Université des Antilles et de la Guyane

THESE

pour obtenir le grade de

Docteur de l'Université des Antilles et de la Guyane

Spécialité : **Mathématiques**

présentée par

Mathilde Colombeau

ETUDE MATHEMATIQUE D'EQUATIONS AUX DERIVEES PARTIELLES NON LINEAIRES PRESENTANT DES SOLUTIONS SINGULIERES

A mathematical study of nonlinear partial differential equations exhibiting irregular solutions

Soutenue le 6 décembre 2011 après avis de

M. EGOROV Youri	Université Paul Sabatier-Toulouse III
M. SHELKOVICH Vladimir	Université de Saint-Petersbourg, Russie
M. STRUPPA Daniele C.	Chapman University, Orange, USA

devant le jury composé de

M. DELCROIX Atoine	Université des Antilles et de la Guyane
M. EGOROV Youri	Université Paul Sabatier-Toulouse III
M. MERIL Alex	Université des Antilles et de la Guyane
M. N'GUEREKATA Gaston	Morgan State University, Baltimore, USA
M. SHELKOVICH Vladimir	Université de Saint-Petersbourg, Russie
M. VALMORIN Vincent	Université des Antilles et de la Guyane

Résumé

Cette thèse a pour objet l'étude théorique et numérique de solutions singulières apparaissant dans des équations aux dérivées partielles non linéaires de la physique, en particulier en dynamique des fluides. La présence de discontinuités dans les solutions de ces équations complique la compréhension mathématique des phénomènes mis en jeu et leur traitement numérique, notamment en vue de simulations informatiques.

Les discontinuités étudiées dans cette thèse sont principalement de trois types. Les ondes de choc, qui peuvent apparaître spontanement au cours du temps ou être imposées en condition initiale. C'est par exemple la brusque variation de pression lorsqu'un avion dépasse le mur du son. Les delta-ondes qui sont des discontinuités surmontées par une masse de Dirac. Elles apparaîssent notamment dans les systèmes de la dynamique des fluides sans pression. Les chocs singuliers sont, quant à eux, des solutions de forme non classique qui restent à être élucidées mathématiquement.

Nous étudions ces équations par une méthode de régularisation dans un espace fonctionel approprié. L'idée de base qui a servi à l'élaboration de ce travail est la suivante : lorsque des schémas numériques construits par des méthodes différentes conduisent à des résultats identiques, ceci jusque dans leurs moindres détails, il semble naturel de s'interroger dans quelle mesure ces suites de solutions numériques constituent une approximation d'une solution des équations étudiées. Nous construisons des suites de solutions approchées à partir d'un schéma numérique original, stable et suffisament simple pour démontrer que ces suites constituent une méthode asymptotique de Maslov au sens des distributions en dimension trois d'espace. La technique employée consiste à étendre les variables réelles du problème (domaine physique) en des variables complexes (domaine non physique), ce qui nous permet de construire des familles de solutions particulières que l'on ramène au cas réel en faisant tendre un petit paramètre vers 0. Les solutions physiques recherchées apparaîssent alors comme valeurs au bord de fonctions holomorphes.

Nous illustrons les résultats obtenus par des applications en dimension deux d'espace en cosmologie dans les cadres Newtonien et relativistes pour des systèmes sans pression, puis avec pression et auto-gravitation, ainsi que pour le système des gaz parfaits.

Mots-clés : Ondes non linéaires ; Solutions singulères ; Méthode itérative ; Méthode asymptotique de Maslov ; Valeurs aux bord de fonctions holomorphes ; Equations d'Euler compressibles ; Dynamique des fluides Newtoniens et relativistes ; Cosmologie ; Gaz parfaits.

Abstract

This thesis is devoted to the theoretical and numerical study of irregular solutions appearing in nonlinear partial differential equations of physics, more specifically in fluid dynamics. The mathematical understanding of the phenomena under concern and their numerical treatment, in particular in view of computer simulations, is made difficult by the presence of discontinuities in the solutions of these equations.

The discontinuities concerned in this thesis range mainly into three kinds. The schock waves, which can appear sponteanously as time passes or be imposed in the initial conditions. For instance the sudden variation of pressure when a plane bypasses the sound speed. The delta waves are discontinuities linked to a Dirac mass. They appear in particular in pressureless fluid dynamics. The singular shocks are solutions having a nonclassical shape that are not completely elucidated.

We study these equations by a regularization method in a convenient functional space. The basic idea at the origin of this work is the following : when numerical schemes from very different numerical methods give identical results, up to the smallest details, it seems natural to ask oneself to what extent these sequences of numerical solutions approximate a solution - in a sense to be made precise - of the equations under study. We construct sequences of approximate solutions from an original numerical scheme which is stable and simple enough to prove that these sequences form a weak asymptotic method in the sense of distributions in three space dimension. The regularization in use consists in extending the real physical variables into complex variables, which permits to construct families of particular solutions that are physically interpreted by letting a small parameter tends to zero. The sought physical solutions appear as boundary values of holomorphic functions.

Results are illusttrated by applications in two spaces dimension in cosmology in the Newtonian and relativistic domains for pressueless systems, then for systems with pressure and self-gravitation, as well as for the system of ideal gases.

Keywords : Nonlinear waves ; Irregular solutions ; Iterative method ; Weak asymptotic method ; Boundary values of holomorphic functions ; Compressible Euler equations ; Newtonian and relativistic fluid dynamics ; Cosmology ; Ideal gases.

Remerciements

Je remercie en tout premier lieu Alex Méril et Vincent Valmorin qui ont accepté d'être mes directeurs de recherche pendant ces trois années.

Je suis très reconnaissante envers Youri Egorov, Vladimir Shelkovich et Daniele Stuppa d'avoir bien voulu être les rapporteurs de ma thèse, et pour l'intérêt qu'ils ont porté à ce travail. Je remercie aussi Youri Egorov pour son écoute attentive et le temps qu'il a pris pour répondre à toutes mes questions lors d'un séjour à Toulouse, ainsi que Vladimir Shelkovich pour ses nombreuses suggestions lors de la soutenance. Ma gratitude va également à Antoine Delcroix et Gaston N'Guerekata qui ont accepté de participer au jury.

Je remercie chaleureusement l'ensemble des membres du CEREGMIA. J'étends volontier ces remerciements aux membres et au personnel administratif du secrétariat de l'école doctorale. Les diverses formations doctorales m'ont permis de rencontrer des doctorants de disciplines différentes ainsi que d'acquérir des compétences complémentaires à mon projet de recherche. Je remercie tous les thésards que j'ai cotoyés lors de ces enseignements, ainsi que les intervenants. C'est aussi pour moi l'occasion d'exprimer toute ma reconnaissance envers tous ceux qui, de près ou de loin, m'ont permis de surmonter les moments de doutes inhérents à la recherche scientifique.

Je remercie enfin ma famille, avec une pensée toute particulière pour mes grand-mères et mon oncle. Cette thèse leur est dédiée.

0.1 Foreword.

The equations of fluid dynamics have applications in numerous domains : cosmology and astrophysics, oceanography, meteorology and climatology, industry and petroleum.... The aim of this work is an attempt to contribute to a theoretical and numerical study of some basic equations of compressible fluid dynamics. One main difficulty in dealing with these equations is that solutions of the Cauchy problem, even those starting from analytic initial data, usually develop singularities in a finite time such as shock waves, delta waves, contact discontinuities, concentrations of matter and void regions, among other irregular solutions. Therefore we are particularly interested in the case where these equations provide irregular solutions. One also faces a severe problem of lack of uniqueness for these irregular solutions.

We study these equations by a regularization method.
• This method consists firstly of exhibiting approximate solutions from a suitable original numerical scheme which is shown to be stable and consistent.
• Secondly of interpreting these approximate solutions in a convenient functional space which permits to regularize them so that they could satisfy the equations.
• Thirdly one shows that one can pass to the limit in this functional space on a sequence of approximate solutions. Then the limit can be considered as a solution of the equations even if it is irregular.
• Finally this solution is concretely put in evidence as a finite set of Radon measures which is an interpretation of the genuine function solution, by letting the regularization variable tend to 0. In the cases solutions are known (for instance in the system of ideal gases : Sod, Woodward-Colella, Toro, Lax, ...) we observe that the concrete solutions obtained in this way are exactly the same as the solutions previously obtained by all authors and widely accepted. In the case of previously unknown solutions we obtain the solutions compatible with physics (large structure formation in cosmology, evolution of rotating dust clouds looking like formation of solar systems, Jeans theory,...).

As pointed out by P.D. Lax in [25] and [26] numerical methods often give good results. When several completely different numerical methods give the same results up to the smallest details one can reasonably expect that these numerical results suggest the existence of a mathematical solution of the equations. This idea was the basis and the main motivation of this work : *use as auxiliary tool a numerical method (to be found so as to be valid and efficient in any space dimension, and to be suitable for proofs of stability and consistency) and use it in an appropriate functional space in which one could prove the convergence of the approximate solutions to a "solution" of the equations in a natural sense, from a result "stability and consistency imply convergence"*. Therefore the aspect of this solution is approximated by the results from the scheme. This method unfortunately does not bring abstract results of uniqueness.

In the first part of the thesis one considers successively basic equations of fluid dynamics (chapters 1 to 3) : pressureless fluid dynamics, presence of self-gravitation and/or presence of pressure, ideal gases. These three systems model a large variety of physical situations. We offer a numerical scheme valid in any space dimension. This scheme is simple and therefore it is only of order one in the space step, but its simplicity is a great mathematical advantage in that it allows us to obtain mathematically rigorous proofs of consistency and convergence in important cases. More precisely let us consider a system of the form

$$U_t + (F(U))_x + (G(U))_y + (H(U))_z = 0, \quad u = (u_1, u_2, \ldots, u_n)^T.$$

Let $(U_h),\ h \to 0$, be a family of step functions issued from a numerical scheme. We say that the scheme "is consistent in the sense of distributions" on $\mathbb{R}^3\times]0,T[$ iff $\forall \psi \in \mathcal{C}_c^\infty(\mathbb{R}^3\times]0,T[)$,

$$\int \{U_h\psi_t + F(U_h)\psi_x + G(U_h)\psi_y + H(U_h)\psi_z\}dxdydzdt = O(h^\alpha)$$

for some $\alpha > 0$ when $h \to 0$. This means that the approximate step functions U_h tend to satisfy the equations when $h \to 0$ within an approximation of order α in h in the sense of distributions. This concept of consistency provides a "weak asymptotic method" obtained from the numerical scheme. The concept of weak asymptotic methods and their relevance have been first put in evidence by V.G. Danilov, G.A. Omel'yanov, and V.M. Shelkovich in [12] by explicit calculations and by reducing the problem of the description of interaction of nonlinear waves to the one of solving some systems of ODEs.

We prove the consistency above for the system of pressureless fluids in 3-D and for the system of self-gravitating pressureless fluids in 1-D (chapters 1 and 2). We can presume that for the systems involving pressure such as the classical system of ideal gases the numerical tests done so far indicate that this consistency would be true (chapters 2 and 3).

This leads to the natural question [26] p. 144 : to what extent do the results above indicate that the existence of the flow that we are approximately calculating exists as a mathematical object that could be qualified as a mathematical solution of the equations ? A tentative answer to this question has been proposed from an analysis of the singular shocks solutions of the Keyfitz-Kranzer system (chapter 5) [22], [21], [33], as well as from explicit calculations on systems of relativistic cosmology (chapter 4) [8], [30]. To take into account the full shape of singular shocks we study them in an appropriate functional space in which convergence can be obtained. The functional space in which the equations are considered remains unchanged in 2-D and 3-D and convergence holds as well.

We will construct elements $\widetilde{U}_h$ in the functional space defined with a regularizing parameter from the step functions in the numerical scheme.Then we can extract a convergent subsequence $(\widetilde{U}_{h_p})_p$. Let $\widetilde{U}$ be its limit. Then we will show that $\widetilde{U}$ is solution of the equations in a natural weak sense close to the classical concept of a weak solution, whose aspect is the numerical solution observed from the scheme.

The weak solutions obtained have the well-known defects of classical weak solutions, in particular a strong problem of lack of uniqueness. Fortunately explicit calculations in very particular cases (shock waves) put in evidence the existence in these cases of stronger solutions for which some uniqueness can be obtained. In chapter 7 we show that in some particular linear cases existence-uniqueness can be obtained by adaptation of the classical method based on coercivity for elliptic boundary value problems. This does not provide even an hint for the above problem of uniqueness for equations of fluid dynamics but shows that extension of classical general results of existence-uniqueness to the case of irregular solutions makes sense. Other existence-uniqueness results are presented in appendix 2 as a work which is presently being investigated.

As a clarification let us divide the methodology into successive steps :

1) find a 3-D numerical scheme that should be rather general to be applied at least to significant equations of fluid dynamics and rather simple to be the starting point of mathematical proofs. As a consequence the (original as far as the author knows) scheme we present is only of

order one for the 3-D usual systems such as the system of ideal gases, the shallow water equations, the systems of collisional and collisionless self-gravitating fluids,.... However it seems this scheme could be useful in numerical practice since in 1-D it gives results similar to the classical Godunov scheme. We show that this scheme extends at once to systems of a large number of equations, and to 3-D problems without dimensional splitting and without any loss of accuracy relatively to 1-D problems.

2) use this scheme for proofs of stability (L^1 stability in density follows from the scheme) and proofs of consistency. Consistency consists in a proof that the approximate solutions from the scheme tend to satisfy the equations when the space step tends to 0. Consistency is rigorously proved as much as possible (3-D pressureless fluids without gravitation, 1-D pressureless fluids with gravitation) and when a proof is lacking (3-D collisional self-gravitating fluids, ideal gases, shallow water equations,...) consistency is reduced to very simple criteria which are verified as convincingly as possible from numerical tests.

3) from an analysis of irregular solutions we can prove the convergence of a sequence of approximate solutions from the scheme to this object. In short this is no more than a version of the familiar fact that "stability and consistency imply convergence". This step has mainly been made possible from an analysis of the singular shock solutions of the Keyfitz-Kranzer equations.

The contents of chapters 1 and 2 have been published in [9] and [10].

Table des matières

0.2 Sum up of equations of fluid dynamics under study.

• **Equations of pressureless fluid dynamics**

In 1-D

$$\rho_t + (\rho u)_x = 0, \quad (\rho u)_t + (\rho u^2)_x = 0, \tag{1}$$

in 3-D

$$\rho_t + \vec{\nabla}.(\rho\vec{u}) = \vec{0}, \quad (\rho\vec{u})_t + \vec{\nabla}.(\rho\vec{u}\otimes\vec{u}) = \vec{0}. \tag{2}$$

We recall that the notation $\vec{\nabla}.(\rho\vec{u}\otimes\vec{u})$ means the vector of components $((\rho u^2)_x + (\rho uv)_y + (\rho uw)_z, (\rho uv)_x + (\rho v^2)_y + (\rho uw)_z, (\rho uw)_x + (\rho vw)_y + (\rho w^2)_z)$ if (u, v, w) are the components of $\vec{u}$.

• **Equations of self-gravitating pressureless fluids**

In 1-D

$$\rho_t + (\rho u)_x = 0, \quad (\rho u)_t + (\rho u^2)_x + \rho\Phi_x = 0, \Phi_{xx} = 4\pi G\rho, \tag{3}$$

in 3-D

$$\rho_t + \vec{\nabla}.(\rho\vec{u}) = \vec{0}, \quad (\rho\vec{u})_t + \vec{\nabla}.(\rho\vec{u}\otimes\vec{u}) + \rho\vec{\nabla\Phi} = \vec{0}, \quad \Delta\Phi = 4\pi G\rho. \tag{4}$$

• **Equations of collisional self-gravitating fluids**

In 1-D

$$\rho_t + (\rho u)_x = 0, \quad (\rho u)_t + (\rho u^2)_x + p_x + \rho\Phi_x = 0, \Phi_{xx} = 4\pi G\rho, p = k\rho, \tag{5}$$

in 3-D

$$\rho_t + \vec{\nabla}.(\rho\vec{u}) = \vec{0}, \ (\rho\vec{u})_t + \vec{\nabla}.(\rho\vec{u}\otimes\vec{u}) + \vec{\nabla}p + \rho\vec{\nabla\Phi} = \vec{0}, \Delta\Phi = 4\pi G\rho, p = k\rho. \tag{6}$$

• **Equations of perfect gases**

In 1-D

$$\begin{gathered}\rho_t + (\rho u)_x = 0, \quad (\rho u)_t + (\rho u^2)_x + p_x = 0, \\ (\rho e)_t + (\rho e u)_x + (pu)_x = 0, \quad p = (\gamma - 1)(\rho e - \rho\frac{u^2}{2}),\end{gathered} \tag{7}$$

in 3-D

$$\begin{gathered}\frac{\partial\rho}{\partial t} + \vec{\nabla}.(\rho\vec{u}) = 0, \quad \frac{\partial}{\partial t}(\rho\vec{u}) + \vec{\nabla}.(\rho\vec{u}\otimes\vec{u}) + \vec{\nabla}p = \vec{0}, \\ \frac{\partial}{\partial t}(\rho e) + \vec{\nabla}.[(\rho e + p)\vec{u}] = 0, \quad p = (\gamma - 1)(\rho e - \rho\frac{\vec{u}^2}{2}).\end{gathered} \tag{8}$$

• **Equations of special relativistic fluid dynamics**

In 1-D [8]
$\rho_t + ((\rho + \frac{p}{c^2})u)_x = 0, \quad (\rho + \frac{p}{c^2})(u_t + uu_x) + p_x + (\rho + \frac{p}{c^2})\Phi_x = 0,$

$$\Phi_{xx} = 4\pi G(\rho + 3\frac{p}{c^2}), p = k\rho, \tag{9}$$

in 3-D [8]
$\rho_t + \vec{\nabla}.((\rho + \frac{p}{c^2})\vec{u}) = \vec{0}, \ (\rho + \frac{p}{c^2})[\vec{u}_t + (\vec{\nabla}.\vec{u}))\vec{u}] + \vec{\nabla}p + (\rho + \frac{p}{c^2})\nabla\vec{\Phi} = \vec{0},$

$$\Delta\Phi = 4\pi G(\rho + 3\frac{p}{c^2}), p = k\rho, \tag{10}$$

with $k = \frac{1}{3}c^2$ for pure radiation where c is the velocity of light.

In 1-D [30]
$\rho_t + ((\rho + \frac{p}{c^2})u)_x = 0, \quad (\rho + \frac{p}{c^2})(u_t + uu_x) + p_x + up_t + (\rho + \frac{p}{c^2})\Phi_x = 0,$

$$\Phi_{xx} = 4\pi G(\rho + 3\frac{p}{c^2}), p = k\rho, \tag{11}$$

in 3-D [30]
$\rho_t + \vec{\nabla}.((\rho + \frac{p}{c^2})\vec{u}) = \vec{0}, \ (\rho + \frac{p}{c^2})[\vec{u}_t + (\vec{\nabla}.\vec{u}))\vec{u}] + \vec{\nabla}p + \frac{\partial p}{\partial t}\vec{u} + (\rho + \frac{p}{c^2})\nabla\vec{\Phi} = \vec{0},$

$$\Delta\Phi = 4\pi G(\rho + 3\frac{p}{c^2}), p = k\rho. \tag{12}$$

In all these equations ρ is the density of matter (the density of energy in the relativistic case 10-13), $\vec{u}$ is the velocity vector, p the pressure, e the density of total energy per unit mass in (7,9), Φ is the gravitation potential and G is the gravitation constant.

