proposition 4.15. *Le schéma (4.66) est well-balanced, c'est-à-dire que si $\forall i \in \mathbb{Z}$ on a*

$$u_i^n = 0 \quad et \quad \frac{(h_{i+1}^n)^2\theta_{i+1}^n}{2} - \frac{(h_i^n)^2\theta_i^n}{2} + \frac{h_i^n + h_{i+1}^n}{2}\frac{\theta_i^n + \theta_{i+1}^n}{2}(Z_{i+1} - Z_i) = 0,$$

alors $w_i^{n+1} = w_i^n$, $\forall i \in \mathbb{Z}$.

Démonstration. Considérons deux états (w_L, Z_L) et (w_R, Z_R) à l'équilibre local. Puisque q^* est défini de la même façon que pour le schéma (4.57), on en déduit que $q^* = 0$. Comme $u_L = 0$ et $u_R = 0$, les équations (4.58) et (4.60) permettent de conclure que $h_L^* = h_L$ et $h_R^* = h_R$. La fin de la démonstration est identique à la preuve de la Proposition 4.12. □

On montre maintenant que le schéma (4.66) est robuste.

Proposition 4.16. *Si pour tout $i \in \mathbb{Z}$, on choisit $|\lambda_L|$ et λ_R suffisamment grands, alors le schéma (4.66) est robuste, c'est-à-dire*

$$\forall i \in \mathbb{Z}, \quad w_i^n \in \Omega \quad \Rightarrow \quad \forall i \in \mathbb{Z}, \quad w_i^{n+1} \in \Omega.$$

Démonstration. De la même manière que dans la preuve de la Proposition 4.13, il suffit de montrer que les hauteurs intermédiaires h_L^* et h_R^* sont strictement positives. En injectant la définition (4.47) de q^* dans (4.64) et (4.65), on peut écrire

$$h_L^* = h_L + \frac{\lambda_R}{\lambda_L}\frac{h_R u_R - h_L u_L}{\lambda_R - \lambda_L} - \frac{C}{\lambda_L(\lambda_R - \lambda_L)},$$

$$h_R^* = h_R + \frac{\lambda_L}{\lambda_R}\frac{h_R u_R - h_L u_L}{\lambda_R - \lambda_L} - \frac{C}{\lambda_R(\lambda_R - \lambda_L)},$$

où C est défini par

$$C = \left[hu^2 + gh^2\theta/2\right] + g\overline{h}\,\overline{\theta}[Z].$$

On définit alors $\alpha = \frac{\lambda_R}{\lambda_R - \lambda_L}$. Notons que α est toujours dans l'intervalle $[0, 1]$. Les hauteurs intermédiaires h_L^* et h_R^* deviennent

$$h_L^* = h_L + \frac{\alpha}{\lambda_L}[hu] + \frac{(1-\alpha)C}{\lambda_L^2},$$

$$h_R^* = h_R - \frac{1-\alpha}{\lambda_R}[hu] - \frac{\alpha C}{\lambda_R^2}.$$

On voit facilement que h_L^* converge vers $h_L > 0$ quand λ_L tend vers $-\infty$ et que h_R^* converge vers $h_R > 0$ quand λ_R tend vers $+\infty$. En choisissant ces vitesses suffisamment grandes (en valeur absolue), les hauteurs h_L^* et h_R^* sont alors strictement positives. □

Conclusion

Le schéma (4.66) que nous venons de construire vérifie les mêmes propriétés que le schéma (4.57). Il est well-balanced et robuste. La positivité de h est même plus simple à préserver que dans le schéma (4.57) car la condition sur les vitesses d'ondes est moins contraignante. L'autre avantage du schéma (4.66), comme nous allons le voir maintenant, est qu'il peut facilement être étendu aux équations d'Euler avec gravité.

4.3.3 Extension aux équations d'Euler avec gravité

Le but de cette partie est de généraliser le schéma (4.66) aux équations d'Euler avec gravité (4.11). Pour simplifier le développement du schéma qui va suivre, on suppose dans cette partie que la pression suit la loi des gaz parfaits

$$p = (\gamma - 1)\left(E - \frac{1}{2}\rho u^2\right),$$

où $\gamma \in]1, 3]$ est le coefficient adiabatique du gaz. On insiste sur le fait que cette simplification est uniquement valable pour cette partie et qu'un schéma well-balanced pour une loi de pression générale sera développé plus tard.

Le solveur de Riemann

On considère un solveur de Riemann approché simple $\widetilde{W}(\xi, w_L, Z_L, w_R, Z_R)$ contenant trois ondes de vitesse $\lambda^L < 0 < \lambda^R$ séparant deux états intermédiaires. Il y a cette fois six inconnues : ρ_L^*, u_L^*, p_L^*, ρ_R^*, u_R^* et p_R^* (voir Figure 4.4). On doit donc déterminer six équations.

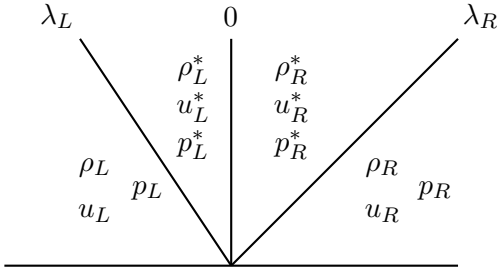


FIGURE 4.4 – Inconnues dans le solveur de Riemann simple pour Euler

Pour les inconnues ρ et u, nous allons utiliser la même approche que pour Ripa. On considère l'équation de consistance avec la formulation intégrale pour ρu qui s'écrit, d'après le Lemme 4.5,

$$\lambda_L\left(\rho_L u_L - \rho_L^* u_L^*\right) + \lambda_R\left(\rho_R^* u_R^* - \rho_R u_R\right) = -\left[\rho u^2 + p\right] - \frac{g}{\Delta t}\int_{-\Delta x/2}^{\Delta x/2}\int_0^{\Delta t} \rho_{\mathcal{R}}\left(\frac{x}{t}, w_L, Z_L, w_R, Z_R\right)\partial_x Z dt dx. \quad (4.67)$$

On cherche maintenant une approximation de l'intégrale du terme source. Pour cela, supposons momentanément que le schéma soit well-balanced. On considère alors deux états (w_L, Z_L) et (w_R, Z_R) à l'équilibre local défini par (4.14). L'équation (4.67) et la définition de l'équilibre local (4.14) entrainent immédiatement

$$\frac{1}{\Delta t}\int_{-\Delta x/2}^{\Delta x/2}\int_0^{\Delta t} \rho_{\mathcal{R}}\left(\frac{x}{t}, w_L, Z_L, w_R, Z_R\right)\partial_x Z dt dx = \overline{\rho}[Z].$$

Dans le cas général, on utilise par conséquent l'approximation suivante :

$$\frac{1}{\Delta t}\int_{-\Delta x/2}^{\Delta x/2}\int_{0}^{\Delta t}\rho_{\mathcal{R}}\left(\frac{x}{t},w_L,Z_L,w_R,Z_R\right)\partial_x Z dt dx \approx \overline{\rho}[Z].$$

Cela revient à choisir comme composante en ρu du terme source numérique

$$S^{\rho u}(w_L,w_R) = -g\overline{\rho}. \tag{4.68}$$

En utilisant cette définition, l'équation (4.67) devient

$$\lambda_L\left(\rho_L u_L - \rho_L^* u_L^*\right) + \lambda_R\left(\rho_R^* u_R^* - \rho_R u_R\right) = -\left[\rho u^2 + p\right] - g\bar{\rho}[Z].$$

On utilise ensuite la relation de saut de Rankine-Hugoniot pour ρ à travers chacune des trois ondes, ce qui nous donne les trois équations suivantes :

$$\rho_L^* u_L^* - \rho_L u_L = \lambda_L\left(\rho_L^* - \rho_L\right),$$

$$\rho_R^* u_R^* = \rho_L^* u_L^*,$$

$$\rho_R u_R - \rho_R^* u_R^* = \lambda_R\left(\rho_R - \rho_R^*\right).$$

Il reste à trouver deux équations pour l'énergie totale. L'équation de consistance avec la forme intégrale pour l'énergie est

$$\begin{aligned}\lambda_L\left(E_L - E_L^*\right) + \lambda_R\left(E_R^* - E_R\right) = &-\left[u(E+p)\right]\\ &-\frac{g}{\Delta t}\int_{-\Delta x/2}^{\Delta x/2}\int_{0}^{\Delta t}\rho_{\mathcal{R}}\left(\frac{x}{t},w_L,Z_L,w_R,Z_R\right)u_{\mathcal{R}}\left(\frac{x}{t},w_L,Z_L,w_R,Z_R\right)\partial_x Z dt dx.\end{aligned}$$

Le choix naturel pour approcher l'intégrale du terme source est

$$\frac{1}{\Delta t}\int_{-\Delta x/2}^{\Delta x/2}\int_{0}^{\Delta t}\rho_{\mathcal{R}}\left(\frac{x}{t},w_L,Z_L,w_R,Z_R\right)u_{\mathcal{R}}\left(\frac{x}{t},w_L,Z_L,w_R,Z_R\right)\partial_x Z dt dx \approx \bar{\rho}\bar{u}[Z],$$

Cela revient à choisir comme composante en E du terme source numérique

$$S^{E}(w_L,w_R) = -g\bar{\rho}\bar{u}, \tag{4.69}$$

ce qui nous donne la relation

$$\lambda_L\left(E_L - E_L^*\right) + \lambda_R\left(E_R^* - E_R\right) = -\left[u(E+p)\right] - g\bar{\rho}\bar{u}[Z].$$

Puisqu'il n'y a pas de terme source dans l'équation en ρ, on a défini complètement le terme source numérique :

$$S(w_L,w_R) = (0, -g\bar{\rho}, -g\bar{\rho}\bar{u})^T. \tag{4.70}$$

On voit immédiatement que S est consistant avec le terme source $s(w) = (0, -g\rho, -g\rho u)^T$.

Pour la dernière équation, on choisit la relation venant de la définition de l'équilibre local (4.14) en adoptant la linéarisation suivante :

$$p_R^* - p_L^* = -g\bar{\rho}[Z], \tag{4.71}$$

qui se réécrit en termes d'énergie

$$E_R^* - E_L^* = \rho_R^* {u_R^*}^2/2 - \rho_L^* {u_L^*}^2/2 - \frac{g\bar{\rho}[Z]}{\gamma - 1}.$$

Pour simplifier les notations, on introduit les états intermédiaires associés au solveur HLL

$$q^{HLL} = \frac{\lambda_R \rho_R u_R - \lambda_L \rho_L u_L}{\lambda_R - \lambda_L} - \frac{1}{\lambda_R - \lambda_L} \left[\rho u^2 + p\right],$$

$$E^{HLL} = \frac{\lambda_R E_R - \lambda_L E_L}{\lambda_R - \lambda_L} - \frac{1}{\lambda_R - \lambda_L} \left[u(E+p)\right]. \tag{4.72}$$

On obtient ainsi le système composé des six équations suivantes

$$\rho_L^* u_L^* - \rho_L u_L = \lambda_L \left(\rho_L^* - \rho_L\right), \tag{4.73}$$

$$\rho_R^* u_R^* = \rho_L^* u_L^*, \tag{4.74}$$

$$\rho_R u_R - \rho_R^* u_R^* = \lambda_R \left(\rho_R - \rho_R^*\right), \tag{4.75}$$

$$\frac{\lambda_R \rho_R^* u_R^* - \lambda_L \rho_L^* u_L^*}{\lambda_R - \lambda_L} = q^{HLL} - \frac{g}{\lambda_R - \lambda_L} \bar{\rho}[Z], \tag{4.76}$$

$$\frac{\lambda_R E_R^* - \lambda_L E_L^*}{\lambda_R - \lambda_L} = E^{HLL} - \frac{g}{\lambda_R - \lambda_L} \bar{\rho}\bar{u}[Z], \tag{4.77}$$

$$E_R^* - E_L^* = \rho_R^* {u_R^*}^2/2 - \rho_L^* {u_L^*}^2/2 - \frac{g\bar{\rho}[Z]}{\gamma - 1}. \tag{4.78}$$

Résolution du système et schéma complet

La résolution pour les inconnues ρ et u s'effectue de la même façon que pour Ripa. On définit $q^* = \rho_L^* u_L^* = \rho_R^* u_R^*$, le moment de l'état intermédiaire. À partir de l'équation (4.76), on trouve

$$q^* = q^{HLL} - \frac{g}{\lambda_R - \lambda_L} \bar{\rho}[Z]. \tag{4.79}$$

Les équations (4.73) et (4.75) impliquent

$$\rho_L^* = \rho_L + \frac{1}{\lambda_L}(q^* - \rho_L u_L), \tag{4.80}$$

$$\rho_R^* = \rho_R + \frac{1}{\lambda_R}(q^* - \rho_R u_R). \tag{4.81}$$

Concernant les énergies intermédiaires, le système composé de (4.77) et (4.78) est un système linéaire de deux équations à deux inconnues E_L^* et E_R^* dont la résolution nous donne

$$E_L^* = E^{HLL} + \frac{\lambda_R}{\lambda_R - \lambda_L} \left(\rho_L^* \left(u_L^*\right)^2/2 - \rho_R^* \left(u_R^*\right)^2/2\right) + \frac{g\bar{\rho}[Z]}{\lambda_R - \lambda_L} \left(\frac{\lambda_R}{\gamma - 1} - \bar{u}\right),$$

$$E_R^* = E^{HLL} + \frac{\lambda_L}{\lambda_R - \lambda_L} \left(\rho_L^* \left(u_L^*\right)^2/2 - \rho_R^* \left(u_R^*\right)^2/2\right) + \frac{g\bar{\rho}[Z]}{\lambda_R - \lambda_L} \left(\frac{\lambda_L}{\gamma - 1} - \bar{u}\right).$$

On en déduit les pressions correspondantes

$$\begin{aligned} \frac{p_L^*}{\gamma - 1} &= E_L^* - \rho_L^* {u_L^*}^2/2 \\ &= E^{HLL} + \frac{\lambda_L \rho_L^* {u_L^*}^2/2 - \lambda_R \rho_R^* {u_R^*}^2/2}{\lambda_R - \lambda_L} + \frac{g\bar{\rho}[Z]}{\lambda_R - \lambda_L} \left(\frac{\lambda_R}{\gamma - 1} - \bar{u}\right) \end{aligned} \tag{4.82}$$

$$\begin{aligned}\frac{p_R^*}{\gamma-1} &= E_R^* - \rho_R^* {u_R^*}^2/2 \\ &= E^{HLL} + \frac{\lambda_L \rho_L^* {u_L^*}^2/2 - \lambda_R \rho_R^* {u_R^*}^2/2}{\lambda_R - \lambda_L} + \frac{g\bar{\rho}[Z]}{\lambda_R - \lambda_L}\left(\frac{\lambda_L}{\gamma-1} - \bar{u}\right). \end{aligned} \tag{4.83}$$

On passe maintenant au schéma numérique associé au solveur de Riemann approché $\widetilde{W}$ qui est défini par (4.25). Pour la composante ρ, l'absence de terme source nous permet d'utiliser la Proposition 1.6 afin d'écrire le flux numérique sous la forme

$$\begin{aligned} F^\rho(w_L, Z_L, w_R, Z_R) &= \rho_L u_L + \frac{\Delta x}{2\Delta t}\rho_L - \frac{1}{\Delta t}\int_{-\Delta x/2}^0 \widetilde{\rho}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx \\ &= \rho_L u_L + \frac{\Delta x}{2\Delta t}\rho_L - \left(\frac{\Delta x}{2\Delta t} + \lambda_L\right)\rho_L + \lambda_L \rho_L^* \\ &= \rho_L u_L + \lambda_L(\rho_L^* - \rho_L). \end{aligned} \tag{4.84}$$

Concernant ρu, la présence du terme source nous oblige à utiliser la formulation (4.15)–(4.16). D'après la Proposition 4.6, le flux numérique s'écrit

$$\begin{aligned} F^{\rho u}(w_L, Z_L, w_R, Z_R) = \overline{\rho u^2 + p} - \frac{\Delta x}{4\Delta t}\Bigg(&\rho_R u_R - \frac{2}{\Delta x}\int_0^{\Delta x/2} \widetilde{\rho u}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx \\ &-\rho_L u_L + \frac{2}{\Delta x}\int_{-\Delta x/2}^0 \widetilde{\rho u}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx\Bigg). \end{aligned}$$

les deux intégrales se calculent aisément :

$$\begin{aligned} \frac{2}{\Delta x}\int_0^{\Delta x/2} \widetilde{\rho u}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx &= 2\lambda_R \frac{\Delta t}{\Delta x} q^* + \left(1 - 2\lambda_R \frac{\Delta t}{\Delta x}\right)\rho_R u_R, \\ \frac{2}{\Delta x}\int_{-\Delta x/2}^0 \widetilde{\rho u}\left(\frac{x}{\Delta t}, w_L, Z_L, w_R, Z_R\right) dx &= \left(1 + 2\lambda_L \frac{\Delta t}{\Delta x}\right)\rho_L u_L - 2\lambda_L \frac{\Delta t}{\Delta x} q^*. \end{aligned}$$

On déduit des trois dernières équations

$$F^{\rho u}(w_L, Z_L, w_R, Z_R) = \overline{\rho u^2 + p} + \frac{\lambda_L}{2}(q^* - \rho_L u_L) + \frac{\lambda_R}{2}(q^* - \rho_R u_R). \tag{4.85}$$

Enfin, pour la composante E, un calcul similaire donne

$$F^E(w_L, Z_L, w_R, Z_R) = \overline{u(E+p)} + \frac{\lambda_L}{2}(E_L^* - E_L) + \frac{\lambda_R}{2}(E_R^* - E_R). \tag{4.86}$$

On peut donc écrire le schéma complet :

$$\begin{cases} \rho_i^{n+1} = \rho_i^n - \dfrac{\Delta t}{\Delta x}\left(F^\rho_{i+1/2} - F^\rho_{i-1/2}\right), \\ \rho_i^{n+1} u_i^{n+1} = \rho_i^n u_i^n - \dfrac{\Delta t}{\Delta x}\left(F^{\rho u}_{i+1/2} - F^{\rho u}_{i-1/2}\right) \\ \qquad\qquad + \dfrac{\Delta t}{2}\left(S^{\rho u}\left(w_{i-1}^n, w_i^n\right)\dfrac{Z_i - Z_{i-1}}{\Delta x} + S^{\rho u}\left(w_i^n, w_{i+1}^n\right)\dfrac{Z_{i+1} - Z_i}{\Delta x}\right), \\ E_i^{n+1} = E_i^n - \dfrac{\Delta t}{\Delta x}\left(F^E_{i+1/2} - F^E_{i-1/2}\right) \\ \qquad\qquad + \dfrac{\Delta t}{2}\left(S^E\left(w_{i-1}^n, w_i^n\right)\dfrac{Z_i - Z_{i-1}}{\Delta x} + S^E\left(w_i^n, w_{i+1}^n\right)\dfrac{Z_i - Z_{i-1}}{\Delta x}\right), \end{cases} \tag{4.87}$$

où l'on a posé

$$F_{i+1/2} = F(w_L, Z_L, w_R, Z_R).$$

Les composantes du flux numériques sont définies par (4.84) pour ρ, par (4.85) pour ρu et par (4.86) pour E. Le terme source numérique $S(w_L, w_R)$ est défini par (4.70).

Propriétés du schéma

On commence par établir que le schéma est well-balanced.

Proposition 4.17. *Le schéma (4.87) est well-balanced, c'est-à-dire que si $\forall i \in \mathbb{Z}$, on a*

$$u_i^n = 0 \quad et \quad p_{i+1}^n - p_i^n + g\frac{\rho_i^n + \rho_{i+1}^n}{2}(Z_{i+1} - Z_i) = 0,$$

alors $w_i^{n+1} = w_i^n$, $\forall i \in \mathbb{Z}$.

Démonstration. On considère deux états (w_L, Z_L) et (w_R, Z_R) à l'équilibre local, défini par (4.14). Les vitesses u_L et u_R étant nulles, on a immédiatement

$$q^{HLL} = -\frac{1}{\lambda_R - \lambda_L}[p].$$

En utilisant l'équilibre local (4.14), on en déduit

$$\begin{aligned} q^* &= -\frac{1}{\lambda_R - \lambda_L}[p] - \frac{g}{\lambda_R - \lambda_L}\overline{\rho}[Z] \\ &= 0. \end{aligned}$$

Les équations (4.80) et (4.81) entrainent alors $\rho_L^* = \rho_L$ et $\rho_R^* = \rho_R$. Il reste à montrer que l'on a $p_L^* = p_L$ et $p_R^* = p_R$. Puisque toutes les vitesses sont nulles, on a, d'après l'équation (4.82),

$$p_L^* = \frac{\lambda_R p_R - \lambda_L p_L}{\lambda_R - \lambda_L} + \frac{\lambda_R}{\lambda_R - \lambda_L} g\bar{\rho}[Z].$$

En utilisant à nouveau l'équilibre local (4.14), on en déduit $p_L^* = p_L$. De la même façon, on trouve $p_R^* = p_R$. Le schéma est donc well-balanced. □

La positivité de la densité ne présente pas de difficulté. En effet, la preuve est la même que pour montrer la positivité de h pour le modèle de Ripa.

Proposition 4.18. *Si pour tout $i \in \mathbb{Z}$, on choisit $|\lambda_L|$ et λ_R suffisamment grands, alors le schéma (4.87) préserve la positivité de ρ, c'est-à-dire que si pour tout $i \in \mathbb{Z}$, on a $\rho_i^n > 0$ et $p_i^n > 0$, alors $\rho_i^{n+1} > 0$.*

Il reste à montrer que le schéma préserve la positivité de la pression.

Proposition 4.19. *Si pour tout $i \in \mathbb{Z}$, on choisit $|\lambda_L|$, λ_R et $\left|\frac{\lambda_R}{\lambda_L}\right|$ suffisamment grands, alors le schéma (4.87) préserve la positivité de p, c'est-à-dire que si pour tout $i \in \mathbb{Z}$, on a $\rho_i^n > 0$ et $p_i^n > 0$, alors $p_i^{n+1} > 0$.*

Avant de montrer cette propriété, remarquons que par rapport à la Proposition 4.13, il manque a priori une condition sur $\left|\frac{\lambda_L}{\lambda_R}\right|$. Cela est dû au fait que pour les équations d'Euler, la gravité est toujours orientée dans le même sens, autrement dit, on a toujours $Z_R - Z_L > 0$ (car $Z(x) = x$), alors que pour Ripa, le signe du terme de topographie $Z_R - Z_L$ n'est pas fixe.

Démonstration. Remarquons tout d'abord que d'après (4.78), on a $p_R^* < p_L^*$, car $Z(x) = x$ et donc $[Z] > 0$. Il suffit alors de montrer que $p_R^* > 0$.

On définit le coefficient $\alpha = \frac{\lambda_R}{\lambda_R - \lambda_L}$. On a alors $1-\alpha = -\frac{\lambda_L}{\lambda_R - \lambda_L}$ et $\alpha \in [0,1]$. Dans un premier temps, on considère la limite quand $\lambda_L \to -\infty$ et $\lambda_R \to +\infty$ tout en gardant α constant. D'après l'équation (4.72), on a

$$\lim_{\substack{\lambda_L \to -\infty \\ \lambda_R \to +\infty \\ \alpha = \text{cst}}} E^{HLL} = \alpha E_R + (1-\alpha) E_L.$$

D'autre part, on a d'après (4.79)

$$\lim_{\substack{\lambda_L \to -\infty \\ \lambda_R \to +\infty \\ \alpha = \text{cst}}} q^* = \alpha \rho_R u_R + (1-\alpha) \rho_L u_L$$

et d'après (4.80) et (4.81), on a

$$\lim_{\substack{\lambda_L \to -\infty \\ \lambda_R \to +\infty \\ \alpha = \text{cst}}} \rho_L^* = \rho_L,$$

$$\lim_{\substack{\lambda_L \to -\infty \\ \lambda_R \to +\infty \\ \alpha = \text{cst}}} \rho_R^* = \rho_R.$$

En utilisant (4.83), on déduit de toutes ces limites

$$\lim_{\substack{\lambda_L \to -\infty \\ \lambda_R \to +\infty \\ \alpha = \text{cst}}} \frac{p_R^*}{\gamma - 1} = \alpha \left(E_R - \frac{\left(\alpha \rho_R u_R + (1-\alpha)\rho_L u_L\right)^2}{2\rho_R} \right) + (1-\alpha) \left(E_L - \frac{\left(\alpha \rho_R u_R + (1-\alpha)\rho_L u_L\right)^2}{2\rho_L} \right) - (1-\alpha) \frac{g\bar{\rho}[Z]}{\gamma - 1}.$$

On considère à présent la limite quand α tend vers 1, c'est-à-dire quand $\left|\frac{\lambda_R}{\lambda_L}\right|$ tend vers $+\infty$ et on obtient

$$\lim_{\substack{\lambda_L \to -\infty \\ \lambda_R \to +\infty \\ \alpha \to 1}} \frac{p_R^*}{\gamma - 1} = E_R - \rho_R u_R^2/2 = \frac{p_R}{\gamma - 1} > 0.$$

Par conséquent, en choisissant $|\lambda_L|$, λ_R et $\left|\frac{\lambda_R}{\lambda_L}\right|$ suffisamment grand, on obtient $p_R^* > 0$. □

Conclusion

Nous avons réussi à généraliser le schéma well-balanced construit pour le système de Ripa en un schéma well-balanced pour les équations d'Euler avec gravité. Ce schéma préserve la positivité de la densité et de la pression. Cependant la positivité de la pression réintroduit une contrainte très forte sur les vitesses d'ondes qui rend le schéma très diffusif. D'autre part, de manière pratique, il est difficile de trouver les vitesses qui conviennent. Nous allons donc nous tourner vers une approche différente en utilisant des schémas de relaxation.

Pour la résolution du solveur de Riemann approché, on a supposé que la pression vérifiait la loi des gaz parfaits. Il serait probablement possible de généraliser cette approche à une loi de pression générale au prix d'une résolution plus compliquée du système d'équations gouvernant le solveur. Ce n'est pas le but de ce travail et on ne s'étendra pas plus sur ce point. Par ailleurs, comme on va le voir dans la partie suivante, les méthodes de relaxation vont fournir des schémas indépendants de la loi de pression utilisée.

4.4 Méthodes de relaxation

Le principe des méthodes de relaxation consiste à utiliser un système d'EDP élargi avec une perturbation singulière qui approche le système d'EDP initial (voir [31, 70, 38, 11, 13, 30]). Le système élargi, appelé système de relaxation, doit redonner le système initial dans la limite vers zéro d'un paramètre de relaxation. On utilise alors ce système approchant pour construire un schéma numérique pour le système initial. Pour que cette méthode soit intéressante, le système de relaxation doit être en un certain sens plus simple que le système initial. Le plus souvent, on demande que le système élargi ne comporte que des champs linéairement dégénérés, ce qui permet d'utiliser le schéma de Godunov sur le système de relaxation.

Dans un premier temps, on introduit le formalisme des méthodes de relaxation. On s'intéresse ensuite à des méthodes de relaxation pour approcher les solutions du système de Saint-Venant. On commence par rappeler le système de relaxation de Suliciu (voir [20]). La résolution exacte du problème de Riemann pour ce modèle est délicate. En effet, le terme source de topographie introduit de fortes non-linéarités dans le modèle de Suliciu et de plus l'ordre des valeurs propres n'est pas déterminé *a priori*. Par conséquent, on propose une modification du modèle de Suliciu qui consiste à « transporter » le terme source à la vitesse du fluide. On fait ainsi disparaitre artificiellement l'onde stationnaire en la collapsant arbitrairement avec l'onde de contact naturelle du modèle. Malheureusement, cette modification rend le problème de Riemann sous-déterminé. Afin d'assurer le caractère bien posé du problème de Riemann, on rajoute une linéarisation de l'équation décrivant l'équilibre local pour fermer le système. On montre enfin que la solution du « problème de Riemann » obtenue peut se réécrire comme la solution d'un nouveau système de relaxation complètement déterminé.

On étend ensuite aisément cette approche au système de Ripa et aux équations d'Euler avec gravité.

4.4.1 Formalisme des méthodes de relaxation

Système de relaxation

On utilise le formalisme des systèmes de relaxation introduit dans [31, 16]. Considérons un système de d lois de conservations avec terme source de la forme

$$\partial_t w + \partial_x f(w) = s(w)\partial_x Z, \tag{4.88}$$

où $w : \mathbb{R} \times \mathbb{R}^+ \to \Omega$ est le vecteur inconnue, $f : \Omega \to \mathbb{R}^d$ est la fonction flux, $s : \Omega \to \mathbb{R}^d$ est le terme source, $\Omega \subset \mathbb{R}^d$ est l'ouvert convexe des états admissibles et $Z : \mathbb{R} \to \mathbb{R}$ est une fonction régulière donnée. On introduit un système de relaxation composé de N lois de conservations, avec $N > d$, de la forme

$$\partial_t W + \partial_x F(W) = S(W, Z) + \frac{1}{\varepsilon} R(W, Z), \tag{4.89}$$

où $W : \mathbb{R}^+ \times \mathbb{R} \to \mathcal{O}$ est le vecteur des inconnues à valeur dans un ouvert $\mathcal{O} \subset \mathbb{R}^N$ supposé convexe, $F : \mathcal{O} \to \mathbb{R}^N$ est la fonction flux, $S : \mathcal{O} \times \mathbb{R} \to \mathbb{R}^N$ est le terme source, $R : \mathcal{O} \times \mathbb{R} \to \mathbb{R}^N$ est le terme de relaxation et $\varepsilon > 0$ est le paramètre de relaxation.

Tout au long de cette partie, on utilisera les conventions d'écriture suivantes pour rendre plus claire la présentation. Les variables de $\mathbb{R}^d$ et les fonctions de $\mathbb{R}^d$ dans $\mathbb{R}^d$, c'est-à-dire tous les objets associés au système original, seront notées en minuscules. On notera en majuscule toutes les variables de $\mathbb{R}^N$ et les fonctions de $\mathbb{R}^N$ dans lui-même, autrement dit, les objets associés au système de relaxation. Enfin les opérateurs allant de $\mathbb{R}^d$ dans $\mathbb{R}^N$ ou inversement seront notés en lettre calligraphiées.

On suppose l'existence d'une matrice constante $\mathcal{Q}$ de taille $d \times N$ et de rang d telle que

$$\mathcal{Q}R(W,Z) = 0, \quad \forall W \in \mathcal{O}, \forall Z \in \mathbb{R},$$
$$\mathcal{Q}\mathcal{O} = \Omega.$$

On suppose de plus que pour chaque couple $(w,Z) \in \Omega \times \mathbb{R}$, il existe un unique équilibre $\mathcal{E}(w,Z)$ vérifiant

$$\mathcal{Q}\mathcal{E}(w,Z) = w, \quad R(\mathcal{E}(w,Z),Z) = 0. \tag{4.90}$$

On introduit alors la variété d'équilibre, définie par

$$\mathcal{M} = \{W = \mathcal{E}(w,Z), w \in \Omega, Z \in \mathbb{R}\}. \tag{4.91}$$

Le lemme suivant sera utile pour caractériser la variété d'équilibre.

Lemme 4.20. *Soit W un vecteur de $\mathcal{O}$. Les trois propositions suivantes sont équivalentes :*

(i) $W \in \mathcal{M}$;

(ii) il existe $Z \in \mathbb{R}$ *tel que* $R(W,Z) = 0$;

(iii) il existe $Z \in \mathbb{R}$ *tel que* $\mathcal{E}(\mathcal{Q}W,Z) = W$.

Démonstration.
(i) $\Rightarrow$ (ii) : soit $W \in \mathcal{M}$. Par la définition (4.91) de $\mathcal{M}$, il existe $w \in \Omega$ et $Z \in \mathbb{R}$ tels que $W = \mathcal{E}(w,Z)$. D'après (4.90), on a alors

$$R(W,Z) = R(\mathcal{E}(w,Z),Z) = 0.$$

(ii) $\Rightarrow$ (iii) : supposons qu'il existe $Z \in \mathbb{R}$ tel que $R(W,Z) = 0$ et notons $w = \mathcal{Q}W$. On a alors

$$\begin{cases} \mathcal{Q}\mathcal{E}(w,Z) = w \\ R(\mathcal{E}(w,Z),Z) = 0 \end{cases} \quad \text{et} \quad \begin{cases} \mathcal{Q}W = w \\ R(W,Z) = 0. \end{cases}$$

Les vecteurs $\mathcal{E}(w,Z)$ et W vérifient donc tous les deux la définition (4.90) de l'équilibre. Puisque l'on a supposé que l'équilibre était unique, on a $W = \mathcal{E}(w,Z)$, donc $W = \mathcal{E}(\mathcal{Q}W,Z)$.
(iii) $\Rightarrow$ (i) : immédiat d'après la définition (4.91) de $\mathcal{M}$. □

Pour assurer la « consistance » entre le système de relaxation (4.89) et le système original (4.88), on demande que les relations de compatibilité suivantes soient vérifiées :

$$\mathcal{Q}F(\mathcal{E}(w,Z)) = f(w), \quad \forall w \in \Omega, \forall Z \in \mathbb{R}. \tag{4.92}$$
$$\mathcal{Q}S(\mathcal{E}(w,Z),Z) = s(w)\partial_x Z, \quad w \in \Omega, \forall Z \in \mathbb{R}.$$

Si W est une solution du système de relaxation (4.89) à valeurs dans la variété d'équilibre $\mathcal{M}$, la fonction $w = \mathcal{Q}W$ est alors solution du système (4.88). En effet, en multipliant l'équation (4.89) par $\mathcal{Q}$, on obtient

$$\partial_t w + \partial_x \mathcal{Q}F(W) = \mathcal{Q}S(W,Z).$$

Si W est à valeurs dans $\mathcal{M}$, le lemme 4.20 nous assure que $W = \mathcal{E}(\mathcal{Q}W,Z)$ et on en déduit

$$\mathcal{Q}F(W) = \mathcal{Q}F(\mathcal{E}(w,Z)) = f(w),$$
$$\mathcal{Q}S(W,Z) = \mathcal{Q}S(\mathcal{E}(w,Z),Z) = s(w)\partial_x Z.$$

On conclut que w vérifie l'équation

$$\partial_t w + \partial_x f(w) = s(w)\partial_x Z.$$

En général, il n'y a pas de raison pour que les solutions de (4.89) soient à valeurs dans $\mathcal{M}$. Cependant, lorsque l'on fait tendre le paramètre de relaxation ε vers 0, le terme de relaxation $R(W)$ tend formellement vers 0 et les solutions de (4.89) sont de plus en plus proches de la variété d'équilibre. C'est dans ce sens que le modèle de relaxation (4.89) « approche » le système initial (4.88) (voir [31, 70, 62]).

