UNIVERSITÉ PARIS-EST

THÈSE

Pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITÉ PARIS-EST

ÉCOLE DOCTORALE MATHEMATIQUES ET SCIENCES ET TECHNOLOGIE DE L'INFORMATION ET DE LA COMMUNICATION

DISCIPLINE : MATHÉMATIQUES

Présentée par

Xavier Lhébrard

Analyse de quelques schémas numériques pour des problèmes de shallow water

---

Directeur de thèse : **François Bouchut**

Soutenue le 27 avril 2015
Devant le jury composé de

M. Christophe Berthon	Université de Nantes	Rapporteur
M. François Bouchut	Univ. Paris-Est Marne-la-Vallée	Directeur de thèse
M. Bruno Després	Univ. Pierre et Marie Curie	Examinateur
M. Robert Eymard	Univ. Paris-Est Marne-La-Vallée	Examinateur
M. Christian Klingenberg	University of Würzburg	Rapporteur
M. Jacques Sainte-Marie	Univ. Pierre et Marie Curie	Examinateur

Thèse préparée au
Laboratoire LAMA CNRS UMR 8050
Université de Paris-Est Marne-la-Vallée
5, boulevard Descartes, Champs-sur-Marne
77454 Marne-la-Vallée cedex 2, France

Remerciements

La tradition veut que je commence par remercier mon directeur de thèse François Bouchut, pour la disponibilité et la patience dont il a fait preuve durant ces trois années. Ses conseils avisés m'auront beaucoup appris sur les mathématiques et sur les exigences du métier d'enseignant-chercheur. Je lui adresse ma plus profonde gratitude pour son aide, ses encouragements et sa bienveillance.

J'exprime ma plus vive reconnaisssance envers Christhophe Berthon et Kristian Klingenberg pour avoir accepté de rapporter ce manuscrit. C'est un grand honneur pour moi qu'ils aient relu mon travail avec tant d'attention. Je remercie également Bruno Després, Robert Eymard et Jacques Sainte-Marie d'avoir montré leur intérêt pour cette thèse en acceptant de faire partie du jury.

Je souhaite remercier tous ceux qui ont participé à mon apprentissage des mathématiques depuis le lycée jusqu'à l'achèvement de cette thèse en passant par les classes préparatoires, en particulier merci à madame Dozias et monsieur Boutemy.

Mes études à l'ENS de Cachan, antenne de Bretagne, m'ont permis de m'épanouir pleinement sur le plan intellectuel grâce à des enseignants de grande qualité. En particulier, merci à Erwan Faou, Mohammed Lemou, Olivier Ley, Florian Méhats, Marc Quicampoix, Nicoletta Tchou, Gregory Vial, Julien Vovelle et San Vu Ngoc. Un remerciement encore plus particulier à Arnaud Debussche, qui m'a fortement encouragé dans le choix de cette thèse.

La thèse, ce n'est pas que des maths ! Je tiens à remercier les doctorants du LAMA avec qui j'ai passé de bons moments, Bertrand, Christelle, Harry, Kirushna, Marie, Marwa, Paolo, Pierre Y, Pierre F, Rania , Sandrine, Sebastian. Merci également à Ludovic, pour nos discussions toujours intéressantes.

Je remercie également mes amis de Rennes. En particulier je pense à Annabelle, Laurent S et Pierre Mi, je garde en mémoire l'année de prépa agreg qu'on a passée ensemble, studieuse mais aussi pleine de moments de rigolades. Merci aussi à mon ami Thomas, pour tous nos délires pendant nos études à Rennes. Merci à mon ami Pierre J, pour son enthouiasme et sa bonne humeur. Merci à Sinziana, pour sa bonne humeur et nos discussions apaisantes dans les moments de doute. Merci à mon ami Laurent P, qui est toujours de bons conseils, concernant les mathématiques et les choix d'orientation.

Merci pour leur soutien à mes amis de Tours, Adrien, Anaïs, Benjamin, Jessica, Jean-Baptiste, Mathieu, Maxime, Mélanie, Nicolas, Perrine, Rémi, Stéphanie.

Merci à mes amis Visanais, que j'ai perdu de vu ces derniers temps mais que je retrouverai avec plaisir prochainement, Alex, François, François, Fanny, Maël, Jérémy, Coralie, Titi, Thibault, Pauline, Anne, Laurent, Lolasse, Pierre Mo, Margaux, Clémence, Marie, Julien, Judith, Oli, Cécile, Manon, Perrine, Ludovic, Hélianthe.
Je tiens à remercier particulièrement mes parents, qui m'encouragent, me soutiennent et sans qui je ne serai pas allé aussi loin. Merci à ma soeur, qui montre son intérêt pour mes travaux et qui est toujours là pour me changer les idées. Merci à ma Clara, qui elle aussi est là pour me changer les idées, m'encourager et s'intéresser à mes travaux. Je remercie également chaleureusement mon parrain et ma marraine qui m'ont beaucoup soutenu pendant mes études à Rennes, je ne les oublie pas.

Résumé

Nous élaborons et analysons mathématiquement des approximations numériques par des méthodes de type volumes finis de solutions faibles de systèmes hyperboliques pour des écoulements géophysiques.
Dans une première partie nous approchons les solutions du système de la magnétohydrodynamique en faible épaisseur avec un fond plat. Nous développons un schéma de type Godunov utilisant un solveur de Riemann approché défini via une méthode de relaxation. Des expressions explicites sont établies pour les vitesses de relaxation, qui permettent d'obtenir un schéma satisfaisant un ensemble de bonnes propriétés de consistance et de stabilité. Il conserve la masse, préserve la positivité de la hauteur de fluide, vérifie une inégalité d'entropie discrète, résout les discontinuités de contact même résonantes, donne des vitesses de propagations contrôlées par les données initiales. Des test numériques sont effectués, validant les résultats théoriques énoncés.
Dans une seconde partie nous approchons les solutions du système de la magnétohydrodynamique en faible épaisseur avec fond variable. Nous développons un schéma équilibre pour certains états stationnaires au repos. Nous utilisons la méthode de reconstruction hydrostatique, avec des états reconstruits pour la hauteur d'eau et les composantes du champ magnétique. Nous trouvons des termes correctifs pour les flux numériques par rapport au cadre habituel, et nous prouvons que le schéma obtenu préserve la positivité de la hauteur d'eau, vérifie une inégalité d'entropie semi-discrète et est consistant. Des test numériques sont effectués, validant les résultats théoriques.
Dans une troisième partie nous établissons la convergence d'un schéma cinétique avec reconstruction hydrostatique pour le système de Saint-Venant avec topographie. De nouvelles estimations sur le gradient des solutions approchées sont obtenues par l'analyse de la dissipation d'énergie. La convergence est obtenue par la méthode de compacité par compensation, sous des hypothèses sur les données initiales et la régularité du fond.

<u>Mots clés :</u> schémas volumes finis, schémas de relaxation, solveur de Riemann approché, shallow water, magnétohydrodynamique, discontinuités de contacts, inégalité d'entropie discrète, schémas équilibre, reconstruction hydrostatique, schéma cinétique, convergence, dissipation.

Abstract

We build and analyze mathematically numerical approximations by finite volume methods of weak solutions to hyperbolic systems for geophysical flows.

In a first part we approximate the solutions of the shallow water magnetohydrodynamics system with flat bottom. We develop a Godunov scheme using an approximate Riemann solver defined via a relaxation method. Explicit formulas are established for the relaxation speeds, that lead to a scheme satisfying good properties of consistency and stability. It preserves mass, positivity of the fluid height, satisfies a discrete entropy inequality, resolves contact discontinuities, and involves propagation speeds controlled by the initial data. Several numerical tests are performed, endorsing the theoretical results.

In a second part we approximate the solutions of the shallow water magnetohydrodynamics system with non-flat bottom. We develop a well-balanced scheme for several steady states at rest. We use the hydrostatic reconstruction method, with reconstructed states for the fluid height and the magnetic field. We get some new corrective terms for the numerical fluxes with respect to the classical framework, and we prove that the obtained scheme preserves the positivity of height, satisfies a semi-discrete entropy inequality, and is consistent. Several numerical tests are presented, endorsing the theoretical results.

In a third part we prove the convergence of a kinetic scheme with hydrostatic reconstruction for the Saint-Venant system with topography. Some new estimates on the gradient of approximate solutions are established, by the analysis of energy dissipation. The convergence is obtained by the compensated compactness method, under some hypotheses concerning the initial data and the regularity of the topography.

Key words : finite volume schemes, relaxation schemes, approximate Riemann solver, shallow water, magnetohydrodynamics, contact discontinuities, discrete entropy inequality, well-balanced schemes, hydrostatic reconstruction, kinetic scheme, convergence, dissipation.

Table des matières

Introduction

0.1 Introduction générale

0.1.1 Modèles de type Saint-Venant

Dans ce travail de thèse on s'intéresse à des écoulements gravitaires à surface libre en couche fine, il est donc intéressant de commencer par présenter le système de Navier-Stokes, qui décrit de manière complète un écoulement gravitaire à surface libre se déplaçant sur un fond avec topographie $z_b(x,t)$. Pour simplifier négligeons les effets de viscosité et de friction, le système de Navier Stokes se réduit alors au système d'Euler. En notant les directions horizontales et verticales x et z respectivement, ce système s'écrit

$$\partial_x u + \partial_z w = 0, \tag{0.1.1}$$
$$\partial_t u + \partial_x \left(u^2\right) + \partial_z (uw) + \partial_x p = \partial_x (2\mu\partial_x u) + \partial_z (\mu (\partial_z u + \partial_x w)), \tag{0.1.2}$$
$$\partial_t w + \partial_x (uw) + \partial_z \left(w^2\right) + \partial_z p = -g + \partial_x (\mu (\partial_z u + \partial_x w)) + \partial_z (2\mu\partial_z w), \tag{0.1.3}$$

ces équations étant définies pour

$$t > t_0, \quad x \in \mathbb{R}, \quad z_b(x,t) \leq z \leq \eta(x,t), \tag{0.1.4}$$

où $\eta(x,t)$ représente l'élévation de la surface libre, $\mathbf{u} = (u,w)^T$ le vecteur vitesse, ρ la densité, g l'accélération gravitationnelle et μ un coefficient de viscosité.

Rappelons deux difficultés pour la résolution de ce système. La première concerne l'utilisation d'un maillage mobile pour discrétiser la surface libre (apparition, disparition de zones sèches et topologie du domaine fluide). La deuxième se situe au niveau de la nature riche du système Navier-Stokes (turbulences, tourbillons, instabilités d'écoulements). On souhaiterait donc utiliser des descriptions qui d'une part soient moins riches mais plus explicites, et d'autre part qui permettraient d'utiliser un maillage fixe. Les systèmes de type Saint-Venant font partie de ces descriptions.

Le système de Saint-Venant, aussi appelé système Shallow Water, a été initialement introduit dans [50]. C'est un système hyperbolique décrivant l'écoulement d'eau dans un canal rectiligne à fond plat en terme de hauteur d'eau $h(t,x)$ et de vitesse moyennée selon la direction verticale $u(t,x)$:

$$\partial_t h + \partial_x(hu) = 0, \tag{0.1.5}$$

$$\partial_t(hu) + \partial_x\left(hu^2 + \frac{1}{2}gh^2\right) = 0, \tag{0.1.6}$$

où g désigne la gravité. La seconde variable conservative $q(t,x) = h(t,x)u(t,x)$ considérée dans ce modèle désigne le débit. Dans sa version bi-dimensionnelle, ce modèle permet de modéliser bon nombre de phénomènes impliquant des fluides géophysiques à surface libre en écoulement "peu profond".

On s'intéresse désormais à des modèles dits " de type Saint-Venant", plus complexes que le modèle original (0.1.5)-(0.1.6), ils sont obtenus à partir du système de Navier-Stokes, dans lequel on peut faire des hypothèses simplificatrices pour des écoulements en couche fine, telles que

- la pression est hydrostatique ou de manière équivalente l'accélération du fluide peut être négligée par rapport aux effects gravitationnels,
- la vitesse horizontale du fluide est approché de manière satisfaisante par sa moyenne verticale.

Concernant les travaux déjà effectués sur la dérivation du modèle, dans [61, 65, 77], le système de Saint-Venant est obtenu à partir d'un développement asymptotique pour des ε petits dans les équations de Navier-Stokes en deux dimensions, où ε représente le rapport de la longueur caractéristique de l'écoulement sur sa hauteur caractéristique. Il est supposé également que la friction au fond et la viscosité sont d'ordre de grandeur ε.

Ajoutons à cela que le système de Saint-Venant a un intérêt au niveau de son cadre mathématique car le système sans friction ni viscosité peut être vu comme un cas particulier du système pour la dynamique des gaz. L'existence des solutions pour ce système a été démontré, voir [75] et les références qui s'y trouvent.

Mentionnons maintenant quelques uns de ces modèles en couche mince.

- Le modèle Ripa introduit dans [53, 87, 88] prend en compte les variations de température et modélisent les courants océaniques. Le point délicat du point de vue approximation numérique est de préserver les états d'équilibre de ce modèle, pour plus de détails, on pourra lire [46, 55, 56] et les références qui y sont citées.
- Des modèles pour des écoulements gravitaire à surface libre de fluides élastiques (fluides non-Newtonien) [26, 35, 76, 79].
- Des modèles de type Savage Hutter [31, 60, 95], où l'on relache les hypothèses sur la pente de l'écoulement par rapport au cas Saint-Venant avec en vue des applications pour les écoulements de matériaux granulaires par exemple des avalanches avec débris.
- Le modèle de Saint-Venant Exner [5, 69], où les équations de Saint-Venant sont couplées avec une loi de conservation de sédiment pour étudier les mouvements de sédiment dans les écoulements fluviaux.
- Le modèle de shallow water magnetohydrodynamic, où l'on applique un régime d'eaux peu profondes aux équations magnétohydrodynamique incompressibles avec gravité, afin d'étudier les mouvements de plasmas en couche fine, c'est un des modèles étudiés dans cette thèse, nous donnons des références sur ce modèle un peu plus loin.
- Un modèle shallow water non-hydrostatique [38], avec des termes dérivées d'ordre élevé, qui se rapproche du modèle Boussinesq et qui permettent de mieux observer les propagation d'ondes longues, ce qui est utile dans l'optique de prévision des vagues de tsunami par exemple.
- Des modèle multi-couches, qui permettent d'enrichir la discrétisation de la vitesse verticale [1, 9, 10, 106],
- Modèle avec une nouvelle approche de dérivation à partir des équations d'Euler [93].

Il y a donc un intérêt persistant autour des systèmes de type Saint-Venant, et les chercheurs voudraient affaiblir voire s'affranchir de certaines hypothèses faites lors de la dérivation à partir des équations de Navier-Stokes. De plus on souhaiterait trouver des modèles sur lesquels on pourrait adapter des méthodes numériques développées précédemment sur d'anciens modèles.

Par conséquent un équilibre est à trouver entre la dérivation de nouveaux modèles plus sophistiqués et la perspective d'une bonne approximation des solutions de ces nouveaux modèles.

Dans ce travail de thèse, seule la partie approximation a été exploré, la dérivation du modèle est soit considérée comme acquise, soit effectuée de manière formelle, la justification rigoureuse reste une question ouverte, pouvant faire l'objet d'un travail ultérieur.

0.1.2 Schémas de type volumes finis pour des lois de conservation

Nous faisons le choix ici d'une discrétisation par des méthodes de type volumes finis, intimement liées au caractère hyperbolique des système de type Saint-Venant et au fait qu'il s'agit - au moins partiellement - de lois de conservation, qui peuvent s'écrire sous la forme suivante :

$$\partial_t U + \partial_x F(U) = 0. \tag{0.1.7}$$

La méthode des volumes finis pour des systèmes de lois de conservation [59, 67, 74, 99] présente le grand intérêt d'être intrinsèquement conservative et s'adapte très bien à l'aspect discontinu des solutions. Il s'agit ici de découper l'espace en cellules puis d'intégrer le système (0.1.7) sur chaque cellule et sur un pas de temps. En une dimension d'espace, un schéma volume fini explicite peut donc s'écrire sous la forme générale suivante

$$U_i^{n+1} - U_i^n + \frac{\Delta t}{\Delta x_i}(F_{i+1/2} - F_{i-1/2}) = 0, \tag{0.1.8}$$

où Δt, Δx_i représentent respectivement le pas de temps et le pas d'espace, U_i^n représente la moyenne de la solution sur la cellule C_i au temps t^n, et où $F_{i+1/2}$ représente le flux numérique à l'interface entre les cellules C_i et C_{i+1}. Ces flux aux interfaces font naturellement intervenir les valeurs des variables aux interfaces, qui ne sont pas connues. Les flux aux interfaces pour un schéma à trois points prennent alors la forme suivante

$$F_{i+1/2} = F(U_i^n, U_{i+1}^n). \tag{0.1.9}$$

Développer une méthode volumes finis consiste donc à définir, à partir des valeurs des variables aux noeuds du maillage, un flux numérique qui soit consistant avec le système continu, et, si possible, préserve ses propriétés de stabilité.

Pour les systèmes homogènes de type Saint-Venant, nous nous concentrons sur deux propriétés de stabilités, conserver la hauteur de fluide positive et vérifier une inégalité d'entropie discrète.

Cette dernière propriété permet à la fois de calculer numériquement des discontinuités admissibles mais également d'avoir une stabilité globale, du fait qu'une quantité mesurant la taille globale des données doit décroître. On cherche donc à obtenir une inégalité de la forme suivante :

$$\eta\left(U_i^{n+1}\right) - \eta\left(U_i^n\right) + \frac{\Delta t}{\Delta x_i}(G_{i+1/2} - G_{i-1/2}) \leq 0 \tag{0.1.10}$$

avec η une entropie du système et $G_{i+1/2}$ un flux d'entropie discret qu'on souhaite être consistant avec le flux d'entropie du système. De plus lorsque l'on établit des propriétés de stabilité rigoureusement, une des difficultés est de ne pas obtenir trop de diffusion numérique, qui supprimerait tout intérêt pratique du schéma.

D'autres difficultés peuvent survenir en fonction de la complexité des équations. Par exemple on souhaite obtenir une condition de Courant (CFL) qui ne soit pas trop restrictive. On cherche également des formules explicites car les méthodes itératives peuvent être coûteuse et induire un temps de calcul trop important. Enfin, pour les systèmes qui ont des termes sources, on veut résoudre des solutions particulières, ce qui permet d'améliorer la précision et la stabilité des schémas obtenus. Les méthodes reconnues satisfaisant ces critères sont les méthodes utilisant un solveur approché (Roe [36, 73, 78, 89]) ou les méthodes avec un système de relaxation (HLL, HLLC, relaxation, Suliciu, cinétique, [24, 45, 49, 70, 72, 82, 100, 101]).

0.1.3 Objectifs numériques pour des écoulements en couche mince

Les objectifs numériques que l'on se fixe afin d'approcher de manière satisfaisante les solutions de systèmes de type Saint-Venant sont

- préserver la positivité de la hauteur de fluide,
- vérifier une inégalité d'entropie discrète, qui est l'analogue d'une inégalité d'énergie dans le cas continu,
- trouver un compromis entre précision et stabilité pour la résolution des discontinuités, on souhaite être optimal au niveau de la diffusion numérique par la résolution de discontinuités de contact,
- gérer des problèmes additionnels de consistance dans le cas d'un système non-conservatif,
- vérifier une propriété d'équilibre entre termes différentiels et termes sources dans le cas d'un système non-homogène.

Nous allons maintenant introduire un modèle de type Saint-Venant, le modèle shallow water magnetohydrodynamic, puis décrire la démarche qui nous a permis d'obtenir un schéma remplissant les objectifs numériques énoncés précédemment.

0.2 Solveur approché multi-ondes par relaxation pour shallow water MHD

0.2.1 Modèle utilisé et état de l'art

On considère les équations shallow water magnetohydrostatique, que l'on notera shallow water MHD ou SWMHD par la suite. Dans le cas où l'état du système dépend d'une variable spatiale x en une dimension et du temps t, ce système s'écrit

$$\partial_t h + \partial_x(hu) = 0, \tag{0.2.1}$$

$$\partial_t(hu) + \partial_x(hu^2 + P) = 0, \tag{0.2.2}$$

$$\partial_t(hv) + \partial_x(huv + P_\perp) = 0, \tag{0.2.3}$$

$$\partial_t(ha) + u\partial_x(ha) = 0, \tag{0.2.4}$$

$$\partial_t(hb) + \partial_x(hbu - hva) + v\partial_x(ha) = 0, \tag{0.2.5}$$

avec

$$P = g\frac{h^2}{2} - ha^2, \quad P_\perp = -hab. \tag{0.2.6}$$

On dispose également d'une inégalité d'énergie

$$\begin{aligned} &\partial_t(\frac{1}{2}h(u^2+v^2) + \frac{1}{2}gh^2 + \frac{1}{2}h(a^2+b^2)) \\ &+\partial_x([\frac{1}{2}h(u^2+v^2) + gh^2 + \frac{1}{2}h(a^2+b^2)]u - ha(au+bv)]) \le 0. \end{aligned} \tag{0.2.7}$$

Les équations SWMHD sont dérivées à partir partir des équations de magnétohydrodynamique incompressible avec gravité [107], à l'aide de deux hypothèses :

- approximation uniforme dans la direction verticale, qui permet remplacer les variables de vitesses et de champs magnétiques par leurs valeurs moyennes entre le fond et la surface libre,

- l'hypothèse de pression magnétohydrostatique, qui consiste à considérer que l'accélération verticale est faible, ce qui permet d'avoir une relation d'équilibre sur le terme de pression totale.

Le modèle shallow water MHD ([51, 52, 66]) a été introduit afin de modéliser une couche fine de l'atmosphère du soleil, appelée tachocline. Cette couche du soleil est positionnée entre la couche interne, à rotation solide, du soleil et la couche externe, à rotation convective. L'étude de cette dernière est un enjeu pour mieux comprendre le mécanisme de dynamo du soleil ainsi que les éruptions solaires.
Concernant les travaux effectués sur ce modèle, quelques tests numériques ont été effectués, via une méthode par ondes de propagation [90], qui apparaît comme une méthode précise mais qui ne permet pas d'étudier la stabilité du schéma, qui risque donc de perdre la positivité de la couche de fluide ou de résoudre des chocs non admissibles. D'autres tests ont été effectués via une méthode cinétique [85], ce type de méthode étant reconnu comme stable mais diffusive, manquant précision dans le calcul de certaines discontinuités. A la connaissance de l'auteur, aucun schéma numérique prenant en compte de conserver la positivité du fluide, de résoudre exactement les contacts et de prendre en compte la topographie n'avait été étudié. C'est une des contributions nouvelle de ce travail de thèse.

0.2.2 Solveur de Rieman approché

Une méthode maintenant classique pour approcher les solutions de (0.1.7) est de suivre l'approche de Godunov ([68, 70]). On considère des approximations constantes par morceaux de la variable U, que l'on note U_i^n et en invoquant un solveur de Riemann aux interfaces $R(x/t, U_i^n, U_{i+1}^n)$ entre deux mailles.
Notre travail s'inscrit dans la continuité des travaux effectués sur les solveurs approchés de Riemann, qui ont d'abord été construits pour des systèmes de taille deux [70], puis trois [58, 100], et maintenant étendu à des systèmes de tailles supérieurs([3, 24, 26, 27]). Nous rappelons qu'une des difficultés pour ces solveurs est la résolution exacte des discontinuités de contact. En effet ces dernières ne bénéficient pas de phénomène de compression et nécessite d'être résolues exactement. Cela a été rendue possible grâce à l'utilisation de solveurs avec des structures multi-ondes. Il est également possible de modifier certaines de ces méthodes pour qu'elles résolvent en plus des chocs stationnaires [33, 42].
Parmis ces méthodes on distingue celles utilisant un système de relaxation, par exemple le système Suliciu pour le système Euler qui a permis d'être étendu à d'autres systèmes ([13, 24, 26, 27, 47, 96, 97]). Ces méthodes sont en fait des cas particuliers de méthodes de relaxation qui une structure simple et permettent une analyse des conditions d'entropicité des schémas

associés ([22, 24, 45, 48, 72]). Ce choix de discrétisation sera privilégié par l'auteur car il permet à la fois une résolution exacte des contact et de faire une analyse d'entropicité qui conduit à des schémas qui préserve la positivité de la hauteur de fluide et vérifie une inégalité d'entropie discrète.

0.2.3 Système SWMHD, système SWMHD relaxation, principaux résultats

Ici on applique la méthode Sulicu pour Euler, on dérive à partir de solutions régulières du système (0.2.1)-(0.2.5) les équations sur P, $P_\perp$ (0.2.6) suivantes :

$$\begin{aligned} \partial_t(hP) + \partial_x(hPu) + h^2\left(a^2 + gh\right)\partial_x u = 0, \\ \partial_t(hP_\perp) + \partial_x(hP_\perp u) + (ha)^2\partial_x v = 0, \end{aligned}$$

Ensuite, on introduit de nouvelles variables π et $\pi_\perp$, et deux paramètres c^2 et c_a^2

$$\begin{aligned} \partial_t(h\pi) + \partial_x(h\pi u) + c^2\partial_x u = 0, \\ \partial_t(h\pi_\perp) + \partial_x(h\pi_\perp u) + c_a^2\partial_x v = 0, \end{aligned}$$

On obtient le système de relaxation suivant :

$$\partial_t h + \partial_x(hu) = 0, \tag{0.2.8}$$
$$\partial_t(hu) + \partial_x(hu^2 + \pi) = 0, \tag{0.2.9}$$
$$\partial_t(hv) + \partial_x(huv + \pi_\perp)) = 0, \tag{0.2.10}$$
$$\partial_t(ha) + u\partial_x(ha) = 0, \tag{0.2.11}$$
$$\partial_t(hb) + \partial_x(hbu - hav) + v\partial_x(ha) = 0, \tag{0.2.12}$$
$$\partial_t(h\pi + \partial_x(h\pi u) + c^2\partial_x u = 0, \tag{0.2.13}$$
$$\partial_t(h\pi_\perp) + \partial_x(h\pi_\perp u) + c_a^2\partial_x v = 0. \tag{0.2.14}$$

Un solveur approché de Riemann peut être alors défini à partir des valeurs gauche et droite à une interface de la variable

$$U = (h, hu, hv, ha, hb), \tag{0.2.15}$$

de la manière suivante :

- on résout (0.2.8)-(0.2.14) avec les valeurs initiales de pression relaxée π et $\pi_\perp$ suivantes :

$$\begin{aligned} \pi_l &= (gh^2/2 + ha^2)_l &&= P(h_l, (ha)_l), \\ \pi_r &= (gh^2/2 + ha^2)_r &&= P(h_r, (ha)_r), \\ (\pi_\perp)_l &= (-hab)_l &&= P_\perp(h_l, (ha)_l, (hb)_l) \\ (\pi_\perp)_r &= (-hab)_r &&= P_\perp(h_r, (ha)_r, (hb)_r) \end{aligned} \tag{0.2.16}$$

  et des valeurs pour $c_{a,l}$, c_l, $c_{a,r}$ and c_r que nous préciserons par la suite.

- on ne retient de cette solution que les variables h, hu, hv, ha, hb. Le résultat est un solveur approché de Riemann.

Cette approche par relaxation nous permet d'obtenir les propriétés suivantes :

- toutes les valeurs du système de relaxation (0.2.8)-(0.2.14) sont linéairement dégénérées. Cette propriété permet d'obtenir un solveur simple, i.e. constant par morceaux en la variable $\xi = x/t$. Le système SWMHD sans relaxation était consitué de 2 valeurs propres extrèmes non-linéaires et de 3 valeurs propres internes linéairement dégénérés. La résolution de son problème de Riemann aurait entrainé un solveur impliquant des chocs, des détentes et des contacts, ce qui aurait été bien plus difficile à résoudre, voire impossible d'obtenir des formules implémentables.

- la stabilité (entropicité, positivité de la hauteur de fluide) et l'optimalité de la diffusion numérique (résolution des contacts), grâce à l'ajustement des paramètres c et c_a.

En effet, on obtient également à partir des solutions régulières de (0.2.8)-(0.2.14) et en combinant ces différentes équations, l' équation conservative de type entropie suivante :

$$\partial_t \left(h \left[\frac{u^2}{2} + \frac{v^2}{2} + \frac{\pi^2}{2c^2} + \frac{\pi_\perp^2}{2c_a^2} + \phi\left(\frac{1}{h} + \frac{\pi}{c^2}, ha, b - \frac{ha\pi_\perp}{c_a^2} \right) \right] \right) \tag{0.2.17}$$

$$+\partial_x \left(h \left[\frac{u^2}{2} + \frac{v^2}{2} + \frac{\pi^2}{2c^2} + \frac{\pi_\perp^2}{2c_a^2} + \phi\left(\frac{1}{h} + \frac{\pi}{c^2}, ha, b - \frac{ha\pi_\perp}{c_a^2} \right) \right] u \right) \tag{0.2.18}$$

$$+\partial_x (\pi u + \pi_\perp v) = 0 \tag{0.2.19}$$

Cette dernière équation permet d'analyser finement les conditions d'entropicité du schéma et de trouver les valeurs explicites suivantes :

$$c_{a,l} = h_| a|_l, \quad c_{a,r} = h_r |a|_r, \tag{0.2.20}$$

$$\begin{aligned} c_l &= h_l s_l + \frac{3}{2} h_l \left((u_l - u_r)_+ + \frac{(\pi_r - \pi_l)_+}{h_l s_l + h_r s_r} \right), \\ c_r &= h_r s_r + \frac{3}{2} h_r \left((u_l - u_r)_+ + \frac{(\pi_l - \pi_r)_+}{h_l s_l + h_r s_r} \right), \end{aligned} \tag{0.2.21}$$

avec

$$s = \sqrt{gh + a^2}. \tag{0.2.22}$$

Ces valeurs ont permis de démontrer le résultat suivant :

Theorem 0.1. *Le choix de vitesse de relaxation* (0.2.20)-(0.2.21) *définit un solveur de Riemann qui vérifie une inégalité d'entropie discrète, conserve la positivité de la hauteur de fluide, résout exactement les discontinuités de contact, implique une vitesse de propagation bornée en fonction des données initiales.*

Ce résultat consistitue une des contributions novatrices de cette thèse en ce qui concerne l'approximation des solutions du système Shallow Water MHD.

0.3 Schéma équilibre pour shallow water MHD avec topographie

0.3.1 Schéma équilibres

Afin de décrire l'approwimation du système SWMHD avec topographie, considérons un système de lois de conservation non homogène avec source

$$\partial_t U + \partial_x F(U) = -B(U)\partial_x Z, \tag{0.3.1}$$

Pour ce type de système, il est reconnu la majorité des applications, le régime d'écoulement intéressant est le régimes stationnaire, qu'on peut écrire de la manière suivante :

$$\partial_x F(U) = -B(U)\partial_x Z. \tag{0.3.2}$$

On souhaite donc trouver des méthodes numériques qui conservent ces états d'équilibre. On peut faire un parrallèle entre ce problème et le problème de résolution des contacts que l'on a rencontré pour les systèmes homogènes. En effet en considérent le système (0.3.1) sous forme une quasi-linéaire on s'aperçoit que les solutions stationnaires que l'on souhaite conserver sont

en fait des contacts associés à la valeur propre 0.
Une méthode reconnu pour satisfaire cette propriété est la reconstruction hydrostatique ([6, 24]). Cette méthode suppose que l'on ait approché le système sans terme source au moyen d'une méthode de volumes finis avec un certain flux numérique, notons le $F_{HOMOGENE}(U_l, U_r)$ pour la suite. Ensuite l'idée est d'introduire des états reconstruits $U_l^\#, U_r^\#$ qui satisfont une propriété d'équilibre. Enfin on peut définir un nouveau flux numérique $F_{SOURCE}(U_l, U_r)$ de la manière suivante :

$$F_{SOURCE}(U_l, U_r) = F_{HOMOGENE}(U_l^\#, U_r^\#) + C(U_l, U_r, U_l^\#, U_r^\#). \tag{0.3.3}$$

où $C(U_l, U_r, U_l^\#, U_r^\#)$ est un terme de correction à déterminer, afin d'assurer la consistance et la stabilité de la méthode.

0.3.2 Traitement de la topographie pour le système shallow water MHD

Dans le cas d'un système non-homogène, nous avons vu précédemment qu'un des objectifs étaient de résoudre certaines solutions stationnaires ou de manière équivalente des contacts associés à la valeur propre 0. Pour le système SWMHD, la particularité est qu'il y a quatre valeurs propres auxquelles sont associées des discontinutés de contact :

$$u, \quad u - |a|, \quad u + |a|, \quad 0,$$

qualifiés respectivement de matériel, d'Aflven gauche, d'Alfven droite et de contact associé au fond. Les cas qui nous intéressent sont les cas de résonnances avec le contact au fond, i.e. les cas où différentes valeurs propres deviennent égales à 0. Parmi les différents cas possibles nous nous intéressons aux équilibres au repos suivant :

- le cas d'une résonance matériel et Alfven ($u = a = 0$), la relation d'équilibre est alors :

$$h + z = cst, \quad u = 0, \quad a = 0, \tag{0.3.4}$$

- le cas d'une résonance matériel ($u = 0$ and $a \neq 0$), la relation d'équilibre est alors :

$$\begin{aligned} u = 0, \quad v = cst, \quad h + z = cst, \\ \sqrt{h}\; a = cst, \quad \sqrt{h}\; b = cst. \end{aligned} \tag{0.3.5}$$

En suivant l'approche de reconstruction hydrostatique à partir de la variable $U = (h, hu, hv, ha, hb)$ nous recontruisons les variables $h^\#, a^\#, b^\#$ de la manière suivante :

$$h_l^\# = (h_l - (\Delta z)_+)_+, \quad h_r^\# = (h_r - (-\Delta z)_+)_+, \tag{0.3.6}$$

avec $x_+ \equiv \max(0, x)$. On définit également

$$a_l^\# = \kappa_l a_l, \quad a_r^\# = \kappa_r a_r, \tag{0.3.7}$$

$$b_l^\# = \kappa_l b_l, \quad b_r^\# = \kappa_r b_r, \tag{0.3.8}$$

avec

$$\kappa_l = \min\left(\sqrt{\frac{h_l}{h_l^\#}}, \gamma\right), \kappa_r = \min\left(\sqrt{\frac{h_r}{h_r^\#}}, \gamma\right), \gamma \geq 1. \tag{0.3.9}$$

Nous avons introduit un paramètre seuil γ afin de résoudre les contacts du type (0.3.5), tout en évitant une explosion du coefficient κ lorsque $h \to 0$.

Une des contributions novatrices de cette thèse concernant le traitement du terme source pour les équations SWMHD est le résultat suivant :

Theorem 0.2. *Les valeurs reconstruites $h^\#, a^\#, b^\#$ (0.3.6)-(0.3.9) permettent de définir un flux numérique du type (0.3.3) avec un terme corecteur explicite. Le schéma volumes finis qui en résulte est consistant, vérifie une inégalité d'entropie semi-discrète, conserve la positivité de la hauteur de fluide, et est un schéma d'équilibre pour les relations (0.3.4), (0.3.5).*

0.4 Système SV, preuve CV d'un schéma cinétique avec HR

0.4.1 Interprétation cinétique de Saint Venant

Dans cette partie on s'intéresse à l'approximation du sytème Saint Venant

$$\begin{aligned} \partial_t h + \partial_x(hu^2) &= 0, \\ \partial_t(hu) + \partial_x(hu^2 + \frac{g}{2}h^2) &= -gh\partial_x z, \end{aligned} \tag{0.4.1}$$

$h(t,x) \geq 0$, $u(t,x) \in \mathbb{R}$, $g > 0$ est la constante de gravité, et la topographie $z(x)$ est donnée. Le système est complété avec une inégalité d'entropie

$$\partial_t \left(h\frac{u^2}{2} + g\frac{h^2}{2} \right) + \partial_x \left(\left(h\frac{u^2}{2} + gh^2 \right) u \right) \leq 0. \tag{0.4.2}$$

0.4.2 Schéma cinétique avec reconstruction hydrostatique

On considère le schéma suivant

$$U_i^{n+1} = U_i^n - \frac{\Delta t}{\Delta x} \left(F_{i+1/2-} - F_{i-1/2+} \right), \tag{0.4.3}$$

avec

$$F_{i+1/2-} = \mathcal{F}(U_{i+1/2-}, U_{i+1/2+}) - S_{i+1/2-}, \tag{0.4.4}$$

$$F_{i-1/2+} = \mathcal{F}(U_{i-1/2-}, U_{i-1/2+}) + S_{i-1/2+}. \tag{0.4.5}$$

Le flux numérique est constitué d'un terme $\mathcal{F}$ qui est le flux numérique pour le système sans topographie. Les seconds termes $S_{i+1/2-}$, $S_{i-1/2+}$ sont des termes sources défnis par

$$S_{i+1/2-} = \begin{pmatrix} 0 \\ g\frac{h_{i+1/2-}^2}{2} - g\frac{h_i^2}{2} \end{pmatrix}, \quad S_{i-1/2+} = \begin{pmatrix} 0 \\ g\frac{h_i^2}{2} - g\frac{h_{i-1/2+}^2}{2} \end{pmatrix}. \tag{0.4.6}$$

Les états reconstruits

$$U_{i+1/2-} = (h_{i+1/2-}, h_{i+1/2-}u_i), \quad U_{i+1/2+} = (h_{i+1/2+}, h_{i+1/2+}u_i) \tag{0.4.7}$$

sont définis par

$$h_{i+1/2-} = (h_i + z_i - z_{i+1/2})_+, \quad h_{i+1/2+} = (h_{i+1} + z_{i+1} - z_{i+1/2})_+ \tag{0.4.8}$$

et

$$z_{i+1/2} = \max(z_i, z_{i+1}). \tag{0.4.9}$$

Pour le terme homogène, on utilisera un flux numérique cinétique $\mathcal{F}$, introdduit dans [83] et définit par

$$\mathcal{F}(U_l, U_r) = F^+(U_l) + F^-(U_r), \tag{0.4.10}$$

$$F^+(U) = \int_{\mathbb{R}} \xi \mathbb{1}_{\xi>0} \begin{pmatrix} 1 \\ \xi \end{pmatrix} M(U,\xi)d\xi, \qquad (0.4.11)$$
$$F^-(U) = \int_{\mathbb{R}} \xi \mathbb{1}_{\xi<0} \begin{pmatrix} 1 \\ \xi \end{pmatrix} M(U,\xi)d\xi,$$

$$M(U) = \frac{1}{g\pi}\left(2gh - (\xi-u)^2\right)_+^{1/2}, \quad U = (h, hu). \qquad (0.4.12)$$

0.4.3 Résultat de convergence

Theorem 0.3. *Soit* $(U_i^n) = (h_i^n, h_i^n u_i^n)$ *les valeurs du schéma* (0.4.3)-(0.4.12). *On suppose l'existence de* $h, m, h_M, u_M > 0$ *tels que*

$$h_m \le h_i^n \le h_M, \quad 0 \le |u_i^n| \le u_M. \qquad (0.4.13)$$

On interpole les valeurs du schéma (U_i^n) *sur chaque maille et entre deux instants* t_n *et* t_{n+1} *et on note* U_Δ *la solution approchée de* (0.4.1) *ainsi obtenue.*
Alors il existe U *une solution faible de* (0.4.1) *telle que* $U_\Delta \underset{\Delta t, \Delta x \to 0}{\to} U$ *pour presque tout* $(t,x) \in [0,T]\times\mathbb{R}$ *et* $\forall t \in [0,T], U_\Delta(t,.) \underset{\Delta t, \Delta x \to 0}{\to} U(t,.)$ *dans* $L^\infty_{x,w*}(\mathbb{R})$.

Ce résultat est établi sous les hypothèses d'hauteur h non nul, de topo graphie z bornée, lipschitz et son approximation (z_i^n) vérifie la relation $|\Delta z_{i+1/2}| < \min(h_i, h_{i+1})$. On suppose également une condition cfl inverse du type $\dfrac{\Delta x}{\Delta t} < \beta < 1$, $\beta > 0$.

0.4.4 Difficultés et étapes de démonstration

On utilise un résultat classique de compacité par compensation ([57, 75, 98]). On veut donc montrer que

$$\partial_t \eta(U_\Delta) + \partial_x G(U_\Delta) \qquad (0.4.14)$$

est compact dans H^{-1}_{loc} pour tout couple (η, G) d'entropie/ flux d'entropie, ce qui permet de conclure à la compacité de la suite U_Δ dans L^p_{loc} .
L'ingrédient principal pour obtenir (0.4.14) est d'obtenir des estimations à priori sur $\partial_x U_\Delta$ et $\partial_t U_\Delta$, qui l'étape principale dans la méthode de Di Perna. Au cours de ce travail de thèse, nous

avons pu trouver les estimations suivantes :

$$\left(\int_0^T \int_{x_{i_0-1/2}}^{x_{i_1+1/2}} |\partial_x U_\Delta|^2 dxdt\right)^{1/2} \leq \frac{C_1}{\sqrt{\Delta x}}, \tag{0.4.15}$$

$$\left(\int_0^T \int_{x_{i_0-1/2}}^{x_{i_1+1/2}} |\partial_t U_\Delta|^2 dxdt\right)^{1/2} \leq \frac{C_2}{\sqrt{\Delta x}}. \tag{0.4.16}$$

Citons le travail [15], qui concerne la convergence au niveau continu de solutions de modèles BGK vers des solutions du système de dynamique des gaz. En conséquence il a été établi des conditions suffisantes de convergence de schéma à flux séparés pour des systèmes hyperboliques sans terme source[14]. Par conséquent une première idée pour la démonstration de notre résultat de convergence a été de vérifier si les conditions suffisantes données dans [14] sont vérifiés pour le schéma (0.4.3)-(0.4.12) en l'absence de topographie. En particulier, une des conditions suffisantes pour ces estimations était d'avoir F^+ ou $-F^-$ (voir (0.4.11)) η-dissipative. Cette propriété n'est pas vérifié pour (0.4.11), par manque de dissipation du schéma cinétique.
Néanmoins, un des résultats novateurs et difficile de ce travail de thèse est d'avoir établi que $F^+ - F^-$ est η-dissipative, ce qui correspond à l'estimation suivante :

$$\begin{aligned} &\int_{\mathbb{R}} |\xi| \left(H(M_2) - H(M_1) - \eta'(U_1) \begin{pmatrix} 1 \\ \xi \end{pmatrix} (M_2 - M_1) \right) d\xi \\ &\geq \alpha \left(\eta(U_2) - \eta(U_1) - \eta'(U_1)\left(U_2 - U_1\right)\right). \end{aligned} \tag{0.4.17}$$

Ce résultat n'est néanmoins valide que sur un ouvert borné convexe $\mathcal{U} \subset \mathbb{R}^2$, excluant les valeurs nulles pour la hauteur, et la preuve d'existence de la constante est non constructive.
Un autre ingrédient important de la preuve est l'inégalité d'entropie discrète établie dans [7], qui s'écrit avec des termes d'erreurs d'ordre Δx^2, sous l'hypothèse d'une topographie de régularité Lipschitz.
Donnons ici pour simplifier une version de cette inégalité lorsque $z = cst$,

$$\begin{aligned} \eta(U_i^{n+1}) \leq \eta(U_i^n) &- \frac{\Delta t}{\Delta x}\left(\widetilde{G}_{i+1/2} - \widetilde{G}_{i-1/2}\right) \\ &- \nu_\beta \frac{\Delta t}{\Delta x}\int_{\mathbb{R}} |\xi| \frac{g^2\pi^2}{6}\Big(\mathbb{1}_{\xi<0}\left(M_{i+1/2+} + M_{i+1/2-}\right)\left(M_{i+1/2+} - M_{i+1/2-}\right)^2 \\ &+ \mathbb{1}_{\xi>0}\left(M_{i-1/2+} + M_{i-1/2-}\right)\left(M_{i-1/2+} - M_{i-1/2-}\right)^2\Big) d\xi, \end{aligned} \tag{0.4.18}$$

où le terme intégré en ξ a le bon signe, renforce l'inégalité d'entropie et permet d'avoir une estimation a priori. Dans le cas où z n'est pas constant, on a des termes additionnels à gérer, mais l'esprit de la démonstration est inchangé.
Après multiplication par Δx et sommation sur i de (0.4.18), suivit de quelques manipulations techniques, on fait apparaître (0.4.17) et on montre (0.4.15), (0.4.16).