0.3 Sum up and main results.

Part I is made of three chapters in which one considers successively in 1-D, 2-D and 3-D the system of pressureless fluids (1)-(2), the system of collisional (5)-(6) (and collisionless, (3)-(4), as a particular case) self-gravitating fluids and the system of ideal gases (7)-(8). These are among the most classical systems of fluid dynamics. *The aim is to find a convenient numerical scheme valid in 1-D, 2-D and 3-D that gives of course good numerical results, but whose aim is to serve as basic starting point for a theoretical study motivated by the questions raised by P.D. Lax [25], [26] and other authors [17], [28], [34].* This scheme will provide a weak asymptotic method, i.e. an asymptotic method whose discrepancy is intended in the sense of distributions, as introduced by V.D. Danilov, G.A. Omel'yanov and V.M. Shelkovich [12] as an extension in the sense of distributions of Maslov's method. This is some kind of consistency between the scheme and the equations : it provides a sequence of approximate solutions that are plugged into the equations stated in the sense of distributions. Therefore we also call this property "consistency in the sense of distributions" as stated in the foreword. To prove this property, the scheme has to be as simple as possible : indeed it is an order 1 scheme only. L^1 stability and positiveness of the density follow at once from the scheme. This scheme is obtained in three steps : transport step from "free streaming" originating from cosmology and studied in chapter 1 for pressureless fluids, averaging step and finally correction step from a centered discretization. The averaging step is needed to eliminate oscillations due to the centered discretization in the correction step. This scheme is inspired from a convection-correction splitting of equations introduced by Le Roux et al [2]. In Part I, besides the statement of the scheme for the various systems and careful numerical tests showing its accuracy (Sod, Woodward-Colella, Toro, Lax and coworkers,...), one proves consistency in the sense of distributions as far as possible. In the case of the 3-D system of pressureless fluids this proof is completely rigorous as well as in the case of the 1-D system of pressureless self-gravitating fluids. In the other cases one obtains very simple criteria to be checked numerically for a family of values of the space step h that tend to 0. Of course one cannot test an infinity of values of h and the tests are limited to values of h as small as possible. If one admits the verification should hold as well when $h \to 0$ then one can apply the consistency result. All tests give a strong impression one can put faith in this extrapolation. If one limits to the finite number of tested values of h then the proof gives an approximation result showing that the numerical solution from the scheme satisfies the equations up to a small deviation of order 1 in the space step. Now let us describe the contents of each chapter.

• In chapter 1 we consider the system of pressureless fluids. In 1-D the solution of the Riemann problem may contain delta waves. The Godunov method consists in taking an average in each cell (projection step) of the solution of the Riemann problems at the interfaces of cells. This average creates a discontinuity that looks somewhat in contradiction with the concept of cosmic fluid since the delta waves can be close to the interface of cells or even change their location according to minor details of calculation when they are located very close to an interface. We introduce a continuous sharing of these delta waves between left and right cells from the observation of the case in which there is coexistence of a physical solution made of discontinuities and of an unphysical delta wave. The scheme so obtained corresponds exactly to the physical intuition : let the constant state fluid in adjacent cells interpenetrate (which represents exactly the free streaming of the cosmic fluid), then average over each cell in order to have well definite values at each time t_n (the averaging corresponds to the sticking of close enough particles). We observe that we obtain very good numerical results by letting the free streaming take place between several cells (2 or 3) before the averaging, which permits unusually large CFL conditions. In the case one decides that the free streaming is allowed only through a single interface the scheme had already

been noticed in mathematics as a very simple kinetic scheme [4]. In this case we have been able to fully prove the consistency of the scheme which is particularly technical for arbitrary signs of velocity in 2-D (each cell has 8 neighbors) and in 3-D (each cell has 26 neighbors) : therefore direct evaluation is impossible and should be replaced by abstract reasoning. We have obtained

Theorem 1.5.1. *Let the initial conditions be L^1 in density $\rho^0 \geq 0$, more generally a positive bounded Radon measure, and L^∞ in velocity. Then the scheme for system (1)-(2) is well defined, L^1-stable and consistent in the sense of distributions for all positive time in 1-D, 2-D, and 3-D.*

Various numerical simulations are presented : in particular structure formation in 2-D, and the fact that structure formation is far less efficient in presence of expansion, and frozen by too fast expansion (Meszaros effect). These results are obtained with a very large CFL condition $r\|u\|_\infty \leq 2.5$ if u is the velocity and $r = \frac{\Delta t}{\Delta x}$.

• In chapter 2 we consider the system of self-gravitating collisional (presence of pressure) and collisionless (absence of pressure) fluids in 1-D, 2-D and 3-D. In absence of pressure the proof of consistency can be extended in 1-D at the price of a different proof since gravitation can increase the velocity. This proof extends in 2-D and 3-D under the assumption that the gradient of the gravitation potential is bounded, which always holds in 1-D, but not always in 2-D (point concentration of matter) and in 3-D (concentration of matter in points and strings). We have obtained

Theorem 1.7.1. *Let the initial conditions be L^1 in density $\rho^0 \geq 0$, more generally a positive bounded Radon measure, and L^∞ in velocity. Then the scheme for the self-gravitating pressureless system (3)-(4) is well defined, L^1-stable and consistent in the sense of distributions for all positive time in 1-D. This result remains true in 2-D and 3-D as long as the gradient of the gravitation potential remains bounded.*

Anyway, we observe numerically that the scheme works even in cases the gravitation potential is unbounded. Consistency can be proved in presence of pressure as long as the CFL condition $r\|u\|_\infty \leq 1$ holds and the gradient of the gravitation potential remains bounded :

Theorem 1.8.1. *Let the initial conditions be L^1 in density $\rho^0 \geq 0$ and L^∞ in velocity. Then in 1-D, 2-D, and 3-D the scheme for the collisional system (5)-(6) is well defined, L^1-stable and consistent in the sense of distributions as long as the velocity remains bounded (in the CFL condition) and the gradient of the gravitation potential remains bounded.*

These properties are checked numerically in all tests since in presence of pressure one observes that collapse to a point or string appears impossible. One applies the scheme to the numerical simulation of gravitational collapse of clouds of gas : in absence of pressure gravitational collapse to a point is observed in absence of expansion or slow expansion. In the case of a fast expansion one observes absence of gravitational collapse (Meszaros effect). In 2-D when the cloud of gas is rotating one observes creation of some structure looking like a solar system : most matter agglomerates in the center and there appear smaller agglomerations of matter accompanied by some clouds, that rotate around the "sun" located in the center. In presence of pressure one observes numerically Jeans theory : a cloud of gas whose size is large enough collapses gravitationally in spite of pressure while a smaller cloud is smeared and dissipated by pressure.

• In chapter 3 we apply the scheme to the 2-D Riemann problems for ideal gases considered by

P.D. Lax in [25] and [26]. We prove consistency as long as the CFL condition holds (boundedness of the velocity) and as long as the density of total energy remains ≥ 0 (which has always been observed in all tests).

Theorem 2.3.1. *Let the initial conditions ρ^0 and e^0 be positive L^1 functions and the velocity $\vec{u^0}$ be L^∞. Assume that on some time interval $[0,T]$ the velocity is bounded (in the CFL condition) and that the density e of total energy remains ≥ 0. Then concerning the conservation laws the scheme is consistent in the sense of distributions. The consistency in the sense of distributions of the state law takes place in regions in which ρ is strictly positive and in which the approximate solution has the familiar aspect of piecewise $\mathcal{C}^1$ functions having limits on both sides of the surfaces of discontinuity : shock waves, contact discontinuities, rarefaction waves, for instance.*

These assumptions on the boundedness of velocity and positiveness of the density of total energy are immediate to check throughout the iterations and have always been satisfied in all tests. 1-D numerical tests : Sod, Woodward-Colella, Toro, show that the scheme gives the correct result with arbitrary precision and with efficiency, although it is only of order one. For the six 2-D Riemann problems considered by P.D. Lax [25] and [26] the scheme gives exactly the results obtained by the other authors with completely different numerical methods. The proof of consistency of the scheme shows that the numerical results obtained by all schemes (the one in this paper and the schemes mentioned by P.D. Lax in [25] and [26]) represent some approximate solution of the equations of ideal gases with a possible deviation of order one in the space step. This is an encouragement and it suggests to go on the investigations in order to put in evidence a mathematically exact solution of the equations, as developped in Part II.

Part II. After Part I that provides a method to obtain approximate solutions as a mathematical tool for theoretical investigation, Part II consists in using this method to answer to the mathematical problems that motivated this work : *put in evidence rigorously defined mathematical objects that could be proposed as solutions of the equations, prove they correspond to the known or classically accepted solutions, study their existence, uniqueness and numerical approximation.* Of course, once the stability and consistency of the scheme have been proved in Part I, it remains to put in evidence a convenient functional space in which one could pass to the limit by compactness in the approximate solutions from the scheme, show that the limit so obtained is "solution" of the equations in a natural sense, then try to study from "abstract mathematical methods" the problem of existence-uniqueness of solutions in the functional space previously put in evidence. From an analysis of the special relativistic equations (10)-(13) in chapter 4 and, above all, in chapter 5, from an analysis of the singular shocks solutions of the Keyfitz-Kranzer equations [22], [21], [33], [35], [36] one will put in evidence a space of germs of holomorphic functions on a boundary of the real space $\mathbb{R}^n$, which can be interpreted as a particular regularization procedure as well as a holomorphic version of the Egorov spaces of generalized functions [15]. In this space of holomorphic germs one can use convergence and compactness. The approximate solutions from the scheme are extended as holomorphic germs. An analog of the classical result "stability and consistency imply convergence" is proved by compactness. The results in Part I permit to apply this convergence result and obtain convergence to a proposed solution of the equations. Unfortunately the problem of uniqueness remains open and our various theoretical attempts to state the equations in a more precise way -on physical ground- so as to guarantee existence and uniqueness for the Cauchy problem have failed so far. Existence-uniqueness results have been obtained but not for the above equations or, when the above equations are concerned,

we only recover the known cases of regular solutions. In short we can mathematically justify -to some reasonable extent- the numerical facts that are observed, but we have failed on the problem of uniqueness, even in the search of more precise formulations of the equations on physical ground that would ensure existence and uniqueness of the irregular solutions. Now we describe the contents of the various chapters.

• In chapter 4 we consider the equations that rule a radiation dominated universe as our universe during the period from soon after the Big Bang to the time of decoupling 380000 years later (13 billion years before the present time) when the cosmic microwave background was created. The importance of these equations come from the fact that the seeds of the to-day universe were created during this period. General Relativity is not indispensible since the fields are weak. Therefore the equations proposed in cosmology [30], [42] are issued from special relativity. The universe is approximately regarded as a perfect fluid because of the very large scale of length used by the observers. The Euler equation in (10)-(13) are in nonconservative form which makes an important difference with the equations considered up to now in Part I. For equations in nonconservative form the classical Rankine-Hugoniot conditions do not hold as in the conservative case. Discontinuous solutions of these equations do not make sense within the theory of distributions and one cannot obtain the jump conditions by a mere integration as in the classical case. In order to obtain jump conditions for equations (10)-(13) we propose a regularization that permits to give a mathematical sense to the equations when the solutions are the regularized objects. The classical Heaviside function $H(x)$ is replaced by a function $H(x,\epsilon)$ where ϵ is a regularizing parameter so that $H(x,\epsilon)$ tends to $H(x), x \neq 0$, when $\epsilon \to 0$. The calculations on the regularized objects make sense. The problem to obtain well defined jump conditions at the limit $\epsilon \to 0$ is not directly solved by the regularization. It has been solved on physical ground simply by observing that physicists do nonlinear calculations to obtain the relativistic continuity and Euler equations, so that we state these equations in a stronger form than the state law whose validity appears to be far less precise. This statement gives nonambiguous jump conditions for systems (10)-(11) and (12)-(13) :

Theorem 4.4.1. *The system (10)-(11) of special-relativistic fluids, with G=0 and in one space dimension, admits step functions solutions when stated in the following form, where the state law is satisfied only in a weak sense*

$$\rho_t + ((\rho + \frac{p}{c^2})u)_x = 0,\ (\rho + \frac{p}{c^2})(u_t + uu_x) + p_x = 0,\ p \overset{weak}{=} k\rho.$$

Besides the classical jump condition of the conservative continuity equation, the shock waves satisfy the nonclassical jump condition

$$V\Delta p = c^2(\rho_l + \frac{p_l}{c^2})(V - u_l)(exp\frac{V\Delta u}{c^2} - 1)$$

which follows from the nonconservative Euler equation. As a consequence the Euler equation can equivalently be stated in the form $u_t + uu_x + \frac{p_x}{\rho + \frac{p}{c^2}} = 0$ *(these two formulations are found in texts of cosmology). Similar results with a different second formula hold for the system (12)-(13).*

We check that adaptations of the numerical scheme used in Part I give the explicit jump conditions so obtained, with a very good approximation in the physical domain under concern. We observe that on this domain the two approximate systems (10)-(11) and (12)-(13) give approximately the same numerical solution. It is convenient to consider $H(x,\epsilon)$ as a real analytic function in x and ϵ in order to benefit of the uniqueness of analytic continuation in the statement of the

space of germs that we introduce for the explicit calculations. This shows the relevance of holomorphy in this context and opens the way for the next chapter.

• The main purpose of chapter 5 is to put in evidence a functional space in which we will find by compactness mathematical objects that could satisfy the equations in a natural sense and appear as limits of the numerical scheme. In Part I it has been rigorously proved or observed from numerical calculations that L^1-stability and consistency hold. It remains to prove in a suitable functional space that "L^1-stability and consistency imply convergence". The singular shock solutions of the Keyfitz-Kranzer equation will permit to put in evidence such a functional space, as an improvement of the space of holomorphic germs introduced in the previous chapter for the need of explicit calculations. The scheme of Part I provides approximate solutions (chapter 6 below). In a singular shock solution of the Keyfitz-Kranzer equations

$$u_t + (u^2 - v)_x = 0, v_t + (\frac{1}{3}u^3 - u)_x = 0,$$

the function v is a delta wave i.e. it carries a Dirac delta measure over the discontinuity. The function u is a mere discontinuity with very small peaks of measure 0 located on the discontinuity. Therefore in distribution theory the function u is equivalent to a mere discontinuity. The facts that the function u is a mere discontinuity and that the function v is a delta wave, both travelling with the same constant speed, are incompatible with the equations : for instance in the first one $u^2 - v$ would contain the Dirac function of v that would not be compensated in u_t. This shows that one has to dissociate the numerical solution u observed in the sense of distributions, which is a mere discontinuity in distribution theory, from the "genuine" solution u, which is not a distribution. Indeed one observes two small peaks in the discontinuity of u, that are negligible in the sense of distributions, but whose participation in u^2, u^3 become essential to compensate the delta wave in the variable v. In the space of holomorphic germs suggested in chapter 4 for the study of explicit solutions of equations of relativistic cosmology these small peaks make sense and permit the equations to hold because they have a significative contribution in u^2 and u^3 and can compensate the Dirac delta distribution in v. In these holomorphic germs one can define a family of Banach spaces and a concept of compactness so that, to any L^1 bounded sequence of step functions which are approximate solutions (for instance the approximate solutions obtained from the scheme) we can associate holomorphic germs which are solutions of the equations in a natural sense provided the given sequence of step functions satisfies the property of consistency, i.e. we obtained a result of the form "L^1-stability and consistency imply convergence". For simplicity we state the theorem in the case of a system of two scalar conservation laws

$$u_t + (f(u,v))_x = 0, \quad v_t + (g(u,v))_x = 0.$$

Theorem 5.3.1. *Under the assumptions of L^1-stability and consistency in the sense of distributions the approximate solutions, denoted (u_n, v_n), satisfy the following :*
there exists a subsequence of the sequence (u_n, v_n), still denoted (u_n, v_n) to simplify the notation, two sequences $(U_n), (V_n)$ of elements of the functional space on $\mathbb{R}\times]0,T[$ and a pair U, V of elements of the functional space such that
i)$\forall n$, *U_n, V_n have the "real interpretations" u_n, v_n respectively (the real interpretations are obtained by letting the regularization parameter tend to 0),*
ii) U, V have the "real interpretation" u, v respectively,
iii) $U_n \to U, V_n \to V$ in the functional space,
iv) the pair (U, V) is a weak solution of the equations in a natural sense.

This applies to an arbitrary number of equations in 1-D, 2-D and 3-D, in particular to all equations considered up to now and in any dimension. When using the scheme considered in this work the numerical results -which have always been observed to approximate the exact ones when exact solutions are known and to be in agreement with physics in absence of previously known solutions- are approximations of the exact "solution" put in evidence in theorem 6. The problem is that these proposed "solutions" are some kind of weak solutions and therefore as usual for weak solutions they suffer from a lack of uniqueness. Various attempts have failed to solve this problem.

• In chapter 6 we extend the scheme considered in Part I and its consistency proof (or consistency criterion when proofs are replaced by numerical tests) to a rather large family of conservation laws, those of the form

$$(u_i)_t + (u_i \Phi(U))_x = (A(U))_x,$$

$U = (u_1, \ldots, u_n)^T$, and their natural multidimensional extensions. The left member is a degenerate system in which $\Phi(U)$ plays the role of the numerical velocity that we consider in the transport step. As in Part I, the scheme is completed by a centered discretization of the right hand side members with, in between, an averaging step imposed by the centered discretization. The scheme applies to the Keyfitz-Kranzer equations.

$$u_t + (u^2 - v)_x = 0, v_t + (\frac{1}{3}u^3 - u)_x = 0,$$

stated in the form above with $\Phi(U) = u$.

• In chapter 7 we extend a classical method to the functional space used in theorem 6 : one introduces generalized Sobolev spaces in which one can use standard tools such as coercivity, the Lax-Milgram theorem, minimisation of convex functionals. A nontrivial extension is possible and yields an extension of the classical results of existence-uniqueness for elliptic boundary value problems in case of possibly very irregular data (our Sobolev spaces contain the distributions with compact support for instance). This was motivated by the presence of the Poisson equation inside the equations of self-gravitating fluids and it yields existence-uniqueness for other linear equations without solution in a more classical context. This extension of the classical methods has been done only in 1-D. The theorems look like the classical ones but take place in our generalized Sobolev spaces. It is clear all the results could be extended to several dimensions taking inspiration from the classical theory. This shows that the space of holomorphic germs is suitable for extensions of classical methods but up to now these results are limited to linear equations.

Première partie

Basic equations of fluid dynamics : a numerical scheme and consistency in the sense of distributions

Chapitre 1

Pressureless fluid dynamics

Some systems of PDE's, such as the one of pressureless fluid dynamics, show delta waves in the solution of the Riemann problem. A method of projection of these delta waves in Godunov's scheme is proposed. It provides a modification of the original *Le Roux et al.* scheme in case of changes in sign of velocity. Stability and convergence of the scheme are proved in one space dimension for the system of pressureless fluids. As an application, this method has been extended to classical systems of fluid dynamics, used for the numerical simulation of large structure formation in cosmology, in presence of expansion of the background. This method of projection of delta waves can also be applied to systems of conservation laws that have delta waves in the solution of the Riemann problem.