Schéma de relaxation

Dans cette partie, nous allons voir comment dériver un schéma numérique pour le système original (4.88) à partir du modèle de relaxation (4.89). On considère le problème de Riemann pour le système de relaxation sans le terme source de relaxation, c'est-à-dire

$$\begin{cases} \partial_t W + \partial_x F(W) = S(W,Z), \\ W(x,0) = \begin{cases} W_L & \text{si } x < 0, \\ W_R & \text{si } x > 0, \end{cases} \end{cases} \tag{4.93}$$

où Z est une fonction discontinue donnée de la forme suivante :

$$Z(x) = \begin{cases} Z_L & \text{si } x < 0, \\ Z_R & \text{si } x > 0. \end{cases} \tag{4.94}$$

On suppose connue la solution exacte $W_{\mathcal{R}}\left(\frac{x}{t}, W_L, Z_L, W_R, Z_R\right)$ de (4.93)–(4.94).

À la date t^n, on considère une approximation de la solution du système original (4.88), constante par morceaux,

$$w^n_{\Delta x}(x) = w^n_i, \quad \text{si } x \in K_i.$$

Le principe d'un schéma de relaxation consiste à considérer cette solution comme une approximation du modèle de relaxation appartenant à la variété d'équilibre

$$W^n_i = \mathcal{E}(w^n_i, Z_i), \quad i \in \mathbb{Z}. \tag{4.95}$$

Puisque la solution exacte $W_{\mathcal{R}}$ du problème de Riemann (4.93) est connue, on utilise le schéma de Godunov pour obtenir une approximation à la date $t^n + \Delta t$ de la solution du système de relaxation sans le terme source de relaxation

$$\partial_t W + \partial_x F(W) = S(W,Z). \tag{4.89}_{\varepsilon=+\infty}$$

Enfin on projette la solution obtenue sur la variété d'équilibre pour obtenir une approximation du système initial au temps $t^n + \Delta t$.

Le schéma de relaxation repose donc sur deux étapes que nous détaillons maintenant.

Évolution en temps ($t^n \to t^{n+1,-}$)
Connaissant une approximation constante par morceaux $w^n_{\Delta x}$ pour le système original (4.88), on définit une approximation à l'équilibre $W^n_{\Delta x}$ pour le modèle de relaxation de la façon suivante :

$$W^n_{\Delta x}(x) = W^n_i, \quad \text{si } x \in K_i,$$

où W^n_i est défini par (4.95). On résout alors le problème $(4.89)_{\varepsilon=+\infty}$ avec pour donnée initiale $W^n_{\Delta x}(x)$ et en omettant le terme source de relaxation. Localement, on a un problème de Riemann à chaque interface $x_{i+1/2}$. On note $W_{\Delta x}(x, t^n + t)$ la solution du problème de Cauchy pour le système $(4.89)_{\varepsilon=+\infty}$ muni de la donnée initiale $W_{\Delta x}(x, t^n) = W^n_{\Delta x}(x)$. Sous la condition CFL

$$\frac{\Delta t}{\Delta x} \max_{i \in \mathbb{Z}} \left| \lambda^{\pm}\left(W^n_i, Z_i, W^n_{i+1}, Z_{i+1}\right)\right| \leq \frac{1}{2}, \tag{4.96}$$

la solution $W_{\Delta x}$ est constituée de la juxtaposition sans interaction de la solution des problèmes de Riemann à chaque interface $x_{i+1/2}$, c'est-à-dire

$$W_{\Delta x}(x, t^n + t) = W_{\mathcal{R}}\left(\frac{x - x_{i+1/2}}{t}, W^n_i, Z_i, W^n_{i+1}, Z_{i+1}\right), \quad \text{si } x \in [x_i, x_{i+1}[.$$

Comme pour le schéma de Godunov, on projette alors cette solution sur l'espace des fonctions constantes sur chaque K_i pour obtenir

$$W_i^{n+1,-} = \frac{1}{\Delta x} \int_{K_i} W_{\Delta x}(x, t^n + \Delta t) dx. \tag{4.97}$$

Relaxation ($t^{n+1,-} \to t^{n+1}$)
Cette étape du schéma est dédiée à la prise en compte du terme source de relaxation qui a été négligé au cours de l'étape d'évolution. On résout l'équation différentielle

$$\partial_t W = \frac{1}{\varepsilon} R(W, Z) \tag{4.98}$$

avec pour condition initiale $W_i^{n+1,-}$ et dans la limite $\varepsilon \to 0$. On note W_i^{n+1} la solution de cette équation différentielle. La mise à jour de l'approximation est alors donnée par

$$w_i^{n+1} = \mathcal{Q} W_i^{n+1}.$$

Pour conclure la présentation du schéma de relaxation, remarquons qu'en pratique, il n'est pas nécessaire de résoudre l'équation différentielle (4.98). En effet, en multipliant l'équation (4.98) par $\mathcal{Q}$, on obtient

$$\partial_t \mathcal{Q} W = 0.$$

On en déduit

$$w_i^{n+1} = \mathcal{Q} W_i^{n+1,-}. \tag{4.99}$$

Ainsi la mise à jour du schéma de relaxation est obtenue directement à partir de l'état $W_i^{n+1,-}$ résultant de l'étape d'évolution.

Il nous reste à trouver des conditions simples pour que le schéma de relaxation ainsi défini vérifie les propriétés classiques.

Propriétés du schéma de relaxation

On commence par présenter une condition pour que le schéma de relaxation soit well-balanced.

Lemme 4.21. *Supposons que pour tous états (w_L, Z_L) et (w_R, Z_R) à l'équilibre local pour le système original (4.88), la solution du problème de Riemann pour le système de relaxation (4.89) vérifie*

$$W_{\mathcal{R}}(\xi, \mathcal{E}(w_L, Z_L), Z_L, \mathcal{E}(w_R, Z_R), Z_R) = \begin{cases} \mathcal{E}(w_L, Z_L) & \text{si } \xi < 0, \\ \mathcal{E}(w_R, Z_R) & \text{si } \xi > 0, \end{cases}$$

alors le schéma de relaxation défini par (4.99) est well-balanced.

Démonstration. Considérons une solution discrète stationnaire constante par morceaux $(w_i^n)_{i \in \mathbb{Z}}$ pour le système (4.88). En particulier, les états (w_{i-1}^n, Z_{i-1}) et (w_i^n, Z_i) d'une part et les états (w_i^n, Z_i) et (w_{i+1}^n, Z_{i+1}) d'autre part sont à l'équilibre local. L'hypothèse affirme alors que

$$W_{\mathcal{R}}\left(\xi, \mathcal{E}(w_{i-1}^n), Z_{i-1}, \mathcal{E}(w_i^n), Z_i\right) = \mathcal{E}(w_i^n, Z_i), \quad \forall \xi > 0,$$

$$W_{\mathcal{R}}\left(\xi, \mathcal{E}(w_i^n), Z_i, \mathcal{E}(w_{i+1}^n), Z_{i+1}\right) = \mathcal{E}(w_i^n, Z_i), \quad \forall \xi < 0.$$

L'équation (4.97) implique donc que $W_i^{n+1,-} = W_i^n$. On a alors

$$w_i^{n+1} = \mathcal{Q} W_i^{n+1,-} = \mathcal{Q} W_i^n = Q \mathcal{E}(w_i^n, Z_i).$$

De plus, par définition de la fonction équilibre $\mathcal{E}$, on a

$$\mathcal{Q}\mathcal{E}(w_i^n, Z_i) = w_i^n.$$

Par conséquent, le schéma (4.99) est well-balanced.

□

On a de même un lemme permettant d'assurer la robustesse du schéma.

Lemme 4.22. *Supposons que la solution du problème de Riemann pour le système de relaxation (4.89) vérifie*

$$W_{\mathcal{R}}\left(\xi, W_L, Z_L, W_R, Z_R\right) \in \mathcal{O}, \quad \forall \xi \in \mathbb{R}, \forall W_L, W_R \in \mathcal{O}, \forall Z_L, Z_R \in \mathbb{R}.$$

Alors, sous la condition CFL (4.96), le schéma de relaxation défini par (4.99) préserve les états admissibles, c'est-à-dire

$$\forall i \in \mathbb{Z}, \quad w_i^n \in \Omega \quad \Rightarrow \quad \forall i \in \mathbb{Z}, \quad w_i^{n+1} \in \Omega.$$

Démonstration. En utilisant l'hypothèse et la convexité de $\mathcal{O}$, l'équation (4.97) nous assure que $W_i^{n+1} \in \mathcal{O}$. Puisque $\mathcal{Q}\mathcal{O} = \Omega$, on déduit de (4.99) que w_i^{n+1} est dans Ω. En conséquence, la mise à jour w_i^{n+1} est bien admissible. □

4.4.2 Schéma de relaxation avec transport de la topographie pour les équations de Saint-Venant

On s'intéresse, dans cette partie, à des schémas de relaxation well-balanced permettant d'approcher les équations de Saint-Venant (4.3). Dans un premier temps, on rappelle le modèle de relaxation de Suliciu [98, 20]. Ce modèle admet des invariants de Riemann très non linéaires associés à l'onde stationnaire. De plus, l'ordre des ondes n'est pas toujours le même. Ces difficultés rendent la résolution exacte du problème de Riemann délicate. Pour remédier à ce problème, on introduit une modification du modèle de Suliciu en transportant artificiellement le terme source de topographie à la vitesse du fluide. Il n'y a alors plus d'onde stationnaire mais il manque un invariant de Riemann. En conséquence, le système d'équations régissant la solution du problème de Riemann est sous-déterminé. Pour fermer le système, on choisit d'ajouter une linéarisation de l'équation définissant l'équilibre local. Pour justifier ce choix, on montre que le schéma ainsi obtenu dérive en fait d'un nouveau modèle de relaxation complètement déterminé.

Modèle de relaxation de Suliciu pour les équations de Saint-Venant

Pour les équations de Saint-Venant (4.3), plusieurs choix de système de relaxation sont possibles. On présente ici le modèle de Suliciu qui ne modifie que la pression, responsable des champs vraiment non linéaires dans le système original. Le système de Suliciu pour les équations de Saint-Venant comporte trois équations ($N = 3$) et s'écrit

$$\begin{cases} \partial_t h + \partial_x hu = 0, \\ \partial_t hu + \partial_x (hu^2 + \pi) = -gh\partial_x Z, \\ \partial_t h\pi + \partial_x (u(h\pi + \nu^2)) = \dfrac{h}{\varepsilon}\left(gh^2/2 - \pi\right). \end{cases} \tag{4.100}$$

Le paramètre ν est ici une linéarisation de l'impédance acoustique qui doit vérifier la condition sous-caractéristique de Whitham [104] :

$$\nu^2 > \rho^2 c^2, \tag{4.101}$$

afin d'éviter des instabilités dans la procédure de relaxation, lorsque ε tend vers 0.

Remarquons que ce système s'écrit sous la forme (4.88), avec

$$W = (h, hu, h\pi)^T,$$

$$F(W) = (hu, hu^2 + \pi, u(h\pi + \nu^2))^T,$$

$$S(W, Z) = (0, -gh\partial_x Z, 0)^T,$$

$$R(W, Z) = \left(0, 0, 0, h\left(gh^2/2 - \pi\right)\right)^T.$$

L'ensemble des états admissibles est

$$\mathcal{O} = \left\{W = (h, hu, h\pi)^T \in \mathbb{R}^3, h > 0\right\}.$$

La matrice $\mathcal{Q}$ représente la projection sur les deux premières composantes de W :

$$\mathcal{Q} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

La fonction déterminant l'équilibre est $\mathcal{E}(w, Z) = (h, hu, gh^3/2)^T$ et la variété d'équilibre est définie par $\mathcal{M} = \left\{W = (h, hu, h\pi)^T, \pi = gh^2/2\right\}$. On peut facilement vérifier que ces objets vérifient toutes les propriétés requises introduites dans la partie précédente.

Pour étudier ce système, il est pratique de faire intervenir le vecteur des grandeurs physiques $U = (h, u, \pi, Z)^T$. On peut alors écrire le système $(4.100)_{\varepsilon=+\infty}$ sous la forme quasi-linéaire

$$\partial_t U + A(U)\partial_x U = 0, \tag{4.102}$$

où la matrice $A(U)$ est définie par

$$A(U) = \begin{pmatrix} u & h & 0 & 0 \\ 0 & u & 1/h & g \\ 0 & \nu^2/h & u & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Ce système admet quatre valeurs propres simples :

$$u, \quad 0, \quad u \pm \frac{\nu}{h},$$

avec pour vecteurs propres respectifs

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} h \\ -u \\ \nu^2/h \\ \frac{u^2 - \nu^2/h^2}{g} \end{pmatrix}, \quad \begin{pmatrix} h^2 \\ \pm\nu \\ \nu^2 \\ 0 \end{pmatrix}.$$

Le système (4.102) est hyperbolique en tout point de $\mathcal{O}$ sauf aux points vérifiant $u = \pm\frac{\nu}{h}$ pour lesquels la base de diagonalisation est perdue. Toutes les valeurs propres sont associées à des champs linéairement dégénérés. Les invariants de Riemann pour le système de Suliciu sont pour la valeur propre u :

$$u, \quad \pi, \quad Z,$$

pour la valeur propre 0 :

$$hu, \quad \pi + \frac{\nu^2}{h}, \quad gZ + \frac{u^2}{2} - \frac{\nu^2}{2h^2}.$$

pour la valeur propre $u \pm \frac{\nu}{h}$:

$$Z, \quad u \pm \frac{\nu}{h}, \quad \pi \mp \nu u,$$

En théorie, on peut en déduire la solution exacte du problème de Riemann pour le système de Suliciu (voir [20]). Il y a cependant deux difficultés à cette approche. Premièrement, l'ordre des valeurs propres n'est pas toujours le même, ce qui fait qu'il y a plusieurs cas à distinguer. Deuxièmement, les invariants de Riemann associés à la valeur propre 0 sont très non linéaires, ce qui rend la résolution du système difficile. La résolution exacte du problème de Riemann est donc possible, mais elle doit être effectuée au cas par cas, ce qui s'avère extrêmement coûteux. On ne s'attardera pas plus sur la résolution du problème de Riemann pour ce système.

Modèle de relaxation avec transport de la topographie

Comme on vient de le voir, la difficulté dans le modèle de relaxation de Suliciu provient de la non-linéarité des invariants de Riemann pour l'onde stationnaire. On choisit de modifier le modèle de Suliciu en introduisant une nouvelle variable a qui est transportée à la vitesse u et qui sera ensuite relaxée vers Z. On obtient un modèle de relaxation comportant quatre équations ($N = 4$) :

$$\begin{cases} \partial_t h + \partial_x hu = 0, \\ \partial_t hu + \partial_x (hu^2 + \pi) = -gh\partial_x a, \\ \partial_t h\pi + \partial_x (u(h\pi + \nu^2)) = \frac{h}{\varepsilon}(gh^2/2 - \pi), \\ \partial_t ha + \partial_x hua = \frac{h}{\varepsilon}(Z - a). \end{cases} \tag{4.103}$$

Avant de montrer que ce système entre dans le formalisme de la Partie 4.4.1, signalons que sept modèles de relaxation sont introduits dans ce chapitre. Pour plus de lisibilité, on utilise les mêmes notations W, F, S, $\mathcal{O}$, $\mathcal{Q}$, $\mathcal{E}$, R et $\mathcal{M}$ pour tous ces modèles. Il n'y a cependant pas de confusion possible.

On peut écrire le système (4.103) sous la forme (4.88), avec

$$W = (h, hu, h\pi, ha)^T,$$

$$F(W) = (hu, hu^2 + \pi, u(h\pi + \nu^2), hua)^T,$$

$$S(W) = (0, -gh\partial_x a, 0, 0)^T,$$

$$R(W, Z) = (0, 0, 0, h(gh^2/2 - \pi), h(Z - a))^T.$$

L'ensemble des états admissibles pour ce système est

$$\mathcal{O} = \left\{ W = (h, hu, h\pi, ha)^T \in \mathbb{R}^4, h > 0 \right\}.$$

La matrice $\mathcal{Q}$ représente encore la projection sur les deux premières composantes de W :

$$\mathcal{Q} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

La fonction déterminant l'équilibre est $\mathcal{E}(w, Z) = (h, hu, gh^3/2, hZ)^T$ et la variété d'équilibre est définie par $\mathcal{M} = \left\{ W = (h, hu, h\pi, ha)^T, \pi = gh^2/2, a = Z \right\}$.

Remarquons que le terme $S(w)$ est un produit non conservatif. Le terme de flux et le terme source vont être traités simultanément dans le schéma qui va suivre, sans utiliser de splitting d'opérateur. Le S est ici simplement une notation consistante avec la formulation (4.89). On précise également que dans tous les modèles de relaxation qui vont suivre, le terme source ne dépend plus de Z, contrairement au modèle de Suliciu. Par conséquent, pour simplifier les notations, on omet à partir de maintenant la dépendance en Z dans le terme source S.

On introduit le vecteur des grandeurs physiques $U = (h, u, \pi, a)^T$. On peut alors réécrire le système $(4.103)_{\varepsilon=+\infty}$ sous forme quasi-linéaire

$$\partial_t U + A(U)\partial_x U = 0, \tag{4.104}$$

où la matrice $A(U)$ est donnée par

$$A(U) = \begin{pmatrix} u & h & 0 & 0 \\ 0 & u & 1/h & g \\ 0 & \nu^2/h & u & 0 \\ 0 & 0 & 0 & u \end{pmatrix}.$$

Cette matrice admet une valeur propre double u et deux valeurs propres simples $u \pm \frac{\nu}{h}$. Les vecteurs propres associés à la valeur propre u sont :

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ gh \\ -1 \end{pmatrix}.$$

Le vecteur propre associé à la valeur propre $u \pm \frac{\nu}{h}$ est :

$$\begin{pmatrix} h^2 \\ \pm\nu \\ \nu^2 \\ 0 \end{pmatrix}.$$

Le système (4.104) est hyperbolique en tout point de Ω. Notons que contrairement au système de Suliciu, les valeurs propres ne peuvent pas se croiser et donc l'ordre de celles-ci est toujours le même. Par ailleurs, les champs associés aux trois valeurs propres sont encore linéairement dégénérés. Les invariants de Riemann pour ce système sont pour la valeur propre $u \pm \frac{\nu}{h}$:

$$a, \quad u \pm \frac{\nu}{h}, \quad \pi \mp \nu u$$

et pour la valeur propre u :

$$u.$$

La valeur propre u étant de multiplicité deux, on s'attend à ce que le champ associé ait deux invariants de Riemann linéairement indépendants. Le fait qu'il y ait un invariant manquant rend le problème de Riemann sous-déterminé. Il est donc nécessaire de rajouter une équation au système pour rendre bien posé le problème de Riemann.

On admet que la solution « exacte » du problème de Riemann pour le système de relaxation (4.103) est composé de deux états intermédiaires W_L^* et W_R^* séparés par trois discontinuités de contact de vitesses $u_L - \frac{\nu}{h_L}$, u^* et $u_R + \frac{\nu}{h_R}$ (voir Figure 4.5). Puisque u est continue à travers l'onde du milieu, on note $u^* = u_L^* = u_R^*$. Ceci revient à éliminer l'équation provenant de l'invariant de Riemann trivial u pour l'onde associée à la valeur propre u. La solution du problème de Riemann pour (4.103) s'écrit alors

$$W_{\mathcal{R}}\left(\frac{x}{t}, W_L, Z_L, W_R, Z_R\right) = \begin{cases} (h_L, h_L u_L, h_L \pi_L, h_L a_L)^T & \text{si } \frac{x}{t} < u_L - \frac{\nu}{h_L}, \\ (h_L^*, h_L^* u^*, h_L^* \pi_L^*, h_L^* a_L)^T & \text{si } u_L - \frac{\nu}{h_L} < \frac{x}{t} < u^*, \\ (h_R^*, h_R^* u^*, h_R^* \pi_R^*, h_R^* a_R)^T & \text{si } u^* < \frac{x}{t} < u_R + \frac{\nu}{h_R}, \\ (h_R, h_R u_R, h_R \pi_R, h_R a_R)^T & \text{si } u_R + \frac{\nu}{h_R} < \frac{x}{t}. \end{cases}$$

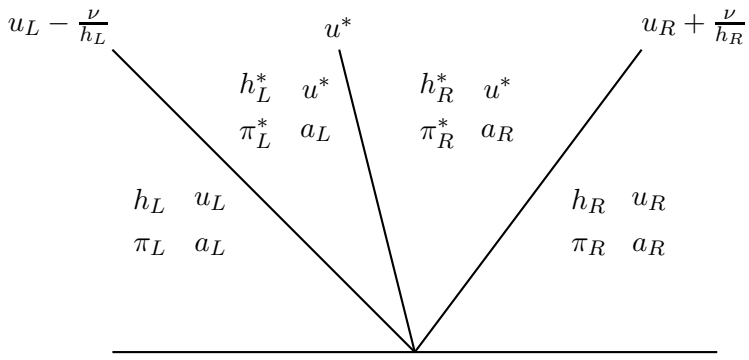


FIGURE 4.5 – Structure de la solution « exacte » du problème de Riemann pour le système de relaxation (4.103)

Il reste alors cinq inconnues intervenant dans ce problème de Riemann : h_L^*, u^*, π_L^*, h_R^* et π_R^*. Or il n'y a que quatre équations pour assurer la continuité des invariants de Riemann. Il manque bien une équation pour fermer le système.

Afin d'assurer que le solveur obtenu est well-balanced (ce qui sera vérifié plus loin dans les propriétés du schéma), on choisit de rajouter l'équation provenant de la définition de l'équilibre local (4.5), c'est-à-dire

$$\pi_R^* - \pi_L^* = -g\frac{h_L + h_R}{2}(a_R - a_L). \tag{4.105}$$

On obtient alors le système composé des cinq équations suivantes :

$$\begin{aligned}
u_L - \frac{\nu}{h_L} &= u^* - \frac{\nu}{h_L^*},\\
u_R + \frac{\nu}{h_R} &= u^* + \frac{\nu}{h_R^*},\\
\pi_L + \nu u_L &= \pi_L^* + \nu u^*,\\
\pi_R - \nu u_R &= \pi_R^* - \nu u^*,\\
\pi_R^* - \pi_L^* &= -g\frac{h_L + h_R}{2}(a_R - a_L).
\end{aligned}$$

On peut alors résoudre le système et on trouve

$$u^* = \overline{u} - \frac{[\pi]}{2\nu} - g\overline{h}\frac{[a]}{2\nu}, \tag{4.106}$$

$$\pi_L^* = \pi_L + \nu(u_L - u^*), \tag{4.107}$$

$$\pi_R^* = \pi_R + \nu(u^* - u_R), \tag{4.108}$$

$$\frac{1}{h_L^*} = \frac{1}{h_L} + \frac{u^* - u_L}{\nu}, \tag{4.109}$$

$$\frac{1}{h_R^*} = \frac{1}{h_R} + \frac{u_R - u^*}{\nu}. \tag{4.110}$$

Le choix de l'équation (4.105) pour fermer le système peut paraitre arbitraire. Avant de présenter le schéma associé à ce solveur de Riemann et de montrer qu'il possède toutes les propriétés requises, on va justifier le choix de l'équation (4.105) en introduisant un nouveau modèle de relaxation.

Reformulation en un modèle complètement déterminé

Comme on l'a vu précédemment, la notion de solution du système $(4.103)_{\varepsilon=+\infty}$ est ambigüe. Cela est lié au fait qu'il manque un invariant de Riemann au système. Nous allons introduire un nouveau modèle de relaxation plus large, qui possède un ensemble complet d'invariants de Riemann et qui admet la « même » solution du problème de Riemann.

Pour cela, on introduit deux nouvelles variables X^- et X^+ qui vont toutes les deux être relaxées vers h. Ces deux variables sont transportées respectivement à la vitesse $u-\delta$ et $u+\delta$, où $\delta > 0$ est un paramètre suffisamment petit. On obtient le modèle de relaxation suivant composé de six équations ($N=6$)

$$\left\{\begin{array}{l} \partial_t h + \partial_x hu = 0, \\ \partial_t hu + \partial_x (hu^2 + \pi) = -g\dfrac{X^- + X^+}{2}\partial_x a, \\ \partial_t h\pi + \partial_x (u(h\pi + \nu^2)) = \dfrac{h}{\varepsilon}(gh^2/2 - \pi), \\ \partial_t ha + \partial_x hua = \dfrac{h}{\varepsilon}(Z - a), \\ \partial_t hX^- + \partial_x huX^- = \delta h\partial_x X^- + \dfrac{h}{\varepsilon}(h - X^-), \\ \partial_t hX^+ + \partial_x huX^+ = -\delta h\partial_x X^+ + \dfrac{h}{\varepsilon}(h - X^+). \end{array}\right. \tag{4.111}$$

On peut écrire ce système sous la forme (4.89) avec

$$W = (h, hu, h\pi, ha, hX^-, hX^+)^T,$$

$$F(W) = (hu, hu^2 + \pi, u(h\pi + \nu^2), hua, huX^-, huX^+)^T,$$

$$S(W) = \left(0, -g\frac{X^- + X^+}{2}\partial_x a, 0, 0, \delta h\partial_x X^-, -\delta h\partial_x X^+\right)^T,$$

$$R(W, Z) = (0, 0, 0, h(gh^2/2 - \pi), h(Z - a), h(\rho - X^-), h(\rho - X^+))^T.$$

L'ensemble des vecteurs admissibles est

$$\mathcal{O} = \left\{W = (h, hu, h\pi, ha, hX^-, hX^+)^T \in \mathbb{R}^6, h > 0\right\}.$$

La matrice $\mathcal{Q}$ représente à nouveau la projection sur les deux premières coordonnées. La fonction déterminant l'équilibre est $\mathcal{E}(w, Z) = (h, hu, gh^3/2, hZ, h^2, h^2)^T$ et la variété d'équilibre est définie par

$$\mathcal{M} = \left\{W = (h, hu, h\pi, ha, hX^-, hX^+)^T, \pi = gh^2/2, a = Z, X^- = h, X^+ = h\right\}.$$

On introduit le vecteur des grandeurs physiques $U = (h, u, \pi, a, X^-, X^+)^T$. Le système de relaxation $(4.111)_{\varepsilon=+\infty}$ se réécrit alors sous la forme quasi-linéaire

$$\partial_t U + A(U)\partial_x U = 0,$$

où la matrice $A(U)$ est donnée par

$$A(U) = \begin{pmatrix} u & h & 0 & 0 & 0 & 0 \\ 0 & u & 1/h & g\frac{X^- + X^+}{2h} & 0 & 0 \\ 0 & \nu^2/h & u & 0 & 0 & 0 \\ 0 & 0 & 0 & u & 0 & 0 \\ 0 & 0 & 0 & 0 & u - \delta & 0 \\ 0 & 0 & 0 & 0 & 0 & u + \delta \end{pmatrix}.$$

Si δ est choisi suffisamment petit, ce système est hyperbolique en tout point de $\mathcal{O}$. Il admet u comme valeur propre double et $u \pm \delta$ et $u \pm \frac{\nu}{h}$ comme valeurs propres simples. Les vecteurs propres associés à la valeur propre u sont :

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ g\frac{X^- + X^+}{2} \\ -1 \\ 0 \\ 0 \end{pmatrix}.$$

Le vecteur propre associé à la valeur propre $u \pm \frac{\nu}{h}$ est :

$$\begin{pmatrix} h^2 \\ \pm\nu \\ \nu^2 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

alors que ceux associés aux valeurs propres $u - \delta$ et $u + \delta$ sont respectivement

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Pour que les valeurs propres ne se croisent pas, il suffit de choisir δ tel que $\delta < \frac{\nu}{h}$. L'ordre des valeurs propres est alors fixe :

$$u - \frac{\nu}{h} < u - \delta < u < u + \delta < u + \frac{\nu}{h}.$$

Toutes les valeurs propres sont associées à des champs linéairement dégénérés. Les invariants de Riemann sont pour la valeur propre u :

$$u, \quad X^-, \quad X^+, \quad \pi + g\frac{X^- + X^+}{2}a,$$

pour la valeur propre $u \pm \frac{\nu}{h}$:

$$a, \quad X^-, \quad X^+, \quad u \pm \frac{\nu}{h}, \quad \pi \mp \nu u,$$

pour la valeur propre $u - \delta$:

$$h, \quad u, \quad \pi, \quad a, \quad X^+,$$

pour la valeur propre $u - \delta$:

$$h, \quad u, \quad \pi, \quad a, \quad X^-,$$

Contrairement au système (4.103) , ce système a donc un ensemble complet d'invariants de Riemann. Ceux-ci vont donc permettre de déterminer de manière complète la solution du problème de Riemann.

La solution du problème de Riemann pour le système $(4.111)_{\varepsilon=+\infty}$ est composée de quatre états intermédiaires séparés par cinq discontinuités de contact (voir Figure 4.6). Après avoir

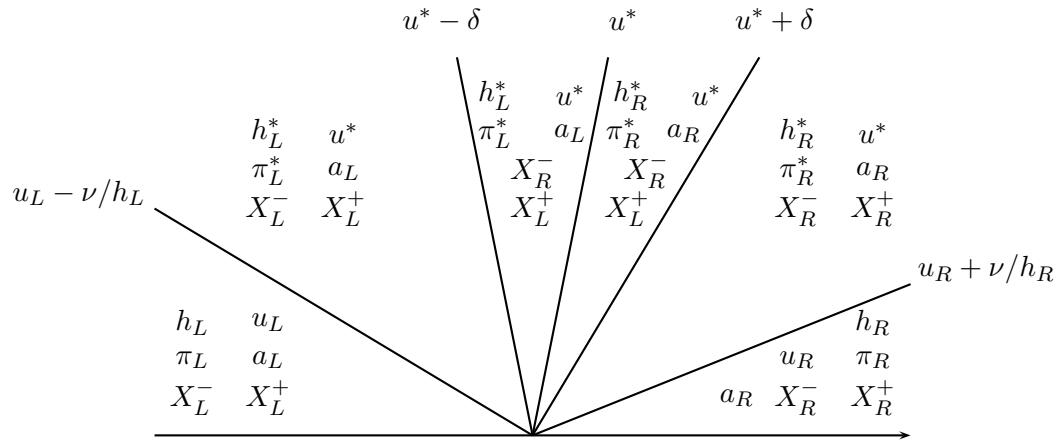


FIGURE 4.6 – Structure de la solution exacte du problème de Riemann pour le système de relaxation (4.111)

éliminé les relations liées aux invariants de Riemann triviaux , il reste cinq inconnues : h_L^*, h_R^*, u^*, π_L^* et π_R^*. Les équations données par les invariants de Riemann $u \pm \frac{\nu}{h}$ et $\pi \mp \nu u$ pour les ondes $u \pm \frac{\nu}{\rho}$ nous donnent quatre équations qui étaient déjà présentes dans le précédent modèle. Le dernier invariant de Riemann non trivial, $\pi + g\frac{X^- + X^+}{2}a$ pour l'onde u nous donne

$$\pi_R^* - \pi_L^* = -g\frac{X_R^- + X_L^+}{2}(a_R - a_L). \tag{4.112}$$

Si l'on suppose que la condition initiale appartient à la variété d'équilibre $\mathcal{M}$, alors on a $X_L^+ = h_L$ et $X_R^- = h_R$ et l'équation (4.112) est identique à la dernière équation (4.105) du système précédent. Par conséquent, les vecteurs

$$\mathcal{Q}W_{\mathcal{R}}\left(\frac{x}{t}, \mathcal{E}(w_L, Z_L), Z_L, \mathcal{E}(w_R, Z_R), Z_R\right)$$

pour les systèmes de relaxation (4.103) et (4.111) sont égaux. Notons que dans la dérivation du schéma de relaxation, la donnée est supposée à l'équilibre au début de chaque pas de temps. Les deux système mènent donc au même schéma numérique.

Le schéma de relaxation

Maintenant que l'on connait la solution exacte du problème de Riemann pour le système de relaxation (4.111), on en déduit un schéma numérique en suivant la technique présentée dans la partie 4.4.1. Notons qu'en pratique, le paramètre ν est choisi localement pour chaque problème de Riemann $W_{\mathcal{R}}(\xi, W_L, Z_L, W_R, Z_R)$, de manière à satisfaire la condition sous-caractéristique de Whitham (4.101), ainsi que les conditions nécessaires à la robustesse qui seront présentées plus loin. On notera $\nu_{i+1/2}$ le paramètre utilisé dans le problème de Riemann $W_{\mathcal{R}}\left(\xi, W_i^n, Z_i, W_{i+1}^n, Z_{i+1}\right)$.

Le schéma de relaxation est décrit par la proposition suivante.

Proposition 4.23. *Le schéma de relaxation associé au modèle de relaxation (4.111) s'écrit*

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(f_{i+1/2} - f_{i-1/2}\right) + \frac{\Delta t}{2}\left(s(w_{i-1}^n, w_i^n)\frac{Z_i - Z_{i-1}}{\Delta x} + s(w_i^n, w_{i+1}^n)\frac{Z_{i+1} - Z_i}{\Delta x}\right), \tag{4.113}$$

où le flux numérique est défini par

$$f_{i+1/2} = f(w_i^n, Z_i, w_{i+1}^n, Z_{i+1}), \tag{4.114}$$

$$f(w_L, Z_L, w_R, Z_R) = \begin{cases} \left(h_L u_L, h_L u_L^2 + \pi_L - \frac{1}{2} g\overline{h}[Z]\right)^T & si\ u_L - \frac{\nu}{h_L} > 0, \\ \left(h_L^* u^*, h_L^* (u^*)^2 + \pi_L^* - \frac{1}{2} g\overline{h}[Z]\right)^T & si\ u_L - \frac{\nu}{h_L} < 0 < u^*, \\ \left(h_R^* u^*, h_R^* (u^*)^2 + \pi_R^* + \frac{1}{2} g\overline{h}[Z]\right)^T & si\ u^* < 0 < u_R + \frac{\nu}{h_R}, \\ \left(h_R u_R, h_R u_R^2 + \pi_R + \frac{1}{2} g\overline{h}[Z]\right)^T & si\ u_R + \frac{\nu}{h_R} < 0, \end{cases} \tag{4.115}$$

et le terme source numérique est défini par

$$s(w_L, w_R) = \left(0, -g\overline{h}\right)^T. \tag{4.116}$$

De plus, le schéma (4.113) est consistant avec (4.3).