Chapter 1

A 5-wave relaxation solver for the shallow water MHD system

Abstract

The shallow water magnetohydrodynamic system describes the thin layer evolution of the solar tachocline. It is obtained from the three dimensional incompressible magnetohydrodynamic system similarly as the classical shallow water system is obtained from the incompressible Navier-Stokes equations. The system is hyperbolic and has two additional waves with respect to the shallow water system, the Alfven waves. These are linearly degenerate, and thus do not generate dissipation. In the present work we introduce a 5-wave approximate Riemann solver for the shallow water magnetohydrodynamic system, that has the property to be non dissipative on Alfven waves. It is obtained by solving a relaxation system of Suliciu type, and is similar to HLLC type solvers. The solver is positive and entropy satisfying, ensuring its robustness. It has sharp wave speeds, and does not involve any iterative procedure.

Keywords: Shallow water magnetohydrodynamics, approximate Riemann solver, relaxation, contact discontinuities, entropy inequality

Mathematics Subject Classification: 76W05, 76M12, 35L65

1.1 Introduction

The shallow water magnetohydrodynamic (SWMHD) system has been introduced in [66] to describe the thin layer evolution of the solar tachocline. It is written in 2d in the tangent plane approximation as

$$\partial_t h + \nabla \cdot (h\mathbf{u}) = 0, \tag{1.1.1}$$

$$\partial_t (h\mathbf{u}) + \nabla \cdot (h\mathbf{u} \otimes \mathbf{u} - h\mathbf{b} \otimes \mathbf{b}) + \nabla(gh^2/2) = 0, \tag{1.1.2}$$

$$\partial_t (h\mathbf{b}) + \nabla \cdot (h\mathbf{b} \otimes \mathbf{u} - h\mathbf{u} \otimes \mathbf{b}) + \mathbf{u}\nabla \cdot (h\mathbf{b}) = 0, \tag{1.1.3}$$

where $g > 0$ is the gravity constant, $h \geq 0$ is the thickness of the fluid, $\mathbf{u} = (u, v)$ is the velocity, $\mathbf{b} = (a, b)$ is the magnetic field, and the notation $\nabla \cdot (\mathbf{b} \otimes \mathbf{u})$ is for the vector with index i given by $\sum_j \partial_j (b_i u_j)$. The system should be complemented with Coriolis force and topography, but these sources will not be considered in this paper.

The system (1.1.1)-(1.1.3) is endowed with an entropy (energy) inequality

$$\begin{aligned} &\partial_t \Big(\frac{1}{2} h|\mathbf{u}|^2 + \frac{1}{2} gh^2 + \frac{1}{2} h|\mathbf{b}|^2 \Big) \\ &+\nabla \cdot \Big(\big(\frac{1}{2} h|\mathbf{u}|^2 + gh^2 + \frac{1}{2} h|\mathbf{b}|^2 \big)\mathbf{u} - h\mathbf{b}(\mathbf{b} \cdot \mathbf{u}) \Big) \leq 0, \end{aligned} \tag{1.1.4}$$

that becomes an equality in the absence of shocks.

The induction equation (1.1.3) implies, by taking its divergence, that

$$\partial_t \Big(\nabla \cdot (h\mathbf{b}) \Big) + \nabla \cdot \Big(\mathbf{u}\nabla \cdot (h\mathbf{b}) \Big) = 0, \tag{1.1.5}$$

meaning that $\nabla \cdot (h\mathbf{b})$ is transported at velocity $\mathbf{u}$. In particular, $\nabla \cdot (h\mathbf{b})$ remains identically zero if it vanishes initially. This situation $\nabla \cdot (h\mathbf{b}) = 0$ (that cancels the last term in (1.1.3)) is indeed the physically relevant one, but for numerical purposes it is convenient to relax this constraint and consider general data. The particular form (1.1.3) has been introduced in [52] for the SWMHD, and in [71] for the full MHD system. It enables to use one-dimensional solvers in two dimensions, indeed this is why the term $\mathbf{u}\nabla \cdot (h\mathbf{b})$ has been added in (1.1.3).

Multidimensional simulations of the SWMHD system have been performed in [91, 92]. The system is closely related to the MHD system, to which many works have been devoted. An important issue in multidimensional simulations is to minimize the numerical viscosity by using accurate solvers, in particular on contact discontinuities; while being robust, see for example [12, 62, 103].

If dependency is only in one spatial variable x, the system simplifies to

$$\partial_t h + \partial_x (hu) = 0, \tag{1.1.6}$$

$$\partial_t (hu) + \partial_x (hu^2 + P) = 0, \tag{1.1.7}$$

$$\partial_t (hv) + \partial_x (huv + P_\perp) = 0, \tag{1.1.8}$$

$$\partial_t (ha) + u\partial_x (ha) = 0, \tag{1.1.9}$$

$$\partial_t (hb) + \partial_x (hbu - hav) + v\partial_x (ha) = 0, \tag{1.1.10}$$

with

$$P = g\frac{h^2}{2} - ha^2, \quad P_\perp = -hab, \tag{1.1.11}$$

and the energy inequality (1.1.4) becomes

$$\begin{aligned} &\partial_t\Big(\frac{1}{2}h(u^2+v^2) + \frac{1}{2}gh^2 + \frac{1}{2}h(a^2+b^2)\Big) \\ &+\partial_x\Big(\Big(\tfrac{1}{2}h(u^2+v^2) + gh^2 + \tfrac{1}{2}h(a^2+b^2)\Big)u - ha(au+bv)\Big) \le 0. \end{aligned} \tag{1.1.12}$$

The eigenvalues of the system (1.1.6)-(1.1.10) are $u, u \pm |a|, u \pm \sqrt{a^2+gh}$. The associated waves are called respectively material (or divergence) waves, Alfven waves and magnetogravity waves, see [51, 107]. Some of these waves will have the same speed when a or h vanishes, hence the system is nonstrictly hyperbolic.

The system has three types of contact discontinuities corresponding to linearly degenerate eigenvalues: the material contacts associated to the eigenvalue u, the left Alfven contacts associated to $u - |a|$, and the right Alfven contacts associated to $u + |a|$. The jump relations associated to these contact discontinuities are as follows. Across a material contact, the quantities $u, v, g\frac{h^2}{2} - ha^2, hab$ are constant. Across an Alfven contact, the quantities h, u, a are constant, and moreover for a left Alfven contact $b \operatorname{sgn} a - v$ is constant, while for a right Alfven contact $b \operatorname{sgn} a + v$ is constant.

The system (1.1.6)-(1.1.10) is nonconservative in the variables ha, hb. However, ha jumps only through the material contacts, where u and v are continuous. Therefore, there is indeed no ambiguity in the non conservative products $u\partial_x(ha)$ and $v\partial_x(ha)$, that are well-defined.

A finite volume scheme for the quasilinear system (1.1.6)-(1.1.10) can be classically built following Godunov's approach, considering piecewise constant approximations of

$$U = (h, hu, hv, ha, hb), \tag{1.1.13}$$

and invoking an approximate Riemann solver at the interface between two cells, see for example [70] or [24, Section 2.3]. A difficulty is however that the system is not conservative. The energy is nevertheless obviously convex with respect to U. The SWMHD system is closely related to the compressible MHD system, for which several entropy schemes are known [21, 28, 54, 63]. In this paper we apply the relaxation approach of [22, 24, 27, 48] to the SWMHD system, in order to get an approximate Riemann solver that is entropy satisfying, ensuring robustness, while being exact on isolated Alfven contacts. The relaxation system is of Suliciu type as introduced in [97], and the approximate Riemann solver belongs to the family of HLLC solvers, as in [12, 17, 18, 24, 62, 70, 94].

The paper is organized as follows. In Section 1.2 we describe the relaxation approximate solver and its entropy property. In Section 1.3 we derive explicit optimal choices of the speeds that enable to obtain stability and accuracy. In Section 1.4 we state our main theorem giving the properties of our relaxation approximate Riemann solver. Finally, in Section **??** we perform numerical tests.

1.2 Approximate Riemann solver

1.2.1 Relaxation approach

In order to get an approximate Riemann solver for (1.1.6)-(1.1.10), we use a standard relaxation approach, used for example in [24] for the Euler equations, in [28] for the MHD equations and in [25] for shallow elastic fluids. An abstract general description can be found in [48], and related works are [42, 44]. The approach enables to naturally handle the energy inequality (1.1.12), and also preserves the positivity of density. Its structure has however to be well-chosen in order to resolve exactly isolated Alfven contacts.

We introduce new variables $\pi, \pi_\perp$, the relaxed pressures, and c_a, c intended to parametrize the speeds. The form of the relaxation system is as follows,

$$\partial_t h + \partial_x(hu) = 0, \tag{1.2.1}$$

$$\partial_t(hu) + \partial_x(hu^2 + \pi) = 0, \tag{1.2.2}$$

$$\partial_t(hv) + \partial_x(huv + \pi_\perp) = 0, \tag{1.2.3}$$

$$\partial_t(ha) + u\partial_x(ha) = 0, \tag{1.2.4}$$

$$\partial_t(hb) + \partial_x(hbu - hav) + v\partial_x(ha) = 0, \tag{1.2.5}$$

$$\partial_t(h\pi) + \partial_x(h\pi u) + c^2\partial_x u = 0, \tag{1.2.6}$$

$$\partial_t(h\pi_\perp) + \partial_x(h\pi_\perp u) + c_a^2\partial_x v = 0, \tag{1.2.7}$$

$$\partial_t c + u\partial_x c = 0, \tag{1.2.8}$$

$$\partial_t c_a + u\partial_x c_a = 0. \tag{1.2.9}$$

The approximate Riemann solver can be defined as follows, starting from left and right values U_l, U_r at an interface.

- Solve the Riemann problem for (1.2.1)-(1.2.9) with initial data obtained by completing U_l, U_r by the equilibrium relations

$$\begin{array}{lclcl} \pi_l & = & P_l & \equiv & (gh^2/2 - ha^2)_l, \\ \pi_r & = & P_r & \equiv & (gh^2/2 - ha^2)_r, \\ (\pi_\perp)_l & = & (P_\perp)_l & \equiv & (-hab)_l, \\ (\pi_\perp)_r & = & (P_\perp)_r & \equiv & (-hab)_r, \end{array} \tag{1.2.10}$$

  and with suitable positive values of c_l, c_r, $c_{a,l}$, $c_{a,r}$ that will be discussed further on, essentially in Section 1.3.

- Retain in the solution only the variables h, hu, hv, ha, hb. The result is a vector called $R(x/t, U_l, U_r)$.

We can remark that the relaxation system (1.2.1)-(1.2.9) is identical to the 5-wave relaxation system in [28, equations (5.5)-(5.7) with $b = 0$], with the identification $B_x = ha$, $B_\perp = hb$. However, the initialization (1.2.10) differs from that of the MHD equations, [28, equation (2.10)]. Indeed, the homogeneity in magnetic terms is different in the SWMHD and MHD systems.

Intuitively, the solver is consistent because of the equations (1.2.1)-(1.2.5), that are consistent with (1.1.6)-(1.1.10). The specific values used for c, c_a do not play any role in this consistency. However, if we require the solver to have the highest accuracy, i.e. to be "tangent" to the original system, one has to take the speeds $c > c_a > 0$ as approximations of $h\sqrt{a^2+gh}$ and $h|a|$ respectively, in the limit when U_l, U_r are close to a common value U. This is because, as can be checked with straightforward computations, smooth solutions to (1.1.6)-(1.1.10) verify

$$\begin{aligned}\partial_t(hP) + \partial_x(hPu) + h^2(a^2+gh)\partial_x u = 0,\\ \partial_t(hP_\perp) + \partial_x(hP_\perp u) + h^2a^2\partial_x v = 0,\end{aligned} \tag{1.2.11}$$

that have to be compared with (1.2.6), (1.2.7). The accuracy of the solver on isolated contacts is described by the following lemma.

Lemma 1.1. *The approximate Riemann solver $R(x/t, U_l, U_r)$ solves exactly:*

(i) material contact discontinuities,

(ii) left Alfven contact discontinuities under the condition $c_{a,l} = (h|a|)_l$,

(iii) right Alfven contact discontinuities under the condition $c_{a,r} = (h|a|)_r$.

Proof. Material contacts are solutions to the SWMHD system (1.1.6)-(1.1.10) with u, v, P, $P_\perp$ constant. These solutions are obviously solutions to the relaxation system (1.2.1)-(1.2.9) with $\pi = P$, $\pi_\perp = P_\perp$. Thus for these data, R coincides with the exact solver, which proves (i).

Alfven contacts are solutions to (1.1.6)-(1.1.10) with h, u, a, $b\,\mathrm{sgn}\,a - v$ (for a left contact), $b\,\mathrm{sgn}\,a + v$ (for a right contact) constant. As previously, it is enough to prove that these solutions, completed with $\pi = P$, $\pi_\perp = P_\perp$, are solutions to the relaxation system (1.2.1)-(1.2.9). One can see that only (1.2.7) is not immediately satisfied. Comparing (1.2.7) to the second line of (1.2.11), we get the condition $c_a^2 = h^2a^2$ where v jumps. Note that according to (1.2.4) and (1.2.9), ha and c_a are both continuous through the Alfven waves (assuming $a \neq 0$). This yields (ii) and (iii). □

Following the Godunov approach, the numerical scheme can be defined by the approximate Riemann solver as follows. We consider a mesh of cells $(x_{i-1/2}, x_{i+1/2})$, $i \in \mathbb{Z}$, of length $\Delta x_i = x_{i+1/2} - x_{i-1/2}$, discrete times t_n with $t_{n+1} - t_n = \Delta t$, and cell values U_i^n approximating the average of U over the cell i at time t_n. We can then define an approximate solution $U_{appr}(t,x)$ for $t_n \le t < t_{n+1}$ and $x \in \mathbb{R}$ by

$$U_{appr}(t,x) = R\Big(\frac{x - x_{i+1/2}}{t - t_n}, U_i^n, U_{i+1}^n\Big) \text{ for } x_i < x < x_{i+1}, \tag{1.2.12}$$

where $x_i = (x_{i-1/2} + x_{i+1/2})/2$. This definition is coherent under a half CFL condition, formulated as

$$\begin{aligned} x/t &< -\frac{\Delta x_i}{2\Delta t} \Rightarrow R(x/t, U_i, U_{i+1}) = U_i, \\ x/t &> \frac{\Delta x_{i+1}}{2\Delta t} \Rightarrow R(x/t, U_i, U_{i+1}) = U_{i+1}. \end{aligned} \tag{1.2.13}$$

The new values at time t_{n+1} are defined by

$$U_i^{n+1} = \frac{1}{\Delta x_i} \int_{x_{i-1/2}}^{x_{i+1/2}} U_{appr}(t_{n+1} - 0, x)dx. \tag{1.2.14}$$

Notice that it is only in this averaging procedure that the choice of the particular pseudo-conservative variable U as (1.1.13) is involved. We can follow the computations of [24, Section 2.3], the only difference being that the system is not conservative. We obtain the update formula

$$U_i^{n+1} = U_i^n - \frac{\Delta t}{\Delta x_i}(\mathcal{F}_l(U_i^n, U_{i+1}^n) - \mathcal{F}_r(U_{i-1}^n, U_i^n)), \tag{1.2.15}$$

where

$$\begin{aligned} \mathcal{F}_l(U_l, U_r) &= F(U_l) - \int_{-\infty}^{0} (R(\xi, U_l, U_r) - U_l)d\xi, \\ \mathcal{F}_r(U_l, U_r) &= F(U_r) + \int_{0}^{\infty} (R(\xi, U_l, U_r) - U_r)\, d\xi, \end{aligned} \tag{1.2.16}$$

the variable ξ stands for x/t, and the pseudo-conservative flux is chosen as

$$F(U) \equiv (hu, hu^2 + P, huv + P_\perp, 0, hbu - hav). \tag{1.2.17}$$

In (1.2.17), the two last components could be chosen differently since the two magnetic equations in our system are not conservative. We can remark that the choice of F has no influence on the update formula (1.2.15).

1.2.2 Energy inequality

Here we do not use the entropy extension defined in [24, Definition 2.14], because the minimization principle is a bit too restrictive. We instead follow the strategy used in [25, 27].

We define the left and right numerical energy fluxes as

$$\begin{aligned}\mathcal{G}_l(U_l,U_r) &= G(U_l) - \int_{-\infty}^{0}\Big(E\big(R(\xi,U_l,U_r)\big) - E(U_l)\Big)\,d\xi,\\ \mathcal{G}_r(U_l,U_r) &= G(U_r) + \int_{0}^{\infty}\Big(E\big(R(\xi,U_l,U_r)\big) - E(U_r)\Big)\,d\xi,\end{aligned} \tag{1.2.18}$$

where E and G are respectively the energy and the energy flux from (1.1.12),

$$\begin{aligned}E(U) &= \frac{1}{2}h(u^2+v^2) + \frac{1}{2}gh^2 + \frac{1}{2}h(a^2+b^2),\\ G(U) &= \Big(\frac{1}{2}h(u^2+v^2) + gh^2 + \frac{1}{2}h(a^2+b^2)\Big)u - ha(au+bv).\end{aligned} \tag{1.2.19}$$

Following [24, Section 2.3], a sufficient condition for the scheme to be energy satisfying is that

$$\mathcal{G}_r(U_l,U_r) - \mathcal{G}_l(U_l,U_r) \le 0. \tag{1.2.20}$$

When this is satisfied, because of the convexity of E with respect to U and of the CFL condition (1.2.13), one has the discrete energy inequality

$$E(U_i^{n+1}) - E(U_i^n) + \frac{\Delta t}{\Delta x_i}\Big(\mathcal{G}(U_i^n,U_{i+1}^n) - \mathcal{G}(U_{i-1}^n,U_i^n)\Big) \le 0, \tag{1.2.21}$$

where the numerical energy flux $\mathcal{G}(U_l,U_r)$ is any function satisfying $\mathcal{G}_r(U_l,U_r) \le \mathcal{G}(U_l,U_r) \le \mathcal{G}_l(U_l,U_r)$. In order to analyze the condition (1.2.20), let us introduce the sum of the gravitational potential energy and the magnetic energy

$$e = \frac{1}{2}gh + \frac{1}{2}(a^2+b^2), \tag{1.2.22}$$

that enables to rewrite the energy as

$$E = \frac{1}{2}h(u^2+v^2) + he. \tag{1.2.23}$$

Then, while solving the relaxation system (1.2.1)-(1.2.9), we solve simultaneously the equation for a new variable $\hat{e}$,

$$\partial_t(\hat{e} - \pi^2/2c^2 - \pi_\perp^2/2c_a^2) + u\partial_x(\hat{e} - \pi^2/2c^2 - \pi_\perp^2/2c_a^2) = 0, \tag{1.2.24}$$

where $\hat{e}$ has left and right initial data $e_l = e(U_l)$, $e_r = e(U_r)$. The reason for writing (1.2.24) is that combined with (1.2.1)-(1.2.9) it implies

$$\partial_t\left(\frac{1}{2}h(u^2+v^2)+h\hat{e}\right)+\partial_x\left(\left(\frac{1}{2}h(u^2+v^2)+h\hat{e}\right)u+\pi u+\pi_\perp v\right)=0. \tag{1.2.25}$$

Indeed, (1.2.25) can be obtained as follows. From (1.2.2), (1.2.3), (1.2.6), (1.2.7) (combined with (1.2.1)), we get

$$\begin{aligned}\partial_t u+u\partial_x u+\frac{1}{h}\partial_x\pi&=0,\\ \partial_t v+u\partial_x v+\frac{1}{h}\partial_x\pi_\perp&=0,\\ \partial_t\pi+u\partial_x\pi+\frac{c^2}{h}\partial_x u&=0,\\ \partial_t\pi_\perp+u\partial_x\pi_\perp+\frac{c_a^2}{h}\partial_x v&=0.\end{aligned} \tag{1.2.26}$$

Multiplying these equations respectively by u, v, π, $\pi_\perp$, we obtain

$$\begin{aligned}\partial_t u^2/2+u\partial_x u^2/2+\frac{u}{h}\partial_x\pi&=0,\\ \partial_t v^2/2+u\partial_x v^2/2+\frac{v}{h}\partial_x\pi_\perp&=0,\\ \partial_t\pi^2/2+u\partial_x\pi^2/2+\frac{\pi c^2}{h}\partial_x u&=0,\\ \partial_t\pi_\perp^2/2+u\partial_x\pi_\perp^2/2+\frac{\pi_\perp c_a^2}{h}\partial_x v&=0.\end{aligned} \tag{1.2.27}$$

Using the advection equations (1.2.8), (1.2.9) for c and c_a, the two last equations of (1.2.27) give

$$\begin{aligned}\partial_t\frac{\pi^2}{2c^2}+u\partial_x\frac{\pi^2}{2c^2}+\frac{\pi}{h}\partial_x u&=0,\\ \partial_t\frac{\pi_\perp^2}{2c_a^2}+u\partial_x\frac{\pi_\perp^2}{2c_a^2}+\frac{\pi_\perp}{h}\partial_x v&=0.\end{aligned} \tag{1.2.28}$$

Adding up the two first equations of (1.2.27) and (1.2.28) yields

$$\partial_t\left(\frac{u^2+v^2}{2}+\frac{\pi^2}{2c^2}+\frac{\pi_\perp^2}{2c_a^2}\right)+u\partial_x\left(\frac{u^2+v^2}{2}+\frac{\pi^2}{2c^2}+\frac{\pi_\perp^2}{2c_a^2}\right)+\frac{1}{h}\partial_x(\pi u+\pi_\perp v)=0. \tag{1.2.29}$$

Then, adding this to (1.2.24) and using (1.2.1) finally gives our stated identity (1.2.25).

Using the value of the Riemann solution to the relaxation system at $x/t=0$, we define

$$\mathcal{G}(U_l,U_r)=\left(\left(\frac{1}{2}h(u^2+v^2)+h\hat{e}\right)u+\pi u+\pi_\perp v\right)_{x/t=0}. \tag{1.2.30}$$

Lemma 1.2. *If for all values of x/t the solution to (1.2.1)-(1.2.9) has nonnegative height h and satisfies*

$$\hat{e} \geq e(U), \tag{1.2.31}$$

where here $U = R(x/t, U_l, U_r)$, $e(U)$ is defined in (1.2.22) and $\hat{e}$ is defined by (1.2.24), then $\mathcal{G}_r(U_l, U_r) \leq \mathcal{G}(U_l, U_r) \leq \mathcal{G}_l(U_l, U_r)$, and the discrete energy inequality (1.2.21) holds under the CFL condition (1.2.13).

Proof. Since (1.2.25) is a conservative equation, and its conserved quantity and flux reduce to E and G on the left and right states, integrating it over rectangles one gets

$$\begin{aligned} \mathcal{G}(U_l, U_r) \quad &= G(U_l) - \int_{-\infty}^{0} \Big(\big(\frac{1}{2} h(u^2 + v^2) + h\hat{e} \big)(\xi) - E(U_l) \Big)\, d\xi \\ &= G(U_r) + \int_{0}^{\infty} \Big(\big(\frac{1}{2} h(u^2 + v^2) + h\hat{e} \big)(\xi) - E(U_r) \Big)\, d\xi. \end{aligned} \tag{1.2.32}$$

Therefore, comparing to (1.2.18), we see that in order to get $\mathcal{G}_r \leq \mathcal{G} \leq \mathcal{G}_l$ it is enough that for all ξ

$$E(R(\xi, U_l, U_r)) \leq \big(\frac{1}{2} h(u^2 + v^2) + h\hat{e} \big)(\xi), \tag{1.2.33}$$

which is equivalent to (1.2.31). □

1.2.3 Intermediate states

In this subsection we describe the solution to the Riemann problem for (1.2.1)-(1.2.9) with initial data completed by the relations (1.2.10). The analysis is similar to that in [27, 28] for the full MHD system, and to [25] for shallow elastic fluids.

The quasilinear system (1.2.1)-(1.2.9) has the property of having a quasi diagonal form

$$\partial_t(\pi + cu) + (u + c/h)\partial_x(\pi + cu) - \frac{u}{h}c\partial_x c = 0, \tag{1.2.34}$$

$$\partial_t(\pi - cu) + (u - c/h)\partial_x(\pi - cu) - \frac{u}{h}c\partial_x c = 0, \tag{1.2.35}$$

$$\partial_t(1/h + \pi/c^2) + u\partial_x(1/h + \pi/c^2) = 0, \tag{1.2.36}$$

$$\partial_t(\pi_\perp + c_a v) + (u + c_a/h)\partial_x(\pi_\perp + c_a v) - \frac{v}{h}c_a\partial_x c_a = 0, \tag{1.2.37}$$

$$\partial_t(\pi_\perp - c_a v) + (u - c_a/h)\partial_x(\pi_\perp - c_a v) - \frac{v}{h}c_a\partial_x c_a = 0, \tag{1.2.38}$$

$$\partial_t(b + \frac{ha}{c_a^2}\pi_\perp) + u\partial_x(b + \frac{ha}{c_a^2}\pi_\perp) = 0, \tag{1.2.39}$$

$$\partial_t(ha) + u\partial_x(ha) = 0, \tag{1.2.40}$$

$$\partial_t c + u\partial_x c = 0, \tag{1.2.41}$$

$$\partial_t c_a + u\partial_x c_a = 0. \tag{1.2.42}$$

One deduces its eigenvalues, which are

$$u - \frac{c}{h}, \quad u - \frac{c_a}{h}, \quad u, \quad u + \frac{c_a}{h}, \quad u + \frac{c}{h}, \tag{1.2.43}$$

the central eigenvalue u having multiplicity 5 and the other being simple. From the above form one also checks easily that the system is hyperbolic, with all eigenvalues linearly degenerate. As a consequence, Rankine-Hugoniot conditions are well defined (the weak Riemann invariants do not jump through the associated discontinuity), and are equivalent to any conservative formulation. In the solution to the Riemann problem, the speeds corresponding to the previous eigenvalues will be denoted by

$$\Sigma_1 < \Sigma_2 < \Sigma_3 < \Sigma_4 < \Sigma_5, \tag{1.2.44}$$

the speed Σ_3 corresponding to the eigenvalue u. Thus we get a 5-wave solver with four intermediate states. The variables take the values "l" for $x/t < \Sigma_1$, "l**" for $\Sigma_1 < x/t < \Sigma_2$, "l*" for $\Sigma_2 < x/t < \Sigma_3$, "r*" for $\Sigma_3 < x/t < \Sigma_4$, "r**" for $\Sigma_4 < x/t < \Sigma_5$, "r" for $\Sigma_5 < x/t$, see Figure 1.1. There are 5 strong Riemann invariants associated to the central wave (i.e. quantities that lie in the kernel of $\partial_t + u\partial_x$), which are

$$c, \quad c_a, \quad ha, \quad \frac{1}{h} + \frac{\pi}{c^2}, \quad b + \frac{ha}{c_a^2}\pi_\perp. \tag{1.2.45}$$

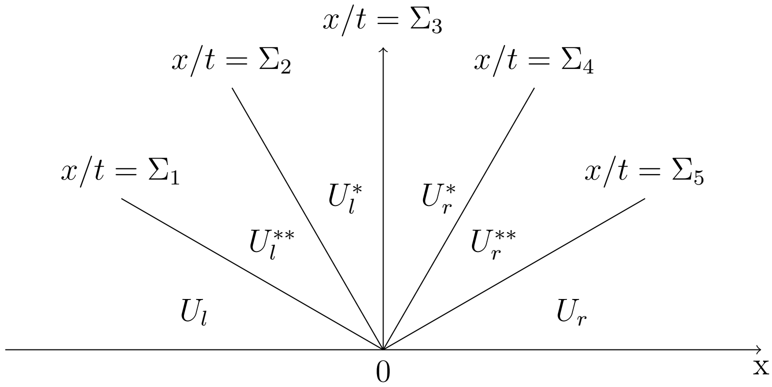


Figure 1.1 – Intermediate states in the Riemann solution

These quantities are thus weak Riemann invariants for the other waves. Four weak Riemann invariants for the central waves are u, v, π, $\pi_\perp$. They take the same value on the left and on the right of this central wave, we shall denote these values by u^*, v^*, π^*, $\pi_\perp^*$. The remaining weak Riemann invariants for the left and right waves are found to be

$$\begin{array}{llll} u - c/h : & \pi + cu, & \pi_\perp, & v, \\ u - c_a/h : & \pi_\perp + c_a v, & \pi, & u, \\ u + c_a/h : & \pi_\perp - c_a v, & \pi, & u, \\ u + c/h : & \pi - cu, & \pi_\perp, & v. \end{array} \tag{1.2.46}$$

We notice that the equations (1.2.1), (1.2.2), (1.2.6), (1.2.8) form a closed system of equations. Therefore, the variables h, u, π, c can be resolved independently of the knowledge of v, $\pi_\perp$, a, b, c_a, and in particular they do not jump through the Σ_2 and Σ_4 waves. This means that for these unknowns h, u, π, c, the "l**" and "l*" states are identical, as well as the "r**" and "r*" states. In particular, u and π take the constant values u^* and π^* on the whole fan $\Sigma_1 < x/t < \Sigma_5$,

$$(u,\pi)_l^{**} = (u,\pi)_l^* = (u,\pi)^*, \quad (u,\pi)_r^{**} = (u,\pi)_r^* = (u,\pi)^*. \tag{1.2.47}$$

The second velocity-pressure set of variables v, $\pi_\perp$ jump only though the c_a waves Σ_2, Σ_4, thus

$$\begin{gathered} (v,\pi_\perp)_l^{**} = (v,\pi_\perp)_l, \quad (v,\pi_\perp)_r^{**} = (v,\pi_\perp)_r, \\ (v,\pi_\perp)_l^* = (v,\pi_\perp)_r^* = (v,\pi_\perp)^*. \end{gathered} \tag{1.2.48}$$

Then, because of (1.2.45), the variables a, c_a do not jump through the Σ_2 and Σ_4 waves, as h, u, π, c. Moreover, b does not jump through the Σ_1 and Σ_5 waves, as $v, \pi_\perp$, see Figure 1.2.

In addition, computations using the Riemann invariants (1.2.45), (1.2.46) give the values as in

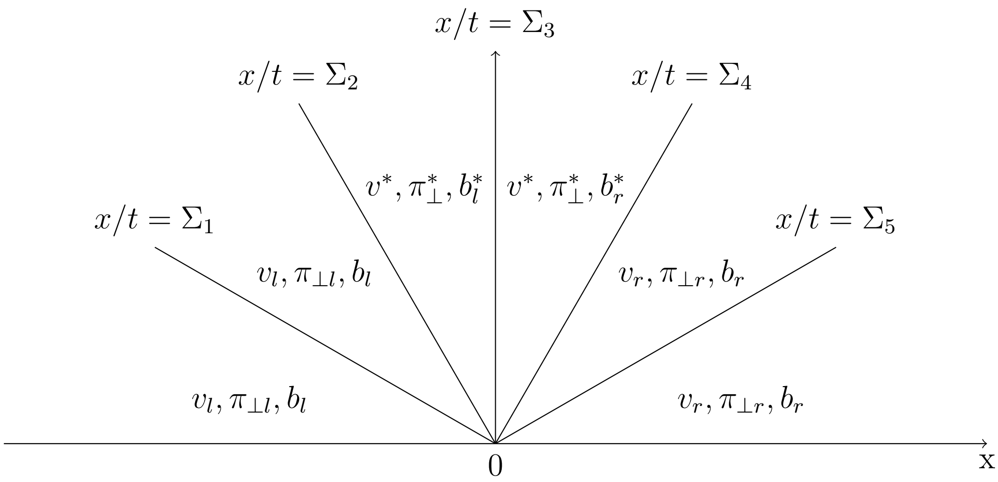


Figure 1.2 – Intermediate states for the variable v, $\pi_\perp$, b

[24, 28]

$$\begin{aligned}
u^* &= \frac{c_l u_l + c_r u_r + \pi_l - \pi_r}{c_l + c_r}, \\
\pi^* &= \frac{c_r \pi_l + c_l \pi_r - c_l c_r (u_r - u_l)}{c_l + c_r}, \\
v^* &= \frac{c_{al} v_l + c_{ar} v_r + \pi_{\perp l} - \pi_{\perp r}}{c_{al} + c_{ar}}, \\
\pi_\perp^* &= \frac{c_{ar} \pi_{\perp l} + c_{al} \pi_{\perp r} - c_{al} c_{ar} (v_r - v_l)}{c_{al} + c_{ar}},
\end{aligned} \tag{1.2.49}$$

and

$$\begin{aligned}
\frac{1}{h_l^*} &= \frac{1}{h_l} + \frac{c_r(u_r - u_l) + \pi_l - \pi_r}{c_l(c_l + c_r)}, \\
\frac{1}{h_r^*} &= \frac{1}{h_r} + \frac{c_l(u_r - u_l) + \pi_r - \pi_l}{c_r(c_l + c_r)}.
\end{aligned} \tag{1.2.50}$$

Next, using the invariants in (1.2.45) that involve a, b, we get

$$a_l^* = a_l \frac{h_l}{h_l^*}, \quad a_r^* = a_r \frac{h_r}{h_r^*}, \tag{1.2.51}$$

and

$$\begin{aligned}
b_l^* &= b_l + \frac{h_l a_l}{c_{al}(c_{al} + c_{ar})} \Big(\pi_{\perp l} - \pi_{\perp r} + c_{ar}(v_r - v_l) \Big), \\
b_r^* &= b_r + \frac{h_r a_r}{c_{ar}(c_{al} + c_{ar})} \Big(\pi_{\perp r} - \pi_{\perp l} + c_{al}(v_r - v_l) \Big).
\end{aligned} \tag{1.2.52}$$

Finally, using the previous formulas one can compute the speeds,

$$\begin{gathered}\Sigma_1 = u_l - c_l/h_l, \quad \Sigma_2 = u^* - c_{al}/h_l^*, \quad \Sigma_3 = u^*, \\ \Sigma_4 = u^* + c_{ar}/h_r^*, \quad \Sigma_5 = u_r + c_r/h_r.\end{gathered} \tag{1.2.53}$$

1.3 Analysis of the energy inequality and choice of the speeds

1.3.1 Sufficient stability conditions for a fixed intermediate state

In this subsection, we derive sufficient conditions for the inequality (1.2.31) in Lemma 1.2 to hold, for a fixed intermediate state

$$U^* = (h^*, h^*u^*, h^*v^*, h^*a^*, h^*b^*), \tag{1.3.1}$$

where the star '*' stands for any of "l**", "l*", "r*", "r**". These values are completed with those of the relaxation variables π^*, $\pi_\perp^*$, c, c_a, and with $\hat{e}^*$ resulting from (1.2.24). Note that the notation '*' differs here slightly from the one in the previous paragraph, in particular, v^* and $\pi_\perp^*$ do not coincide with the values in (1.2.49) in case of the "l**" and "r**" states. It is convenient to denote by the subscript 'l/r' any data evaluated on the initial state on the same side of the central wave as the intermediate state considered. The short notation c, c_a will mean that these quantities are evaluated locally, i.e. at the '*' state, or equivalently at the 'l/r' location since these quantities are strong invariants for the central wave.

In order to analyze (1.2.31), we use the same strategy as in [24, Lemma 2.20], and [27, Lemma 3.2], that consists in using a decomposition of $e(U^*) - \hat{e}^*$ in elementary entropy dissipation terms along each waves. This idea was introduced in [22] in the case of constant speeds.

Lemma 1.3. *With the preceding notations, we have the identity*

$$\begin{aligned} e(U^*) - \hat{e}^* = {} & D_{--}\Big(U^*, u^* - \frac{\pi^*}{c}\Big) + D_-\Big(U^*, v^* - \frac{\pi_\perp^*}{c_a}\Big) \\ & + D_0\Big(U^*, U_{l/r}\Big) + D_+\Big(U^*, v^* + \frac{\pi_\perp^*}{c_a}\Big) + D_{++}\Big(U^*, u^* + \frac{\pi^*}{c}\Big), \end{aligned} \tag{1.3.2}$$

where we set for any state U and any scalar Λ, with P and $P_\perp$ defined in (1.1.11),

$$D_{--}(U,\Lambda) = \frac{1}{4}(u-\frac{P}{c})^2 - \frac{1}{4}\Lambda^2 - (-P,u)\cdot\left(\frac{1}{2c}(u-\frac{P}{c}-\Lambda), \frac{1}{2}(u-\frac{P}{c}-\Lambda)\right), \tag{1.3.3}$$

$$D_{++}(U,\Lambda) = \frac{1}{4}(u+\frac{P}{c})^2 - \frac{1}{4}\Lambda^2 - (-P,u)\cdot\left(-\frac{1}{2c}(u+\frac{P}{c}-\Lambda), \frac{1}{2}(u+\frac{P}{c}-\Lambda)\right), \tag{1.3.4}$$

$$D_{-}(U,\Lambda) = \frac{1}{4}(v-\frac{P_\perp}{c_a})^2 - \frac{1}{4}\Lambda^2 - (b,v)\cdot\left(\frac{ha}{2c_a}(v-\frac{P_\perp}{c_a}-\Lambda), \frac{1}{2}(v-\frac{P_\perp}{c_a}-\Lambda)\right), \tag{1.3.5}$$

$$D_{+}(U,\Lambda) = \frac{1}{4}(v+\frac{P_\perp}{c_a})^2 - \frac{1}{4}\Lambda^2 - (b,v)\cdot\left(-\frac{ha}{2c_a}(v+\frac{P_\perp}{c_a}-\Lambda), \frac{1}{2}(v+\frac{P_\perp}{c_a}-\Lambda)\right), \tag{1.3.6}$$

$$D_0(U_1,U_2) = e(U_1) - \frac{P_1^2}{2c^2} - \frac{P_{1\perp}^2}{2c_a^2} - e(U_2) + \frac{P_2^2}{2c^2} + \frac{P_{2\perp}^2}{2c_a^2} - (-P_1,b_1)\cdot\begin{pmatrix} 1/h_1 + P_1/c^2 - 1/h_2 - P_2/c^2 \\ b_1 + \frac{h_1a_1}{c_a^2}P_{1\perp} - b_2 - \frac{h_2a_2}{c_a^2}P_{2\perp}\end{pmatrix}. \tag{1.3.7}$$

Proof. We have to sum up the contributions in the right-hand side of (1.3.2). We look first at terms that are in factor of $P^* \equiv P(U^*)$,

$$\frac{1}{2c}\left(u^* - \frac{P^*}{c} - u^* + \frac{\pi^*}{c}\right) \quad \text{in } D_{--}, \tag{1.3.8}$$

$$\frac{-1}{2c}\left(u^* + \frac{P^*}{c} - u^* - \frac{\pi^*}{c}\right) \quad \text{in } D_{++}, \tag{1.3.9}$$

$$\left(\frac{1}{h^*} + \frac{P^*}{c^2}\right) - \left(\frac{1}{h_{l/r}} + \frac{P_{l/r}}{c^2}\right) \quad \text{in } D_0. \tag{1.3.10}$$

Then, since $\frac{1}{h} + \frac{\pi}{c^2}$ is a strong invariant associated to the eigenvalue u,

$$\frac{1}{h_{l/r}} + \frac{P_{l/r}}{c^2} = \frac{1}{h^*} + \frac{\pi^*}{c^2}. \tag{1.3.11}$$

Thus the sum of (1.3.8), (1.3.9), (1.3.10) equals zero. Then, the terms in factor of u^* in D_{--},

D_{++} also sum up to zero, as well as the terms in factor of v^* in D_-, D_+. Then we look at the terms in factor of $-b^*$,

$$\frac{h^*a^*}{2c_a}\left(v^* - \frac{P_\perp^*}{c_a} - v^* + \frac{\pi_\perp^*}{c_a}\right) \quad \text{in } D_-, \tag{1.3.12}$$

$$\frac{-h^*a^*}{2c_a}\left(v^* + \frac{P_\perp^*}{c_a} - v^* - \frac{\pi_\perp^*}{c_a}\right) \quad \text{in } D_+, \tag{1.3.13}$$

$$\left(b^* + \frac{h^*a^*}{c_a^2}P_\perp^*\right) - \left(b_{l/r} + \frac{(ha)_{l/r}}{c_a^2}P_{\perp l/r}\right) \quad \text{in } D_0. \tag{1.3.14}$$

But since $b + \frac{ha}{c_a^2}\pi_\perp$ is a strong invariant associated to the eigenvalue u, one has

$$b_{l/r} + \frac{(ha)_{l/r}}{c_a^2}P_{\perp l/r} = b^* + \frac{h^*a^*}{c_a^2}\pi_\perp^*. \tag{1.3.15}$$

Thus we get that the sum of (1.3.12), (1.3.13), (1.3.14) equals zero.