1.1 Introduction.

In [1] the authors noticed that the solution of the Riemann problem for the system of pressureless fluid dynamics

$$\rho_t + (\rho u)_x = 0, \tag{1.1}$$

$$(\rho u)_t + (\rho u^2)_x = 0, \tag{1.2}$$

shows a delta wave located on the discontinuity of the solution. Nevertheless, they succeeded to extend the Godunov scheme to this case, and obtained excellent numerical results. After the pioneering article [2], various numerical methods have been proposed for the numerical solution of system (1.1)-(1.2). References are given in [3], [4], [7] and [27]. In the Godunov methods the delta waves are projected on the cell in which they are located. This method of projection lacks continuity relatively to the initial conditions, since infinitesimal variations can change the location of the delta waves when they are close to the interfaces of meshes. The method in [2] has been modified in this chapter to fit with physical intuition at the level of cells for the case of a cosmic fluid modelled by pressureless material [8] p.34, p.210. The idea developped here is the following : in this case the cosmic fluid is made of collisionless particles that interact through gravitation only. Therefore the above lack of continuity looks irrealistic at the level of cells in the case of large structure formation in cosmology.

To obtain a continuous flux as in usual Godunov schemes, in the projection step of the Godunov method, one has to share the delta waves into left-hand-side and right-hand-side contri-

butions. In the solution of the Riemann problem there occur two cases. In the first case, we have only one solution of the Riemann problem. It is made of a physically meaningful delta wave, that we do not know *a priori* how to share. In the second case, one has two possible solutions : a physical one that has a classical form (step functions without delta waves), and an unphysical one involving a delta wave. In this second case, one obtains a Godunov scheme from the physical solution in form of step functions, which permits to compute the formulas governing the sharing of the unphysical delta wave that would lead to the same scheme. The method in this chapter consists in applying the same formulas in the first case, when the unique solution is in form of a delta wave. This method gives back, after some calculations, a very natural scheme from the physical viewpoint (Theorem 1.5.1) : let the constant-state fluids in adjacent cells "interpenetrate", then average over the overlapping states.

One could conjecture that this method works for pressureless fluid dynamics because by chance it gives this very natural scheme. An example is given in which the formulas of the sharing of delta waves are different from those in the case of pressureless fluids and it is checked that the scheme gives the exact solution (figure 1.9.4). The above method of sharing delta waves can be applied to systems of PDE's for which there is some coexistence of delta waves and classical waves as described above in case of system (1.1)-(1.2).

In this chapter one proves stability and consistency of order one of the scheme, i.e. that the scheme provides a weak asymptotic method of order one in the sense [12] for any configuration of the velocity field in one space dimension with initial condition any positive Radon measure of finite mass in density and any L^∞ function in velocity. The proof is new and relies completely on the very specific form of the scheme. The proof extends to two and three space dimension and also in expanding background.

1.2 Description of the numerical scheme in [2].

Standard 1D notation is used : the space cells are the segments $[ih - \frac{h}{2}, ih + \frac{h}{2}], i \in \mathbb{Z}$, the space step is denoted by h and the time step by Δt ; we set $t_n = n\Delta t$ and $r = \frac{\Delta t}{h}$. The constant values of ρ and u on the cell $[ih - \frac{h}{2}, ih + \frac{h}{2}]$ at time t_n are denoted by ρ_i^n and u_i^n. In the scheme in [2] the passage from $(\rho_i^n, u_i^n)_{i\in\mathbb{Z}}$ to $(\rho_i^{n+1}, u_i^{n+1})_{i\in\mathbb{Z}}$ is done as follows. One introduces three intermediate values (attached to the junctions of cells)

$$w_{i+\frac{1}{2}}^n = \sqrt{\rho_i^n} u_i^n + \sqrt{\rho_{i+1}^n} u_{i+1}^n \tag{1.3}$$

and

$$u_{i+\frac{1}{2}}^n, \rho_{i+\frac{1}{2}}^n,$$

defined by :

- if $u_i^n \geq 0$ and $u_{i+1}^n \geq 0$ then $u_{i+\frac{1}{2}}^n = u_i^n, \rho_{i+\frac{1}{2}}^n = \rho_i^n$,
- if $u_i^n > 0$ and $u_{i+1}^n < 0$,

if $w_{i+\frac{1}{2}}^n > 0$ then $u_{i+\frac{1}{2}}^n = u_i^n, \rho_{i+\frac{1}{2}}^n = \rho_i^n$,
if $w_{i+\frac{1}{2}}^n < 0$ then $u_{i+\frac{1}{2}}^n = u_{i+1}^n, \rho_{i+\frac{1}{2}}^n = \rho_{i+1}^n$,

- if $u_i^n \leq 0$ and $u_{i+1}^n \geq 0$ then $u_{i+\frac{1}{2}}^n = 0, \rho_{i+\frac{1}{2}}^n = 0$,

- if $u_i^n \leq 0$ and $u_{i+1}^n \leq 0$ then $u_{i+\frac{1}{2}}^n = u_{i+1}^n, \rho_{i+\frac{1}{2}}^n = \rho_{i+1}^n$.

Finally one computes the values $(\rho_i^{n+1}, u_i^{n+1})_{i\in\mathbb{Z}}$ from the formulas

$$\rho_i^{n+1} = \rho_i^n - r\rho_{i+\frac{1}{2}}^n u_{i+\frac{1}{2}}^n + r\rho_{i-\frac{1}{2}}^n u_{i-\frac{1}{2}}^n, \tag{1.4}$$

$$(\rho u)_i^{n+1} = \rho_i^n u_i^n - r\rho_{i+\frac{1}{2}}^n (u_{i+\frac{1}{2}}^n)^2 + r\rho_{i-\frac{1}{2}}^n (u_{i-\frac{1}{2}}^n)^2, \tag{1.5}$$

$$u_i^{n+1} = \frac{(\rho u)_i^{n+1}}{\rho_i^{n+1}}. \tag{1.6}$$

In the case $u_i^n > 0$ and $u_{i+1}^n < 0$, if $w_{i+\frac{1}{2}}^n = 0$, then the two possible values of ρ_i^{n+1} differ by a significative quantity. As an example, let the values of ρ_{i-1}^n, ρ_i^n and ρ_{i+1}^n be equal $(= \rho)$, $u_{i-1}^n = 1, u_i^n = 1$ and $u_{i+1}^n = -1 - \epsilon$; then one computes that the scheme gives $\rho_i^{n+1} = (1 + 2r + r\epsilon)\rho$. Now, if one changes u_i^n and u_{i+1}^n into $u_i^n = 1 + \epsilon$ and $u_{i+1}^n = -1$, one computes $\rho_i^{n+1} = (1 - r\epsilon)\rho$, which differs from the previous value by a quantity $2r\rho$ (when $\epsilon \to 0$, which makes the two possibilities undistinguishable while $2r\rho$ is not at all small). In the case $w_{i+\frac{1}{2}}^n = 0$ one can take an average.

In the case $u_i^n > 0$ and $u_{i+1}^n < 0$ there is a collision of two volumes of fluid. From formulas (1.8)-(1.13) below, in this case the solution of the Riemann problem is made of a delta wave whose velocity has the sign of $w_{i+\frac{1}{2}}^n$. It is easy to check that in the scheme above this delta wave is projected on the cell in which it is located. This looks physical in hydrodynamics [2] but not in cosmology. As an example in the collision of two galaxies there is no star collision but interpenetration of the two galaxies.

1.3 Solution of the Riemann problem.

The formulas of the solution of the Riemann problem for the system of pressureless fluid dynamics can be found in [4] and [27]. The values of (ρ, u) are (ρ_l, u_l) on the left-hand-side of the initial discontinuity located at $x = 0$ and (ρ_r, u_r) on the right-hand-side. If w is any variable, we set $\Delta w = w_r - w_l$. We set

$$u(x,t) = u_l + \Delta u H(x - ct), \tag{1.7}$$

$$\rho(x,t) = \rho_l + \Delta\rho H(x - ct) + \alpha t\delta(x - ct), \tag{1.8}$$

$$(\rho u)(x,t) = (\rho u)_l + \Delta(\rho u) H(x - ct) + \beta t\delta(x - ct), \tag{1.9}$$

$$(\rho u)_l = \rho_l u_l, (\rho u)_r = \rho_r u_r, u = \frac{(\rho u)}{\rho}, \tag{1.10}$$

where H is the Heaviside function and δ is the Dirac delta function. The velocity u is discontinuous at $x = ct$, while ρ and ρu display a δ-peak on the discontinuity, which is proportional to time.

Calculations give [4] and [27] :

$$c = \frac{\sqrt{\rho_r}u_r + \sqrt{\rho_l}u_l}{\sqrt{\rho_r} + \sqrt{\rho_l}}, \tag{1.11}$$

$$\alpha = -\sqrt{\rho_l \rho_r}\Delta u, \tag{1.12}$$

$$\beta = c\alpha. \tag{1.13}$$

In the case $u_l > u_r$ one has $\Delta u < 0$, therefore $\alpha > 0$, as requested since the density ρ cannot be < 0. But in the case $u_l < u_r,\ \Delta u > 0$, therefore $\alpha < 0$, which is not acceptable for a density. Therefore the solution (1.7)-(1.10) is not physically acceptable in the case $u_l < u_r$ (one also finds it is unstable). Fortunately, in this case, one finds another solution, which is physically acceptable [2], [4] and [27] :

- if $x < u_l t$ then $u(x,t) = u_l, \rho(x,t) = \rho_l$ (left-hand-side region),
- if $u_l t < x < u_r t$ then $u(x,t)$ undefined, $\rho(x,t) = 0$ (void region),
- if $x > u_r t$ then $u(x,t) = u_r, \rho(x,t) = \rho_r$ (right-hand-side region).

(1.14)

This solution corresponds to the physics of the problem : in absence of pressure the two sides depart each other with their respective velocities.

1.4 Projection of delta waves.

When a function is regular enough, say L^∞, one usually projects it on a discretization lattice by taking its mean value on each cell $[ih - \frac{h}{2}, ih + \frac{h}{2}]$. This method lacks continuity when a delta wave is located close to an interface. Such a delta wave that, within the unavoidable uncertainty, would be located on the interface, could be as well attributed to any side. In presence of delta waves, the knowledge of the function itself is not sufficient to permit a correct projection on a discretization lattice. The presence of delta waves in the solution (1.7)-(1.9) of Riemann problems for the equations (1.1)-(1.2) and for the systems of physics in [8], [30] and [31] therefore makes the projection step of a Godunov scheme nontrivial. The delta waves from the Riemann problems should have non trivial right-hand-side and left-hand-side contributions to be discovered.

How can we treat the delta wave in the projection step of a Godunov scheme in the case $u_l > u_r$? The idea developed here is the following :

• In the case $u_l < u_r$, one applies the classical Godunov scheme using the solution (1.14) which has the usual form of step functions.

• Still in this case $u_l < u_r$, one seeks how to share the (unphysical) delta-waves in ρ and ρu in $(1.8) - (1.9)$, so as to obtain the classical Godunov scheme. The delta wave in ρ is assumed to contribute to the left-hand-side cell by a factor λ_l and to the right-hand-side cell by a factor λ_r, with $\lambda_l + \lambda_r = 1$. Same for the delta wave in (ρu) whose explicit contributions are proportional to factors μ_l and μ_r, with $\mu_l + \mu_r = 1$. We compute explicitely the values $\lambda_l, \lambda_r, \mu_l$ and μ_r that give the same numerical scheme as the one from the step functions solution.

• Now, in the case $u_l > u_r$, for each configuration of the waves, one adopts the same formulas for $\lambda_l, \lambda_r, \mu_l$ and μ_r to share the delta waves into left-hand-side and right-hand-side contributions.

In the sequel of this section the sharing coefficients $\lambda_l, \lambda_r, \mu_l$ and μ_r are calculated in the case $u_l < u_r$ as functions of the variables u_l, u_r, ρ_l and ρ_r. We denote by $\overline{w_l}$ (respectively $\overline{w_r}$) the mean value of a variable w on the segment $[-\frac{h}{2}, 0]$ (resp. $[0, \frac{h}{2}]$).

• Case : $0 < u_l < u_r$. In this case $u_l < c < u_r$ from (11). Projection of the step functions (two discontinuities of velocities u_l and u_r, provided the CFL condition $u_r\Delta t < \frac{h}{2}$ i.e. $ru_r < \frac{1}{2}$) gives from (1.14) :

$$\begin{aligned}
\overline{\rho_l} &= \rho_l, \\
\overline{(\rho u)_l} &= (\rho u)_l, \\
\overline{\rho_r} = \frac{\rho_l u_l \Delta t + \rho_r(\frac{h}{2} - u_r\Delta t)}{\frac{h}{2}} &= \rho_r + 2r\rho_l u_l - 2r\rho_r u_r, \\
\overline{(\rho u)_r} &= (\rho u)_r + 2r\rho_l u_l^2 - 2r\rho_r u_r^2.
\end{aligned}$$

Projection of the (unphysical) delta wave gives from (1.8)-(1.10) :

$$\begin{aligned}
\overline{\rho_l} = \frac{\rho_l \frac{h}{2} + \lambda_l \alpha \Delta t}{\frac{h}{2}} &= \rho_l + 2r\lambda_l\alpha, \\
\overline{(\rho u)_l} = \frac{(\rho u)_l \frac{h}{2} + \mu_l \beta \Delta t}{\frac{h}{2}} &= (\rho u)_l + 2r\mu_l\beta, \\
\overline{\rho_r} = \frac{\rho_l c\Delta t + \lambda_r \alpha \Delta t + \rho_r(\frac{h}{2} - c\Delta t)}{\frac{h}{2}} &= \rho_r + 2rc(\rho_l - \rho_r) + 2r\lambda_r\alpha, \\
\overline{(\rho u)_r} = \frac{(\rho u)_l c\Delta t + \mu_r \beta \Delta t + (\rho u)_r(\frac{h}{2} - c\Delta t)}{\frac{h}{2}} &= (\rho u)_r + 2rc((\rho u)_l - (\rho u)_r) + 2r\mu_r\beta.
\end{aligned}$$

Identification of the two sets of formulas gives

$$\begin{aligned}
\lambda_l &= 0, \\
\mu_l &= 0, \\
\rho_l u_l - \rho_r u_r &= (\rho_l - \rho_r)c + \lambda_r\alpha, \\
\rho_r u_r^2 - \rho_l u_l^2 &= ((\rho u)_l - (\rho u)_r)c + \mu_r\beta.
\end{aligned}$$

Using (1.11)-(1.13), the last two formulas give, after immediate calculation, $\lambda_r = 1$ and $\mu_r = 1$. Therefore, in this case, the sharing coefficients are

$$\lambda_l = 0, \lambda_r = 1, \mu_l = 0, \mu_r = 1.$$

This means that in this case the delta wave contributes only to the right-hand-side.

• Case $u_l < u_r < 0$. In this case one obtains similarly as above

$$\lambda_l = 1, \lambda_r = 0, \mu_l = 1, \mu_r = 0.$$

• Case $u_l < 0 < u_r$ and $c > 0$. Projection of the step functions (provided the CFL condition $max(|u_l|, u_r)\Delta t < \frac{h}{2}$ i.e. $rmax(|u_l|, u_r) < \frac{1}{2}$) gives from (1.14) :

$$\begin{aligned}
\overline{\rho_l} = \frac{\rho_l(\frac{h}{2} + u_l\Delta t)}{\frac{h}{2}} &= \rho_l + 2r\rho_l u_l, \\
\overline{(\rho u)_l} = \frac{(\rho u)_l(\frac{h}{2} + u_l\Delta t)}{\frac{h}{2}} &= (\rho u)_l + 2r(\rho u)_l u_l, \\
\overline{\rho_r} = \frac{\rho_r(\frac{h}{2} - u_r\Delta t)}{\frac{h}{2}} &= \rho_r - 2r\rho_r u_r, \\
\overline{(\rho u)_r} = \frac{(\rho u)_r(\frac{h}{2} - u_r\Delta t)}{\frac{h}{2}} &= (\rho u)_r - 2r(\rho u)_r u_r.
\end{aligned}$$

Projection of the (unphysical) delta wave gives from (1.7)-(1.13) exactly the same results as in the first case obtained above in the case $0 < u_l < u_r$.

Identification of the two sets of formulas gives

$$\begin{aligned}\rho_l u_l &= \lambda_l \alpha,\\ \rho_l u_l^2 &= \mu_l \beta,\\ -\rho_r u_r &= \rho_l c + \lambda_r \alpha - \rho_r c,\\ -(\rho u)_r u_r &= (\rho u)_l c + \mu_r \beta - (\rho u)_r c.\end{aligned}$$

Thus one obtains the formulas for the left-hand-side and right-hand-side contributions of the delta wave

$$\lambda_l = \frac{\rho_l u_l}{\alpha}, \lambda_r = \frac{-\rho_r u_r + c(\rho_r - \rho_l)}{\alpha}, \mu_l = \frac{\rho_l u_l^2}{\beta}, \mu_r = \frac{-\rho_r u_r^2 + c(\rho_r u_r - \rho_l u_l)}{\beta}. \tag{1.15}$$

- Case $u_l < 0 < u_r$ and $c < 0$. Similar calculations give

$$\lambda_l = \frac{\rho_l u_l + c(\rho_r - \rho_l)}{\alpha},\ \lambda_r = \frac{-\rho_r u_r}{\alpha},\ \mu_l = \frac{\rho_l u_l^2 + c(\rho_r u_r - \rho_l u_l)}{\beta},\ \mu_r = \frac{-\rho_r u_r^2}{\beta}.$$

In summary, the rule of splitting of the unphysical delta wave observed in the case $u_l < u_r$. The splitting of the unphysical delta wave into a left-hand-side contribution and a right-hand-side contribution depends on the (left or right-hand-side) positions of the three waves under concern : the discontinuities of velocities u_l, u_r and the delta wave of velocity c. Looking at the above four cases in which $u_l < u_r$ one arrives at the conclusion that the following rule always hold to evaluate the λ_l and λ_r factors in the contribution of the delta wave in ρ :
- λ-contribution to the side where the physical discontinuity of velocity u_l is located : $\frac{\rho_l u_l}{\alpha}$;
- λ-contribution to the side where the physical discontinuity of velocity u_r is located : $-\frac{\rho_r u_r}{\alpha}$;
- λ-contribution to the side where the delta wave of velocity c is located : $c\frac{\rho_r - \rho_l}{\alpha}$.

Note that the respective contributions are null if $u_r = 0$, $u_l = 0$ or $c = 0$: there is no ambiguity when a wave lies at the interface.