Démonstration. Écrivons le système $(4.111)_{\varepsilon=+\infty}$ sous la forme

$$\partial_t W + \partial_x F(W) = S(W). \tag{4.117}$$

On note $W_{\Delta x}$ la solution exacte de ce système pour la condition initiale

$$W_{\Delta x}(x, t^n) = W_i^n = \mathcal{E}(w_i^n, Z_i) \quad \text{si } x \in K_i.$$

On peut alors intégrer (4.117) sur le rectangle $K_i \times [t^n, t^n + \Delta t]$ pour obtenir

$$\begin{aligned} \int_{K_i} W_{\Delta x}(x, t^n + \Delta t) dx = \int_{K_i} W_{\Delta x}(x, t^n) dx - \int_{t^n}^{t^n + \Delta t} F\left(W_{\Delta x}(x_{i+1/2}, t)\right) dt \\ + \int_{t^n}^{t^n + \Delta t} F\left(W_{\Delta x}(x_{i-1/2}, t)\right) dt + \int_{K_i} \int_{t^n}^{t^n + \Delta t} S\left(W_{\Delta x}(x, t)\right) dt dx. \end{aligned}$$

En utilisant l'auto-similarité de la solution du problème de Riemann, on en déduit

$$\begin{aligned} W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x} \left(F\left(W_{\mathcal{R}}\left(0, W_i^n, Z_i, W_{i+1}^n, Z_{i+1}\right)\right) - F\left(W_{\mathcal{R}}\left(0, W_{i-1}^n, Z_{i-1}, W_i^n, Z_i\right)\right)\right) \\ + \frac{1}{\Delta x} \int_{K_i} \int_{t^n}^{t^n + \Delta t} S\left(W_{\Delta x}(x, t)\right) dt dx. \end{aligned}$$

On multiplie cette équation par $\mathcal{Q}$ et, puisque $w_i^n = \mathcal{Q} W_i^n$ et $w_i^{n+1} = \mathcal{Q} W_i^{n+1}$, on trouve

$$\begin{aligned} w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x} \left(\mathcal{Q} F\left(W_{\mathcal{R}}\left(0, W_i^n, Z_i, W_{i+1}^n, Z_{i+1}\right)\right) - \mathcal{Q} F\left(W_{\mathcal{R}}\left(0, W_{i-1}^n, Z_{i-1}, W_i^n, Z_i\right)\right)\right) \\ + \frac{1}{\Delta x} \int_{K_i} \int_{t^n}^{t^n + \Delta t} \mathcal{Q} S\left(W_{\Delta x}(x, t)\right) dt dx, \end{aligned} \tag{4.118}$$

Avant d'évaluer le terme source, on rappelle la formule des sauts (voir par exemple [110]). Soit f une fonction C^1 par morceaux dont les discontinuités sont situées aux points $(x_i)_{i \in I}$. Pour chaque $i \in I$, on note $\sigma_i = f(x_i^+) - f(x_i^-)$ le saut de la fonction f au point x_i. On note $\{f'\}$ la fonction définie presque partout et qui est égale à la dérivée au sens usuel de f en chaque point ou celle-ci est dérivable. Alors la dérivée de f au sens des distributions, notée f', vérifie la formule des sauts :

$$f' = \{f'\} + \sum_{i \in I} \sigma_i \delta_{x_i}, \tag{4.119}$$

où δ_{x_i} est la distribution de Dirac au point x_i.

On va ensuite découper l'intégrale du terme source en deux :

$$\frac{1}{\Delta x}\int_{K_i}\int_{t^n}^{t^n+\Delta t}\mathcal{Q}S\left(W_{\Delta x}(x,t)\right)dtdx = \Delta t\left(S^+(W_{i-1}^n,Z_{i-1},W_i^n,Z_i)+S^-(W_i^n,Z_i,W_{i+1}^n,Z_{i+1})\right) \tag{4.120}$$

où les fonctions $S^{\pm}$ sont définies par

$$S^-(W_L,Z_L,W_R,Z_R)=\frac{1}{\Delta t\Delta x}\int_{-\Delta x/2}^{0}\int_0^{\Delta t}\mathcal{Q}S\left(W_{\mathcal{R}}\left(\frac{x}{t},W_L,Z_L,W_R,Z_R\right)\right)dtdx,$$

$$S^+(W_L,Z_L,W_R,Z_R)=\frac{1}{\Delta t\Delta x}\int_{0}^{\Delta x/2}\int_0^{\Delta t}\mathcal{Q}S\left(W_{\mathcal{R}}\left(\frac{x}{t},W_L,Z_L,W_R,Z_R\right)\right)dtdx.$$

En utilisant les définitions de S et de $\mathcal{Q}$ dans le cas du système (4.111), on a

$$\mathcal{Q}S(W)=\left(0,-g\frac{X^-+X^+}{2}\partial_x a\right)^T.$$

Dans un premier temps, on s'intéresse à la solution du problème de Riemann dans le rectangle $[-\Delta x/2,0]\times[0,\Delta t]$. Si $u^*>0$ alors a est constante, donc

$$S^-(W_L,Z_L,W_R,Z_R)=(0,0)^T.$$

Si $u^*<0$, alors a n'est discontinue que le long de la droite de vitesse u^* et X^- et X^+ restant constantes le long de cette droite (voir Figure 4.7), on trouve par la formule des sauts

$$S^-(W_L,Z_L,W_R,Z_R)=\left(0,-g\frac{X_L^-+X_R^+}{2}\frac{a_R-a_L}{\Delta x}\right)^T.$$

Pour des états W_L et W_R dans la variété d'équilibre $\mathcal{M}$, on a $X_L^+=h_L$, $X_R^-=h_R$, $a_L=Z_L$ et $a_R=Z_R$. Les états $\mathcal{E}(w_L,Z_L)$ et $\mathcal{E}(w_R,Z_R)$ étant par définition dans la variété d'équilibre $\mathcal{M}$, on a donc

$$S^-\left(\mathcal{E}(w_L,Z_L),Z_L,\mathcal{E}(w_R,Z_R),Z_R\right)=\left(0,-g\overline{h}\frac{[Z]}{\Delta x}\right)^T.$$

On peut unifier les cas $u^*<0$ et $u^*>0$ par la formule

$$S^-\left(\mathcal{E}(w_L,Z_L),Z_L,\mathcal{E}(w_R,Z_R),Z_R\right)=\left(0,\frac{\mathrm{sgn}(u^*)-1}{2}g\overline{h}\frac{[Z]}{\Delta x}\right)^T,$$

où l'on a noté

$$\mathrm{sgn}(x)=\begin{cases}-1 & \text{si } x<0,\\ 1 & \text{si } x>0.\end{cases}$$

On procède de la même façon dans le rectangle $[0,\Delta x/2]\times[0,\Delta t]$. Si $u^*<0$, l'inconnue a est constante dans le rectangle , donc

$$S^+(W_L,Z_L,W_R,Z_R)=(0,0)^T.$$

Si $u^*>0$, on trouve par la formule des sauts

$$S^+(W_L,Z_L,W_R,Z_R)=\left(0,-g\frac{X_L^-+X_R^+}{2}\frac{a_R-a_L}{\Delta x}\right)^T.$$

FIGURE 4.7 – L'inconnue a dans le problème de Riemann pour le système (4.111) – Gauche : cas $u^* < 0$. Droite : cas $u^* > 0$

Puisque $\mathcal{E}(w_L, Z_L)$ et $\mathcal{E}(w_R, Z_R)$ sont dans la variété d'équilibre $\mathcal{M}$, on en déduit

$$S^+ \left(\mathcal{E}(W_L, Z_L), Z_L, \mathcal{E}(W_R, Z_R), Z_R\right) = \left(0, -g\overline{h}\frac{[Z]}{\Delta x}\right)^T.$$

On trouve finalement la formule générique pour S^+ :

$$S^+ \left(\mathcal{E}(w_L, Z_L), Z_L, \mathcal{E}(w_R, Z_R), Z_R\right) = \left(0, -\frac{\text{sgn}(u^*)+1}{2}g\overline{h}\frac{[Z]}{\Delta x}\right)^T.$$

En rappelant que $W_i^n = \mathcal{E}(w_i^n, Z_i)$, on déduit de (4.120) que l'intégrale du terme source s'écrit

$$\begin{aligned}\frac{1}{\Delta x}\int_{K_i}\int_{t^n}^{t^n+\Delta t} \mathcal{QS}\left(W_{\Delta x}(x,t)\right) dtdx = \frac{\Delta t}{2}\left(s(w_{i-1}^n, w_i^n)\frac{Z_i - Z_{i-1}}{\Delta x} + s(w_i^n, w_{i+1}^n)\frac{Z_{i+1}-Z_i}{\Delta x}\right)\\ - \frac{\Delta t}{\Delta x}\left(0, \frac{\text{sgn}(u^*_{i-1/2})}{2}g\frac{h_{i-1}^n + h_i^n}{2}(Z_i - Z_{i-1}) - \frac{\text{sgn}(u^*_{i+1/2})}{2}g\frac{h_i^n + h_{i+1}^n}{2}(Z_{i+1}-Z_i)\right)^T,\end{aligned}$$

où $s(w_L, w_R)$ est défini par (4.116). En injectant cette équation dans (4.118), on trouve

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(f_{i+1/2} - f_{i-1/2}\right) + \frac{\Delta t}{2}\left(s(w_{i-1}^n, w_i^n)\frac{Z_i - Z_{i-1}}{\Delta x} + s(w_i^n, w_{i+1}^n)\frac{Z_{i+1}-Z_i}{\Delta x}\right),$$

où le flux numérique est défini par (4.114) et

$$f(w_L, Z_L, w_R, Z_R) = \mathcal{QF}\left(W_{\mathcal{R}}\left(0, \mathcal{E}(w_L, Z_L), Z_L, \mathcal{E}(w_R, Z_R), Z_R\right)\right) - \left(0, \frac{\text{sgn}(u^*)}{2}g\overline{h}[Z]\right)^T.$$

Il reste à montrer que ce flux numérique peut se réécrire sous la forme (4.115), ce qui se voit facilement d'après la structure du problème de Riemann (voir Figure 4.5).

Enfin, la consistance du terme source est immédiate d'après (4.116). Concernant la consistance du flux numérique, si $w_L = w_R = w$ et $Z_L = Z_R = Z$, on déduit aisément des équations (4.106) à (4.110) que $w_L^* = w_R^* = w$. La formulation (4.115) du flux numérique implique alors

$$f(w, Z, w, Z) = f(w).$$

□

Propriétés vérifiées par le schéma

Commençons par montrer que le schéma est well-balanced.

Proposition 4.24. *Le schéma (4.113) est well-balanced, c'est-à-dire que si* $\forall i \in \mathbb{Z}$ *on a*

$$u_i^n = 0 \quad et \quad h_{i+1}^n + Z_{i+1} = h_i^n + Z_i,$$

alors $w_i^{n+1} = w_i^n$, $\forall i \in \mathbb{Z}$.

Démonstration. Soient (w_L, Z_L) et (w_R, Z_R) deux états vérifiant l'équilibre local (4.5). D'après le lemme 4.21, il suffit de montrer que

$$W_{\mathcal{R}}(\xi, \mathcal{E}(w_L, Z_L), Z_L, \mathcal{E}(w_R, Z_R), Z_R) = \begin{cases} \mathcal{E}(w_L, Z_L) & \text{si } \xi < 0, \\ \mathcal{E}(w_R, Z_R) & \text{si } \xi > 0. \end{cases}$$

Par définition de la fonction $\mathcal{E}$ et puisque $[h] = [Z]$, les états $W_L = \mathcal{E}(w_L, Z_L)$ et $W_R = \mathcal{E}(w_R, Z_R)$ vérifient

$$\begin{aligned} \pi_R - \pi_L &= gh_R^2/2 - gh_L^2/2 \\ &= -g[h][Z] \\ &= -g[h][a]. \end{aligned}$$

D'autre part, on a $u_L = u_R = 0$. On déduit de l'équation (4.106) que $u^* = 0$. Les équations (4.107) à (4.110) impliquent alors

$$\pi_L^* = \pi_L, \quad \pi_R^* = \pi_R, \quad h_L^* = h_L, \quad h_R^* = h_R,$$

donc $w_L^* = \mathcal{E}(w_L, Z_L)$ et $w_R^* = \mathcal{E}(w_R, Z_R)$, ce qui conclut la preuve. □

On montre maintenant que le schéma est robuste sous certaines conditions sur la constante ν.

Proposition 4.25. *Supposons que la constante* $\nu_{i+1/2}$ *assure que les valeurs propres du système vérifient l'ordre suivant :*

$$u_i^n - \frac{\nu_{i+1/2}}{h_i^n} < u_{i+1/2}^* < u_{i+1}^n + \frac{\nu_{i+1/2}}{h_{i+1}^n}. \tag{4.121}$$

Si la condition CFL

$$\frac{\Delta t}{\Delta x} \max_{i \in \mathbb{Z}} \left| u_i^n \pm \frac{\nu_{i \mp 1/2}}{h_i^n} \right| \leq \frac{1}{2}$$

est vérifiée, alors le schéma (4.113) est robuste.

Démonstration. D'après le lemme 4.22, il suffit de montrer que pour tous W_L et W_R dans $\mathcal{O}$ et tous Z_L et Z_R dans $\mathbb{R}$, on a $W_L^* \in \mathcal{O}$ et $W_R^* \in \mathcal{O}$. Cela revient à montrer que $h_L^* > 0$ et $h_R^* > 0$.

En utilisant la continuité des invariants de Riemann $u \pm \frac{\nu}{h}$ pour les ondes $u \pm \frac{\nu}{h}$, on déduit que l'hypothèse (4.121) se réécrit de façon équivalente

$$u^* - \frac{\nu}{h_L^*} < u^* < u^* + \frac{\nu}{h_R^*}.$$

En conséquence, imposer (4.121) revient à imposer $h_L^* > 0$ et $h_R^* > 0$. On en déduit immédiatement la robustesse attendue. □

Remarquons qu'imposer les inéquations (4.121) revient à assurer la positivité de deux polynômes du second degré en ν tendant vers $+\infty$ quand ν tend vers $+\infty$.

4.4.3 Schéma de relaxation avec transport de topographie pour les équations de Ripa

On utilise maintenant la technique précédente pour construire un schéma de relaxation well-balanced pour approcher les équations de Ripa (4.6). On ne présente pas ici le modèle de Suliciu qui comporte les mêmes difficultés que pour les équations de Saint-Venant. On introduit donc directement la version modifiée de ce modèle où l'on transporte le terme de topographie à la vitesse u. Il manque alors une équation pour déterminer la solution du problème de Riemann et l'on considère une linéarisation de l'équation régissant l'équilibre local. On montre enfin que le schéma obtenu est en fait dérivé d'un autre modèle de relaxation complètement déterminé.

Modèle de relaxation avec transport de topographie

On introduit une nouvelle variable a qui sera transportée à la vitesse u et relaxée vers la topographie Z. On obtient alors le modèle de relaxation suivant :

$$\begin{cases} \partial_t h + \partial_x hu = 0, \\ \partial_t hu + \partial_x (hu^2 + \pi) = -gh\theta\partial_x a, \\ \partial_t h\theta + \partial_x hu\theta = 0, \\ \partial_t h\pi + \partial_x (u(h\pi + \nu^2)) = \dfrac{h}{\varepsilon}(gh^2\theta/2 - \pi), \\ \partial_t ha + \partial_x hua = \frac{h}{\varepsilon}(Z - a). \end{cases} \tag{4.122}$$

Ce système entre dans le formalisme introduit dans la partie 4.4.1 avec

$$W = (h, hu, h\theta, h\pi, ha)^T,$$

$$F(W) = (hu, hu^2 + \pi, hu\theta, u(h\pi + \nu^2), hua)^T,$$

$$S(W) = (0, -gh\theta\partial_x a, 0, 0, 0)^T,$$

$$R(W, Z) = (0, 0, 0, h(gh^2\theta/2 - \pi), h(Z - a))^T.$$

L'ensemble des états admissibles pour ce système est

$$\mathcal{O} = \left\{ W = (h, hu, h\theta, h\pi, ha)^T \in \mathbb{R}^5, h > 0, \theta > 0 \right\}.$$

La matrice $\mathcal{Q}$ représente la projection sur les trois premières composantes de W :

$$\mathcal{Q} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

La fonction déterminant l'équilibre est

$$\mathcal{E}(w, Z) = (h, hu, h\theta, gh^3\theta/2, hZ)^T$$

et la variété d'équilibre est définie par

$$\mathcal{M} = \left\{ W = (h, hu, h\theta, h\pi, ha)^T, \pi = gh^2\theta/2, a = Z \right\}.$$

On introduit le vecteur des grandeurs physiques $U = (h, u, \theta, \pi, a)^T$. On peut alors réécrire le système $(4.122)_{\varepsilon=+\infty}$ sous forme quasi-linéaire

$$\partial_t U + A(U)\partial_x U = 0,$$

où la matrice $A(U)$ est donnée par

$$A(U) = \begin{pmatrix} u & h & 0 & 0 & 0 \\ 0 & u & 0 & 1/h & g\theta \\ 0 & 0 & u & 0 & 0 \\ 0 & \nu^2/h & 0 & u & 0 \\ 0 & 0 & 0 & 0 & u \end{pmatrix}.$$

Ce système est hyperbolique en tout point de $\mathcal{O}$, avec pour valeurs propres u de multiplicité 3 et $u \pm \frac{\nu}{h}$, chacune de multiplicité 1. L'ordre des valeurs propres est toujours le même et les champs qui leur sont associés sont tous linéairement dégénérés. Les invariants de Riemann pour ce système sont pour la valeur propre $u \pm \frac{\nu}{h}$:

$$\theta, \quad a, \quad u \pm \frac{\nu}{h}, \quad \pi \mp \nu u$$

et pour la valeur propre u :

$$u.$$

Comme dans le cas des équations de Saint-Venant, il manque un invariant de Riemann pour l'onde u. En effet, celle-ci est de multiplicité 3 et on pourrait s'attendre à ce qu'elle ait deux invariants de Riemann. Une équation supplémentaire sera donc nécessaire pour que le système soit complètement déterminé

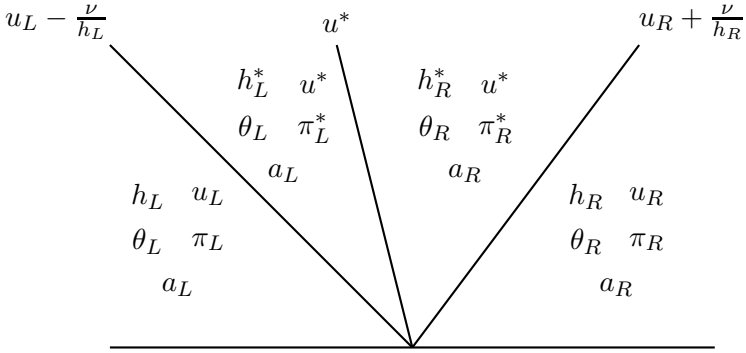


FIGURE 4.8 – Structure de la solution « exacte » du problème de Riemann pour le système de relaxation (4.122)

On admet que la solution « exacte » du problème de Riemann pour le système de relaxation $(4.122)_{\varepsilon=+\infty}$ est composé de deux états intermédiaires W_L^* et W_R^* séparés par trois discontinuités de contact de vitesses $u_L - \frac{\nu}{h_L}$, u^* et $u_R + \frac{\nu}{h_R}$ (voir Figure 4.8).

Après avoir éliminé les invariants de Riemann triviaux (dont celui pour l'onde u qui amène à définir $u^* = u_L^* = u_R^*$), il reste cinq inconnues : h_L^*, u^*, π_L^*, h_R^* et π_R^*. La solution du problème de Riemann pour (4.122) s'écrit

$$W_{\mathcal{R}}\left(\frac{x}{t}, W_L, Z_L, W_R, Z_R\right) = \begin{cases} (h_L, h_L u_L, h_L \theta_L, h_L \pi_L, h_L a_L)^T & \text{si } \frac{x}{t} < u_L - \frac{\nu}{h_L} \\ (h_L^*, h_L^* u^*, h_L^* \theta_L, h_L^* \pi_L^*, h_L^* a_L)^T & \text{si } u_L - \frac{\nu}{h_L} < \frac{x}{t} < u^*, \\ (h_R^*, h_R^* u^*, h_R^* \theta_R, h_R^* \pi_R^*, h_R^* a_R)^T & \text{si } u^* < \frac{x}{t} < u_R + \frac{\nu}{h_R}, \\ (h_R, h_R u_R, h_R \theta_R, h_R \pi_R, h_R a_R)^T & \text{si } u_R + \frac{\nu}{h_R} < \frac{x}{t}. \end{cases}$$

Le système est pour le moment composé des quatre équations provenant des invariants de Riemann $u \pm \frac{\nu}{h}$ et $\pi \mp \nu u$.

On choisit de rajouter l'équation provenant de la définition de l'équilibre local, c'est-à-dire

$$\pi_R^* - \pi_L^* = -g\bar{h}\bar{\theta}[a]. \tag{4.123}$$

On obtient le système composé des cinq équations suivantes :

$$\begin{aligned}
u_L - \frac{\nu}{h_L} &= u^* - \frac{\nu}{h_L^*},\\
u_R + \frac{\nu}{h_R} &= u^* + \frac{\nu}{h_R^*},\\
\pi_L + \nu u_L &= \pi_L^* + \nu u^*,\\
\pi_R - \nu u_R &= \pi_R^* - \nu u^*,\\
\pi_R^* - \pi_L^* &= -g\bar{h}\bar{\theta}[a].
\end{aligned}$$

On peut alors résoudre le système et on trouve

$$u^* = \overline{u} - \frac{[\pi]}{2\nu} - g\bar{h}\bar{\theta}\frac{[a]}{2\nu}, \tag{4.124}$$

$$\pi_L^* = \pi_L + \nu(u_L - u^*), \tag{4.125}$$

$$\pi_R^* = \pi_R + \nu(u^* - u_R), \tag{4.126}$$

$$\frac{1}{h_L^*} = \frac{1}{h_L} + \frac{u^* - u_L}{\nu}, \tag{4.127}$$

$$\frac{1}{h_R^*} = \frac{1}{h_R} + \frac{u_R - u^*}{\nu}. \tag{4.128}$$

Avant de présenter le schéma associé à ce solveur de Riemann et de montrer qu'il possède toutes les propriétés requises, on va justifier le choix de l'équation (4.123) en introduisant un nouveau modèle de relaxation.

Reformulation en un modèle complètement déterminé

On présente maintenant un nouveau système de relaxation complètement déterminé et qui va permettre de justifier le choix de l'équation (4.123).

On introduit quatre nouvelles variables X^-, X^+, Y^- et Y^+. Les variables X^- et Y^- vont être transportées à la vitesse $u-\delta$ et les variables X^+ et Y^+ seront transportées à la vitesse $u+\delta$. Ici, $\delta > 0$ est un paramètre suffisamment petit. On va alors relaxer les variables $X^\pm$ vers h et les variables $Y^\pm$ vers θ. On obtient le modèle de relaxation suivant composé de neuf équations ($N = 9$)

$$\left\{\begin{array}{l}
\partial_t h + \partial_x hu = 0,\\
\partial_t hu + \partial_x(hu^2 + \pi) = -g\dfrac{X^- + X^+}{2}\dfrac{Y^- + Y^+}{2}\partial_x a,\\
\partial_t h\theta + \partial_x hu\theta = 0,\\
\partial_t h\pi + \partial_x(u(h\pi + \nu^2)) = \dfrac{h}{\varepsilon}(gh^2\theta/2 - \pi),\\
\partial_t ha + \partial_x hua = \dfrac{h}{\varepsilon}(Z - a),\\
\partial_t hX^- + \partial_x huX^- = \delta h\partial_x X^- + \dfrac{h}{\varepsilon}(h - X^-),\\
\partial_t hX^+ + \partial_x huX^+ = -\delta h\partial_x X^+ + \dfrac{h}{\varepsilon}(h - X^+),\\
\partial_t hY^- + \partial_x huY^- = \delta h\partial_x Y^- + \dfrac{h}{\varepsilon}(\theta - Y^-),\\
\partial_t hY^+ + \partial_x huY^+ = -\delta h\partial_x Y^+ + \dfrac{h}{\varepsilon}(\theta - Y^+).
\end{array}\right. \tag{4.129}$$

Ce système rentre dans le formalisme introduit dans la partie 4.4.1 en posant

$$W = (h, hu, h\theta, h\pi, ha, hX^-, hX^+, hY^-, hY^+)^T,$$

$$F(W) = (hu, hu^2 + \pi, hu\theta, u(h\pi + \nu^2), hua, huX^-, huX^+, huY^-, huY^+)^T,$$

$$S(W) = \left(0, -g\frac{X^- + X^+}{2}\frac{Y^- + Y^+}{2}\partial_x a, 0, 0, 0, \delta h\partial_x X^-, -\delta h\partial_x X^+, \delta h\partial_x Y^-, -\delta h\partial_x Y^+\right)^T,$$

$$R(W, Z) = (0, 0, 0, h(gh^2\theta/2 - \pi), h(Z - a), h(\rho - X^-), h(\rho - X^+), h(\theta - Y^-), h(\theta - Y^+))^T.$$

L'ensemble des vecteurs admissibles est

$$\mathcal{O} = \left\{W = (h, hu, h\theta, h\pi, ha, hX^-, hX^+, hY^-, hY^+)^T \in \mathbb{R}^9, h > 0, \theta > 0\right\}.$$

La matrice $\mathcal{Q}$ représente la projection sur les trois premières composantes de W. La fonction déterminant l'équilibre est

$$\mathcal{E}(w, Z) = (h, hu, h\theta, gh^3\theta/2, hZ, h^2, h^2, h\theta, h\theta)^T$$

et la variété d'équilibre est définie par

$$\mathcal{M} = \left\{W \in \mathcal{O}, \pi = gh^2\theta/2, a = Z, X^- = h, X^+ = h, Y^- = \theta, Y^+ = \theta\right\}.$$

On introduit le vecteur des grandeurs physiques $U = (h, u, \theta, \pi, a, X^-, X^+, Y^-, Y^+)^T$. Le système de relaxation $(4.129)_{\varepsilon=+\infty}$ se réécrit alors sous la forme quasi-linéaire

$$\partial_t U + A(U)\partial_x U = 0,$$

où la matrice $A(U)$ est donnée par

$$A(U) = \begin{pmatrix}
u & h & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & u & 0 & 1/h & g\frac{X^- + X^+}{2h}\frac{Y^- + Y^+}{2} & 0 & 0 & 0 & 0 \\
0 & 0 & u & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \nu^2/h & 0 & u & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & u & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & u - \delta & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & u + \delta & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & u - \delta & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & u - \delta
\end{pmatrix}.$$

Ce système est hyperbolique en tout point de $\mathcal{O}$ pour δ suffisamment petit. Ses valeurs propres sont :

$$u \text{ (triple)}, \quad u \pm \frac{\nu}{h} \text{ (simples)}, \quad u \pm \delta \text{ (doubles)}. \tag{4.130}$$

Elles sont toutes associées à des champs linéairement dégénérés. Il suffit de choisir δ tel que $\delta < \frac{\nu}{h}$ pour assurer que les valeurs propres ne se croisent pas. Les invariants de Riemann de ce système sont pour la valeur propre u :

$$u, \quad X^-, \quad X^+, \quad Y^-, \quad Y^+, \quad \pi + g\frac{X^- + X^+}{2}\frac{Y^- + Y^+}{2}a,$$

pour la valeur propre $u \pm \frac{\nu}{h}$:

$$a, \quad \theta, \quad X^-, \quad X^+, \quad Y^-, \quad Y^+, \quad u \pm \frac{\nu}{h}, \quad \pi \mp \nu u,$$

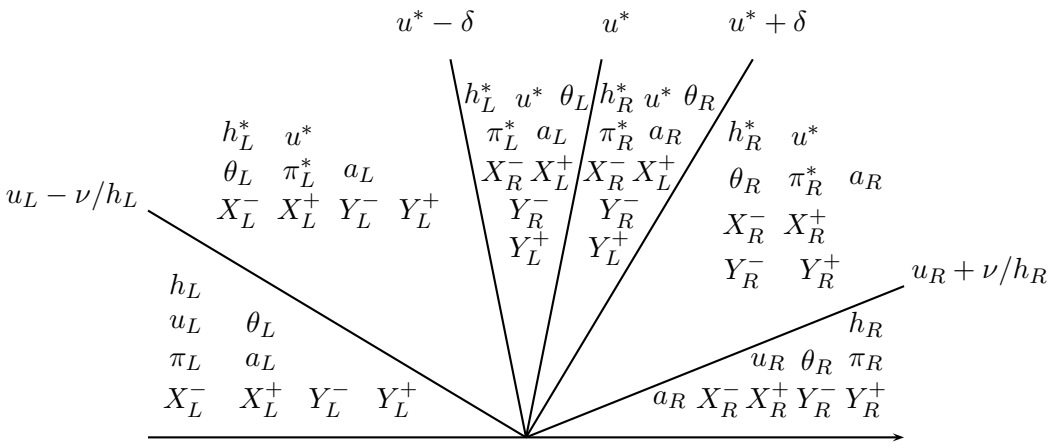


FIGURE 4.9 – Structure de la solution exacte du problème de Riemann pour le système de relaxation (4.129)

pour la valeur propre $u - \delta$:

$$h, \quad u, \quad \theta, \quad \pi, \quad a, \quad X^+, \quad Y^+,$$

pour la valeur propre $u + \delta$:

$$h, \quad u, \quad \theta, \quad \pi, \quad a, \quad X^-, \quad Y^-.$$

Le système (4.129) possède un ensemble complet d'invariants de Riemann qui permet de caractériser de manière unique la solution du problème de Riemann.

La solution du problème de Riemann pour le système $(4.111)_{\varepsilon=+\infty}$ est composée de quatre états intermédiaires séparés par cinq discontinuités de contact (voir Figure 4.6). Après élimination des invariants de Riemann triviaux, il reste cinq inconnues : h_L^*, h_R^*, u^*, π_L^* et π_R^*. Les invariants de Riemann $u \pm \frac{\nu}{h}$ et $\pi \mp \nu u$ pour les ondes $u \pm \frac{\nu}{\rho}$ nous donnent clairement quatre équations qui étaient déjà présentes dans le précédent modèle. Le dernier invariant de Riemann non trivial, $\pi + g\frac{X^- + X^+}{2}\frac{Y^- + Y^+}{2}a$ pour l'onde u nous donne

$$\pi_R^* - \pi_L^* = -g\frac{X_R^- + X_L^+}{2}\frac{Y_R^- + Y_L^+}{2}(a_R - a_L). \tag{4.131}$$

Cette équation coïncide avec la dernière équation (4.123) du précédent modèle dès que la condition initiale appartient à la variété d'équilibre $\mathcal{M}$. Les solution du problème de Riemann de chacun des deux modèles sont donc « identiques » pour des données à l'équilibre.

Le schéma de relaxation

Connaissant la solution du problème de Riemann pour le système $(4.129)_{\varepsilon=+\infty}$, on en déduit un schéma de relaxation que l'on présente dans la proposition suivante.

Proposition 4.26. *Le schéma de relaxation associé au modèle de relaxation (4.129) s'écrit*

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(f_{i+1/2} - f_{i-1/2}\right) + \frac{\Delta t}{2}\left(s(w_{i-1}^n, w_i^n)\frac{Z_i - Z_{i-1}}{\Delta x} + s(w_i^n, w_{i+1}^n)\frac{Z_{i+1} - Z_i}{\Delta x}\right), \tag{4.132}$$

où le flux numérique est défini par

$$f_{i+1/2} = f(w_i^n, Z_i, w_{i+1}^n, Z_{i+1}), \tag{4.133}$$

$$f(w_L,Z_L,w_R,Z_R)=\begin{cases}\left(h_Lu_L,h_Lu_L^2+\pi_L-\frac{1}{2}g\bar{h}\bar{\theta}[Z],h_Lu_L\theta_L\right)^T & \text{si } u_L-\frac{\nu}{h_L}>0,\\ \left(h_L^*u^*,h_L^*(u^*)^2+\pi_L^*-\frac{1}{2}g\bar{h}\bar{\theta}[Z],h_L^*u^*\theta_L\right)^T & \text{si } u_L-\frac{\nu}{h_L}<0<u^*,\\ \left(h_R^*u^*,h_R^*(u^*)^2+\pi_R^*+\frac{1}{2}g\bar{h}\bar{\theta}[Z],h_R^*u^*\theta_R\right)^T & \text{si } u^*<0<u_R+\frac{\nu}{h_R},\\ \left(h_Ru_R,h_Ru_R^2+\pi_R+\frac{1}{2}g\bar{h}\bar{\theta}[Z],h_Ru_R\theta_R\right)^T & \text{si } u_R+\frac{\nu}{h_R}<0,\end{cases} \tag{4.134}$$

et le terme source numérique est défini par

$$s(w_L,w_R)=\left(0,-g\bar{h}\bar{\theta},0\right)^T. \tag{4.135}$$

De plus, le schéma (4.132) est consistant avec (4.6).