Now it remains to sum up the first lines from (1.3.3)-(1.3.7). Summing up the first lines from D_{--} and D_{++} we get

$$\frac{1}{4}(u^* - \frac{P^*}{c})^2 - \frac{1}{4}(u^* - \frac{\pi^*}{c})^2 + \frac{1}{4}(u^* + \frac{P^*}{c})^2 - \frac{1}{4}(u^* + \frac{\pi^*}{c})^2 = \frac{(P^*)^2}{2c^2} - \frac{(\pi^*)^2}{2c^2}. \tag{1.3.16}$$

Summing up the first lines from D_- and D_+ we get

$$\frac{1}{4}(v^* - \frac{P_\perp^*}{c_a})^2 - \frac{1}{4}(v^* - \frac{\pi_\perp^*}{c_a})^2 + \frac{1}{4}(v^* + \frac{P_\perp^*}{c_a})^2 - \frac{1}{4}(v^* + \frac{\pi_\perp^*}{c_a})^2 = \frac{(P_\perp^*)^2}{2c_a^2} - \frac{(\pi_\perp^*)^2}{2c_a^2}. \tag{1.3.17}$$

The last terms are those from the first line of D_0,

$$\left(e(U^*) - \frac{(P^*)^2}{2c^2} - \frac{(P_\perp^*)^2}{2c_a^2}\right) - \left(e(U) - \frac{P^2}{2c^2} - \frac{P_\perp^2}{2c_a^2}\right)_{l/r}. \tag{1.3.18}$$

Moreover, according to (1.2.24), $\hat{e} - \frac{\pi^2}{2c^2} - \frac{\pi_\perp^2}{2c_a^2}$ is a strong invariant associated to the eigenvalue u, which gives

$$\left(e(U) - \frac{P^2}{2c^2} - \frac{P_\perp^2}{2c_a^2}\right)_{l/r} = \left(\hat{e}^* - \frac{(\pi^*)^2}{2c^2} - \frac{(\pi_\perp^*)^2}{2c_a^2}\right). \tag{1.3.19}$$

Summing up (1.3.16), (1.3.17) and (1.3.18), we get $e(U^*) - \hat{e}^*$, which proves the lemma. □

We notice now that all dissipations excepted the central one in Lemma 1.3 can be written as opposite of squares,

$$D_{--}(U,\Lambda) = -\frac{1}{4}(u - \frac{P}{c} - \Lambda)^2,\ D_{++}(U,\Lambda) = -\frac{1}{4}(u + \frac{P}{c} - \Lambda)^2, \tag{1.3.20}$$

$$D_{-}(U,\Lambda) = -\frac{1}{4}(v - \frac{P_\perp}{c_a} - \Lambda)^2,\ D_{+}(U,\Lambda) = -\frac{1}{4}(v + \frac{P_\perp}{c_a} - \Lambda)^2. \tag{1.3.21}$$

Thus they are all nonpositive, and in order to obtain a sufficient stability condition via Lemma 1.2 and Lemma 1.3, we need only to prove that D_0 from (1.3.7) is nonpositive.

In order to analyze $D_0(U^*, U_{l/r})$, we group the terms in factor of $1/2c^2$, $1/2c_a^2$ and the terms where b is involved, because they are squares, which gives using the expression of $P_\perp$ from (1.1.11) and the relation $h^*a^* = (ha)_{l/r}$,

$$\begin{aligned} D_0(U^*, U_{l/r}) = {}& \frac{gh^*}{2} - \frac{gh_{l/r}}{2} + \frac{(a^*)^2}{2} - \frac{a_{l/r}^2}{2} + P^*(\frac{1}{h^*} - \frac{1}{h_{l/r}}) \\ &+ \frac{1}{2c^2}(P^* - P_{l/r})^2 + \frac{1}{2c_a^2}(P_\perp^* - P_{\perp l/r})^2 - \frac{1}{2}(b^* - b_{l/r})^2. \end{aligned} \tag{1.3.22}$$

Next, we use the expression of P and $P_\perp$ from (1.1.11) to obtain

$$\begin{aligned} D_0(U^*, U_{l/r}) = {}& \frac{gh^*}{2} - \frac{gh_{l/r}}{2} + \frac{g(h^*)^2}{2}(\frac{1}{h^*} - \frac{1}{h_{l/r}}) \\ &+ \frac{(a^*)^2}{2} - \frac{a_{l/r}^2}{2} - h^*(a^*)^2(\frac{1}{h^*} - \frac{1}{h_{l/r}}) \\ &+ \frac{1}{2c^2}\left[\left(\frac{g(h^*)^2}{2} - \frac{g(h_{l/r})^2}{2}\right)^2 + (ha)_{l/r}^2(a^* - a_{l/r})^2 \right.\\ &\left. -2(ha)_{l/r}(a^* - a_{l/r})\left(\frac{g(h^*)^2}{2} - \frac{g(h_{l/r})^2}{2}\right)\right] \\ &- \frac{1}{2}\left(1 - \frac{(ha)_{l/r}^2}{c_a^2}\right)(b^* - b_{l/r})^2. \end{aligned} \tag{1.3.23}$$

Obviously, under the condition $c_a \geq h_{l/r}|a_{l/r}|$, the last line is nonpositive. The following result is a particular case of [27, Lemma 3.3].

Lemma 1.4. *For any $h_* > 0$, $h_{l/r} > 0$ one has*

$$\frac{gh_*}{2} - \frac{gh_{l/r}}{2} + \frac{gh_*^2}{2}(\frac{1}{h_*} - \frac{1}{h_{l/r}}) + \frac{1}{2}\frac{1}{(gh^3)_{*,l/r}}\left(\frac{gh_*^2}{2} - \frac{gh_{l/r}^2}{2}\right)^2 \leq 0, \tag{1.3.24}$$

with

$$(gh^3)_{*,l/r} = \sup_{h\in[h_{l/r},h_*]} gh^3. \tag{1.3.25}$$

Using the inequality (1.3.24), we get an upper bound on the first line of the right-hand side of (1.3.23),

$$\begin{aligned} D_0(U^*, U_{l/r}) \leq & -\frac{1}{2}\frac{1}{(gh^3)_{*,l/r}}\left(\frac{g(h^*)^2}{2} - \frac{g(h_{l/r})^2}{2}\right)^2 \\ & -\frac{(ha)^2_{l/r}}{2}\left(\frac{1}{h^*} - \frac{1}{h_{l/r}}\right)^2 \\ & +\frac{1}{2c^2}\left[\left(\frac{g(h^*)^2}{2} - \frac{g(h_{l/r})^2}{2}\right)^2 + (ha)^4_{l/r}\left(\frac{1}{h^*} - \frac{1}{h_{l/r}}\right)^2\right. \\ & \left. -2(ha)^2_{l/r}\left(\frac{1}{h^*} - \frac{1}{h_{l/r}}\right)\left(\frac{g(h^*)^2}{2} - \frac{g(h_{l/r})^2}{2}\right)\right]. \end{aligned} \tag{1.3.26}$$

Finally, we can rewrite the right-hand side of (1.3.26) as a quadratic form,

$$\begin{aligned} D_0(U^*, U_{l/r}) \leq & -\left(\frac{1}{2(gh^3)_{*,l/r}} - \frac{1}{2c^2}\right)\left(\frac{g(h^*)^2}{2} - \frac{g(h_{l/r})^2}{2}\right)^2 \\ & -\frac{1}{2}(ha)^2_{l/r}\left(1 - \frac{(ha)^2_{l/r}}{c^2}\right)\left(\frac{1}{h^*} - \frac{1}{h_{l/r}}\right)^2 \\ & -\frac{(ha)^2_{l/r}}{c^2}\left(\frac{1}{h^*} - \frac{1}{h_{l/r}}\right)\left(\frac{g(h^*)^2}{2} - \frac{g(h_{l/r})^2}{2}\right). \end{aligned} \tag{1.3.27}$$

It leads to the following proposition stating that the entropy condition reduces to subcharacteristic conditions.

Proposition 1.5. *In order to have $e(U^*) - \hat{e}^* \leq 0$ at the intermediate state U^* (ensuring the discrete energy inequality according to Lemma 1.2), it is enough that $h^* > 0$ and*

$$(ha)^2_{l/r} + (gh^3)_{*,l/r} \leq c^2 \tag{1.3.28}$$

and

$$(ha)^2_{l/r} \leq c_a^2, \tag{1.3.29}$$

where $(gh^3)_{*,l/r}$ is defined in (1.3.25), and $(ha)_{l/r}$, c, c_a are evaluated on the same side of the central wave as is U^*.

Proof. We use Lemma 1.3 and the previous analysis. With the condition (1.3.29), it yields (1.3.27). It is now sufficient that the right-hand side in (1.3.27) is a non positive quadratic form. Thus it is sufficient that

$$\left(\frac{(ha)^2_{l/r}}{c^2}\right)^2 \leq (ha)^2_{l/r}\left(1-\frac{(ha)^2_{l/r}}{c^2}\right)\left(\frac{1}{(gh^3)_{*,l/r}}-\frac{1}{c^2}\right), \tag{1.3.30}$$

or

$$\frac{(ha)^2_{l/r}}{c^2} \leq \left(1-\frac{(ha)^2_{l/r}}{c^2}\right)\left(\frac{c^2}{(gh^3)_{*,l/r}}-1\right). \tag{1.3.31}$$

Developing the right-hand side and simplifying, we obtain

$$\frac{(ha)^2_{l/r}}{(gh^3)_{*,l/r}} \leq \frac{c^2}{(gh^3)_{*,l/r}}-1. \tag{1.3.32}$$

Multiplying by $(gh^3)_{*,l/r}$ we get (1.3.28), which concludes the proof. □

1.3.2 Choice of signal speeds

In this subsection we derive explicit values for the signal speeds c, c_a that are sufficient for having positivity of height and (1.3.28), (1.3.29), that yield the energy inequality. Such values have been found in [24, Proposition 2.18] for Euler equations and in [28] for MHD equations. We use here the approach of [28], that enables to treat negative pressures π. We make the following a priori choice of the relaxation speeds c_l, c_r,

$$\begin{aligned} c_l &= h_l s_l + \frac{3}{2}h_l\left((u_l-u_r)_+ + \frac{(\pi_r-\pi_l)_+}{h_l s_l + h_r s_r}\right), \\ c_r &= h_r s_r + \frac{3}{2}h_r\left((u_l-u_r)_+ + \frac{(\pi_l-\pi_r)_+}{h_l s_l + h_r s_r}\right), \end{aligned} \tag{1.3.33}$$

with

$$s = \sqrt{a^2+gh}. \tag{1.3.34}$$

This choice for c_l and c_r implies in particular that $c_l \geq h_l s_l$ and $c_r \geq h_r s_r$. We set

$$X_l = \frac{1}{s_l}\left((u_l-u_r)_+ + \frac{(\pi_r-\pi_l)_+}{h_l s_l + h_r s_r}\right), \tag{1.3.35}$$

so that by (1.3.33), one has

$$\frac{c_l}{h_l} = s_l(1+\frac{3}{2}X_l). \tag{1.3.36}$$

Now we observe that we have two intermediate states on the left, but they have the same density h_l^*. Thus the conditions (1.3.28), (1.3.29) are identical for the "l**" and "l*" states, and refer to the height h_l^*. We estimate it from (1.2.50),

$$\frac{1}{h_l^*} = \frac{1}{h_l} + \frac{c_r(u_r - u_l) + \pi_l - \pi_r}{c_l(c_l + c_r)} \tag{1.3.37}$$

$$\geq \frac{1}{h_l} - \frac{c_r(u_l - u_r)_+}{c_l(c_l + c_r)} - \frac{(\pi_r - \pi_l)_+}{c_l(c_l + c_r)} \tag{1.3.38}$$

$$\geq \frac{1}{h_l} - \frac{(u_l - u_r)_+}{c_l} - \frac{(\pi_r - \pi_l)_+}{c_l(h_l s_l + h_r s_r)}. \tag{1.3.39}$$

Using (1.3.35) and (1.3.36), one gets

$$\frac{1}{h_l^*} \geq \frac{1}{h_l}\left(1 - \frac{X_l}{1 + \frac{3}{2}X_l}\right). \tag{1.3.40}$$

Thus $h_l^* > 0$, and

$$0 < h_l^* \leq h_l/x_l, \tag{1.3.41}$$

with

$$x_l = 1 - \frac{X_l}{1 + \frac{3}{2}X_l} \in (1/3, 1]. \tag{1.3.42}$$

This allows to estimate the supremum in (1.3.25),

$$\sqrt{(gh^3)_{*,l/r}} \leq \frac{h_l}{x_l}\sqrt{\frac{gh_l}{x_l}}. \tag{1.3.43}$$

We can estimate the right-hand side using the following result from [28, Lemma 3.1].

Lemma 1.6. *Consider a pressure law $p(\rho)$ defined for $\rho > 0$, satisfying*

$$\frac{d}{d\rho}\left(\rho\sqrt{p'}\right) > 0, \tag{1.3.44}$$

$$\frac{d}{d\rho}\left(\rho\sqrt{p'}\right) \leq \alpha\sqrt{p'}, \qquad \textit{for some constant } \alpha > 1. \tag{1.3.45}$$

Let $x = 1 - X/(1 + \alpha X)$ for some $X \geq 0$. Then for all $\rho > 0$,

$$\frac{\rho}{x}\sqrt{p'(\frac{\rho}{x})} \leq \rho\sqrt{p'(\rho)}(1 + \alpha X). \tag{1.3.46}$$

The assumptions concerning the pressure (1.3.44)-(1.3.45) are satisfied with $p(\rho) = g\rho^2/2$ and $\alpha = 3/2$. Therefore we can apply this result with $\rho = h_l$, $X = X_l$, which gives

$$\frac{h_l}{x_l}\sqrt{\frac{gh_l}{x_l}} \leq h_l\sqrt{gh_l}(1 + \frac{3}{2}X_l). \tag{1.3.47}$$

Thus, using (1.3.43) and (1.3.47), for getting (1.3.28) it is enough that

$$(ha)_l^2 + gh_l^3(1 + \frac{3}{2}X_l)^2 \leq c_l^2. \tag{1.3.48}$$

But using (1.3.36) and the definition of s_l from (1.3.34),

$$c_l^2 = s_l^2 h_l^2(1 + \frac{3}{2}X_l)^2 \tag{1.3.49}$$

$$= (ha)_l^2(1 + \frac{3}{2}X_l)^2 + gh_l^3(1 + \frac{3}{2}X_l)^2, \tag{1.3.50}$$

which yields (1.3.48). The same analysis is valid on the right, with

$$X_r = \frac{1}{s_r}\left((u_l - u_r)_+ + \frac{(\pi_l - \pi_r)_+}{h_l s_l + h_r s_r}\right), \quad x_r = 1 - \frac{X_r}{1 + \frac{3}{2}X_r}. \tag{1.3.51}$$

Proposition 1.7. *The solver is positive in height and entropy satisfying for the choice of c_l, c_r given by (1.3.33) and c_{al}, c_{ar} given by*

$$c_{al} = (h|a|)_l, \quad c_{ar} = (h|a|)_r. \tag{1.3.52}$$

Proof. We apply Proposition 1.5. The above computations show that $h_l^*, h_r^* > 0$ and that (1.3.28) holds. The choice (1.3.52) gives obviously (1.3.29). □

Lemma 1.8 (Bounds on the propagation speeds)**.** *The formulas (1.3.33) ensure the following estimate on propagation speeds:*

$$\max\left(\frac{c_l}{h_l}, \frac{c_r}{h_r}\right) \leq C\Big((u_l - u_r)_+ + s_l + s_r\Big), \tag{1.3.53}$$

where $C > 0$ is an absolute constant.

Proof. We have $|P| \leq hs^2$ with $P = g\frac{h^2}{2} - ha^2$ and $s^2 = a^2 + gh$. Since $\pi_l = P_l$ and $\pi_r = P_r$, the result follows obviously. $\square$

1.4 Properties of the relaxation approximate Riemann solver

Before stating our main result, we have to explicit the numerical fluxes and the CFL condition. The solution to the Riemann problem for the relaxation system (1.2.1)-(1.2.9) has five wave speeds $\Sigma_1 < \Sigma_2 < \Sigma_3 < \Sigma_4 < \Sigma_5$, that can be computed by (1.2.53). The intermediate states l^{**}, l^*, r^*, r^{**} have been determined in Section 1.2.3. We would like now to compute the left/right numerical fluxes $\mathcal{F}_l$, $\mathcal{F}_r$ that are involved in the update formula (1.2.15).

All components of the system except ha and hb are conservative, thus classical computations give the associated numerical fluxes,

$$\mathcal{F}_l = (\mathcal{F}^h,\ \mathcal{F}^{hu},\ \mathcal{F}^{hv},\ \mathcal{F}_l^{ha},\ \mathcal{F}_l^{hb}), \quad \mathcal{F}_r = (\mathcal{F}^h,\ \mathcal{F}^{hu},\ \mathcal{F}^{hv},\ \mathcal{F}_r^{ha},\ \mathcal{F}_r^{hb}), \tag{1.4.1}$$

where the conservative part involves the Riemann solution evaluated at x/t=0,

$$\begin{aligned} \mathcal{F}^h &= (hu)_{x/t=0}, \\ \mathcal{F}^{hu} &= (hu^2 + \pi)_{x/t=0}, \\ \mathcal{F}^{hv} &= (huv + \pi_\perp)_{x/t=0}. \end{aligned} \tag{1.4.2}$$

More explicitly (1.4.2) yields that the quantities between parentheses are evaluated at "l" if $\Sigma_1 \geq 0$, at "l^{**}" if $\Sigma_1 \leq 0 \leq \Sigma_2$, at "$l^*$" if $\Sigma_2 \leq 0 \leq \Sigma_3$, at "$r^*$" if $\Sigma_3 \leq 0 \leq \Sigma_4$, at "$r^{**}$" if $\Sigma_4 \leq 0 \leq \Sigma_5$, at "$r$" if $\Sigma_5 \leq 0$. As usual there is no ambiguity when equality occurs in these conditions. Similarly, the numerical energy flux is computed according to (1.2.30).

We complete these formulas by computing the left/right numerical fluxes for the variables ha and hb from (1.2.16),

$$\begin{aligned} \mathcal{F}_l^{ha} &= 0 + \min(0, \Sigma_3)\Big((ha)_r - (ha)_l\Big), \\ \mathcal{F}_r^{ha} &= 0 - \max(0, \Sigma_3)\Big((ha)_r - (ha)_l\Big), \end{aligned} \tag{1.4.3}$$

$$
\begin{aligned}
\mathcal{F}_l^{hb} = (hbu - hav)_l &+ \min(0,\Sigma_1)\Big((hb)_l^{**} - (hb)_l\Big) \\
&+ \min(0,\Sigma_2)\Big((hb)_l^{*} - (hb)_l^{**}\Big) + \min(0,\Sigma_3)\Big((hb)_r^{*} - (hb)_l^{*}\Big) \\
&+ \min(0,\Sigma_4)\Big((hb)_r^{**} - (hb)_r^{*}\Big) + \min(0,\Sigma_5)\Big((hb)_r - (hb)_r^{**}\Big),
\end{aligned}
\tag{1.4.4}
$$

$$
\begin{aligned}
\mathcal{F}_r^{hb} = (hbu - hav)_r &- \max(0,\Sigma_1)\Big((hb)_l^{**} - (hb)_l\Big) \\
&- \max(0,\Sigma_2)\Big((hb)_l^{*} - (hb)_l^{**}\Big) - \max(0,\Sigma_3)\Big((hb)_r^{*} - (hb)_l^{*}\Big) \\
&- \max(0,\Sigma_4)\Big((hb)_r^{**} - (hb)_r^{*}\Big) - \max(0,\Sigma_5)\Big((hb)_r - (hb)_r^{**}\Big).
\end{aligned}
\tag{1.4.5}
$$

Using the computation performed in [28, Subsection 5.3], these fluxes can be also written

$$
\text{If } \Sigma_3 \geq 0 \text{ then } \left\{
\begin{aligned}
\mathcal{F}_l^{hb} &= (hbu - hav)_{x/t=0-}, \\
\mathcal{F}_r^{hb} &= (hbu - hav)_{x/t=0-} - v^*\Big((ha)_r - (ha)_l\Big),
\end{aligned}\right.
\tag{1.4.6}
$$

$$
\text{If } \Sigma_3 \leq 0 \text{ then } \left\{
\begin{aligned}
\mathcal{F}_l^{hb} &= (hbu - hav)_{x/t=0+} + v^*\Big((ha)_r - (ha)_l\Big), \\
\mathcal{F}_r^{hb} &= (hbu - hav)_{x/t=0+},
\end{aligned}\right.
\tag{1.4.7}
$$

where v^* is the central value of v defined in (1.2.49) (and indeed $\Sigma_3 = u^*$).

The maximal propagation speed is then

$$
A(U_l, U_r) = \max\left(|\Sigma_1|, |\Sigma_2|, |\Sigma_3|, |\Sigma_4|, |\Sigma_5|\right) = \max\left(|\Sigma_1|, |\Sigma_5|\right). \tag{1.4.8}
$$

The CFL condition (1.2.13) becomes

$$
\Delta t A(U_i, U_{i+1}) \leq \frac{1}{2}\min(\Delta x_i, \Delta x_{i+1}). \tag{1.4.9}
$$

Note that with (1.3.53) and (1.2.53) we get

$$
A(U_l, U_r) \leq C(|u_l| + |u_r| + s_l + s_r) \tag{1.4.10}
$$

with C an absolute constant, bounding the propagation speed of the approximate Riemann solver by the left and right true speeds. This property is also valid in [25] for shallow water elastic fluids and is more general than the possibility of treating data with vacuum considered in [24]. Note that for getting (1.4.10), no restriction on the ratio h_l/h_r is required.

We are now able to obtain our main result.

Theorem 1.9. *Assume that the initial data U_l, U_r satisfy $h_l > 0$, $h_r > 0$ and use the choice of the speeds (1.3.33), (1.3.52) with $\pi_l = P(U_l)$, $\pi_r = P(U_r)$. Under the CFL condition (1.4.8), (1.4.9), the Riemann solver defined by the intermediate states and speeds Σ_i computed in Subsection 1.2.3 and the numerical fluxes $\mathcal{F}_l(U_l, U_r)$, $\mathcal{F}_r(U_l, U_r)$ defined via (1.4.1)-(1.4.7), has the following properties.*

(i) It keeps the positivity of h,

(ii) The approximate Riemann solver satisfies the discrete energy inequality (1.2.21), with numerical energy flux (1.2.30),

(iii) Isolated material contact discontinuities are resolved exactly,

(iv) Isolated Alfven contact discontinuities are resolved exactly,

(v) Resonant material-Alfven contact discontinuities defined by

$$h = cst, u = cst, a = 0, \tag{1.4.11}$$

are resolved exactly,

(vi) Data with bounded propagation speeds give finite numerical propagation speed, according to (1.4.10),

(vii) The numerical viscosity is sharp, in the sense that the propagation speeds Σ_i of the approximate Riemann solver tend to the exact propagation speeds when the left and right states U_l, U_r tend to a common value,

(viii) The variables h, hu, hv are conservative, and the scheme satisfies an asymptotic consistency with the non-conservative part of the system for smooth data.

Proof. The items (i), (ii) are consequences of Proposition 1.7 and Lemma 1.2. The items (iii), (iv) hold according to Lemma 1.1. The proof of (v) is straightforward with the formulas of Section 1.2.3, since in this case $\Sigma_2 = \Sigma_3 = \Sigma_4$. The item (vi) has been proved with Lemma 1.8. The item (vii) is obvious with (1.3.33), (1.3.52). The property that h, hu, hv are conservative is also obvious.

Thus it remains to prove the statement of (viii) concerning the consistency. Denote as in the introduction $\mathbf{b} = (a, b)$, $\mathbf{u} = (u, v)$. Then the magnetic equations of the original system

(1.1.9), (1.1.10) can be written

$$\partial_t(h\mathbf{b}) + \partial_x(h\mathbf{b}u - ha\mathbf{u}) + \mathbf{u}\partial_x(ha) = 0. \tag{1.4.12}$$

At the discrete level, the $h\mathbf{b}$ components of the update formula (1.2.15) involve the numerical fluxes from (1.4.1)

$$\mathcal{F}_l^{h\mathbf{b}} \equiv (\mathcal{F}_l^{ha},\ \mathcal{F}_l^{hb}), \quad \mathcal{F}_r^{h\mathbf{b}} \equiv (\mathcal{F}_r^{ha},\ \mathcal{F}_r^{hb}). \tag{1.4.13}$$

Using the formulas for the fluxes (1.4.3) and (1.4.6), (1.4.7), we observe that

$$\mathcal{F}_r^{h\mathbf{b}} - \mathcal{F}_l^{h\mathbf{b}} = -\mathbf{u}^*\Big((ha)_r - (ha)_l\Big). \tag{1.4.14}$$

Thus, asymptotically when $U_l, U_r \to U$, one has

$$\mathcal{F}_r^{h\mathbf{b}}(U_l,U_r) - \mathcal{F}_l^{h\mathbf{b}}(U_l,U_r) = -\mathbf{u}\Big((ha)_r - (ha)_l\Big) + o(|U_l - U| + |U_r - U|). \tag{1.4.15}$$

Since obviously $\mathcal{F}_l^{h\mathbf{b}}(U,U) = \mathcal{F}_r^{h\mathbf{b}}(U,U) = h\mathbf{b}u - ha\mathbf{u}$, we conclude the asymptotic consistency with (1.4.12) for smooth solutions (in the sense proposed in [24, Section 4.2]), which concludes the proof of (viii). □

1.5 Numerical tests

In this section we perform numerical computations in order to evaluate the properties of the scheme, in relation with Theorem 1.9. First and second order methods in time and space are evaluated, the latter using an ENO reconstruction, as described in [24, section 4.13]. The conservative variable is U as in (1.1.13), and the slope limitations are performed on the variables h, u, v, ha, b.

We take 200 points, and plot a reference solution obtained by a first order computation with 10000 points. The CFL-number is taken 1/2 in all tests. The space variable x is taken in $[0, 1]$, $g = 9.81$. Neumann boundary conditions are applied. Three test cases are investigated:

- Test 1 is a Riemann problem with $(ha)_l = (ha)_r$,
- Test 2 is a generic Riemann problem with positive height,
- Test 3 is a Riemann problem where the initial heights are taken positive on the left side and vanishing on the right side.

The numerical values for Test 1 are given in Table 1.1. In this case, $ha = 1/2$ remains constant, in accordance with (1.1.9). The first order method in time and space is evaluated, with our numerical fluxes defined by (1.4.1)-(1.4.7). Note that ha remains stricly constant at the discrete level, because of (1.4.3). Our results are compared to those obtained with the HLL flux, see for example [24, Equation (2.111)], applied here to the system (1.1.6)-(1.1.8), (1.1.10) with $ha = 1/2$, which is conservative. We ommit to plot the ha component, since it is constant in both methods. The reference solution is plotted in Figure 1.3. It consists of, from left to right, a left rarefaction wave, a left Alfven contact, a right Alfven contact, a right shock. There is no material contact. We observe as expected that the components h, u only vary through fast waves whereas the components v and b only vary through Alfven waves. In Figure 1.4 we observe that the performances on fast waves are extremely similar for both methods, whereas for Alfven waves, our 5-wave solver shows a much better resolution than the 2-wave HLL solver.

The numerical values for Test 2 are given in Table 1.2. The reference solution is plotted in Figure 1.5. It consists of, from left to right, a left rarefaction wave, a left Alfven contact, a material contact, a right Alfven contact, and a right shock. We observe in Figure 1.6 that the second order resolution improves the sharpness of contact discontinuities, but sometimes gives rise to slight instabilities.

The numerical values for Test 3 are given in Table 1.3. The reference solution is plotted in Figure 1.7. It consists of, from left to right, a left rarefaction wave, a left Alfven contact, a right rarefaction wave. There is no material contact, nor right Alfven contact. The numerical results are shown in Figure 1.8. We notice that on the right the height h vanishes, and the variables u, v, a, b take eventually non-physical values, due to the fact that only the conservative variables hu, hv, ha, hb make sense. Taking this into account, we observe that the computed solution achieves a good accuracy.

Values of x	h	u	v	a	b
x≤0.5	1.	0.2	0.7	0.5	0.4
0.5<x≤1	0.5	-0.1	0.3	1.	0.1

Tableau 1.1 – Initial data for Test 1, the Riemann problem with $ha = 1/2$

Values of x	h	u	v	a	b
x≤0.5	1.4	0.2	0.6	1	0.4
0.5<x≤1	0.2	-0.1	0.3	1.2	0.1

Tableau 1.2 – Initial data for Test 2, the Riemann problem with positive height

Values of x	h	u	v	a	b
x≤0.5	2.	1.	2.5	0.8	0.4
0.5<x≤1	0.	0.	0.	0.	0.

Tableau 1.3 – Initial data for Test 3, the Riemann problem with vanishing height

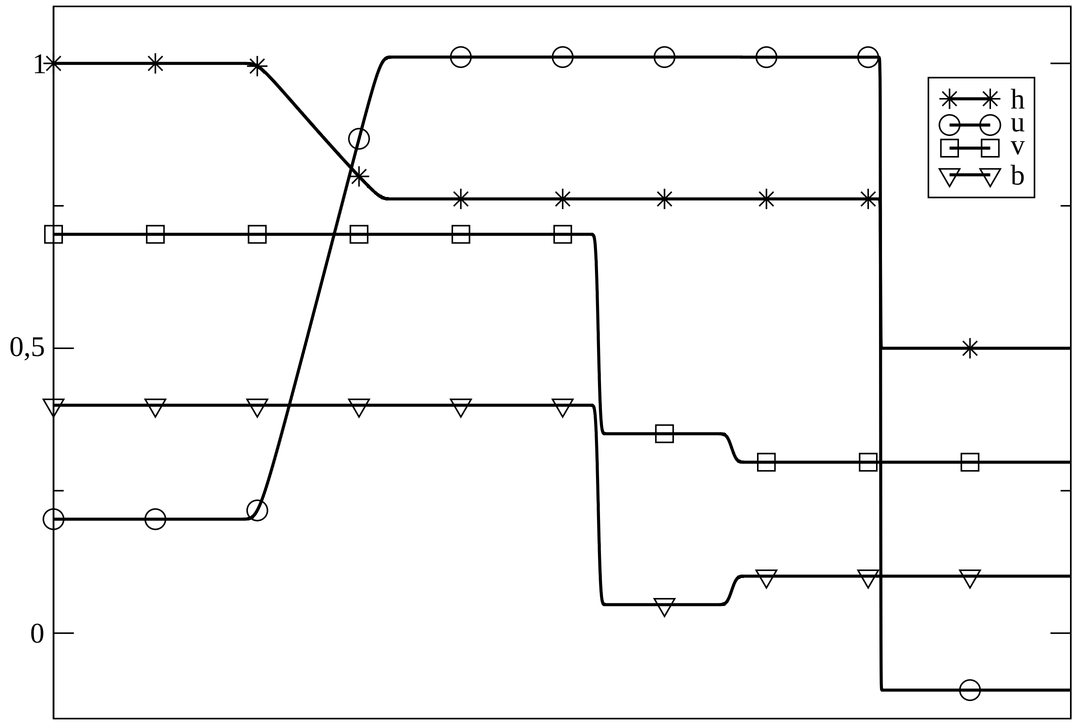


Figure 1.3 – Riemann solution for Test 1 at time $t = 0.1$ computed with the HLL flux or the 5-wave solver with 10000 points

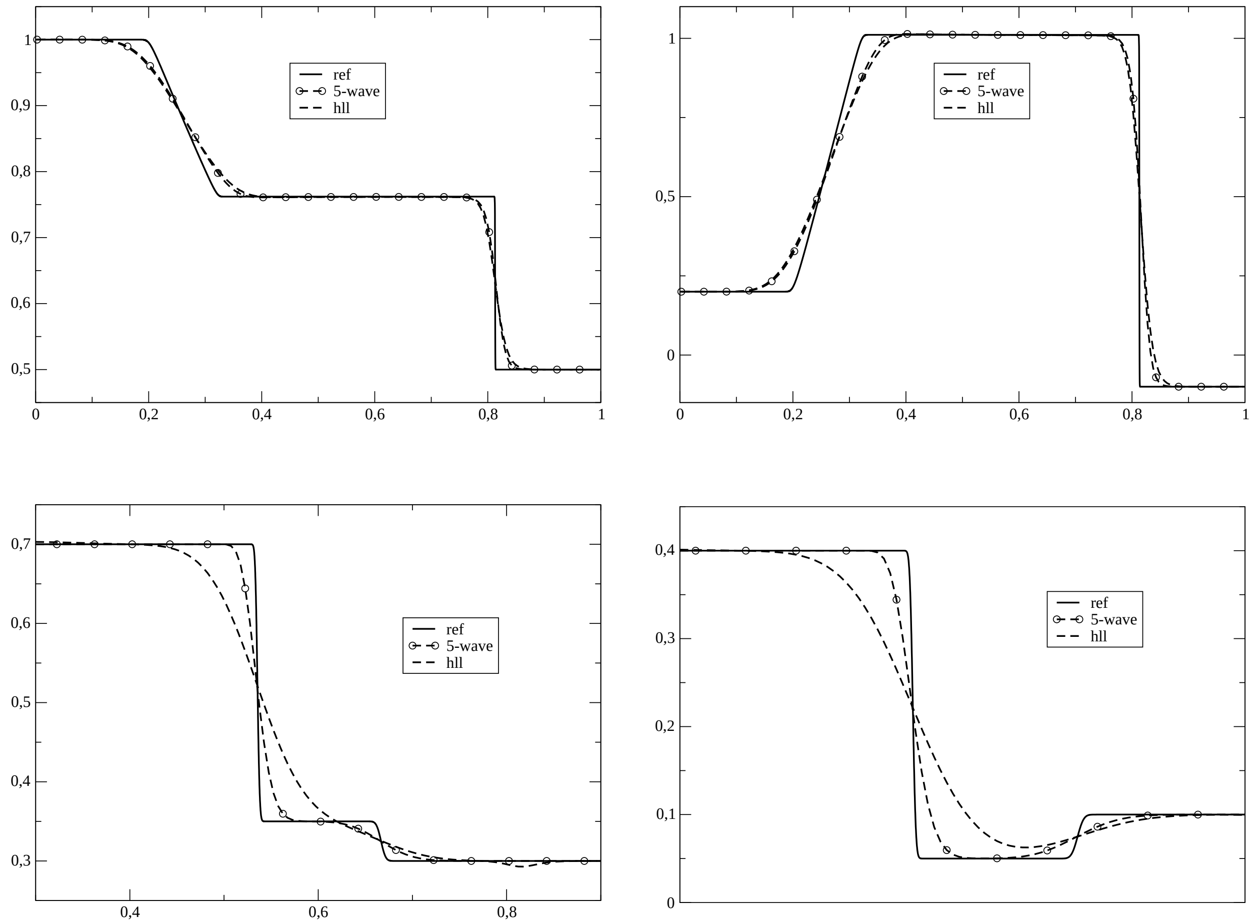


Figure 1.4 – Components h, u, v, b for Test 1 at time $t = 0.1$ computed with the HLL flux and the 5-wave solver with 200 points. The reference solution is the continuous line.

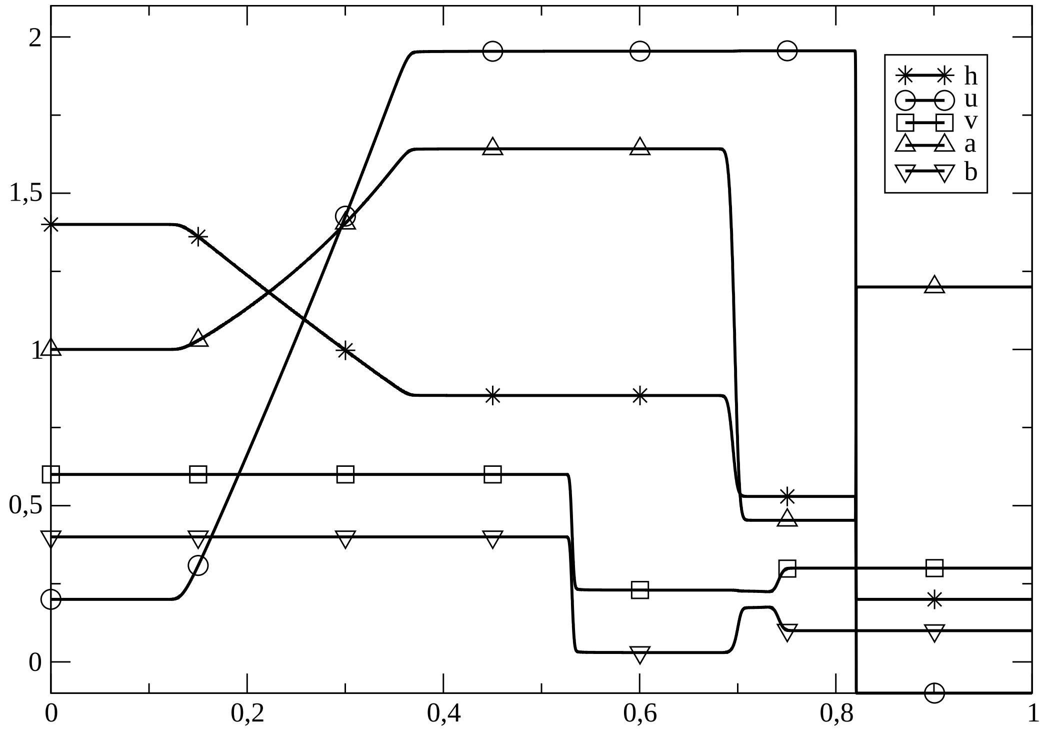


Figure 1.5 – Reference solution for Test 2 at time $t = 0.1$ computed at first order with 10000 points

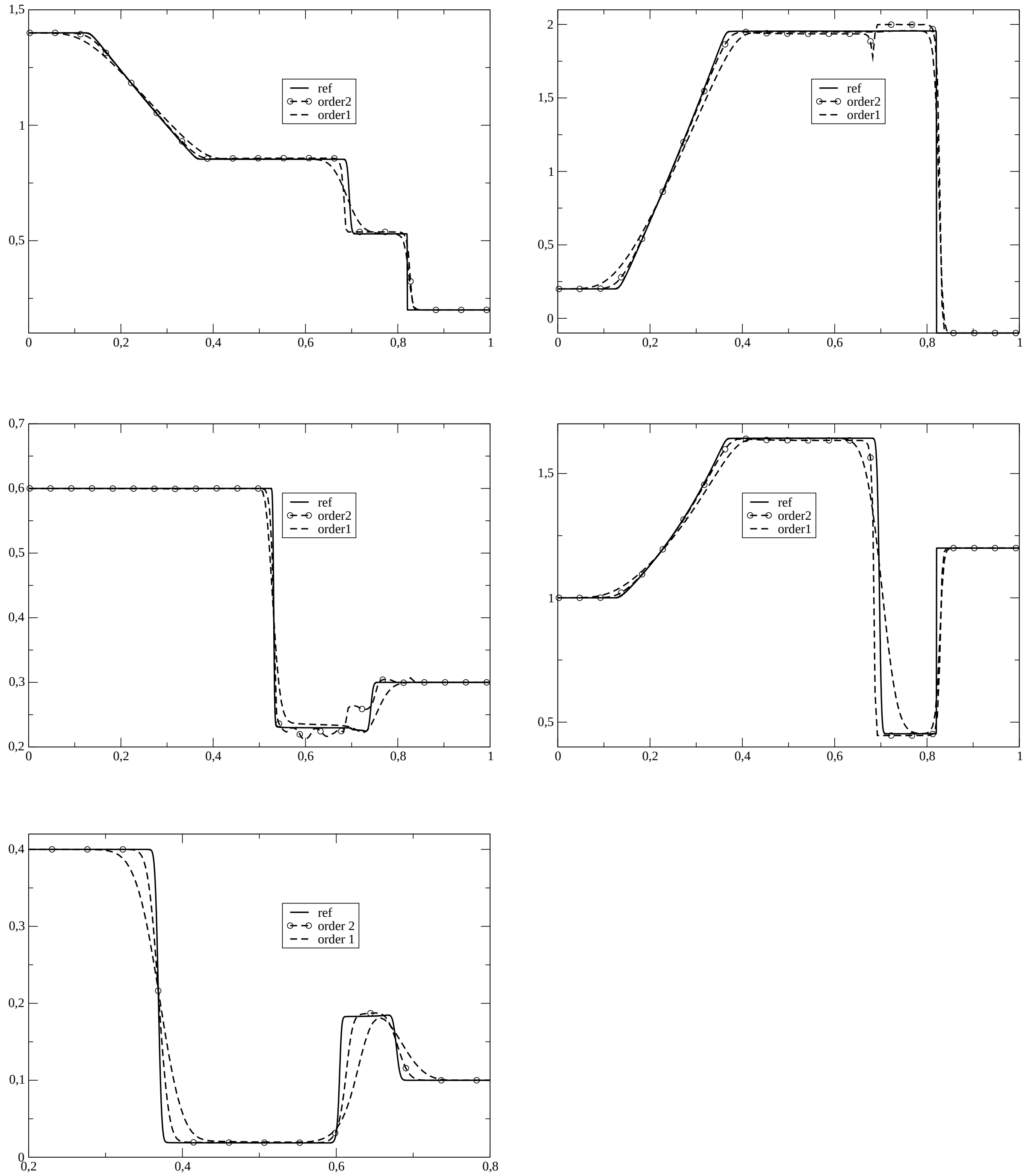


Figure 1.6 – Components h, u, v, a, b for Test 2 at time $t = 0.1$ computed at first and second order with 200 points. The reference solution is the continuous line.