For example, in the case $u_l < 0 < u_r$ and $c > 0$,
- the wave of velocity u_r contributes to the right-hand-side, i.e. to λ_r, by $-\frac{\rho_r u_r}{\alpha}$;

the wave of velocity u_l contributes to the left-hand-side, i.e. to λ_l, by $\frac{\rho_l u_l}{\alpha}$;
the wave of velocity c contributes to λ_r by $c\frac{\rho_r - \rho_l}{\alpha}$.
To summarize all contributions : $\lambda_l = \frac{\rho_l u_l}{\alpha}$ and $\lambda_r = -\frac{\rho_r u_r}{\alpha} + c\frac{\rho_r - \rho_l}{\alpha}$. One recovers (1.15).

For the μ_l and μ_r factors in the contribution of the delta wave in ρu the rule is :
- μ-contribution to the side where the wave of velocity u_l is located : $\frac{\rho_l u_l^2}{\beta}$,
- μ-contribution to the side where the wave of velocity u_r is located : $-\frac{\rho_r u_r^2}{\beta}$,
- μ-contribution to the side where the delta wave of velocity c is located : $c\frac{\rho_r u_r - \rho_l u_l}{\beta}$.

The method in this chapter consists in adopting this rule (obtained in the classical case $u_l < u_r$) in the (unknown) case $u_l > u_r$ for the splitting of delta waves. How can it be justified ? One could think that the proper formulas for the projection of delta waves are the same whether they are physical or unphysical. This rule will be validated by the physical interpretation (Theorem 1.5.1), the numerical tests (section 1.8) and the convergence proof of the scheme (Theorem 1.8.1). Then this modified Godunov method can be exploited in physics (section 1.9) and applied to other systems (end of section 1.7). For some of them the splitting formulas are different from those obtained with the system of pressureless fluids (figure 1.9.4).

1.5 Interpretation of the splitting rule.

In the case $u_r > u_l$, in which the formulas were obtained, the solution displays a void region separated by two discontinuities of velocities u_l and u_r. In the case $u_r < u_l$ one has instead some phenomenon looking intuitively like a collision of two volumes of fluid. Does the adopted splitting rule allow an intuitive interpretation in the collision case ?

Theorem 1.5.1. *The Godunov type scheme with delta wave splitting defined above amounts to the following method in the variables ρ and ρu : a free streaming step followed by an averaging step.*

The free streaming step consists in letting matter from any cell cross freely the interfaces with the neighbor cells and penetrate freely through the matter in these cells. The averaging step consists in taking an average of the matter present in each cell.

The simplest version of the scheme (CFL : $r\|u\|_{L^\infty} \leq 1$) consists in letting the matter enter only in the immediate neighbor cells (on left and right). Then numerical tests showed the relevance of letting the matter cross p successive cells in the first step, thus giving CFL conditions $r\|u\|_{L^\infty} \leq p$. The scheme so obtained gives usually good numerical results for $p = 2$ and $p = 3$ (figures 1.9.1 and 1.9.2). The scheme degenerates for larger values of p.

proof. For case $0 < u_r < u_l$, from (1.11) one has $0 < u_r\Delta t < c\Delta t < u_l\Delta t < \frac{h}{2}$. The projection in case of interpenetration of the two fluids gives $\overline{\rho_l} = \rho_l$ and $\overline{\rho_r} = \frac{\rho_l u_l \Delta t + \rho_r(\frac{h}{2} - u_r\Delta t)}{\frac{h}{2}} = \rho_r + 2r\rho_l u_l - 2r\rho_r u_r$. The splitting rule gives $\lambda_r = \frac{\rho_l u_l}{\alpha} - \frac{\rho_r u_r}{\alpha} + \frac{c(\rho_r - \rho_l)}{\alpha}$, and then $\overline{\rho_r} = \frac{\rho_l c\Delta t + \lambda_r \alpha \Delta t + (\frac{h}{2} - c\Delta t)\rho_r}{\frac{h}{2}}$. Same formulas hold for ρu. Same results are obtained in the case $u_r < u_l < 0$.

For case $u_r < 0 < u_l$ and $c > 0$, the splitting rule gives $\lambda_l = -\frac{\rho_r u_r}{\alpha}$, $\lambda_r = \frac{\rho_l u_l + c(\rho_r - \rho_l)}{\alpha}$, $\mu_l = -\frac{\rho_r u_r^2}{\beta}$ and $\mu_r = \frac{\rho_l u_l^2 + c(\rho_r u_r - \rho_l u_l)}{\beta}$.

Then, from the splitting rule,

$$\begin{aligned}
\overline{\rho_l} &= \frac{\rho_l \frac{h}{2} + \lambda_l \alpha \Delta t}{\frac{h}{2}} = \rho_l - 2r\rho_r u_r, \\
\overline{\rho_r} &= \frac{\rho_r(\frac{h}{2} - c\Delta t) + c\Delta t \rho_l + \lambda_r \alpha \Delta t}{\frac{h}{2}} = \rho_r - 2r\rho_l u_l, \\
\overline{(\rho u)_l} &= \frac{(\rho u)_l \frac{h}{2} + \mu_l \beta \Delta t}{\frac{h}{2}} = (\rho u)_l - 2r\rho_r u_r^2, \\
\overline{(\rho u)_r} &= \frac{(\rho u)_r(\frac{h}{2} - c\Delta t) + c\Delta t (\rho u)_l + \mu_r \beta \Delta t}{\frac{h}{2}} = (\rho u)_r + 2r\rho_l u_l^2.
\end{aligned}$$

The projection in case of interpenetration gives

$$\begin{aligned}
\overline{\rho_l} &= \frac{-u_r\Delta t \rho_r + \frac{h}{2}\rho_l}{\frac{h}{2}} = \rho_l - 2r\rho_r u_r, \\
\overline{\rho_r} &= \frac{\frac{h}{2}\rho_r + u_l \Delta t \rho_l}{\frac{h}{2}} = \rho_r + 2r\rho_l u_l, \\
\overline{(\rho u)_l} &= \frac{-u_r\Delta t (\rho u)_r + \frac{h}{2}(\rho u)_l}{\frac{h}{2}} = \rho_l u_l - 2r\rho_r u_r^2, \\
\overline{(\rho u)_r} &= \frac{\frac{h}{2}\rho_r u_r + u_l \Delta t \rho_l u_l}{\frac{h}{2}} = \rho_r u_r + 2r\rho_l u_l^2.
\end{aligned}$$

We proceed the same way in the case $u_r < 0 < u_l$ and $c < 0$.□

1.6 A free streaming numerical scheme.

In this section the general scheme in static background is described. At first it is stated in one space dimension. The statement of the scheme depends on the parameter $p = 1, 2, 3, ...$ that represents the number of cells crossed before the averaging step. For $p = 1$ the scheme admits the CFL condition $r\|u\|_{L^\infty} \leq 1$. The scheme stated with the number p, that we will call p-scheme, amounts to allow free streaming up to the p^{th} neighbor cell (both into the left and into the right cells) before the averaging step. Then its CFL condition is $r\|u\|_{L^\infty} \leq p$. Of course for $r\|u\|_{L^\infty} \leq p-1$ it coincides with the (p-1)-scheme. Hence the title of this section. The motivation of the introduction of these variants is that it is observed that the 2 and 3-scheme usually give far better results than the 1-scheme. But a degenerescence has been observed for $p \geq 3$ or 4 according to the test.

As usual the space cells in one dimension are the segments $[ih - \frac{h}{2}, ih + \frac{h}{2}], i \in \mathbb{Z}$. One sets $\Delta t = rh$ and $t_n = n\Delta t$. The constant values of ρ and u on the cell $[ih - \frac{h}{2}, ih + \frac{h}{2}]$ at time t_n are denoted by ρ_i^n and u_i^n respectively. If $a < b$ one sets

$$L(a, b) = max(0, min(1, b) - max(0, a)) \tag{1.16}$$

which is the length of $[0, 1] \cap [a, b]$. The p-scheme is :

$$\rho_i^{n+1} = \sum_{-p \leq \lambda \leq p} \rho_{i+\lambda}^n L(\lambda + ru_{i+\lambda}^n, \lambda + 1 + ru_{i+\lambda}^n), \tag{1.17}$$

$$(\rho u)_i^{n+1} = \sum_{-p \leq \lambda \leq p} (\rho u)_{i+\lambda}^n L(\lambda + ru_{i+\lambda}^n, \lambda + 1 + ru_{i+\lambda}^n), \tag{1.18}$$

$$u_i^{n+1} = \frac{(\rho u)_i^{n+1}}{\rho_i^{n+1}}. \tag{1.19}$$

The notation L allows a synthetic formulation of the transport, without being forced to distinguish several cases depending on the signs of the numerical velocities.Take $p = 1$ for brevity. Then (1.17) can be rewritten as

$$\rho_i^{n+1} := \rho_{i-1}^n L(-1 + ru_{i-1}^n, ru_{i-1}^n) + \rho_i^n L(ru_i^n, 1 + ru_i^n) + \rho_{i+1}^n L(1 + ru_{i+1}^n, 2 + ru_{i+1}^n), \tag{1.20}$$

When the CFL condition $r|u_i^n| \leq 1 \quad \forall i, \ \forall n$ is satisfied, the first term, when multiplied by h, represents the quantity ρ issued from the cell I_{i-1} between times t_n and t_{n+1} that lie in the cell I_i at time t_{n+1}. Indeed, the cell $I_{i-1} = [(i - \frac{3}{2})h, (i - \frac{1}{2})h]$ has been transported according to the vector $ru_{i-1}^n h$, since u_{i-1}^n is the numerical velocity and the duration time is rh. The overlap with the fixed cell $I_i = [(i - \frac{1}{2})h, (i + \frac{1}{2})h]$ has a length of $ru_{i-1}^n h$ if $u_{i-1}^n \geq 0$, 0 if $u_{i-1}^n \leq 0$, taking into account the CFL condition $r|u_{i-1}^n| \leq 1$. From (1.16), one finds $L(-1+ru_{i-1}^n, ru_{i-1}^n) = ru_{i-1}^n$ if $u_{i-1}^n \geq 0$, 0 if $u_{i-1}^n \leq 0$. Division by h is due to the fact that ρ_i^{n+1} is a mean value on cells of length h.

The second term, when multiplied by h, represents the quantity ρ issued from the cell I_i that remain in I_i at time t_{n+1}. Indeed the cell $[(i - \frac{1}{2})h, (i + \frac{1}{2})h]$ has been transported by the vector $ru_i^n h$. The overlap with the fixed cell $[(i - \frac{1}{2})h, (i + \frac{1}{2})h]$ is $h - ru_i^n h$ if $u_i^n \geq 0$, $h + ru_i^n h$ if $u_i^n \leq 0$.

From (1.16) one finds $L(ru_i^n, 1+ru_i^n) = 1 - ru_i^n$ if $u_i^n \geq 0$, $1+ru_i^n$ if $u_i^n \leq 0$.

The third term is similar to the first one : it concerns the quantity ρ issued from the cell I_{i+1} that lies in the cell I_i at time t_{n+1}, with the same verification as above.

One observes void regions so that in numerical practice the denominator in (1.19) is replaced by $\max(\rho_i^{n+1}, 10^{-100})$. The CFL condition is

$$r \max(|u_i^n|) \leq p. \tag{1.21}$$

From the proposition below u_i^n is undefinite in the vacuum points and this undefinite value enters only into factors of the value 0 in (1.17) and (1.18).

Proposition 1.6.1. *$\rho_i^n = 0$ implies $(\rho u)_i^n = 0$.*

proof. From the fact the initial condition is a pair (ρ^0, u^0) then $\rho_i^0 = 0$ implies $(\rho u)_i^0 := \rho_i^0 u_i^0 = 0$. Note from (1.16) that all $L_{i,\lambda}^n := L(\lambda + ru_{i+\lambda}^n, \lambda + 1 + ru_{i+\lambda}^n)$ are positive or null and from (1.17) that all $\rho_{i+\lambda}^n$ are positive or null (immediate induction from the initial condition which is a positive Radon measure). From the positiveness of $\rho_{i+\lambda}^0$ and $L_{i,\lambda}^0$, (1.17) and the assumption, $\rho_i^1 = 0$ imply that $\rho_{i+\lambda}^0 = 0$ or $L_{i,\lambda}^0 = 0$ for $-p \leq \lambda \leq p$. From the above remark that $\rho_i^0 = 0$ implies $(\rho u)_i^0 = 0$ this implies that $(\rho u)_{i+\lambda}^0 = 0$ or $L_{i,\lambda}^0 = 0$, for $-p \leq \lambda \leq p$. From formula (1.18) this implies that $(\rho u)_i^1 = 0$. The lemma is proved for n=1. The result is immediate by induction on n.□

The two dimensional space (x, y) is divided into square cells $C_{i,j}$ of side h and centers $(ih, jh)_{i\in\mathbb{Z}}$: $C_{i,j}$ is the set of all (x, y) such that $ih - \frac{h}{2} < x < ih + \frac{h}{2}$ and $jh - \frac{h}{2} < y < jh + \frac{h}{2}$. We set

$$A(a, b) = L(a, 1+a).L(b, 1+b) \tag{1.22}$$

which is the area of the intersection of the square of vertices $(0,0), (0,1), (1,0), (1,1)$ with the square of vertices $(a, b), (1+a, b), (a, 1+b), (1+a, 1+b)$. The p-scheme permits free streaming through p successive cells and it has the CFL condition $rmax(|u_{i,j}^n|, |v_{i,j}^n|) \leq p$. The formulas of the p-scheme are

$$\rho_{i,j}^{n+1} = \sum_{-p\leq\lambda,\mu\leq p} \rho_{i+\lambda,j+\mu}^n A(\lambda + ru_{i+\lambda,j+\mu}^n, \mu + rv_{i+\lambda,j+\mu}^n), \tag{1.23}$$

$$(\rho u)_{i,j}^{n+1} = \sum_{-p\leq\lambda,\mu\leq p} (\rho u)_{i+\lambda,j+\mu}^n A(\lambda + ru_{i+\lambda,j+\mu}^n, \mu + rv_{i+\lambda,j+\mu}^n), \tag{1.24}$$

$$(\rho v)_{i,j}^{n+1} = \sum_{-p\leq\lambda,\mu\leq p} (\rho v)_{i+\lambda,j+\mu}^n A(\lambda + ru_{i+\lambda,j+\mu}^n, \mu + rv_{i+\lambda,j+\mu}^n), \tag{1.25}$$

$$u_{i,j}^{n+1} = \frac{(\rho u)_{i,j}^{n+1}}{\rho_{i,j}^{n+1}},\ v_{i,j}^{n+1} = \frac{(\rho v)_{i,j}^{n+1}}{\rho_{i,j}^{n+1}}. \tag{1.26}$$

The undefinedness of u, v in vacuum points is no problem (same proof as in the one dimensional case). The scheme in three space dimension is similar. Let $C_{i,j,k}$ be the cube of all (x, y, z) such that $(i - \frac{1}{2})h < x < (i + \frac{1}{2})h, (j - \frac{1}{2})h < y < (j + \frac{1}{2})h, (k - \frac{1}{2})h < z < (k + \frac{1}{2})h$. Let

$$V(a, b, c) = L(a, 1+a).L(b, 1+b).L(c, 1+c) \tag{1.27}$$

be the volume of the intersection of the cube of vertices (i,j,k), $i,j,k=0$ or 1, with the cube of vertices $(a+i,b+j,c+k)$, $i,j,k=0$ or 1. For the p-scheme, if $\omega=\rho,\rho u,\rho v,\rho w$ one sets

$$\omega_{i,j,k}^{n+1}=$$

$$\sum_{-p\leq\lambda,\mu,\nu\leq p}\omega_{i+\lambda,j+\mu,k+\nu}^{n}V(\lambda+ru_{i+\lambda,j+\mu,k+\nu}^{n},\mu+rv_{i+\lambda,j+\mu,k+\nu}^{n},\nu+rw_{i+\lambda,j+\mu,k+\nu}^{n}). \quad (1.28)$$

Stability for the p-scheme (for any p) and consistency for the 1-scheme (i.e. the p-scheme in the case $r\|\vec{u}\|_{L^\infty}\leq 1$) can be proved in three space dimension as an immediate adaptation of the proof of Theorem 1.7.1 for stability, and in section 1.11 for consistency.

1.7 Stability of the p-scheme.

The aim of this section is to give the proof of stability. To simplify the exposition the proof is given in the case $p=1$. It is identical for any p. For convenience let us recall the 1-scheme

$$\rho_i^{n+1}=\rho_{i-1}^{n}L(-1+ru_{i-1}^{n},ru_{i-1}^{n})+\rho_i^{n}L(ru_i^{n},1+ru_i^{n})+\rho_{i+1}^{n}L(1+ru_{i+1}^{n},2+ru_{i+1}^{n}), \quad (1.29)$$

$$(\rho u)_i^{n+1}=(\rho u)_{i-1}^{n}L(-1+ru_{i-1}^{n},ru_{i-1}^{n})+$$

$$(\rho u)_i^{n}L(ru_i^{n},1+ru_i^{n})+(\rho u)_{i+1}^{n}L(1+ru_{i+1}^{n},2+ru_{i+1}^{n}), \quad (1.30)$$

$$u_i^{n+1}=\frac{(\rho u)_i^{n+1}}{\rho_i^{n+1}} \quad (1.31)$$

if $\rho_i^{n+1}\neq 0$. In case $\rho_i^{n+1}=0$ it has been proved in the proposition of section 1.5 that any arbitrary value u_i^{n+1} fits. For convenience in statement of results one chooses a value between $\min_i(u_i^n)$ and $\max_i(u_i^n)$.

The CFL condition is

$$rmax(|u_i^n|)\leq 1. \quad (1.32)$$

lemma 1.7.1. The maximum principle in u holds

$$\min(u_0)\leq u_i^n\leq\max(u_0)\ \forall i. \quad (1.33)$$

proof. If $a\leq u_i^n\leq b\ \forall i$, one proves that

$$a\leq u_i^{n+1}\leq b\ \forall i. \quad (1.34)$$

If $\rho_i^{n+1}=0$ this is the above choice. If $\rho_i^{n+1}>0$ one has to prove from (1.28-1.30) that

$$a\leq\frac{(\rho u)_{i-1}^{n}L(-1+ru_{i-1}^{n},ru_{i-1}^{n})+(\rho u)_i^{n}L(ru_i^{n},1+ru_i^{n})+(\rho u)_{i+1}^{n}L(1+ru_{i+1}^{n},2+ru_{i+1}^{n})}{\rho_{i-1}^{n}L(-1+ru_{i-1}^{n},ru_{i-1}^{n})+\rho_i^{n}L(ru_i^{n},1+ru_i^{n})+\rho_{i+1}^{n}L(1+ru_{i+1}^{n},2+ru_{i+1}^{n})}\leq b.$$

Indeed since $\rho_i^n\geq 0\ \forall i$ (from (1.28)) and since $(\rho u)_i^n=\rho_i^n u_i^n$ (from (1.30) if $\rho_i^n>0$ and from the proposition of section 1.5 for vacuum points) the assumption implies

$$a\rho_i^n\leq(\rho u)_i^n\leq b\rho_i^n\ \ \forall i. \quad (1.35)$$

Since the coefficients L's are positive and the same in the numerator and in the denominator, inequalities (1.34) with $i-1, i$ and $i+1$ prove (1.33). It suffices to consider an induction on n to conclude. □

Theorem 1.7.1 : stability. *Assume that $r\|u_0\|_{L^\infty} \leq 1$. Then on $\mathbb{R}$, ρ is positive and L^1 stable, u is L^∞ stable and satisfies the maximum principle, ρu and ρu^2 are L^1 stable.*

proof. The L^1 stability of ρ follows from (1.28) which is merely a transport. The L^∞ stability of u follows from the lemma. The stabilities of ρ and u imply the stabilities of ρu and ρu^2.□

1.8 Consistency of the 1-scheme in one space dimension.

For $\omega = \rho, u, \rho u$ and ρu^2 we denote by ω_h the step function on $\mathbb{R} \times [0, +\infty[$ whose values on the rectangles $[ih - \frac{h}{2}, ih + \frac{h}{2}[\times[nrh, (n+1)rh[$ are ω_i^n. We skip the indices h to shorten the notation.