Démonstration. Écrivons le système $(4.129)_{\varepsilon=+\infty}$ sous la forme

$$\partial_tW+\partial_xF(W)=S(W). \tag{4.136}$$

On note $W_{\Delta x}$ la solution exacte de ce système pour la condition initiale

$$W_{\Delta x}(x,t^n)=W_i^n=\mathcal{E}(w_i^n,Z_i)\quad \text{si } x\in K_i.$$

On peut alors intégrer (4.136) sur le rectangle $K_i\times[t^n,t^n+\Delta t]$ pour obtenir

$$\begin{aligned}\int_{K_i}W_{\Delta x}(x,t^n+\Delta t)dx=&\int_{K_i}W_{\Delta x}(x,t^n)dx-\int_{t^n}^{t^n+\Delta t}F\left(W_{\Delta x}(x_{i+1/2},t)\right)dt\\&+\int_{t^n}^{t^n+\Delta t}F\left(W_{\Delta x}(x_{i-1/2},t)\right)dt+\int_{K_i}\int_{t^n}^{t^n+\Delta t}S\left(W_{\Delta x}(x,t)\right)dtdx.\end{aligned}$$

En utilisant l'auto-similarité de la solution du problème de Riemann, on en déduit

$$\begin{aligned}W_i^{n+1}=W_i^n-\frac{\Delta t}{\Delta x}\left(F\left(W_{\mathcal{R}}\left(0,W_i^n,Z_i,W_{i+1}^n,Z_{i+1}\right)\right)-F\left(W_{\mathcal{R}}\left(0,W_{i-1}^n,Z_{i-1},W_i^n,Z_i\right)\right)\right)\\+\frac{1}{\Delta x}\int_{K_i}\int_{t^n}^{t^n+\Delta t}S\left(W_{\Delta x}(x,t)\right)dtdx.\end{aligned}$$

On multiplie cette équation par $\mathcal{Q}$ et, puisque $w_i^n=\mathcal{Q}W_i^n$ et $w_i^{n+1}=\mathcal{Q}W_i^{n+1}$, on trouve

$$\begin{aligned}w_i^{n+1}=w_i^n-\frac{\Delta t}{\Delta x}\left(\mathcal{Q}F\left(W_{\mathcal{R}}\left(0,W_i^n,Z_i,W_{i+1}^n,Z_{i+1}\right)\right)-\mathcal{Q}F\left(W_{\mathcal{R}}\left(0,W_{i-1}^n,Z_{i-1},W_i^n,Z_i\right)\right)\right)\\+\frac{1}{\Delta x}\int_{K_i}\int_{t^n}^{t^n+\Delta t}\mathcal{Q}S\left(W_{\Delta x}(x,t)\right)dtdx,\end{aligned} \tag{4.137}$$

On découpe ensuite l'intégrale du terme source en deux :

$$\frac{1}{\Delta x}\int_{K_i}\int_{t^n}^{t^n+\Delta t}\mathcal{Q}S\left(W_{\Delta x}(x,t)\right)dtdx=\Delta t\left(S^+(W_{i-1}^n,Z_{i-1},W_i^n,Z_i)+S^-(W_i^n,Z_i,W_{i+1}^nZ_{i+1})\right) \tag{4.138}$$

où les fonctions $S^\pm$ sont définies par

$$S^-(W_L,Z_L,W_R,Z_R)=\frac{1}{\Delta t\Delta x}\int_{-\Delta x/2}^0\int_0^{\Delta t}\mathcal{Q}S\left(W_{\mathcal{R}}\left(\frac{x}{t},W_L,Z_L,W_R,Z_R\right)\right)dtdx, \tag{4.139}$$

$$S^+(W_L,Z_L,W_R,Z_R)=\frac{1}{\Delta t\Delta x}\int_0^{\Delta x/2}\int_0^{\Delta t}\mathcal{Q}S\left(W_{\mathcal{R}}\left(\frac{x}{t},W_L,Z_L,W_R,Z_R\right)\right)dtdx.$$

En utilisant les définitions de S et de $\mathcal{Q}$ dans le cas du système (4.129), on a

$$\mathcal{Q}S(W)=\left(0,-g\frac{X^-+X^+}{2}\frac{Y^-+Y^+}{2}\partial_x a,0\right)^T.$$

Dans un premier temps, on s'intéresse à la solution du problème de Riemann dans le rectangle $[-\Delta x/2,0]\times[0,\Delta t]$. Si $u^*>0$ alors a est constante, donc

$$S^-(W_L,Z_L,W_R,Z_R)=(0,0,0)^T.$$

Si $u^*<0$, alors a n'est discontinue que le long de la droite de vitesse u^* et les inconnues X^-, X^+, Y^- et Y^+ restant constantes le long de cette droite, on trouve par la formule des sauts

$$S^-(W_L,Z_L,W_R,Z_R)=\left(0,-g\frac{X_L^-+X_R^+}{2}\frac{Y_L^++Y_R^-}{2}\frac{a_R-a_L}{\Delta x},0\right)^T.$$

Pour des états W_L et W_R dans la variété d'équilibre $\mathcal{M}$, on a $X_L^+=h_L$, $X_R^-=h_R$, $Y_L^+=\theta_L$, $Y_R^-=\theta_R$, $a_L=Z_L$ et $a_R=Z_R$. Les états $\mathcal{E}(w_L,Z_L)$ et $\mathcal{E}(w_R,Z_R)$ étant par définition dans la variété d'équilibre $\mathcal{M}$, on a donc

$$S^-\left(\mathcal{E}(w_L,Z_L),Z_L,\mathcal{E}(w_R,Z_R),Z_R\right)=\left(0,-g\bar{h}\bar{\theta}\frac{[Z]}{\Delta x},0\right)^T.$$

On peut unifier les cas $u^*<0$ et $u^*>0$ par la formule

$$S^-\left(\mathcal{E}(w_L,Z_L),Z_L,\mathcal{E}(w_R,Z_R),Z_R\right)=\left(0,\frac{\operatorname{sgn}(u^*)-1}{2}g\bar{h}\bar{\theta}\frac{[Z]}{\Delta x},0\right)^T. \tag{4.140}$$

On procède de la même façon dans le rectangle $[0,\Delta x/2]\times[0,\Delta t]$ pour trouver

$$S^+\left(\mathcal{E}(w_L,Z_L),Z_L,\mathcal{E}(w_R,Z_R),Z_R\right)=\left(0,-\frac{\operatorname{sgn}(u^*)+1}{2}g\bar{h}\bar{\theta}\frac{[Z]}{\Delta x},0\right)^T.$$

En rappelant que $W_i^n=\mathcal{E}(w_i^n,Z_i)$, on déduit de (4.138) que l'intégrale du terme source s'écrit

$$\begin{aligned}\frac{1}{\Delta x}\int_{K_i}\int_{t^n}^{t^n+\Delta t}\mathcal{Q}S\left(W_{\Delta x}(x,t)\right)dtdx=&\frac{\Delta t}{2}\left(s(w_{i-1}^n,w_i^n)\frac{Z_i-Z_{i-1}}{\Delta x}+s(w_i^n,w_{i+1}^n)\frac{Z_{i+1}-Z_i}{\Delta x}\right)\\&-\frac{\Delta t}{\Delta x}\left(0,\frac{\operatorname{sgn}(u^*_{i-1/2})}{2}g\frac{h_{i-1}^n+h_i^n}{2}\frac{\theta_{i-1}^n+\theta_i^n}{2}(Z_i-Z_{i-1})\right.\\&\left.-\frac{\operatorname{sgn}(u^*_{i+1/2})}{2}g\frac{h_i^n+h_{i+1}^n}{2}\frac{\theta_i^n+\theta_{i+1}^n}{2}(Z_{i+1}-Z_i),0\right)^T,\end{aligned}$$

où $s(w_L,w_R)$ est défini par (4.135). En injectant cette équation dans (4.137), on trouve

$$w_i^{n+1}=w_i^n-\frac{\Delta t}{\Delta x}\left(f_{i+1/2}-f_{i-1/2}\right)+\frac{\Delta t}{2}\left(s(w_{i-1}^n,w_i^n)\frac{Z_i-Z_{i-1}}{\Delta x}+s(w_i^n,w_{i+1}^n)\frac{Z_{i+1}-Z_i}{\Delta x}\right),$$

où le flux numérique est défini par (4.114) et

$$f(w_L,Z_L,w_R,Z_R)=\mathcal{Q}F\left(W_{\mathcal{R}}\left(0,\mathcal{E}(w_L,Z_L),Z_L,\mathcal{E}(w_R,Z_R),Z_R\right)\right)-\left(0,\frac{\operatorname{sgn}(u^*)}{2}g\bar{h}\bar{\theta}[Z],0\right)^T. \tag{4.141}$$

Il reste à montrer que ce flux numérique peut se réécrire sous la forme (4.134), ce qui se voit facilement d'après la structure du problème de Riemann (voir Figure 4.8).

Enfin, la consistance du terme source est immédiate d'après (4.135). Concernant la consistance du flux numérique, si $w_L = w_R = w$ et $Z_L = Z_R = Z$, on déduit aisément des équations (4.124) à (4.128) que $w_L^* = w_R^* = w$. La formulation (4.134) du flux numérique implique alors

$$f(w, Z, w, Z) = f(w).$$

□

Pour conclure la présentation du schéma, on montre un résultat qui sera utile pour l'extension en deux dimensions d'espace.

Lemme 4.27. *Supposons que la condition CFL*

$$\frac{\Delta t}{\Delta x} \max\left\{|u_L - \nu/h_L|, |h_R + \nu/h_R|\right\} \leq \frac{1}{2} \tag{4.142}$$

est vérifiée, où ν est la linéarisation de l'impédance acoustique utilisée dans le problème de Riemann $W_{\mathcal{R}}(\xi, W_L, Z_L, W_R, Z_R)$.

Alors le flux numérique f et le terme source numérique s vérifient l'identité suivante :

$$\begin{aligned} f(w_L, Z_L, w_R, Z_R) - \frac{s(w_L, w_R)}{2}[Z] = f(w_L) + \frac{\Delta x}{2\Delta t} w_L \\ - \frac{1}{\Delta t} \int_{-\Delta x/2}^{0} \mathcal{Q} W_{\mathcal{R}}\left(\frac{x}{\Delta t}, \mathcal{E}(w_L, Z_L), Z_L, \mathcal{E}(w_R, Z_R), Z_R\right) dx. \end{aligned} \tag{4.143}$$

Démonstration. Pour simplifier les notations, on introduit $W_L = \mathcal{E}(w_L, Z_L)$ et $W_R = \mathcal{E}(w_R, Z_R)$. Les définitions (4.141) du flux numérique f et (4.135) du terme source numérique s impliquent

$$\begin{aligned} f(w_L, Z_L, w_R, Z_R) - \frac{s(w_L, w_R)}{2}[Z] = \mathcal{Q}F\left(W_{\mathcal{R}}\left(0, W_L, Z_L, W_R, Z_R\right)\right) - \left(0, \frac{\operatorname{sgn}(u^*)}{2} g\bar{h}\bar{\theta}[Z], 0\right)^T \\ - \left(0, -\frac{1}{2} g\bar{h}\bar{\theta}[Z], 0\right)^T, \end{aligned}$$

que l'on peut réécrire sous la forme

$$\begin{aligned} f(w_L, Z_L, w_R, Z_R) - \frac{s(w_L, w_R)}{2}[Z] = \mathcal{Q}F\left(W_{\mathcal{R}}\left(0, W_L, Z_L, W_R, Z_R\right)\right) \\ - \left(0, \frac{\operatorname{sgn}(u^*) - 1}{2} g\bar{h}\bar{\theta}[Z], 0\right)^T. \end{aligned}$$

En utilisant les équations (4.139) et (4.140), on obtient

$$\begin{aligned} f(w_L, Z_L, w_R, Z_R) - \frac{s(w_L, w_R)}{2}[Z] = \mathcal{Q}F\left(W_{\mathcal{R}}\left(0, W_L, Z_L, W_R, Z_R\right)\right) \\ - \frac{1}{\Delta t} \int_{-\Delta x/2}^{0} \int_{0}^{\Delta t} \mathcal{Q}S\left(W_{\mathcal{R}}\left(\frac{x}{t}, W_L, Z_L, W_R, Z_R\right)\right) dt dx. \end{aligned} \tag{4.144}$$

D'autre part, en intégrant la solution exacte du problème de Riemann pour le système (4.129) sur le rectangle $[-\Delta x/2, 0] \times [0, \Delta t]$, on trouve grâce à la condition CFL (4.142) :

$$\begin{aligned} \int_{-\Delta x/2}^{0} W_{\mathcal{R}}\left(\frac{x}{\Delta t}, W_L, Z_L, W_R, Z_R\right) dx - \frac{\Delta x}{2} W_L + \int_{0}^{\Delta t} F\left(W_{\mathcal{R}}\left(0, W_L, Z_L, W_R, Z_R\right)\right) dt \\ - \Delta t F(W_L) = \int_{-\Delta x/2}^{0} \int_{0}^{\Delta t} S\left(W_{\mathcal{R}}\left(\frac{x}{t}, W_L, Z_L, W_R, Z_R\right)\right) dt dx. \end{aligned}$$

En multipliant à gauche par $\frac{1}{\Delta t}\mathcal{Q}$, on en déduit

$$\begin{aligned}\mathcal{Q}F\left(W_{\mathcal{R}}\left(0,W_L,Z_L,W_R,Z_R\right)\right) - \frac{1}{\Delta t}\int_{-\Delta x/2}^{0}\int_{0}^{\Delta t}\mathcal{Q}S\left(W_{\mathcal{R}}\left(\frac{x}{t},W_L,Z_L,W_R,Z_R\right)\right)dtdx \\ = \mathcal{Q}F(W_L) + \frac{\Delta x}{2\Delta t}\mathcal{Q}W_L - \frac{1}{\Delta t}\int_{-\Delta x/2}^{0}\mathcal{Q}W_{\mathcal{R}}\left(\frac{x}{\Delta t},W_L,Z_L,W_R,Z_R\right)dx.\end{aligned}$$

Les propriétés (4.92) et (4.90) permettent alors d'écrire

$$\begin{aligned}\mathcal{Q}F\left(W_{\mathcal{R}}\left(0,W_L,Z_L,W_R,Z_R\right)\right) - \frac{1}{\Delta t}\int_{-\Delta x/2}^{0}\int_{0}^{\Delta t}\mathcal{Q}S\left(W_{\mathcal{R}}\left(\frac{x}{t},W_L,Z_L,W_R,Z_R\right)\right)dtdx \\ = f(w_L) + \frac{\Delta x}{2\Delta t}w_L - \frac{1}{\Delta t}\int_{-\Delta x/2}^{0}\mathcal{Q}W_{\mathcal{R}}\left(\frac{x}{\Delta t},W_L,Z_L,W_R,Z_R\right)dx.\end{aligned}$$

En injectant cette relation dans (4.144), on trouve le résultat attendu. □

Propriétés vérifiées par le schéma

On établit d'abord que le schéma est well-balanced.

Proposition 4.28. *Le schéma (4.132) est well-balanced, c'est-à-dire que si $\forall i \in \mathbb{Z}$ on a*

$$u_i^n = 0 \quad et \quad \frac{(h_{i+1}^n)^2\theta_{i+1}^n}{2} - \frac{(h_i^n)^2\theta_i^n}{2} + \frac{h_i^n + h_{i+1}^n}{2}\frac{\theta_i^n + \theta_{i+1}^n}{2}(Z_{i+1} - Z_i) = 0,$$

alors $w_i^{n+1} = w_i^n$, $\forall i \in \mathbb{Z}$.

Démonstration. Soient (w_L, Z_L) et (w_R, Z_R) deux états vérifiant l'équilibre local (4.10). D'après le lemme 4.21, il suffit de montrer que

$$W_{\mathcal{R}}(\xi, \mathcal{E}(w_L,Z_L), Z_L, \mathcal{E}(w_R,Z_R), Z_R) = \begin{cases}\mathcal{E}(w_L,Z_L) & \text{si } \xi < 0,\\ \mathcal{E}(w_R,Z_R) & \text{si } \xi > 0.\end{cases}$$

Par définition de la fonction $\mathcal{E}$ et puisque $[h^2\theta/2] = -\bar{h}\bar{\theta}[Z]$, les états $W_L = \mathcal{E}(w_L, Z_L)$ et $W_R = \mathcal{E}(w_R, Z_R)$ vérifient

$$\begin{aligned}[\pi] &= [gh^2\theta/2]\\ &= -g\bar{h}\bar{\theta}[Z]\\ &= -g\bar{h}\bar{\theta}[a].\end{aligned}$$

D'autre part, on a $u_L = u_R = 0$. On déduit de l'équation (4.124) que $u^* = 0$. Les équations (4.125) à (4.128) impliquent alors

$$\pi_L^* = \pi_L, \quad \pi_R^* = \pi_R, \quad h_L^* = h_L, \quad h_R^* = h_R,$$

ce qui conclut la preuve. □

Montrons maintenant que le schéma est robuste.

Proposition 4.29. *Supposons que la constante $\nu_{i+1/2}$ assure que les valeurs propres vérifient l'ordre suivant :*

$$u_i^n - \frac{\nu_{i+1/2}}{h_i^n} < u_{i+1/2}^* < u_{i+1}^n + \frac{\nu_{i+1/2}}{h_{i+1}^n}. \tag{4.145}$$

Si la condition CFL

$$\frac{\Delta t}{\Delta x}\max_{i\in\mathbb{Z}}\left|u_i^n \pm \frac{\nu_{i\mp1/2}}{h_i^n}\right| \leq \frac{1}{2}$$

est vérifiée, alors le schéma (4.132) est robuste.

Démonstration. D'après le lemme 4.22, il suffit de montrer que pour tous W_L et W_R dans $\mathcal{O}$ et tous Z_L et Z_R dans $\mathbb{R}$, on a $W_L^* \in \mathcal{O}$ et $W_R^* \in \mathcal{O}$. Puisque les valeurs de θ dans les états intermédiaires sont toujours soit θ_L, soit θ_R, il suffit de montrer que $h_L^* > 0$ et $h_R^* > 0$. En utilisant la continuité des invariants de Riemann $u \pm \frac{\nu}{h}$ pour les ondes $u \pm \frac{\nu}{h}$, on déduit que l'hypothèse (4.145) se réécrit de façon équivalente

$$u^* - \frac{\nu}{h_L^*} < u^* < u^* + \frac{\nu}{h_R^*}.$$

En conséquence, imposer (4.145) revient à imposer $h_L^* > 0$ et $h_R^* > 0$. On en déduit immédiatement la robustesse attendue. □

4.4.4 Schéma de relaxation avec transport de gravité pour les équations d'Euler avec gravité

Le but de cette partie est de construire un schéma de relaxation well-balanced pour les équations d'Euler avec gravité (4.11) pour une loi de pression générale (4.12). Le modèle de Suliciu présentant les mêmes difficultés que pour les autres systèmes, on introduit la version modifiée où l'on transporte le terme de gravité à la vitesse u.

Modèle de relaxation avec transport de gravité

On introduit une nouvelle variable a qui va être transportée à la vitesse u et relaxée vers le terme de gravité $Z(x) = x$. Cela mène au modèle de relaxation suivant :

$$\begin{cases} \partial_t \rho + \partial_x \rho u = 0, \\ \partial_t \rho u + \partial_x (\rho u^2 + \pi) = -g\rho\partial_x a, \\ \partial_t E + \partial_x (u(E+\pi)) = -g\rho u \partial_x a, \\ \partial_t \rho\pi + \partial_x (u(\rho\pi + \nu^2)) = \dfrac{h}{\varepsilon}(p(\rho,e) - \pi), \\ \partial_t \rho a + \partial_x \rho u a = \frac{h}{\varepsilon}(Z - a). \end{cases} \tag{4.146}$$

Ce système entre dans le formalisme introduit dans la partie 4.4.1 avec

$$W = (\rho, \rho u, E, \rho\pi, \rho a)^T,$$

$$F(W) = (\rho u, \rho u^2 + \pi, u(E+\pi), u(\rho\pi + \nu^2), \rho u a)^T,$$

$$S(W) = (0, -g\rho\partial_x a, -g\rho u \partial_x a, 0, 0)^T,$$

$$R(W, Z) = (0, 0, 0, h(p(\rho, e) - \pi), h(Z - a))^T.$$

L'ensemble des états admissibles pour ce système est

$$\mathcal{O} = \left\{ W = (\rho, \rho u, E, \rho\pi, \rho a)^T \in \mathbb{R}^5, \rho > 0, E - \frac{1}{2}\rho u^2 > 0 \right\}.$$

La matrice $\mathcal{Q}$ représente la projection sur les trois premières composantes de W :

$$\mathcal{Q} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

La fonction déterminant l'équilibre est

$$\mathcal{E}(w, Z) = (\rho, \rho u, E, \rho p(\rho, e), \rho Z)^T$$

et la variété d'équilibre est définie par

$$\mathcal{M} = \left\{ W = (\rho, \rho u, E, \rho\pi, \rho a)^T, \pi = p(\rho, e), a = Z \right\}.$$

On introduit le vecteur des grandeurs physiques $U = (\rho, u, e, \pi, a)^T$. On peut alors réécrire le système $(4.146)_{\varepsilon=+\infty}$ sous forme quasi-linéaire

$$\partial_t U + A(U)\partial_x U = 0,$$

où la matrice $A(U)$ est donnée par

$$A(U) = \begin{pmatrix} u & \rho & 0 & 0 & 0 \\ 0 & u & 0 & 1/\rho & g \\ 0 & \pi/\rho & u & 0 & 0 \\ 0 & \nu^2/\rho & 0 & u & 0 \\ 0 & 0 & 0 & 0 & u \end{pmatrix}.$$

Ce système est hyperbolique en tout point de $\mathcal{O}$, avec pour valeurs propres u de multiplicité 3 et $u \pm \frac{\nu}{h}$, chacune de multiplicité 1. L'ordre des valeurs propres est toujours le même et les champs qui leur sont associés sont tous linéairement dégénérés. Les invariants de Riemann pour ce système sont pour la valeur propre $u \pm \frac{\nu}{\rho}$:

$$a, \quad u \pm \frac{\nu}{\rho}, \quad \pi \mp \nu u, \quad \nu^2 e - \frac{\pi^2}{2}$$

et pour la valeur propre u :

$$u.$$

Il manque un invariant de Riemann pour l'onde u. En effet, celle-ci est de multiplicité 3 et l'on pourrait s'attendre à ce qu'elle ait deux invariants de Riemann. Une équation supplémentaire sera donc nécessaire pour que le système soit complètement déterminé

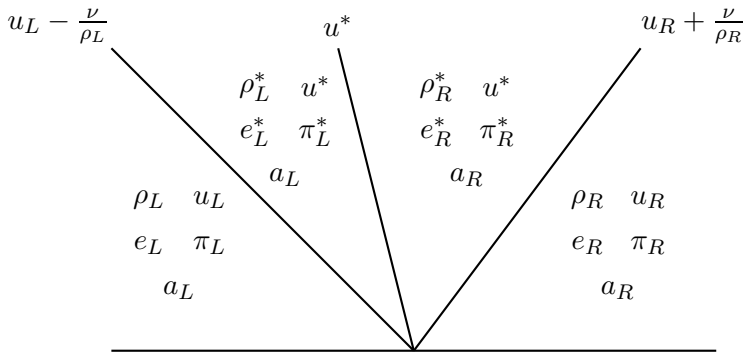


FIGURE 4.10 – Structure de la solution « exacte » du problème de Riemann pour le système de relaxation (4.146)

On admet que la solution « exacte » du problème de Riemann pour le système de relaxation $(4.146)_{\varepsilon=+\infty}$ est composée de deux états intermédiaires W_L^* et W_R^* séparés par trois discontinuités de contact de vitesses $u_L - \frac{\nu}{\rho_L}$, u^* et $u_R + \frac{\nu}{\rho_R}$ (voir Figure 4.10).

Après avoir éliminé les invariants de Riemann triviaux (dont celui pour l'onde u qui amène à définir $u^* = u_L^* = u_R^*$), il reste sept inconnues : h_L^*, u^*, e_L^*, π_L^*, h_R^*, e_R^* et π_R^*. La solution du

problème de Riemann pour (4.146) s'écrit

$$W_{\mathcal{R}}\left(\frac{x}{t}, W_L, Z_L, W_R, Z_R\right) = \begin{cases} (\rho_L, \rho_L u_L, E_L, \rho_L \pi_L, \rho_L a_L)^T & \text{si } \frac{x}{t} < u_L - \frac{\nu}{\rho_L}, \\ (\rho_L^*, \rho_L^* u^*, E_L^*, \rho_L^* \pi_L^*, \rho_L^* a_L)^T & \text{si } u_L - \frac{\nu}{\rho_L} < \frac{x}{t} < u^*, \\ (\rho_R^*, \rho_R^* u^*, E_R^*, \rho_R^* \pi_R^*, \rho_R^* a_R)^T & \text{si } u^* < \frac{x}{t} < u_R + \frac{\nu}{\rho_R}, \\ (\rho_R, \rho_R u_R, E_R, \rho_R \pi_R, \rho_R a_R)^T & \text{si } u_R + \frac{\nu}{\rho_R} < \frac{x}{t}. \end{cases}$$

On choisit de rajouter l'équation provenant de la définition de l'équilibre local, c'est-à-dire

$$\pi_R^* - \pi_L^* = -g\overline{\rho}[a]. \tag{4.147}$$

On obtient le système composé des sept équations suivantes :

$$u_L - \frac{\nu}{\rho_L} = u^* - \frac{\nu}{\rho_L^*},$$

$$u_R + \frac{\nu}{\rho_R} = u^* + \frac{\nu}{\rho_R^*},$$

$$\pi_L + \nu u_L = \pi_L^* + \nu u^*,$$

$$\pi_R - \nu u_R = \pi_R^* - \nu u^*,$$

$$\nu^2 e_L - \frac{\pi_L^2}{2} = \nu^2 e_L^* - \frac{(\pi_L^*)^2}{2},$$

$$\nu^2 e_R - \frac{\pi_R^2}{2} = \nu^2 e_R^* - \frac{(\pi_R^*)^2}{2},$$

$$\pi_R^* - \pi_L^* = -g\overline{\rho}[a].$$

On peut alors résoudre le système et on trouve

$$u^* = \overline{u} - \frac{[\pi]}{2\nu} - g\overline{\rho}\frac{[a]}{2\nu}, \tag{4.148}$$

$$\pi_L^* = \pi_L + \nu(u_L - u^*), \tag{4.149}$$

$$\pi_R^* = \pi_R + \nu(u^* - u_R), \tag{4.150}$$

$$\frac{1}{\rho_L^*} = \frac{1}{\rho_L} + \frac{u^* - u_L}{\nu}, \tag{4.151}$$

$$\frac{1}{\rho_R^*} = \frac{1}{\rho_R} + \frac{u_R - u^*}{\nu}, \tag{4.152}$$

$$e_L^* = e_L + \frac{{\pi_L^*}^2 - {\pi_L}^2}{2\nu^2}, \tag{4.153}$$

$$e_R^* = e_R + \frac{{\pi_R^*}^2 - {\pi_R}^2}{2\nu^2}. \tag{4.154}$$

Avant de présenter le schéma associé à ce solveur de Riemann et de montrer qu'il possède toutes les propriétés requises, on va justifier le choix de l'équation (4.147) en introduisant un nouveau modèle de relaxation.

Reformulation en un modèle complètement déterminé

On présente maintenant le système de relaxation avec un ensemble complet d'invariants de Riemann qui permet de justifier du choix de l'équation (4.147). Pour cela, on introduit deux nouvelles variables X^- et X^+ qui vont être relaxées vers ρ. Ces deux variables sont transportées respectivement à la vitesse $u-\delta$ et $u+\delta$, où $\delta>0$ est un paramètre suffisamment petit. On obtient le modèle de relaxation suivant composé de sept équations

$$\left\{\begin{array}{l}\partial_t\rho+\partial_x\rho u=0,\\ \partial_t\rho u+\partial_x(\rho u^2+\pi)=-g\dfrac{X^-+X^+}{2}\partial_x a,\\ \partial_t E+\partial_x(u(E+\pi))=-g\dfrac{X^-+X^+}{2}u\partial_x a,\\ \partial_t\rho\pi+\partial_x(u(\rho\pi+\nu^2))=\dfrac{h}{\varepsilon}(p(\rho,e)-\pi),\\ \partial_t\rho a+\partial_x\rho u a=\dfrac{h}{\varepsilon}(Z-a),\\ \partial_t\rho X^-+\partial_x\rho u X^-=\delta\rho\partial_x X^-+\dfrac{h}{\varepsilon}(\rho-X^-),\\ \partial_t\rho X^++\partial_x\rho u X^+=-\delta\rho\partial_x X^++\dfrac{h}{\varepsilon}(\rho-X^+).\end{array}\right. \tag{4.155}$$

On peut écrire ce système sous la forme (4.89) avec

$$W=(\rho,\rho u,E,\rho\pi,\rho a,\rho X^-,\rho X^+)^T,$$
$$F(W)=(\rho u,\rho u^2+\pi,u(E+\pi),u(\rho\pi+\nu^2),\rho ua,\rho uX^-,\rho uX^+)^T,$$
$$S(W)=\left(0,-g\frac{X^-+X^+}{2}\partial_x a,-g\frac{X^-+X^+}{2}u\partial_x a,0,0,\delta\rho\partial_x X^-,-\delta\rho\partial_x X^+\right)^T,$$
$$R(W,Z)=(0,0,0,h(p(\rho,e)-\pi),h(Z-a),h(\rho-X^-),h(\rho-X^+))^T.$$

L'ensemble des vecteurs admissibles est

$$\mathcal{O}=\left\{W=(\rho,\rho u,E,\rho\pi,\rho a,\rho X^-,\rho X^+)^T\in\mathbb{R}^7,\rho>0,E-\frac{1}{2}\rho u^2>0\right\}.$$

La matrice $\mathcal{Q}$ représente à nouveau la projection sur les trois premières composantes de W. La fonction décrivant l'équilibre est

$$\mathcal{E}(w,Z)=(\rho,\rho u,E,\rho p(\rho,e),\rho Z,\rho^2,\rho^2)^T$$

et la variété d'équilibre est définie par

$$\mathcal{M}=\left\{W=(\rho,\rho u,E,\rho\pi,\rho a,\rho X^-,\rho X^+)^T,\pi=p(\rho,e),a=Z,X^-=\rho,X^+=\rho\right\}.$$

On introduit le vecteur des grandeurs physiques $U=(\rho,u,e,\pi,a,X^-,X^+)^T$. Le système de relaxation $(4.155)_{\varepsilon=+\infty}$ se réécrit alors sous la forme quasi-linéaire

$$\partial_t U+A(U)\partial_x U=0,$$

où la matrice $A(U)$ est donnée par

$$A(U)=\begin{pmatrix}u&\rho&0&0&0&0&0\\0&u&0&1/\rho&g\frac{X^-+X^+}{2\rho}&0&0\\0&\pi/\rho&u&0&0&0&0\\0&\nu^2/\rho&0&u&0&0&0\\0&0&0&0&u&0&0\\0&0&0&0&0&u-\delta&0\\0&0&0&0&0&0&u+\delta\end{pmatrix}.$$

Si δ est choisi suffisamment petit, ce système est hyperbolique en tout point de $\mathcal{O}$. Ses valeurs propres sont $u \pm \delta$ et $u \pm \frac{\nu}{\rho}$, chacune de multiplicité 1 et u de multiplicité 3. Pour que les valeurs propres ne se croisent pas, il suffit de choisir δ tel que $\delta < \frac{\nu}{\rho}$. L'ordre des valeurs propres est alors fixe :

$$u - \frac{\nu}{\rho} < u - \delta < u < u + \delta < u + \frac{\nu}{\rho}.$$

Toutes les valeurs propres sont associées à des champs linéairement dégénérés. Les invariants de Riemann sont pour la valeur propre u :

$$u, \quad \pi, \quad X^-, \quad X^+, \quad \pi + g\frac{X^- + X^-}{2}a,$$

pour la valeur propre $u \pm \frac{\nu}{\rho}$:

$$a, \quad X^-, \quad X^+, \quad u \pm \frac{\nu}{\rho}, \quad \pi \mp \nu u, \quad \nu^2 e - \frac{\pi^2}{2},$$

pour la valeur propre $u - \delta$:

$$\rho, \quad u, \quad e, \quad \pi, \quad a, \quad X^+,$$

pour la valeur propre $u + \delta$:

$$\rho, \quad u, \quad e, \quad \pi, \quad a, \quad X^-.$$

Il n'y a pas d'invariant manquant pour ce système. Le problème de Riemann a donc une unique solution déterminée par la continuité des invariants de Riemann à travers leur onde respective.

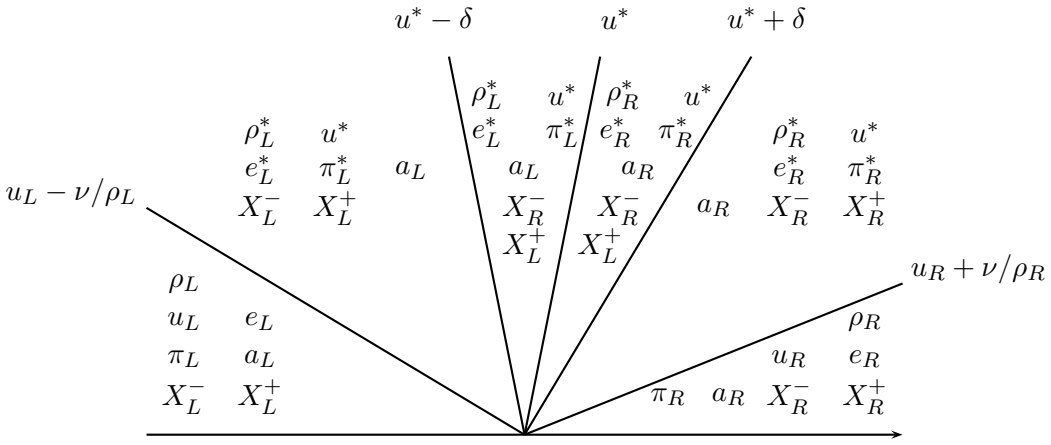


FIGURE 4.11 – Structure de la solution exacte du problème de Riemann pour le système de relaxation (4.155)

La solution du problème de Riemann pour le système $(4.155)_{\varepsilon=+\infty}$ est composée de quatre états intermédiaires séparés par cinq discontinuités de contact (voir Figure 4.11). Après avoir éliminé les relations liées aux invariants de Riemann triviaux , il reste sept inconnues : ρ_L^*, ρ_R^*, u^*, e_L^*, e_R^*, π_L^* et π_R^*. Les équations données par les invariants de Riemann $u \pm \frac{\nu}{\rho}$, $\pi \mp \nu u$ et $\nu^2 e - \frac{\pi^2}{2}$ pour les ondes $u \pm \frac{\nu}{\rho}$ nous donnent six équations qui étaient déjà présentes dans le précédent système. Le dernier invariant de Riemann non trivial, $\pi + g\frac{X^- + X^+}{2}a$ pour l'onde u nous donne

$$\pi_R^* - \pi_L^* = -g\frac{X_R^- + X_L^+}{2}(a_R - a_L). \tag{4.156}$$

Si l'on suppose que la condition initiale appartient à la variété d'équilibre, alors on a $X_L^+ = \rho_L$ et $X_R^- = \rho_R$ et l'équation (4.156) est la même que la dernière équation (4.147) du système précédent. Par conséquent, les deux modèles possèdent la « même » solution du problème de Riemann pour une donnée à l'équilibre.

Le schéma de relaxation

Connaissant la solution du problème de Riemann pour le système $(4.155)_{\varepsilon=+\infty}$, on déduit un schéma numérique pour le système original (4.11) qui est décrit dans la proposition suivante.

Proposition 4.30. *Le schéma de relaxation associé au modèle de relaxation (4.155) s'écrit*

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(f_{i+1/2} - f_{i-1/2}\right) + \frac{\Delta t}{2}\left(s(w_{i-1}^n, w_i^n)\frac{Z_i - Z_{i-1}}{\Delta x} + s(w_i^n, w_{i+1}^n)\frac{Z_{i+1} - Z_i}{\Delta x}\right), \tag{4.157}$$

où le flux numérique est défini par

$$f_{i+1/2} = f(w_i^n, Z_i, w_{i+1}^n, Z_{i+1}), \tag{4.158}$$

$$f(w_L, Z_L, w_R, Z_R) = \begin{cases} \left(\rho_L u_L, \rho_L u_L^2 + \pi_L - \frac{1}{2}g\overline{\rho}[Z], u_L(E_L + \pi_L) - \frac{1}{2}g\overline{\rho}u^*[Z]\right)^T & \text{si } u_L - \frac{\nu}{\rho_L} > 0, \\ \left(\rho_L^* u^*, \rho_L^*(u^*)^2 + \pi_L^* - \frac{1}{2}g\overline{\rho}[Z], u^*(E_L^* + \pi_L^*) - \frac{1}{2}g\overline{\rho}u^*[Z]\right)^T & \text{si } u_L - \frac{\nu}{\rho_L} < 0 < u^*, \\ \left(\rho_R^* u^*, \rho_R^*(u^*)^2 + \pi_R^* + \frac{1}{2}g\overline{\rho}[Z], u^*(E_R^* + \pi_R^*) + \frac{1}{2}g\overline{\rho}u^*[Z]\right)^T & \text{si } u^* < 0 < u_R + \frac{\nu}{\rho_R}, \\ \left(\rho_R u_R, \rho_R u_R^2 + \pi_R + \frac{1}{2}g\overline{\rho}[Z], u_R(E_R + \pi_R) + \frac{1}{2}g\overline{\rho}u^*[Z]\right)^T & \text{si } u_R + \frac{\nu}{\rho_R} < 0, \end{cases} \tag{4.159}$$

et le terme source numérique est défini par

$$s(w_L, w_R) = (0, -g\overline{\rho}, -g\overline{\rho}u^*)^T. \tag{4.160}$$

De plus, le schéma (4.157) est consistant avec (4.11).