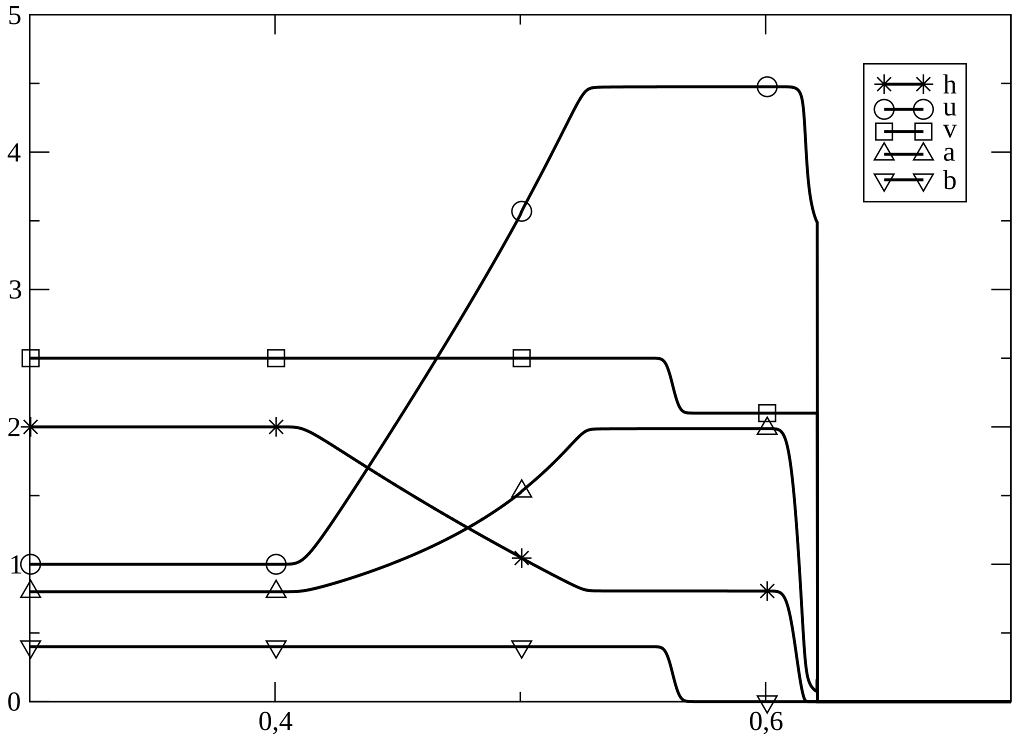


Figure 1.7 – Reference solution for Test 3 at time $t = 0.05$ computed at first order with 10000 points

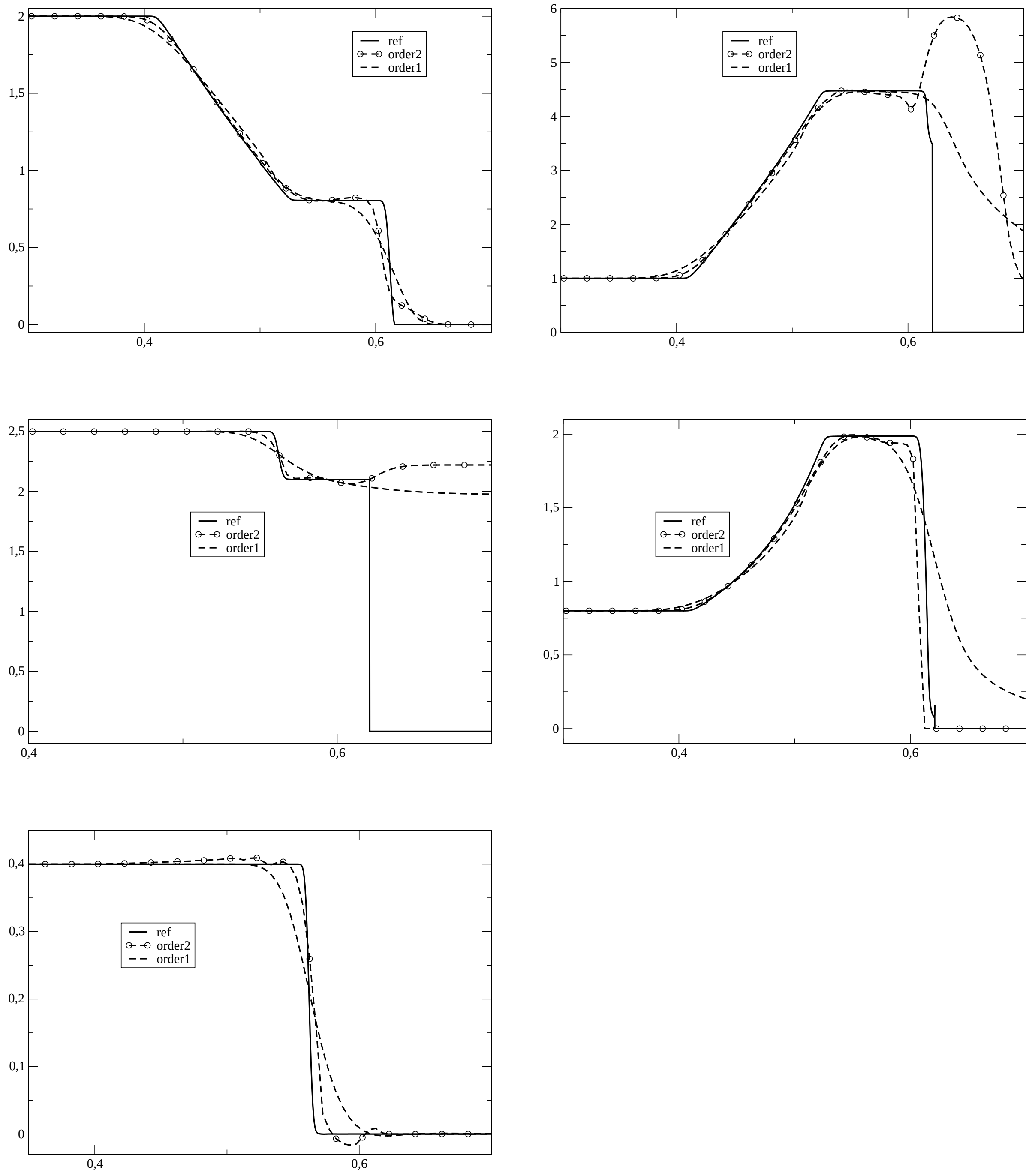


Figure 1.8 – Components h, u, v, a, b for Test 3 at time $t = 0.05$ computed at first and second order with 200 points. The reference solution is the continuous line.

Chapter 2

A multi well-balanced scheme for the shallow water MHD system with topography

Abstract

The shallow water magnetohydrodynamic system involves different families of physically relevant steady states. In this paper, we design a well-balanced numerical scheme for the shallow water magnetohydrodynamic system with topography, that resolves exactly a large family of steady states. It is obtained by a generalized hydrostatic reconstruction algorithm involving the magnetic field. It is positive in height and semi-discrete entropy satisfying, which ensures the robustness of the scheme.

Keywords: Shallow water magnetohydrodynamics, topography, well-balanced scheme, hydrostatic reconstruction, semi-discrete entropy inequality.

Mathematics Subject Classification: 76W05, 76M12, 35L65

2.1 Introduction

The shallow water magnetohydrodynamic (SWMHD) system has been introduced in [66] to describe the thin layer evolution of the solar tachocline. It is written in 2d in the tangent plane approximation as

$$\partial_t h + \nabla \cdot (h\mathbf{u}) = 0, \tag{2.1.1}$$

$$\partial_t(h\mathbf{u}) + \nabla\cdot(h\mathbf{u}\otimes\mathbf{u} - h\mathbf{b}\otimes\mathbf{b}) + \nabla(gh^2/2) + gh\nabla z + fh\mathbf{u}^{\perp} = 0, \tag{2.1.2}$$

$$\partial_t(h\mathbf{b}) + \nabla\cdot(h\mathbf{b}\otimes\mathbf{u} - h\mathbf{u}\otimes\mathbf{b}) + \mathbf{u}\nabla\cdot(h\mathbf{b}) = 0, \tag{2.1.3}$$

where $g > 0$ is the gravity constant, $h \geq 0$ is the thickness of the fluid, $\mathbf{u} = (u, v)$ is the velocity, $\mathbf{b} = (a, b)$ is the magnetic field, $z(x)$ is the topography, $f(x)$ is the Coriolis parameter, and $\mathbf{u}^{\perp}$ denotes the vector obtained from $\mathbf{u}$ by a rotation of angle $\pi/2$. The notation $\nabla\cdot(\mathbf{b}\otimes\mathbf{u})$ is for the vector with index i given by $\sum_j \partial_j(b_i u_j)$. The system has to be completed with the entropy (energy) inequality

$$\begin{aligned}&\partial_t\Big(\frac{1}{2}h|\mathbf{u}|^2 + \frac{1}{2}gh^2 + \frac{1}{2}h|\mathbf{b}|^2 + ghz\Big)\\ &+\nabla\cdot\Big(\Big(\frac{1}{2}h|\mathbf{u}|^2 + gh^2 + \frac{1}{2}h|\mathbf{b}|^2 + ghz\Big)\mathbf{u} - h\mathbf{b}(\mathbf{b}\cdot\mathbf{u})\Big) \leq 0,\end{aligned} \tag{2.1.4}$$

that becomes an equality in the absence of shocks. We recall that extra term $\mathbf{u}\nabla\cdot(h\mathbf{b})$ in the induction equation (2.1.3), that has been proposed in [52], is put for 2d numerical purposes only, while the physically relevant situation is $\nabla\cdot(h\mathbf{b}) = 0$.

In the one and a half dimensional setting, i.e. if dependency is only in one spatial variable x, the system simplifies to

$$\partial_t h + \partial_x(hu) = 0, \tag{2.1.5}$$

$$\partial_t(hu) + \partial_x(hu^2 + P) + gh\partial_x z - fhv = 0, \tag{2.1.6}$$

$$\partial_t(hv) + \partial_x(huv + P_{\perp}) + fhu = 0, \tag{2.1.7}$$

$$\partial_t(ha) + u\partial_x(ha) = 0, \tag{2.1.8}$$

$$\partial_t(hb) + \partial_x(hbu - hav) + v\partial_x(ha) = 0, \tag{2.1.9}$$

with

$$P = g\frac{h^2}{2} - ha^2, \quad P_{\perp} = -hab, \tag{2.1.10}$$

and the energy inequality (2.1.4) becomes

$$\begin{aligned}&\partial_t\Big(\frac{1}{2}h(u^2+v^2) + \frac{1}{2}gh^2 + \frac{1}{2}h(a^2+b^2) + ghz\Big)\\ &+\partial_x\Big(\Big(\tfrac{1}{2}h(u^2+v^2) + gh^2 + \tfrac{1}{2}h(a^2+b^2) + ghz\Big)u - ha(au+bv)\Big) \leq 0.\end{aligned} \tag{2.1.11}$$

According to [51], the eigenvalues of the system (2.1.5)-(2.1.9) are $u, u \pm |a|, u \pm \sqrt{a^2 + gh}$. The associated waves are called respectively material (or divergence) waves, Alfven waves and magnetogravity waves. It is classical in shallow water systems to consider the topography z as an additional variable to the system, satisfying $\partial_t z = 0$. In this setting there is an additional eigenvalue which is 0, and we shall call the associated wave the topography wave. The presence of the zero-order Coriolis terms proportional to f induces indeed more complex nonlinear waves [107]. These are studied numerically in [30]. In the present work, from now on we shall always assume that $f \equiv 0$.

The system (2.1.5)-(2.1.9) is nonconservative in the variables ha, hb. However, ha jumps only through the material contacts, where u and v are continuous. Therefore, there is indeed no ambiguity in the non conservative products $u\partial_x(ha)$ and $v\partial_x(ha)$, that are well-defined. Concerning the nonconservative term $h\partial_x z$ in (2.1.6), it is well-defined for continuous topography z. Piecewise constant discontinuous z is considered however for discrete approximations.

A striking property of the system (2.1.5)-(2.1.9) is that four out of six of the waves are contact discontinuities, corresponding to linearly degenerate eigenvalues: the material contacts associated to the eigenvalue u, the left Alfven contacts associated to $u - |a|$, the right Alfven contacts associated to $u + |a|$, and the topography contacts associated to the eigenvalue 0. Resonance can occur, which means that these waves can collapse. It happens in particular when $u = 0$ or $u \pm |a| = 0$.

Multidimensional simulations of the SWMHD system have been performed in [86, 91, 92]. As for the compressible MHD system, one-dimensional solvers that are accurate on contact waves are needed in order to reduce significantly the numerical diffusion in complex and multidimensional settings, that generically involve Alfven waves, see for example [12, 62]. At the same time, the robustness of the scheme must be maintained.

Well-balanced finite volume schemes for solving shallow water type models with topography have been extensively developed, see [24] and the references therein. A main principle in such schemes is to resolve exactly some steady states, in order to reduce significantly the numerical viscosity. The same question arises for hydrodynamic systems without topography, when linearly degenerate eigenvalues are involved. Indeed, in the numerical simulation of conservation laws, shocks are generally better resolved than contact discontinuities because of their compressive nature. This is why it is important to resolve well the contact discontinuities, that do not benefit of any compressive effect. In the SWMHD system (2.1.5)-(2.1.9), we have at the same time "dynamic" linearly degenerate eigenvalues (material and Alfven contact waves), and the "static" linearly degenerate eigenvalue (steady topography contact waves). The aim of this paper is to build a well-balanced scheme for the SWMHD system (2.1.5)-(2.1.9) that is accurate on all these contact waves. It follows [29], where we built an entropy satisfying approximate Riemann solver for the SWMHD system without topography that is accurate on all contact waves.

A generic tool for building well-balanced schemes that we use is the hydrostatic reconstruction method, that has been introduced in [6]. One of its strengths is that it enforces a semi-discrete entropy inequality, ensuring the robustness of the scheme and the computation of entropic shocks. Several variants and extensions have been proposed in [24, 25, 32, 34, 40]. Other approaches are the Roe method [11, 39, 41, 80, 81], the approximate Riemann solver method [19, 43, 64]. Central schemes are used also, and can handle multi steady states [46]. Higher-order extensions are reviewed in [105].

The paper is organized as follows. In Section 2.2 we describe the steady states of the SWMHD system with topography. In Section 2.3 we write down our numerical scheme, and in particular the reconstruction involved in the numerical fluxes, and our main result Theorem 2.1. Section 2.4 is devoted to the proof of this theorem. Finally, in Section 2.5 we perform numerical tests.

2.2 Steady states

As mentioned above, the system with topography (2.1.5)-(2.1.9) with $f \equiv 0$ has four linearly degenerate eigenvalues $u-|a|$, u, $u+|a|$ and 0, that can be resonant. We would like to build a scheme that is well-balanced for some contact waves for the eigenvalue 0, that are in particular steady states. Several cases can be considered. For each of them, it is straightforward to check that the following relations define steady states.

- Non-resonant case ($u \neq 0$ and $u \pm a \neq 0$). The relations are

$$\begin{gathered} hu = cst \; (\neq 0), \quad ha = cst \; (\neq \pm hu), \quad v = cst, \quad b = cst, \\ \frac{u^2}{2} - \frac{a^2}{2} + g(h+z) = cst. \end{gathered} \tag{2.2.1}$$

  As in the classical shallow water system, we shall not consider these steady states for the well-balanced property, because they are too complicate to handle (see however [32]).

- Material resonant case ($u = 0$ and $a \neq 0$). The differential relations are

$$\begin{gathered} u = 0, \quad v = cst, \quad hab = cst, \\ \partial_x\Big(g\frac{h^2}{2} - ha^2\Big) + gh\partial_x z = 0. \end{gathered} \tag{2.2.2}$$

Note that in contrast with the other cases, the second line in (2.2.2) is not integrable. It implies that for discontinuous data, this differential relation can have different possible interpretations in terms of nonconservative products.

We shall consider indeed a particular subfamily of steady states from (2.2.2), characterized by the relation $\sqrt{h}\ a = cst$, which yields

$$u = 0, \quad v = cst, \quad h + z = cst, \quad \sqrt{h}\, a = cst\ (\neq 0), \quad \sqrt{h}\, b = cst. \tag{2.2.3}$$

- Alfven resonant case ($u \neq 0$ and $u \pm a = 0$). The relations are

$$hu = cst\ (\neq 0), \quad ha = \mp hu, \quad h + z = cst, \quad v \pm b = cst. \tag{2.2.4}$$

- Material and Alfven resonant case ($u = a = 0$). The relations are

$$u = 0, \quad a = 0, \quad h + z = cst. \tag{2.2.5}$$

2.3 Formulas for the numerical fluxes, and main result

A finite volume scheme for the nonconservative system (2.1.5)-(2.1.9) with $f \equiv 0$ can be written

$$U_i^{n+1} = U_i^n - \frac{\Delta t}{\Delta x_i}\Big(F_l(U_i^n, U_{i+1}^n, \Delta z_{i+1/2}) - F_r(U_{i-1}^n, U_i^n, \Delta z_{i-1/2})\Big), \tag{2.3.1}$$

where

$$U = (h, hu, hv, ha, hb), \tag{2.3.2}$$

and as usual n stands for the time index, i for the space location, and $\Delta z_{i+1/2} = z_{i+1} - z_i$. Thus we need to define the left and right numerical fluxes $F_l(U_l, U_r, \Delta z)$, $F_r(U_l, U_r, \Delta z)$, for all left and right values U_l, U_r, z_l, z_r with $\Delta z = z_r - z_l$.

We use the hydrostatic reconstruction method of [6]. Denoting the left and right states by $U_l = (h_l, h_l u_l, h_l v_l, h_l a_l, h_l b_l)$, $U_r = (h_r, h_r u_r, h_r v_r, h_r a_r, h_r b_r)$, we define the reconstructed heights

$$h_l^{\#} = \Big(h_l - (\Delta z)_+\Big)_+, \quad h_r^{\#} = \Big(h_r - (-\Delta z)_+\Big)_+, \tag{2.3.3}$$

with the notation $x_+ \equiv \max(0, x)$. We also define new reconstructed magnetic states

$$a_l^{\#} = \kappa_l a_l, \quad a_r^{\#} = \kappa_r a_r, \tag{2.3.4}$$

$$b_l^\# = \kappa_l b_l, \quad b_r^\# = \kappa_r b_r, \tag{2.3.5}$$

with

$$\kappa_l = \min\left(\sqrt{\frac{h_l}{h_l^\#}}, \gamma\right), \quad \kappa_r = \min\left(\sqrt{\frac{h_r}{h_r^\#}}, \gamma\right), \tag{2.3.6}$$

and where $\gamma \geq 1$ is a cutoff parameter. We define finally the left and right reconstructed states as

$$U_l^\# = \left(h_l^\#, h_l^\# u_l, h_l^\# v_l, h_l^\# a_l^\#, h_l^\# b_l^\#\right), \quad U_r^\# = \left(h_r^\#, h_r^\# u_r, h_r^\# v_r, h_r^\# a_r^\#, h_r^\# b_r^\#\right). \tag{2.3.7}$$

Note that we use the notation $\#$ instead of $*$ in order to avoid confusions with intermediate states of Riemann solvers. Then the numerical fluxes are defined by

$$\begin{aligned}
F_l(U_l, U_r, \Delta z) = {} & \mathcal{F}_l(U_l^\#, U_r^\#) \\
& + \left(0,\ g\frac{h_l^2}{2} - g\frac{h_l^{\#2}}{2},\ 0,\ \left(\kappa_l (ha)_l^\# - (ha)_l\right) u_l,\ \left(\kappa_l (ha)_l^\# - (ha)_l\right) v_l\right) \\
& + (\kappa_l - 1)\ \left(0,\ 0,\ 0, \mathcal{F}_l^{ha}(U_l^\#, U_r^\#), \mathcal{F}_l^{hb}(U_l^\#, U_r^\#)\right) \\
& + \mathcal{F}^h(U_l^\#, U_r^\#)\left(0,\ 0,\ 0,\ \frac{a_l}{2}(1-\kappa_l^2),\ \frac{b_l}{2}(1-\kappa_l^2)\right), \\
F_r(U_l, U_r, \Delta z) = {} & \mathcal{F}_r(U_l^\#, U_r^\#) \\
& + \left(0,\ g\frac{h_r^2}{2} - g\frac{h_r^{\#2}}{2},\ 0,\ \left(\kappa_r (ha)_r^\# - (ha)_r\right) u_r,\ \left(\kappa_r (ha)_r^\# - (ha)_r\right) v_r\right) \\
& + (\kappa_r - 1)\ \left(0,\ 0,\ 0, \mathcal{F}_r^{ha}(U_l^\#, U_r^\#), \mathcal{F}_r^{hb}(U_l^\#, U_r^\#)\right) \\
& + \mathcal{F}^h(U_l^\#, U_r^\#)\left(0,\ 0,\ 0,\ \frac{a_r}{2}(1-\kappa_r^2),\ \frac{b_r}{2}(1-\kappa_r^2)\right),
\end{aligned} \tag{2.3.8}$$

where $\mathcal{F}_l$ and $\mathcal{F}_r$ are the numerical fluxes of [29] associated to the problem without topography, and $\mathcal{F}^h$ is its common left/right height flux.

Theorem 2.1. *The scheme (2.3.1) with the numerical fluxes F_l, F_r defined by (2.3.8) satisfies the following properties.*

(i) It is conservative in the variables h and hv,

(ii) It is consistent with (2.1.5)-(2.1.9) for smooth solutions,

(iii) It keeps the positivity of h under the CFL condition $\Delta t A(U_l^\#, U_r^\#) \leq \frac{1}{2}\min(\Delta x_l, \Delta x_r)$ with $A(.,.)$ the maximum speed of the homogeneous solver, defined by [29, eq. (4.8)],

(iv) It satisfies a semi-discrete energy inequality associated to (2.1.11),

(v) It is well-balanced with respect to steady material and Alfven contact discontinuities without jump in topography,

(vi) Under the condition

$$\gamma \geq \max\left(\sqrt{\frac{h_l}{h_l^\#}}, \sqrt{\frac{h_r}{h_r^\#}}\right) \tag{2.3.9}$$

on the parameter γ appearing in (2.3.6), it is well-balanced with respect to the steady states (2.2.3) *corresponding to material resonance,*

(vii) It is well-balanced with respect to the steady states (2.2.5) *corresponding to material and Alfven resonance.*

The proof of Theorem 2.1 is given in Section 2.4, and we give here some comments on this result.

- The formulas (2.3.8) for the numerical fluxes are defined exactly so that the proof of the entropy inequality is an identity. Then it follows that the scheme is consistent.
- The cutoff parameter γ is put in order to prevent κ_l, κ_r to be too large. Without it one would have possibly large $a_l^\#$, $a_r^\#$ and as a consequence a very restrictive CFL condition stated in (iii).
- The particular values (2.3.6) of κ_l, κ_r are involved only in the well-balanced property (vi), and do not matter for the other properties. We only need that their value is 1 when $\Delta z = 0$. In particular, if $\gamma = 1$, we get $\kappa_l = \kappa_r = 1$ (but then we loose the property (vi)). One can use also different formulas like

$$\kappa_l = \min\left(\frac{h_l}{h_l^\#}, \gamma\right), \quad \kappa_r = \min\left(\frac{h_r}{h_r^\#}, \gamma\right), \tag{2.3.10}$$

  the idea being to have, if γ is large enough, $\kappa_l = h_l/h_l^\#$, $\kappa_r = h_r/h_r^\#$, $h_l^\# a_l^\# = h_l a_l$, $h_r^\# a_r^\# = h_r a_r$. However, with (2.3.10) or with (2.3.6), the scheme does not leave invariant the data satisfying $ha = cst$, unfortunately.

- Instead of (2.2.3), one can consider another subfamily of steady states from (2.2.2), characterized by the relation $ha = cst$. It leads to consider the steady states

$$u = 0, \quad v = cst, \quad ha = cst \ (\neq 0), \quad b = cst, \quad h - \frac{a^2}{2g} + z = cst, \tag{2.3.11}$$

which are indeed the limit of (2.2.1) when $hu \to 0$. An interesting question would be to design a scheme that is well-balanced with respect to this family instead of (2.2.3).

2.4 Proof of Theorem 2.1

The proof of (i), i.e. $F_l^h = F_r^h$, $F_l^{hv} = F_r^{hv}$, is obvious from formulas (2.3.8) since the homogeneous solver already satisfies this property. We omit the proof of (iii), which follows the proof of Proposition 4.14 in [24].

The property (v) is inherited from the homogeneous solver that is described in [29], since

$$\Delta z = 0 \quad \text{implies} \quad F_l(U_l, U_r, 0) = \mathcal{F}_l(U_l, U_r),\ F_r(U_l, U_r, 0) = \mathcal{F}_r(U_l, U_r). \tag{2.4.1}$$

We recall more explicitly that, defining

$$F(U) = (hu, hu^2 + P, huv + P_\perp, 0, hbu - hav), \tag{2.4.2}$$

this property of well-balancing for the homogeneous solver means that $\mathcal{F}_l(U_l, U_r) = F(U_l)$ and $\mathcal{F}_r(U_l, U_r) = F(U_r)$ for all data of the form:

$$u_l = u_r = 0,\ v_l = v_r,\ P(U_l) = P(U_r),\ P_\perp(U_l) = P_\perp(U_r), \tag{2.4.3}$$

or

$$h_l = h_r,\ a_l = a_r \neq 0,\ u_l = u_r = |a_l|,\ b_l \operatorname{sgn}(a_l) - v_l = b_r \operatorname{sgn}(a_r) - v_r, \tag{2.4.4}$$

or

$$h_l = h_r,\ a_l = a_r \neq 0,\ u_l = u_r = -|a_l|,\ b_l \operatorname{sgn}(a_l) + v_l = b_r \operatorname{sgn}(a_r) + v_r, \tag{2.4.5}$$

or

$$h_l = h_r,\ u_l = u_r = 0,\ a_l = a_r = 0. \tag{2.4.6}$$

For the proof of (vi), consider data U_l, U_r, z_l, z_r satisfying (2.2.3), i.e. $u_l = u_r = 0$, $v_l = v_r$,

$h_l + z_l = h_r + z_r$, $\sqrt{h_l}a_l = \sqrt{h_r}a_r$, $\sqrt{h_l}b_l = \sqrt{h_r}b_r$. Then from (2.3.3) we get

$$h_l^\# = h_r^\# \equiv h^\#, \tag{2.4.7}$$

the common value $h^\#$ being h_r if $\Delta z \geq 0$, or h_l if $\Delta z \leq 0$. Using condition (2.3.9), according to (2.3.4), (2.3.5), (2.3.6), we get $\kappa_l = \sqrt{h_l/h_l^\#}$, $\kappa_r = \sqrt{h_r/h_r^\#}$, $\sqrt{h_l^\#}a_l^\# = \sqrt{h_l}a_l$, $\sqrt{h_r^\#}a_r^\# = \sqrt{h_r}a_r$, $\sqrt{h_l^\#}b_l^\# = \sqrt{h_l}b_l$, $\sqrt{h_r^\#}b_r^\# = \sqrt{h_r}b_r$. Thus

$$\sqrt{h_l^\#}a_l^\# = \sqrt{h_r^\#}a_r^\#, \quad \sqrt{h_l^\#}b_l^\# = \sqrt{h_r^\#}b_r^\#. \tag{2.4.8}$$

Using (2.4.7), (2.4.8), we get

$$U_l^\# = U_r^\# \equiv U^\# \equiv (h^\#, 0, h^\# v^\#, h^\# a^\#, h^\# b^\#). \tag{2.4.9}$$

We observe that then $\mathcal{F}_l(U_l^\#, U_r^\#) = \mathcal{F}_r(U_l^\#, U_r^\#) = F(U^\#)$, and that indeed

$$F(U^\#) = \left(0, g(h^\#)^2/2 - h^\#(a^\#)^2, -h^\# a^\# b^\#, 0, -h^\# a^\# v^\#\right). \tag{2.4.10}$$

The formulas (2.3.8) yield

$$F_l = \Big(0, g(h_l)^2/2 - h_l(a_l)^2, -h_l a_l b_l, 0, -h_l a_l v_l\Big) = F(U_l), \tag{2.4.11}$$

$$F_r = \Big(0, g(h_r)^2/2 - h_r(a_r)^2, -h_r a_r b_r, 0, -h_r a_r v_r\Big) = F(U_r), \tag{2.4.12}$$

which proves the claim.

For the proof of (vii), consider data U_l, U_r, z_l, z_r satisfying (2.2.5), i.e. $u_l = u_r = 0$, $h_l + z_l = h_r + z_r$, $a_l = a_r = 0$. Then we get $h_l^\# = h_r^\#$, $u_l^\# = u_r^\# = 0$, $a_l^\# = a_r^\# = 0$, and the fluxes $\mathcal{F}_l$, $\mathcal{F}_r$ are evaluated on states $U_l^\#$, $U_r^\#$ of the type (2.4.6). Thus $\mathcal{F}_l(U_l^\#, U_r^\#) = F(U_l^\#)$ and $\mathcal{F}_r(U_l^\#, U_r^\#) = F(U_r^\#)$. Plugging this in (2.3.8), we obtain $F_l = F(U_l)$, $F_r = F(U_r)$, which proves the claim.

2.4.1 Consistency

In order to get (ii) in Theorem 2.1 in the sense of Definition 4.2 in [24], we need to prove that

$$F_l(U, U, 0) = F_r(U, U, 0) = F(U), \tag{2.4.13}$$

and that as $U_l \to U$, $U_r \to U$, $\Delta z \to 0$,

$$\begin{aligned} F_r(U_l, U_r, \Delta z) - F_l(U_l, U_r, \Delta z) = -B(u,v)\Big((ha)_r - (ha)_l\Big) \\ +\Big(0, -gh\Delta z, 0, 0, 0\Big) + o\Big(|U_l - U| + |U_r - U| + |\Delta z|\Big), \end{aligned} \tag{2.4.14}$$

with

$$B(u,v) = (0,0,0,u,v). \tag{2.4.15}$$

The consistency with the exact flux (2.4.13) is obviously satisfied because of the property (2.4.1). In order to prove the consistency with the source (2.4.14), we write

$$\begin{aligned} &F_r(U_l, U_r, \Delta z) - F_l(U_l, U_r, \Delta z) \\ =\ &\mathcal{F}_r(U_l^\#, U_r^\#) - \mathcal{F}_l(U_l^\#, U_r^\#) \\ &+ B(u_r, v_r)\Big(\kappa_r (ha)_r^\# - (ha)_r\Big) - B(u_l, v_l)\Big(\kappa_l (ha)_l^\# - (ha)_l\Big) \\ &+ (\kappa_r - 1)\ \Big(0,\ 0,\ 0, \mathcal{F}_r^{ha}(U_l^\#, U_r^\#), \mathcal{F}_r^{hb}(U_l^\#, U_r^\#)\Big) \\ &- (\kappa_l - 1)\ \Big(0,\ 0,\ 0, \mathcal{F}_l^{ha}(U_l^\#, U_r^\#), \mathcal{F}_l^{hb}(U_l^\#, U_r^\#)\Big) \\ &+ \mathcal{F}^h(U_l^\#, U_r^\#)\left(0,\ 0,\ 0,\ \frac{a_r}{2}(1-\kappa_r^2),\ \frac{b_r}{2}(1-\kappa_r^2)\right) \\ &- \mathcal{F}^h(U_l^\#, U_r^\#)\left(0,\ 0,\ 0,\ \frac{a_l}{2}(1-\kappa_l^2),\ \frac{b_l}{2}(1-\kappa_l^2)\right) \\ &+ \left(0, g\tfrac{h_l^{\#2}}{2} - g\tfrac{h_l^2}{2} + g\tfrac{h_r^2}{2} - g\tfrac{h_r^{\#2}}{2}, 0, 0, 0\right). \end{aligned} \tag{2.4.16}$$

Let us denote $\Delta = |U_l - U| + |U_r - U| + |\Delta z|$. When $U_l, U_r \to U$ and $\Delta z \to 0$ one has from (2.3.3)-(2.3.7) $\kappa_l - 1 = O(\Delta)$, $\kappa_r - 1 = O(\Delta)$, and thus $U_l^\# - U = O(\Delta)$, $U_r^\# - U = O(\Delta)$ (we consider only the case $h > 0$ here). Then the consistency of the numerical flux without source obtained in [29] gives

$$\mathcal{F}_r(U_l^\#, U_r^\#) - \mathcal{F}_l(U_l^\#, U_r^\#) = -B(u,v)\left((ha)_r^\# - (ha)_l^\#\right) + o(\Delta). \tag{2.4.17}$$

Next, we have

$$B(u_r, v_r)\Big(\kappa_r (ha)_r^\# - (ha)_r\Big) = B(u,v)\Big(\kappa_r (ha)_r^\# - (ha)_r\Big) + o(\Delta), \tag{2.4.18}$$

and

$$B(u_l, v_l)\Big(\kappa_l (ha)_l^\# - (ha)_l\Big) = B(u,v)\Big(\kappa_l (ha)_l^\# - (ha)_l\Big) + o(\Delta). \tag{2.4.19}$$

Summing up (2.4.17), (2.4.18), (2.4.19), we obtain

$$
\begin{aligned}
&\mathcal{F}_r(U_l^\#, U_r^\#) - \mathcal{F}_l(U_l^\#, U_r^\#) \\
&\quad + B(u_r, v_r)\Big(\kappa_r (ha)_r^\# - (ha)_r\Big) - B(u_l, v_l)\Big(\kappa_l (ha)_l^\# - (ha)_l\Big) \\
&= B(u,v)(\kappa_r - 1)(ha)_r^\# - B(u,v)(\kappa_l - 1)(ha)_l^\# \\
&\quad - B(u,v)\Big((ha)_r - (ha)_l\Big) + o(\Delta). \\
&= B(u,v)(\kappa_r - 1)(ha) - B(u,v)(\kappa_l - 1)(ha) \\
&\quad - B(u,v)\Big((ha)_r - (ha)_l\Big) + o(\Delta).
\end{aligned}
\tag{2.4.20}
$$

Now we look at the terms in the right-hand side of (2.4.16) from the third to the sixth line. Using that $\mathcal{F}_l^{ha}(U,U) = \mathcal{F}_r^{ha}(U,U) = 0$ and $\mathcal{F}_l^{hb}(U,U) = \mathcal{F}_r^{hb}(U,U) = hbu - hav$, we deduce

$$
\begin{aligned}
&(\kappa_r - 1)\ \Big(0,\ 0,\ 0, \mathcal{F}_r^{ha}(U_l^\#, U_r^\#), \mathcal{F}_r^{hb}(U_l^\#, U_r^\#)\Big) \\
&= (\kappa_r - 1)\ \ (0,\ 0,\ 0, 0, hbu - hav) + o(\Delta),
\end{aligned}
\tag{2.4.21}
$$

and

$$
\begin{aligned}
&-(\kappa_l - 1)\ \Big(0,\ 0,\ 0, \mathcal{F}_l^{ha}(U_l^\#, U_r^\#), \mathcal{F}_l^{hb}(U_l^\#, U_r^\#)\Big) \\
&= -(\kappa_l - 1)\ \ (0,\ 0,\ 0, 0, hbu - hav) + o(\Delta).
\end{aligned}
\tag{2.4.22}
$$

Writing $1 - \kappa_r^2 = (1 + \kappa_r)(1 - \kappa_r)$, we get asymptotically

$$
\frac{a_r}{2}(1 - \kappa_r^2) = a(1 - \kappa_r) + o(\Delta). \tag{2.4.23}
$$

Similarly, we have

$$
\frac{a_l}{2}(1 - \kappa_l^2) = a(1 - \kappa_l) + o(\Delta), \tag{2.4.24}
$$

$$
\frac{b_r}{2}(1 - \kappa_r^2) = b(1 - \kappa_r) + o(\Delta), \tag{2.4.25}
$$

$$
\frac{b_l}{2}(1 - \kappa_l^2) = b(1 - \kappa_l) + o(\Delta). \tag{2.4.26}
$$

Using (2.4.23), (2.4.24), (2.4.25), (2.4.26) and the property $\mathcal{F}^h(U,U) = hu$, we obtain

$$
\begin{aligned}
&\mathcal{F}^h(U_l^\#, U_r^\#)\left(0,\ 0,\ 0,\ \frac{a_r}{2}(1 - \kappa_r^2),\ \frac{b_r}{2}(1 - \kappa_r^2)\right) \\
&= \Big(0,\ 0,\ 0,\ hua(1 - \kappa_r),\ hub(1 - \kappa_r)\Big) + o(\Delta),
\end{aligned}
\tag{2.4.27}
$$

$$\begin{aligned}&-\mathcal{F}^h(U_l^\#,U_r^\#)\left(0,\ 0,\ 0,\ \frac{a_l}{2}(1-\kappa_l^2),\ \frac{b_l}{2}(1-\kappa_l^2)\right)\\&=-\Big(0,\ 0,\ 0,\ hua(1-\kappa_l),\ hub(1-\kappa_l)\Big)+o(\Delta).\end{aligned}\tag{2.4.28}$$

The sum of (2.4.21), (2.4.22), (2.4.27), (2.4.28) gives the asymptotic formula

$$\begin{aligned}&(\kappa_r-1)\ \Big(0,\ 0,\ 0,\mathcal{F}_r^{ha}(U_l^\#,U_r^\#),\mathcal{F}_r^{hb}(U_l^\#,U_r^\#)\Big)\\&-(\kappa_l-1)\ \Big(0,\ 0,\ 0,\mathcal{F}_l^{ha}(U_l^\#,U_r^\#),\mathcal{F}_l^{hb}(U_l^\#,U_r^\#)\Big)\\&+\mathcal{F}^h(U_l^\#,U_r^\#)\left(0,\ 0,\ 0,\ \frac{a_r}{2}(1-\kappa_r^2),\ \frac{b_r}{2}(1-\kappa_r^2)\right)\\&-\mathcal{F}^h(U_l^\#,U_r^\#)\left(0,\ 0,\ 0,\ \frac{a_l}{2}(1-\kappa_l^2),\ \frac{b_l}{2}(1-\kappa_l^2)\right)\\&=-B(u,v)(\kappa_r-1)(ha)+B(u,v)(\kappa_l-1)(ha)+o(\Delta).\end{aligned}\tag{2.4.29}$$

Adding (2.4.20) and (2.4.29), we obtain the consistency of the nonconservative magnetic terms

$$\begin{aligned}&\mathcal{F}_r(U_l^\#,U_r^\#)-\mathcal{F}_l(U_l^\#,U_r^\#)\\&+B(u_r,v_r)\Big(\kappa_r(ha)_r^\#-(ha)_r\Big)-B(u_l,v_l)\Big(\kappa_l(ha)_l^\#-(ha)_l\Big)\\&+(\kappa_r-1)\ \Big(0,\ 0,\ 0,\mathcal{F}_r^{ha}(U_l^\#,U_r^\#),\mathcal{F}_r^{hb}(U_l^\#,U_r^\#)\Big)\\&-(\kappa_l-1)\ \Big(0,\ 0,\ 0,\mathcal{F}_l^{ha}(U_l^\#,U_r^\#),\mathcal{F}_l^{hb}(U_l^\#,U_r^\#)\Big)\\&+\mathcal{F}^h(U_l^\#,U_r^\#)\left(0,\ 0,\ 0,\ \frac{a_r}{2}(1-\kappa_r^2),\ \frac{b_r}{2}(1-\kappa_r^2)\right)\\&-\mathcal{F}^h(U_l^\#,U_r^\#)\left(0,\ 0,\ 0,\ \frac{a_l}{2}(1-\kappa_l^2),\ \frac{b_l}{2}(1-\kappa_l^2)\right)\\&=-B(u,v)\Big((ha)_r-(ha)_l\Big)+o(\Delta).\end{aligned}\tag{2.4.30}$$

Finally, as in the unmodified hydrostatic reconstruction scheme, the last line in (2.4.16) gives the nonconservative topography term

$$\left(0,g\frac{h_l^{\#2}}{2}-g\frac{h_l^2}{2}+g\frac{h_r^2}{2}-g\frac{h_r^{\#2}}{2},0,0,0\right)=\Big(0,-gh\Delta z,0,0,0\Big)+o(\Delta).\tag{2.4.31}$$

With (2.4.30), all the terms in (2.4.16) have been expanded, and we get (2.4.14).

2.4.2 Entropy inequality

Let us finally prove the property (iv) in Theorem 2.1. At the continous level, the energy inequality (2.1.11) can be written

$$\partial_t \tilde{E} + \partial_x \tilde{G} \leq 0, \tag{2.4.32}$$

with

$$\tilde{E}(U,z) = E(U) + ghz, \quad \tilde{G}(U,z) = G(U) + ghzu, \tag{2.4.33}$$

and

$$\begin{aligned} E(U) &= \frac{1}{2}h(u^2+v^2) + \frac{1}{2}gh^2 + \frac{1}{2}h(a^2+b^2), \\ G(U) &= E(U)u + P(U)u + P_\perp(U)v. \end{aligned} \tag{2.4.34}$$

As before, $U = (h, hu, hv, ha, hb)$ and P, $P_\perp$ are defined by (2.1.10). As proved in [29], the scheme without topography satisfies a fully discrete energy inequality. According to [24, section 2.2.2], it implies that it satisfies also a semi-discrete energy inequality, under the form

$$\begin{aligned} &G(U_r) + E'(U_r)\left(\mathcal{F}_r(U_l,U_r) - F(U_r)\right) \leq \mathcal{G}(U_l,U_r), \\ &\mathcal{G}(U_l,U_r) \leq G(U_l) + E'(U_l)\left(\mathcal{F}_l(U_l,U_r) - F(U_l)\right), \end{aligned} \tag{2.4.35}$$

for all values of U_l, U_r, where E' is the derivative of E with respect to U, F is defined in (2.4.2), and $\mathcal{G}(U_l,U_r)$ is a consistent energy flux.