Theorem 1.8.1 : consistency. *For $\forall \psi \in C_c^\infty(\mathbb{R}\times]0, +\infty))$ the following limits hold when $h \to 0$*

$$\int (\rho\psi_t + \rho u\psi_x)dxdt = O(h), \tag{1.36}$$

$$\int (\rho u\psi_t + \rho u^2\psi_x)dxdt = O(h), \tag{1.37}$$

i.e. the scheme provides a weak asymptotic method of order one.

proof. First step. If $\omega = \rho$ or ρu a direct calculation gives that $\forall \psi \in C_c^\infty(\mathbb{R}\times]0, +\infty))$

$$\int (\omega\psi_t + \omega u\psi_x)dxdt = -h\sum_{i,n}[\omega_i^n - \omega_i^{n-1} + r((\omega u)_i^{n-1} - (\omega u)_{i-1}^{n-1})]\psi_i^n + hO(1), \tag{1.38}$$

where $\psi_i^n = \psi(ih, nrh)$. Indeed (1.37) is easily obtained using the stability results in Theorem 1.1.7 and Taylor's formula for ψ. The intermediate steps are

$\int (\omega\psi_t + \omega u\psi_x)dxdt = rh^2 \sum_{i,n}(\omega_i^n \frac{\psi_i^{n+1} - \psi_i^n}{rh} + (\omega u)_i^n \frac{\psi_{i+1}^n - \psi_i^n}{h}) + hO(1).$

Then

$\int (\omega\psi_t + \omega u\psi_x)dxdt = -h \sum_{i,n}[\omega_i^n - \omega_i^{n-1} + r((\omega u)_i^n - (\omega u)_{i-1}^n)]\psi_i^n + hO(1).$

Then, one uses that

$\sum_{i,n}[(\omega u)_i^n - (\omega u)_{i-1}^n)]\psi_i^n = \sum_{i,n}[(\omega u)_i^{n-1} - (\omega u)_{i-1}^{n-1})]\psi_i^n + \sum_{i,n}(\omega u)_i^n(\psi_i^n - \psi_{i+1}^n - \psi_i^{n-1} + \psi_{i+1}^{n-1})$, where the last factor is $O(h^2)$.

Remark : positive signs of velocities. If $u_i^{n-1} \geq 0\ \forall i$ then (1.28)-(1.29) give

$$\omega_i^n = \omega_{i-1}^{n-1} r u_{i-1}^{n-1} + \omega_i^{n-1}(1 - r u_i^{n-1}) = \omega_i^{n-1} - r((\omega u)_i^{n-1} - (\omega u)_{i-1}^{n-1}). \tag{1.39}$$

Therefore the first term in the sum in (37) is null, which proves (35)-(36).

Second step : arbitrary signs of velocities. For given index i_0 the value $u_{i_0}^{n-1}$ can be ≥ 0 or ≤ 0.

• If $u_{i_0}^{n-1} \leq 0$ then the "matter ω" in the cell $[(i_0 - \frac{1}{2})h, (i_0 + \frac{1}{2})h]$ at time t_{n-1} goes to the left between t_{n-1} and t_n.
Therefore, after division by h, the amount of matter in the cell i_0-1 is $\omega_{i_0-1}^n = \omega_{i_0}^{n-1}(-ru_{i_0}^{n-1})$+terms not involving $\omega_{i_0}^{n-1}$, and the amount of matter in the cell i_0 is $\omega_{i_0}^n = \omega_{i_0}^{n-1}(1 + ru_{i_0}^{n-1})$+terms not involving $\omega_{i_0}^{n-1}$, where the $\omega_{i_0}^{n-1}$ terms concern respectively received and remaining matter.

Therefore, for fixed n, in the sum $\sum_i \omega_i^n \psi_i^n$ the term $\omega_{i_0}^{n-1}$ occurs in (and only in) $\omega_{i_0}^{n-1}(-ru_{i_0}^{n-1})\psi_{i_0-1}^n + \omega_{i_0}^{n-1}(1 + ru_{i_0}^{n-1})\psi_{i_0}^n$.

Consequently, in the sum $\sum_i [\omega_i^n - \omega_i^{n-1} + r((\omega u)_i^{n-1} - (\omega u)_{i-1}^{n-1})]\psi_i^n$ the term involving $\omega_{i_0}^{n-1}$ is

$\omega_{i_0}^{n-1}(-ru_{i_0}^{n-1})\psi_{i_0-1}^n + \omega_{i_0}^{n-1}(1+ru_{i_0}^{n-1})\psi_{i_0}^n - \omega_{i_0}^{n-1}\psi_{i_0}^n + r(\omega u)_{i_0}^{n-1}\psi_{i_0}^n - r(\omega u)_{i_0}^{n-1}\psi_{i_0+1}^n = r\omega_{i_0}^{n-1}u_{i_0}^{n-1}(-\psi_{i_0-1}^n + \psi_{i_0}^n + \psi_{i_0}^n - \psi_{i_0+1}^n) = r\omega_{i_0}^{n-1}u_{i_0}^{n-1}O(h^2)$

from Taylor's formula applied to ψ.

• Similarly, if $u_{i_0}^{n-1} \geq 0$ one checks that one has again $O(h^2)$ ($O(h^2)$ is the value 0 in this case) in factor of $\omega_{i_0}^{n-1}u_{i_0}^{n-1}$. This is done as follows. Now the matter goes to the right. Therefore

$\omega_{i_0+1}^n = \omega_{i_0}^{n-1}ru_{i_0}^{n-1}$+terms not involving $\omega_{i_0}^{n-1}$,
$\omega_{i_0}^n = \omega_{i_0}^{n-1}(1 - ru_{i_0}^{n-1})$+terms not involving $\omega_{i_0}^{n-1}$.

Therefore in the sum $\sum_i \omega_i^n \psi_i^n$ the term $\omega_{i_0}^{n-1}$ occurs in (and only in) $\omega_{i_0}^{n-1}ru_{i_0}^{n-1}\psi_{i_0+1}^n + \omega_{i_0}^{n-1}(1 - ru_{i_0}^{n-1})\psi_{i_0}^n$.
As a consequence in the sum $\sum_i (\omega_i^n - \omega_i^{n-1} + r((\omega u)_i^{n-1} - (\omega u)_{i-1}^{n-1}))\psi_i^n$ the term involving $\omega_{i_0}^{n-1}$ is

$\omega_{i_0}^{n-1}ru_{i_0}^{n-1}\psi_{i_0+1}^n + \omega_{i_0}^{n-1}(1-ru_{i_0}^{n-1})\psi_{i_0}^n - \omega_{i_0}^{n-1}\psi_{i_0}^n + r(\omega u)_{i_0}^{n-1}\psi_{i_0}^n - r(\omega u)_{i_0}^{n-1}\psi_{i_0+1}^n = r\omega_{i_0}^{n-1}u_{i_0}^{n-1}(\psi_{i_0+1}^n - \psi_{i_0}^n + \psi_{i_0}^n - \psi_{i_0+1}^n) = 0.$

Finally from these two cases the second member of (1.37) appears as
$-h\sum_{i_0,n} \omega_{i_0}^{n-1}u_{i_0}^{n-1}O(h^2) + hO(1) = hO(1)$ from the L^1 stability in ωu.□

This proof extends to three space dimension in which case it becomes rather technical : section 1.11.□

The initial condition. The initial values $(\rho_i^0, u_i^0)_{i\in\mathbb{Z}}$ are obtained as mean values of the data ρ^0, u^0 on the cells in x-space. Since $u^0 \in L^\infty$ and since ρ^0 is a positive Radon measure of finite mass, i.e. a positive continuous linear map on the Banach space of continuous bounded functions on $\mathbb{R}$, there is a difficulty to interpret the product $\rho^0 u^0$ (think at a Dirac mass located on a discontinuity of u^0). The physically significant quantities are the mass (of density ρ^0) and the momentum (of density $\rho^0 u^0$) ; of course the velocity of a concentration of matter is physically well defined which eliminates the above ambiguity and permits a well defined discretization of ρ^0 and $\rho^0 u^0$ on the cells. When a concentration of ρ^0 is located on an interface of cells it is shared arbitrarily into left and right, sharing in the same way $\rho^0 u^0$. Then u^0 is well defined on the non void cells. In the void cells the values u^0 do not matter (from the scheme). One can choose them in between the min and the max of u^0 in nonvoid cells for a more precise formulation of the

maximum principle.

Let $\rho_h^0(x), u_h^0(x)$ denote the step functions on $\mathbb{R}$ which are the discretizations of the initial condition on the cells and let $\rho_h(x,t), u_h(x,t)$ be the step functions solution from the scheme. The initial condition is satisfied in the following natural sense :

Proposition 1.8.1. $\forall \psi \in C_c^\infty(\mathbb{R})$ *one has the following :*

$$\int [\rho_h(x,t) - \rho_h^0(x)]\psi(x)dx = tO(1) + hO(1), \tag{1.40}$$

$$\int [(\rho_h u_h)(x,t) - (\rho_h u_h)^0(x)]\psi(x)dx = tO(1) + hO(1). \tag{1.41}$$

proof. For simplicity in notation we drop again the indices h. Let $\omega = \rho_h$ or $\rho_h u_h$. If $t \in](n-\frac{1}{2})rh, (n+\frac{1}{2})rh[$ let

$$I = \int [\omega(x,t) - \omega^0(x)]\psi(x)dx.$$

Then $I = \sum_i \int_{[(i-\frac{1}{2})h,(i+\frac{1}{2})h]} (\omega_i^n - \omega_i^0)\psi(x)dx = \sum_i (\omega_i^n - \omega_i^0)\psi_i h + hO(1)$ using the L^1 stability in ω. Then

$I = \sum_{p=1}^n \sum_i (\omega_i^p - \omega_i^{p-1})\psi_i h + hO(1)$.

From the following bound of the second member of (1.37) with only $\sum_i$ instead of $\sum_{i,n}$ obtained in the third step in the above proof of convergence with $\psi(x)$ in place of $\psi(x,t)$:

$\sum_i (\omega_i^p - \omega_i^{p-1})\psi_i = \sum_i -r[(\omega u)_i^{p-1} - (\omega u)_{i-1}^{p-1}]\psi_i + \sum_i r(\omega u)_i^{p-1} O(h^2)$,

one has

$I = \sum_{p=1}^n \sum_i -r[(\omega u)_i^{p-1} - (\omega u)_{i-1}^{p-1}]\psi_i h + \sum_{p=1}^n \sum_i r(\omega u)_i^{p-1} O(h^3) + hO(1)$.

Then from the L^1 stability in ωu

$I = \sum_{p=1}^n \sum_i r(\omega u)_i^{p-1}(\psi_{i+1} - \psi_i)h + hO(1) = \sum_{p=1}^n hO(1) + hO(1)$.

Since for fixed t one has $n =$ the integer part of $\frac{t}{rh}$, $I = tO(1) + hO(1)$, which ends the proof. □

Remark 1. The method of projection of delta waves can be applied to systems of conservation laws of the form

$$u_t + (u^2)_x = 0, \ v_t + (f(u))_x + (uv)_x = 0. \tag{1.42}$$

The case $f = 0$ has been studied in [24], [38] and [44] as a system whose solution of the Riemann problem contains delta waves. Plugging

$$u(x,t) = u_l + \Delta u H(x - ct), \ v(x,t) = v_l + \Delta v H(x - ct) + \alpha t \delta(x - ct) \tag{1.43}$$

into (1.41) one obtains a solution of the Riemann problem in the form (1.42) where the formulas of c and α are :

$$c = u_l + u_r,\ \alpha = c\Delta v - \Delta f - \Delta(uv). \tag{1.44}$$

The projection method described in section 1.1 can be used because when $u_l \leq u_r$ the stable solution in u for equation $u_t + (u^2)_x = 0$ is a rarefaction wave therefore there is no longer a delta wave in v. To simplify the calculation this rarefaction wave is replaced by a suitable pair of nonentropic waves having the usual form of discontinuities with constant speed. In this case $u_l \leq u_r$ one considers (unstable) solutions of the Riemann problem of the form :

$$\begin{aligned} u(x,t) &= u_l + (\bar{u} - u_l)H(x - c_l t) + (u_r - \bar{u})H(x - c_r t), \\ v(x,t) &= v_l + (\bar{v} - v_l)H(x - c_l t) + (v_r - \bar{v})H(x - c_r t). \end{aligned} \tag{1.45}$$

The jump formulas are obtained as usual :

$$c_l = u_l + \bar{u}, c_r = u_r + \bar{u}, v_r\bar{u} - u_r\bar{v} + f(\bar{u}) = f(u_r), v_l\bar{u} - u_l\bar{v} + f(\bar{u}) = f(u_l). \tag{1.46}$$

In the case $f(u) = u$ one obtains $\bar{u} = 0$ and $\bar{v} = -1$. Then the formulas for the contributions are respectively :

$$\frac{u_l(v_l + 1)}{\alpha}, -\frac{u_r(v_r + 1)}{\alpha}, \frac{c(v_r - v_l)}{\alpha}. \tag{1.47}$$

The projection gives a scheme different from the one described in Theorem 1.5.1. It is compared with the exact solution in figure 1.9.4.

1.9 Numerical tests.

The scheme in this chapter is compared with other schemes. We also propose a test showing the applicability of the method to other systems.

In figure 1.9.1 the left figures show comparisons with the exact solution for three different large CFL conditions when this solution is a piecewise continuous curve. In the right figures the exact solution is a delta wave at $i = 250$. One first observes that it is perfectly located. From top to bottom the CFL conditions are $r\|u\|_{L^\infty} = 1.035, 2.07, 3.57$ (relevant of values $p = 2, 3, 4$). The support of the delta wave encompasses $10, 3, 10$ cells respectively. In this test the sticky particle method of [7] gives a support located on one mesh only (figure 6 in [7]). The isolated points that are observed in the left figures are parasite values due to some unphysical numerical inconsistencies at sonic points, i.e. points in which the wave speed equals 0 and usually changes sign. These isolated points have been observed in figure 1 of [4], figure 1 in [3], figure 1.9.3 in this chapter from the scheme in [2] and are mentioned in section 5 of [27] when the velocity changes sign in regions where the density is smoothly varying. The method in [4] makes them disappear in an order 2 scheme (figure 2 in [4]). They are absent in the sticky particle method of [7]. For $p < 1$ there is only one of them (indeed the numerical result from the scheme in this chapter is identical to the one in figure 1.9.3 top-left). Best results are obtained for $r\|u\|_{L^\infty}$ between 2 and 2.5 as it is observed in the middle figures. A beginning of degenerescence is observed on the bottom figures for $r\|u\|_{L^\infty} = 3.57$. The computations are very fast : from 0.05 second to 0.08 second on a standard PC. The initial conditions are [7] : if $-\pi \leq x \leq \pi,\ \rho^0(x) = 2 - sinx, u(x) = 1 - x$. Elsewhere the initial values are $\rho^0 = 0 = u^0$. The values $x = 0$ and $x = \pi$ correspond respectively to $i = 200$ and $i = 357$. On the left figures the product Nr is chosen equal to 25 where N is the number of time steps. On the right figures Nr=50.

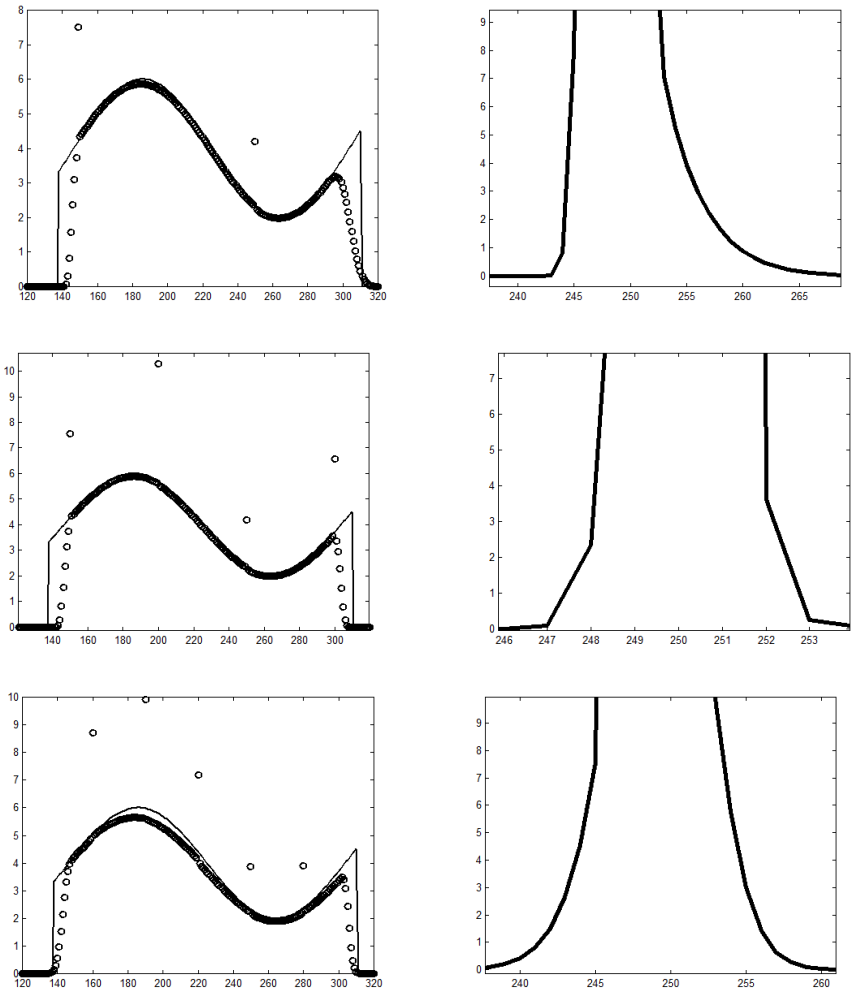
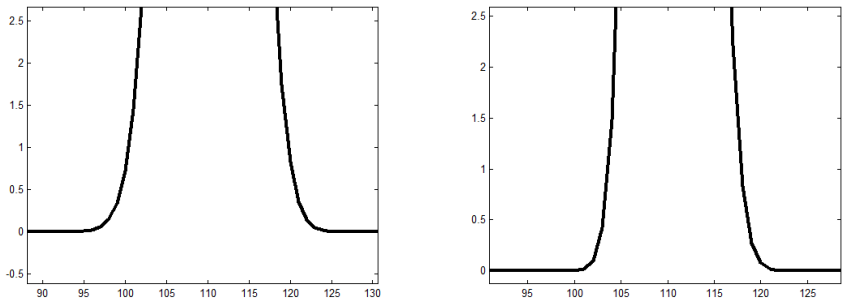


Figure 1.9.1. Comparaison of p-versions of the scheme.