Démonstration. Écrivons le système $(4.155)_{\varepsilon=+\infty}$ sous la forme

$$\partial_t W + \partial_x F(W) = S(W). \tag{4.161}$$

On note $W_{\Delta x}$ la solution exacte de ce système pour la condition initiale

$$W_{\Delta x}(x, t^n) = W_i^n = \mathcal{E}(w_i^n, Z_i) \quad \text{si } x \in K_i.$$

On peut alors intégrer (4.161) sur le rectangle $K_i \times [t^n, t^n + \Delta t]$ pour obtenir

$$\begin{aligned}\int_{K_i} W_{\Delta x}(x, t^n + \Delta t)dx = \int_{K_i} W_{\Delta x}(x, t^n)dx - \int_{t^n}^{t^n+\Delta t} F\left(W_{\Delta x}(x_{i+1/2}, t)\right)dt \\ + \int_{t^n}^{t^n+\Delta t} F\left(W_{\Delta x}(x_{i-1/2}, t)\right)dt + \int_{K_i}\int_{t^n}^{t^n+\Delta t} S\left(W_{\Delta x}(x, t)\right)dtdx.\end{aligned}$$

En utilisant l'auto-similarité de la solution du problème de Riemann, on en déduit

$$\begin{aligned}W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x}\left(F\left(W_{\mathcal{R}}\left(0, W_i^n, Z_i, W_{i+1}^n, Z_{i+1}\right)\right) - F\left(W_{\mathcal{R}}\left(0, W_{i-1}^n, Z_{i-1}, W_i^n, Z_i\right)\right)\right) \\ + \frac{1}{\Delta x}\int_{K_i}\int_{t^n}^{t^n+\Delta t} S\left(W_{\Delta x}(x, t)\right)dtdx.\end{aligned}$$

On multiplie cette équation par $\mathcal{Q}$ et, puisque $w_i^n = \mathcal{Q}W_i^n$ et $w_i^{n+1} = \mathcal{Q}W_i^{n+1}$, on trouve

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(\mathcal{Q}F\left(W_{\mathcal{R}}\left(0, W_i^n, Z_i, W_{i+1}^n, Z_{i+1}\right)\right) - \mathcal{Q}F\left(W_{\mathcal{R}}\left(0, W_{i-1}^n, Z_{i-1}, W_i^n, Z_i\right)\right)\right) + \frac{1}{\Delta x}\int_{K_i}\int_{t^n}^{t^n+\Delta t} \mathcal{Q}S\left(W_{\Delta x}(x,t)\right) dtdx, \quad (4.162)$$

On va ensuite découper l'intégrale du terme source en deux :

$$\frac{1}{\Delta x}\int_{K_i}\int_{t^n}^{t^n+\Delta t} \mathcal{Q}S\left(W_{\Delta x}(x,t)\right) dtdx = \Delta t\left(S^+(W_{i-1}^n, Z_{i-1}, W_i^n, Z_i) + S^-(W_i^n, Z_i, W_{i+1}^n Z_{i+1})\right) \quad (4.163)$$

où les fonctions $S^\pm$ sont définies par

$$S^-(W_L, Z_L, W_R, Z_R) = \frac{1}{\Delta t \Delta x}\int_{-\Delta x/2}^0 \int_0^{\Delta t} \mathcal{Q}S\left(W_{\mathcal{R}}\left(\frac{x}{t}, W_L, Z_L, W_R, Z_R\right)\right) dtdx,$$

$$S^+(W_L, Z_L, W_R, Z_R) = \frac{1}{\Delta t \Delta x}\int_0^{\Delta x/2} \int_0^{\Delta t} \mathcal{Q}S\left(W_{\mathcal{R}}\left(\frac{x}{t}, W_L, Z_L, W_R, Z_R\right)\right) dtdx.$$

En utilisant les définitions de S et de $\mathcal{Q}$ dans le cas du système (4.155), on a

$$\mathcal{Q}S(W) = \left(0, -g\frac{X^- + X^+}{2}\partial_x a, -g\frac{X^- + X^+}{2}u\partial_x a\right)^T.$$

Dans un premier temps, on s'intéresse à la solution du problème de Riemann dans le rectangle $[-\Delta x/2, 0] \times [0, \Delta t]$. Si $u^* > 0$ alors a est constante, donc

$$S^-(W_L, Z_L, W_R, Z_R) = (0,0,0)^T.$$

Si $u^* < 0$, alors a n'est discontinue que le long de la droite de vitesse u^* et les inconnues X^-, X^+ et u restant constantes le long de cette droite, on trouve par la formule des sauts

$$S^-(W_L, Z_L, W_R, Z_R) = \left(0, -g\frac{X_L^- + X_R^+}{2}\frac{a_R - a_L}{\Delta x}, -g\frac{X_L^- + X_R^+}{2}u^*\frac{a_R - a_L}{\Delta x}\right)^T.$$

Pour des états W_L et W_R dans la variété d'équilibre $\mathcal{M}$, on a $X_L^+ = \rho_L$, $X_R^- = \rho_R$, $a_L = Z_L$ et $a_R = Z_R$. Les états $\mathcal{E}(w_L, Z_L)$ et $\mathcal{E}(w_R, Z_R)$ étant par définition dans la variété d'équilibre $\mathcal{M}$, on a donc

$$S^-\left(\mathcal{E}(w_L, Z_L), Z_L, \mathcal{E}(w_R, Z_R), Z_R\right) = \left(0, -g\overline{\rho}\frac{[Z]}{\Delta x}, -g\overline{\rho}u^*\frac{[Z]}{\Delta x}\right)^T.$$

On peut unifier les cas $u^* < 0$ et $u^* > 0$ par la formule

$$S^-\left(\mathcal{E}(w_L, Z_L), Z_L, \mathcal{E}(w_R, Z_R), Z_R\right) = \left(0, \frac{\mathrm{sgn}(u^*) - 1}{2}g\overline{\rho}\frac{[Z]}{\Delta x}, \frac{\mathrm{sgn}(u^*) - 1}{2}g\overline{\rho}u^*\frac{[Z]}{\Delta x}\right)^T.$$

On procède de la même façon dans le rectangle $[0, \Delta x/2] \times [0, \Delta t]$ pour trouver

$$S^+\left(\mathcal{E}(w_L, Z_L), Z_L, \mathcal{E}(w_R, Z_R), Z_R\right) = \left(0, -\frac{\mathrm{sgn}(u^*) + 1}{2}g\overline{\rho}\frac{[Z]}{\Delta x}, -\frac{\mathrm{sgn}(u^*) + 1}{2}g\overline{\rho}u^*\frac{[Z]}{\Delta x}\right)^T.$$

En rappelant que $W_i^n = \mathcal{E}(w_i^n, Z_i)$, on déduit de (4.163) que l'intégrale du terme source s'écrit

$$\frac{1}{\Delta x}\int_{K_i}\int_{t^n}^{t^n+\Delta t} \mathcal{Q}S\left(W_{\Delta x}(x,t)\right)dtdx = \frac{\Delta t}{2}\left(s(w_{i-1}^n, w_i^n)\frac{Z_i - Z_{i-1}}{\Delta x} + s(w_i^n, w_{i+1}^n)\frac{Z_{i+1}-Z_i}{\Delta x}\right)$$
$$- \frac{\Delta t}{\Delta x}\left(0, \frac{\mathrm{sgn}(u^*_{i-1/2})}{2} g\frac{\rho_{i-1}^n + \rho_i^n}{2}(Z_i - Z_{i-1}) - \frac{\mathrm{sgn}(u^*_{i+1/2})}{2} g\frac{\rho_i^n + \rho_{i+1}^n}{2}(Z_{i+1} - Z_i),\right.$$
$$\left.\frac{\mathrm{sgn}(u^*_{i-1/2})}{2} g\frac{\rho_{i-1}^n + \rho_i^n}{2} u^*_{i-1/2}(Z_i - Z_{i-1}) - \frac{\mathrm{sgn}(u^*_{i+1/2})}{2} g\frac{\rho_i^n + \rho_{i+1}^n}{2} u^*_{i+1/2}(Z_{i+1} - Z_i)\right)^T,$$

où $s(w_L, w_R)$ est défini par (4.160). En injectant cette équation dans (4.162), on trouve

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(f_{i+1/2} - f_{i-1/2}\right) + \frac{\Delta t}{2}\left(s(w_{i-1}^n, w_i^n)\frac{Z_i - Z_{i-1}}{\Delta x} + s(w_i^n, w_{i+1}^n)\frac{Z_{i+1}-Z_i}{\Delta x}\right),$$

où le flux numérique est défini par (4.158) et

$$f(w_L, Z_L, w_R, Z_R) = \mathcal{Q}F\left(W_{\mathcal{R}}\left(0, \mathcal{E}(w_L, Z_L), Z_L, \mathcal{E}(w_R, Z_R), Z_R\right)\right)$$
$$- \left(0, \frac{\mathrm{sgn}(u^*)}{2} g\overline{\rho}[Z], \frac{\mathrm{sgn}(u^*)}{2} g\overline{\rho}u^*[Z]\right)^T.$$

Il reste à montrer que ce flux numérique peut se réécrire sous la forme (4.159), ce qui se voit facilement d'après la structure du problème de Riemann (voir Figure 4.10).

Enfin, la consistance du terme source est immédiate d'après (4.160). Concernant la consistance du flux numérique, si $w_L = w_R = w$ et $Z_L = Z_R = Z$, on déduit aisément des équations (4.148) à (4.154) que $w_L^* = w_R^* = w$. La formulation (4.159) du flux numérique implique alors

$$f(w, Z, w, Z) = f(w).$$

□

Propriétés vérifiées par le schéma

On établit d'abord que le schéma est well-balanced.

Proposition 4.31. *Le schéma (4.157) est well-balanced, c'est-à-dire que si $\forall i \in \mathbb{Z}$ on a*

$$u_i^n = 0 \quad et \quad p_{i+1}^n - p_i^n + g\frac{\rho_i^n + \rho_{i+1}^n}{2}(Z_{i+1} - Z_i) = 0,$$

alors $w_i^{n+1} = w_i^n$, $\forall i \in \mathbb{Z}$.

Démonstration. Soient (w_L, Z_L) et (w_R, Z_R) deux états vérifiant l'équilibre local (4.14). D'après le Lemme 4.21, il suffit de montrer que

$$W_{\mathcal{R}}(\xi, \mathcal{E}(w_L, Z_L), Z_L, \mathcal{E}(w_R, Z_R), Z_R) = \begin{cases} \mathcal{E}(w_L, Z_L) & \text{si } \xi < 0, \\ \mathcal{E}(w_R, Z_R) & \text{si } \xi > 0. \end{cases}$$

Par définition de la fonction $\mathcal{E}$, les états $W_L = \mathcal{E}(w_L, Z_R)$ et $W_R = \mathcal{E}(w_R, Z_R)$ vérifient

$$[\pi] + g\overline{\rho}[a] = 0.$$

D'autre part, on a $u_L = u_R = 0$. On déduit de l'équation (4.148) que $u^* = 0$. Les équations (4.149) à (4.154) impliquent alors

$$\pi_L^* = \pi_L, \quad \pi_R^* = \pi_R, \quad \rho_L^* = \rho_L, \quad \rho_R^* = \rho_R, \quad e_L^* = e_L, \quad e_R^* = e_R,$$

ce qui conclut la preuve. □

Montrons maintenant que le schéma est robuste.

Proposition 4.32. *Supposons que la constante $\nu_{i+1/2}$ vérifie les inégalités suivantes :*

$$u_i^n - \frac{\nu_{i+1/2}}{\rho_i^n} < u_{i+1/2}^* < u_{i+1}^n + \frac{\nu_{i+1/2}}{\rho_{i+1}^n}, \tag{4.164}$$

$$e_i^n + \frac{\left(\pi_{i+1/2}^{L,*}\right)^2 - (p_i^n)^2}{2\nu_{i+1/2}^2} > 0, \quad e_{i+1}^n + \frac{\left(\pi_{i+1/2}^{R,*}\right)^2 - (p_{i+1}^n)^2}{2\nu_{i+1/2}^2} > 0. \tag{4.165}$$

Si la condition CFL

$$\frac{\Delta t}{\Delta x} \max_{i\in\mathbb{Z}} \left| u_i^n \pm \frac{\nu_{i\mp 1/2}}{\rho_i^n} \right| \leq \frac{1}{2}$$

est vérifiée, alors le schéma (4.157) est robuste.

Remarquons que les inégalités (4.164) sont les mêmes que pour les systèmes de Saint-Venant et Ripa. Les inégalités (4.165) sont également équivalentes à assurer la positivité d'un polynôme du second degré en ν tendant vers $+\infty$ quand ν tend vers $+\infty$.

Démonstration. La preuve de la positivité de ρ_L^* et ρ_R^* est identique à celle pour Saint-Venant. Pour la positivité de l'énergie interne, les équations (4.153) et (4.154) et l'hypothèse impliquent que $e_L^* > 0$ et $e_R^* > 0$. On en déduit que W_L^* et W_R^* sont dans $\mathcal{O}$. Le Lemme 4.22 permet alors de conclure. □

4.5 Extension à l'ordre deux pour les équations d'Euler avec gravité

Le but de cette partie est de construire des schémas well-balanced d'ordre deux de type MUSCL. Pour les équations de Saint-Venant, de nombreux schémas well-balanced d'ordre élevé ont été développés (voir par exemple [5, 20, 85, 15]). En ce qui concerne le système de Ripa, plusieurs difficultés rendent extrêmement délicate l'extension à l'ordre deux. Celles-ci seront brièvement présentées dans la Partie 4.5.3. On se concentre ici sur les équations d'Euler avec gravité (4.11). Pour simplifier, on considère que le système est fermé par la loi des gaz parfaits

$$p = (\gamma - 1)\left(E - \frac{1}{2}\rho u^2\right), \tag{4.166}$$

où $\gamma \in]1, 3]$ est le coefficient adiabatique du gaz.

4.5.1 Solutions discrètes stationnaires affines par morceaux

Dans un premier temps, il est essentiel de souligner qu'il n'est pas envisageable de construire des schémas d'ordre deux well-balanced au sens de la Définition 4.9. En effet, si l'on impose au schéma de préserver exactement les solutions discrètes constantes par morceaux, cela reviendra à forcer la préservation d'une approximation d'ordre un entraînant ainsi la perte de l'ordre deux. On doit donc utiliser une définition différente de schéma well-balanced pouvant prendre en compte des approximations affines par morceaux. Pour cela, on va introduire la notion de solution discrète stationnaire affine par morceaux qui est une extension à l'ordre deux des solutions discrètes stationnaires constantes par morceaux définies dans la Partie 4.2.3.

On considère une reconstruction affine par morceaux donnée par

$$\widetilde{w}_i^n(x) = w_i^n + \sigma_i^n(x - x_i), \quad \text{si } x \in K_i = [x_{i-1/2}, x_{i+1/2}[,$$

où $\sigma_i^n = \left(\sigma_i^\rho, \sigma_i^{\rho u}, \sigma_i^E\right)^T \in \mathbb{R}^3$ désigne le vecteur des pentes d'une reconstruction sur la cellule K_i.

On enrichit la Définition 4.8 par des conditions d'ordre deux pour les équations d'Euler avec gravité.

Définition 4.33 (Solution discrète stationnaire affine par morceaux)**.** On dit que l'approximation $(w_i^n, \sigma_i^n)_{i\in\mathbb{Z}}$ définit une solution discrète stationnaire affine par morceaux si pour tout $i \in \mathbb{Z}$, on a

$$u_i^n = 0, \tag{4.167}$$

$$p_{i+1}^n - p_i^n + g\frac{\rho_i^n + \rho_{i+1}^n}{2}(Z_{i+1} - Z_i) = 0, \tag{4.168}$$

$$\sigma_i^{\rho u} = 0, \tag{4.169}$$

$$\sigma_i^p + g\rho_i^n = 0, \tag{4.170}$$

où $\sigma_i^p = (\gamma - 1)\sigma_i^E$.

On souligne que la Définition 4.33 est consistante avec l'équation (4.13) qui décrit les états d'équilibre au repos pour les équations d'Euler avec gravité. En effet, les équations (4.167) et (4.168) étaient déjà présentes dans la définition 4.8 d'une solution discrète stationnaire constante par morceaux. Leur consistance a déjà été prouvée. L'équation (4.169) est naturelle puisque l'on ne considère que les états d'équilibre au repos. Enfin, si l'on suppose que $\rho_i^n = \rho(x) + O(\Delta x)$ et que $\sigma_i^p = \partial_x p(x) + O(\Delta x)$, on déduit immédiatement de (4.170) la relation suivante :

$$\partial_x p(x) + g\rho(x)\partial_x Z(x) = O(\Delta x),$$

puisque $Z(x) = x$. Par conséquent, l'équation (4.170) est bien consistante avec (4.14).

On remarque qu'en général, la reconstruction en pression n'est pas affine. En effet, celle-ci s'écrit

$$\widetilde{p}_i^n(x) = (\gamma - 1)\widetilde{E}_i^n(x) - \frac{\left(\widetilde{\rho u}_i^n(x)\right)^2}{2\widetilde{\rho}_i^n(x)},$$

qui n'a aucune raison, en général, d'être affine en x. Cependant, dans le cadre de la définition (4.33), on a $u_i^n = 0$ et $\sigma_i^{\rho u} = 0$, donc $\widetilde{\rho u}_i^n(x) = 0$. On obtient ainsi

$$\begin{aligned}\widetilde{p}_i^n(x) &= (\gamma - 1)\widetilde{E}_i^n(x) \\ &= (\gamma - 1)E_i^n + (\gamma - 1)\sigma_i^E(x - x_i) \\ &= p_i^n + \sigma_i^p(x - x_i).\end{aligned}$$

Par conséquent, dans ce cas, la reconstruction en pression est affine de pente σ_i^p.

On définit maintenant les états reconstruits à chaque interface par

$$w_i^{n,\pm} = \widetilde{w}_i^n(x_{i\pm1/2}) = w_i^n \pm \frac{\Delta x}{2}\sigma_i^n. \tag{4.171}$$

On introduit également des valeurs intermédiaires de la fonction $Z(x) = x$:

$$Z_{i+1/2} = x_{i+1/2}.$$

On peut alors établir un lien simple entre la Définition 4.33 et la notion d'équilibre local introduite dans la définition 4.3.

Proposition 4.34. *L'approximation $(w_i^n, \sigma_i^n)_{i\in\mathbb{Z}}$ est une solution discrète stationnaire affine par morceaux si et seulement si pour tout $i \in \mathbb{Z}$, les états $(w_i^{n,-}, Z_{i-1/2})$ et $(w_i^{n,+}, Z_{i+1/2})$ d'une part et les états $(w_i^{n,+}, Z_{i+1/2})$ et $(w_{i+1}^{n,-}, Z_{i+1/2})$ d'autre part sont à l'équilibre local défini par (4.14).*

Démonstration. Supposons dans un premier temps que l'approximation $(w_i^n, \sigma_i^n)_{i\in\mathbb{Z}}$ est une solution discrète stationnaire affine par morceaux. On montre d'abord que $(w_i^{n,-}, Z_{i-1/2})$ et $(w_i^{n,+}, Z_{i+1/2})$ sont à l'équilibre local (4.14). Par définition, on a $u_i^n = 0$ et $\sigma_i^{\rho u} = 0$, pour tout $i \in \mathbb{Z}$. On en déduit immédiatement

$$u_i^{n,-} = u_i^{n,+} = 0, \quad \forall i \in \mathbb{Z}.$$

D'autre part, puisque $Z_{i+1/2} - Z_{i-1/2} = \Delta x$, d'après l'équation (4.170), on a

$$\begin{aligned} p_i^{n,+} - p_i^{n,-} + g\frac{\rho_i^{n,-} + \rho_i^{n,+}}{2}(Z_{i+1/2} - Z_{i-1/2}) &= \Delta x \sigma_i^p + g\rho_i^n \Delta x \\ &= 0. \end{aligned}$$

En conséquence, $(w_i^{n,-}, Z_{i-1/2})$ et $(w_i^{n,+}, Z_{i+1/2})$ satisfont l'équilibre local (4.14). De même, on montre maintenant que $(w_i^{n,+}, Z_{i+1/2})$ et $(w_{i+1}^{n,-}, Z_{i+1/2})$ vérifient l'équilibre local (4.14). Par définition des états reconstruits, on a

$$p_{i+1}^{n,-} - p_i^{n,+} + g\frac{\rho_i^{n,+} + \rho_{i+1}^{n,-}}{2}(Z_{i+1/2} - Z_{i+1/2}) = p_{i+1}^n - p_i^n - \frac{\Delta x}{2}(\sigma_{i+1}^p + \sigma_i^p).$$

Puisque $\sigma_i^p = -g\rho_i^n$ par (4.170), on en déduit

$$p_{i+1}^{n,-} - p_i^{n,+} + g\frac{\rho_i^{n,+} + \rho_{i+1}^{n,-}}{2}(Z_{i+1/2} - Z_{i+1/2}) = p_{i+1}^n - p_i^n + g\frac{\rho_i^n + \rho_{i+1}^n}{2}\Delta x.$$

En utilisant à nouveau $Z_{i+1} - Z_i = \Delta x$ et l'équation (4.168), on trouve

$$p_{i+1}^{n,-} - p_i^{n,+} + g\frac{\rho_i^{n,+} + \rho_{i+1}^{n,-}}{2}(Z_{i+1/2} - Z_{i+1/2}) = 0.$$

Il en résulte que les états $(w_i^{n,+}, Z_{i+1/2})$ et $(w_{i+1}^{n,-}, Z_{i+1/2})$ vérifient l'équilibre local (4.14).

Inversement, supposons que les états $(w_i^{n,-}, Z_{i-1/2})$ et $(w_i^{n,+}, Z_{i+1/2})$ d'une part et les états $(w_i^{n,+}, Z_{i+1/2})$ et $(w_{i+1}^{n,-}, Z_{i+1/2})$ d'autre part vérifient l'équilibre local (4.14). On a $u_i^{n,-} = u_i^{n,+} = 0$, donc $u_i^n = \sigma_i^{\rho u} = 0$. Ensuite, puisque $(w_i^{n,-}, Z_{i-1/2})$ et $(w_i^{n,+}, Z_{i+1/2})$ sont à l'équilibre local, on a

$$p_i^{n,+} - p_i^{n,-} + g\frac{\rho_i^{n,-} + \rho_i^{n,+}}{2}(Z_{i+1/2} - Z_{i-1/2}) = 0.$$

On en déduit

$$\Delta x \sigma_i^p + g\rho_i^n(Z_{i+1/2} - Z_{i-1/2}) = 0.$$

Puisque $Z_{i+1/2} - Z_{i-1/2} = \Delta x$, on obtient

$$\sigma_i^p + g\rho_i^n = 0. \tag{4.172}$$

Enfin, puisque $(w_i^{n,+}, Z_{i+1/2})$ et $(w_{i+1}^{n,-}, Z_{i+1/2})$ sont à l'équilibre local, on a

$$p_{i+1}^{n,-} - p_i^{n,+} + g\frac{\rho_i^{n,+} + \rho_{i+1}^{n,-}}{2}(Z_{i+1/2} - Z_{i+1/2}) = 0.$$

On en déduit

$$p_{i+1}^n - p_i^n - \frac{\Delta x}{2}(\sigma_i^p + \sigma_{i+1}^p) = 0.$$

En utilisant (4.172), on trouve finalement

$$p_{i+1}^n - p_i^n - \frac{\rho_i^n + \rho_{i+1}^n}{2}(Z_{i+1} - Z_i) = 0,$$

ce qui prouve que la discrétisation $(w_i^n, \sigma_i^n)_{i\in\mathbb{Z}}$ est une solution discrète stationnaire affine par morceaux.

□

Pour conclure cette partie, on définit ce qu'est un schéma d'ordre deux well-balanced.

Définition 4.35. Un schéma d'ordre deux est dit well-balanced si pour toute solution discrète stationnaire affine par morceaux $(w_i^n, \sigma_i^n)_{i\in\mathbb{Z}}$, le schéma vérifie $w_i^{n+1} = w_i^n$, pour tout $i \in \mathbb{Z}$.

4.5.2 Schéma MUSCL well-balanced

On présente maintenant le schéma MUSCL associé au schéma de relaxation (4.157). Rappelons qu'en l'absence de terme source, le schéma MUSCL peut s'écrire comme la moyenne de deux schémas d'ordre un définis sur des demi-cellules (voir la preuve du Lemme 1.9). Pour prendre en compte le terme source, la méthode usuelle consiste à *définir* le schéma MUSCL comme la moyenne de deux schémas d'ordre un avec terme source définis sur des demi-cellules.

Plus précisément, on considère l'approximation au temps t^n donnée par

$$\left(W_{\Delta x}^n, Z_{\Delta x}\right)(x) = \begin{cases} (w_i^{n,-}, Z_{i-1/2}), & \text{si } x \in [x_{i-1/2}, x_i[, \\ (w_i^{n,+}, Z_{i+1/2}), & \text{si } x \in [x_i, x_{i+1/2}[. \end{cases}$$

On fait évoluer cette approximation par le schéma d'ordre un (4.157), ce qui nous donne les deux états intermédiaires suivants :

$$w_i^{n+1,-} = w_i^{n,-} - \frac{\Delta t}{\Delta x/2}\left(F\left(w_i^{n,-}, Z_{i-1/2}, w_i^{n,+}, Z_{i+1/2}\right) - F\left(w_{i-1}^{n,+}, Z_{i-1/2}, w_i^{n,-}, Z_{i-1/2}\right)\right) + \frac{\Delta t}{2} S\left(w_i^{n,-}, w_i^{n,+}\right)\frac{Z_{i+1/2} - Z_{i-1/2}}{\Delta x/2}, \quad (4.173)$$

$$w_i^{n+1,+} = w_i^{n,+} - \frac{\Delta t}{\Delta x/2}\left(F\left(w_i^{n,+}, Z_{i+1/2}, w_{i+1}^{n,-} Z_{i+1/2}\right) - F\left(w_i^{n,-}, Z_{i-1/2}, w_i^{n,+}, Z_{i+1/2}\right)\right) + \frac{\Delta t}{2} S\left(w_i^{n,-}, w_i^{n,+}\right)\frac{Z_{i+1/2} - Z_{i-1/2}}{\Delta x/2}. \quad (4.174)$$

La mise à jour du schéma MUSCL est alors donnée par la moyenne suivante :

$$w_i^{n+1} = \frac{1}{2}\left(w_i^{n+1,-} + w_i^{n+1,+}\right), \quad (4.175)$$

qui se réécrit

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(F\left(w_i^{n,+}, Z_{i+1/2}, w_{i+1}^{n,-} Z_{i+1/2}\right) - F\left(w_{i-1}^{n,+}, Z_{i-1/2}, w_i^{n,-}, Z_{i-1/2}\right)\right) + \Delta t S\left(w_i^{n,-}, w_i^{n,+}\right)\frac{Z_{i+1/2} - Z_{i-1/2}}{\Delta x}. \quad (4.176)$$

On montre maintenant que le schéma MUSCL est well-balanced.

Proposition 4.36. *Le schéma MUSCL (4.176) est well-balanced, c'est-à-dire que pour toute solution discrète stationnaire affine par morceaux $(w_i^n, \sigma_i^n)_{i\in\mathbb{Z}}$, le schéma vérifie*

$$w_i^{n+1} = w_i^n, \quad \forall i \in \mathbb{Z}.$$

Démonstration. Soit $(w_i^n, \sigma_i^n)_{i\in\mathbb{Z}}$ une solution discrète stationnaire affine par morceaux. D'après la Proposition 4.34, cela signifie que les états $(w_i^{n,-}, Z_{i-1/2})$ et $(w_i^{n,+}, Z_{i+1/2})$ d'une part et les états $(w_i^{n,+}, Z_{i+1/2})$ et $(w_{i+1}^{n,-}, Z_{i+1/2})$ d'autre part sont à l'équilibre local (4.14). Puisque le

schéma de relaxation (4.157) est well-balanced au sens de la Définition 4.9 (voir Proposition 4.31), les états intermédiaires définis par les schémas d'ordre un (4.173) et (4.174) vérifient

$$w_i^{n+1,-} = w_i^{n,-} \quad \text{et} \quad w_i^{n+1,+} = w_i^{n,+}.$$

On déduit alors des équations (4.175) et (4.171)

$$\begin{aligned} w_i^{n+1} &= \frac{1}{2}\left(w_i^{n,-} + w_i^{n,+}\right) \\ &= w_i^n \end{aligned}$$

et le schéma (4.176) est well-balanced. □

Avant de montrer que le schéma est robuste, on introduit ν_i la linéarisation de l'impédance acoustique utilisée dans le problème de Riemann $W_{\mathcal{R}}(\xi, W_i^{n,-}, Z_{i-1/2}, W_i^{n,+}, Z_{i+1/2})$ et $\nu_{i+1/2}$ celle utilisée dans le problème de Riemann $W_{\mathcal{R}}(\xi, W_i^{n,+}, Z_{i+1/2}, W_{i+1}^{n,-}, Z_{i+1/2})$.

Proposition 4.37. *Supposons que les paramètres ν_i et $\nu_{i+1/2}$ vérifient les inégalités (4.164) et (4.165) pour tout $i \in \mathbb{Z}$. Si la condition CFL*

$$\frac{\Delta t}{\Delta x} \max_{i\in\mathbb{Z}} \left(\left| u_i^{n,\pm} \pm \frac{\nu_i}{\rho_i^{n,\pm}} \right|, \left| u_i^{n,\pm} \mp \frac{\nu_{i\pm1/2}}{\rho_i^{n,\pm}} \right| \right) \leq \frac{1}{4}$$

est vérifiée, alors le schéma (4.176) est robuste.

Démonstration. Les états intermédiaires $w_i^{n+1,-}$ et $w_i^{n+1,+}$ sont obtenus par des schémas de relaxation d'ordre un (4.173) et (4.174). D'après la Proposition 4.32, on en déduit que les états $w_i^{n+1,\pm}$ sont dans Ω. L'état w_i^{n+1} étant défini comme la moyenne de $w_i^{n+1,-}$ et $w_i^{n+1,+}$ et l'ensemble Ω étant convexe, on en déduit que w_i^{n+1} est dans Ω. □

4.5.3 Difficultés pour le modèle de Ripa

On explique ici brièvement pourquoi il s'avère beaucoup plus complexe de construire un schéma d'ordre deux well-balanced pour les équations de Ripa que pour les équations d'Euler avec gravité. La simplicité de l'approche suivie pour les équations d'Euler repose sur la Proposition 4.34 qui fait le lien entre les solutions discrètes stationnaires affines par morceaux et la notion d'équilibre local. En définissant le schéma MUSCL comme une combinaison convexe de schémas d'ordre un, la propriété well-balanced s'étend alors naturellement du schéma d'ordre un au schéma d'ordre deux. Concernant le modèle de Ripa, il est beaucoup plus dur de définir une notion de solution discrète stationnaire affine par morceaux qui puisse être connectée simplement à l'équilibre local. Il y a deux raisons principales à cette difficulté.

La première raison concerne les grandeurs physiques intervenant dans la définition de l'équilibre local. Pour les équations d'Euler, les grandeurs intervenant dans (4.14) sont toutes conservatives. En effet, comme mentionné précédemment, puisque la vitesse est identiquement nulle pour les solutions stationnaires, la pression est alors proportionnelle à l'énergie totale et donc conservative. Cela implique que la reconstruction pour toutes ces grandeurs est affine. Dans la définition (4.10) de l'équilibre local pour les équations de Ripa, les grandeurs θ et $h^2\theta$ intervenant sont des fonctions non linéaires des grandeurs conservatives. Par conséquent, la reconstruction de ces grandeurs n'a aucune raison d'être affine, ce qui rend beaucoup plus compliquée l'obtention de relations analogues à (4.167)–(4.170).

D'autre part, dans les équations d'Euler, la fonction Z est extrêmement simple ($Z(x) = x$). La reconstruction de cette fonction est donc triviale et l'on a utilisé plusieurs fois le fait que $Z_{i+1} - Z_i = \Delta x$. Dans les équations de Ripa, la fonction Z peut être n'importe quelle fonction régulière. Cela rend le choix des valeurs reconstruites de Z plus délicat et les calculs beaucoup plus lourds.

4.6 Extension en deux dimensions d'espace

Dans cette partie, on utilise les schémas de relaxation construits dans la Partie 4.4 pour construire des schémas numériques well-balanced en deux dimensions d'espace. On se concentre ici sur les équations de Ripa 2D. Tout ce qu'on va présenter s'applique cependant très simplement au système de Saint-Venant et aux équations d'Euler avec gravité. Il serait même possible de présenter la méthode pour les trois systèmes simultanément, au prix de notations légèrement plus lourdes (voir [56, 14]).

Dans un premier temps, on présente brièvement le système de Ripa 1D avec une vitesse tangentielle. On constate que cette inconnue supplémentaire ne modifie pas la structure du système et qu'on peut construire sans difficulté un schéma de relaxation similaire à celui présenté dans la Partie 4.4.3. On introduit ensuite le modèle de Ripa en deux dimensions d'espace. On décrit en particulier les états d'équilibre au repos pour ce système. On définit ensuite le caractère well-balanced pour le système 2D avant de dériver un schéma numérique par des techniques classiques (voir [14, 15]). On prouve finalement que ce schéma est bien well-balanced et robuste.

4.6.1 Le modèle de Ripa 1D avec vitesse tangentielle

On considère le modèle de Ripa 1D, dans lequel on rajoute une inconnue : la vitesse tangentielle v. Celle-ci est simplement transportée à la vitesse normale u, sans terme source de topographie. Ce modèle ne présente pas vraiment d'intérêt en lui-même, mais il sera très utile dans la présentation en deux dimensions qui va suivre.