Then, for the scheme with topography, the characterization of the semi-discrete energy inequality writes

$$\begin{aligned} &\tilde{G}(U_r,z_r) + \tilde{E}'(U_r,z_r)\left(F_r - F(U_r)\right) \leq \tilde{\mathcal{G}}(U_l,U_r,z_l,z_r), \\ &\tilde{\mathcal{G}}(U_l,U_r,z_l,z_r) \leq \tilde{G}(U_l,z_l) + \tilde{E}'(U_l,z_l)\left(F_l - F(U_l)\right), \end{aligned} \tag{2.4.36}$$

where $\tilde{E}$ and $\tilde{G}$ are defined by (2.4.33), $\tilde{E}'$ is the derivative of $\tilde{E}$ with respect to U, and $\tilde{\mathcal{G}}$ is an unknown consistent numerical energy flux. Let us choose

$$\tilde{\mathcal{G}}(U_l,U_r,z_l,z_r) = \mathcal{G}(U_l^\#,U_r^\#) + \mathcal{F}^h(U_l^\#,U_r^\#)gz^\#, \tag{2.4.37}$$

where $\mathcal{F}^h$ is the common h-component of $\mathcal{F}_l$ and $\mathcal{F}_r$, and for some $z^\#$ that is defined below. Then, noticing that $\tilde{E}'(U,z) = E'(U) + gz(1,0,0,0,0)$, we can write the desired inequalities (2.4.36) as

$$\begin{aligned} &G(U_r) + E'(U_r)\left(F_r - F(U_r)\right) + \mathcal{F}^h(U_l^\#,U_r^\#)gz_r \\ &\leq \mathcal{G}(U_l^\#,U_r^\#) + \mathcal{F}^h(U_l^\#,U_r^\#)gz^\#, \end{aligned} \tag{2.4.38}$$

$$\begin{aligned}&\mathcal{G}(U_l^\#, U_r^\#) + \mathcal{F}^h(U_l^\#, U_r^\#)gz^\# \\ &\leq G(U_l) + E'(U_l)\left(F_l - F(U_l)\right) + \mathcal{F}^h(U_l^\#, U_r^\#)gz_l.\end{aligned} \tag{2.4.39}$$

By using (2.4.35) evaluated at $U_l^\#$, $U_r^\#$ and comparing the result with (2.4.38) and (2.4.39), we get the sufficient conditions

$$\begin{aligned}&G(U_r) + E'(U_r)\left(F_r - F(U_r)\right) + \mathcal{F}^h(U_l^\#, U_r^\#)gz_r \\ &\leq G(U_r^\#) + E'(U_r^\#)\Big(\mathcal{F}_r(U_l^\#, U_r^\#) - F(U_r^\#)\Big) + \mathcal{F}^h(U_l^\#, U_r^\#)gz^\#,\end{aligned} \tag{2.4.40}$$

$$\begin{aligned}&G(U_l^\#) + E'(U_l^\#)\Big(\mathcal{F}_l(U_l^\#, U_r^\#) - F(U_l^\#)\Big) + \mathcal{F}^h(U_l^\#, U_r^\#)gz^\# \\ &\leq G(U_l) + E'(U_l)\left(F_l - F(U_l)\right) + \mathcal{F}^h(U_l^\#, U_r^\#)gz_l.\end{aligned} \tag{2.4.41}$$

Let us focus on (2.4.40), that can be rewritten as

$$\begin{aligned}\Big[G - E'F\Big]_{r\#}^{r} + E'(U_r)F_r - E'(U_r^\#)\mathcal{F}_r(U_l^\#, U_r^\#)& \\ +g(z_r - z^\#)\mathcal{F}^h(U_l^\#, U_r^\#) &\leq 0,\end{aligned} \tag{2.4.42}$$

with

$$\Big[G - E'F\Big]_{r\#}^{r} \equiv \Big(G(U_r) - E'(U_r)F(U_r)\Big) - \Big(G(U_r^\#) - E'(U_r^\#)F(U_r^\#)\Big). \tag{2.4.43}$$

We compute now

$$E'(U) = \Big(-\left(u^2 + v^2\right)/2 + gh - \left(a^2 + b^2\right)/2, u, v, a, b\Big), \tag{2.4.44}$$

and using (2.4.34), (2.4.2), we deduce the identity

$$G(U) - E'(U)F(U) = -g\frac{h^2}{2}u + ha(au + bv) = -P(U)u - P_\perp(U)v. \tag{2.4.45}$$

Then, according to the definition (2.3.8) of F_r,

$$\begin{aligned}E'(U_r)F_r = E'(U_r)&\mathcal{F}_r(U_l^\#, U_r^\#) \\ &+ E'(U_r)\Bigg(0,\ g\frac{h_r^2}{2} - g\frac{h_r^{\#2}}{2},\ 0, \\ &\quad\Big(\kappa_r(ha)_r^\# - (ha)_r\Big)u_r,\ \Big(\kappa_r(ha)_r^\# - (ha)_r\Big)v_r\Bigg) + Q_r,\end{aligned} \tag{2.4.46}$$

with

$$
\begin{aligned}
Q_r = {} & E'(U_r)(\kappa_r - 1)\left(0,\ 0,\ 0, \mathfrak{F}_r^{ha}(U_l^{\#}, U_r^{\#}), \mathfrak{F}_r^{hb}(U_l^{\#}, U_r^{\#})\right) \\
& + E'(U_r)\mathfrak{F}^h(U_l^{\#}, U_r^{\#})\left(0,\ 0,\ 0,\ \frac{a_r}{2}(1-\kappa_r^2),\ \frac{b_r}{2}(1-\kappa_r^2)\right).
\end{aligned}
\tag{2.4.47}
$$

Using (2.4.44) and (2.4.45), we can rewrite (2.4.46) as

$$
E'(U_r)F_r = E'(U_r)\mathfrak{F}_r(U_l^{\#}, U_r^{\#}) - \Big[G - E'F\Big]_{r\#}^{r} + Q_r. \tag{2.4.48}
$$

Thus the required inequality (2.4.42) simplifies to

$$
\Big(E'(U_r) - E'(U_r^{\#})\Big)\mathfrak{F}_r(U_l^{\#}, U_r^{\#}) + Q_r + g(z_r - z^{\#})\mathfrak{F}^h(U_l^{\#}, U_r^{\#}) \le 0. \tag{2.4.49}
$$

Now, one the one side, one can compute

$$
\begin{aligned}
Q_r = {} & (\kappa_r - 1)a_r\mathfrak{F}_r^{ha}(U_l^{\#}, U_r^{\#}) + (\kappa_r - 1)b_r\mathfrak{F}_r^{hb}(U_l^{\#}, U_r^{\#}) \\
& + (1-\kappa_r^2)\mathfrak{F}^h(U_l^{\#}, U_r^{\#})\frac{a_r^2 + b_r^2}{2}.
\end{aligned}
\tag{2.4.50}
$$

On the other side, according to (2.4.44), we have

$$
\begin{aligned}
& E'(U_r) - E'(U_r^{\#}) \\
= & \left(g(h_r - h_r^{\#}) - \frac{a_r^2 + b_r^2}{2} + \frac{(a_r^{\#})^2 + (b_r^{\#})^2}{2}, 0, 0, a_r - a_r^{\#}, b_r - b_r^{\#}\right) \\
= & \left(g(h_r - h_r^{\#}) - (1-\kappa_r^2)\frac{a_r^2 + b_r^2}{2}, 0, 0, (1-\kappa_r)a_r, (1-\kappa_r)b_r\right).
\end{aligned}
\tag{2.4.51}
$$

Using both (2.4.50) and (2.4.51), we get

$$
\Big(E'(U_r) - E'(U_r^{\#})\Big)\mathfrak{F}_r(U_l^{\#}, U_r^{\#}) + Q_r = g(h_r - h_r^{\#})\mathfrak{F}^h(U_l^{\#}, U_r^{\#}). \tag{2.4.52}
$$

Plugging this in (2.4.49), we obtain the sufficient right inequality

$$
g(h_r - h_r^{\#} + z_r - z^{\#})\mathfrak{F}^h(U_l^{\#}, U_r^{\#}) \le 0. \tag{2.4.53}
$$

A symmetric analysis for the left inequality (2.4.41) gives similarly

$$
g(h_l - h_l^{\#} + z_l - z^{\#})\mathfrak{F}^h(U_l^{\#}, U_r^{\#}) \ge 0. \tag{2.4.54}
$$

We choose $z^\# = \max(z_l, z_r)$, so that (2.4.53), (2.4.54) can be finally put under the form

$$\begin{aligned} g(h_r - h_r^\# - (-\Delta z)_+)\mathcal{F}^h(U_l^\#, U_r^\#) \le 0, \\ g(h_l - h_l^\# - (\Delta z)_+)\mathcal{F}^h(U_l^\#, U_r^\#) \ge 0. \end{aligned} \tag{2.4.55}$$

Taking into account (2.3.3), we observe that if $h_l - (\Delta z)_+ \ge 0$ then the second line of (2.4.55) is trivial. Otherwise $h_l^\# = 0$ and the second inequality of (2.4.55) holds because $\mathcal{F}^h(0, U_r^\#) \le 0$ by the h-nonnegativity of the numerical flux. The same argument is valid for the first inequality of (2.4.55), which concludes the proof of Theorem 2.1.

2.5 Numerical tests

In this section we perform numerical computations in order to evaluate the properties of the scheme, in relation with Theorem 2.1. First and second order methods in time and space are evaluated, the latter using an ENO reconstruction, as described in [24, section 4.13]. The conservative variable is U as in (2.3.2), and the slope limitations are performed on the variables h, $h+z$, u, v, ha, b. We also compare different values of the parameter γ, which is a key to obtain the well-balanced property for steady states of material resonance. We take 200 points, and plot a reference solution obtained by a second order computation with 3300 points. The CFL-number is taken 1/2 in all tests.

Test case - The space variable x is taken in $[0, 1]$, $g = 9.81$. Neumann boundary conditions are applied. The test consists of two steady states:

- On $[0, 1/2)$, we take initial data corresponding to a steady state in the case of material resonance.
- On $(1/2, 1]$, we take initial data corresponding to a steady state in the case of material and Alfven resonance.

The initial data is sketched on Figure 2.1 and the numerical values are given in Tables 2.1 and 2.2. The solution consists of, from left to right, a material contact, a left rarefaction wave, a left Alfven contact, a resonant material - right Alfven contact, and a right shock. We observe that the second order resolution improves the sharpness of contact discontinuities. On Figure 2.10, we observe that the solution computed with $\gamma = 1$ looses the well-balanced property for the resonant material contact, whereas with $\gamma = 2$ it is well-balanced, which is coherent with point (vi) of Theorem 2.1.

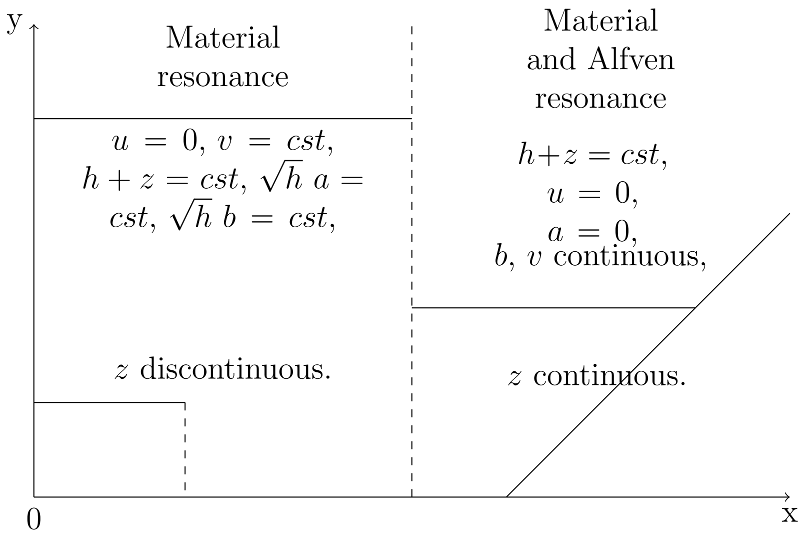


Figure 2.1 – Initial data configuration

Values of x	z	h	u	v	a	b
x≤0.2	0.5	1.5	0.0	2.0	$1/\sqrt{1.5}$	$2/\sqrt{1.5}$
0.2<x≤0.5	0.0	2.0	0.0	2.0	$1/\sqrt{2}$	$2/\sqrt{2}$

Tableau 2.1 – Initial data for Material resonance

Values of x	z	h	u	v	a	b
0.5<x ≤0.625	0.0	0.5	0.0	0.5	0.0	1.0
0.625<x≤1	$d(x)$	$(0.5 - d(x))_+$	0.0	0.5+$d(x)$	0.0	1.0+$d(x)$

Tableau 2.2 – Initial data for Material and Alfven resonance, $d(x) = 4.0\ (x - 0.625)$

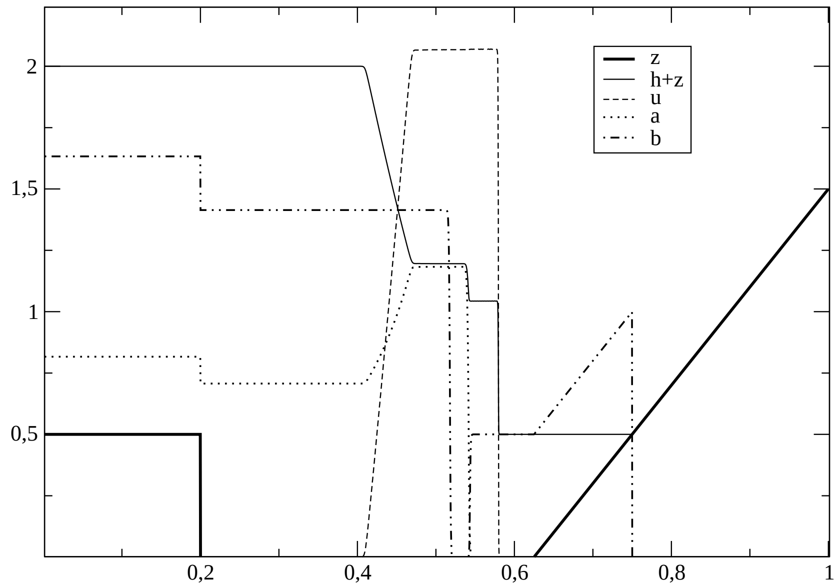


Figure 2.2 – Reference solution at time $t = 0.02$ computed at second order with 3300 points

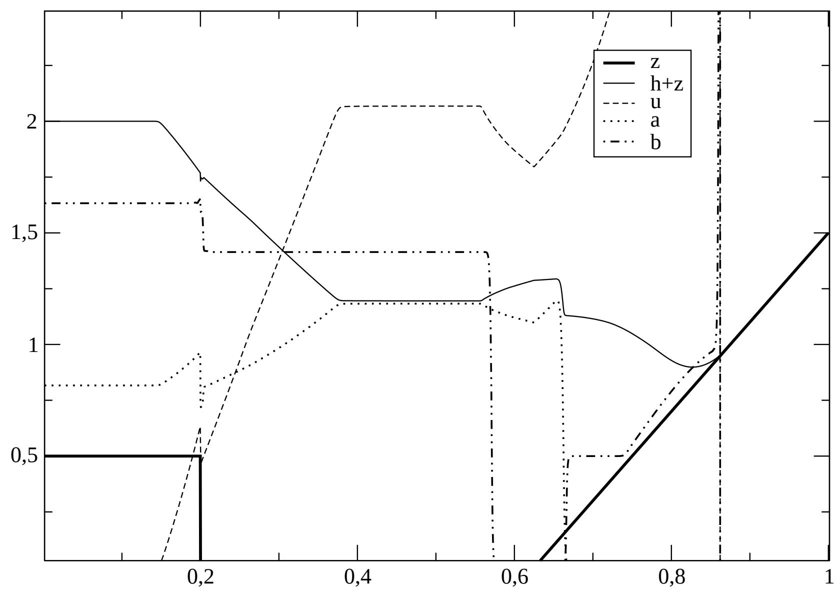


Figure 2.3 – Reference solution at time $t = 0.08$ computed at second order with 3300 points

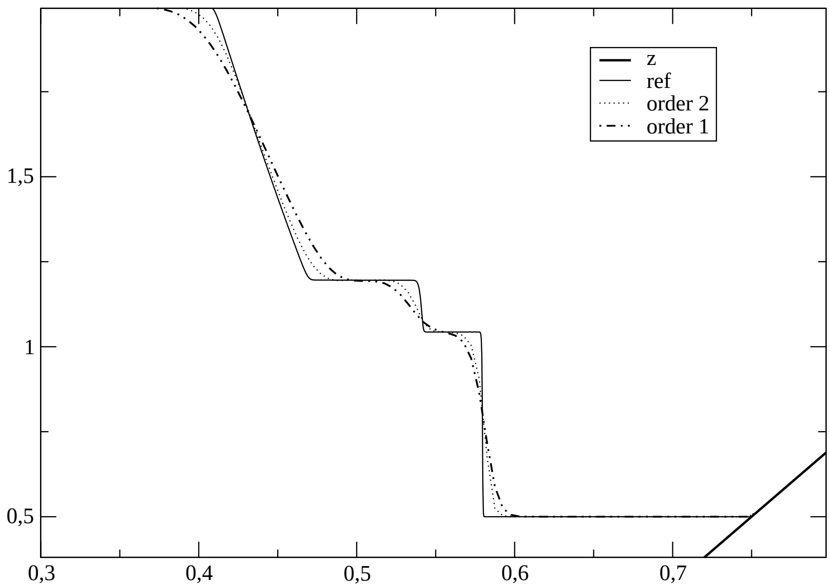


Figure 2.4 – Solution $h + z$ at time $t = 0.02$ computed at first and second order with 200 points. The reference solution is the continuous line.

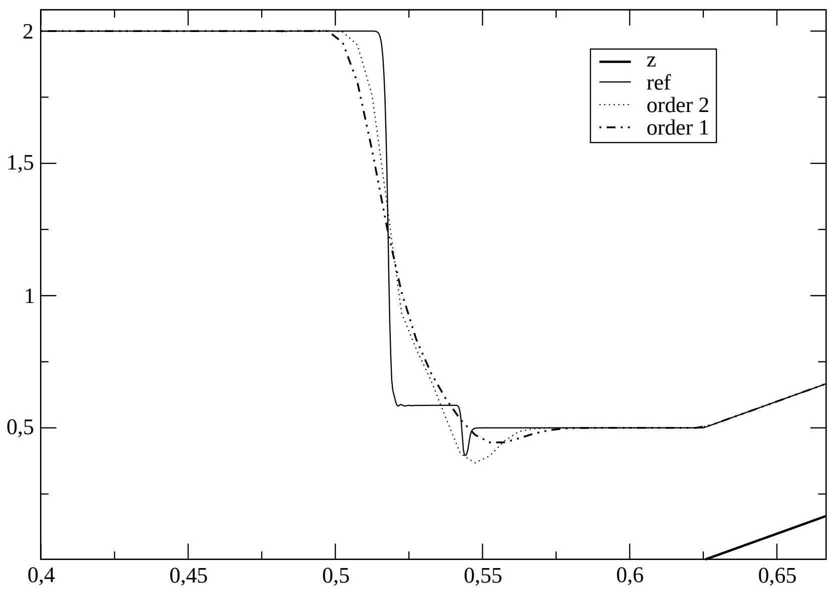


Figure 2.5 – Solution v at time $t = 0.02$ computed at first and second order with 200 points. The reference solution is the continuous line.

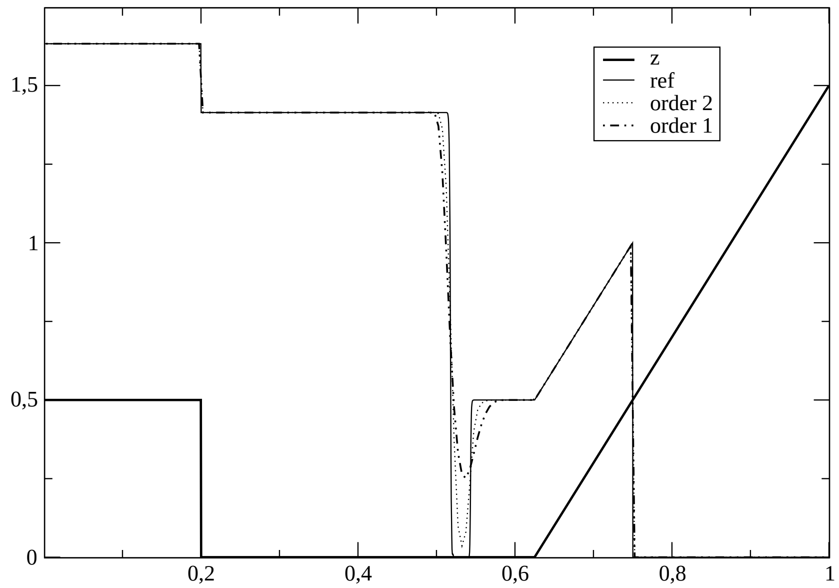


Figure 2.6 – Solution b at time $t = 0.02$ computed at first and second order with 200 points. The reference solution is the continuous line.

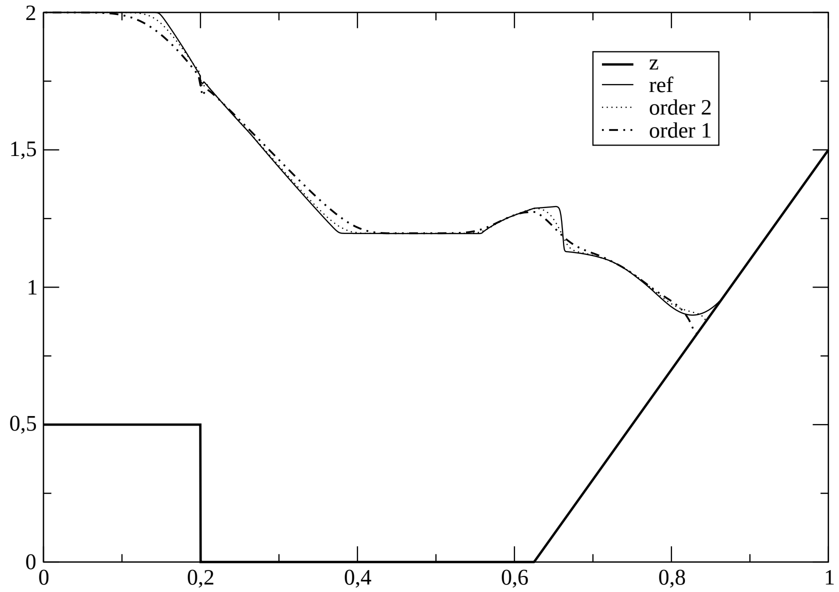


Figure 2.7 – Solution $h + z$ at time $t = 0.08$ computed at first and second order with 200 points. The reference solution is the continuous line.

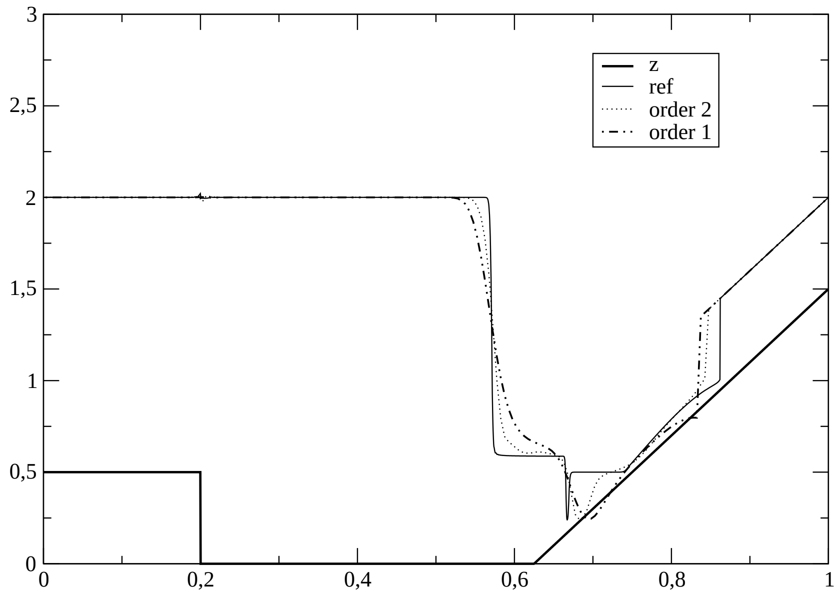


Figure 2.8 – Solution v at time $t = 0.08$ computed at first and second order with 200 points. The reference solution is the continuous line.

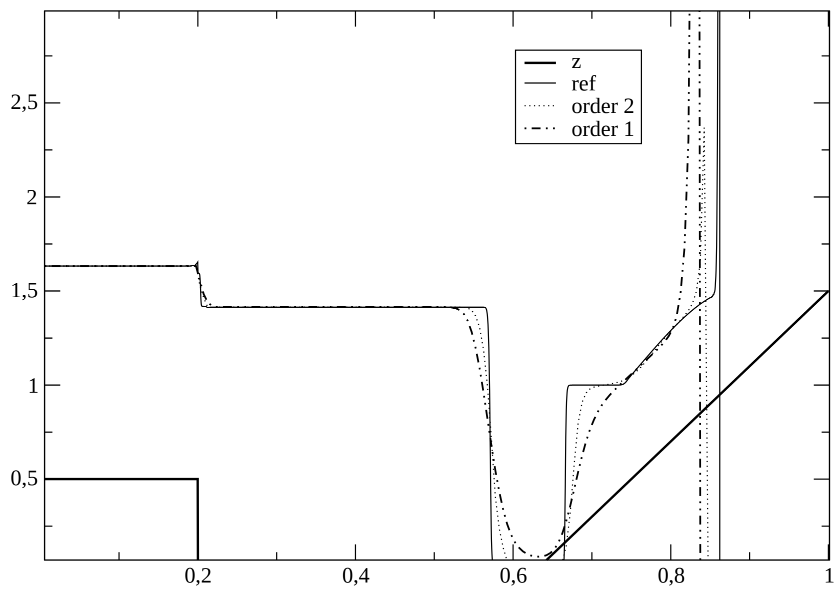


Figure 2.9 – Solution b at time $t = 0.08$ computed at first and second order with 200 points. The reference solution is the continuous line.

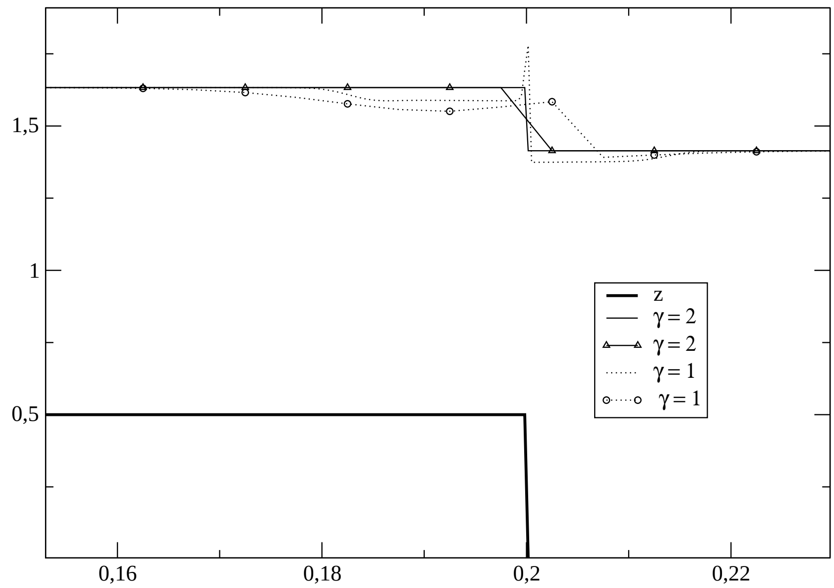


Figure 2.10 – Solution b at time $t = 0.02$ computed at first order with 200 points with different values of γ. The reference solutions are the continuous lines.

Chapter 3

Convergence of the kinetic hydrostatic reconstruction scheme for the Saint Venant system

Abstract

We prove the convergence of the kinetic scheme with hydrostatic reconstruction for the Saint-Venant system with topography. Using a sharp analysis of the dissipation recently proposed in [7], we establish an estimate in the inverse of the square root of the space increment Δx on the L^2 norm of gradient of approximate solution. It allows by the method of Diperna to prove the convergence towards weak entropy solutions.

3.1 Introduction and main result

We consider the Saint Venant system

$$\begin{aligned} \partial_t h + \partial_x(hu^2) = 0, \\ \partial_t(hu) + \partial_x(hu^2 + \frac{g}{2}h^2) = -gh\partial_x z, \end{aligned} \tag{3.1.1}$$

where $h(t,x) \geq 0$, $u(t,x) \in \mathbb{R}$, $g > 0$ is the gravity constant, and the topography $z(x)$ is given. The system is completed with an entropy inequality

$$\partial_t \left(h\frac{u^2}{2} + g\frac{h^2}{2} + ghz \right) + \partial_x \left(\left(h\frac{u^2}{2} + gh^2 + ghz \right) u \right) \leq 0. \tag{3.1.2}$$

We shall denote $U = (h, hu)$, $h \geq 0$ and

$$\eta(U) = h\frac{u^2}{2} + g\frac{h^2}{2}, \quad G(U) = \left(h\frac{u^2}{2} + gh^2 \right) u, \tag{3.1.3}$$

the entropy and entropy fluxes without topography.
For this system we have some existence and stability results [57, 75, 98, 102]. Concerning the approximation of this system, several schemes have been investigated [4, 6, 8, 16, 20, 32, 37, 83] and results dealing with the consistency, the stability or the convergence of those schemes [2, 14, 15, 23, 24, 84].w were established. Notice that the presence of a discontinuous topography induces non-uniqueness of the solution [4, 104].

This paper we give a proof of convergence for the hydrostatic reconstruction scheme [6] with kinetic flux [83]. Our result uses the work [7], that states that the hydrostatic reconstruction scheme, used with the classical kinetic solver, satisfies a fully discrete entropy inequality with an error term. In the case without topography, the error terms vanish and we have the following inequality:

$$\begin{aligned}
\eta(U_i^{n+1}) \leq \eta(U_i^n) &- \frac{\Delta t}{\Delta x}\left(\widetilde{G}_{i+1/2} - \widetilde{G}_{i-1/2}\right) \\
&- \nu_\beta \frac{\Delta t}{\Delta x}\int_{\mathbb{R}} |\xi| \frac{g^2\pi^2}{6}\Big(\mathbb{1}_{\xi<0}\left(M_{i+1/2+} + M_{i+1/2-}\right)\left(M_{i+1/2+} - M_{i+1/2-}\right)^2 \\
&\qquad + \mathbb{1}_{\xi>0}\left(M_{i-1/2+} + M_{i-1/2-}\right)\left(M_{i-1/2+} - M_{i-1/2-}\right)^2 \Big) d\xi.
\end{aligned} \tag{3.1.4}$$

In the time-only discrete case and without topography, this single energy inequality that holds for the kinetic scheme ensures the convergence [15]. The fully-discrete case (still without topography) is treated in [14] and the result is given under the dissipation assumption that F^+ or $-F^-$ (defined in (3.1.25)) are strictly η-dissipative, this notion being defined in [14, 23] . Unfortunately this property does not hold for the scheme we considerer, there is a lack of strong dissipativity of the kinetic scheme. Thus the new contribution of this paper is to give a proof for the convergence in the case of non constant topography, under weaker dissipation assumptions. Let us give here some of the main ideas of our proof. The first step is to establish a weak dissipation property that enables us to prove the convergence. Indeed we get that it is enough that $F^+ - F^-$ is strictly η-dissipative, which corresponds to the following inequality

$$\int_{\mathbb{R}} |\xi| \left(H(M_2) - H(M_1) - \eta'(U_1) \begin{pmatrix} 1 \\ \xi \end{pmatrix} (M_2 - M_1) \right) d\xi$$
$$\geq \alpha \left(\eta(U_2) - \eta(U_1) - \eta'(U_1) (U_2 - U_1) \right). \tag{3.1.5}$$

A rigorous statement of this result can be founded in Lemma 3.6 and we can point out that it is only valid on an open bounded convex set which does not content zero value for the height, and the constant α is not explicit.
In order to go further in the proof we multiply by a time increment Δt the inquality (3.1.4) and sum over indices i and n. Then after doing some re-sum up on the gradient of the approximate solution we are able to use (3.1.5). As a consequence we get an estimate and we conclude by a compensated compactness result. Indeed we recall that the compensated compactness theory [98] gives the compactness on a bounded sequence of approximate (U_ε) solutions of the system which satisfies the properties

$$\partial_t \eta(U_\varepsilon) + \partial_x G(U_\varepsilon) \text{ is compact in } H^{-1}_{loc}, \tag{3.1.6}$$

for a sufficiently large family of entropies η. In our work, we prove an estimate on the gradient of the approximate numerical solutions (U_Δ):

$$\|\partial_t U_\Delta\|_{L^2_{tx}} \leq \frac{C}{\sqrt{\Delta x}}, \quad \|\partial_x U_\Delta\|_{L^2_{tx}} \leq \frac{C}{\sqrt{\Delta x}}, \tag{3.1.7}$$

where Δx is the space increment. This is the key point of the method. Indeed, the estimates (3.1.7) are enough as in Di Perna approximation technique [57] to control the entropy dissipation as (3.1.6).

3.1.1 Kinetic representation

Before going into discretised models, we recall the classical kinetic Maxwellian equilibrium, used in [83] for example, at the continuous level. The kinetic Maxwellian is given by

$$M(U, \xi) = \frac{1}{g\pi} \left(2gh - (\xi - u)^2 \right)_+^{1/2}, \tag{3.1.8}$$

where $\xi \in \mathbb{R}$ and $x_+ \equiv \max(0, x)$ for any $x \in \mathbb{R}$. It satisfies the following moment relation,

$$\int_{\mathbb{R}} \begin{pmatrix} 1 \\ \xi \end{pmatrix} M(U, \xi) d\xi = U. \tag{3.1.9}$$

The interest of this particular form lies in its link with a kinetic entropy. Consider the kinetic entropy,

$$H(f, \xi, z) = \frac{\xi^2}{2} f + \frac{g^2 \pi^2}{6} f^3 + gzf, \tag{3.1.10}$$

where $f \geq 0$, $\xi \in \mathbb{R}$ and $z \in \mathbb{R}$, and its version without topography

$$H_0(f, \xi) = \frac{\xi^2}{2} f + \frac{g^2 \pi^2}{6} f^3. \tag{3.1.11}$$

Then one can check the relations

$$\int_{\mathbb{R}} H(M(U, \xi), \xi, z) d\xi = \eta(U) + ghz, \tag{3.1.12}$$

$$\int_{\mathbb{R}} \xi H(M(U, \xi), \xi, z) d\xi = G(U) + ghzu. \tag{3.1.13}$$

Moreover, for any $f(\xi) \geq 0$, setting $h = \int f(\xi) d\xi$, $hu = \int \xi f(\xi) d\xi$ (assumed finite), one has the following entropy minimization principle [7],

$$\eta(U) = \int_{\mathbb{R}} H_0(M(U, \xi), \xi) d\xi \leq \int_{\mathbb{R}} H_0(f(\xi), \xi) d\xi. \tag{3.1.14}$$

3.1.2 Hydrostatic reconstruction scheme and kinetic flux

We consider a time-step Δt and an uniform grid $(x_{i+1/2})_{i \in \mathbb{Z}}$ with space increment $\Delta x = x_{i+1/2} - x_{i-1/2}$, we set $x_{i-1/2} = i\Delta x$ and $t_n = n\Delta t$. Let $U^0 = (h^0, h^0 u^0)$, $h^0 \geq 0$, $h^0, u^0 \in L^\infty(\mathbb{R})$ and $x \mapsto z(x)$, assumed Lipschitz continuous, be an initial data. We define the discretization of the initial data as

$$U_i^0 = \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} U^0(y) dy, \tag{3.1.15}$$

and

$$z_i \text{ an approximation of } z(x_i), \tag{3.1.16}$$

where $x_i = \left(x_{i+1/2} + x_{i-1/2}\right)/2$. Then the scheme writes

$$U_i^{n+1} = U_i^n - \frac{\Delta t}{\Delta x} \left(F_{i+1/2-} - F_{i-1/2+}\right), \tag{3.1.17}$$

with

$$F_{i+1/2-} = \mathcal{F}(U_{i+1/2-}, U_{i+1/2+}) - S_{i+1/2-}, \tag{3.1.18}$$

$$F_{i-1/2+} = \mathcal{F}(U_{i-1/2-}, U_{i-1/2+}) + S_{i-1/2+}, \tag{3.1.19}$$

with $\mathcal{F}$ is a numerical flux for the system without topography. The source terms $S_{i+1/2-}$, $S_{i-1/2+}$ are defined by

$$S_{i+1/2-} = \begin{pmatrix} 0 \\ g\frac{h_{i+1/2-}^2}{2} - g\frac{h_i^2}{2} \end{pmatrix}, \quad S_{i-1/2+} = \begin{pmatrix} 0 \\ g\frac{h_i^2}{2} - g\frac{h_{i-1/2+}^2}{2} \end{pmatrix}. \tag{3.1.20}$$

The reconstructed states

$$U_{i+1/2-} = (h_{i+1/2-}, h_{i+1/2-}u_i), \quad U_{i+1/2+} = (h_{i+1/2+}, h_{i+1/2+}u_i) \tag{3.1.21}$$

are defined by

$$h_{i+1/2-} = (h_i + z_i - z_{i+1/2})_+, \quad h_{i+1/2+} = (h_{i+1} + z_{i+1} - z_{i+1/2})_+ \tag{3.1.22}$$

and

$$z_{i+1/2} = \max(z_i, z_{i+1}). \tag{3.1.23}$$

We will use in this paper a kinetic numerical flux $\mathcal{F}$ introduced in [83]

$$\mathcal{F}(U_l, U_r) = F^+(U_l) + F^-(U_r), \tag{3.1.24}$$

$$\begin{aligned} F^+(U) &= \int_{\mathbb{R}} \xi \mathbb{1}_{\xi>0} \begin{pmatrix} 1 \\ \xi \end{pmatrix} M(U,\xi)d\xi, \\ F^-(U) &= \int_{\mathbb{R}} \xi \mathbb{1}_{\xi<0} \begin{pmatrix} 1 \\ \xi \end{pmatrix} M(U,\xi)d\xi, \end{aligned} \tag{3.1.25}$$

with $M(U,\xi)$ defined by (3.1.8).
We consider the velocity $v_m \geq 0$ such that for all i,

$$M(U_i, \xi) > 0 \Leftrightarrow |\xi| \leq v_m. \tag{3.1.26}$$

This means equivalently that $|u_i| + \sqrt{2gh_i} \leq v_m$. We consider a CFL condition strictly less than one,

$$v_m \frac{\Delta t}{\Delta x} \leq \beta < 1, \tag{3.1.27}$$

where β is a given constant.

3.1.3 Convergence result

Let (U_i^n, z_i) be the scheme defined by (3.1.15)-(3.1.25). We define the approximate solution, at fixed Δx, by

$$\begin{aligned} &U_\Delta(t,x) \\ &= \frac{1}{\Delta t}\left[\frac{U_{i+1}^{n+1} - U_{i-1}^{n+1} - U_{i+1}^n + U_{i-1}^n}{2\Delta x}(x - x_{i-1/2}) + U_i^{n+1} - U_i^n\right](t - t_n) \\ &\qquad + \frac{U_{i+1}^n - U_{i-1}^n}{2\Delta x}(x - x_{i-1/2}) + U_i^n \\ &\text{for } x_{i-1/2} < x < x_{i+1/2} \text{ and } t_n \leq t < t_{n+1}, \end{aligned} \tag{3.1.28}$$

and we set

$$z_\Delta(x) = \frac{z_{i+1} - z_i}{\Delta x}(x - x_i) + z_i, \qquad \text{for } x_i < x < x_{i+1}. \tag{3.1.29}$$

Moreover, for $h_M > h_m > 0$ and $u_M \geq 0$, we set

$$\mathcal{U}_{h_m,h_M,u_M} = \{(h,u) \in \mathbb{R}^2, \quad h_m \leq h \leq h_M, \quad |u| \leq u_M\} \tag{3.1.30}$$

which is a convex set. We state now the main result of this article, which is the proof of the convergence of the numerical scheme.

Theorem 3.1. *Let $U^0 = (h^0, h^0u^0)$, $h^0 \geq 0$, $h^0, u^0 \in L^\infty(\mathbb{R})$, be an initial data and let $z \in L^\infty(\mathbb{R}) \cap C^1(\mathbb{R})$ and Lipschitz continuous be the given topography. Let (U_i^n, z_i) be the scheme defined by (3.1.15)-(3.1.25). Let U_Δ be the continuous approximate solution to (3.1.1) defined by (3.1.28) and z_Δ the approximate topography defined by (3.1.29). We assume that*

$$1 \leq v^* \frac{\Delta t}{\Delta x}, \tag{3.1.31}$$

for some $v^* > 0$. Then let $h_m > 0$, $h_M > 0$ and $u_M > 0$. We assume that

$$\sup_i(|z_{i+1} - z_i|) < h_m, \tag{3.1.32}$$

which enables us to set $\tilde{h}_m = h_m - \sup_i(|z_{i+1} - z_i|) > 0$, and make the assumption

$$\forall i, n, \quad U_i^n \in \mathcal{U}_{\tilde{h}_m, h_M, u_M}, \tag{3.1.33}$$

with $\mathcal{U}_{h_m,h_M,u_M}$ defined by (3.1.30). Moreover we assume

$$\mathrm{Lip}(z_\Delta) \leq C, \tag{3.1.34}$$

and that

$$z_\Delta \underset{\substack{\Delta t \to 0 \\ \Delta x \to 0}}{\longrightarrow} z, \quad \frac{\mathrm{d}z_\Delta}{\mathrm{d}x} \underset{\substack{\Delta t \to 0 \\ \Delta x \to 0}}{\longrightarrow} \frac{\mathrm{d}z}{\mathrm{d}x}, \text{ uniformly.} \tag{3.1.35}$$

Then, under the CFL condition (3.1.27), up to a subsequence, $U_\Delta \to U$ a.e. in (t, x) and in $C_t([0,T], L^\infty_{x,w*}(\mathbb{R}))$ as $\Delta t \to 0$ and $\Delta x \to 0$ where U is a weak solution to (3.1.1) with initial data U^0 satisfying the entropy condition

$$\partial_t \eta(U) + \partial_x G(U) \in \mathcal{M}_{loc}, \tag{3.1.36}$$

for all suitable couple entropy-entropy flux (η, G) and the inequality (3.1.2).

The outline of the paper is as follows. In Section 3.2, we establish estimates on the gradient of the approximate solution as we mentioned in (3.1.7). In Section 3.3, we introduce some interpolation functions and prove some regularity estimates on the approximate solution. In Section 3.4 we prove Theorem 3.1, first we obtain (3.1.6) by combining the gradient estimate and the regularity estimate, then we complete the proof by applying a compensated compactness result. The appendix is devoted to several technical results.

3.2 Estimate on the gradient of the approximate solution

This section is devoted to the proof of Proposition 3.2.
We have the following estimate on the approximate solution.

Proposition 3.2. *Let $U^0 = (h^0, h^0 u^0)$, $h^0 \geq 0$, $h^0, u^0 \in L^\infty(\mathbb{R})$, be an initial data and let $z \in L^\infty(\mathbb{R}) \cap C^1(\mathbb{R})$ and Lipschitz continuous be the given topography. Let (U_i^n, z_i) be the* *scheme defined by (3.1.15)-(3.1.25). Let U_Δ be the continuous approximate solution to (3.1.1) defined by (3.1.28) and z_Δ the approximate topography defined by (3.1.29). Let $\beta > 0$ and v^*, h_m, h_M, $u_M>0$, involved respectively in assumption (3.1.27) and (3.1.31)-(3.1.34). We define for all $U = (h, hu)$,*

$$|U|^2 = g\frac{h^2}{2} + \frac{u^2h^2}{2h_m}. \tag{3.2.1}$$

Let $N \in \mathbb{N}$, $T = N\Delta t$, i_0, $i_1 \in \mathbb{N}$ such that $i_0 < i_1$. For all $i < j \in \mathbb{N}$, we set

$$I_{i,j}^{v^*} = (x_{i-1/2} - v^*T, x_{j+1/2} + v^*T). \tag{3.2.2}$$

Then there exists some constants C_1, C_2, C_3 such that

$$\sum_{n=0}^{N-1}\sum_{i=i_0}^{i_1-1} \Delta t|U_{i+1}^n - U_i^n|^2 \leq C_1. \tag{3.2.3}$$

$$\sum_{n=0}^{N-1}\sum_{i=i_0}^{i_1-1} \Delta t|U_i^{n+1} - U_i^n|^2 \leq C_1\frac{\Delta t^2}{\Delta x^2}v_m^2\left(1+v_m^2\right), \tag{3.2.4}$$

$$\left(\int_0^T\int_{x_{i_0-1/2}}^{x_{i_1+1/2}} |\partial_x U_\Delta|^2 dxdt\right)^{1/2} \leq \frac{C_2}{\sqrt{\Delta x}}. \tag{3.2.5}$$

$$\left(\int_0^T\int_{x_{i_0-1/2}}^{x_{i_1+1/2}} |\partial_t U_\Delta|^2 dxdt\right)^{1/2} \leq \frac{C_3}{\sqrt{\Delta x}}. \tag{3.2.6}$$

The constants C_1, C_2, C_3 depend only on gravity constant g, h_m, h_M, u_M, v_m, β, on final time T, on $|x_{i_0-1/2} - x_{i_1+1/2}|$, on $\|z\|_{L^\infty}$, $\|\eta(U_0)\|_{L^1(I^{v^*}_{i_0,i_1})}$ and $\|h^0\|_{L^1(I^{v^*}_{i_0,i_1})}$.

We are able to find thoses estimates on $\partial_t U_\Delta$ and $\partial_x U_\Delta$ using recent results on discrete kinetic inequalities founded in [7]. The proof we will be developing is rather technical and we will use several lemmas in Section 3.2.3. We put their demonstrations in the appendix in order to keep clarity of the demonstration.