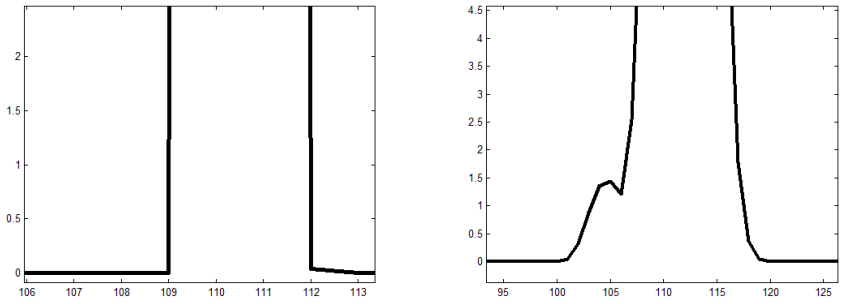


Figure 1.9.2. Comparison of p-versions of the scheme. Delta wave from collision of two finite dust clouds.

In figue 1.9.2 one observes the collapse at $i = 110$, corresponding to $rN = 351$, in form of a single delta wave of the kind "double-rarefaction adjacent to vacuum states" [27] section 6. One first observes it is perfectly located. From top left to bottom right $r\|u\|_{L^\infty} = 1, 2, 3, 4$. The support of the delta wave encompasses $20, 15, 2, 15$ cells respectively, showing that the CFL condition $r\|u\|_{L^\infty} = 3$ gives the best result. The initial conditions are [27] : for $i = 50$ to $100, \rho(i) = 2, u(i) = 1$ and for $i = 200$ to $400, \rho(i) = 1, u(i) = \quad 1$. Elsewhere $\rho = 0$ and $u = 0$.

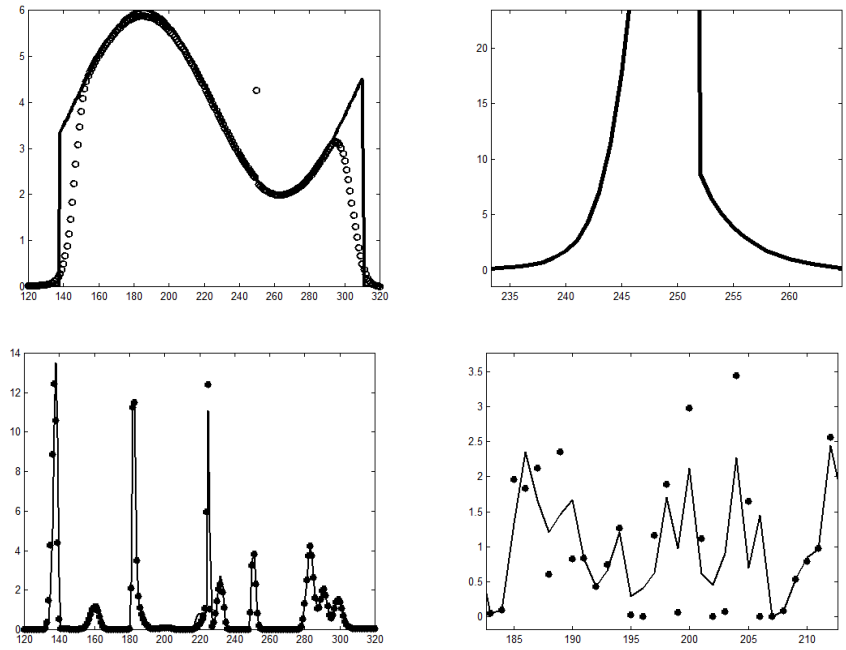


Figure 1.9.3. Comparaison with the original Godunov scheme.

In figure 1.9.3 the top figure, which reproduces the test in figure 1.9.1 with the scheme in [2], gives the correct result, but only under the CFL condition $r\|u\|_{L^\infty} \leq 1$. With the same value of r the scheme in this chapter gives identical results. The same observation holds for the test in

figure 1.9.2. On the bottom figures the initial conditions are chosen at random, as in figure 1.9.5 below and $r\|u\|_{L^\infty} \leq 1$. On the bottom left figure the scheme in [2] (bullets) and the scheme in this chapter (continuous line) are compared after 100 iterations : one observes that the results are practically identical . The differences observed after a few iterations (bottom right : 10 iterations) disappear soon when matter coalesces in peaks. As observed numerically the Godunov scheme [2] forbids interpenetration (case of fluids) while the scheme in this chapter corresponds rather to rarefied gases and cosmology where some significative interpenetration takes place. The scheme [2] differs from the scheme in this chapter at interfaces $i+\frac{1}{2}$ when $u_i^n > 0$ and $u_{i+1}^n < 0$ which are rare events in classical tests such as those in figure 1.9.1 and in figure 1.9.2. In figure 1.9.3 these events are not rare but a difference that appears in the first iterations disappears after a significative number of iterations.

The scheme with sharing of delta waves according to (1.46) is used in figure 1.9.4 with 300 time steps and $r = 0.5$. The left figure shows the numerical solution of the Riemann problem when $u_l = 2, v_l = 1, u_r = 1$ and $v_r = 12$. One observes two discontinuities corresponding to sets of values $(c, u_l, v_l, u_r, v_r) = (2, 2, 1, 2, 25)$ for the left discontinuity and $(3, 2, 25, 1, 12)$ for the right discontinuity. For both discontinuities one checks at once that the jump condition (1.42) with $\alpha = 0$ is satisfied. The right figure shows the numerical solution of the Riemann problem $u_l = 2, v_l = 1, u_r = \quad 1$ and $v_r = 1$. The numerical solution is a delta wave located at $i = 750$ and $i = 751$, of heights 300 and 600 respectively. One observes that $ct = 150$, i.e. $t = 150$ since $c = 1$; $\alpha t = 600 + 300 = 900$ then gives $\alpha = 6$; one checks that (1.42) is satisfied. In these two tests the scheme from the sharing formulas (1.46) gives exactly a solution of the equations.

Conclusion. Degenerescence of the results for a large enough number of time steps has been observed for a CFL $r\|u\|_{L^\infty} = 2.6$ (in the test of figure 1.9.1), or more, depending on the test. The programs in this chapter are typed with $p = 3$ or 4 and one avoids to have $r\|u\|_{L^\infty} > 2.5$ without careful tests that this is possible. In the performed numerical tests the original Godunov scheme [2] and the scheme in this chapter always give the same final result when $h \to 0$ despite the fact that no convergence has been proved for the original Godunov scheme. It does not work as soon as $r\|u\|_{L^\infty} > 1$ while the p-schemes in this chapter allow far larger values of r and adapt without splitting to any space dimension.

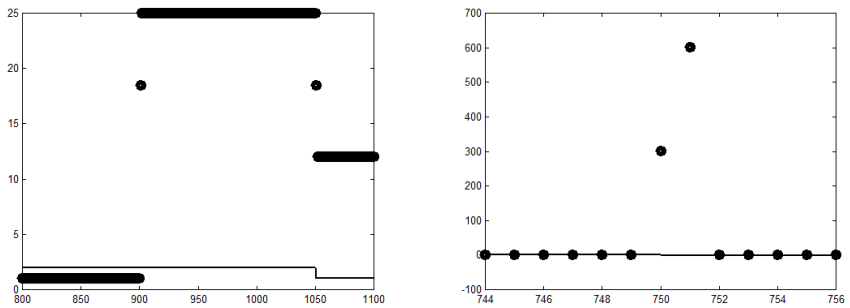


Figure 1.9.4. A numerical scheme for the system in Remark 1 in the case f(u)=u with the projections from the sharing formulas in Remark 1.

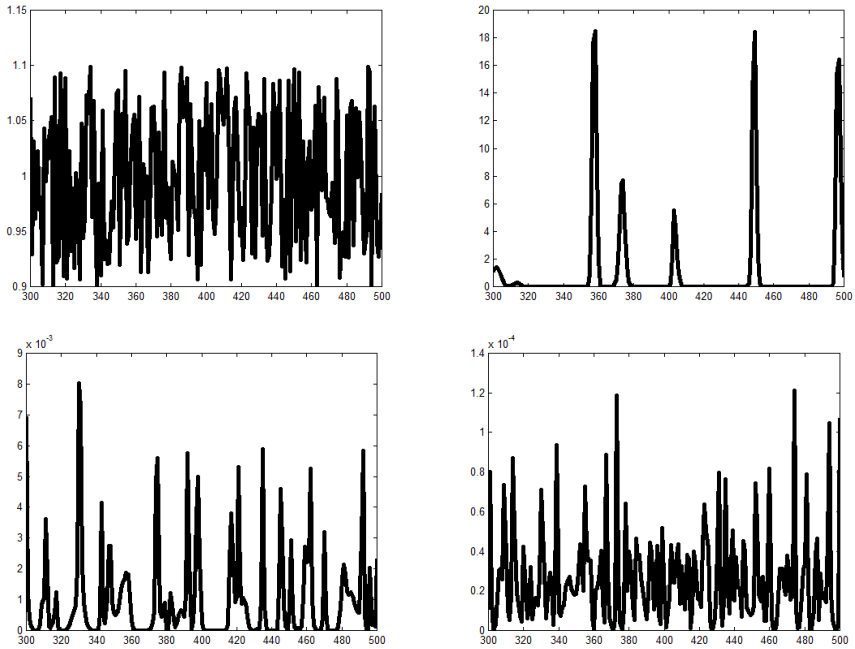


Figure 1.9. 5. Structure formation at same time according to the expansion rate

1.10 Numerical simulations.

In this section two numerical simulations are presented. The consideration of expanding background is justified from the fact one can reproduce steps of the theory of large structure formation in cosmology for matter dominated universes [8], [30] and [31], with a numerical scheme whose consistency (and from chapter 5 convergence to a solution of the equations) has been proved (Theorem 1.8.3). The fully nonlinear but numerical results presented below are replaced in books of cosmology by perturbation theory, i.e. linearization of the equations around a suitable solution, transformation of the linear PDEs into linear ODEs by Fourier transform, and explicit solutions of these ODEs. The need to develop fully nonlinear techniques by solving numerically the nonlinear equations considered in this chapter is pointed out in [8] pp. 287-322.

The expansion of the background is described by the scale factor a(t) which is a smooth strictly positive given function of the time t : a physical distance unity at time 0 becomes a(t) at time t. The equations are given in comoving coordinates, i.e. spatial coordinates whose unit of length follows the expansion of the background : the spatial physical coordinates at time t are obtained by multiplying the comoving coordinates by the scale factor $a(t)$. In one space dimension the equations are (see [8] p. 294 and [30] p. 233) :

$$\rho_t + 3\frac{\dot{a}(t)}{a(t)}\rho + \frac{1}{a(t)}(\rho u)_x = 0, \tag{1.48}$$

$$(\rho u)_t + 4\frac{\dot{a}(t)}{a(t)}\rho u + \frac{1}{a(t)}(\rho u^2)_x = 0. \tag{1.49}$$

This system is equivalent to the one in static background :

Proposition 1.10.1. *Let $\rho(x,t), u(x,t)$ be solution of the static background system (1,2). Let $\phi(t) = \int_0^t \frac{d\tau}{a(\tau)^2}$. Then $\bar{\rho}(x,t) = a(t)\ {}^3\rho(x,\phi(t))$, $\bar{u}(x,t) = a(t)\ {}^1u(x,\phi(t))$ is solution of the system in expanding background (1.47)-(1.48).*

proof. It is a direct verification.□

Therefore the scheme in expanding background is an easy extension of the scheme in static background, with same stability and convergence properties.

The simulation in figure 1.10.1 shows how structure formation depends on the expansion. The top-left figure represents initial matter whose density and velocity are randomly distributed on each cell : $0.9 \leq \rho \leq 1.1, \quad 0.5 \leq u \leq 0.5$. The background expansion is of the form $a(t) = 1 + ct$ with various coefficients c. There are 100 time steps in the 3 other figures. The top-right figure in static background shows an efficient structure formation : randomly distributed matter is agglomerated into peaks. In the bottom-left figure the scale factor has been multiplied by 6, then structure formation is far less efficient. In the bottom-right figure the scale factor has been multiplied by 41, one observes a near absence of structure formation. One notices that the maximum values of density peaks change very much mainly because the physical sizes of the cells have been multiplied by a^3. The conclusion is that structure formation is very sensitive to the expansion rate, and made impossible by too fast expansion.

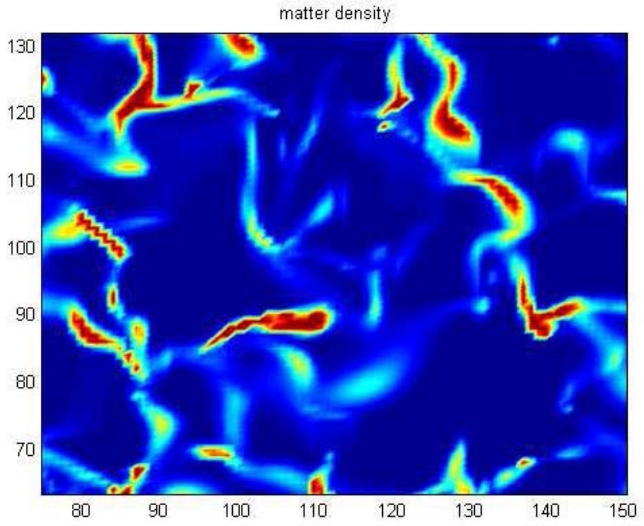


Figure 1.10. 1. Typical patchwork of voids, clusters and filaments of matter in two dimension.

In figure 1.10.2 he initial conditions are at random around the value 1 for density and the value 0 for velocities as in figure 1.9.5. The background has expanded by a factor 1.1 in 50 iterations with the large CFL $r\|u\|_{L^\infty} = 2$. Calculations have been done on a standard PC in a 200×200 window in 3 minutes. One observes the typical structure of filament-cluster-void network observed in [8] p. 308 and p. 333, [30] cover and [31] cover, p. 490, p. 458. The time

evolution one observes is exactly similar to the one depicted in [8] p.308. The 2D tests can be done on any standard PC.

1.11 End of the proof of consistency in 2-D and 3-D.

For brevity we set $\omega := \rho, \rho u, \rho v, \rho w$. Following the 1-D proof in 2-D and 3-D the respective extension of (1.37) is :

$$I_2 := -h^2 \sum_{i,j,n} [\omega_{i,j}^{n+1} - \omega_{i,j}^n + r((\omega u)_{i,j}^n - (\omega u)_{i-1,j}^n) + r((\omega v)_{i,j}^n - (\omega v)_{i,j-1}^n)]\psi_{i,j}^n, \tag{1.50}$$

and

$$I_3 := -h^3 \sum_{i,j,k,n} [\omega_{i,j,k}^{n+1} - \omega_{i,j,k}^n + r((\omega u)_{i,j,k}^n - (\omega u)_{i-1,j,k}^n) + r((\omega v)_{i,j,k}^n - (\omega v)_{i,j-1,k}^n) \tag{1.51}$$

$+r((\omega w)_{i,j,k}^n - (\omega w)_{i,j,k-1}^n)]\psi_{i,j,k}^n$.

We will prove that I_2 and I_3 equal $O(h)$ which will prove that the scheme is a weak asymptotic method of order one in 2-D and 3-D.

First step : two dimension and positive velocities. In this first step let us assume that $\forall i, j, n \quad (u)_{i,j}^n \geq 0$ and $(v)_{i,j}^n \geq 0$. The induction formulas (1.28)-(1.29) are an evaluation of the transports that take place between times t_n and t_{n+1} : the cell $C_{i,j}$ looses part of its contents (wich has been transported at velocity $(u_{i,j}^n, v_{i,j}^n)$) and has received matter from the cells $C_{i-1,j}, C_{i,j-1}, C_{i-1,j-1}$ only since we assume positiveness of velocities. The following contributions are obvious from pictures of the overlapping transported cells with the fixed cell $C_{i,j}$ in the four cases given in the Appendix.

From figure 1.11.1 we obtain

$$h^2 \omega_{i,j}^{n+1} = T_{i,j} + T_{i-1,j} + T_{i,j-1} + T_{i-1,j-1}$$

where
• $T_{i,j} := \omega_{i,j}^n (h - rhu_{i,j}^n)(h - rhv_{i,j}^n)$ denotes the matter that remains at time t_{n+1} in the fixed cell $C_{i,j}$, from an evaluation of the area of the intersection of the transported cell $C_{i,j}$ with the fixed cell $C_{i,j}$;
• $T_{i-1,j} := \omega_{i-1,j}^n rhu_{i-1,j}^n (h - rhv_{i-1,j}^n)$ denotes the matter that comes from the cell $C_{i-1,j}$ from an evaluation of the area of the intersection of the transported cell $C_{i-1,j}$ with the fixed cell $C_{i,j}$;
• $T_{i,j-1} := \omega_{i,j-1}^n (h - rhu_{i,j-1}^n) rhv_{i,j-1}^n$ denotes the matter that comes from the cell $C_{i,j-1}$ from an evaluation of the area of the intersection of the transported cell $C_{i,j-1}$ with the fixed cell $C_{i,j}$;
• $T_{i-1,j-1} := \omega_{i-1,j-1}^n rhu_{i-1,j-1}^n rhv_{i-1,j-1}^n$ denotes the matter that comes from the cell $C_{i-1,j-1}$ from an evaluation of the area of the intersection of the transported cell $C_{i-1,j-1}$ with the fixed cell $C_{i,j}$.