Le système s'écrit de la façon suivante :

$$\begin{cases} \partial_t h + \partial_x hu = 0, \\ \partial_t hu + \partial_x (hu^2 + gh^2\theta/2) = -gh\theta\partial_x Z, \\ \partial_t hv + \partial_x huv = 0, \\ \partial_t h\theta + \partial_x hu\theta = 0. \end{cases} \tag{4.177}$$

On peut écrire ce système sous la forme condensée

$$\partial_t w + \partial_x f(w) = s(w)\partial_x Z, \tag{4.178}$$

où le vecteur d'état est défini par

$$w = (h, hu, hv, h\theta)^T,$$

la fonction flux est définie par

$$f(w) = \left(hu, hu^2 + gh^2\theta/2, huv, hu\theta\right)^T,$$

et le terme source est donné par

$$s(w) = (0, -gh\theta, 0, 0)^T.$$

L'ensemble des états admissibles est

$$\Omega = \{w \in \mathbb{R}^4, h > 0, \theta > 0\}.$$

Le système (4.177) est très proche du système de Ripa 1D (4.6). En effet, la vitesse tangentielle v n'est discontinue qu'à travers l'onde linéairement dégénérée de vitesse u. D'autre part, la variable v n'intervient pas dans l'évolution des autres grandeurs. Par conséquent, il s'agit d'une variable « passive » qui ne modifie pas la structure du système.

La définition de l'équilibre local pour ce système est très similaire à celle du système 1D.

Définition 4.38 (Équilibre local pour Ripa avec vitesse tangentielle)**.** Deux états (w_L, Z_L) et (w_R, Z_R) sont dits à l'équilibre local pour le système de Ripa avec vitesse tangentielle (4.177) si

$$\begin{cases} u_L = u_R = 0, \\ v_L = v_R = 0, \\ \left[h^2\theta/2\right] + \bar{h}\bar{\theta}[Z] = 0. \end{cases} \tag{4.179}$$

On s'intéresse maintenant à un schéma de relaxation pour approcher les solutions de (4.177). Il serait possible de dériver rigoureusement ce schéma en suivant la procédure présentée dans la Partie 4.4.3. Cette procédure étant quasiment identique, on se contente ici de donner directement le schéma. Celui-ci s'écrit

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(f_{i+1/2} - f_{i-1/2}\right) + \frac{\Delta t}{2}\left(s(w_{i-1}^n, w_i^n)\frac{Z_i - Z_{i-1}}{\Delta x} + s(w_i^n, w_{i+1}^n)\frac{Z_{i+1} - Z_i}{\Delta x}\right), \tag{4.180}$$

où le flux numérique est défini par

$$f_{i+1/2} = f(w_i^n, Z_i, w_{i+1}^n, Z_{i+1}), \tag{4.181}$$

$$f(w_L, Z_L, w_R, Z_R) = \begin{cases} \left(h_L u_L, h_L u_L^2 + \pi_L - \dfrac{g}{2}\bar{h}\bar{\theta}[Z], h_L u_L v_L, h_L u_L \theta_L\right)^T & \text{si } \lambda_L > 0, \\ \left(h_L^* u^*, h_L^* (u^*)^2 + \pi_L^* - \dfrac{g}{2}\bar{h}\bar{\theta}[Z], h_L^* u^* v_L h_L^* u^* \theta_L\right)^T & \text{si } \lambda_L < 0 < u^*, \\ \left(h_R^* u^*, h_R^* (u^*)^2 + \pi_R^* + \dfrac{g}{2}\bar{h}\bar{\theta}[Z], h_R^* u^* v_R, h_R^* u^* \theta_R\right)^T & \text{si } u^* < 0 < \lambda_R, \\ \left(h_R u_R, h_R u_R^2 + \pi_R + \dfrac{g}{2}\bar{h}\bar{\theta}[Z], h_R u R v_R, h_R u_R \theta_R\right)^T & \text{si } \lambda_R < 0, \end{cases} \tag{4.182}$$

où $\lambda_L = u_L - \nu/h_L$ et $\lambda_R = u_R + \nu/h_R$, et le terme source numérique est défini par

$$s(w_L, w_R) = \left(0, -g\bar{h}\bar{\theta}, 0, 0\right)^T. \tag{4.183}$$

Les variables $h_{L,R}^*$, u^* et $\pi_{L,R}^*$ sont définies par les relations (4.124)–(4.128). On peut aisément vérifier que ce schéma est well-balanced et robuste sous les mêmes conditions que le schéma de relaxation (4.132) pour le modèle sans vitesse tangentielle.

4.6.2 Le modèle de Ripa en deux dimensions d'espace

On s'intéresse maintenant au modèle de Ripa 2D qui s'écrit

$$\begin{cases} \partial_t h + \partial_x hu + \partial_y hv = 0, \\ \partial_t hu + \partial_x(hu^2 + gh^2\theta/2) + \partial_y huv = -gh\theta\partial_x Z, \\ \partial_t hv + \partial_x huv + \partial_y(hv^2 + gh^2\theta/2) = -gh\theta\partial_y Z, \\ \partial_t h\theta + \partial_x hu\theta + \partial_y hv\theta = 0, \end{cases} \tag{4.184}$$

où h est la hauteur d'eau, $(u, v) \in \mathbb{R}^2$ est le vecteur vitesse, θ est la température, $Z : \mathbb{R}^2 \to \mathbb{R}$ est la topographie du fond qui est une fonction régulière donnée et g est la constante de gravité.

On peut écrire ce système sous la forme condensée

$$\partial_t w + \partial_x f(w) + \partial_y g(w) = s_x(w)\partial_x Z + s_y(w)\partial_y Z, \tag{4.185}$$

où le vecteur d'état est défini par

$$w = (h, hu, hv, h\theta)^T,$$

les fonctions flux sont définies par

$$f(w) = \left(hu, hu^2 + gh^2\theta/2, huv, hu\theta\right)^T \quad \text{et} \quad g(w) = \left(hv, huv, hv^2 + gh^2\theta/2, hv\theta\right)^T,$$

et les termes source sont donnés par

$$s_x(w) = (0, -gh\theta, 0, 0)^T \quad \text{et} \quad s_y(w) = (0, 0, -gh\theta, 0)^T.$$

Notons que la composante $s_x(w)$ est égale au terme source $s(w)$ du modèle de Ripa 1D avec vitesse tangentielle (4.177). L'ensemble des états admissibles est

$$\Omega = \{w \in \mathbb{R}^4, h > 0, \theta > 0\}.$$

Les états d'équilibre au repos pour le système (4.184) sont décrits par

$$\begin{cases} (u, v) \equiv (0, 0), \\ \nabla h^2\theta/2 = -h\theta\nabla Z. \end{cases}$$

Cette équation n'est pas intégrable et l'on ne peut pas obtenir d'expression analytique de tous les états d'équilibre pour le système (4.184). Il existe cependant deux états d'équilibre remarquables vérifiant une équation algébrique. Le premier est obtenu en imposant la température θ constante. Cela nous donne la solution dite du lac au repos :

$$\begin{cases} (u, v) \equiv (0, 0), \\ \theta = \text{ constante}, \\ h + Z = \text{ constante}. \end{cases}$$

Le deuxième état d'équilibre remarquable est obtenu en imposant la topographie Z constante. On trouve la solution suivante :

$$\begin{cases} (u, v) \equiv 0, \\ Z = \text{ constante}, \\ h^2\theta = \text{ constante}. \end{cases}$$

L'équation décrivant les états d'équilibre au repos est la même que pour le système de Ripa 1D avec vitesse tangentielle. Par conséquent, on utilise également la Définition 4.38 pour l'équilibre local.

4.6.3 Schéma numérique pour le système de Ripa 2D

On commence par introduire quelques notations relatives aux maillages en deux dimensions d'espace. On considère un maillage de $\mathbb{R}^2$ constitué de cellules polygonales $(K_i)_{i\in\mathbb{Z}}$. Pour chaque $i \in \mathbb{Z}$, on note $\gamma(i)$ l'ensemble des indices des cellules voisines de K_i. Pour $j \in \gamma(i)$, on appelle e_{ij} le côté commun à K_i et K_j et on note ν_{ij} la normale extérieure à e_{ij} (voir Figure 4.12 à gauche). On définit respectivement $|K_i|$ et $|e_{ij}|$, l'aire de la cellule K_i et la longueur du côté e_{ij}. Il sera également utile dans la suite de définir le triangle T_{ij}, formé par le côté e_{ij} et le centre de masse de la cellule K_i (voir Figure 4.12 à droite). L'aire de ce triangle est notée $|T_{ij}|$.

On introduit l'approximation suivante de la fonction Z :

$$Z_i = \frac{1}{|K_i|}\int_{K_i} Z(x)dx, \quad \forall i \in \mathbb{Z}.$$

On note w_i^n une approximation constante de la solution de (4.184) au temps t^n sur la cellule K_i. On définit maintenant les solutions discrètes stationnaires constantes par morceaux en deux dimensions d'espace.

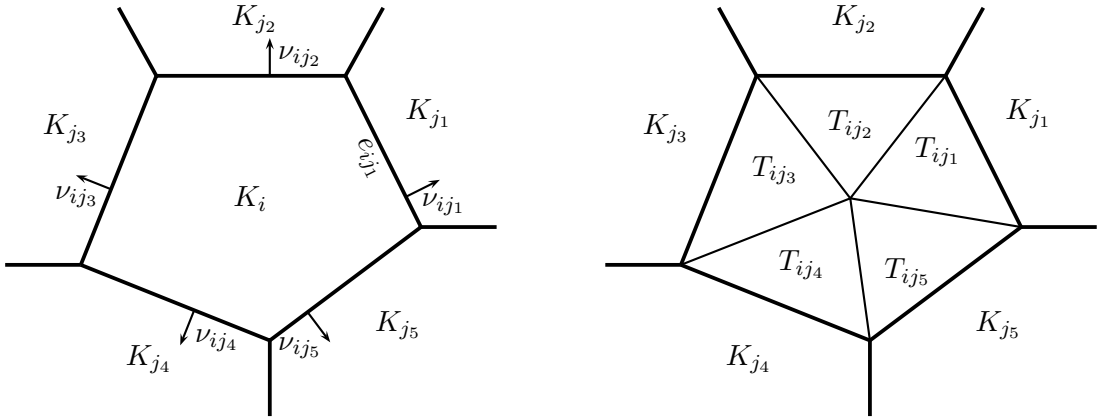


FIGURE 4.12 – Géométrie de la cellule K_i

Définition 4.39 (Solution discrète stationnaire constante par morceaux en 2D)**.** Si pour tout $i \in \mathbb{Z}$ et pour tout $j \in \gamma(i)$, les états (w_i^n, Z_i) et (w_j^n, Z_j) sont à l'équilibre local (4.179), alors on dit que l'approximation $(w_i^n)_{i\in\mathbb{Z}}$ définit une solution discrète stationnaire constante par morceaux.

La définition d'un schéma well-balanced est alors naturelle.

Définition 4.40. Un schéma est dit well-balanced si pour toute solution discrète stationnaire constante par morceaux $(w_i^n)_{i\in\mathbb{Z}}$, le schéma vérifie $w_i^{n+1} = w_i^n$, pour tout $i \in \mathbb{Z}$.

Le but est maintenant de construire un schéma numérique well-balanced pour approcher les solutions faibles de (4.184). On commence par introduire quelques notations. Pour un vecteur unitaire $\nu = (\nu_x, \nu_y)^T \in \mathbb{S}^1$, on définit le flux dans la direction ν par

$$h_\nu(w) = \nu_x f(w) + \nu_y g(w). \tag{4.186}$$

Il est également pratique de faire intervenir la rotation de $\mathbb{R}^4$ définie pour $\nu = (\nu_x, \nu_y)^T \in \mathbb{S}^1$ par la matrice

$$R_\nu = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \nu_x & -\nu_y & 0 \\ 0 & \nu_y & \nu_x & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

On vérifie alors aisément la propriété suivante qui traduit l'invariance par rotation du système :

$$h_\nu(w) = R_\nu f\left(R_\nu^{-1} w\right). \tag{4.187}$$

Dans un premier temps, on suppose que la topographie est plate, $Z =$ constante, autrement dit, il n'y a pas de terme source. La technique usuelle pour dériver un schéma numérique 2D consiste à utiliser le schéma numérique pour le système 1D combiné avec l'invariance par rotation du système. Plus précisément, le schéma numérique pour le système de Ripa avec une topographie plate s'écrit

$$\left(w^{\text{flat}}\right)_i^{n+1} = w_i^n - \frac{\Delta t}{|K_i|} \sum_{j\in\gamma(i)} |e_{ij}| \varphi\left(w_i^n, w_j^n, \nu_{ij}\right), \tag{4.188}$$

où le flux numérique φ est défini par

$$\varphi(w_L, w_R, \nu) = R_\nu F\left(R_{\nu^{-1}} w_L, R_{\nu^{-1}} w_R\right). \tag{4.189}$$

Ici, F est un flux numérique pour le système de Ripa 1D avec vitesse tangentielle et une topographie plate. On rappelle le résultat suivant.

Lemme 4.41. *Le schéma (4.188) s'écrit comme une combinaison convexe de schémas 1D écrits dans les directions ν_{ij}, $j \in \gamma(i)$:*

$$\left(w^{\text{flat}}\right)_i^{n+1} = \sum_{j\in\gamma(i)} \frac{|T_{ij}|}{|K_i|} \widetilde{w}_{ij}^{n+1}, \tag{4.190}$$

où l'on a posé

$$\widetilde{w}_{ij}^{n+1} = w_i^n - \frac{\Delta t}{|T_{ij}|}|e_{ij}| \left(\varphi\left(w_i^n, w_j^n, \nu_{ij}\right) - \varphi\left(w_i^n, w_i^n, \nu_{ij}\right)\right). \tag{4.191}$$

Démonstration. En combinant (4.190) et (4.191), on obtient

$$\left(w^{\text{flat}}\right)_i^{n+1} = \sum_{j\in\gamma(i)} \frac{|T_{ij}|}{|K_i|} w_i^n - \frac{\Delta t}{|K_i|} \sum_{j\in\gamma(i)} |e_{ij}| \varphi\left(w_i^n, w_j^n, \nu_{ij}\right) + \frac{\Delta t}{|K_i|} \sum_{j\in\gamma(i)} |e_{ij}| \varphi\left(w_i^n, w_i^n, \nu_{ij}\right), \tag{4.192}$$

La définition (4.189) du flux numérique 2D φ, la consistance du flux numérique 1D f et l'invariance par rotation du flux (4.187) nous donnent immédiatement

$$\begin{aligned}
\varphi\left(w_i^n, w_i^n, \nu_{ij}\right) &= R_{\nu_{ij}} F\left(R_{\nu_{ij}}^{-1} w_i^n, R_{\nu_{ij}}^{-1} w_i^n\right) \\
&= R_{\nu_{ij}} f\left(R_{\nu_{ij}}^{-1} w_i^n\right) \\
&= h_{\nu_{ij}}(w_i^n).
\end{aligned}$$

En appliquant la formule de Green, on trouve

$$\sum_{j\in\gamma(i)} |e_{ij}| h_{\nu_{ij}}(w_i^n) = 0.$$

Par conséquent, le dernier terme de (4.192) s'annule. On en déduit

$$\left(w^{\text{flat}}\right)_i^{n+1} = w_i^n - \frac{\Delta t}{|K_i|} \sum_{j\in\gamma(i)} |e_{ij}| \varphi\left(w_i^n, w_j^n, \nu_{ij}\right)$$

et on retrouve le schéma (4.188). □

On revient maintenant au cas général, avec une topographie quelconque. On introduit le flux numérique 2D donné par

$$\varphi(w_L, Z_L, w_R, Z_R, \nu) = R_\nu f\left(R_{\nu^{-1}} w_L, Z_L, R_{\nu^{-1}} w_R, Z_R\right), \tag{4.193}$$

où f est le flux numérique de relaxation (4.182) pour le système de Ripa 1D avec vitesse tangentielle. On introduit également le terme source numérique 2D défini par

$$\sigma(w_L, w_R, \nu) = R_\nu s\left(R_{\nu^{-1}} w_L, R_{\nu^{-1}} w_R\right), \tag{4.194}$$

où s est le terme source numérique (4.183) du schéma de relaxation pour le système de Ripa 1D avec vitesse tangentielle.

On choisit de définir le schéma numérique pour (4.184) en s'inspirant de la formulation (4.190)–(4.191). On considère donc le schéma donné par la combinaison convexe

$$w_i^{n+1} = \sum_{j\in\gamma(i)} \frac{|T_{ij}|}{|K_i|} \widetilde{w}_{ij}^{n+1}, \tag{4.195}$$

où les états intermédiaires sont donnés par

$$\widetilde{w}_{ij}^{n+1} = w_i^n - \frac{\Delta t}{|T_{ij}|}|e_{ij}|\left(\varphi\left(w_i^n, Z_i, w_j^n, Z_j, \nu_{ij}\right) - \varphi\left(w_i^n, Z_i, w_i^n, Z_i\nu_{ij}\right)\right) + \frac{\Delta t}{2}\sigma(w_i^n, w_j^n, \nu_{ij})\frac{Z_j - Z_i}{|T_{ij}|/|e_{ij}|}. \quad (4.196)$$

En combinant les deux dernières relations et la formule de Green, on peut écrire le schéma sous la forme

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{|K_i|}\sum_{j\in\gamma(i)}|e_{ij}|\varphi\left(w_i^n, Z_i, w_j^n, Z_j, \nu_{ij}\right) + \frac{\Delta t}{2}\sum_{j\in\gamma(i)}\sigma\left(w_i^n, w_j^n, \nu_{ij}\right)\frac{Z_j - Z_i}{|K_i|/|e_{ij}|}. \quad (4.197)$$

4.6.4 Propriétés du schéma

On commence par montrer que le schéma est well-balanced.

Proposition 4.42. *Le schéma numérique (4.197) est well-balanced, c'est-à-dire que si pour tout $i \in \mathbb{Z}$ et pour tout $j \in \gamma(i)$, on a*

$$(u_i^n, v_i^n) = (0,0) \quad \textit{et} \quad \frac{(h_j^n)^2\theta_j^n}{2} - \frac{(h_i^n)^2\theta_i^n}{2} + \frac{h_i^n + h_j^n}{2}\frac{\theta_i^n + \theta_j^n}{2}(Z_j - Z_i) = 0,$$

alors $w_i^{n+1} = w_i^n$, $\forall i \in \mathbb{Z}$.

Démonstration. D'après la formulation en combinaison convexe (4.195), il suffit de montrer que $\widetilde{w}_{ij}^{n+1} = w_i^n$, pour tout $i \in \mathbb{Z}$ et pour tout $j \in \gamma(i)$.

Pour un $j \in \gamma(i)$ fixé, on pose $\Delta x = |T_{ij}|/|e_{ij}|$ et $\overline{w} = R_{\nu_{ij}}^{-1}w$. En utilisant les définitions (4.193) et (4.194) du flux numérique et du terme source numérique, on peut réécrire la relation (4.196) sous la forme

$$R_{\nu_{ij}}^{-1}\widetilde{w}_{ij}^{n+1} = \overline{w}_i^n - \frac{\Delta t}{\Delta x}\left(f\left(\overline{w}_i^n, Z_i, \overline{w}_j^n, Z_j\right) - f\left(\overline{w}_i^n, Z_i, \overline{w}_i^n, Z_i\right)\right) + \frac{\Delta t}{2}\left(s\left(\overline{w}_i^n, \overline{w}_j^n\right)\frac{Z_j - Z_i}{\Delta x} + s\left(\overline{w}_i^n, \overline{w}_i^n\right)\frac{Z_i - Z_i}{\Delta x}\right).$$

On a ainsi écrit l'état $R_{\nu_{ij}}^{-1}\widetilde{w}_{ij}^{n+1}$ comme une mise à jour du schéma 1D (4.180).

D'autre part, la notion d'équilibre local est invariante par la rotation $R_{\nu_{ij}}$. L'hypothèse assure donc que $(\overline{w}_i^n, Z_i)$ et $(\overline{w}_j^n, Z_j)$ sont à l'équilibre local (4.179). Les états $(\overline{w}_i^n, Z_i)$ et $(\overline{w}_i^n, Z_i)$ vérifient également trivialement l'équilibre local (4.179). Le schéma 1D (4.177) étant well-balanced, on en déduit

$$R_{\nu_{ij}}^{-1}\widetilde{w}_{ij}^{n+1} = \overline{w}_i^n.$$

En composant à gauche par $R_{\nu_{ij}}$, on obtient le résultat attendu.

□

Il reste à prouver la robustesse du schéma.

Proposition 4.43. *On note respectivement ν_{ij} et u_{ij}^* la linéarisation de l'impédance acoustique et la vitesse intermédiaire utilisées dans le problème de Riemann $W_{\mathcal{R}}(\xi, W_i^n, Z_i, W_j^n, Z_j)$. Supposons que la constante ν_{ij} assure que les valeurs propres vérifient l'ordre suivant :*

$$u_i^n - \frac{\nu_{ij}}{h_i^n} < u_{ij}^* < u_j^n + \frac{\nu_{ij}}{h_j^n}. \quad (4.198)$$

Si la condition CFL

$$\Delta t \frac{\mathcal{P}_i}{|K_i|} \max_{j\in\gamma(i)} \left\{ |u_i^n - \nu_{ij}/h_i^n|, |u_j^n + \nu_{ij}/h_j^n| \right\} \leq 1, \quad \forall i \in \mathbb{Z} \tag{4.199}$$

est vérifiée, alors le schéma (4.197) est robuste.

Démonstration. En utilisant les définitions (4.193) du flux numérique et (4.194) du terme source numérique, on peut réécrire le schéma (4.197) sous la forme

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{|K_i|} \sum_{j\in\gamma(i)} |e_{ij}| R_{\nu_{ij}} \left(f\left(R_{\nu_{ij}}^{-1} w_i^n, Z_i, R_{\nu_{ij}}^{-1} w_j^n, Z_j \right) - \frac{s\left(R_{\nu_{ij}}^{-1} w_i^n, R_{\nu_{ij}}^{-1} w_j^n \right)}{2} (Z_j - Z_i) \right). \tag{4.200}$$

On utilise maintenant le Lemme 4.27. Bien que celui-ci n'ait pas été montré pour le système avec vitesse tangentielle, l'extension est immédiate. Notons qu'en fixant

$$\Delta x = 2|K_i|/\mathcal{P}_i, \quad w_L = R_{\nu_{ij}} w_i^n, \quad w_R = R_{\nu_{ij}} w_j^n, \quad Z_L = Z_i \quad \text{et} \quad Z_R = Z_j,$$

les conditions CFL (4.142) et (4.199) coïncident. Par conséquent, on peut utiliser l'identité (4.143) qui s'écrit ici :

$$\begin{aligned} f\left(R_{\nu_{ij}}^{-1} w_i^n, Z_i, R_{\nu_{ij}}^{-1} w_j^n, Z_j \right) - \frac{1}{2} s\left(R_{\nu_{ij}}^{-1} w_i^n, R_{\nu_{ij}}^{-1} w_j^n \right) (Z_j - Z_i) &= f\left(R_{\nu_{ij}}^{-1} w_i^n \right) + \frac{|K_i|}{\mathcal{P}_i} R_{\nu_{ij}}^{-1} w_i^n \\ &\quad - \frac{1}{\Delta t} \int_{-|K_i|/\mathcal{P}_i}^0 \mathcal{Q} W_{\mathcal{R}} \left(\frac{x}{\Delta t}, \mathcal{E}(R_{\nu_{ij}} w_i^n, Z_i), Z_i, \mathcal{E}(R_{\nu_{ij}} w_j^n, Z_j), Z_j \right) dx. \end{aligned}$$

En injectant cette identité dans (4.200), on obtient

$$\begin{aligned} w_i^{n+1} = &\left(1 - \frac{\Delta t}{\mathcal{P}_i} \sum_{j\in\gamma(i)} |e_{ij}| \right) w_i^n - \frac{\Delta t}{|K_i|} \sum_{j\in\gamma(i)} |e_{ij}| R_{\nu_{ij}} f\left(R_{\nu_{ij}}^{-1} w_i^n \right) \\ &+ \frac{1}{|K_i|} \sum_{j\in\gamma(i)} |e_{ij}| \int_{-|K_i|/\mathcal{P}_i}^0 \mathcal{R}_{\nu_{ij}} \mathcal{Q} W_{\mathcal{R}} \left(\frac{x}{\Delta t}, \mathcal{E}(R_{\nu_{ij}} w_i^n, Z_i), Z_i, \mathcal{E}(R_{\nu_{ij}} w_j^n, Z_j), Z_j \right) dx. \end{aligned} \tag{4.201}$$

Le premier terme du second membre est évidemment nul. En utilisant (4.186), on a

$$R_{\nu_{ij}} f\left(R_{\nu_{ij}}^{-1} w_i^n \right) = h_{\nu_{ij}}(w_i^n)$$

et en appliquant la formule de Green, on voit que le deuxième terme du second membre de (4.201) est également nul. On introduit les états intermédiaires

$$w_{ij}^{n+1} = \frac{\mathcal{P}_i}{|K_i|} \int_{-|K_i|/\mathcal{P}_i}^0 \mathcal{R}_{\nu_{ij}} \mathcal{Q} W_{\mathcal{R}} \left(\frac{x}{\Delta t}, \mathcal{E}(R_{\nu_{ij}} w_i^n, Z_i), Z_i, \mathcal{E}(R_{\nu_{ij}} w_j^n, Z_j), Z_j \right) dx.$$

La condition (4.198) sur l'ordre des valeurs propres assure que les états intermédiaires du solveur de Riemann $W_{\mathcal{R}}$ sont dans $\mathcal{O}$. On utilise ensuite le fait que $\mathcal{Q}\mathcal{O} = \Omega$, l'invariance par rotation de Ω et la convexité de Ω pour déduire que w_{ij}^{n+1} est dans Ω. On voit, à partir de l'équation (4.201), que w_i^{n+1} s'écrit

$$w_i^{n+1} = \sum_{j\in\gamma(i)} \frac{|e_{ij}|}{\mathcal{P}_i} w_{ij}^{n+1}.$$

L'état w_i^{n+1} est donc une combinaison convexe d'état appartenant à Ω, il est donc lui-même dans Ω. □

4.7 Résultats numériques

On présente maintenant quelques expériences numériques pour valider les schémas construits dans ce chapitre. On se concentre sur le système de Ripa et sur les équations d'Euler avec gravité.

4.7.1 Modèle de Ripa

Il existe peu d'article dédiés à l'approximation numériques du modèle de Ripa dans la littérature. On réalise ici quelques expériences introduites par Chertock et al. [32]. Les résultats que l'on obtient ne peuvent toutefois pas être comparés avec ceux de [32] puisque le schéma utilisé dans cet article est d'ordre deux, alors que les schémas dérivés dans ce travail ne sont que d'ordre un. Cependant, à la différence du schéma utilisé dans [32] qui ne préserve que les deux équilibres remarquables (4.8) et (4.9), les schémas que l'on a dérivés préservent tous les états d'équilibre au repos. On réalise donc une dernière expérience pour illustrer cette propriété.

Rupture de barrage sur un fond plat

Afin de valider le schéma de relaxation construit dans la Partie 4.4, on considère un problème de rupture de barrage sur un fond plat ($Z \equiv 0$). La donnée initiale est donnée par

$$(h, u, \theta)(x, 0) = \begin{cases} (5, 0, 3), & \text{si } x < 0, \\ (1, 0, 5), & \text{si } x > 0. \end{cases}$$

On présente, sur la Figure 4.13, les résultats obtenus au temps $T = 0.2$ pour la hauteur h, la température θ et la pression $p = h^2\theta/2$. On utilise 200 cellules pour discrétiser le domaine de calcul $[-1, 1]$. Les solutions sont comparées à une solution de référence calculée avec $20\,000$ cellules.

On observe que le comportement général de la solution est bien capturé. L'approximation est cependant assez diffuse, ce qui est normal pour un schéma d'ordre un.

Rutpture de barrage avec topographie

On s'intéresse de nouveau à une rupture de barrage, mais cette fois en présence d'un terme de topographie non constant, donné par (voir le haut de la Figure 4.14)

$$Z(x) = \begin{cases} 2(\cos(10\pi(x + 0.3)) + 1), & \text{si } -0.4 \leq x \leq -0.2, \\ 0.5(\cos(10\pi(x - 0.3)) + 1), & \text{si } 0.2 \leq x \leq 0.4, \\ 0, & \text{sinon.} \end{cases}$$

La donnée initiale pour ce problème s'écrit

$$(h, u, \theta)(x, 0) = \begin{cases} (5 - Z(x), 0, 1), & \text{si } x < 0, \\ (1 - Z(x), 0, 5), & \text{si } x > 0. \end{cases}$$

On calcule une solution approchée sur le domaine $[-1, 1]$ en utilisant 200 cellules. Les résultats en hauteur totale $h + Z$, en température θ et en pression p sont affichés Figure 4.14. On présente également une solution de référence calculée avec $20\,000$ cellules.

Les résultats montent que le schéma se comporte bien en présence d'un terme de topographie. Bien que la solution reste peu précise au voisinage des discontinuités à cause de la viscosité numérique du schéma, l'allure générale est similaire aux résultats obtenus dans [32].

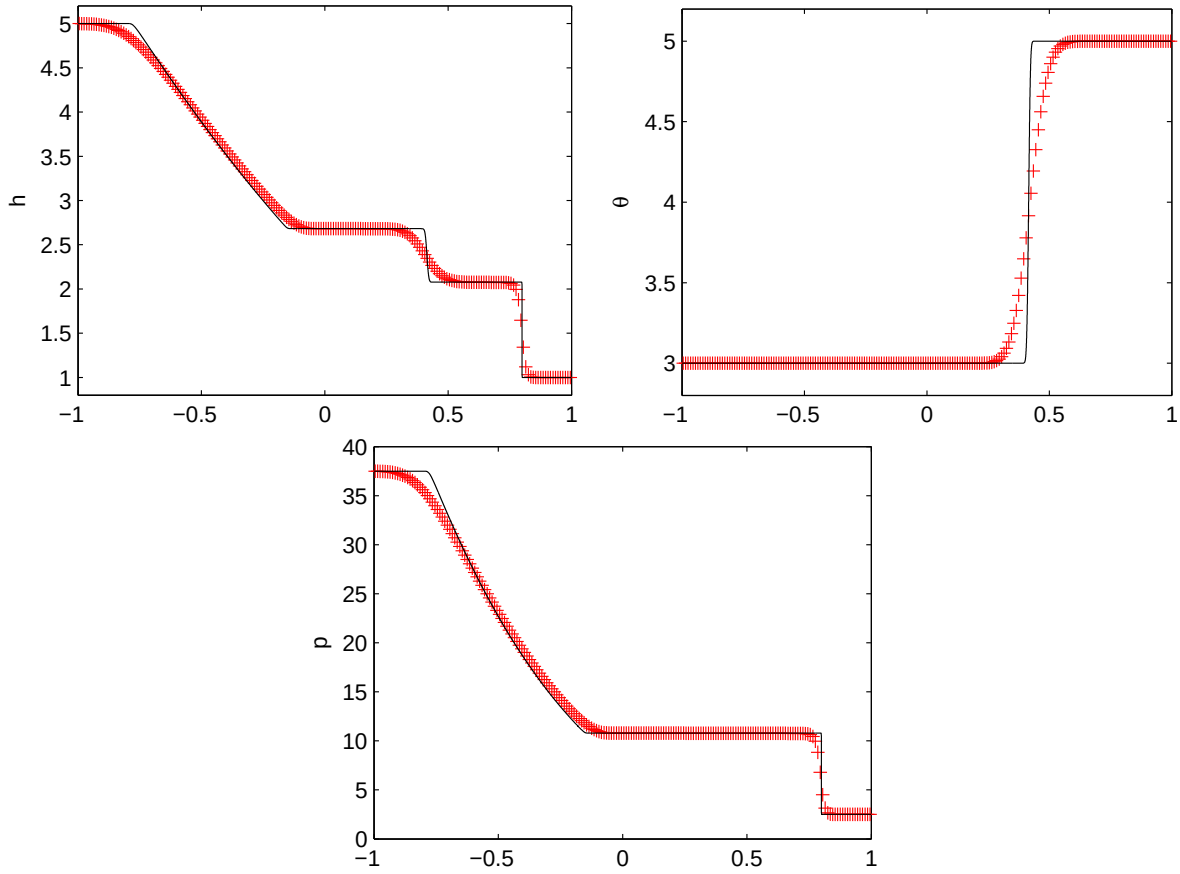


FIGURE 4.13 – Rupture de barrage sur un fond plat. Résultats au temps $T = 0.2$ avec 200 cellules. Les lignes noires représentent une solution de référence calculée avec 20 000 cellules – Haut, gauche : hauteur h. Haut, droite : température θ. Bas : pression p

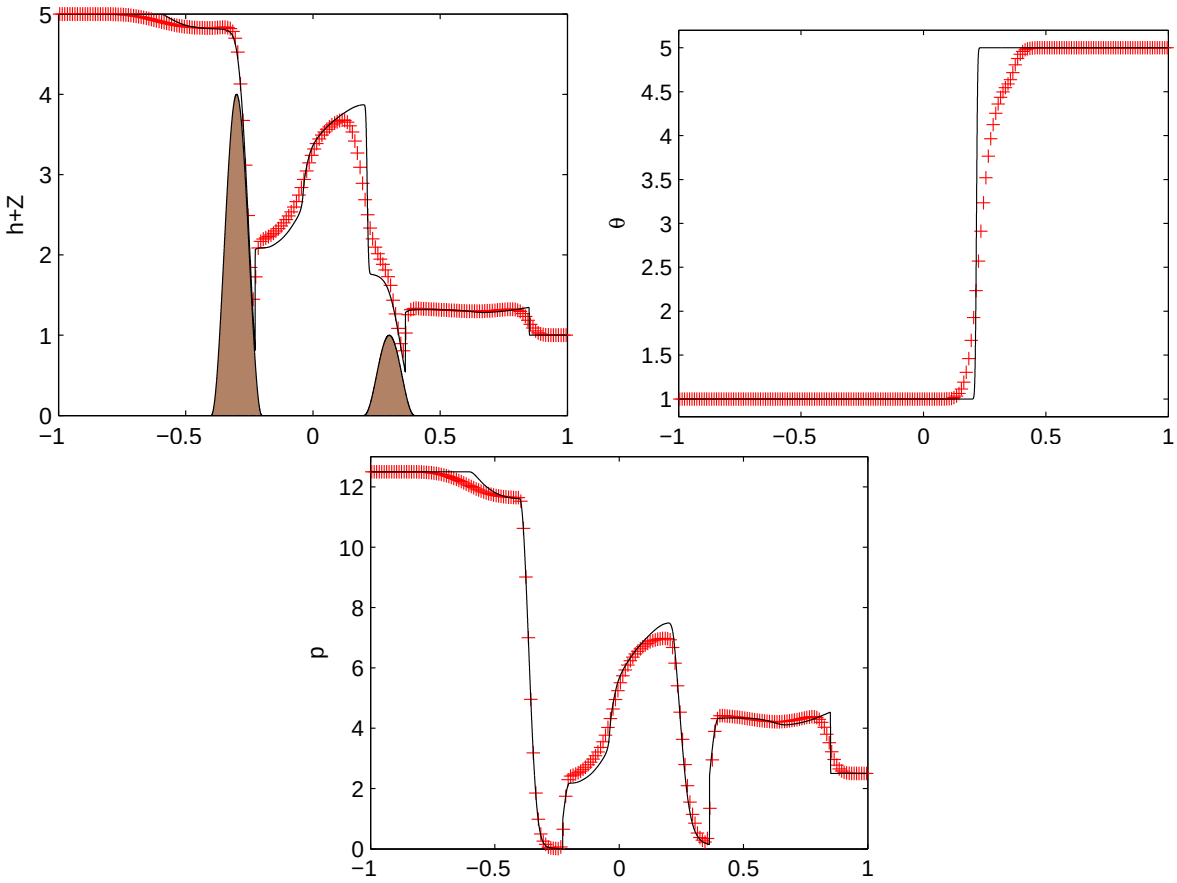


FIGURE 4.14 – Rupture de barrage avec topographie. Résultats au temps $T = 0.3$ avec 200 cellules. Les lignes noires représentent une solution de référence calculée avec 20 000 cellules – Haut, gauche : hauteur totale $h + Z$. Haut, droite : température θ. Bas : pression p

Petite perturbation d'un lac au repos

On considère ici une topographie contenant deux bosses isolées (voir Figure 4.15) :

$$Z(x) = \begin{cases} 0.85(\cos(10\pi(x+0.9))+1), & \text{si } -1.0 \leq x \leq -0.8, \\ 1.25(\cos(10\pi(x-0.4))+1), & \text{si } 0.3 \leq x \leq 0.5, \\ 0, & \text{sinon.} \end{cases}$$

On voit aisément que la solution

$$(h_s, u_s, \theta_s)(x) = \begin{cases} (6 - Z(x), 0, 4), & \text{si } x < 0, \\ (4 - Z(x), 0, 9), & \text{si } x > 0 \end{cases}$$

est un état d'équilibre, constitué de deux lacs au repos connectés par une discontinuité de contact stationnaire. On perturbe cette solution en considérant une donnée initiale donnée par

$$(h, u, \theta)(x, 0) = (h_s, u_s, \theta_s)(x) + (0.1, 0, 0)\chi_{[-1.5,-1.4]}(x),$$

où $\chi_{[-1.5,-1.4]}$ est la fonction indicatrice du segment $[-1.5, -1.4]$ (voir Figure 4.15). Cette perturbation va créer deux ondes se propageant dans les directions opposées. L'onde se propageant vers la droite rencontre successivement la première bosse, la discontinuité de contact stationnaire, puis la deuxième bosse.