3.2.1 Bounded propagation on the space integral of the height

Here we found some bound on $\sum_{i=i_0}^{i_1} \Delta x h_i^N$.

$$h_i^{n+1} = h_i^n - \frac{\Delta t}{\Delta x}\left(F_{i+1/2}^h - F_{i-1/2}^h\right), \tag{3.2.7}$$

with

$$F^h_{i+1/2} = \int_{\mathbb{R}} \xi \mathbb{1}_{\xi>0} M(U_i,\xi)d\xi + \int_{\mathbb{R}} \xi \mathbb{1}_{\xi<0} M(U_{i+1},\xi)d\xi. \tag{3.2.8}$$

We multiply by Δx and sum over index i and we obtain

$$\sum_{i=i_0}^{i_1} \Delta x h_i^{n+1} = \sum_{i=i_0}^{i_1} \Delta x h_i^n - \Delta t \left(F^h_{i_1+1/2} - F^h_{i_0-1/2}\right). \tag{3.2.9}$$

Then we notice that

$$-\Delta t F^h_{i_1+1/2} \le \Delta t v_m h_{i_1+1}, \quad \Delta t F^h_{i_0-1/2} \le \Delta t v_m h_{i_0-1}. \tag{3.2.10}$$

With CFL condition (3.1.27) we have

$$\sum_{i=i_0}^{i_1} \Delta x h_i^{n+1} \le \sum_{i=i_0-1}^{i_1+1} \Delta x h_i^n. \tag{3.2.11}$$

Let $N \in \mathbb{N}$ and $T = N\Delta t$, using the previous inequality we get

$$\sum_{i=i_0}^{i_1} \Delta x h_i^N \le \sum_{i=i_0-N}^{i_1+N} \Delta x h_i^0 = \int_{x_{i_0-1/2-N}}^{x_{i_1+1/2+N}} h^0(x)dx. \tag{3.2.12}$$

The last equality holds because $h_i^0 = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} h^0(x)dx$. Moreover we have

$$x_{i_0-1/2-N} = x_{i_0-1/2} - N\Delta x = x_{i_0-1/2} - T\frac{\Delta x}{\Delta t}, \tag{3.2.13}$$

$$x_{i_1+1/2+N} = x_{i_1+1/2} + N\Delta x = x_{i_1+1/2} + T\frac{\Delta x}{\Delta t}. \tag{3.2.14}$$

Therefore by assumption (3.1.31) we get

$$\sum_{i=i_0}^{i_1} \Delta x h_i^N \le \int_{x_{i_0-1/2}-Tv^*}^{x_{i_1+1/2}+Tv^*} h^0(x)dx = \|h^0\|_{L^1(I^{v^*}_{i_0,i_1})}, \tag{3.2.15}$$

with $I^{v^*}_{i_0,i_1}$ defined in (3.2.2).

3.2.2 From kinetic to macroscopic discrete entropy inequality

We use the notations introduced in Proposition 3.2. Using CFL condition (3.1.27) we can use and integrate kinetic entropy inequality [7, Theorem 3.7] with respect to ξ and we obtain

$$\begin{aligned}\eta(U_i^{n+1}) + gz_ih_i^{n+1} \leq \eta(U_i^n) + gz_ih_i - \sigma_i\left(\widetilde{G}_{i+1/2} - \widetilde{G}_{i-1/2}\right)\\
- \nu_\beta \frac{\Delta t}{\Delta x}\int_{\mathbb{R}} |\xi| \frac{g^2\pi^2}{6}\Big(\mathbb{1}_{\xi<0}\left(M_{i+1/2+} + M_{i+1/2-}\right)(M_{i+1/2+} - M_{i+1/2-})^2\\
+ \mathbb{1}_{\xi>0}\left(M_{i-1/2+} + M_{i-1/2-}\right)(M_{i-1/2+} - M_{i-1/2-})^2\Big)d\xi\\
+ C_\beta\left(\frac{\Delta t}{\Delta x}v_m\right)^2 \frac{g^2\pi^2}{6}\int_{\mathbb{R}} M_i\Big((M_i - M_{i+1/2-})^2 + (M_i - M_{i-1/2+})^2\Big)d\xi,\end{aligned} \tag{3.2.16}$$

with

$$\widetilde{G}_{i+1/2} = \int_{\xi<0} \xi H(M_{i+1/2+}, \xi, z_{i+1/2})d\xi + \int_{\xi>0} \xi H(M_{i+1/2-}, \xi, z_{i+1/2})d\xi, \tag{3.2.17}$$

the constant $\nu_\beta > 0$ is a dissipation constant depending only on β, and $C_\beta \geq 0$ is a constant depending only on β. Using (3.1.9) and technical resultat over maxwellian functions (3.5.106), we get that

$$\begin{aligned}\eta(U_i^{n+1}) + gz_ih_i^{n+1} \leq \eta(U_i^n) + gz_ih_i - \sigma_i\left(\widetilde{G}_{i+1/2} - \widetilde{G}_{i-1/2}\right)\\
- \nu_\beta \frac{\Delta t}{\Delta x}\int_{\mathbb{R}} |\xi| \frac{g^2\pi^2}{6}\Big(\mathbb{1}_{\xi<0}\left(M_{i+1/2+} + M_{i+1/2-}\right)(M_{i+1/2+} - M_{i+1/2-})^2\\
+ \mathbb{1}_{\xi>0}\left(M_{i-1/2+} + M_{i-1/2-}\right)(M_{i-1/2+} - M_{i-1/2-})^2\Big)d\xi\\
+ C_\beta\left(\frac{\Delta t}{\Delta x}v_m\right)^2 \Big(g(h_i - h_{i+1/2-})^2 + g(h_i - h_{i-1/2+})^2\Big).\end{aligned} \tag{3.2.18}$$

Using the definition (3.1.22) we get that

$$0 \leq h_i - h_{i+1/2-} \leq |z_{i+1} - z_i|, \tag{3.2.19}$$

$$0 \leq h_i - h_{i-1/2+} \leq |z_i - z_{i-1}|, \tag{3.2.20}$$

and we deduce that

$$\begin{aligned}\eta(U_i^{n+1}) + gz_i h_i^{n+1} \leq \eta(U_i^n) + gz_i h_i - \sigma_i\left(\widetilde{G}_{i+1/2} - \widetilde{G}_{i-1/2}\right)\\ - \nu_\beta \frac{\Delta t}{\Delta x}\int_{\mathbb{R}} |\xi| \frac{g^2\pi^2}{6}\Big(\mathbb{1}_{\xi<0}\left(M_{i+1/2+} + M_{i+1/2-}\right)\left(M_{i+1/2+} - M_{i+1/2-}\right)^2\\ + \mathbb{1}_{\xi>0}\left(M_{i-1/2+} + M_{i-1/2-}\right)\left(M_{i-1/2+} - M_{i-1/2-}\right)^2\Big)d\xi\\ + gC_\beta\left(\frac{\Delta t}{\Delta x}v_m\right)^2\left(|z_{i+1} - z_i|^2 + |z_i - z_{i-1}|^2\right).\end{aligned} \tag{3.2.21}$$

Then we follow the computations over height done in Subsection 3.2.1. Thus we multiply by Δx, take the sum over i and make a translation over index i in order to obtain

$$\begin{aligned}\sum_{i=i_0}^{i_1} \Delta x\left(\eta(U_i^{n+1}) + gz_i h_i^{n+1}\right) \leq \sum_{i=i_0}^{i_1} \Delta x\left(\eta(U_i^n) + \ gz_i h_i\right)\\ - \Delta t\widetilde{G}_{i_1+1/2} + \Delta t\widetilde{G}_{i_0-1/2}\\ - \nu_\beta\Delta t\sum_{i=i_0}^{i_1-1}\int_{\mathbb{R}} |\xi|\frac{g^2\pi^2}{6}\left(M_{i+1/2+} + M_{i+1/2-}\right)\left(M_{i+1/2+} - M_{i+1/2-}\right)^2 d\xi\\ - \nu_\beta\Delta t\int_{\mathbb{R}} |\xi|\mathbb{1}_{\xi<0}\frac{g^2\pi^2}{6}\left(M_{i_1+1/2+} + M_{i_1+1/2-}\right)\left(M_{i_1+1/2+} - M_{i_1+1/2-}\right)^2 d\xi\\ - \nu_\beta\Delta t\int_{\mathbb{R}} |\xi|\mathbb{1}_{\xi>0}\frac{g^2\pi^2}{6}\left(M_{i_0-1/2+} + M_{i+1/2-}\right)\left(M_{i_0-1/2+} - M_{i_0-1/2-}\right)^2 d\xi\\ + \sum_{i=i_0-1}^{i_1} 2gC_\beta\frac{\Delta t^2}{\Delta x}v_m\left|z_{i+1} - z_i\right|^2.\end{aligned} \tag{3.2.22}$$

We notice according to (3.2.17) that we have

$$-\Delta t\widetilde{G}_{i_1+1/2} \leq v_m\Delta t\eta(U_{i_1+1/2+}) + v_m\Delta t g h_{i+1/2+}z_{i+1/2}, \tag{3.2.23}$$

and

$$-\Delta t\widetilde{G}_{i_0-1/2} \leq v_m\Delta t\eta(U_{i_0-1/2-}) + v_m\Delta t g h_{i-1/2-}z_{i-1/2}, \tag{3.2.24}$$

with (3.1.27), $h_{i_1+1/2+} \leq h_{i+1}$ and $|z_{i_1+1/2} - z_{i_1+1}| \leq |z_{i_1+1} - z_{i_1}|$, it leads to

$$-\Delta t\widetilde{G}_{i_1+1/2} \leq \Delta x\eta(U_{i_1+1}) + v_m\Delta t g h_{i_1+1}z_{i_1+1} + gh_M\Delta t|z_{i_1+1} - z_{i_1}|, \tag{3.2.25}$$

and similarly we get

$$-\Delta t\widetilde{G}_{i_0-1/2} \leq \Delta x\eta(U_{i_0-1}) + v_m\Delta tgh_{i_0-1}z_{i_0-1} + gh_M\Delta t|z_{i_0} - z_{i_0-1}|. \tag{3.2.26}$$

From (3.2.22), noticing that the last two integrals are nonpositive and using (3.2.25),(3.2.26), we obtain

$$\begin{aligned}\sum_{i=i_0}^{i_1} \Delta x \left(\eta(U_i^{n+1}) + gz_ih_i^{n+1}\right) \leq \sum_{i=i_0-1}^{i_1+1} \Delta x \left(\eta(U_i^n) + \ gz_ih_i\right)\\ - \nu_\beta\Delta t\sum_{i=i_0}^{i_1-1}\int_{\mathbb{R}}|\xi|\frac{g^2\pi^2}{6}\left(M_{i+1/2+} + M_{i+1/2-}\right)(M_{i+1/2+} - M_{i+1/2-})^2 d\xi\\ + gh_M\Delta t|z_{i_0} - z_{i_0-1}| + gh_M\Delta t|z_{i_1+1} - z_{i_1}| + \sum_{i=i_0}^{i_1} 2gC_\beta\frac{\Delta t^2}{\Delta x}v_m\left|z_{i+1} - z_i\right|^2.\end{aligned} \tag{3.2.27}$$

Next, we sum now over index n and we use that $T = N\Delta t$ and that, by assumption (3.1.34) and (3.1.29), we have

$$|z_{i+1} - z_i| \leq C\Delta x, \tag{3.2.28}$$

and therefore we get

$$gTh_M|z_{i_0} - z_{i_0-1}| + gTh_M|z_{i_1+1} - z_{i_1}| + \sum_{n=0}^{N-1}\sum_{i=i_0}^{i_1} 2gC_\beta\frac{\Delta t^2}{\Delta x}v_m\left|z_{i+1} - z_i\right|^2 \leq C. \tag{3.2.29}$$

Thus we get

$$\begin{aligned}\sum_{i=i_0}^{i_1} \Delta x \left(\eta(U_i^N) + gz_ih_i^N\right)\\ + \nu_\beta\sum_{n=0}^{N-1}\Delta t\sum_{i=i_0}^{i_1-1}\int_{\mathbb{R}}|\xi|\frac{g^2\pi^2}{6}\left(M_{i+1/2+} + M_{i+1/2-}\right)(M_{i+1/2+} - M_{i+1/2-})^2 d\xi\\ \leq \sum_{i=i_0-N}^{i_1+N} \Delta x \left(\eta(U_i^0) + \ gz_ih_i^0\right) + C\end{aligned} \tag{3.2.30}$$

with C depending on g, T, h_M, β, v_m, $|x_{i_0-1/2} - x_{i_1+1/2}|$. Now we will see that the integral in LHS of (3.2.30) is underestimated by a term proportionnal to $\sum_{n=0}^{N-1}\sum_{i=i_0-1}^{i_1+1}\Delta t|U_{i+1/2+} - U_{i-1/2+}|^2$.

3.2.3 Lower estimate of dissipation terms

First we notice that

$$\begin{aligned}&\int_{\mathbb{R}} |\xi| \frac{g^2\pi^2}{6}\left(M_{i+1/2+}+M_{i+1/2-}\right)(M_{i+1/2+}-M_{i+1/2-})^2 d\xi\\ &\geq \frac{1}{2}\int_{\mathbb{R}} |\xi| \frac{g^2\pi^2}{6}\left(2M_{i+1/2+}+M_{i+1/2-}\right)(M_{i+1/2+}-M_{i+1/2-})^2 d\xi. \end{aligned} \tag{3.2.31}$$

Now using Lemma 3.8, we obtain that there exists some $C>0$ depending only on g, h_m, h_M, u_M such that

$$\begin{aligned}&\int_{\mathbb{R}} |\xi| \frac{g^2\pi^2}{6}(2M_1+M_2)(M_1-M_2)^2 d\xi\\ \geq & C\left(g\frac{(h_2-h_1)^2}{2}+h_m\frac{(u_2-u_1)^2}{2}\right)\end{aligned} \tag{3.2.32}$$

for every U_1 , $U_2 \in \mathfrak{U}_{h_m,h_M,u_M}$.
Next we notice using the definitions (3.1.21) and the assumption (3.1.33) we get that

$$U_{i+1/2+}, U_{i+1/2-} \in \mathfrak{U}_{h_m,h_M,u_M}. \tag{3.2.33}$$

Thus from (3.2.31) and applying the last estimate (3.2.32) with $U_1 = U_{i+1/2+}$ and $U_2 = U_{i+1/2-}$, there exists some constant $C>0$ depending only on g, h_m, h_M, u_M such that

$$\begin{aligned}\int_{\mathbb{R}} |\xi| \frac{g^2\pi^2}{6}\left(M_{i+1/2+}+M_{i+1/2-}\right)(M_{i+1/2+}-M_{i+1/2-})^2 d\xi&\\ \geq C|U_{i+1/2+}-U_{i+1/2-}|^2&\end{aligned} \tag{3.2.34}$$

where $|\cdot|$ is defined in (3.2.1).

3.2.4 Estimate of the discrete gradient

Now we use (3.2.34) in (3.2.30) and we get

$$\begin{aligned}&\nu_\beta C\sum_{n=0}^{N-1}\sum_{i=i_0}^{i_1-1}\Delta t|U_{i+1/2+}-U_{i+1/2-}|^2\\ &\qquad\leq \sum_{i=i_0-N}^{i_1+N}\Delta x\left(\eta(U_i^0)+\ gz_ih_i\right)-\sum_{i=i_0}^{i_1}\Delta x\left(\eta(U_i^N)+gz_ih_i^N\right)+C.\end{aligned} \tag{3.2.35}$$

with C depending on g, T, h_M, β, v_m, $|x_{i_0-1/2} - x_{i_1+1/2}|$.
Then we notice that $\eta(U_i^N) \geq 0$ and we get

$$\nu_\beta C \sum_{n=0}^{N-1} \sum_{i=i_0}^{i=i_1-1} \Delta t |U_{i+1/2+} - U_{i+1/2-}|^2$$
$$\leq \sum_{i=i_0-N}^{i_1+N} \Delta x \left(\eta(U_i^0) + \ gz_i h_i\right) + \sum_{i=i_0}^{i_1} \Delta x \left(-gz_i h_i^N\right) + C. \quad (3.2.36)$$

Next, using (3.1.16) and the fact we assumed $z \in L^\infty(\mathbb{R})$ we get $\forall i,\ z_i \leq \|z\|_\infty$, we have

$$gz_i h_i^0 \leq \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} g\|z\|_\infty h^0(x)\ dx, \quad (3.2.37)$$

Moreover, by convexity of $(h, hu) \mapsto \eta(U)$, we have

$$\eta(U_i^0) \leq \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \eta(U^0(x))\ dx. \quad (3.2.38)$$

Combining the last two results and summing over n we get

$$\sum_{i=i_0-N}^{i_1+N} \Delta x \left(\eta(U_i^0) + \ gz_i h_i^0\right) \qquad \leq \int_{x_{i_0-1/2-N}}^{x_{i_1+1/2+N}} \eta\left(U^0(x)\right) + g\|z\|_\infty h^0(x)\ dx. \quad (3.2.39)$$

One notice that $x_{i_0-1/2-N} = x_{i_0} - N\Delta x = x_{i_0} - T\frac{\Delta x}{\Delta t}$ and by finite propagation hypothesis (3.1.31) one deduce that

$$\sum_{i=i_0-N}^{i_1+N} \Delta x \left(\eta(U_i^0) + \ gz_i h_i^0\right) \leq \int_{x_{i_0-1/2-v^*T}}^{x_{i_1+1/2+v^*T}} \eta\left(U^0(x)\right) + g\|z\|_\infty h^0(x)\ dx \quad (3.2.40)$$
$$= \|\eta(U_i^0)\|_{L^1(I^{v^*}_{i_0,i_1})} + g\|z\|_\infty \|h^0\|_{L^1(I^{v^*}_{i_0,i_1})}, \quad (3.2.41)$$

with $I^{v^*}_{i_0,i_1}$ defined in (3.2.2). In addition, by preliminary computation (3.2.15), we have

$$\sum_{i=i_0}^{i_1} \Delta x \left(-gz_i h_i^N\right) \leq g\|z\|_\infty \sum_{i=i_0}^{i_1} \Delta x h_i^N \leq g\|z\|_\infty \|h^0\|_{L^1(I^{v^*}_{i_0,i_1})}. \quad (3.2.42)$$

Using together (3.2.40), (3.2.42) and (3.2.29) in (3.2.36), we get

$$\sum_{n=0}^{N-1} \sum_{i=i_0}^{i_1-1} \Delta t |U_{i+1/2+} - U_{i+1/2-}|^2 \leq C \quad (3.2.43)$$

where C depends on g, h_m, h_M, u_M, v_m, β, T, on $|x_{i_0-1/2} - x_{i_1+1/2}|$, $\|z\|_{L^\infty}$, $\|\eta(U_0)\|_{L^1(I^{v^*}_{i_0,i_1})}$ and $\|h^0\|_{L^1(I^{v^*}_{i_0,i_1})}$. Moreover using triangle inequality and (3.1.21)-(3.1.23),(3.2.1)(3.2.20),(3.2.19) there exist some absolute constant C such that

$$\begin{aligned}
&|U_{i+1} - U_i|^2 \\
\leq & C|U_{i+1/2+} - U_{i+1/2-}|^2 + C|U_{i+1/2+} - U_{i+1}|^2 + C|U_{i+1/2-} - U_i|^2, \\
\leq & C|U_{i+1/2+} - U_{i+1/2-}|^2 + C|z_{i+1} - z_i|^2 + C|z_i - z_{i-1}|^2, \\
\leq & C|U_{i+1/2+} - U_{i+1/2-}|^2 + 2C\Delta x^2.
\end{aligned} \tag{3.2.44}$$

Last inequality holds because of (3.2.28). With (3.2.43), we get (3.2.3) of Proposition 3.2. In addition, using (3.1.17), (3.5.112), (3.5.113) and (3.2.3), we get (3.2.4) of Proposition 3.2.

3.2.5 End of the proof of Proposition 3.2: estimate the gradient of the approximate solution

Now from (3.1.28) we compute

$$\partial_x U_\Delta = \frac{t - t_n}{\Delta t} \frac{U^{n+1}_{i+1} - U^{n+1}_{i-1} - U^n_{i+1} + U^n_{i-1}}{2\Delta x} + \frac{U^n_{i+1} - U^n_{i-1}}{2\Delta x} \tag{3.2.45}$$

and using the triangle inequality we obtain that

$$\begin{aligned}
&\int_{t_n}^{t_{n+1}} \int_{x_{i-1/2}}^{x_{i+1/2}} |\partial_x U_\Delta|^2 dx dt \\
\leq & C\frac{\Delta t}{\Delta x} \left[|U^{n+1}_{i+1} - U^{n+1}_i|^2 + |U^{n+1}_i - U^{n+1}_{i-1}|^2 + |U^n_{i+1} - U^n_i|^2 + |U^n_i - U^n_{i-1}|^2 \right].
\end{aligned} \tag{3.2.46}$$

with $C > 0$ an absolute constant. In consequence, by using (3.2.3) we get (3.2.5) by summing over i and n. Similarly, from (3.1.28) we compute

$$\partial_t U_\Delta = \frac{1}{\Delta t} \left[\frac{U^{n+1}_{i+1} - U^{n+1}_{i-1} - U^n_{i+1} + U^n_{i-1}}{2\Delta x} (x - x_{i-1/2}) + U^{n+1}_i - U^n_i \right] \tag{3.2.47}$$

thus

$$\begin{aligned}
&\int_{t_n}^{t_{n+1}} \int_{x_{i-1/2}}^{x_{i+1/2}} |\partial_t U_\Delta|^2 dxdt \\
&\leq \frac{\Delta x}{\Delta t} \left[|U_{i+1}^{n+1} - U_{i+1}^n|^2 + |U_{i-1}^n - U_{i-1}^{n+1}|^2 + |U_i^{n+1} - U_i^n|^2 \right] \\
&= \frac{1}{\Delta x} \cdot \frac{\Delta x^2}{\Delta t} \left[|U_{i+1}^{n+1} - U_{i+1}^n|^2 + |U_{i-1}^n - U_{i-1}^{n+1}|^2 + |U_i^{n+1} - U_i^n|^2 \right]
\end{aligned} \tag{3.2.48}$$

in consequence, by using (3.2.4) we get (3.2.6) by summing over i and n.
This concludes the proof of Proposition 3.2

3.3 Regularity estimates

Before going into the proof of Theorem 3.1 , we give some regularity estimate.

3.3.1 Definition of interpolation functions $\widetilde{U}_\Delta$ and $\widetilde{F}_\Delta$

We define $\widetilde{U}_\Delta(t)$ a piecewise linear function by

$$\widetilde{U}_\Delta(t) = U_i - \frac{t - t_n}{\Delta x} \left(F_{i+1/2-} - F_{i-1/2+} \right) \tag{3.3.1}$$

for $t_n \leq t < t_{n+1}$, with $F_{i+1/2-}$, $F_{i-1/2+}$ defined in (3.1.18).
Let us remark that

$$\forall n \in [|0, N|] \,, \quad \widetilde{U}_\Delta(t^n) = U_i^n, \qquad \lim_{\substack{t \to t^{n+1} \\ t < t^{n+1}}} \widetilde{U}_\Delta(t) = U_i^{n+1} = \widetilde{U}_\Delta(t^{n+1}) \tag{3.3.2}$$

i.e. $\widetilde{U}_\Delta(t)$ is continuous.
We also define $\widetilde{F}_\Delta(x) \in C(\mathbb{R})$ by

$$\begin{aligned}
\widetilde{F}_\Delta(x) = &\frac{x - x_{i-1/2}}{\Delta x} \left(F^+(U_{i+1/2-}) + F^-(U_{i+1/2+}) \right) \\
&+ \frac{x_{i+1/2} - x}{\Delta x} \left(F^+(U_{i-1/2-}) + F^-(U_{i-1/2+}) \right)
\end{aligned} \tag{3.3.3}$$

for $x_{i-1/2} \leq x < x_{i+1/2}$, with F^+, F^- defined in (3.1.25), $U_{i-1/2-}$, $U_{i-1/2+}$ defined in (3.1.21).
Let us remark that

$$\forall i \in \mathbb{Z}, \quad \widetilde{F}_\Delta(x_{i-1/2}) = F\left(U_{i-1/2-}, U_{i-1/2+} \right). \tag{3.3.4}$$

and

$$\lim_{\substack{x\to x_{i+1/2}\\ x<x_{i+1/2}}} \widetilde{F}_\Delta(x) = F\left(U_{i+1/2-}, U_{i+1/2+}\right) = \widetilde{F}_\Delta(x_{i+1/2}) \tag{3.3.5}$$

They satisfy a partial differential equation

$$\partial_t \widetilde{U}_\Delta + \partial_x \widetilde{F}_\Delta = \widetilde{S}_\Delta \tag{3.3.6}$$

with

$$\widetilde{S}_\Delta(t,x) = \frac{1}{\Delta x}\left(S_{i+1/2-} + S_{i-1/2+}\right) \tag{3.3.7}$$

for $t_n \le t \le t_{n+1}$ and $x_{i-1/2} \le x \le x_{i+1/2}$, with $h_{i+1/2-}$, $h_{i+1/2+}$ defined in (3.1.22) and $S_{i+1/2-}$, $S_{i+1/2+}$ defined in (3.1.20).

3.3.2 Estimate of $\int_0^T \int_{x_{i_0-1/2}}^{x_{i_1+1/2}} |U_\Delta - \widetilde{U}_\Delta|^2 \, dtdx$

We will see later on that, in order to prove compactness of the sequence $\partial_t \eta(U_\Delta) + \partial_x G(U_\Delta)$ in a convenient space, we will need an estimate on $U_\Delta - \widetilde{U}_\Delta$. It is the following proposition:

Lemma 3.3. *Let $\beta > 0$ and v^*, h_m, h_M, u_M>0, involved respectively in assumption (3.1.27) and (3.1.31)-(3.1.34). Let $N \in \mathbb{N}$, $T = N\Delta t$, i_0, $i_1 \in \mathbb{N}$ such that $i_0 < i_1$. Let U_Δ be the approximate solution (3.1.28) and $\widetilde{U}_\Delta$ defined by (3.3.1). Then*

$$\left(\int_0^T \int_{x_{i_0-1/2}}^{x_{i_1+1/2}} |U_\Delta - \widetilde{U}_\Delta|^2 \, dtdx\right)^{\frac{1}{2}} \le C\sqrt{\Delta x} \tag{3.3.8}$$

with $|\cdot|$ defined by (3.2.1). The constant C depends only on g, h_m, h_M, u_M, v_m, β, T, $|x_{i_0-3/2} - x_{i_1+1/2}|$, $\|z\|_{L^\infty}$, $\|\eta(U_0)\|_{L^1(I^{v^*}_{i_0-1,i_1+1})}$ and $\|h^0\|_{L^1(I^{v^*}_{i_0-1,i_1+1})}$, $I^{v^*}_{i_0-1,i_1+1}$ defined in (3.2.2).

Proof. On the one hand we use (3.1.28), the definition of U_Δ, and we write

$$\begin{aligned} &U_\Delta - U_i^n \\ =&\frac{1}{\Delta t}\left[\frac{U_{i+1}^{n+1} - U_{i-1}^{n+1} - U_{i+1}^n + U_{i-1}^n}{2\Delta x}(x - x_{i-1/2}) + U_i^{n+1} - U_i^n\right](t - t_n) \\ &+ \frac{U_{i+1}^n - U_{i-1}^n}{2\Delta x}(x - x_{i-1/2}), \end{aligned} \tag{3.3.9}$$

for all $x_{i-1/2} < x < x_{i+1/2}$ and $t_n \le t < t_{n+1}$.

Using the triangle inequality, we obtain

$$\begin{aligned}|U_\Delta - U_i^n| \leq \frac{1}{2}|U_{i+1}^{n+1} - U_i^{n+1}| + \frac{1}{2}|U_i^{n+1} - U_{i-1}^{n+1}| \\ +\frac{1}{2}|U_{i+1}^{n} - U_i^{n}| + \frac{1}{2}|U_i^{n} - U_{i-1}^{n}| + |U_i^{n+1} - U_i^n| \\ +\frac{1}{2}|U_{i+1}^{n} - U_i^{n}| + \frac{1}{2}|U_i^{n} - U_{i-1}^{n}|.\end{aligned} \tag{3.3.10}$$

Thus,

$$\begin{aligned}\int_{t_n}^{t_{n+1}} \int_{x_{i-1/2}}^{x_{i+1/2}} |U_\Delta - U_i^n|^2 dxdt \leq C_1 \Delta t \Delta x \Big(|U_{i+1}^{n+1} - U_i^{n+1}|^2 + |U_i^{n+1} - U_{i-1}^{n+1}|^2 \\ +|U_{i+1}^{n} - U_i^{n}|^2 + |U_i^{n} - U_{i-1}^{n}|^2 + |U_i^{n+1} - U_i^n| \Big).\end{aligned} \tag{3.3.11}$$

with $C_1 > 0$ an absolute constant.
Next, we set

$$U_\Delta^1(t,x) = U_i^n, \tag{3.3.12}$$

for $x_{i-1/2} < x < x_{i+1/2}, t^n < t < t^{n+1}$. Now, taking the sum over n and i and making substitutions of indices, we get

$$\begin{aligned}\int_0^T \int_{x_{i_0-1/2}}^{x_{i_1+1/2}} |U_\Delta - U_\Delta^1|^2 dxdt \leq 4C_1 \Delta x \sum_{n=0}^{N+1} \sum_{i=i_0-1}^{i_1+1} \Delta t |U_{i+1}^n - U_i^n|^2 \\ +C_1 \Delta x \sum_{n=0}^{N+1} \sum_{i=i_0-1}^{i_1+1} \Delta t |U_i^{n+1} - U_i^n|^2.\end{aligned} \tag{3.3.13}$$

Then we use the discrete gradient estimates (3.2.3), (3.2.4) and CFL condition (3.1.27) in order to get

$$\int_0^T \int_{x_{i_0-1/2}}^{x_{i_1+1/2}} |U_\Delta - U_\Delta^1|^2 dxdt \leq C_2 \Delta x, \tag{3.3.14}$$

with C_2 a constant depending on g, h_m, h_M, u_M, v_m, β, T, $|x_{i_0-3/2} - x_{i_1+1/2}|$, $\|z\|_{L^\infty}$, $\|\eta(U_0)\|_{L^1(I^{v*}_{i_0-1,i_1+1})}$ and $\|h^0\|_{L^1(I^{v*}_{i_0-1,i_1+1})}$.

On the other hand we use (3.3.1), the definition of $\widetilde{U}_\Delta$, and we get

$$\begin{aligned} &U_i^n - \widetilde{U}_\Delta \\ &= \frac{t-t_n}{\Delta x}\Bigg(F^+(U_{i+1/2-}) + F^-(U_{i+1/2+}) - F^+(U_{i-1/2-}) - F^-(U_{i-1/2+}) \\ &\qquad - \frac{g}{2}\begin{pmatrix} 0 \\ h_i^2 - h_{i+1/2-}^2 - (h_{i+1}^2 - h_{i+1/2+}^2) \end{pmatrix}\Bigg), \end{aligned} \tag{3.3.15}$$

for all $t_n \leq t < t_{n+1}$, with F^+, F^- defined in (3.1.25), $U_{i+1/2-}$, $U_{i+1/2+}$ defined in (3.1.21), $h_{i+1/2+}$, $h_{i+1/2-}$ defined in (3.1.22).

Then, using that F^+ and F^- are Lipschitz continuous, see (3.5.112) and (3.5.113), with the CFL condition (3.1.27) we obtain that there exists $C_3 > 0$, depending on g, h_M, u_M and v_m such that

$$\begin{aligned} &|U_i^n - \widetilde{U}_\Delta| \\ \leq &C_3\Bigg(|U_{i+1/2-} - U_{i-1/2-}| + |U_{i+1/2+} - U_{i-1/2+}| \\ &\qquad + \frac{g}{2}\left|h_i^2 - h_{i+1/2-}^2 - (h_{i+1}^2 - h_{i+1/2+}^2)\right|\Bigg), \end{aligned} \tag{3.3.16}$$

for all $t_n \leq t < t_{n+1}$.

Moreover using (3.1.21)-(3.1.23), (3.2.1), (3.2.20), (3.2.19), we get that there exists $C > 0$, depending only on g and h_m such that

$$|U_{i+1/2-} - U_{i-1/2-}| \leq C\left(|h_i - h_{i-1}| + |z_{i+1} - z_i| + |z_i - z_{i-1}|\right), \tag{3.3.17}$$

$$|U_{i+1/2+} - U_{i-1/2+}| \leq C\left(|h_{i+1} - h_i| + |z_{i+1} - z_i| + |z_i - z_{i-1}|\right). \tag{3.3.18}$$

In addition, (3.5.133), (3.5.133),we deduce that there exists C, depending on h_M such that

$$\frac{g}{2}\left|h_i^2 - h_{i+1/2-}^2 - (h_{i+1}^2 - h_{i+1/2+}^2)\right| \leq C|z_{i+1} - z_i| + C|z_i - z_{i-1}|. \tag{3.3.19}$$

Thus, from (3.3.16) using the triangle inequality with (3.3.17), (3.3.18) and (3.3.19), there exists

C depending on g, h_m, h_M such that

$$\begin{aligned}&\int_{t_n}^{t_{n+1}}\int_{x_{i-1/2}}^{x_{i+1/2}}|U_i^n-\widetilde{U}_\Delta|^2dtdx\\ \le\ & C\Delta t\Delta x\left(|h_i-h_{i-1}|^2+|h_{i+1}-h_i|^2+|z_{i+1}-z_i|^2+|z_i-z_{i-1}|^2\right).\end{aligned}\tag{3.3.20}$$

Now, taking the sum over n and i and making substitutions of indices, we get

$$\begin{aligned}&\int_0^T\int_{x_{i_0-1/2}}^{x_{i_1+1/2}}|U_\Delta^1-\widetilde{U}_\Delta|^2dxdt\\ \le\ & C\Delta x\left(\sum_{n=0}^{N+1}\sum_{i=i_0-1}^{i_1+1}\Delta t|U_{i+1}^n-U_i^n|^2+\sum_{n=0}^{N+1}\sum_{i=i_0-1}^{i_1+1}\Delta t|z_{i+1}-z_i|^2\right),\end{aligned}\tag{3.3.21}$$

with U_Δ^1 defined in (3.3.12) and C is a constant depending on g, h_m, h_M. Next, using (3.2.28) and the gradient estimate (3.2.3), we get

$$\int_0^T\int_{x_{i_0-1/2}}^{x_{i_1+1/2}}|U_\Delta^1-\widetilde{U}_\Delta|^2dxdt\le C_2\Delta x,\tag{3.3.22}$$

with C_2 a constant depending on g, h_m, h_M, u_M, v_m, β, T, $|x_{i_0-3/2}-x_{i_1+1/2}|$, $\|z\|_{L^\infty}$, $\|\eta(U_0)\|_{L^1(I^{v^*}_{i_0-1,i_1+1})}$ and $\|h^0\|_{L^1(I^{v^*}_{i_0-1,i_1+1})}$.

Finally, noticing that $U_\Delta-\widetilde{U}_\Delta=\left(U_\Delta-U_\Delta^1\right)+\left(U_\Delta^1-\widetilde{U}_\Delta\right)$, we get

$$\begin{aligned}&\int_0^T\int_{x_{i_0-1/2}}^{x_{i_1+1/2}}|U_\Delta-\widetilde{U}_\Delta|^2dtdx\\ \le\ & 2\left(\int_0^T\int_{x_{i_0-1/2}}^{x_{i_1+1/2}}|U_\Delta-U_\Delta^1|^2dtdx+\int_0^T\int_{x_{i_0-1/2}}^{x_{i_1+1/2}}|U_\Delta^1-\widetilde{U}_\Delta|^2dtdx\right).\end{aligned}\tag{3.3.23}$$

With (3.3.14), (3.3.22) we get (3.3.8), which concludes the proof. $\square$

3.3.3 Estimate of $\int_0^T\int_{x_{i_0-1/2}}^{x_{i_1+1/2}}|F(U_\Delta)-\widetilde{F}_\Delta|^2\ dtdx$

We will see later on that, in order to prove compactness of the sequence $\partial_t\eta(U_\Delta)+\partial_x G(U_\Delta)$ in H^{-1}_{loc}, we will need an estimate on $F(U_\Delta)-\widetilde{F}_\Delta$. It is the following proposition:

Lemma 3.4. *Let $\beta>0$ and v^*, h_m, h_M, $u_M>0$, involved respectively in assumption (3.1.27) and (3.1.31)-(3.1.34). Let $N\in\mathbb{N}$, $T=N\Delta t$, i_0, $i_1\in\mathbb{N}$ such that $i_0<i_1$. Let U_Δ be the*

approximate solution (3.1.28) *and* $\widetilde{F}_\Delta$ *defined by* (3.3.3). *Then*

$$\int_0^T \int_{x_{i_0-1/2}}^{x_{i_1+1/2}} |F(U_\Delta) - \widetilde{F}_\Delta|^2 \, dtdx \leq C\Delta x \tag{3.3.24}$$

with $|\cdot|$ *defined by* (3.2.1). *The constant* C *depends only on* g, h_m, h_M, u_M, v_m, β, T, $|x_{i_0-3/2} - x_{i_1+3/2}|$, $\|z\|_{L^\infty}$, $\|\eta(U_0)\|_{L^1(I^{v^*}_{i_0-1,i_1+1})}$ *and* $\|h^0\|_{L^1(I^{v^*}_{i_0-1,i_1+1})}$, $I^{v^*}_{i_0-1,i_1+1}$ *defined in* (3.2.2).

Proof. We recall here (3.3.3)

$$\begin{aligned}\widetilde{F}_\Delta(x) = {}& \frac{x - x_{i-1/2}}{\Delta x}\left(F^+(U_{i+1/2-}) + F^-(U_{i+1/2+})\right) \\ &+ \frac{x_{i+1/2} - x}{\Delta x}\left(F^+(U_{i-1/2-}) + F^-(U_{i-1/2+})\right),\end{aligned} \tag{3.3.25}$$

for all $x_{i-1/2} < x < x_{i+1/2}$. Moreover, we have

$$F(U_\Delta) = F^+(U_\Delta) + F^-(U_\Delta). \tag{3.3.26}$$

Thus, using triangle inequality, for all $x_{i-1/2} < x < x_{i+1/2}$, we get

$$\begin{aligned}&|\widetilde{F}_\Delta(x) - F(U_\Delta)| \\ &\leq \frac{1}{2}\left|F^+(U_{i+1/2-}) - F^+(U_\Delta)\right| + \frac{1}{2}\left|F^-(U_{i+1/2+}) - F^-(U_\Delta)\right| \\ &\quad + \frac{1}{2}\left|F^+(U_{i-1/2-}) - F^+(U_\Delta)\right| + \frac{1}{2}\left|F^-(U_{i+1/2+}) - F^-(U_\Delta)\right|.\end{aligned} \tag{3.3.27}$$

Then, using that F^+ and F^- are Lipschitz continuous, see (3.5.112) and (3.5.113), with the CFL condition (3.1.27) we obtain that there exists $C > 0$, depending on g, h_m, h_M, u_M and v_m such that

$$\begin{aligned}&|\widetilde{F}_\Delta(x) - F(U_\Delta)| \\ &\leq C\left(\left|U_{i+1/2-} - U_\Delta\right| + \left|U_{i+1/2+} - U_\Delta\right| + \left|U_{i-1/2-} - U_\Delta\right| + \left|U_{i-1/2+} - U_\Delta\right|\right)\end{aligned} \tag{3.3.28}$$

Moreover using (3.1.21), (3.2.19), (3.2.20), we get

$$\begin{aligned}&|\widetilde{F}_\Delta(x) - F(U_\Delta)| \\ &\leq C\left(2\left|U_i - U_\Delta\right| + \left|U_{i+1} - U_\Delta\right| + \left|U_{i-1} - U_\Delta\right| + \left|z_{i+1} - z_i\right| + \left|z_i - z_{i-1}\right|\right)\end{aligned} \tag{3.3.29}$$

with $C > 0$, depending on g, h_m, h_M, u_M and v_m.

Thus we get

$$\begin{aligned}
&\int_{t_n}^{t_{n+1}}\int_{x_{i-1/2}}^{x_{i+1/2}} |\widetilde{F}_\Delta(x) - F(U_\Delta)|^2 dtdx \\
&\leq C\int_{t_n}^{t_{n+1}}\int_{x_{i-1/2}}^{x_{i+1/2}} |U_i - U_\Delta|^2\, dtdx \\
+C\Delta t\Delta x &\left(|U_{i+1}-U_i|^2 + |U_{i-1}-U_i|^2 + |z_{i+1}-z_i|^2 + |z_i - z_{i-1}|^2\right),
\end{aligned} \tag{3.3.30}$$

with $C > 0$ an absolute contant.

Now, taking the sum over n and i and making substitutions of indices, we get

$$\begin{aligned}
\int_0^T\int_{x_{i_0-1/2}}^{x_{i_1+1/2}} |\widetilde{F}_\Delta(x) - F(U_\Delta)|^2 dxdt &\leq \int_0^T\int_{x_{i_0-1/2}}^{x_{i_1+1/2}} |U_\Delta - U_\Delta^1|^2 dxdt \\
+C\Delta x\left(\sum_{n=0}^{N+1}\sum_{i=i_0-1}^{i_1+1}\Delta t|U_{i+1}^n - U_i^n|^2 + \sum_{n=0}^{N+1}\sum_{i=i_0-1}^{i_1+1}\Delta t|z_{i+1}-z_i|^2\right).
\end{aligned} \tag{3.3.31}$$

Finally, using (3.2.28), the gradient estimate (3.2.3) and previous estimation (3.3.14), involving $U_\Delta - U_i^n$, we get (3.3.24), which concludes the proof. □

3.4 Proof of Theorem 3.1

Using (3.4.1) we compute

$$\partial_t U_\Delta + \partial_x F(U_\Delta) = \partial_t(U_\Delta - \widetilde{U}_\Delta) + \partial_x\left(F(U_\Delta) - \widetilde{F}_\Delta\right) + \widetilde{S}_\Delta, \tag{3.4.1}$$

with $U_\Delta(t,x)$ defined in (3.1.28). We multiply (3.4.1) by $\eta'(U_\Delta)$ and we get, for any entropy-entropy flux (η, G), the following decomposition

$$\begin{aligned}
\partial_t\eta(U_\Delta) + \partial_x G(U_\Delta) = \;& \eta'(U_\Delta)\cdot\partial_t(U_\Delta - \widetilde{U}_\Delta) \\
&+\eta'(U_\Delta)\cdot\partial_x\left(F(U_\Delta) - \widetilde{F}_\Delta\right) \\
&+\;\eta'(U_\Delta)\cdot\widetilde{S}_\Delta \\
&= R_1 + M_1 + R_2 + M_2 - \eta'(U_\Delta)\cdot\widetilde{S}_\Delta
\end{aligned} \tag{3.4.2}$$

where

$$\begin{aligned} R_1 &= \partial_t \left(\eta'(U_\Delta)\cdot(U_\Delta - \widetilde{U}_\Delta)\right), \\ M_1 &= -\eta''(U_\Delta)\cdot\partial_t U_\Delta\cdot\left(U_\Delta - \widetilde{U}_\Delta\right), \\ R_2 &= \partial_x \left(\eta'(U_\Delta)\cdot\left(F(U_\Delta) - \widetilde{F}_\Delta\right)\right), \\ M_2 &= -\eta''(U_\Delta)\cdot\partial_x U_\Delta\cdot\left(F(U_\Delta) - \widetilde{F}_\Delta\right). \end{aligned} \tag{3.4.3}$$

First we have, using (3.3.24)

$$\begin{aligned} &\int_0^T \int_{-R}^R \left|\eta'(U_\Delta)\cdot\left(F(U_\Delta) - \widetilde{F}_\Delta\right)\right|^2 dxdt \\ &\leq \|\eta'(U_\Delta)\|_{L^\infty(]0,T[\times]-R,R[)} \int_0^T \int_{-R}^R \left|\left(F(U_\Delta) - \widetilde{F}_\Delta\right)\right|^2 dxdt \\ &\leq C_R\sqrt{\Delta x} \end{aligned} \tag{3.4.4}$$

thus R_2 goes to zero in H_{loc}^-1 as $\Delta x \to 0$. Similarly, using (3.3.8), R_1 goes to zero in H_{loc}^-1 as $\Delta x \to 0$.