Developping and dividing by h^2 one obtains the formula

$$\omega_{i,j}^{n+1} - \omega_{i,j}^n + r[(\omega u)_{i,j}^n + (\omega v)_{i,j}^n - (\omega u)_{i-1,j}^n - (\omega v)_{i,j-1}^n] =$$

$$r^2((\omega uv)_{i,j}^n - (\omega uv)_{i-1,j}^n - (\omega uv)_{i,j-1}^n + (\omega uv)_{i-1,j-1}^n). \tag{1.52}$$

Therefore, from (1.49)

$$I_2 = -r^2h^2 \sum_{i,j,n} [(\omega uv)^n_{i,j} - (\omega uv)^n_{i-1,j} - (\omega uv)^n_{i,j-1} + (\omega uv)^n_{i-1,j-1}]\psi^n_{i,j}. \tag{1.53}$$

A change in indices gives

$$I_2 = -r^2h^2 \sum_{i,j,n} (\omega uv)^n_{i,j}(\psi^n_{i,j} - \psi^n_{i+1,j} - \psi^n_{i,j+1} + \psi^n_{i+1,j+1}). \tag{1.54}$$

From the L^1 stability of ωuv and Taylor's formula in ψ, which gives a $O(h^2)$ bound depending only on ψ, $I_2 = O(h)$.□

Second step : three dimension and positive velocities. From the transport formula (1.27) with $p = 1$ and since all velocities are ≥ 0 one has to take into account the cell $C_{i,j,k}$ itself, the three cells $C_{i-1,j,k}, C_{i,j-1,k}, C_{i,j,k-1}$, the three cells $C_{i-1,j-1,k}, C_{i,j-1,k-1}, C_{i-1,j,k-1}$ and finally the cell $C_{i-1,j-1,k-1}$. This gives 27 terms in the second member below. Using the similarity with the 2-D case, one obtains

$$h^3\omega^{n+1}_{i,j,k} = U_0 + U_1 + U_2 + U_3$$

where

$$U_0 = \omega^n_{i,j,k}(h - rhu^n_{i,j,k})(h - rhv^n_{i,j,k})(h - rhw^n_{i,j,k});$$

$$U_1 = \omega^n_{i-1,j,k} rhu^n_{i-1,j,k}(h - rhv^n_{i-1,j,k})(h - rhw^n_{i-1,j,k})$$
$$+\omega^n_{i,j-1,k}(h - rhu^n_{i,j-1,k}) rhv^n_{i,j-1,k}(h - rhw^n_{i,j-1,k})$$
$$+\omega^n_{i,j,k-1}(h - rhu^n_{i,j,k-1})(h - rhv^n_{i,j,k-1}) rhw^n_{i,j,k-1};$$

$$U_2 = \omega^n_{i-1,j-1,k} rhu^n_{i-1,j-1,k} rhv^n_{i-1,j-1,k}(h - rhw^n_{i-1,j-1,k})$$
$$+\omega^n_{i,j-1,k-1}(h - rhu^n_{i,j-1,k-1}) rhv^n_{i,j-1,k-1} rhw^n_{i,j-1,k-1}$$
$$+\omega^n_{i-1,j,k-1} rhu^n_{i-1,j,k-1}(h - rhv^n_{i-1,j,k-1}) rhw^n_{i-1,j,k-1};$$

$$U_3 = \omega^n_{i-1,j-1,k-1} rhu^n_{i-1,j-1,k-1} rhv^n_{i-1,j-1,k-1} rhw^n_{i-1,j-1,k-1}.$$

Developping, dividing by h^3 and setting $A = \omega uv, B = \omega uw, C = \omega vw, D = \omega uvw, E = -\omega uvw$ (it is convenient to consider D and E separately), one obtains

$$\omega^{n+1}_{i,j,k} = \omega^n_{i,j,k} +$$

$$r(-(\omega u)^n_{i,j,k} - (\omega v)^n_{i,j,k} - (\omega w)^n_{i,j,k} + (\omega u)^n_{i-1,j,k} + (\omega v)^n_{i,j-1,k} + (\omega w)^n_{i,j,k-1}) \tag{1.55}$$

$$+r^2(A^n_{i,j,k} - A^n_{i-1,j,k} - A^n_{i,j-1,k} + A^n_{i-1,j-1,k} + B^n_{i,j,k} - B^n_{i-1,j,k} - B^n_{i,j-1,k} + B^n_{i-1,j-1,k} \tag{1.56}$$
$$+C^n_{i,j,k} - C^n_{i-1,j,k} - C^n_{i,j-1,k} + C^n_{i-1,j-1,k})$$

$$-r^3(D^n_{i,j,k} - D^n_{i-1,j,k} - D^n_{i,j-1,k} + D^n_{i-1,j-1,k} + E^n_{i,j,k-1} - E^n_{i-1,j,k-1} - E^n_{i,j-1,k-1} + E^n_{i-1,j-1,k-1}). \tag{1.57}$$

The sum (1.54) enters into the sum in the second member of (1.50). Each of the five blocks of four terms in A, B, C, D, E gives a $O(h^2)$ bound after transfer of the lower indices to the smooth function ψ : for instance

$\sum_{i,j,k,n} r^2(A^n_{i,j,k} - A^n_{i-1,j,k} - A^n_{i,j-1,k} + A^n_{i-1,j-1,k})\psi^n_{i,j,k} = \sum_{i,j,k,n} r^2 A^n_{i,j,k}(\psi^n_{i,j,k} - \psi^n_{i+1,j,k} - \psi^n_{i,j+1,k} + \psi^n_{i+1,j+1,k})$.

The conclusion $I_3 = O(h)$ follows from the L^1 stability of A, B, C, D, E and Taylor's formula in ψ.□

Third step : recall of the one dimensional proof with arbitrary signs of velocities. We need to recall the one dimensional proof of Theorem 1.8.3, as a preparation to help for the understanding of the two and three dimensional proofs. Indeed it provides a description of the proof for the first order terms in r in the two and three dimensional cases.

From (1.17)-(1.18), for given index i_0, the quantity $\omega^n_{i_0}$, $\omega = \rho, \rho u$, which lies in the cell $[i_0 h - \frac{h}{2}, i_0 h + \frac{h}{2}]$ at time t_n is in part transported from time t_n to time t_{n+1} at a velocity $u^n_{i_0}$ to one of the two neighbor cells : a quantity $\omega^n_{i_0} u^n_{i_0}(t_{n+1} - t_n) = \omega^n_{i_0} u^n_{i_0} rh$ leaves the cell $[i_0 h - \frac{h}{2}, i_0 h + \frac{h}{2}]$. It contributes to the cell on the left if $u^n_{i_0} < 0$, or to the cell on the right if $u^n_{i_0} > 0$. Therefore, after division by h,

if $u^n_{i_0} > 0$, then (loss of the cell i_0 and gain of the cell $i_0 + 1$) :
$\omega^{n+1}_{i_0} - \omega^n_{i_0} = -r(\omega u)^n_{i_0}$ (and terms not involving $(\omega u)^n_{i_0}$) ,
$\omega^{n+1}_{i_0+1} - \omega^n_{i_0+1} = r(\omega u)^n_{i_0}$ (and terms not involving $(\omega u)^n_{i_0}$).

if $u^n_{i_0} < 0$, then (loss of the cell i_0 and gain of the cell $i_0 - 1$) :
$\omega^{n+1}_{i_0} - \omega^n_{i_0} = r(\omega u)^n_{i_0}$ (and terms not involving $(\omega u)^n_{i_0}$),
$\omega^{n+1}_{i_0-1} - \omega^n_{i_0-1} = -r(\omega u)^n_{i_0}$ (and terms not involving $(\omega u)^n_{i_0}$).

In both cases, from the CFL condition (1.20) with $p = 1$, there are no more terms $(\omega u)^n_{i_0}$ in the sum $\sum_i(\omega^{n+1}_i - \omega^n_i)$. Therefore in the sum $\sum_i(\overline{\omega}_i - \omega^n_i)\psi^n_i$ there are only two occurences of $(\omega u)^n_{i_0}$, namely those in the two above cases : $(\omega u)^n_{i_0}$ appears in this sum as

$$r(\omega u)^n_{i_0}(\psi^n_{i_0+1} - \psi^n_{i_0}) \text{ if } u^n_{i_0} > 0, \quad r(\omega u)^n_{i_0}(\psi^n_{i_0} - \psi^n_{i_0-1}) \text{ if } u^n_{i_0} < 0. \tag{1.58}$$

Applying this result for all i in place of i_0, one obtains

$$\sum_i(\omega^{n+1}_i - \omega^n_i)\psi^n_i + r\sum_i(\omega u)^n_i \psi^n_i - r\sum_i(\omega u)^n_{i-1}\psi^n_i = r\sum_i(\omega u)^n_i Q^n_i$$

where $Q^n_i = \psi^n_{i+1} - \psi^n_i + \psi^n_i - \psi^n_{i+1} = 0$ if $u^n_i \geq 0$
or
$Q^n_i = \psi^n_i - \psi^n_{i-1} + \psi^n_i - \psi^n_{i+1} = O(h^2)$ if $u^n_i \leq 0$.

One obtains

$$I_1 := -h\sum_{i,n}[\omega_i^{n+1} - \omega_i^n + r((\omega u)_i^n - (\omega u)_{i-1}^n)]\psi_i^n = -h\sum_{i,n}(\omega u)_i^n O(h^2) = O(h) \tag{1.59}$$

from the L^1 stability of ωu.□

Fourth step : two dimensional case and arbitrary signs of velocities. In the two dimensional case the one dimensional argument applies without any change in the terms $r(\omega u)_{i,j}^n$, $r(\omega v)_{i,j}^n$ in I_2, (1.49). If, for instance $u_{i_0,j_0}^n > 0$, then the matter ω_{i_0,j_0}^n in the cell C_{i_0,j_0} at time t_n goes to the right. Therefore, at time t_{n+1} it covers in the cell C_{i_0+1,j_0} a region of area $ru_{i_0,j_0}^n h(h - rh|v_{i_0,j_0}^n|) = ru_{i_0,j_0}^n h^2$ (and term in r^2). Therefore the terms in factor of r are exactly the same in each x, y direction as in the one dimensional case, except that the factor h^2 replaces the factor h. The difference with the one dimensional case is the occurence of terms in r^2 that we now consider.

In the (unknown since it depends on the field of velocity at time t_n) formula giving $\omega_{i,j}^{n+1}$ (see (1.51) in the case all velocities are positive) there appear terms $r^2(\omega uv)_{i,j}^n$ that were proved in (1.52)-(1.53) to be unefficient in the case all velocities are positive. Here one has to prove again that these terms are unefficient. For given (i_0, j_0) one can distinguish four cases, depending on the signs in $(u_{i_0,j_0}^n,\ v_{i_0,j_0}^n)$. Let us first consider the case $(u_{i_0,j_0}^n \geq 0,\ v_{i_0,j_0}^n \geq 0)$: one cannot *a priori* use (1.51)-(1.53) because the signs of all the velocities for $(i,j) \neq (i_0, j_0)$ are unknown here.

In (1.51) the quantity $(\omega_{i,j}^{n+1} - \omega_{i,j}^n)$ has been evaluated in the case all velocities are positive. This is no longer the case here since we only know that $u_{i_0,j_0}^n \geq 0, v_{i_0,j_0}^n \geq 0$ (with any possible signs for the other velocities). In the present case, formula (1.51) is no longer valid. The second members of the unknown formulas which here replace the formulas (1.51) depend on the unknown signs of the velocities for all (i,j), but, when considered for all (i,j), they contain the same terms $r^2(\omega uv)_{i_0,j_0}^n$ as the set of all formulas (1.51) since these terms follow from the transport of the cell C_{i_0,j_0} according to the positive velocities $u_{i_0,j_0}^n \geq 0, v_{i_0,j_0}^n \geq 0$. Therefore it suffices to search the terms $(\omega uv)_{i_0,j_0}^n$ in formulas (1.61) written for all (i,j). Therefore from formulas (1.51) the terms $(\omega uv)_{i_0,j_0}^n$ in the series $\sum_{i,j}(\omega_{i,j}^{n+1} - \omega_{i,j}^n)\psi_{i,j}^n$ are

- $+ r^2(\omega uv)_{i_0,j_0}^n \psi_{i_0,j_0}^n$ from (51) with first member $\omega_{i_0,j_0}^{n+1} - \omega_{i_0,j_0}^n + \ldots$;
- $- r^2(\omega uv)_{i_0,j_0}^n \psi_{i_0+1,j_0}^n$ from (51) with first member $\omega_{i_0+1,j_0}^{n+1} - \omega_{i_0+1,j_0}^n + \ldots$;
- $- r^2(\omega uv)_{i_0,j_0}^n \psi_{i_0,j_0+1}^n$ from (51) with first member $\omega_{i_0,j_0+1}^{n+1} - \omega_{i_0,j_0+1}^n + \ldots$;
- $+ r^2(\omega uv)_{i_0,j_0}^n \psi_{i_0+1,j_0+1}^n$ from (51) with first member $\omega_{i_0+1,j_0+1}^{n+1} - \omega_{i_0+1,j_0+1}^n + \ldots$

Their sum gives

$$r^2[(\omega uv)_{i_0,j_0}^n \psi_{i_0,j_0}^n - (\omega uv)_{i_0,j_0}^n \psi_{i_0+1,j_0}^n - (\omega uv)_{i_0,j_0}^n \psi_{i_0,j_0+1}^n + (\omega uv)_{i_0,j_0}^n \psi_{i_0+1,j_0+1}^n]. \tag{1.60}$$

We have obtained : if $u_{i_0,j_0}^n \geq 0$ and $v_{i_0,j_0}^n \geq 0$ the factor $(\omega uv)_{i_0,j_0}^n$ occurs in the sum $\sum_{i,j}(\omega_{i,j}^{n+1} - \omega_{i,j}^n)\psi_{i,j}^n$ as the term

$$r^2(\omega uv)_{i_0,j_0}^n(\psi_{i_0,j_0}^n - \psi_{i_0,j_0+1}^n - \psi_{i_0+1,j_0}^n + \psi_{i_0+1,j_0+1}^n). \tag{1.61}$$

The $\psi' s$ give the $O(h^2)$ bound already noticed in the one dimensional case.

Now it suffices to notice that for each of the three other cases concerning the signs of $(u^n_{i_0,j_0},\ v^n_{i_0,j_0})$ the $O(h^2)$ bound occurs as above by changing the sense of some x, y axis so as to be in the above positiveness case (the scheme is clearly unsensitive to a change in sense of the x, y axis). Finally, in the sum $\sum_{i,j}(\omega^{n+1}_{i,j}-\omega^n_{i,j})\psi^n_{i,j}$, the sum of all terms in r^2 is $\sum_{i,j} r^2(\omega uv)^n_{i,j}O(h^2)$. Therefore the sum of all r^2 terms gives $O(h)$ in (1.49). □

Fifth step : Three dimension and arbitrary signs of velocities. In the three dimensional case (1.50), the occurences in the sum $\sum_{i,j,k}(\omega^{n+1}_{i,j,k} - \omega^n_{i,j,k})\psi^n_{i,j,k}$ of the terms of order 1 in r are similar to those considered in the one dimensional case and the occurences of the terms of order 2 are similar to those considered in the two dimensional case. It remains to evaluate the terms $r^3(\omega uvw)^n_{i_0,j_0,k_0}$. We distinguish eight cases, depending on the signs of $(u^n_{i_0,j_0,k_0},\ v^n_{i_0,j_0,k_0},\ w^n_{i_0,j_0,k_0})$. First, consider the case all signs are positive. As explained in the two dimensional case the terms in $r^3(\omega uvw)^n_{i_0,j_0,k_0}$ contained in the sum $\sum_{i,j,k}(\omega^{n+1}_{i,j,k} - \omega^n_{i,j,k})\psi^n_{i,j,k}$ can be extracted from all formulas (1.56) written for all $(i,\ j,\ k)$ by searching the terms emanating from the cell C_{i_0,j_0,k_0} because the three velocities $u^n_{i_0,j_0,k_0}, v^n_{i_0,j_0,k_0}, w^n_{i_0,j_0,k_0}$ are positive. Considering only the D terms (the proof is the same for the E terms) we recall for convenience the formula (1.56) (only valid in the case all velocities in all cells are positive) under the form

$\sum_{i,j,k}(\omega^{n+1}_{i,j,k}-\omega^n_{i,j,k})\psi^n_{i,j,k} = -r^3\sum_{i,j,k}(D^n_{i,j,k}\psi^n_{i,j,k}-D^n_{i-1,j,k}\psi^n_{i,j,k}-D^n_{i,j-1,k}\psi^n_{i,j,k}+D^n_{i-1,j-1,k}\psi^n_{i,j,k})$
(and terms not involving D).

The terms we seek emanating from the cell C_{i_0,j_0,k_0} are the terms in $D^n_{i_0,j_0,k_0}$:

$-r^3D^n_{i_0,j_0,k_0}(\psi^n_{i_0,j_0,k_0} - \psi^n_{i_0+1,j_0,k_0} - \psi^n_{i_0,j_0+1,k_0} + \psi^n_{i_0+1,j_0+1,k_0})$.

The four $\psi's$ give the requested $O(h^2)$ bound. The seven other cases of signs of velocities are treated by changing the senses of the coordinate axis as in the two dimensional case.□

1.12 Conclusion.

In the case of pressureless fluids the method of splitting of delta waves presented in this chapter has permitted to obtain a numerical scheme which is stable (Theorem 1.7.1) and convergent Theorem 1.8.3. In particular, the p-schemes described in this chapter improve significantly the original Godunov scheme, taking into account they are fast and with a large CFL condition. They extend at once, without dimensional splitting, into 2 and 3 space dimension. This method of projection of delta waves applies to different systems of conservation laws from physics and mathematics, whether they are connected or not with the system of pressureless flows. These results will be transformed, from chapter 5, into a result of convergence to a solution of the equations.

1.13 Appendix.

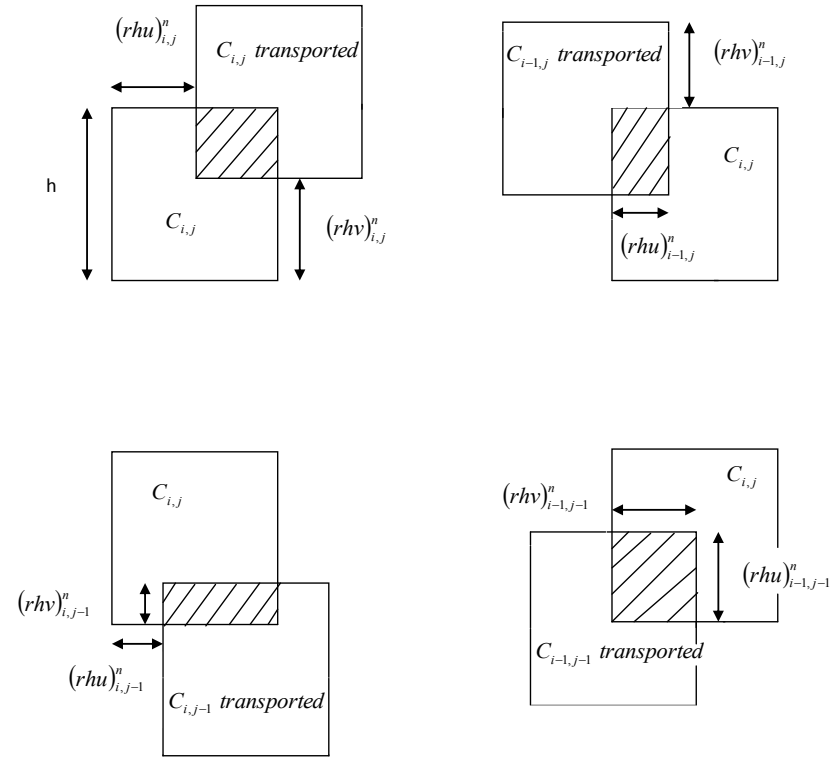


Figure 1.11.1. The four successive evaluations below from overlapping squares are represented by hatched regions.

Chapitre 2

Self-gravitating fluids

In this chapter we present a numerical scheme for the 3-D system of self-gravitating fluid dynamics in the collisional case as well as in the non-collisional case. Consistency of order one in the sense of distributions is proved in 1-D and in absence of pressure. In the other cases consistency is proved under the numerical assumptions of boundedness of the velocity field in the CFL condition and of boundedness of the gradient of the gravitation potential. In 2-D and 3-D, concentrations of matter in strings and points can cause a theoretical difficulty although one observes that the scheme still works. The initial data are L^∞ functions in velocity and L^1 functions in density. Applications are given to situations in cosmology and astrophysics such as the role of dark matter at decoupling time, the formation and repartition of galaxies, the formation of solar systems and Jeans theory which explains the formation of stars.