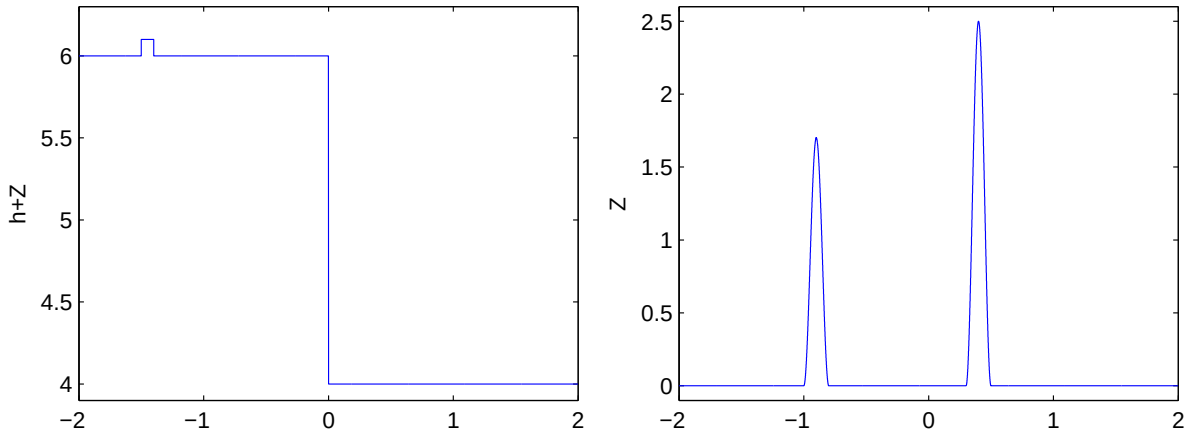


FIGURE 4.15 – Petite perturbation d'un lac au repos – Gauche : perturbation initiale en $h + Z$. Droite : topographie

Les résultats en hauteur totale $h + Z$ et en pression p, obtenus aux temps $T = 0.1$ et $T = 0.4$ avec 100 cellules sont présentés sur la Figure 4.16. On peut tout d'abord remarquer que le schéma de relaxation capture parfaitement la discontinuité de contact stationnaire, malgré la perturbation, ce qui n'est pas le cas de tous les schémas (voir [32]). D'autre part, il n'y a pas d'oscillations non physiques créées par la perturbation et les ondes développées diminuent avec le temps. Cela montre que le schéma de relaxation développé est stable au voisinage de cet état d'équilibre.

Petite perturbation d'un état d'équilibre complexe

On s'intéresse maintenant à un état d'équilibre plus compliqué que le lac au repos décrit précédemment. On se place sur le domaine $[-1, 1]$ et on considère une topographie donnée par

$$Z(x) = 6 - 2\exp(x).$$

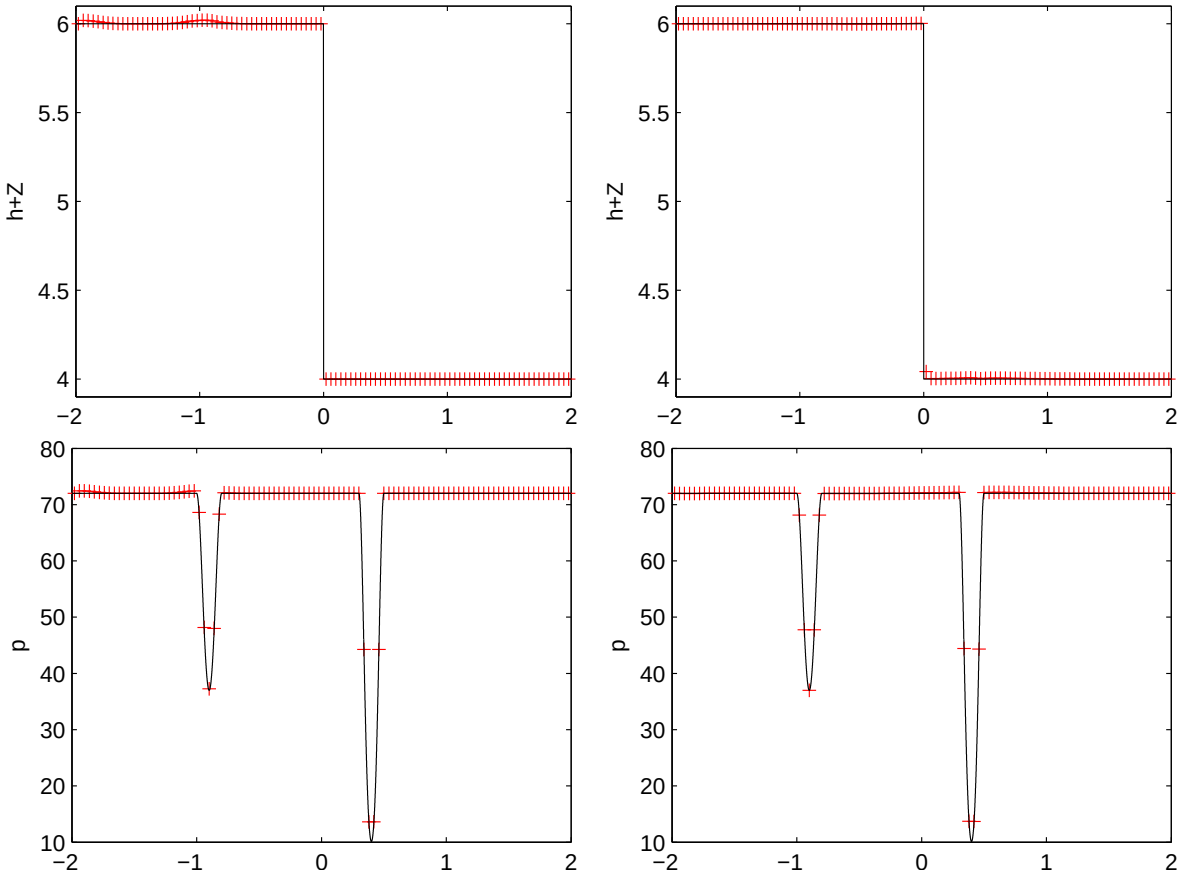


FIGURE 4.16 – Petite perturbation d'un lac au repos. Résultats obtenus avec 100 cellules. Les lignes noires représentent l'état stationnaire, sans la perturbation – Gauche : solution approchée au temps $T = 0.1$. Droite : solution approchée au temps $T = 0.4$. Haut : hauteur totale $h + Z$. Bas : pression p

On vérifie facilement que la solution

$$(h_w, u_s, \theta_x)(x) = (\exp(x), 0, \exp(2x))$$

est un état d'équilibre au repos pour le modèle (4.6). On introduit une perturbation en définissant la donnée initiale donnée par

$$(h, u, \theta)(x, 0) = (h_s, u_s, \theta_s)(x) + 0.1\chi_{[-0.1,0.0]}(x).$$

Les résultats en hauteur totale $h + Z$ et en vitesse u, obtenus aux temps $T = 0.2$ et $T = 0.4$, en utilisant 200 cellules, sont montrés Figure 4.17. On observe deux ondes créées par la perturbation et se propageant dans les directions opposées. Encore une fois, aucune oscillation non physique n'est présente et la perturbation diminue en intensité avec le temps. Cela prouve la stabilité numérique du schéma de relaxation au voisinage de l'état d'équilibre.

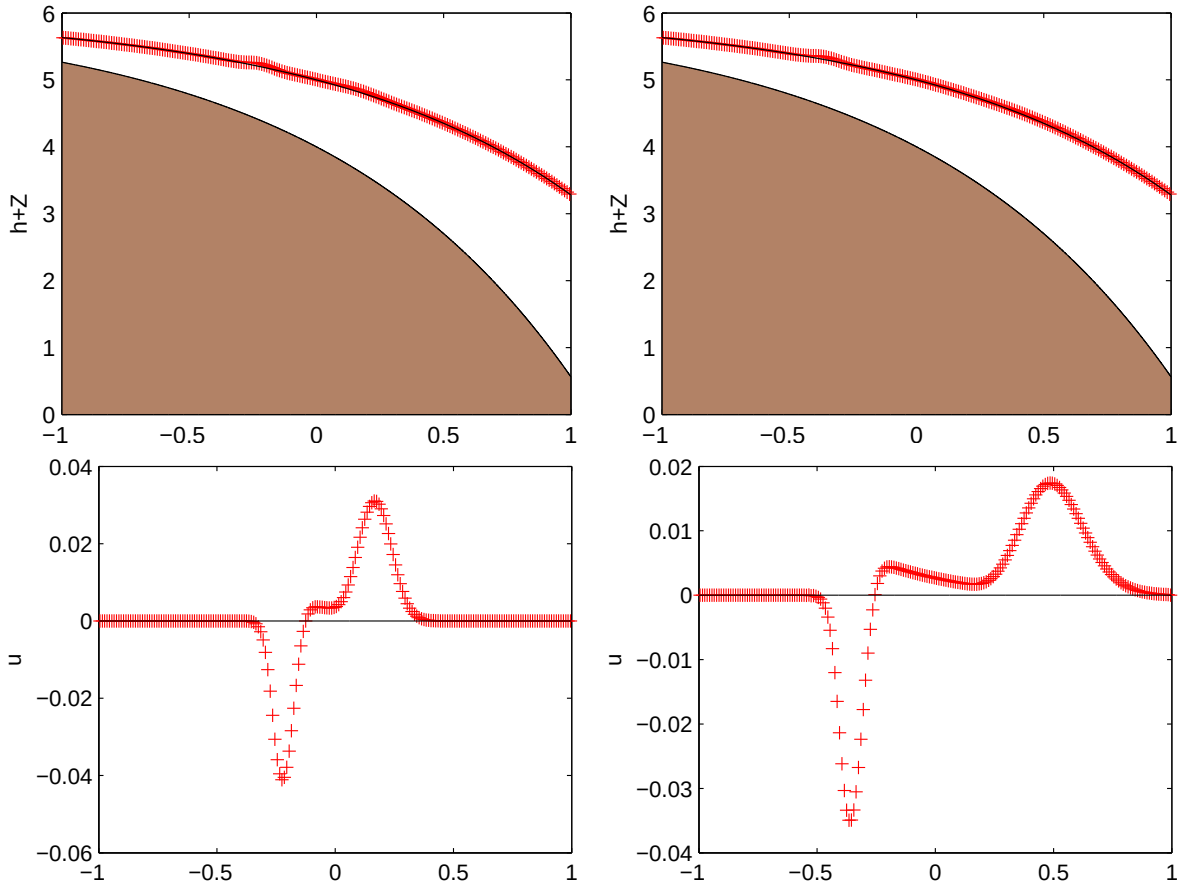


FIGURE 4.17 – Petite perturbation d'un état d'équilibre complexe. Résultats obtenus avec 200 cellules. Les lignes noires représentent l'état stationnaire, sans la perturbation – Gauche : solution approchée au temps $T = 0.2$. Droite : solution approchée au temps $T = 0.4$. Haut : hauteur totale $h + Z$. Bas : vitesse u

4.7.2 Équations d'Euler avec gravité

On passe maintenant à l'approximation numérique des équations d'Euler avec gravité (4.11). Pour simplifier, on considère que le système est fermé par la loi des gaz parfaits (4.166). On présente deux expériences inspirées de [71].

Atmosphère hydrostatique

Le but est ici de vérifier numériquement la propriété well-balanced des schémas construits. On considère pour cela une donnée initiale décrivant une atmosphère hydrostatique, définie par

$$\rho_s(x) = \left(1 - \frac{\gamma - 1}{\gamma} g x\right)^{\frac{1}{\gamma - 1}}, \quad u_s(x) = 0, \quad p_s(x) = \rho_s(x)^\gamma. \tag{4.202}$$

On vérifie aisément qu'il s'agit d'une solution stationnaire au repos pour le système (4.11). Pour les simulations, on fixe $g = 1$, $\gamma = 1.4$ et on considère le domaine de calcul $[0, 2]$ jusqu'au temps $T = 1.5$. La donnée initiale $w_s(x) = (\rho_s(x), u_s(x), p_s(x))^T$ est discrétisée de façon à ce que l'on obtienne une solution stationnaire discrète constante par morceaux, selon la Définition 4.8.

Dans le Tableau 4.1, on présente l'erreur L^1 en pression :

$$err = \|w^{\Delta x}(x, T) - w_s(x)\|_{L^1},$$

obtenue avec le schéma de relaxation d'ordre un, construit dans la Partie 4.4.4 et avec le schéma MUSCL dérivé dans la Partie 4.5. Pour le schéma MUSCL, on compare le limiteur minmod et le limiteur van Leer.

N	**Ordre un**		**MUSCL minmod**		**MUSCL van Leer**	
128	1.26E-16	–	2.89E-5	–	1.25E-7	–
256	1.14E-16	–	7.46E-6	1.95	1.60E-8	2.97
512	2.90E-16	–	1.90E-6	1.97	2.03E-9	2.98
1024	1.92E-16	–	4.78E-7	1.99	2.55E-10	2.99
2048	5.52E-16	–	1.20E-7	1.99	3.20E-11	2.99

Tableau 4.1 – Atmosphère hydrostatique : erreurs L^1 en pression pour le schéma de relaxation d'ordre un et le schéma MUSCL avec les limiteurs minmod et van Leer

Par construction, le schéma de relaxation préserve exactement les solutions discrètes stationnaires constantes par morceaux, ce qui est vérifié numériquement, à l'erreur machine près. Par contre, le schéma MUSCL d'ordre deux n'a *a priori* aucune raison de préserver cette solution stationnaire discrète constante par morceaux. En effet, en partant d'une solution stationnaire discrète constante par morceaux, les limiteurs classiques ne vérifient pas la relation (4.170) de manière exacte. Il n'y a donc aucune raison pour que la donnée initiale soit une solution stationnaire discrète affine par morceaux, qui serait préservée exactement par le schéma d'ordre deux.

On observe cependant que la solution donnée par le schéma MUSCL ne s'écarte pas trop de l'équilibre. En effet, l'erreur obtenue avec le limiteur minmod se comporte en $O(\Delta x^2)$ et celle obtenue avec le limiteur van Leer se comporte en $O(\Delta x^3)$.

Perturbation de l'atmosphère hydrostatique

On utilise à nouveau comme donnée initiale l'atmosphère hydrostatique décrite par (4.202). On introduit une perturbation sous la forme d'une condition de bord périodique appliquée au bord gauche :

$$u(0, t) = 0.1 \sin(6\pi t).$$

On observe alors le comportement de la perturbation

$$\delta w(x, t) = w(x, t) - w_s(x).$$

Le Tableau 4.2 présente les erreurs L^1 en perturbation pour la pression et en perturbation pour la vitesse. Ces erreurs sont calculées par rapport à une solution de référence w_{ref} obtenue avec le schéma d'ordre un en utilisant $32\ 768$ cellules. Elles s'écrivent ainsi

$$err_p = \|\delta p(x,T) - \delta p_{ref}(x,T)\|_{L^1}, \quad err_u = \|\delta u(x,T) - \delta u_{ref}(x,T)\|_{L^1}.$$

Pour le schéma MUSCL, on utilise le limiteur de van Leer, puisque l'on a montré précédemment qu'il était plus précis pour préserver l'atmosphère hydrostatique.

	Ordre un				**MUSCL**			
N	err_p		err_u		err_p		err_u	
128	1.95E-2	–	4.45E-2	–	1.74E-2	–	3.50E-2	–
256	1.23E-2	0.66	2.89E-2	0.62	8.68E-3	1.00	1.72E-2	1.02
512	7.09E-3	0.79	1.70E-2	0.77	4.26E-3	1.03	8.31E-3	1.05
1024	3.84E-3	0.88	9.25E-3	0.88	2.10E-3	1.02	4.05E-3	1.04
2048	1.96E-3	0.97	4.69E-3	0.98	1.05E-3	1.00	2.00E-3	1.02

Tableau 4.2 – Perturbation de l'atmosphère hydrostatique : erreurs L^1 en pression et en vitesse pour le schéma de relaxation d'ordre un et le schéma MUSCL

Le schéma MUSCL s'avère un peu plus précis que le schéma d'ordre un, bien que sa convergence numérique soit seulement en $O(\Delta x)$. Cela s'explique par le fait que les ondes de la perturbation prennent une forme en dents de scie en se propageant à travers l'atmosphère. Les discontinuités qui se forment ainsi empêchent le schéma MUSCL de converger plus vite que $O(\Delta x)$.

La Figure 4.18 montre les solutions obtenues par le schéma d'ordre un et par le schéma MUSCL en perturbation pour la pression et la densité en utilisant $1\ 024$ cellules. Ces solutions sont comparées à la solution de référence w_{ref}. Le schéma MUSCL capture bien les dents de scie mentionnées précédemment, malgré la présence de quelques légères oscillations. Le schéma d'ordre un est moins précis à cause de sa plus grande diffusion numérique.

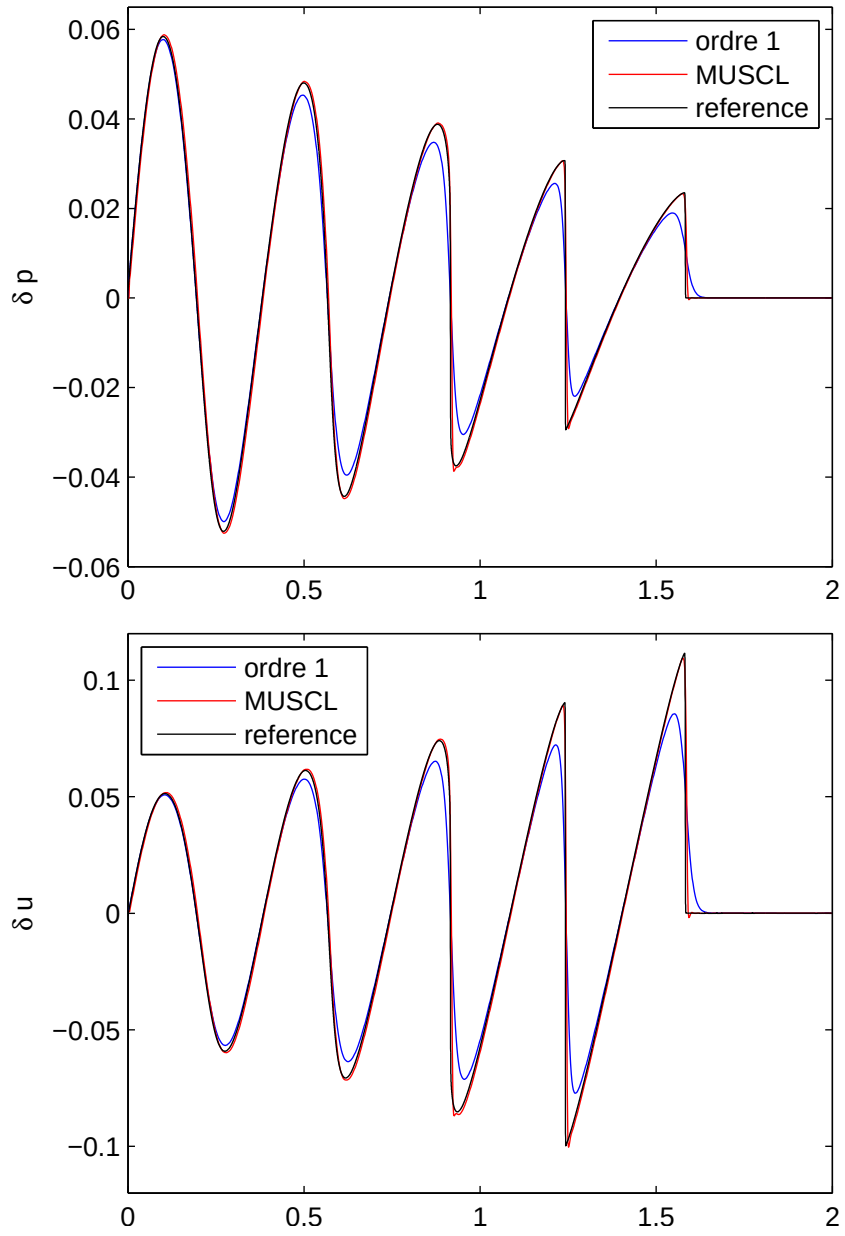


FIGURE 4.18 – Perturbation de l'atmosphère hydrostatique : Perturbation en pression (haut) et en vitesse (bas) – Noir : solution de référence. Bleu : schéma de relaxation d'ordre un. Rouge : schéma MUSCL

A

Preuve du Théorème de Lax-Wendroff

On donne ici la preuve du Théorème de Lax-Wendroff (3.1). La preuve existe dans [78] (voir aussi [55, 51, 80]), mais en considérant uniquement des schémas d'ordre élevé en espace et une solution à valeurs dans $\mathbb{R}^n$ tout entier. On présente ici une extension directe, en considérant une discrétisation d'ordre élevé en temps et une solution à valeurs dans un ensemble convexe $\Omega \subset \mathbb{R}^n$.

Par soucis de complétude, on rappelle les principales notations intervenant dans ce théorème. On cherche à approcher les solutions faibles d'un système hyperbolique de lois de conservation sous la forme condensée

$$\begin{cases} \partial_t w + \partial_x f(w) = 0, \\ w(x, t = 0) = w_0(x), \end{cases} \tag{A.1}$$

où $w : \mathbb{R} \times \mathbb{R}^+ \to \Omega$ est la fonction inconnue, $f : \Omega \to \mathbb{R}^d$ est la fonction flux, $w_0 \in L^1_{\text{loc}}(\mathbb{R}; \Omega) \cap L^\infty(\mathbb{R}; \Omega)$ est la donnée initiale et $\Omega \subset \mathbb{R}^d$ est l'ensemble convexe des états admissibles. Ce système est complété par les inégalités d'entropie

$$\partial_t \eta(w) + \partial_x \mathcal{G}(w) \leq 0, \tag{A.2}$$

où η est une fonction convexe vérifiant

$$\nabla_w f \nabla_w \eta = \nabla_w \mathcal{G}.$$

On considère une discrétisation régulière de l'espace en cellules $K_i = [x_{i-1/2}, x_{i+1/2}[$, avec $\Delta x = x_{i+1/2} - x_{i-1/2}$ le pas d'espace. On approche la donnée initiale par

$$w_i^0 = \frac{1}{\Delta x} \int_{K_i} w_0(x) dx.$$

En notant w_i^n une approximation de la solution au temps t^n sur la cellule K_i, la mise à jour au temps $t^{n+1} = t^n + \Delta t$ est donnée par un schéma en temps de Runge-Kutta à m étapes qui s'écrit

$$\begin{aligned} w_i^{n,(0)} &= w_i^n, \\ w_i^{n,(\ell)} &= w_i^n - \frac{\Delta t}{\Delta x} \sum_{j=0}^{\ell-1} c_{\ell,j} \left(F_{i+1/2}^{n,(j)} - F_{i-1/2}^{n,(j)} \right), \quad \ell = 1, \cdots, m, \\ w_i^{n+1} &= w_i^{n,(m)}. \end{aligned} \tag{A.3}$$

On suppose que les coefficients $c_{\ell,(j)}$ vérifient les propriétés de consistance suivantes

$$c_{\ell,j} \geq 0, \quad \sum_{j=1}^{m-1} c_{m,(j)} = 1. \tag{A.4}$$

Afin de permettre au schéma d'être d'ordre élevé en espace, on considère un flux numérique dépendant d'un large stencil :

$$F_{i+1/2}^{n,(j)} = F^s\left(w_{i-s+1}^{n,(j)}, \cdots, w_{i+s}^{n,(j)}\right), \tag{A.5}$$

où $F^s : \Omega^{2s} \to \mathbb{R}^3$ est continu et consistant :

$$F^s(w, \cdots, w) = f(w).$$

Pour simplifier ce qui va suivre, on définit deux sortes de cellules rectangulaires dans le plan (x, t) :

$$R_i^n = [x_{i-1/2}, x_{i+1/2}[\times[t^n, t^{n+1}[,$$

$$\widetilde{R}_{i+1/2}^n = [x_i, x_{i+1}[\times[t^n, t^{n+1}[$$

et on définit les fonctions constantes par morceaux suivantes :

$$w^\Delta(x,t) = w_i^n, \quad \text{pour } (x,t) \in R_i^n,$$

$$w^{\Delta,(\ell)}(x,t) = w_i^{n,(\ell)}, \quad \text{pour } (x,t) \in R_i^n.$$

Le théorème de Lax-Wendroff s'énonce alors comme suit.

Théorème A.1 (Lax-Wendroff)**.** *Supposons que la suite Δx tend vers 0 tout en préservant le ratio $\Delta t/\Delta x$ constant. On suppose que les hypothèses suivantes sont vérifiées :*

- *il existe un compact $K \subset \Omega$ tel que pour tout $0 \leq \ell \leq m$, la fonction $w^{\Delta,(\ell)}$ est à valeurs dans K ;*
- *la suite w^Δ converge dans $L^1_{loc}(\mathbb{R} \times \mathbb{R}^+; \Omega)$ vers une fonction w.*

Alors w est une solution faible de (A.1).

De plus, s'il existe un flux numérique d'entropie $G^s : \Omega^{2s} \to \mathbb{R}$ qui est lipschitzien et consistant :

$$G^s(w, \cdots, w) = \mathcal{G}(w),$$

et qui vérifie l'inégalité d'entropie discrète suivante :

$$\frac{1}{\Delta t}\left(\eta\left(w_i^{n+1}\right) - \eta\left(w_i^n\right)\right) + \frac{1}{\Delta x}\sum_{j=0}^{m-1} c_{m,j}\left(G_{i+1/2}^{n,(j)} - G_{i-1/2}^{n,(j)}\right) \leq 0, \tag{A.6}$$

où

$$G_{i+1/2}^{n,(j)} = G^s\left(w_{i-s+1}^{n,(j)}, \cdots, w_{i+s}^{n,(j)}\right),$$

alors w est une solution entropique de (A.1).

On définit une autre fonction constante par morceaux qui sera utile dans la preuve :

$$F^{\Delta,(l)}(x,t) = F^{n,(l)}_{i+1/2}, \quad \text{pour } (x,t) \in \widetilde{R}^n_{i+1/2}.$$

Dans la suite, les convergences seront implicitement considérées à sous-suite près. En particulier, on utilisera abusivement le fait que la convergence dans L^1_{loc} implique la convergence p.p. On verra à la fin de la preuve du Théorème A.1 que cela ne pose pas de problème.

La preuve s'organise de la façon suivante : le Lemme A.2 est un résultat technique concernant la convergence d'une suite décalée. Dans le Lemme A.3, on prouve que la convergence de $w^{\Delta,(j)}$ vers w dans L^1_{loc} implique la convergence p.p. de $F^{\Delta,(j)}$ vers $f(w)$. On déduit de ce résultat dans le Lemme A.4 que tous les $w^{\Delta,(j)}$ convergent vers w dans L^1_{loc}. Enfin, grâce aux Lemmes A.3 et A.4, on peut prouver le Théorème A.1.

Lemme A.2. *On considère une suite de fonctions $u^\Delta : \mathbb{R} \to \Omega$ vérifiant les hypothèses suivantes :*

(i) il existe un compact $K \subset \Omega$ tel que u^Δ soit à valeurs dans K ;

(ii) u^Δ converge dans $L^1_{loc}(\mathbb{R};\Omega)$ vers une fonction u.

Alors, pour tout $\xi \in \mathbb{R}$, la suite $u^\Delta(x+\xi\Delta x)$ converge vers $u(x)$ pour presque tout $x \in \mathbb{R}$.

Démonstration. Soit $a < b$ deux réels. On définit

$$I^\Delta = \int_a^b \left\| u^\Delta(x+\xi\Delta x) - u(x) \right\| dx.$$

L'inégalité triangulaire nous donne immédiatement

$$I^\Delta \leq \int_a^b \left\| u^\Delta(x+\xi\Delta x) - u(x+\xi\Delta x) \right\| dx + \int_a^b \left\| u(x+\xi\Delta x) - u(x) \right\| dx.$$

On note respectivement I_1^Δ et I_2^Δ les deux intégrales apparaissant dans le second membre. On va montrer qu'elles tendent toutes les deux vers zéro.

Pour I_1^Δ, un changement de variable donne

$$I_1^\Delta = \int_{a+\xi\Delta x}^{b+\xi\Delta x} \left\| u^\Delta(x) - u(x) \right\| dx,$$

donc on a

$$I_1^\Delta = \int_a^b \left\| u^\Delta(x) - u(x) \right\| dx - \int_a^{a+\xi\Delta x} \left\| u^\Delta(x) - u(x) \right\| dx + \int_b^{b+\xi\Delta x} \left\| u^\Delta(x) - u(x) \right\| dx.$$

Grâce à l'hypothèse (i), on a

$$\int_a^{a+\xi\Delta x} \left\| u^\Delta(x) - u(x) \right\| dx \leq 2\xi\Delta x \sup_{w\in K} \|w\| \to 0$$

et il en est de même pour $\int_b^{b+\xi\Delta x} \left\| u^\Delta(x) - u(x) \right\| dx$. L'intégrale $\int_a^b \left\| u^\Delta(x) - u(x) \right\| dx$ converge quant à elle vers zéro par l'hypothèse (ii). On a ainsi montré $I_1^\Delta \to 0$.

On étudie maintenant la convergence de I_2^Δ. L'ensemble des fonctions continues étant dense dans $L^1(\mathbb{R};\Omega)$, pour tout $\epsilon > 0$, on peut trouver une fonction continue ψ telle que

$$\|\psi - u\|_{L^1([a,b+\xi\Delta x];\Omega)} \leq \epsilon.$$

L'inégalité triangulaire nous donne alors

$$I_2^\Delta \leq \int_a^b \|u(x+\xi\Delta x) - \psi(x+\xi\Delta x)\|\,dx + \int_a^b \|\psi(x+\xi\Delta x) - \psi(x)\|\,dx + \int_a^b \|\psi(x) - u(x)\|\,dx. \tag{A.7}$$

La première intégrale de (A.7) s'écrit

$$\int_a^b \|u(x+\xi\Delta x) - \psi(x+\xi\Delta x)\|\,dx = \int_{a+\xi\Delta x}^{b+\xi\Delta x} \|u(x) - \psi(x))\|\,dx.$$

Elle est donc inférieure à ϵ par définition de ψ, de même que la troisième intégrale de (A.7). Enfin, la deuxième intégrale de (A.7) converge vers zéro par continuité de ψ. On a ainsi montré $I_2^\Delta \to 0$ et donc $I^\Delta \to 0$.

Cela signifie que $x \mapsto u^\Delta(x+\xi\Delta x)$ converge vers u dans L^1_{loc}. Par conséquent, $u^\Delta(x+\xi\Delta x)$ converge vers $u(x)$ pour presque tout $x \in \mathbb{R}$. □

Lemme A.3. *Sous les hypothèses du Théorème A.1, si $w^{\Delta,(j)}$ converge vers w dans $L^1_{loc}(\mathbb{R}\times\mathbb{R}^+;\Omega)$, alors $F^{\Delta,(j)}$ converge vers $f(w)$ p.p.*

Démonstration. On remarque dans un premier temps que

$$x \in \widetilde{R}^n_{i+1/2} \quad \Rightarrow \quad x + (k-1/2)\Delta x \in R^n_{i+k}.$$

Par conséquent, on peut réécrire l'équation (A.5) de la façon suivante :

$$F^{\Delta,(j)}(x,t) = F\left(w^{\Delta,(j)}\left(x-(s-1/2)\Delta x,t\right),\cdots,w^{\Delta,(j)}\left(x+(s-1/2)\Delta x,t\right)\right).$$

La suite de fonctions $x \mapsto w^{\Delta,(j)}(x,t)$ vérifie les hypothèses du Lemme A.2 pour presque tout $t \in \mathbb{R}^+$. On en déduit que pour tout $\xi \in \mathbb{R}$, $w^{\Delta,(j)}(x+\xi\Delta x,t)$ converge vers $w(x,t)$ pour presque tout $(x,t) \in \mathbb{R}\times\mathbb{R}^+$. Grâce à la continuité et à la consistance de F, on conclut que $F^{\Delta,(j)}$ converge vers $f(w)$ p.p. □

Lemme A.4. *Sous les hypothèses du Théorème A.1, la suite $w^{\Delta,(\ell)}$ converge vers w dans $L^1_{loc}(\mathbb{R}\times\mathbb{R}^+;\Omega)$, pour tout $\ell = 0,\cdots,m-1$.*

Démonstration. On prouve ce lemme par récurrence sur ℓ. Le résultat est vrai par hypothèse pour $\ell = 0$, puisque $w^{\Delta,(0)} = w^\Delta$. Supposons que pour tout $j = 0,\cdots,\ell-1$, la suite $w^{\Delta,(j)}$ converge vers w dans $L^1_{\text{loc}}(\mathbb{R}\times\mathbb{R}^+;\Omega)$. Le Lemme A.3 assure alors que $F^{\Delta,(j)}$ converge vers $f(w)$ p.p. pour tout $j = 0,\cdots,\ell-1$. De plus, $F^{\Delta,(j)}$ est à valeurs dans le compact $F(K,\cdots,K)$, donc en utilisant le Lemme A.2, on déduit que $(x,t) \mapsto F^{\Delta,(j)}(x+\Delta x/2,t)$ et $(x,t) \mapsto F^{\Delta,(j)}(x-\Delta x/2)$ convergent tous les deux vers $f(w)$ p.p.