Futhermore, using (3.2.5) and (3.3.24), we have

$$\begin{aligned} &\int_0^T \int_{-R}^R |M_2| dxdt \\ &\leq \|\eta''(U_\Delta)\|_{L^\infty(]0,T[\times]-R,R[)} \left(\iint |\partial_t U_\Delta|^2 dxdt\right)^{1/2} \left(\int_0^T \int_{-R}^R \left|\left(F(U_\Delta) - \widetilde{F}_\Delta\right)\right|^2 dxdt\right)^{1/2} \\ &\leq \|\eta''(U_\Delta)\|_{L^\infty(]0,T[\times]-R,R[)} \frac{C_1}{\sqrt{\Delta x}} C_3\sqrt{\Delta x} \\ &\leq C_R \end{aligned} \tag{3.4.5}$$

thus M_2 is bounded in $\mathcal{M}_{loc}((0,T)\times\mathbb{R})$. Similarly, using (3.2.6) and (3.3.8), M_1 is bounded in $\mathcal{M}_{loc}((0,T)\times\mathbb{R})$.

Using (3.5.133), if $z_{i+1} - z_i \geq 0$, we have

$$\frac{|S_{i+1/2-}|}{\Delta x} \leq C\mathrm{Lip}(z_\Delta)\left(\|U_\Delta\|_\infty + \Delta x \mathrm{Lip}(z)\right), \tag{3.4.6}$$

and if not, $h_i = h_{i+1/2-}$ and the last inequality holds.
Similarly, using (3.5.134), if $z_{i+1} - z_i \geq 0$, we have

$$\frac{|S_{i-1/2+}|}{\Delta x} \leq C\text{Lip}(z_\Delta)\left(\|U_\Delta\|_\infty + \Delta x \text{Lip}(z)\right), \tag{3.4.7}$$

and if not, $h_i = h_{i-1/2+}$ and the last inequality holds. Using (3.4.6), (3.4.7), we get

$$\|\widetilde{S}_\Delta\|_\infty \leq C\text{Lip}(z_\Delta)\left(\|U_\Delta\|_\infty + \Delta x \text{Lip}(z)\right). \tag{3.4.8}$$

Moreover, according to (3.4.2) and (3.4.3), one has

$$\partial_t \eta(U_\Delta) + \partial_x G(U_\Delta) - R_1 - R_2 = M_1 + M_2 + \eta'(U_\Delta) \cdot \widetilde{S}_\Delta \tag{3.4.9}$$

thus $M_1 + M_2 + \eta'(U_\Delta) \cdot \widetilde{S}_\Delta$ is bounded in $W^{-1,p}_{loc} \cap \mathcal{M}_{loc}$, $\forall p$, $1 < p < +\infty$, as a consequence it is compact in H^{-1}_{loc}. At this point, we know that $R_1 + R_2$ and $M_1 + M_2 + \eta'(U_\Delta) \cdot \widetilde{S}_\Delta$ are compact in H^{-1}_{loc}, therefore their sum, which is equal to $\partial_t \eta(U_\Delta) + \partial_x G(U_\Delta)$, is compact in H^{-1}_{loc}. Furthermore, $(U_\Delta)_{\Delta > 0}$ is bounded since we assume that $(U_i^n)_{i,n}$ is a bounded sequence. We are now able to apply the compensated compactness method and we get that up to a subsequence $U_\Delta \to U$ a.e. and in $L^1_{loc,t,x}$ as $\Delta t \to 0$ and $\Delta x \to 0$, see [75].
Moreover, according to Lemma 3.11, $\partial_t U_\Delta$ is bounded in $L^\infty_t(\mathcal{D}'_x)$ and therefore we get

$$d\left((U_\Delta(t_1), U_\Delta(t_2)\right)_{(W^{1,1})'} \leq C\|\partial_t U_\Delta\|_{L^\infty_t(\mathcal{D}'_x)}|t_1 - t_2|, \tag{3.4.10}$$

and we conclude that $U_\Delta \to U$ in $C_t([0,T], L^\infty_{x,w*}(\mathbb{R}))$, Then, knowing that U_Δ converges in L^p_{loc} to U, we can apply Lemma 3.12, which concludes the convergence of the approximate source term $\widetilde{S}_\Delta$ to S.
Finally, we pass to the limit in (3.4.1) using (3.3.8), (3.3.24), and (3.5.131), which enables us to get that the limit U is a solution to the system. Moreover passing to the limit with a test function ϕ in (3.2.16) we get (3.1.2).
This ends the proof of Theorem 3.1.

3.5 Appendix

We prove here some technical results used throughout the paper.

Lemma 3.5. *Let $Uk = (h_k, h_k u_k)$ for $k = 1,\ 2$ with $h_k \geq 0$. Then*

$$\begin{aligned}
&\frac{g^2\pi^2}{6} \left(2M_1 + M_2\right) \left(M_1 - M_2\right)^2 \\
&= H(M_2) - H(M_1) - \eta'(U_1) \begin{pmatrix} 1 \\ \xi \end{pmatrix} (M_2 - M_1) \\
&- \mathbb{1}_{(\xi - u_1)^2 > 2gh_1}\ M_2 \left(\frac{(\xi - u_1)^2}{2} - gh_1 \right),
\end{aligned} \tag{3.5.1}$$

where $M_k \equiv M_k(\xi) \equiv M(U_k, \xi)$, and $M(U, \xi)$ is defined in (3.1.8) and $H(f) \equiv H(f, \xi)$ is defined in (3.1.8).

Proof. Using the identity

$$b^3 - a^3 - 3a^2(b - a) = (b + 2a)(b - a)^2, \tag{3.5.2}$$

one has

$$\frac{g^2\pi^2}{6} \left(2M_1 + M_2\right) \left(M_1 - M_2\right)^2 = H(M_2) - H(M_1) - H'(M_1) \left(M_2 - M_1\right), \tag{3.5.3}$$

where we donote $H'(f, \xi) \equiv \frac{\partial}{\partial f} H(f, \xi)$. Thus it is sufficient to prove

$$\begin{aligned}
&\left(\eta'(U_1) \begin{pmatrix} 1 \\ \xi \end{pmatrix} - H'(M_2) \right) (M_2 - M_1) \\
&= -\mathbb{1}_{(\xi - u_1)^2 > 2gh_1}\ M_2 \left(\frac{(\xi - u_1)^2}{2} - gh_1 \right).
\end{aligned} \tag{3.5.4}$$

On the one hand we compute

$$\eta'(U_1) \begin{pmatrix} 1 \\ \xi \end{pmatrix} = \left(gh_1 - \frac{u_1^2}{2} + u_1 \xi \right). \tag{3.5.5}$$

On the other hand we get

$$\begin{aligned}
H'(M_1) &= \frac{\xi^2}{2} + \frac{g^2\pi^2}{2} M_1^2 \\
&= \frac{\xi^2}{2} + \left(gh_1 - \frac{(\xi - u_1)^2}{2} \right)_+.
\end{aligned} \tag{3.5.6}$$

In consequence, by adding (3.5.5) and (3.5.6) we get

$$\eta'(U_1)\begin{pmatrix}1\\ \xi\end{pmatrix} - H'(M_2) = -\mathbb{1}_{(\xi-u_1)^2>2gh_1}\left(\frac{(\xi-u_1)^2}{2} - gh_1\right). \tag{3.5.7}$$

and therefore

$$\begin{aligned}&\left(\eta'(U_1)\begin{pmatrix}1\\ \xi\end{pmatrix} - H'(M_2)\right)(M_2-M_1)\\ &= -\,\mathbb{1}_{(\xi-u_1)^2>2gh_1}\left(\frac{(\xi-u_1)^2}{2} - gh_1\right)(M_2-M_1).\end{aligned} \tag{3.5.8}$$

Finally we notice that

$$(\xi-u_1)^2 > 2gh_1 \iff M_1 = 0 \tag{3.5.9}$$

and we get (3.5.4), which concludes the proof. □

Lemma 3.6. *There exists some constant $\alpha > 0$, depending only on on gravity constant g, on constants h_m, h_M, u_M, which are involved in (3.1.30), such that*

$$\begin{aligned}&\int_{\mathbb{R}} |\xi|\left(H(M_2) - H(M_1) - \eta'(U_1)\begin{pmatrix}1\\ \xi\end{pmatrix}(M_2-M_1)\right)d\xi\\ &\geq \alpha\left(\eta(U_2) - \eta(U_1) - \eta'(U_1)(U_2-U_1)\right)\end{aligned} \tag{3.5.10}$$

for every U_1, $U_2 \in \mathfrak{U}_{h_m,h_M,u_M}$ defined by (3.1.30) and where $M_k \equiv M_k(\xi) \equiv M(U_k,\xi)$, with $M(U,\xi)$ defined in (3.1.8), $H(f) \equiv H(f,\xi)$ is defined in (3.1.8) and $\eta(U)$ defined in (3.1.3).

Proof. We set

$$\widehat{\mathfrak{U}_m} = \left\{(h,hu) \in \mathbb{R}^2,\ h \geq h_m\right\} \tag{3.5.11}$$

and we first deal with the case

$$U_1 = \begin{pmatrix}h_1\\ h_1u_1\end{pmatrix} \text{ and } U_2 = \begin{pmatrix}h_2\\ h_2u_2\end{pmatrix} \in \widehat{\mathfrak{U}_m}, \text{ such that } |u_1-u_2| \leq \sqrt{gh_m}. \tag{3.5.12}$$

In this case we have

$$\forall t \in [0,1],\ (1-t)\eta'(U_1) + t\eta'(U_2) \in \eta'(\widetilde{\mathfrak{U}_m}). \tag{3.5.13}$$

with

$$\widetilde{\mathcal{U}_m} = \left\{(h, hu) \in \mathbb{R}^2,\ h \geq \frac{h_m}{2}\right\}. \tag{3.5.14}$$

Indeed we notice that

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \in \eta'(\widetilde{\mathcal{U}_m}) \Longleftrightarrow V_1 \geq g\frac{h_m}{2} - \frac{V_2^2}{2}. \tag{3.5.15}$$

Thus (3.5.13) is equivalent to

$$\begin{aligned} &\forall t \in [0,1], \forall h_1, h_2 \geq h_m, \forall u_1, u_2 \in \mathbb{R}, \text{ such that } |u_1 - u_2| \leq \sqrt{gh_m} \\ &(1-t)\left(gh_1 - \frac{u_1^2}{2}\right) + t\left(gh_2 - \frac{u_2^2}{2}\right) \geq g\frac{h_m}{2} - \frac{1}{2}\Big((1-t)u_1 + tu_2\Big)^2. \end{aligned} \tag{3.5.16}$$

Thus it is sufficient to check that

$$\begin{aligned} &\forall t \in [0,1], \forall u_1, u_2 \in \mathbb{R}, \text{ such that } |u_1 - u_2| \leq \sqrt{gh_m} \\ &(1-t)\left(gh_m - \frac{u_1^2}{2}\right) + t\left(gh_m - \frac{u_2^2}{2}\right) \geq g\frac{h_m}{2} - \frac{1}{2}\Big((1-t)u_1 + tu_2\Big)^2. \end{aligned} \tag{3.5.17}$$

This inequality simplifies to

$$\forall t \in [0,1], \forall u_1, u_2 \in \mathbb{R} \quad \frac{gh_m}{2} \geq \frac{t(1-t)}{2}(u_1 - u_2)^2 \tag{3.5.18}$$

which is true if $|u_1 - u_2| \leq 2\sqrt{gh_m}$.
We want now to use property (3.5.13) and define a path $v(t) \in \widetilde{\mathcal{U}_m}$, connecting two states U_1, U_2 satisfying (3.5.12) by

$$\eta'(v(t)) = (1-t)\eta'(U_1) + t\eta'(U_2) \tag{3.5.19}$$

for $0 \leq t \leq 1$. Such a definition is possible because η is strictly convex and η' is a diffeormorphism. It enables us to set

$$\begin{aligned} \phi(t) = \int_{\mathbb{R}} |\xi| \Big(& H(M(v(t),\xi)) - H(M(U_1,\xi)) \\ & - \eta'(U_1)\begin{pmatrix} 1 \\ \xi \end{pmatrix}(M(v(t),\xi) - M(U_1,\xi))\Big) d\xi \\ & - \alpha\left(\eta(v(t)) - \eta(U_1) - \eta'(U_1)(v(t) - U_1)\right). \end{aligned} \tag{3.5.20}$$

We notice that $\phi(0) = 0$, and the result of (3.5.10) is equivalent to $\phi(1) \geq 0$. Thus it is sufficient to prove that ϕ is non-decreasing. Using the fact that

$$\eta'(U) \begin{pmatrix} 1 \\ \xi \end{pmatrix} = H'\left(M(U,\xi),\xi\right), \text{ for all } \xi \in \mathbb{R} \text{ such that } M(U,\xi) > 0, \tag{3.5.21}$$

we can compute

$$\begin{aligned} \phi'(t) = \int_{\mathbb{R}} |\xi| \left(\eta'(v(t)) - \eta'(U_1)\right) \begin{pmatrix} 1 \\ \xi \end{pmatrix} M'(v(t),\xi) \cdot v'(t) d\xi \\ - \left(\eta'(v(t)) - \eta'(U_1)\right) \cdot v'(t). \end{aligned} \tag{3.5.22}$$

Moreover using

$$\eta'(v(t)) - \eta'(U_1) = t\eta''(v(t)) \cdot v'(t) \tag{3.5.23}$$

we get

$$\begin{aligned} \phi'(t) = t \int_{\mathbb{R}} |\xi| \eta''(v(t)) \cdot v'(t) \cdot \begin{pmatrix} 1 \\ \xi \end{pmatrix} M'(v(t),\xi) \cdot v'(t) d\xi \\ - t\,\eta''(v(t)) \cdot v'(t) \cdot v'(t), \end{aligned} \tag{3.5.24}$$

which can be rewritten as

$$\begin{aligned} &\phi'(t) \\ &= -t \int_{\mathbb{R}} |\xi| \left(M'(v(t),\xi)^t \begin{pmatrix} 1 \\ \xi \end{pmatrix}^t \eta''(v(t)) \right) \cdot v'(t) \cdot v'(t) d\xi \\ &\quad - t\,\eta''(v(t)) \cdot v'(t) \cdot v'(t) \\ &= -t \int_{\mathbb{R}} |\xi| M'(v(t),\xi) \otimes \left(\eta''(v(t)) \begin{pmatrix} 1 \\ \xi \end{pmatrix} \right) \cdot v'(t) \cdot v'(t) d\xi \\ &\quad - t\,\eta''(v(t)) \cdot v'(t) \cdot v'(t). \end{aligned} \tag{3.5.25}$$

Thus now it is sufficient for getting (3.5.10) to prove that

$$
\begin{aligned}
&\forall U \in \widetilde{\mathfrak{U}_m}, \forall X \in \mathbb{R}^2 \\
&\int_{\mathbb{R}} |\xi| M'(U,\xi) \otimes \left(\eta''(U) \begin{pmatrix} 1 \\ \xi \end{pmatrix}\right) \cdot X \cdot X d\xi \geq \alpha\, \eta''(U) \cdot X \cdot X.
\end{aligned}
\tag{3.5.26}
$$

For all $U \in \widetilde{\mathfrak{U}_m}$ and $\xi \in \mathbb{R}$ such that $M(U,\xi) > 0$, we compute

$$
\eta'(U) \begin{pmatrix} 1 \\ \xi \end{pmatrix} = H'\left(M(U,\xi), \xi\right)
\tag{3.5.27}
$$

and

$$
\eta''(U) \begin{pmatrix} 1 \\ \xi \end{pmatrix} = H''\left(M(U,\xi), \xi\right) M'(U).
\tag{3.5.28}
$$

Moreover one can check that

$$
H''\left(M(U,\xi)\right) = g^2\pi^2 M(U,\xi)
\tag{3.5.29}
$$

and we obtain

$$
\begin{aligned}
&\int_{\mathbb{R}} |\xi| M'(U,\xi) \otimes \left(\eta''(U) \begin{pmatrix} 1 \\ \xi \end{pmatrix}\right) d\xi \\
=&g^2\pi^2 \int_{M(U,\xi)>0} |\xi|\; M(U,\xi) \cdot M'(U,\xi) \otimes M'(U,\xi) d\xi.
\end{aligned}
\tag{3.5.30}
$$

which is equivalent to

$$
\begin{aligned}
&\int_{\mathbb{R}} |\xi| M'(U,\xi) \otimes (\eta''(U)X) \cdot X \cdot X d\xi \\
=&g^2\pi^2 \int_{M(U,\xi)>0} |\xi|\; M(U,\xi) \left(M'(U,\xi) \cdot X\right)^2 d\xi
\end{aligned}
\tag{3.5.31}
$$

for all $X \in \mathbb{R}^2$.

Now we denote $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and compute $M'(U,\xi) \cdot X$. We recall that

$$
M(U,\xi) = \frac{1}{g\pi} \left(2gh - (\xi - u)^2\right)_+^{1/2}, \quad U = (h, hu)
\tag{3.5.32}
$$

so we can rewrite

$$M(U,\xi) = \frac{1}{g\pi}\left(2gh - \left(\xi - \frac{hu}{h}\right)^2\right)_+^{1/2} \tag{3.5.33}$$

and compute partial derivatives

$$\partial_h M(U,\xi) = \frac{1}{2g\pi}\left(2gh - (\xi-u)^2\right)^{-1/2}\left(2g - 2\frac{u}{h}(\xi-u)\right) \tag{3.5.34}$$

and

$$\partial_{hu} M(U,\xi) = \frac{1}{2g\pi}\left(2gh - (\xi-u)^2\right)^{-1/2}\left(\frac{2}{h}(\xi-u)\right). \tag{3.5.35}$$

Finally it leads to the formula

$$M'(U,\xi)\cdot X = \frac{M(U,\xi)^{-1}}{g^2\pi^2}\left(gx_1 + \frac{(\xi-u)}{h}(x_2 - ux_1)\right). \tag{3.5.36}$$

Now we denote

$$x_3 = \frac{1}{h}(x_2 - ux_1) \tag{3.5.37}$$

in order to write

$$M'(U,\xi)\cdot X = \frac{M(U,\xi)^{-1}}{g^2\pi^2}\left(gx_1 + (\xi-u)x_3\right) \tag{3.5.38}$$

and using (3.5.38) and (3.5.31) we get

$$\begin{aligned}
&\int_{\mathbb{R}} |\xi| M'(U,\xi) \otimes \left(\eta''(U)\begin{pmatrix}1\\ \xi\end{pmatrix}\right)\cdot X\cdot X d\xi\\
&= g^2\pi^2 \int_{M(U,\xi)>0} |\xi|\; M(U,\xi)\left(M'(U,\xi)\cdot X\right)^2 d\xi\\
&= \frac{1}{g^2\pi^2}\int_{M(U,\xi)>0}|\xi|\;\frac{1}{M(U,\xi)}\left(gx_1 + (\xi-u)x_3\right)^2 d\xi\\
&\geq \frac{1}{g\pi\sqrt{2gh}}\int_{|\xi-u|\leq\sqrt{2gh}}|\xi|\left(gx_1+(\xi-u)x_3\right)^2 d\xi := I.
\end{aligned} \tag{3.5.39}$$

Last estimate In order to get the we used the fact that

$$M(U,\xi) = \frac{1}{g\pi}\left(2gh-(\xi-u)^2\right)_+^{1/2} \leq \frac{\sqrt{2gh}}{g\pi}. \tag{3.5.40}$$

Using the substitution $v = \xi - u$ and using the convention that

$$\text{if } u = 0 \text{ then } \operatorname{sgn}(u) = 1 \tag{3.5.41}$$

we obtain

$$I = \frac{1}{g\pi\sqrt{2gh}} \int_{|v|\le\sqrt{2gh}} |v+u| \left(gx_1 + vx_3\right)^2 dv \tag{3.5.42}$$

$$\ge \frac{1}{g\pi\sqrt{2gh}} \int_{|v|\le\sqrt{2gh}, \mathrm{sgn}(v)=\mathrm{sgn}(u)} (|v|+|u|) \left(gx_1 + v\,\mathrm{sgn}(u)x_3\right)^2 dv \tag{3.5.43}$$

$$\ge \frac{1}{g\pi\sqrt{2gh}} \int_0^{\sqrt{2gh}} v \left(gx_1 + v\,\mathrm{sgn}(u)x_3\right)^2 dv \tag{3.5.44}$$

$$\ge \frac{1}{2g\pi} \int_{\frac{\sqrt{2gh}}{2}}^{\sqrt{2gh}} \left(gx_1 + v\,\mathrm{sgn}(u)x_3\right)^2 dv. \tag{3.5.45}$$

Using the substitution $\xi = \frac{v}{\sqrt{2gh}}$ we obtain

$$\frac{1}{2g\pi} \int_{\frac{\sqrt{2gh}}{2}}^{\sqrt{2gh}} \left(gx_1 + v\,\mathrm{sgn}(u)x_3\right)^2 dv$$
$$= \frac{\sqrt{h}}{\sqrt{2g}\pi} \int_{1/2}^{1} \left(gx_1 + \xi\sqrt{2gh}\,\mathrm{sgn}(u)x_3\right)^2 d\xi \tag{3.5.46}$$

which is a positive definite quadratic form with respect to $y_1 = gx_1$ and $y_3 = \mathrm{sgn}(u)\sqrt{gh}x_3$. Thus we have for some absolute constant $C > 0$

$$\frac{\sqrt{h}}{\sqrt{2g}\pi} \int_{1/2}^{1} \left(gx_1 + \xi\sqrt{2gh}\,\mathrm{sgn}(u)x_3\right)^2 d\xi$$
$$\ge C\frac{\sqrt{h}}{\sqrt{2g}\pi} \left((gx_1)^2 + 2ghx_3^2\right)$$
$$\ge C\frac{\sqrt{h}}{\sqrt{2g}\pi} \left((gx_1)^2 + \frac{2g}{h}\left(x_2 - ux_1\right)^2\right)$$
$$= C\frac{\sqrt{gh}}{\sqrt{2}\pi} \left(gx_1^2 + \frac{2}{h}\left(x_2 - ux_1\right)^2\right) \tag{3.5.47}$$

and by (3.5.39) (3.5.46) (3.5.47), we get

$$\int_{\mathbb{R}} |\xi| M'(U,\xi) \otimes \left(\eta''(U) \begin{pmatrix} 1 \\ \xi \end{pmatrix}\right) \cdot X \cdot X d\xi$$
$$\ge C\frac{\sqrt{gh}}{\sqrt{2}\pi} \left(gx_1^2 + \frac{2}{h}\left(x_2 - ux_1\right)^2\right). \tag{3.5.48}$$

Besides, we have

$$\eta(h,q) = \left(\frac{1}{2}\frac{q^2}{h} + g\frac{h^2}{2}\right), \tag{3.5.49}$$

$$\eta'(h,q) = \left(-\frac{1}{2}\frac{q^2}{h^2} + gh, \frac{q}{h}\right), \tag{3.5.50}$$

$$\eta''(h,q) = \begin{pmatrix} \frac{q^2}{h^3} + g & -\frac{q}{h^2} \\ -\frac{q}{h^2} & \frac{1}{h} \end{pmatrix}, \tag{3.5.51}$$

$$\eta''(h,hu) = \begin{pmatrix} \frac{u^2}{h} + g & -\frac{u}{h} \\ -\frac{u}{h} & \frac{1}{h} \end{pmatrix}, \tag{3.5.52}$$

and finally we get

$$\begin{aligned} &\eta''(U)\cdot X\cdot X \\ &= \left(\left(g+\frac{u^2}{h}\right)x_1^2 + \frac{1}{h}x_2^2 - \frac{2u}{h}x_1x_2\right) \\ &= gx_1^2 + \frac{1}{h}\left(x_2 - ux_1\right)^2. \end{aligned} \tag{3.5.53}$$

Thus we find that

$$\int_{\mathbb{R}} |\xi| M'(U) \otimes \left(\eta''(U)\begin{pmatrix}1\\ \xi\end{pmatrix}\right) \geq \frac{C}{\sqrt{2\pi}}\sqrt{gh_m}\eta''(U) \tag{3.5.54}$$

At this point, with the last estimate we obtain that (3.5.26) holds, and therefore we have the result (3.5.10) for all $U_1, U_2 \in \widehat{\mathfrak{U}_m}$ such that $|u_1 - u_2| \leq \sqrt{gh_m}$, with the constant $\alpha_m = \frac{C}{\sqrt{2\pi}}\sqrt{gh_m}$, where $C > 0$ is an absolute constant.
Thus, it is now sufficient to prove that

$$\begin{aligned} &\exists \alpha_1 > 0, \quad \forall U_1, U_2 \in \mathfrak{U}_{h_m,h_M,u_M}, \\ &|u_1 - u_2| > \sqrt{gh_m} \\ &\Rightarrow \int_{\mathbb{R}} |\xi| \left(H(M_2) - H(M_1) - \eta'(U_1)\begin{pmatrix}1\\ \xi\end{pmatrix}(M_2 - M_1)\right) d\xi \geq \alpha_1. \end{aligned} \tag{3.5.55}$$

Indeed, when (3.5.55) holds, we have

$$\begin{aligned}
&(\eta(U_2) - \eta(U_1) - \eta'(U_1)\,(U_2 - U_1)) \\
&= g\frac{(h_2-h_1)^2}{2} + h_2\frac{(u_2-u_1)^2}{2} \\
&\leq C(h_M, u_M) \\
&\leq \frac{C(h_M,u_M)}{\alpha_1}\int_{\mathbb{R}} |\xi| \left(H(M_2) - H(M_1) - \eta'(U_1)\begin{pmatrix}1\\ \xi\end{pmatrix}(M_2 - M_1)\right) d\xi
\end{aligned} \tag{3.5.56}$$

which proves (3.5.10). Using reductio ad absurdum, we suppose (3.5.55) does not hold. Thus

$$\begin{aligned}
&\forall n > 0, \quad \exists U_1^n, U_2^n \in \mathfrak{U}_{h_m,h_M,u_M}, \text{ such that} \\
&|u_1^n - u_2^n| > \sqrt{gh_m} \\
&\text{and } \int_{\mathbb{R}} |\xi| \left(H(M_2^n) - H(M_1^n) - \eta'(U_1^n)\begin{pmatrix}1\\ \xi\end{pmatrix}(M_2^n - M_1^n)\right) d\xi \leq \frac{1}{n}
\end{aligned} \tag{3.5.57}$$

where $M_i^n = M(U_i^n, \xi)$.
As $\mathfrak{U}_{h_m,h_M,u_M}$ is a closed and bounded set, we can take 2 subsequences which we also denote U_1^n, U_2^n such that

$$U_1^n \to U_1 \in, \quad U_2^n \to U_2 \in \mathfrak{U}_m \tag{3.5.58}$$

with

$$|u_1 - u_2| \geq \sqrt{gh_m} \tag{3.5.59}$$

and by dominated converge theorem

$$\int_{\mathbb{R}} |\xi| \left(H(M_2) - H(M_1) - \eta'(U_1)\begin{pmatrix}1\\ \xi\end{pmatrix}(M_2 - M_1)\right) d\xi = 0. \tag{3.5.60}$$

We also know by (3.5.1) that

$$\begin{aligned}
&\int_{\mathbb{R}} |\xi| \left(H(M_2) - H(M_1) - \eta'(U_1)\begin{pmatrix}1\\ \xi\end{pmatrix}(M_2 - M_1)\right) d\xi \\
&\geq \int_{\mathbb{R}} |\xi|\,(2M_1 + M_2)\,(M_1 - M_2)^2 d\xi
\end{aligned} \tag{3.5.61}$$

and therefore we get

$$(2M_1 + M_2)(M_1 - M_2)^2 = 0 \quad \text{almost everywhere} \tag{3.5.62}$$

itself implying that $M_1 = M_2$ a.e. and therefore $U_1 = U_2$, the later being in contradiction with (3.5.59). ☐

Lemma 3.7. *Let $g > 0$ be the gravity constant. One has*

$$\int_{(\xi-u_1)^2>2gh_1} |\xi| M(U_2,\xi)\left(\frac{(\xi-u_1)^2}{2} - gh_1\right) d\xi$$
$$\leq \frac{4\left(|u_2| + \sqrt{2gh_2}\right)}{g\pi\sqrt{gh_2}}\left(g|h_1 - h_2| + (|u_2| + \sqrt{2gh_2})|u_1 - u_2| + \frac{1}{2}|u_1^2 - u_2^2|\right)^{\frac{5}{2}} \tag{3.5.63}$$

for every $U_1 = (h_1, h_1u_1)$, $h_1 > 0$ and $U_2 = (h_2, h_2u_2)$, $h_2 > 0$, where $M(U,\xi)$ is defined in (3.1.8).

Proof. We set

$$K = g|h_1 - h_2| + (|u_2| + \sqrt{2gh_2})|u_1 - u_2| + \frac{1}{2}|u_1^2 - u_2^2|. \tag{3.5.64}$$

Thus we can rewrite (3.5.63) as

$$\int_{(\xi-u_1)^2>2gh_1} |\xi| M(U_2,\xi)\left(\frac{(\xi-u_1)^2}{2} - gh_1\right) d\xi$$
$$\leq \frac{4}{g\pi\sqrt{gh_2}}(|u_2| + \sqrt{2gh_2})K^{\frac{5}{2}} \tag{3.5.65}$$

for every $U_1, U_2 \in \mathfrak{U}_{h_m,h_M,u_M}$ defined by (3.1.30).
We notice that $\forall \xi \in \mathrm{supp}(M_2)$, one has $|\xi| \leq |u_2| + \sqrt{2gh_2}$ and we get

$$\begin{aligned}
&\left| gh_2 - \frac{(\xi-u_2)^2}{2} - \left(gh_1 - \frac{(\xi-u_1)^2}{2}\right)\right| \\
&= \left| g(h_2 - h_1) + \xi(u_2 - u_1) - \frac{1}{2}(u_2^2 - u_1^2)\right| \\
&\leq K
\end{aligned} \tag{3.5.66}$$

with K defined by (3.5.64). Moreover using that $\xi \in \mathrm{supp}(M_1)^c \cap \mathrm{supp}(M_2)$ iff $gh_2 - \frac{(\xi-u_2)^2}{2} \geq 0$

and $\frac{(\xi-u_1)^2}{2} - gh_1 \geq 0$, we get

$$
\begin{aligned}
0 &\leq gh_2 - \frac{(\xi-u_2)^2}{2} \\
&\leq gh_2 - \frac{(\xi-u_2)^2}{2} + \frac{(\xi-u_1)^2}{2} - gh_1 \\
&= gh_2 - \frac{(\xi-u_2)^2}{2} - \left(gh_1 - \frac{(\xi-u_1)^2}{2}\right) \\
&\leq K.
\end{aligned} \tag{3.5.67}
$$

Similarly we obtain

$$
\begin{aligned}
0 \leq & \ \frac{(\xi-u_1)^2}{2} - gh_1 \\
\leq & \ K.
\end{aligned} \tag{3.5.68}
$$

Finally using (3.1.8), (3.5.67) and (3.5.68), we get

$$
\begin{aligned}
& \int_{(\xi-u_1)^2>2gh_1} |\xi| M(U_2,\xi) \left(\frac{(\xi-u_1)^2}{2} - gh_1\right) d\xi \\
&= \frac{\sqrt{2}}{g\pi} \int_{(\xi-u_1)^2>2gh_1} |\xi| \left(gh_2 - \frac{(\xi-u_2)^2}{2}\right)^{1/2} \left(\frac{(\xi-u_1)^2}{2} - gh_1\right) d\xi \\
&\leq \frac{\sqrt{2}}{g\pi}(|u_2| + \sqrt{2gh_2}) \left|\mathrm{supp}(M_1)^c \cap \mathrm{supp}(M_2)\right| K^{3/2}.
\end{aligned} \tag{3.5.69}
$$

Thus it is now sufficient for getting (3.5.65) to prove that

$$
\left|\mathrm{supp}(M_1)^c \cap \mathrm{supp}(M_2)\right| \leq \frac{4K}{\sqrt{2gh_2}}. \tag{3.5.70}
$$

Moreover from (3.5.67) one has for $\xi \in \mathrm{supp}(M_1)^c \cap \mathrm{supp}(M_2)$ that

$$
P(\xi) \leq 0 \tag{3.5.71}
$$

where

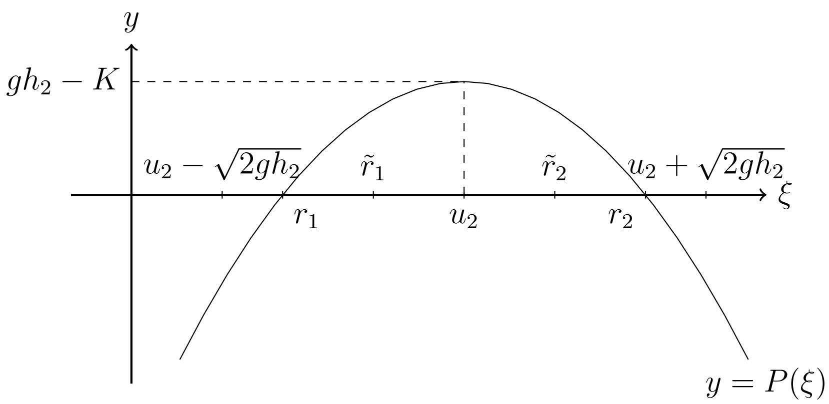
$$
P(\xi) = gh_2 - \frac{(\xi-u_2)^2}{2} - K. \tag{3.5.72}
$$

We notice that when $\xi = u_2$, P reaches a maximum equals to $gh_2 - K$, and we distinguish:

- if $K < gh_2$, then the maximum of P is positive and using (3.5.71) we get that for $\xi \in \mathrm{supp}(M_1)^c \cap \mathrm{supp}(M_2)$ we have

$$\xi \in \left[u_2 - \sqrt{2gh_2}, r_1\right] \bigcup \left[r_2, u_2 + \sqrt{2gh_2}\right] \tag{3.5.73}$$

with $r_1 < u < r_2 \in \mathbb{R}$ are such that $P(r_1) = P(r_2) = 0$, we have $u_2 - \sqrt{2gh_2} < r_1$ because $P(u_2 - \sqrt{2gh_2}) = -K < 0$ and $r_2 < u_2 + \sqrt{2gh_2}$ because $P(u_2 + \sqrt{2gh_2}) = -K < 0$. This configuration is illustrated in the following picture.



Graph of $\xi \mapsto P(\xi)$ when $K < gh_2$

Thus

$$|\,\mathrm{supp}(M_1)^c \cap \mathrm{supp}(M_2)| \leq \left| r_1 - \left(u_2 - \sqrt{2gh_2}\right)\right| + \left|u_2 + \sqrt{2gh_2} - r_2\right|. \tag{3.5.74}$$

We set

$$\tilde{r}_1 = u_2 - \sqrt{2gh_2} + \frac{2K}{\sqrt{2gh_2}} \tag{3.5.75}$$

and we notice that

$$\frac{2K}{\sqrt{2gh_2}} < \sqrt{2gh_2} \tag{3.5.76}$$

because of the assumption $K < gh_2$. Thus we obtain that

$$\tilde{r}_1 < u_2. \tag{3.5.77}$$

Moreover

$$\begin{aligned} & gh_2 - \frac{(\tilde{r}_1 - u_2)^2}{2} \\ & = gh_2 - \frac{1}{2}\left(-\sqrt{2gh_2} + \frac{2K}{\sqrt{2gh_2}}\right)^2 \\ & = -\frac{K^2}{gh_2} + 2K = K\left(-\frac{K}{gh_2} + 2\right) \end{aligned} \tag{3.5.78}$$

and again using the asumption $K < gh_2$ we notice that

$$-\frac{K}{gh_2} + 2 > 1 \tag{3.5.79}$$

therefore

$$gh_2 - \frac{(\tilde{r}_1 - u_2)^2}{2} > K \tag{3.5.80}$$

which means that $P(\tilde{r}_1) > 0$. In consequence, using (3.5.77), we deduce that

$$r_1 < \tilde{r}_1 < u_2. \tag{3.5.81}$$

Similarly we set

$$\tilde{r}_2 = u_2 + \sqrt{2gh_2} - \frac{2K}{\sqrt{2gh_2}} \tag{3.5.82}$$

and by the same arguments we obtain that

$$u_2 < \tilde{r}_2 < r_2. \tag{3.5.83}$$

Putting together (3.5.81) and (3.5.83), we get

$$\begin{aligned} & \left| r_1 - \left(u_2 - \sqrt{2gh_2}\right)\right| + \left| u_2 + \sqrt{2gh_2} - r_2 \right| \\ \leq & \left| \tilde{r}_1 - \left(u_2 - \sqrt{2gh_2}\right)\right| + \left| u_2 + \sqrt{2gh_2} - \tilde{r}_2 \right|. \end{aligned} \tag{3.5.84}$$

Finally, using (3.5.74), we get

$$\begin{aligned}|\operatorname{supp}(M_1)^c \cap \operatorname{supp}(M_2)| &\leq \left|\tilde{r}_1 - \left(u_2 - \sqrt{2gh_2}\right)\right| + \left|u_2 + \sqrt{2gh_2} - \tilde{r}_2\right| \\ &\leq \frac{4K}{\sqrt{2gh_2}}. \end{aligned} \tag{3.5.85}$$

- if $K \geq gh_2$ then

$$|\operatorname{supp}(M_1)^c \cap \operatorname{supp}(M_2)| \leq |\operatorname{supp}(M_2)| \leq 2\sqrt{2gh_2} = 2\frac{2gh_2}{\sqrt{2gh_2}} \leq \frac{4K}{\sqrt{2gh_2}} \tag{3.5.86}$$

which concludes (3.5.70) and the proof.

□

Lemma 3.8. *There exists some $C > 0$ depending only on g, h_m, h_M, u_M such that*

$$\begin{aligned}&\int_{\mathbb{R}} |\xi| \frac{g^2\pi^2}{6} (2M_1 + M_2)(M_1 - M_2)^2 d\xi \\ \geq& C\left(g\frac{(h_2 - h_1)^2}{2} + h_m \frac{(u_2 - u_1)^2}{2}\right). \end{aligned} \tag{3.5.87}$$

for every U_1 , $U_2 \in \mathfrak{U}_{h_m,h_M,u_M}$ defined by (3.1.30) and where $M_k \equiv M_k(\xi) \equiv M(U_k, \xi)$, with $M(U, \xi)$ defined in (3.1.8).