L'équation (A.3) se réécrit

$$w^{\Delta,(\ell)}(x,t) = w^\Delta(x,t) - \frac{\Delta t}{\Delta x}\sum_{j=0}^{\ell-1} c_{\ell,j}\left(F^{\Delta,(j)}(x+\Delta x/2,t) - F^{\Delta,(j)}(x-\Delta x/2)\right).$$

Chaque terme de la somme converge vers zéro p.p. et w^Δ converge vers w p.p., donc $w^{\Delta,(\ell)}$ converge vers w p.p. La suite $w^{\Delta,(\ell)}$ étant uniformément bornée, le théorème de convergence dominée assure que $w^{\Delta,(\ell)}$ converge vers w dans $L^1_{\text{loc}}(\mathbb{R}\times\mathbb{R}^+;\Omega)$. □

On peut maintenant prouver le Théorème A.1.

Preuve du Théorème A.1. Soit $\phi \in C_c^1(\mathbb{R}\times\mathbb{R}^+;\mathbb{R}^d)$ une fonction test régulière à support compact. Pour $i \in \mathbb{Z}$ et $n \in \mathbb{N}$, on définit $\phi_i^n = \phi(x_i, t^n)$. En multipliant la dernière itération ($\ell = m$) du schéma (A.3) par $\Delta x \phi_i^n$ et en sommant sur i et sur n, on obtient

$$\Delta x \sum_{i,n} \left(w_i^{n+1} - w_i^n\right)\cdot \phi_i^n + \Delta t \sum_{i,n} \phi_i^n \cdot \sum_{j=0}^{m-1} c_{m,j}\left(F_{i+1/2}^{n,(j)} - F_{i-1/2}^{n,(j)}\right) = 0.$$

Une intégration par parties donne

$$\Delta x \sum_{i,n} w_i^{n+1}\cdot\left(\phi_i^{n+1} - \phi_i^n\right) + \Delta x \sum_i w_i^0 \cdot \phi_i^0 + \Delta t \sum_{i,n}\left(\phi_{i+1}^n - \phi_i^n\right)\cdot \sum_{j=0}^{m-1} c_{m,j} F_{i+1/2}^{n,(j)} = 0. \quad \text{(A.8)}$$

On définit la fonction constante par morceaux $\phi^\Delta(x,t) = \phi_i^n = \phi(x_i,t^n)$ pour $(x,t) \in R_i^n$. On peut alors mettre l'équation (A.8) sous forme intégrale :

$$\begin{aligned}\int_{\mathbb{R}\times[\Delta t,+\infty[} w^\Delta(x,t)\cdot\frac{\phi^\Delta(x,t)-\phi^\Delta(x,t-\Delta t)}{\Delta t}dxdt + \int_{\mathbb{R}} w_0(x)\cdot\phi^\Delta(x,0)dx \\ + \int_{\mathbb{R}\times\mathbb{R}^+}\frac{\phi^\Delta(x+\Delta x/2,t)-\phi^\Delta(x-\Delta x/2,t)}{\Delta x}\cdot\sum_{j=0}^{m-1} c_{m,j}F^{\Delta,(j)}(x,t)dxdt = 0. \quad \text{(A.9)}\end{aligned}$$

La fonction ϕ étant régulière, ϕ^Δ converge uniformément vers ϕ et puisque w_0 est essentiellement bornée, on a

$$\int_{\mathbb{R}} w_0(x)\cdot\phi^\Delta(x,0)dx \to \int_{\mathbb{R}} w_0(x)\cdot\phi(x,0)dx. \quad \text{(A.10)}$$

On note respectivement I_1^Δ et I_2^Δ la première intégrale de (A.9) qui correspond à la dérivée en temps et la troisième intégrale de (A.9) qui correspond à la dérivée en espace

Convergence de la discrétisation en temps I_1^Δ

L'intégrale I_1^Δ s'écrit

$$I_1^\Delta = \int_{\mathbb{R}\times\mathbb{R}^+} w^\Delta(x,t)\cdot 1_{\mathbb{R}\times[\Delta t,+\infty[}(x,t)\frac{\phi^\Delta(x,t)-\phi^\Delta(x,t-\Delta t)}{\Delta t}dxdt.$$

La fonction $(x,t)\mapsto 1_{\mathbb{R}\times[\Delta t,+\infty[}(x,t)\frac{\phi^\Delta(x,t)-\phi^\Delta(x,t-\Delta t)}{\Delta t}$ converge uniformément vers $\partial_t\phi$ et les fonctions w^Δ sont uniformément essentiellement bornées, donc

$$\int_{\mathbb{R}\times\mathbb{R}^+} w^\Delta(x,t)\cdot\left(1_{\mathbb{R}\times[\Delta t,+\infty[}(x,t)\frac{\left(\phi^\Delta(x,t)-\phi^\Delta(x,t-\Delta t)\right)}{\Delta t} - \partial_t\phi(x,t)\right)dxdt \to 0. \quad \text{(A.11)}$$

D'autre part, comme w^Δ converge vers w dans $L^1_{\text{loc}}(\mathbb{R}\times\mathbb{R}^+)$, on a

$$\int_{\mathbb{R}\times\mathbb{R}^+} w^\Delta(x,t)\cdot\partial_t\phi(x,t)dxdt \to \int_{\mathbb{R}\times\mathbb{R}^+} w(x,t)\cdot\partial_t\phi(x,t)dxdt. \quad \text{(A.12)}$$

On déduit des limites (A.11) et (A.12) que

$$I_1^\Delta \to \int_{\mathbb{R}\times\mathbb{R}^+} w(x,t)\cdot\partial_t\phi(x,t)dxdt. \quad \text{(A.13)}$$

Convergence de la discrétisation en temps I_2^Δ

En utilisant à nouveau le fait que ϕ est régulière, la suite

$$(x,t) \mapsto \frac{\phi^{\Delta}(x+\Delta x/2,t)-\phi^{\Delta}(x-\Delta x/2,t)}{\Delta x}$$

converge uniformément vers $\partial_x\phi$. D'autre part, les fonctions $F^{\Delta,(j)}$ sont toutes uniformément bornées car elles sont à valeurs dans le compact $F(K,\cdots,K)$. Par conséquent, on a

$$\int_{\mathbb{R}\times\mathbb{R}^+}\left(\frac{\phi^{\Delta}(x+\Delta x/2,t)-\phi^{\Delta}(x-\Delta x/2,t)}{\Delta x}-\partial_x\phi(x,t)\right)\cdot\sum_{j=0}^{m-1}c_{m,j}F^{\Delta,(j)}(x,t)dxdt\to 0. \tag{A.14}$$

En combinant les Lemmes A.3 et A.4, on obtient que la suite fonctions $F^{\Delta,(j)}$ converge p.p vers $f(w)$, pour $j=0,\cdots,m-1$. En utilisant (A.4), on en déduit que $\sum_{j=0}^{m-1}c_{m,j}F^{\Delta,(j)}$ converge presque partout vers $f(w)$.

Le théorème de convergence dominée assure alors que

$$\int_{\mathbb{R}\times\mathbb{R}^+}\partial_x\phi(x,t)\cdot\sum_{j=0}^{m-1}c_{m,j}F^{\Delta,(j)}(x,t)dxdt\to\int_{\mathbb{R}\times\mathbb{R}^+}f(w(x,t))\cdot\partial_x\phi(x,t)dxdt. \tag{A.15}$$

On déduit alors des relations (A.14) et (A.15)

$$I_2^{\Delta}\to\int_{\mathbb{R}\times\mathbb{R}^+}f(w(x,t))\cdot\partial_x\phi(x,t)dxdt. \tag{A.16}$$

Les trois limites (A.10), (A.13) et (A.16) sont vraies à une sous-suite près. On peut clairement trouver une sous-suite commune qui vérifie les trois limites. En prenant la limite pour cette sous-suite dans l'équation (A.9), on obtient

$$\int_{\mathbb{R}\times\mathbb{R}^+}w(x,t)\cdot\partial_t\phi(x,t)dxdt+\int_{\mathbb{R}}w_0(x)\cdot\phi(x,0)dx+\int_{\mathbb{R}\times\mathbb{R}^+}f(w(x,t))\cdot\partial_x\phi(x,t)dxdt=0,$$

ce qui prouve que w est une solution faible de (A.1).

La preuve du résultat concernant l'entropie est similaire et ne présente pas de difficulté supplémentaire. □

Liste des tableaux

Table des figures

Bibliographie

[1] R. Abgrall. A review of residual distribution schemes for hyperbolic and parabolic problems : the July 2010 state of the art. *Commun. Comput. Phys.*, 11(4) :1043–1080, 2012. 66

[2] B. Andreianov, M. Bendahmane, and K. H. Karlsen. Discrete duality finite volume schemes for doubly nonlinear degenerate hyperbolic-parabolic equations. *J. Hyperbolic Differ. Equ.*, 7(1) :1–67, 2010. 58

[3] B. Andreianov, F. Boyer, and F. Hubert. Discrete duality finite volume schemes for Leray-Lions-type elliptic problems on general 2D meshes. *Numer. Methods Partial Differential Equations*, 23(1) :145–195, 2007. 11, 30

[4] M. Artola and A.J. Majda. Nonlinear development of instabilities in supersonic vortex sheets. i. the basic kink modes. *Phys. D*, 28(3) :253–281, 1987. 45, 47

[5] E. Audusse, F. Bouchut, M.O. Bristeau, R. Klein, and B. Perthame. A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows. *SIAM J. Sci. Comput.*, 25(6) :2050–2065, 2004. 10, 87, 89, 151

[6] T. Barth and D. Jespersen. The design and application of upwind schemes on unstructured meshes. In *AIAA, Aerospace Sciences Meeting, 27 th, Reno, NV*, 1989. 10, 30, 56

[7] P. Batten, N. Clarke, C. Lambert, and D. M. Causon. On the choice of wavespeeds for the HLLC Riemann solver. *SIAM J. Sci. Comput.*, 18(6) :1553–1570, 1997. 8, 62

[8] M. Ben-Artzi and J. Falcovitz. A second-order Godunov-type scheme for compressible fluid dynamics. *J. Comput. Phys.*, 55(1) :1–32, 1984. 8, 63, 66

[9] M. Berger, M.J. Aftosmis, and S.M. Murman. Analysis of slope limiters on irregular grids. In *43rd AIAA Aerospace Sciences Meeting*, volume NAS Technical Report NAS-05-007, 2005. 10, 30

[10] A. Bermudez and M.E. Vazquez. Upwind methods for hyperbolic conservation laws with source terms. *Comput. & Fluids*, 23(8) :1049–1071, 1994. 10, 87

[11] C. Berthon. Inégalités d'entropie pour un schéma de relaxation. *C. R. Math. Acad. Sci. Paris*, 340(1) :63–68, 2005. 8, 66, 71, 118

[12] C. Berthon. Stability of the MUSCL schemes for the Euler equations. *Commun. Math. Sci.*, 3(2) :133–157, 2005. 8, 9, 12, 25, 30, 63, 64, 67, 71, 76, 80

[13] C. Berthon. Numerical approximations of the 10-moment Gaussian closure. *Math. Comp.*, 75(256) :1809–1831 (electronic), 2006. 66, 118

[14] C. Berthon. Robustness of MUSCL schemes for 2D unstructured meshes. *J. Comput. Phys.*, 218(2) :495–509, 2006. 10, 11, 30, 36, 37, 55, 63, 71, 156

[15] C. Berthon and F. Foucher. Efficient well-balanced hydrostatic upwind schemes for shallow-water equations. *J. Comput. Phys.*, 231(15) :4993–5015, 2012. 10, 87, 151, 156

[16] C. Berthon, P.G. LeFloch, and R. Turpault. Late-time/stiff-relaxation asymptotic-preserving approximations of hyperbolic equations. *Math. Comp.*, 82(282) :831–860, 2013. 13, 118

[17] C. Berthon and F. Marche. A positive preserving high order VFRoe scheme for shallow water equations : a class of relaxation schemes. *SIAM J. Sci. Comput.*, 30(5) :2587–2612, 2008. 8, 63, 66, 84

[18] C. Berthon, F. Marche, and R. Turpault. An efficient scheme on wet/dry transitions for shallow water equations with friction. *Comput. & Fluids*, 48 :192–201, 2011. 89

[19] C. Berthon, C. Sarazin, and R. Turpault. Space-time generalized Riemann problem solvers of order k for linear advection with unrestricted time step. *J. Sci. Comput.*, 55(2) :268–308, 2013. 8

[20] F. Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources*. Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2004. 7, 8, 10, 13, 30, 35, 45, 62, 63, 66, 71, 78, 87, 88, 89, 118, 122, 124, 151

[21] F. Bouchut, C. Bourdarias, and B. Perthame. A MUSCL method satisfying all the numerical entropy inequalities. *Math. Comp.*, 65(216) :1439–1461, 1996. 8, 12, 25, 30, 63, 64, 66, 67, 76

[22] A. Bourgeade, P.G. LeFloch, and P.A. Raviart. Approximate solution of the generalized Riemann problem and applications. In *Nonlinear hyperbolic problems (St. Etienne, 1986)*, volume 1270 of *Lecture Notes in Math.*, pages 1–9. Springer, Berlin, 1987. 8, 63

[23] A. Bourgeade, P.G. LeFloch, and P.A. Raviart. An asymptotic expansion for the solution of the generalized Riemann problem. II. Application to the equations of gas dynamics. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 6(6) :437–480, 1989. 8, 63

[24] T. Buffard and S. Clain. Monoslope and multislope MUSCL methods for unstructured meshes. *J. Comput. Phys.*, 229(10) :3745–3776, 2010. 10, 30

[25] T. Buffard, T. Gallouet, and J.M. Hérard. Un schéma simple pour les équations de saint-venant. *C. R. Acad. Sci. Paris Sér. I Math.*, 326(3) :385–390, 1998. 10, 87

[26] T. Buffard, T. Gallouët, and J.M. Hérard. A sequel to a rough Godunov scheme : application to real gases. *Comput. & Fluids*, 29(7) :813–847, 2000. 8, 63

[27] C. Calgaro, E. Chane-Kane, E. Creusé, and T. Goudon. L^∞-stability of vertex-based MUSCL finite volume schemes on unstructured grids : simulation of incompressible flows with high density ratios. *J. Comput. Phys.*, 229(17) :6027–6046, 2010. 10, 30

[28] C. Calgaro, E. Creusé, T. Goudon, and Y. Penel. Positivity-preserving schemes for Euler equations : sharp and practical CFL conditions. *J. Comput. Phys.*, 234 :417–438, 2013. 10, 30

[29] C. Chalons, F. Coquel, E. Godlewski, P.A. Raviart, and N. Seguin. Godunov-type schemes for hyperbolic systems with parameter-dependent source. the case of euler system with friction. *Math. Models Methods Appl. Sci.*, 20(11) :2109–2166, 2010. 8, 62, 88, 100

[30] C. Chalons and J.F. Coulombel. Relaxation approximation of the euler equations. *J. Math. Anal. Appl.*, 348(2) :872–893, 2008. 12, 71, 78, 80, 118

[31] G.Q. Chen, C.D. Levermore, and T.P. Liu. Hyperbolic conservation laws with stiff relaxation terms and entropy. *Comm. Pure Appl. Math.*, 47(6) :787–830, 1994. 13, 118, 119

[32] A. Chertock, A. Kurganov, and Y. Liu. Central-upwind schemes for the system of shallow water equations with horizontal temperature gradients. soumis. 10, 92, 163, 166

[33] S. Clain and V. Clauzon. L^∞ stability of the MUSCL methods. *Numer. Math.*, 116(1) :31–64, 2010. 8, 10, 30, 63

[34] S. Clain, S. Diot, and R. Loubère. A high-order finite volume method for systems of conservation laws—Multi-dimensional Optimal Order Detection (MOOD). *J. Comput. Phys.*, 230(10) :4028–4050, 2011. 8, 10, 11, 27, 28, 30, 58, 63, 64, 66, 80, 81

[35] P. Colella and P.R. Woodward. The piecewise parabolic method (ppm) for gas-dynamical simulations. *Journal of Computational Physics*, 54(1) :174 – 201, 1984. 8, 63

[36] F. Coquel, P. Helluy, and J. Schneider. Second-order entropy diminishing scheme for the Euler equations. *Internat. J. Numer. Methods Fluids*, 50(9) :1029–1061, 2006. 8, 63

[37] F. Coquel and P.G. LeFloch. An entropy satisfying MUSCL scheme for systems of conservation laws. *Numer. Math.*, 74(1) :1–33, 1996. 8, 63

[38] F. Coquel and B. Perthame. Relaxation of energy and approximate Riemann solvers for general pressure laws in fluid dynamics. *SIAM J. Numer. Anal.*, 35(6) :2223–2249 (electronic), 1998. 35, 45, 88, 118

[39] Y. Coudière and F. Hubert. A 3D discrete duality finite volume method for nonlinear elliptic equations. *SIAM J. Sci. Comput.*, 33(4) :1739–1764, 2011. 59

[40] Y. Coudière, C. Pierre, O. Rousseau, and R. Turpault. A 2D/3D discrete duality finite volume scheme. Application to ECG simulation. *Int. J. Finite Vol.*, 6(1) :24, 2009. 58

[41] J.F. Coulombel and P. Secchi. The stability of compressible vortex sheets in two space dimensions. *Indiana Univ. Math. J.*, 53(4) :941–1012, 2004. 45, 47

[42] R. Courant, K. Friedrichs, and H. Lewy. Über die partiellen Differenzengleichungen der mathematischen Physik. *Math. Ann.*, 100(1) :32–74, 1928. 18

[43] P.H. Cournède, B. Koobus, and A. Dervieux. Positivity statements for a mixed-element-volume scheme on fixed and moving grids. *European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique*, 15(7-8) :767–798, 2006. 10, 30

[44] C.M. Dafermos. *Hyperbolic conservation laws in continuum physics*, volume 325 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 2010. 61

[45] M.S. Darwish and F. Moukalled. Tvd schemes for unstructured grids. *International Journal of heat and mass transfer*, 46(4) :599–611, 2003. 10, 30, 56

[46] O. Delestre, C. Lucas, P.A. Ksinant, F. Darboux, C. Laguerre, T.N.T Vo, F. James, and S. Cordier. Swashes : a compilation of shallow water analytic solutions for hydraulic and environmental studies. *International Journal for Numerical Methods in Fluids*, 72(3) :269–300, May 2013. 40 pages. 10, 87

[47] S. Diot, S. Clain, and R. Loubère. Improved detection criteria for the multi-dimensional optimal order detection (MOOD) on unstructured meshes with very high-order polynomials. *Comput. & Fluids*, 64 :43–63, 2012. 8, 10, 11, 27, 28, 30, 46, 63, 64, 80

[48] R.J. DiPerna. Convergence of approximate solutions to conservation laws. *Arch. Rational Mech. Anal.*, 82(1) :27–70, 1983. 70

[49] K. Domelevo and P. Omnes. A finite volume method for the Laplace equation on almost arbitrary two-dimensional grids. *M2AN Math. Model. Numer. Anal.*, 39(6) :1203–1249, 2005. 11, 30, 42

[50] B. Dubroca. Solveur de Roe positivement conservatif. *C. R. Acad. Sci. Paris Sér. I Math.*, 329(9) :827–832, 1999. 8, 63

[51] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In *Handbook of numerical analysis, Vol. VII*, Handb. Numer. Anal., VII, pages 713–1020. North-Holland, Amsterdam, 2000. 64, 66, 173

[52] G. Gallice. Solveurs simples positifs et entropiques pour les systèmes hyperboliques avec terme source. *C. R. Math. Acad. Sci. Paris*, 334(8) :713–716, 2002. 8, 12, 62, 88, 100

[53] G. Gallice. Positive and entropy stable Godunov-type schemes for gas dynamics and MHD equations in Lagrangian or Eulerian coordinates. *Numer. Math.*, 94(4) :673–713, 2003. 8, 12, 62, 88, 100

[54] R. Ghostine, G. Kesserwani, R. Mosé, J. Vazquez, and A. Ghenaim. An improvement of classical slope limiters for high-order discontinuous galerkin method. *Internat. J. Numer. Methods Fluids*, 59(4) :423–442, 2009. 30

[55] E. Godlewski and P.A. Raviart. *Hyperbolic systems of conservation laws*, volume 3/4 of *Mathématiques & Applications (Paris) [Mathematics and Applications]*. Ellipses, Paris, 1991. 61, 66, 173

[56] E. Godlewski and P.A. Raviart. *Numerical approximation of hyperbolic systems of conservation laws*, volume 118 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1996. 7, 9, 10, 15, 30, 35, 62, 156

[57] S.K. Godunov. A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Mat. Sb. (N.S.)*, 47 (89) :271–306, 1959. 17

[58] L. Gosse. A well-balanced flux-vector splitting scheme designed for hyperbolic systems of conservation laws with source terms. *Comput. Math. Appl.*, 39(9-10) :135–159, 2000. 10, 87

[59] L. Gosse. Computing qualitatively correct approximations of balance laws. *Computing*, 2, 2013. 7, 62

[60] J.M. Greenberg and A.Y. Leroux. A well-balanced scheme for the numerical processing of source terms in hyperbolic equations. *SIAM J. Numer. Anal.*, 33(1) :1–16, 1996. 10, 87

[61] J.M. Greenberg, A.Y. Leroux, R. Baraille, and A. Noussair. Analysis and approximation of conservation laws with source terms. *SIAM J. Numer. Anal.*, 34(5) :1980–2007, 1997. 10, 87

[62] B. Hanouzet and R. Natalini. Global existence of smooth solutions for partially dissipative hyperbolic systems with a convex entropy. *Arch. Ration. Mech. Anal.*, 169(2) :89–117, 2003. 119

[63] A. Harten, P.D. Lax, C.D. Levermore, and W.J. Morokoff. Convex entropies and hyperbolicity for general Euler equations. *SIAM J. Numer. Anal.*, 35(6) :2117–2127 (electronic), 1998. 62

[64] A. Harten, P.D. Lax, and B. van Leer. On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. *SIAM Rev.*, 25(1) :35–61, 1983. 8, 11, 12, 20, 33, 35, 62, 63, 88, 100, 104

[65] F. Hermeline. A finite volume method for the approximation of diffusion operators on distorted meshes. *J. Comput. Phys.*, 160(2) :481–499, 2000. 11, 30

[66] F. Hermeline. Approximation of 2-D and 3-D diffusion operators with variable full tensor coefficients on arbitrary meshes. *Comput. Methods Appl. Mech. Engrg.*, 196(21-24) :2497–2526, 2007. 11, 30, 58

[67] T.Y. Hou and P.G. LeFloch. Why nonconservative schemes converge to wrong solutions : error analysis. *Math. Comp.*, 62(206) :497–530, 1994. 12, 70, 71, 72

[68] X.Y. Hu, N.A. Adams, and C.W. Shu. Positivity-preserving method for high-order conservative schemes solving compressible euler equations. *Journal of Computational Physics*, 242(0) :169 – 180, 2013. 64

[69] M.E. Hubbard. Multidimensional slope limiters for muscl-type finite volume schemes on unstructured grids. *J. Comput. Phys.*, 155(1) :54–74, 1999. 10, 30

[70] S. Jin and Z.P. Xin. The relaxation schemes for systems of conservation laws in arbitrary space dimensions. *Comm. Pure Appl. Math.*, 48(3) :235–276, 1995. 118, 119

[71] R. Käppeli and S. Mishra. Well-balanced schemes for the euler equations with gravitation. Technical report, Seminar für Angewandte Mathematik Eidgenössische Technische Hochschule, Février 2013. 10, 168

[72] B. Keen and S. Karni. A second order kinetic scheme for gas dynamics on arbitrary grids. *J. Comput. Phys.*, 205(1) :108–130, 2005. 8, 30, 63

[73] B. Khobalatte and B. Perthame. Maximum principle on the entropy and second-order kinetic schemes. *Math. Comp.*, 62(205) :119–131, 1994. 8, 63

[74] K. Kitamura and E. Shima. Simple and parameter-free second slope limiter for unstructured grid aerodynamic simulations. *AIAA journal*, 50(6) :1415–1426, 2012. 10, 30, 56

[75] A. Kurganov and E. Tadmor. Solution of two-dimensional Riemann problems for gas dynamics without Riemann problem solvers. *Numer. Methods Partial Differential Equations*, 18(5) :584–608, 2002. 45, 49, 52

[76] P.D. Lax. Shock waves and entropy. In *Contributions to nonlinear functional analysis (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1971)*, pages 603–634. Academic Press, New York, 1971. 7, 29, 62

[77] P.D. Lax. *Hyperbolic systems of conservation laws and the mathematical theory of shock waves*. Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1973. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 11. 7, 29, 62

[78] P.D. Lax and B. Wendroff. Systems of conservation laws. *Comm. Pure Appl. Math.*, 13 :217–237, 1960. 64, 66, 173

[79] P.G. LeFloch and M.D. Thanh. A godunov-type method for the shallow water equations with discontinuous topography in the resonant regime. *J. Comput. Phys.*, 230(20) :7631–7660, 2011. 89

[80] R.J. LeVeque. *Finite volume methods for hyperbolic problems*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2002. 7, 8, 9, 15, 30, 61, 62, 63, 64, 66, 67, 71, 173

[81] W. Li, Y.X. Ren, G. Lei, and H. Luo. The multi-dimensional limiters for solving hyperbolic conservation laws on unstructured grids. *J. Comput. Phys.*, 230(21) :7775–7795, 2011. 10, 30

[82] Q. Liang and F. Marche. Numerical resolution of well-balanced shallow water equations with complex source terms. *Advances in water resources*, 32(6) :873–884, 2009. 10, 30

[83] X.D. Liu. A maximum principle satisfying modification of triangle based adapative stencils for the solution of scalar hyperbolic conservation laws. *SIAM J. Numer. Anal.*, 30(3) :701–716, 1993. 10, 30

[84] K. Michalak and C. Ollivier-Gooch. Limiters for unstructured higher-order accurate solutions of the euler equations. In *Proceedings of the AIAA forty-sixth aerospace sciences meeting*, 2008. 10, 30, 56

[85] S. Noelle, N. Pankratz, G. Puppo, and J.R. Natvig. Well-balanced finite volume schemes of arbitrary order of accuracy for shallow water flows. *J. Comput. Phys.*, 213(2) :474–499, 2006. 10, 87, 151

[86] B. Perthame. Second-order Boltzmann schemes for compressible Euler equations in one and two space dimensions. *SIAM J. Numer. Anal.*, 29(1) :1–19, 1992. 8, 10, 30, 63

[87] B. Perthame and Y. Qiu. A variant of Van Leer's method for multidimensional systems of conservation laws. *J. Comput. Phys.*, 112(2) :370–381, 1994. 9, 10, 11, 25, 30, 36, 37

[88] B. Perthame and C.W. Shu. On positivity preserving finite volume schemes for Euler equations. *Numer. Math.*, 73(1) :119–130, 1996. 8, 9, 10, 11, 25, 30, 31, 36, 63

[89] P. Ripa. Conservation laws for primitive equations models with inhomogeneous layers. *Geophys. Astrophys. Fluid Dynam.*, 70(1-4) :85–111, 1993. 88

[90] P. Ripa. On improving a one-layer ocean model with thermodynamics. *J. Fluid Mech.*, 303 :169–201, 1995. 88

[91] P.L. Roe. Approximate riemann solvers, parameter vectors, and difference schemes. *J. Comput. Phys.*, 43(2) :357–372, 1981. 8, 20, 63

[92] A.J.C. Barré de Saint-Venant. Théorie du mouvement non permanent des eaux, avec application aux crues des rivières et à l'introduction de marées dans leurs lits. *Comptes rendus des seances de l'Academie des Sciences*, 36 :174–154, 1871. 88

[93] D. Serre. *Systems of conservation laws. 1*. Cambridge University Press, Cambridge, 1999. Hyperbolicity, entropies, shock waves, Translated from the 1996 French original by I. N. Sneddon. 7, 61, 62

[94] J. Shi, Y.T. Zhang, and C.W. Shu. Resolution of high order WENO schemes for complicated flow structures. *J. Comput. Phys.*, 186(2) :690–696, 2003. 55, 56

[95] C.W. Shu. Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws. In *Advanced numerical approximation of nonlinear hyperbolic equations (Cetraro, 1997)*, volume 1697 of *Lecture Notes in Math.*, pages 325–432. Springer, Berlin, 1998. 46, 66

[96] C.W. Shu and S. Osher. Efficient implementation of essentially nonoscillatory shock-capturing schemes. *J. Comput. Phys.*, 77(2) :439–471, 1988. 67, 68

[97] C.W. Shu and S. Osher. Efficient implementation of essentially nonoscillatory shock-capturing schemes. II. *J. Comput. Phys.*, 83(1) :32–78, 1989. 67, 68

[98] I. Suliciu. On the thermodynamics of rate-type fluids and phase transitions. i. rate-type fluids. *Internat. J. Engrg. Sci.*, 36(9) :921–947, 1998. 122

[99] E. Tadmor. A minimum entropy principle in the gas dynamics equations. *Appl. Numer. Math.*, 2(3-5) :211–219, 1986. 62, 63

[100] E.F. Toro. *Riemann solvers and numerical methods for fluid dynamics*. Springer-Verlag, Berlin, third edition, 2009. A practical introduction. 7, 8, 9, 30, 62, 71, 78, 80, 83

[101] E.F. Toro, M. Spruce, and W. Speares. Restoration of the contact surface in the hll-riemann solver. *Shock waves*, 4(1) :25–34, 1994. 8, 62, 71, 78, 83

[102] B. van Leer. Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method [J. Comput. Phys. **32** (1979), no. 1, 101–136]. *J. Comput. Phys.*, 135(2) :227–248, 1997. With an introduction by Ch. Hirsch, Commemoration of the 30th anniversary {of J. Comput. Phys.}. 8, 11, 25, 30, 63, 66

[103] V. Venkatakrishnan. Convergence to steady state solutions of the euler equations on unstructured grids with limiters. *Journal of Computational Physics*, 118(1) :120–130, 1995. 10, 30

[104] G.B. Whitham. *Linear and nonlinear waves*. Wiley-Interscience [John Wiley & Sons], New York, 1974. Pure and Applied Mathematics. 122

[105] P.R. Woodward. Simulation of the Kelvin-Helmholtz instability of a supersonic slip surface with the piecewise-parabolic method (PPM). In *Numerical methods for the Euler equations of fluid dynamics (Rocquencourt, 1983)*, pages 493–508. SIAM, Philadelphia, PA, 1985. 45, 47

[106] P.R. Woodward and P. Colella. The numerical simulation of two-dimensional fluid flow with strong shocks. *J. Comput. Phys.*, 54(1) :115–173, 1984. 55, 56

[107] S. Yamamoto and H. Daiguji. Higher-order-accurate upwind schemes for solving the compressible Euler and Navier-Stokes equations. *Comput. & Fluids*, 22(2-3) :259–270, 1993. 66, 84

[108] X. Zhang and C.W. Shu. Positivity-preserving high order finite difference WENO schemes for compressible Euler equations. *J. Comput. Phys.*, 231(5) :2245–2258, 2012. 8, 63

[109] Y. Zhang, X. Zhang, and C.W. Shu. Maximum-principle-satisfying second order discontinuous Galerkin schemes for convection-diffusion equations on triangular meshes. *J. Comput. Phys.*, 234 :295–316, 2013. 8, 63

[110] C. Zuily. *Éléments de distributions et d'équations aux dérivées partielles : cours et problèmes résolus*. Dunod, 2002. 130

l'université nantes angers le mans
PÔLE DE RECHERCHE ET D'ENSEIGNEMENT SUPÉRIEUR

UNIVERSITÉ DE NANTES

Thèse de Doctorat

Vivien DESVEAUX

Contribution à l'approximation numérique des systèmes hyperboliques

Contribution to the numerical approximation of hyperbolic systems

Résumé

Dans ce travail, on s'intéresse à plusieurs aspects de l'approximation numérique des systèmes hyperboliques de lois de conservation. La première partie est dédiée à la construction de schémas d'ordre élevé sur des maillages 2D non structurés. On développe une nouvelle technique de reconstruction de gradients basée sur l'écriture de deux schémas MUSCL sur deux maillages imbriqués. Cette procédure augmente le nombre d'inconnues numériques, mais permet d'approcher la solution avec une grande précision. Dans la deuxième partie, on étudie la stabilité des schémas d'ordre élevé. On montre dans un premier temps que les inégalités d'entropie discrètes usuelles vérifiées par les schémas d'ordre élevé ne sont pas pertinentes pour assurer le bon comportement dans le régime de convergence. On propose alors une extension des techniques de limitation *a posteriori* pour forcer la vérification des inégalités d'entropie discrètes requises. Dans la dernière partie, on s'intéresse à la construction de schémas well-balanced pour le modèle de Saint-Venant, le modèle de Ripa et les équations d'Euler avec gravité. On propose plusieurs stratégies permettant d'obtenir des schémas numériques capables de préserver tous les régimes stationnaires au repos. On développe également des extensions d'ordre élevé.

Mots clés

Systèmes hyperboliques, reconstruction MUSCL, robustesse, condition CFL, inégalités d'entropie discrètes, méthode MOOD, schémas well-balanced, méthodes de relaxation, schémas de type Godunov.

Abstract

This work is devoted to several aspects of the numerical approximation of hyperbolic systems of conservation laws.

The first part is dedicated to the derivation of high-order schemes on 2D unstructured meshes. We develop a new technique to reconstruct gradients based on two MUSCL schemes written on two overlapping meshes. This process increases the number of numerical unknowns, but it allows to approximate the solution very accurately.

In the second part, we study the stability of high-order schemes. First, we show that the usual discrete entropy inequalities satisfied by high-order schemes are not relevant to ensure the good behaviour in the convergence regime. Therefore, we propose to extend the *a posteriori* limitation techniques to enforce the scheme to satisfy the required discrete entropy inequalities.

In the last part, we focus on the derivation of well-balanced schemes for the Shallow water equations, the Ripa model and the Euler equations with gravity. We present several strategies leading to numerical schemes able to preserve all the steady states at rest. We also develop high-order extensions.

Key Words

Hyperbolic systems, MUSCL reconstruction, robustness, CFL condition, discrete entropy inequalities, MOOD method, well-balanced schemes, relaxation methods, Godunov-type schemes.