Proof. Let $U_1, U_2 \in \mathfrak{U}_{h_m,h_M,u_M}$. We use Lemma 3.5 and get

$$\begin{aligned}&\int_{\mathbb{R}} |\xi| \frac{g^2\pi^2}{6} (2M_1 + M_2)(M_1 - M_2)^2 d\xi \\ =& \int_{\mathbb{R}} |\xi| \left(H(M_2) - H(M_1) - \eta'(U_1)\begin{pmatrix}1 \\ \xi\end{pmatrix}(M_2 - M_1)\right) d\xi \\ &- \int_{(\xi-u_1)^2 > 2gh_1} |\xi| M(U_2, \xi)\left(gh_1 - \frac{(\xi - u_1)^2}{2}\right) d\xi. \end{aligned} \tag{3.5.88}$$

We first we deal with the case

$$U_1, U_2 \text{ such that } |h_1 - h_2| \leq \frac{1}{4\widetilde{C}_1^2} \text{ and } |u_1 - u_2| \leq \frac{1}{4\widetilde{C}_2^2} \tag{3.5.89}$$

where $\widetilde{C}_1$, $\widetilde{C}_2$ are positive constants depending on g, h_m, h_M, u_M such that $4\widetilde{C}_2^2 \geq \frac{1}{\sqrt{gh_m}}$. In this case, we are going to estimate the right-hand side of (3.5.88). On the one hand, in order to estimate the first term in the RHS of (3.5.88), we apply Lemma 3.6 and since $4\widetilde{C}_2^2 \geq \frac{1}{\sqrt{gh_m}}$ we are in the case (3.5.12) and we get

$$\begin{aligned}
&\int_{\mathbb{R}} |\xi| \left(H(M_2) - H(M_1) - \eta'(U_1) \begin{pmatrix} 1 \\ \xi \end{pmatrix} (M_2 - M_1) \right) d\xi \\
&\geq \alpha_m \left(\eta(U_2) - \eta(U_1) - \eta'(U_1)\,(U_2 - U_1)\right) \\
&= \alpha_m \left(g\frac{(h_2-h_1)^2}{2} + h_1 \frac{(u_2-u_1)^2}{2} \right) \\
&\geq \alpha_m \left(g\frac{(h_2-h_1)^2}{2} + h_m \frac{(u_2-u_1)^2}{2} \right)
\end{aligned} \tag{3.5.90}$$

with $\alpha_m = \frac{C}{\sqrt{2}\pi}\sqrt{gh_m}$ and $C > 0$ and absolute constant. One may notice that in order to obtain the last inequality we only used the fact that $h_1 \geq h_m$. On the other hand, in order to estimate the second term in the RHS of (3.5.88), we use Lemma 3.7 and obtain

$$\begin{aligned}
&\int_{(\xi-u_1)^2 > 2gh_1} |\xi| M(U_2,\xi) \left(\frac{(\xi-u_1)^2}{2} - gh_1 \right) d\xi \\
&\leq \frac{4\left(|u_2| + \sqrt{2gh_2}\right)}{g\pi\sqrt{gh_2}} \left(g|h_1 - h_2| + (|u_2| + \sqrt{2gh_2})|u_1 - u_2| + \frac{1}{2}|u_1^2 - u_2^2| \right)^{\frac{5}{2}} \\
&\leq C_1(h_m, h_M, u_M)\left(g|h_1 - h_2| + C_2(h_m, h_M, u_M)|u_1 - u_2|\right)^{\frac{5}{2}}
\end{aligned} \tag{3.5.91}$$

with

$$\begin{aligned}
C_1(h_m, h_M, u_M) &= \frac{4\left(|u_M| + \sqrt{2gh_M}\right)}{g\pi\sqrt{gh_m}}, \\
C_2(h_m, h_M, u_M) &= 2|u_M| + \sqrt{2gh_M}
\end{aligned} \tag{3.5.92}$$

where $u_M = \dfrac{u_M}{h_m}$. Using Hölder inequality on $\mathbb{R}^2$ we get that for $a, b > 0$, $(a+b)^{5/2} \leq 2^{3/2}(a^{5/2} + b^{5/2})$, we get

$$\begin{aligned}
&\int_{(\xi-u_1)^2 > 2gh_1} |\xi| M(U_2,\xi) \left(\frac{(\xi-u_1)^2}{2} - gh_1 \right) d\xi \\
&\leq 2^{3/2} C_1(h_m, h_M, u_M) \left(g^{\frac{5}{2}} |h_1 - h_2|^{\frac{5}{2}} + C_2(h_m, h_M, u_M)^{\frac{5}{2}} |u_1 - u_2|^{\frac{5}{2}} \right).
\end{aligned} \tag{3.5.93}$$

Thus, putting together the two estimates (3.5.90) and (3.5.93) of the RHS of (3.5.88), we get

$$\begin{aligned}
&\int_{\mathbb{R}} |\xi| \frac{g^2\pi^2}{6} \left(2M_1 + M_2\right) (M_1 - M_2)^2 d\xi \\
&\geq \alpha_m \left(g\frac{(h_2-h_1)^2}{2} + h_m \frac{(u_2-u_1)^2}{2} \right) \\
&\quad - 2^{3/2} C_1(h_m, h_M, u_M) \left(g^{\frac{5}{2}} |h_1 - h_2|^{\frac{5}{2}} + C_2(h_m, h_M, u_M)^{\frac{5}{2}} |u_1 - u_2|^{\frac{5}{2}} \right) \\
&= \alpha_m \frac{g(h_2-h_1)^2}{2} \left(1 - \widetilde{C}_1 |h_1 - h_2|^{\frac{1}{2}}\right) \\
&\quad + \alpha_m \frac{h_m (u_2-u_1)^2}{2} \left(1 - \widetilde{C}_2 |u_1 - u_2|^{\frac{1}{2}}\right)
\end{aligned} \tag{3.5.94}$$

with

$$\alpha_m = \frac{C}{\sqrt{2\pi}} \sqrt{g h_m}, \quad C > 0 \text{ an absolute constant,} \tag{3.5.95}$$

and

$$\begin{aligned}
\widetilde{C}_1 &= \frac{2^{3/2+1} C_1(h_m, h_M, u_M) g^{\frac{5}{2}}}{\alpha_m g}, \\
\widetilde{C}_2 &= \frac{2^{3/2+1} C_1(h_m, h_M, u_M) g^{\frac{5}{2}} C_2(h_m, h_M, u_M)^{\frac{5}{2}}}{\alpha_m g}.
\end{aligned} \tag{3.5.96}$$

From (3.5.94), using that we deal with U_1, U_2 satisfying (3.5.89), we get

$$\begin{aligned}
&\int_{(\xi-u_1)^2 > 2gh_1} |\xi| M(U_2, \xi) \left(\frac{(\xi - u_1)^2}{2} - g h_1 \right) d\xi \\
&\geq \frac{\alpha_m}{2} \left(g \frac{(h_2-h_1)^2}{2} + h_m \frac{(u_2-u_1)^2}{2} \right).
\end{aligned} \tag{3.5.97}$$

At this point we have the result (3.5.87) for all $U_1, U_2 \in \mathfrak{U}_{h_m,h_M,u_M}$ satisfying (3.5.89). Thus, it is now sufficient to prove that

$$\begin{aligned}
&\exists \alpha_1 > 0, \quad \forall U_1, U_2 \in \mathfrak{U}_{h_m,h_M,u_M} \text{ such that} \\
&|h_1 - h_2| > \frac{1}{4\widetilde{C}_1^2} \text{ or } |u_1 - u_2| > \frac{1}{4\widetilde{C}_2^2}, \\
&\text{ we have } \int_{\mathbb{R}} |\xi| \frac{g^2\pi^2}{6} \left(2M_1 + M_2\right)(M_1 - M_2)^2 d\xi \geq \alpha_1.
\end{aligned} \tag{3.5.98}$$

Indeed, this last inequality implies that

$$\begin{aligned}
&g\frac{(h_2-h_1)^2}{2}+h_m\frac{(u_2-u_1)^2}{2}\\
&\le C(h_M,u_M)\\
&\le \frac{C(h_M,u_M)}{\alpha_1}\int_{\mathbb{R}}|\xi|\frac{g^2\pi^2}{6}\left(2M_1+M_2\right)\left(M_1-M_2\right)^2 d\xi
\end{aligned} \tag{3.5.99}$$

which proves (3.5.87). Using reductio ad absurdum as in the proof of Lemma 3.6, we suppose that (3.5.98) does not hold. Thus

$$\begin{aligned}
&\forall n>0, \quad \exists U_1^n, U_2^n \in \mathfrak{U}_{h_m,h_M,u_M}, \text{ such that}\\
&4\widetilde{C}_1^2|h_1^n-h_2^n|+4\widetilde{C}_2^2|u_1^n-u_2^n|>1\\
&\text{and } \int_{\mathbb{R}}|\xi|\frac{g^2\pi^2}{6}\left(2M_1^n+M_2^n\right)\left(M_1^n-M_2^n\right)^2 d\xi \le \frac{1}{n}
\end{aligned} \tag{3.5.100}$$

where $M_i^n = M(U_i^n,\xi)$. As $\mathfrak{U}_{h_m,h_M,u_M}$ is a closed and bounded set, we can take 2 subsequences which we also denote U_1^n, U_2^n such that

$$U_1^n \to U_1 \in \mathfrak{U}_m, \quad U_2^n \to U_2 \in \mathfrak{U}_m \tag{3.5.101}$$

with

$$4\widetilde{C}_1^2|h_1-h_2|+4\widetilde{C}_2^2|u_1-u_2|\ge 1 \tag{3.5.102}$$

and by dominated converge theorem

$$\int_{\mathbb{R}}|\xi|\frac{g^2\pi^2}{6}\left(2M_1+M_2\right)\left(M_1-M_2\right)^2 d\xi = 0. \tag{3.5.103}$$

Therefore we get

$$\left(2M_1+M_2\right)\left(M_1-M_2\right)^2 = 0 \quad \text{almost everywhere} \tag{3.5.104}$$

itself implying that $M_1 = M_2$ a.e. and therefore $U_1 = U_2$, the later being in contradiction with (3.5.102). □

Lemma 3.9. *Let $U_k = (h_k, h_k u_k)$, $k = 1,2$ with $h_k \ge 0$. Then*

$$\begin{aligned}
&\int_{\mathbb{R}}|(M(U_1,\xi)-M(U_2,\xi)|)\,d\xi\\
&\le \frac{2\sqrt{3}}{\sqrt{g}}\left(g(h_2-h_1)^2+\min(h_1,h_2)(u_2-u_1)^2\right)^{\frac{1}{2}},
\end{aligned} \tag{3.5.105}$$

with $M(U,\xi)$ *defined in* (3.1.8).

Proof. Let us recall some result from [7]

$$\begin{aligned}&\int_{\mathbb{R}} M(U_1,\xi)\left(M(U_1,\xi)-M(U_2,\xi)\right)^2 d\xi\\&\leq \frac{3}{g^2\pi^2}\Big(g(h_2-h_1)^2+\min(h_1,h_2)(u_2-u_1)^2\Big).\end{aligned} \tag{3.5.106}$$

We compute

$$\begin{aligned}&\int_{\mathbb{R}} \left|\left(M(U_1,\xi)-M(U_2,\xi)\right|\right) d\xi\\&\leq \int_{M_1>0}|M_1-M_2|\,d\xi+\int_{M_2>0}|M_1-M_2|\,d\xi\\&\leq \left(\int_{M_1>0}\frac{1}{M_1}d\xi\right)^{1/2}\left(\int_{M_1>0}M_1\left(M_1-M_2\right)^2 d\xi\right)^{1/2}\\&+\left(\int_{M_2>0}\frac{1}{M_2}d\xi\right)^{1/2}\left(\int_{M_2>0}M_2\left(M_1-M_2\right)^2 d\xi\right)^{1/2}\end{aligned} \tag{3.5.107}$$

where last estimate is obtained by using Cauchy-Schwarz inequality.
Using the substitution $v=\frac{\xi-u}{\sqrt{2gh}}$ we get

$$\begin{aligned}&\int_{M(U,\xi)>0}\frac{1}{M(U,\xi)}d\xi=\int_{u-\sqrt{2gh}}^{u+\sqrt{2gh}}\frac{g\pi}{\left(2gh-(\xi-u)^2\right)^{1/2}}d\xi\\&=\int_{-1}^{1}\frac{g\pi\sqrt{2gh}}{\sqrt{2gh}\,(1-v^2)^{1/2}}dv=g\pi\Big[\operatorname{Arcsin}(v)\Big]_{-1}^{1}=g\pi^2.\end{aligned} \tag{3.5.108}$$

Now from (3.5.107), using (3.5.106) and (3.5.108), we get

$$\begin{aligned}&\int_{\mathbb{R}}\left|\left(M(U_1,\xi)-M(U_2,\xi)\right)\right| d\xi\\&\leq 2\cdot\frac{\sqrt{3}}{\sqrt{g}}\Big(g(h_2-h_1)^2+\min(h_1,h_2)(u_2-u_1)^2\Big)^{\frac{1}{2}}\end{aligned} \tag{3.5.109}$$

i.e. we find (3.5.105), which concludes the proof. ∎

Lemma 3.10. *Let* $U_k=(h_k,h_ku_k)$, $k=1,2$ *with* $h_k\geq 0$. *Moreover we set here*

$$C=\max_{v\in\left\{|u_1|+\sqrt{gh_1},|u_2|+\sqrt{gh_2}\right\}}|v|\left(1+v^2\right)^{\frac{1}{2}}. \tag{3.5.110}$$

Then one has

$$|F(U_1)-F(U_2)| \leq \frac{2\sqrt{3}}{\sqrt{g}} C \left(g(h_2-h_1)^2+\min(h_1,h_2)(u_2-u_1)^2\right)^{\frac{1}{2}}, \tag{3.5.111}$$

$$\left|F^+(U_1)-F^+(U_2)\right| \leq \frac{2\sqrt{3}}{\sqrt{g}} C \left(g(h_2-h_1)^2+\min(h_1,h_2)(u_2-u_1)^2\right)^{\frac{1}{2}}, \tag{3.5.112}$$

$$\left|F^-(U_1)-F^-(U_2)\right| \leq \frac{2\sqrt{3}}{\sqrt{g}} C \left(g(h_2-h_1)^2+\min(h_1,h_2)(u_2-u_1)^2\right)^{\frac{1}{2}}. \tag{3.5.113}$$

Proof. We recall that

$$\begin{aligned} F^+(U) &= \int_{\mathbb{R}} \xi \mathbb{1}_{\xi>0} \begin{pmatrix} 1 \\ \xi \end{pmatrix} M(U,\xi)d\xi, \\ F^-(U) &= \int_{\mathbb{R}} \xi \mathbb{1}_{\xi<0} \begin{pmatrix} 1 \\ \xi \end{pmatrix} M(U,\xi)d\xi \\ \text{and} \quad F(U) &= \int_{\mathbb{R}} \xi \begin{pmatrix} 1 \\ \xi \end{pmatrix} M(U,\xi)d\xi. \end{aligned} \tag{3.5.114}$$

Thus the result is an immediate consequence of Lemma 3.9 and the fact that

$$\forall \xi \in \operatorname{supp} M_1 \cap \operatorname{supp} M_2, \qquad \left|\xi \begin{pmatrix} 1 \\ \xi \end{pmatrix}\right| \leq \underbrace{\max_{\xi \in \left\{|u_1|+\sqrt{gh_1},|u_2|+\sqrt{gh_2}\right\}} |\xi| \left(1+\xi^2\right)^{\frac{1}{2}}}_{=C} \tag{3.5.115}$$

with C defined by (3.5.110). □

Lemma 3.11. *Let (U_i^n) defined by (3.1.15)-(3.1.25) and U_Δ defined by (3.1.28). Let $\phi \in \mathcal{D}$ and we assume (3.1.32)-(3.1.34). Then, under the CFL condition (3.1.27) there exists some $C>0$ depending only on $\phi, g, h_m, h_M, u_M, v_m$ such that*

$$\forall t \in [0,T], \quad <\partial_t U_\Delta(t,\cdot), \phi> \leq C. \tag{3.5.116}$$

Proof. Using (3.2.47) we get

$$< \partial_t U_\Delta, \phi >= A + B \tag{3.5.117}$$

with

$$A = \sum_i \frac{1}{\Delta t}\left[\frac{U_{i+1}^{n+1} - U_{i-1}^{n+1} - U_{i+1}^n + U_{i-1}^n}{2\Delta x}\right]\int_{x_{i-1/2}}^{x_{i+1/2}} (x - x_{i-1/2})\phi(x)dx \tag{3.5.118}$$

and

$$B = \sum_i \frac{1}{\Delta t}\left[U_i^{n+1} - U_i^n\right]\int_{x_{i-1/2}}^{x_{i+1/2}} \phi(x)dx. \tag{3.5.119}$$

First we notice that

$$\int_{x_{i-1/2}}^{x_{i+1/2}} (x - x_{i-1/2})\phi(x)dx = \psi(x_{i+1/2})\Delta x - \int_{x_{i-1/2}}^{x_{i+1/2}} \psi(x)dx, \tag{3.5.120}$$

where ψ is an antiderivative of ϕ. Then, using last equality in (3.5.118), we get

$$A = \sum_i \frac{1}{\Delta t}\left[\frac{U_{i+1}^{n+1} - U_{i-1}^{n+1} - U_{i+1}^n + U_{i-1}^n}{2\Delta x}\right]\Delta\psi_i \Delta x, \tag{3.5.121}$$

with $\Delta\psi_i := \psi(x_{i+1/2}) - \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \psi(x)dx$. Moreover, by making substitutions of indices we get

$$A = \sum_i \frac{1}{\Delta t}\left[\frac{U_{i+1}^{n+1} - U_{i+1}^n}{2}\right]\left[\Delta\psi_i - \Delta\psi_{i+2}\right]. \tag{3.5.122}$$

Next using that (U_i^n) is a bounded sequence we get that

$$|U_{i+1}^{n+1} - U_{i+1}^n| \leq \frac{\Delta t}{\Delta x}\left(\|F^+(U)\|_\infty + F^-(U)\|_\infty\right). \tag{3.5.123}$$

Moreover we notice that

$$\left|\frac{\Delta\psi_i - \Delta\psi_{i+2}}{2}\right| \leq \Delta x^2 \mathrm{Lip}(\phi), \tag{3.5.124}$$

which enables us to get

$$|A| \leq 2C\|(U_i^n)\|_{l^\infty}\mathrm{Lip}(\phi)\left(\Delta x|\mathrm{supp}\phi|\right), \tag{3.5.125}$$

with $C > 0$ a constant depending only on ϕ. Finally using CFL, it is bounded. Next, from

(3.5.119), we use (3.1.17) and we get

$$B = \sum_i -\frac{1}{\Delta t}\frac{\Delta t}{\Delta x}\left[F_{i+1/2-} - F_{i-1/2+}\right]\Delta x \phi_i, \tag{3.5.126}$$

with $\phi_i := \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \phi(x)dx$. Using (3.1.18)-(3.1.20) we get

$$\begin{aligned} B &= \sum_i -\left[F_{i+1/2-} - F_{i-1/2+}\right]\phi_i \\ &= \sum_i -\left[\mathcal{F}_{i+1/2} - \mathcal{F}_{i-1/2}\right]\phi_i + \sum_i -\left[S_{i+1/2-} - S_{i-1/2+}\right]\phi_i \\ &= \sum_i -\mathcal{F}_{i+1/2}\left[\phi_i - \phi_{i+1}\right] + \sum_i -\left[S_{i+1/2-} - S_{i-1/2+}\right]\phi_i. \end{aligned} \tag{3.5.127}$$

Moreover

$$|\phi_i - \phi_{i+1}| \leq \Delta x \mathrm{Lip}(\phi) \tag{3.5.128}$$

and

$$|S_{i+1/2-} - S_{i-1/2+}| \leq C|z_{i+1} - z_i| \leq C|\Delta x|. \tag{3.5.129}$$

Furthermore, using that (U_i^n) is a bounded sequence we get

$$|B| \leq C\left(\|F^+(U)\|_\infty + \|F^-(U)\|_\infty\right)\mathrm{Lip}(\phi) + C\|\widetilde{S}_\Delta\|_\infty\|\phi\|_\infty, \tag{3.5.130}$$

with C depending only on ϕ. □

Lemma 3.12. *Let $U_\Delta = (h_\Delta, h_\Delta u_\Delta)$ be the approximate solution of (3.1.1) defined by (3.1.28) and $\widetilde{S}_\Delta$ be the approximate source defined by (3.3.7). We assume that there exists U such that U_Δ tends to U a.e. and in L^p_{loc}, as $\Delta x, \Delta t \to 0$. Then we get that*

$$\forall \phi(t, xt) \in \mathcal{D}(\mathbb{R}^2), \quad \iint \widetilde{S}_\Delta(t,x)\phi dt dx \underset{\Delta x, \Delta t \to 0}{\longrightarrow} \iint S(t,x)\phi(t,x) dt dx, \tag{3.5.131}$$

with $S(t,x) = \begin{pmatrix} 0 \\ -gh\partial_x z \end{pmatrix}$.

Proof. Let $\phi(t,x) \in \mathcal{D}(\mathbb{R}^2)$. We study the following integral

$$\iint \widetilde{S}_\Delta(t,x)\phi dtdx = \sum_{n=0}^{N+1} \sum_{i=i_0-1}^{i_1+1} \Delta t \Delta x \left(S_{i+1/2-} + S_{i-1/2+}\right) \phi_i$$

with $\phi_i^n = \dfrac{1}{\Delta t}\dfrac{1}{\Delta x}\displaystyle\int_{t_n}^{t_{n+1}}\int_{x_{i-1/2}}^{x_{i+1/2}} \phi(t,x)dtdx$. Next we develop and make a translation over index i and we get

$$\begin{aligned}
&\sum_n \sum_i \Delta t \Delta x \left(S_{i+1/2-} + S_{i-1/2+}\right) \phi_i \\
=&\sum_n \Delta t \Delta x \sum_i S_{i+1/2-}\phi_i + \sum_n \Delta t \Delta x \sum_i S_{i-1/2+}\phi_i, \\
=&\sum_n \Delta t \Delta x \sum_i S_{i+1/2-}\phi_i + \sum_n \Delta t \Delta x \sum_i S_{i+1/2+}\phi_{i+1}.
\end{aligned}$$

Then we notice that $|\phi_{i+1}^n - \phi_i^n| \leq C\Delta x$ with constant $C > 0$ and we obtain that

$$\begin{aligned}
&\left| \iint \widetilde{S}_\Delta(t,x)\phi dtdx - \sum_n \sum_i \Delta t \Delta x \left(S_{i+1/2-} + S_{i+1/2+}\right) \phi_i \right| \\
&\qquad \leq C\Delta x \sum_{n, t_n \in supp\phi} \Delta t \sum_{i, x_{i+1/2} \in supp\phi} S_{i+1/2+}\Delta x. \qquad (3.5.132)
\end{aligned}$$

Since $S_{i+1/2+}$ is bounded the RHS tends to 0. Next, for $\Delta x, |z_{i+1} - z_i|$ small enough, we have on the one hand

$$\begin{aligned}
\Delta x S^{hu}_{i+1/2-} =& g\frac{h^2_{i+1/2-}}{2} - g\frac{h_i^2}{2} \\
=& g\frac{(h_i + z_i - z_{i+1/2})^2}{2} - g\frac{h_i^2}{2}, \text{ (by assumption (3.1.32))} \\
=& g(z_i - z_{i+1/2})\left(h_i + \frac{z_i - z_{i+1/2}}{2}\right). \qquad (3.5.133)
\end{aligned}$$

On the other hand, as in (3.5.133), we obtain

$$\Delta x S^{hu}_{i+1/2+} = g\frac{h^2_{i+1}}{2} - g\frac{h^2_{i+1/2+}}{2} = -g(z_{i+1} - z_{i+1/2})\left(h_{i+1} + \frac{z_{i+1} - z_{i+1/2}}{2}\right). \qquad (3.5.134)$$

Moreover noticing that $h_{i+1} = h_i + (h_{i+1} - h_i)$, with (3.5.133),(3.5.134) we get

$$\begin{aligned}&\sum_n \sum_i \Delta t \Delta x \left(S^{hu}_{i+1/2-} + S^{hu}_{i+1/2+} \right) \phi_i \\ &= - \sum_n \sum_i \Delta t \Delta x \, g(z_{i+1} - z_i) h_i \phi_i + \sum_n \sum_i \Delta t \Delta x R^n_i \phi_i \end{aligned} \tag{3.5.135}$$

with

$$R^n_i = -g(z_{i+1} - z_{i+1/2}) \left(h_{i+1} - h_i \right) + g\frac{(z_i - z_{i+1/2})^2}{2} \qquad -g\frac{(z_{i+1} - z_{i+1/2})^2}{2}. \tag{3.5.136}$$

First term in the RHS of (3.5.135) converges to the source term:

$$\begin{aligned}&- \sum_n \sum_i \Delta t g(z_{i+1} - z_i) h_i \phi_i = - \sum_n \sum_i \Delta t \Delta x \, g\frac{z_{i+1} - z_i}{\Delta x} h_i \phi_i \\ &= \iint -g\frac{\mathrm{d}z_\Delta(x)}{\mathrm{d}x} h_\Delta(t,x)\phi(x)dx \to \iint -g\frac{\mathrm{d}z(x)}{\mathrm{d}x} h(x)\phi(x)dx,\end{aligned}$$

the convergence holds because we supposed $h_\Delta \to h$, in L^p_{loc} and $\frac{\mathrm{d}z_\Delta(x)}{\mathrm{d}x} \to \frac{\mathrm{d}z(x)}{\mathrm{d}x}$, in L^∞_{loc}. In order to conclude, according to (3.5.132) we need to prove that the remaining terms $\sum_n \sum_i \Delta t \Delta x R^n_i \phi_i$ tend to 0 as $\Delta t, \Delta x \to 0$.
Using that $0 \leq z_{i+1/2} - z_{i+1} \leq |z_{i+1} - z_i|$, $0 \leq z_{i+1/2} - z_i \leq |z_{i+1} - z_i|$ and $|z_{i+1} - z_i| \leq C\Delta x$, we get:

$$|R^n_i| \leq C\Delta x \, |h_{i+1} - h_i| + C\Delta x^2 \tag{3.5.137}$$

and

$$\left| \sum_n \sum_i \Delta t R^n_i \phi_i \right| \leq C \sum_n \sum_i \Delta t \Delta x \, |h_{i+1} - h_i| + C \sum_n \sum_i \Delta t \Delta x^2 \, |\phi_i| \, . \tag{3.5.138}$$

On the one hand we have

$$\sum_n \sum_i \Delta t \Delta x \, |h_{i+1} - h_i| \, \phi_i = O(\Delta x^{1/2}), \tag{3.5.139}$$

using Cauchy Swartz and $\sum_n \sum_i \Delta t (h_{i+1} - h_i)^2 < C$, see (3.2.3).

On the other hand we have

$$\left|\sum_n \sum_i \Delta t \Delta x^2 \phi_i\right| \leq \Delta x \|\phi\|_{L^1} \tag{3.5.140}$$

Thus

$$\sum_n \sum_i \Delta t R_i^n \phi_i \longrightarrow 0, \tag{3.5.141}$$

which concludes the proof. □

Bibliography

[1] R. Abgrall and S. Karni. Two-layer shallow water system: a relaxation approach. *SIAM Journal on Scientific Computing*, 31:1603–1627, 2009. 3

[2] D. Amadori and L. Gosse. Stringent error estimates for one-dimensional, space-dependent 2×2 relaxation systems. preprint, 2015. 76

[3] A. Ambroso, C. Chalons, and P.-A. Raviart. A godunov-type method for the seven-equation model of compressible two-phase flow. *Computers & Fluids*, 54:67–91, 2012. 7

[4] N. Andrianov. Performance of numerical methods on the non-unique solution to the riemann problem for the shallow water equations. *International Journal for numerical methods in fluids*, 47:825–831, 2005. 76

[5] E. Audusse, C. Berthon, C. Chalons, O. Delestre, N. Goutal, M. Jodeau, J. Sainte-Marie, J. Giesselmann, and G. Sadaka. Sediment transport modelling: Relaxation schemes for saint-venant–exner and three layer models. In *ESAIM: proceedings*, 2012. 3

[6] 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 Journal on Scientific Computing*, 25:2050–2065, 2004. 11, 56, 57, 76

[7] E. Audusse, F. Bouchut, M.-O. Bristeau, and J. Sainte-Marie. Kinetic entropy inequality and hydrostatic reconstruction scheme for the saint-venant system. preprint. 15, 75, 76, 78, 82, 84, 116

[8] E. Audusse and M-O. Bristeau. A well-balanced positivity preserving second-order scheme for shallow water flows on unstructured meshes. *Journal of Computational Physics*, 206:311–333, 2005. 76

[9] E. Audusse, M.-O. Bristeau, M. Pelanti, and J. Sainte-Marie. Approximation of the hydrostatic navier–stokes system for density stratified flows by a multilayer model: kinetic interpretation and numerical solution. *Journal of Computational Physics*, 230:3453–3478, 2011. 3

[10] E. Audusse, M.-O. Bristeau, B. Perthame, and J. Sainte-Marie. A multilayer saint-venant system with mass exchanges for shallow water flows. derivation and numerical validation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 45:169–200, 2011. 3

[11] D.S. Bale, R. J. LeVeque, S. Mitran, and J.A. Rossmanith. A wave propagation method for conservation laws and balance laws with spatially varying flux functions. *SIAM Journal on Scientific Computing*, 24:955–978, 2003. 56

[12] D.S. Balsara, M. Dumbser, and R. Abgrall. Multidimensional hllc riemann solver for unstructured meshes–with application to euler and mhd flows. *Journal of Computational Physics*, 261:172–208, 2014. 21, 22, 55

[13] M. Baudin, C. Berthon, F. Coquel, R. Masson, and Q.H. Tran. A relaxation method for two-phase flow models with hydrodynamic closure law. *Numerische Mathematik*, 99:411–440, 2005. 7

[14] F. Berthelin. Convergence of flux vector splitting schemes with single entropy inequality for hyperbolic systems of conservation laws. *Numerische Mathematik*, 99:585–604, 2005. 15, 76

[15] F. Berthelin and F. Bouchut. Relaxation to isentropic gas dynamics for a bgk system with single kinetic entropy. *Methods and Applications of Analysis*, 9:313–328, 2002. 15, 76

[16] C. Berthon and C. Chalons. A fully well-balanced, positive and entropy-satisfying godunov-type method for the shallow water equations. preprint, 2013. 76

[17] C. Berthon, P. Charrier, and B. Dubroca. An hllc scheme to solve the m 1 model of radiative transfer in two space dimensions. *Journal of Scientific Computing*, 31:347–389, 2007. 22

[18] C. Berthon, B. Dubroca, and A. Sangam. A local entropy minimum principle for deriving entropy preserving schemes. *SIAM Journal on Numerical Analysis*, 50:468–491, 2012. 22

[19] C. Berthon and F. Foucher. Efficient well-balanced hydrostatic upwind schemes for shallow-water equations. *Journal of Computational Physics*, 231. 56

[20] C. Berthon and F. Marche. A positive preserving high order vfroe scheme for shallow water equations: a class of relaxation schemes. *SIAM Journal on Scientific Computing*, 30:2587–2612, 2008. 76

[21] F. Bezard and B. Després. An entropic solver for ideal lagrangian magnetohydrodynamics. *Journal of Computational Physics*, 154:65–89, 1999. 22

[22] F. Bouchut. Entropy satisfying flux vector splittings and kinetic bgk models. *Numerische Mathematik*, 94:623–672, 2003. 8, 22, 32

[23] F. Bouchut. Entropy satisfying flux vector splittings and kinetic bgk models. *Numerische Mathematik*, 94(4):623–672, 2003. 76

[24] F. Bouchut. *Nonlinear Stability of Finite Volume Methods for Hyperbolic Conservation Laws: And Well-Balanced Schemes for Sources*. Birkhäuser, 2004. 5, 7, 8, 11, 22, 25, 26, 31, 32, 37, 41, 43, 44, 55, 56, 60, 61, 65, 68, 76

[25] F. Bouchut and S. Boyaval. A new model for shallow viscoelastic fluids. *Mathematical Models and Methods in Applied Sciences*, 23:1479–1526, 2013. 22, 25, 28, 41, 56

[26] F. Bouchut and S. Boyaval. Unified formal reduction for fluid models of free-surface shallow gravity-flows. *arXiv preprint arXiv:1306.3464*, 2013. 3, 7

[27] F. Bouchut, C. Klingenberg, and K. Waagan. A multiwave approximate riemann solver for ideal mhd based on relaxation. i: theoretical framework. *Numerische Mathematik*, 108:7–42, 2007. 7, 22, 25, 28, 32, 35

[28] F. Bouchut, C. Klingenberg, and K. Waagan. A multiwave approximate riemann solver for ideal mhd based on relaxation ii: numerical implementation with 3 and 5 waves. *Numerische Mathematik*, 115:647–679, 2010. 22, 23, 28, 31, 37, 38, 41

[29] F. Bouchut and X. Lhebrard. A 5-wave relaxation solver for the shallow water mhd system. preprint, 2015. 56, 58, 59, 60, 62, 65

[30] F. Bouchut, C. Lusso, and V. Zeitlin. Geostrophic vs magnetostrophic adjustment and nonlinear magneto-inertia-gravity waves in rotating shallow water magnetohydrodynamics. preprint, 2014. 55

[31] F. Bouchut, Anne Mangeney-Castelnau, B. Perthame, and Jean-Pierre Vilotte. A new model of saint venant and savage–hutter type for gravity driven shallow water flows. *Comptes rendus mathematique*, 336:531–536, 2003. 3

[32] F. Bouchut and De la luna T. Morales. A subsonic-well-balanced reconstruction scheme for shallow water flows. *SIAM Journal on Numerical Analysis*, 48:1733–1758, 2010. 56, 76

[33] F. Bouchut and T. De la luna Morales. Semi-discrete entropy satisfying approximate riemann solvers. the case of the suliciu relaxation approximation. *Journal of Scientific Computing*, 41:483–509, 2009. 7

[34] F. Bouchut and V. Zeitlin. A robust well-balanced scheme for multi-layer shallow water equations. *Discrete and Continuous Dynamical Systems-Series B*, 13:739–758, 2010. 56

[35] D. Bresch, E.D. Fernandez-Nieto, I.R. Ionescu, and P. Vigneaux. Augmented lagrangian method and compressible visco-plastic flows: applications to shallow dense avalanches. In *New directions in mathematical fluid mechanics*. 2010. 3

[36] M. Brio and C.C. Wu. An upwind differencing scheme for the equations of ideal magnetohydrodynamics. *Journal of Computational Physics*, 75:400–422, 1988. 5

[37] M.-O. Bristeau, N. Goutal, and J. Sainte-Marie. Numerical simulations of a non-hydrostatic shallow water model. *Computers & fluids*, 47:51–64, 2011. 76

[38] M.-O. Bristeau and J. Sainte-Marie. Derivation of a non-hydrostatic shallow water model; comparison with saint-venant and boussinesq systems. *Discrete and Continuous Dynamical Systems - Series B*, 10:733–759, 2008. 3

[39] M. Castro, A. Pardo, C. Parés, and E. Toro. On some fast well-balanced first order solvers for nonconservative systems. *Mathematics of Computation*, 79:1427–1472, 2010. 56

[40] M. J. Castro, A. Pardo M., and C. Parés. Well-balanced numerical schemes based on a generalized hydrostatic reconstruction technique. *Mathematical Models and Methods in Applied Sciences*, 17:2055–2113, 2007. 56

[41] M.J. Castro Díaz and E. Fernández-Nieto. A class of computationally fast first order finite volume solvers: Pvm methods. *SIAM Journal on Scientific Computing*, 34:A2173–A2196, 2012. 56

[42] C. Chalons and F. Coquel. Modified suliciu relaxation system and exact resolution of isolated shock waves. *Mathematical Models and Methods in Applied Sciences*, 24:937–971, 2014. 7, 22

[43] 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. *Mathematical Models and Methods in Applied Sciences*, 20. 56

[44] C. Chalons, F. Coquel, and C. Marmignon. Well-balanced time implicit formulation of relaxation schemes for the euler equations. *SIAM Journal on Scientific Computing*, 30:394–415, 2008. 22

[45] C. Chalons and J.-F. Coulombel. Relaxation approximation of the euler equations. *Journal of Mathematical Analysis and Applications*, 348:872–893, 2008. 5, 8

[46] A. Chertock, A. Kurganov, and Y. Liu. Central-upwind schemes for the system of shallow water equations with horizontal temperature gradients. *Numerische Mathematik*, pages 1–45, 2013. 3, 56

[47] F. Coquel, E. Godlewski, B. Perthame, A. In, and P. Rascle. Some new godunov and relaxation methods for two-phase flow problems. In *Godunov methods*. 2001. 7

[48] F. Coquel, E. Godlewski, and N. Seguin. Relaxation of fluid systems. *Mathematical Models and Methods in Applied Sciences*, 22:1250014, 2012. 8, 22

[49] F. Coquel, K. Saleh, and N. Seguin. A robust and entropy-satisfying numerical scheme for fluid flows in discontinuous nozzles. *Mathematical Models and Methods in Applied Sciences*, 24:2043–2083, 2014. 5

[50] A.J.C. De Saint Venant. Théorie du mouvement non-permanent des eaux, avec application aux crues des rivières et à lâintroduction des marées dans leur lit. *Comptes Rendus des séances de l'Académie des Sciences*, 73:237–240, 1871. 2

[51] H. De Sterck. Hyperbolic theory of the shallow waters magnetohydrodynamics equations. *Physics of Plasmas (1994-present)*, 8:3293–3304, 2001. 7, 21, 55

[52] P.J. Dellar. Hamiltonian and symmetric hyperbolic structures of shallow water magnetohydrodynamics. *Physics of Plasmas*, 9:1130–1136, 2002. 7, 20, 54

[53] P.J. Dellar. Common hamiltonian structure of the shallow water equations with horizontal temperature gradients and magnetic fields. *Physics of Fluids (1994-present)*, 15:292–297, 2003. 3

[54] B. Després. Lagrangian systems of conservation laws. *Numerische Mathematik*, 89:99–134, 2001. 22

[55] V. Desveaux. *Contribution à l'approximation numérique des systèmes hyperboliques*. PhD thesis, Université de Nantes, 2013. 3

[56] V. Desveaux, M. Zenk, C. Berthon, and C. Klingenberg. Well-balanced schemes to capture non-explicit steady states. part 1: Ripa model. preprint 2014. 3

[57] R.J. DiPerna. Convergence of approximate solutions to conservation laws. *Archive for Rational Mechanics and Analysis*, 82:27–70, 1983. 14, 76, 77

[58] B. Einfeldt. On godunov-type methods for gas dynamics. *SIAM Journal on Numerical Analysis*, 25:294–318, 1988. 7

[59] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. *Handbook of numerical analysis*, 7:713–1018, 2000. 4

[60] E.D. Fernández-Nieto, F. Bouchut, D. Bresch, M.J. Castro Díaz, and A. Mangeney. A new savage–hutter type model for submarine avalanches and generated tsunami. *Journal of Computational Physics*, 227:7720–7754, 2008. 3

[61] S. Ferrari and F. Saleri. A new two-dimensional shallow water model including pressure effects and slow varying bottom topography. *ESAIM: Mathematical Modelling and Numerical Analysis*, 38:211–234, 2004. 2

[62] F.G. Fuchs, A.D. McMurry, S. Mishra, N.H. Risebro, and K. Waagan. Approximate riemann solvers and stable high-order finite volume schemes for multi-dimensional ideal mhd. *Comm. Comput. Phys.*, 9:324–362, 2011. 21, 22, 55

[63] G. Gallice. Positive and entropy stable godunov-type schemes for gas dynamics and mhd equations in lagrangian or eulerian coordinates. *Numerische Mathematik*, 94:673–713, 2003. 22

[64] D. L. George. Augmented riemann solvers for the shallow water equations over variable topography with steady states and inundation. *Journal of Computational Physics*, 227:3089–3113, 2008. 56

[65] J.-F.. Gerbeau and B. Perthame. Derivation of Viscous Saint-Venant System for Laminar Shallow Water; Numerical Validation. Technical report, INRIA, 2000. 2

[66] P.A. Gilman. Magnetohydrodynamic shallow water equations for the solar tachocline. *The Astrophysical Journal Letters*, 544:L79–L82, 2000. 7, 20, 53

[67] E. Godlewski and P.-A. Raviart. *Numerical approximation of hyperbolic systems of conservation laws*. Springer, 1996. 4

[68] S.K. Godunov. A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Matematicheskii Sbornik*, 89:271–306, 1959. 7

[69] W. Graf. Les équations de saint-venant-exner. *ÖIAZ*, 139:417–422, 1994. 3

[70] A. Harten, P.D. Lax, and B.V. Leer. On upstream differencing and godunov-type schemes for hyperbolic conservation laws. *SIAM review*, 25:35–61, 1983. 5, 7, 22

[71] P. Janhunen. A positive conservative method for magnetohydrodynamics based on hll and roe methods. *Journal of Computational Physics*, 160:649–661, 2000. 20

[72] S. Jin and Z. Xin. The relaxation schemes for systems of conservation laws in arbitrary space dimensions. *Communications on Pure and Applied Mathematics*, 48:235–276, 1995. 5, 8

[73] R.J. LeVeque. Wave propagation algorithms for multidimensional hyperbolic systems. *Journal of Computational Physics*, 131:327–353, 1997. 5

[74] R.J. LeVeque. *Finite volume methods for hyperbolic problems*. Cambridge university press, 2002. 4

[75] P.-L. Lions, B. Perthame, and P.E. Souganidis. Existence and stability of entropy solutions for the hyperbolic systems of isentropic gas dynamics in eulerian and lagrangian coordinates. *Communications on pure and applied mathematics*, 49:599–638, 1996. 3, 14, 76, 98

[76] Y.-G. Lu, C. Klingenberg, L. Rendon, and D.-Y. Zheng. Global solutions for a simplified shallow elastic fluids model. In *Abstract and Applied Analysis*, 2014. 3

[77] F. Marche. Derivation of a new two-dimensional viscous shallow water model with varying topography, bottom friction and capillary effects. *European Journal of Mechanics-B/Fluids*, 26:49–63, 2007. 2

[78] R.S. Myong and P.L. Roe. On godunov-type schemes for magnetohydrodynamics: 1. a model system. *Journal of Computational Physics*, 147:545 – 567, 1998. 5

[79] G Narbona-Reina and D. Bresch. On a shallow water model for non-newtonian fluids. In *Numerical Mathematics and Advanced Applications 2009*. 2010. 3

[80] C. Parés. Numerical methods for nonconservative hyperbolic systems: a theoretical framework. *SIAM Journal on Numerical Analysis*, 44:300–321, 2006. 56

[81] C. Parés and M. Castro. On the well-balance property of roe's method for nonconservative hyperbolic systems. applications to shallow-water systems. *ESAIM: mathematical modelling and numerical analysis*, 38:821–852, 2004. 56

[82] B. Perthame. *Kinetic formulation of conservation laws*. Oxford University Press, 2002. 5

[83] B. Perthame and C. Simeoni. A kinetic scheme for the saint-venant system with a source term. *Calcolo*, 38:201–231, 2001. 13, 76, 77, 79

[84] B. Perthame and C. Simeoni. Convergence of the upwind interface source method for hyperbolic conservation laws. In *Hyperbolic problems: theory, numerics, applications*, pages 61–78. Springer, 2003. 76

[85] S. Qamar and S. Mudasser. A kinetic flux-vector splitting method for the shallow water magnetohydrodynamics. *Computer Physics Communications*, 181:1109–1122, 2010. 7

[86] S. Qamar and G. Warnecke. Application of space–time ce/se method to shallow water magnetohydrodynamic equations. *Journal of computational and applied mathematics*, 196:132–149, 2006. 55

[87] P. Ripa. Conservation laws for primitive equations models with inhomogeneous layers. *Geophysical & Astrophysical Fluid Dynamics*, 70:85–111, 1993. 3

[88] P. Ripa. On improving a one-layer ocean model with thermodynamics. *Journal of Fluid Mechanics*, 303:169–201, 1995. 3

[89] P.L. Roe. Approximate riemann solvers, parameter vectors, and difference schemes. *Journal of computational physics*, 43:357–372, 1981. 5

[90] J.A. Rossmanith. *A wave propagation method with constrained transport for ideal and shallow water magnetohydrodynamics*. PhD thesis, University of WaS.ngton, 2002. 7

[91] J.A. Rossmanith. A constrained transport method for the shallow water mhd equations. In *Hyperbolic Problems: Theory, Numerics, Applications*, pages 851–860. 2003. 20, 55

[92] J.A. Rossmanith, D.S. Bale, and R.J. LeVeque. A wave propagation algorithm for hyperbolic systems on curved manifolds. *Journal of Computational Physics*, 199:631–662, 2004. 20, 55

[93] J. Sainte-Marie. Vertically averaged models for the free surface non-hydrostatic euler system: derivation and kinetic interpretation. *Mathematical Models and Methods in Applied Sciences*, 21:459–490, 2011. 3

[94] A. Sangam. An hllc scheme for ten-moments approximation coupled with magnetic field. *International Journal of Computing Science and Mathematics*, 2:73–109, 2008. 22

[95] S.B. Savage and K. Hutter. The dynamics of avalanches of granular materials from initiation to runout. part i: Analysis. *Acta Mechanica*, 86:201–223, 1991. 3

[96] I. Suliciu. On modelling phase transitions by means of rate-type constitutive equations. shock wave structure. *International Journal of Engineering Science*, 28:829–841, 1990. 7

[97] I. Suliciu. Some stability-instability problems in phase transitions modelled by piecewise linear elastic or viscoelastic constitutive equations. *International journal of engineering science*, 30:483–494, 1992. 7, 22

[98] L. Tartar. Compensated compactness and applications to partial differential equations. In *Nonlinear analysis and mechanics, Heriot-Watt symposium*, 1979. 14, 76, 77

[99] E.F. Toro. *Riemann solvers and numerical methods for fluid dynamics*. Springer, 1999. 4

[100] E.F. Toro, M. Spruce, and W. Speares. Restoration of the contact surface in the hll-riemann solver. *Shock waves*, 4:25–34, 1994. 5, 7

[101] Q.H. Tran, M. Baudin, and F. Coquel. A relaxation method via the born–infeld system. *Mathematical Models and Methods in Applied Sciences*, 19:1203–1240, 2009. 5

[102] A. Vasseur. Well-posedness of scalar conservation laws with singular sources. *Methods and applications of analysis*, 9:291–312, 2002. 76

[103] K. Waagan, C. Federrath, and C. Klingenberg. A robust numerical scheme for highly compressible magnetohydrodynamics: Nonlinear stability, implementation and tests. *Journal of Computational Physics*, 230:3331–3351, 2011. 21

[104] G. Warnecke and N. Andrianov. On the solution to the riemann problem for the compressible duct flow. *SIAM Journal on Applied Mathematics*, 64(3):878–901, 2004. 76

[105] Y. Xing and C.-W. Shu. A survey of high order schemes for the shallow water equations. *Journal of Mathematical Study*, 47:221–249, 2014. 56

[106] V. Zeitlin. *Nonlinear Dynamics of Rotating Shallow Water: Methods and Advances: Methods and Advances*. Elsevier, 2007. 3

[107] V. Zeitlin. Remarks on rotating shallow-water magnetohydrodynamics. *Nonlinear Processes in Geophysics*, 20:893–898, 2013. 6, 21, 55