2.1 Introduction.

We consider the equations governing a self-gravitating fluid [8] p. 207, [30] p. 460, [31] p. 231, [5] p. 49

$$\frac{\partial \rho}{\partial t} + \vec{\nabla}.(\rho \vec{u}) = 0, \tag{2.1}$$

$$\frac{\partial}{\partial t}(\rho \vec{u}) + \vec{\nabla}.(\rho \vec{u} \otimes \vec{u}) + \vec{\nabla p} + \rho \vec{\nabla \Phi} = \vec{0}, \tag{2.2}$$

$$\Delta \Phi = 4\pi G \rho, \tag{2.3}$$

$$p = K\rho, \tag{2.4}$$

where $\rho, \vec{u} = (u, v, w), p, \Phi$ denote respectively the density, the velocity vector, the pressure and the gravitation potential ; G is the gravitation constant and $K \geq 0$ a constant from the state law. These equations are the continuity equation (2.1), the Euler equation (2.2), the Poisson equation (2.3) and an isothermal state law (2.4). These equations are extended to expanding background by a change of variable [8] p. 294, [31] p. 233, for their use in cosmology.

This system is classically treated in cosmology by perturbation theory which consists in linearization of the equations around a known solution, see [8] p. 207, [30] pp. 460-461, [31] pp.

231-232, [5] p. 50. The linearized equations of motion provide an excellent description of gravitational instability when density fluctuations are small. However, the linear regime breaks down as soon as the density fluctuations are not small, which makes a numerical approximate solution of (2.1)-(2.4) indispensible, see [8] pp. 304-332. In the absence of an exact solution to validate the scheme, one needs to prove at least its consistency, i.e. that the step functions from the scheme tend to satisfy the equations when the space step tends to 0. As far as the author is aware this is the first time that a mathematical proof of consistency has been obtained for this system, even in one space dimension and absence of pressure. It is also the first time that the full system (2.1)-(2.4) is studied numerically even in 1-D.

We propose an original 3-D numerical scheme for (2.1)-(2.4) which is consistent in the sense of distributions under natural assumptions whose numerical verification is immediate for a given value of the space step h : a CFL condition $\|u\|_{L^\infty} \frac{\Delta t}{\Delta x} \leq 1$, supplemented by an assumption of boundedness of the gravitation potential. Then, in order to apply the consistency theorem, this property has to be extrapolated when $h \to 0$. If one does not accept this extrapolation, the proof in this chapter shows that whenever these properties hold for a small value of h, then the step functions from the scheme satisfy the equations with a small deviation of order one in the space step (for given test functions ψ, with bounds depending on the size of the support and the sup. of the first and second derivatives). In absence of exact solutions or physical experiments this mathematical result allows us to put faith in the numerical solutions obtained from the scheme, which is interesting since faith in numerical results is a serious problem in cosmology while the equations (2.1)-(2.4) are fully accepted.

In the case of absence of pressure, i.e. $K = 0$ in (2.4), the scheme is simplified. It concerns the system of self-gravitating pressureless fluids. This system has already been considered in [13] and [29] from a theoretical viewpoint. These authors have obtained results of existence of solutions under various assumptions. In [13] the authors consider in particular the case of random initial data needed to explain large structure formation in cosmology (see [8] and [30]). The initial density is either discrete or absolutely continuous with respect to the Lebesgue measure. In [29] the authors use the theory of mass transportation. The initial velocity has to be continuous and square integrable and the initial density has to be a Borel probability on $\mathbb{R}$ with finite two order moment. From the numerical viewpoint cosmologists have developped N-body simulations representing a sample of the universe as a box with periodic boundary conditions containing a large number of point masses interacting through their mutual gravity [8] pp. 304-310, [30] pp. 482-494. There exists a number of numerical codes done by cosmologists. They represent a cosmological fluid as a discrete set of a large number of particles and calculate the gravitational forces between them. They differ mainly in the way gravitation forces on each particle are calculated, [8] pp. 305-310. In absence of exact solutions for their validation, and impossibility of physical experiments, faith in these methods comes only from the fact they mimick the real physical process and reproduce qualitatively the aspect of the universe as it is observed, [8] p. 308. This is the reason which makes a mathematical proof of consistency particularly useful.

As applications we propose four simulations in the pressureless case : gravitational collapse to a point in absence of fast expansion, then impossibility of collapse in presence of fast expansion of the background (Meszaros effect), formation of structures looking like solar systems from gravitational collapse of a rotating disk, agglomeration of baryonic matter on the existing structures of dark matter at decoupling. Then, in presence of pressure we present two simulations of Jeans theory [8] p. 206, [5] p. 44 : Jeans theory asserts that a gas of collisional particles collapses gravitationally besides pressure if its size is large enough ($\geq$ Jeans length), which explains the

formation of stars.

The scheme is obtained from a convection-pressure correction method which was introduced in Le Roux et al. [2]. The authors of [2] used a splitting technique consisting of separation of the convection terms from the pressure terms and showed the good numerical quality of the schemes thus obtained, with a less restrictive CFL condition than the original schemes without splitting.

2.2 Statement of the scheme.

The real line is divided into intervals $I_i =]ih - \frac{1}{2}h, ih + \frac{1}{2}h[, i \in \mathbb{Z}$. We set $r := \frac{\Delta t}{\Delta x}$ and $t_n = nrh$ for r small enough. We will construct step functions $\rho(x,t)$, $u(x,t)$, $p(x,t)$, ... depending on h, which are constant on the rectangles $I_i \times]t_n, t_{n+1}[$, whose step values are denoted $\rho_i^n, u_i^n, p_i^n, \ldots$, respectively. The indices h are skipped to simplify the notation : ρ stands for $\rho_h, \ldots$. From these step functions ρ and u, we define the step functions ρu, $\rho u^2, \ldots$ by $(\rho u)_i^n = \rho_i^n u_i^n$ and $(\rho u^2)_i^n = \rho_i^n (u_i^n)^2, \ldots$. The initial condition (ρ^0, u^0) is discretized on the intervals I_i by taking mean values on these intervals. We always assume that u^0 is a L^∞ function and that ρ^0, e^0 are positive L^1 functions.

Statement of the scheme for self-gravitating fluids in one dimension. In one space dimension the equations (2.1)-(2.4) reduce to

$$\rho_t + (\rho u)_x = 0, \tag{2.5}$$

$$(\rho u)_t + (\rho u^2)_x + p_x + \rho \Phi_x = 0, \tag{2.6}$$

$$\Phi_{xx} = 4\pi G \rho, \tag{2.7}$$

$$p = K\rho. \tag{2.8}$$

We assume the set $\{\rho_i^n, u_i^n, p_i^n\}_{i \in Z}$ is given. The set $\{\rho_i^{n+1}, u_i^{n+1}, p_i^{n+1}\}_{i \in Z}$ is defined as follows.

If $a < b$, one sets

$$L(a,b) := length\ of\ [0,1] \cap [a,b], \tag{2.9}$$

i.e.

$$L(a,b) = max(0, min(1,b) - max(0,a)). \tag{2.10}$$

- *Transport step.* See section 1.5,

$$\overline{\rho}_i := \rho_{i-1}^n L(-1 + ru_{i-1}^n, ru_{i-1}^n) + \rho_i^n L(ru_i^n, 1 + ru_i^n) + \rho_{i+1}^n L(1 + ru_{i+1}^n, 2 + ru_{i+1}^n). \tag{2.11}$$

When the CFL condition (2.37) is satisfied, the first term represents the matter issued from the cell I_{i-1} between times t_n and t_{n+1} that lies in the cell I_i at time t_{n+1}. The second term represents the matter from the cell I_i that remains in I_i at time t_{n+1}. The third term is similar to the first one : it concerns matter issued from the cell I_{i+1} that lies in the cell I_i at time t_{n+1}. Note that $\overline{\rho}_i$ depends on n, which is not explicitely stated to shorten the notation. The same discretization as the one in chapter 1 gives

$$\begin{aligned}\overline{(\rho u)}_i := {} & (\rho u)_{i-1}^n L(-1 + ru_{i-1}^n, ru_{i-1}^n) + \\ & (\rho u)_i^n L(ru_i^n, 1 + ru_i^n) + (\rho u)_{i+1}^n L(1 + ru_{i+1}^n, 2 + ru_{i+1}^n)\end{aligned} \tag{2.12}$$

where $(\rho u)_i^n = \rho_i^n u_i^n$. The state law is set in the form

$$p_i^{n+1} := K\overline{\rho}_i. \tag{2.13}$$

• *Averaging step.* For some value $0 \leq \alpha < 0.5$ chosen in the scheme

$$\rho_i^{n+1} := \alpha\overline{\rho}_{i-1} + (1-2\alpha)\overline{\rho}_i + \alpha\overline{\rho}_{i+1}, \tag{2.14}$$

$$\widetilde{(\rho u)}_i := \alpha\overline{(\rho u)}_{i-1} + (1-2\alpha)\overline{(\rho u)}_i + \alpha\overline{(\rho u)}_{i+1}. \tag{2.15}$$

The averaging step serves to avoid oscillations caused by the centered discretization in pressure in the next step. In absence of pressure one chooses $\alpha = 0$.

• *Pressure correction step.* One can compute Φ from (2.7), considered as a Dirichlet problem with values 0 on the boundary, or as a periodic problem, which gives

$$(\Phi_x)_i^{n+1} := 4\pi G \sum_{j=1}^{i} \rho_j^{n+1} h + \beta \tag{2.16}$$

for some fixed value β. Then a centered discretization of the pressure term gives

$$(\rho u)_i^{n+1} := \widetilde{(\rho u)}_i - \frac{r}{2}(p_{i+1}^{n+1} - p_{i-1}^{n+1}) - rh\rho_i^{n+1}(\Phi_x)_i^{n+1}. \tag{2.17}$$

If $\rho_i^{n+1} \neq 0$, we set

$$u_i^{n+1} := \frac{(\rho u)_i^{n+1}}{\rho_i^{n+1}}, \tag{2.18}$$

if $\rho_i^{n+1} = 0$ then u_i^{n+1} can be given any value from Proposition 2.2.1 below.

Proposition 2.2.1. *$\rho_i^n = 0$ implies $(\rho u)_i^n = 0$ and $p_i^n = 0$.*

proof. The proof is an induction on n. We first give the proof in presence of pressure. For $n = 0$ it holds by construction. Assume the property holds for n. Then, if $\rho_i^{n+1} = 0$, since $\alpha > 0$, formula (2.14) implies

$$\overline{\rho}_{i-1},\ \overline{\rho}_i,\ \overline{\rho}_{i+1} = 0. \tag{2.19}$$

From (2.11)-(2.12) it follows that $\overline{\rho}_i = 0$ implies $\overline{(\rho u)}_i = 0$: indeed from (2.11) $\rho_j^n L(\ldots) = 0$ for $j = i-1, i, i+1$ since each term in (2.11) is ≥ 0. Either $\rho_j^n = 0$ or $L(\ldots) = 0$. From the induction assumption, $\rho_j^n = 0$ implies $(\rho u)_j^n = 0$, therefore one has always $(\rho u)_j^n L(\ldots) = 0, j = i-1, i, i+1$, i.e. each term in (2.12) is null. Therefore, (2.19) implies

$$\overline{(\rho u)}_{i-1} = 0,\ \overline{(\rho u)}_i = 0,\ \overline{(\rho u)}_{i+1} = 0.$$

Then, from (2.15), we obtain $\widetilde{(\rho u)}_i = 0$. From (2.19), formula (2.13) implies $p_{i-1}^{n+1} = 0, p_i^{n+1} = 0, p_{i+1}^{n+1} = 0$. Finally, all terms in (2.17) are null. In the pressureless case one can take $\alpha = 0$. Then, from (2.14) $\overline{\rho}_i = 0$; from the above implication $\overline{(\rho u)}_i = 0$ and from (2.15) with $\alpha = 0$ $\widetilde{(\rho u)}_i = 0$. It suffices to use (2.17) without pressure to conclude.□

It follows from (2.11)-(2.14) that ρ is positive. Since the coefficients L in (2.11) represent transport i.e. a new repartition of matter at time t_{n+1}, Theorem 1.5.1. and section 1.5, one has $\sum_i \rho_i^n h = \sum_i \rho_i^0 h$. From the positiveness of ρ one has the L^1 stability in ρ. The L^1 stability in ρu

follows from the L^1 stability in ρ and the boundedness of u from assumption (2.37). From (2.16) and L^1-stability in ρ, Φ_x is L^∞ bounded : $|(\Phi_x)_i^n| \leq 4\pi G(\|\rho^0\|_{L^1} + |\beta|)$. In one space dimension, assumption (2.38) is always satisfied since $|\Phi_x| \leq const$: the gradient of the gravitation potential is bounded, even on a point concentration of matter.

Statement of the scheme for self-gravitating fluids in two and three dimensions. The equations in the two dimensional case are

$$\rho_t + (\rho u)_x + (\rho v)_y = 0, \tag{2.20}$$

$$(\rho u)_t + (\rho u^2)_x + (\rho uv)_y + p_x + \rho\Phi_x = 0, \tag{2.21}$$

$$(\rho v)_t + (\rho uv)_x + (\rho v^2)_y + p_y + \rho\Phi_y = 0, \tag{2.22}$$

$$p = K\rho, \tag{2.23}$$

$$\Delta\Phi = 4\pi G\rho. \tag{2.24}$$

The two dimensional space (x, y) is divided into square cells $C_{i,j}$ of side h and centers $(ih, jh)_{i,j\in\mathbb{Z}}$: $C_{i,j}$ is the set of all (x, y) such that $ih - \frac{h}{2} < x < ih + \frac{h}{2}$ and $jh - \frac{h}{2} < y < jh + \frac{h}{2}$. We assume the set $\{\rho_{i,j}^n, u_{i,j}^n, v_{i,j}^n, p_{i,j}^n\}_{i,j\in Z}$ is given. The set $\{\rho_{i,j}^{n+1}, u_{i,j}^{n+1}, v_{i,j}^{n+1}, p_{i,j}^{n+1}\}_{i,j\in Z}$ is defined as follows. We set

$$A(a, b) := L(a, 1 + a).L(b, 1 + b) \tag{2.25}$$

which is the area of the intersection of the square of vertices $(0, 0), (0, 1), (1, 0), (1, 1)$ with the square of vertices $(a, b), (1 + a, b), (a, 1 + b), (1 + a, 1 + b)$. Then we set

- *Transport step.* As in the 1D case let

$$\overline{\rho}_{i,j} := \sum_{-1\leq\lambda,\mu\leq 1} \rho_{i+\lambda,j+\mu}^n A(\lambda + ru_{i+\lambda,j+\mu}^n,\ \mu + rv_{i+\lambda,j+\mu}^n), \tag{2.26}$$

$$(\overline{\rho u})_{i,j} := \sum_{-1\leq\lambda,\mu\leq 1} (\rho u)_{i+\lambda,j+\mu}^n A(\lambda + ru_{i+\lambda,j+\mu}^n,\ \mu + rv_{i+\lambda,j+\mu}^n), \tag{2.27}$$

$$(\overline{\rho v})_{i,j} := \sum_{-1\leq\lambda,\mu\leq 1} (\rho v)_{i+\lambda,j+\mu}^n A(\lambda + ru_{i+\lambda,j+\mu}^n,\ \mu + rv_{i+\lambda,j+\mu}^n), \tag{2.28}$$

$$p_{i,j}^{n+1} := K\overline{\rho}_{i,j}. \tag{2.29}$$

Interpretation of (2.26)-(2.29) is a transport in 2-D, see section 1.6, similarly to (2.11)-(2.12) in 1-D.

- *Averaging step.* Let α, $0 \leq \alpha < \frac{1}{20}$, be given in the scheme. Set

$$\begin{aligned}\rho_{i,j}^{n+1} := \alpha(2\overline{\rho}_{i-1,j-1} + 2\overline{\rho}_{i-1,j+1} + 2\overline{\rho}_{i+1,j-1} + 2\overline{\rho}_{i+1,j+1} + 3\overline{\rho}_{i-1,j} + \\ 3\overline{\rho}_{i,j-1} + 3\overline{\rho}_{i,j+1} + 3\overline{\rho}_{i+1,j}) + (1 - 20\alpha)\overline{\rho}_{i,j},\end{aligned} \tag{2.30}$$

$$\begin{aligned}\widetilde{(\rho u)}_{i,j} := \alpha(2\overline{(\rho u)}_{i-1,j-1} + 2\overline{(\rho u)}_{i-1,j+1} + 2\overline{(\rho u)}_{i+1,j-1} + 2\overline{(\rho u)}_{i+1,j+1} + 3\overline{(\rho u)}_{i-1,j} + \\ 3\overline{(\rho u)}_{i,j-1} + 3\overline{(\rho u)}_{i,j+1} + 3\overline{(\rho u)}_{i+1,j}) + (1 - 20\alpha)\overline{(\rho u)}_{i,j}.\end{aligned} \tag{2.31}$$

We set the same formula for $\widetilde{(\rho v)}_{i,j}$, replacing u by v.

• *Pressure correction step.* The values $\Phi_{i,j}^{n+1}$ of the potential are obtained from a numerical solution of the Poisson equation (for instance the Dirichlet problem with null values on the boundary, or the periodic problem) on the mesh of cells $C_{i,j}$ with second member the function $4\pi G\rho$, where ρ here is the step function equal to $\rho_{i,j}^{n+1}$ on $C_{i,j}$. Then a centered discretization gives

$$(\Phi_x)_{i,j}^{n+1} := \frac{1}{2h}(\Phi_{i+1,j}^{n+1} - \Phi_{i-1,j}^{n+1}), \quad (\Phi_y)_{i,j}^{n+1} := \frac{1}{2h}(\Phi_{i,j+1}^{n+1} - \Phi_{i,j-1}^{n+1}), \tag{2.32}$$

$$(\rho u)_{i,j}^{n+1} := \widetilde{\rho u}_{i,j} - \frac{r}{2}(p_{i+1,j}^{n+1} - p_{i-1,j}^{n+1}) - rh\rho_{i,j}^{n+1}(\Phi_x)_{i,j}^{n+1}. \tag{2.33}$$

A similar formula is given for $(\rho v)_{i,j}^{n+1}$, using y-derivatives. If $\rho_{i,j}^{n+1} \neq 0$, then

$$u_{i,j}^{n+1} := \frac{(\rho u)_{i,j}^{n+1}}{\rho_{i,j}^{n+1}},\ v_{i,j}^{n+1} := \frac{(\rho v)_{i,j}^{n+1}}{\rho_{i,j}^{n+1}}, \tag{2.34}$$

if $\rho_{i,j}^{n+1} = 0$, then $u_{i,j}^{n+1}$ can be given any value as in 1-D as it is proved in Proposition 2.2.1 in 1-D.

The scheme in three space dimension is very similar to the scheme in two space dimension (2.25)-(2.34). Let $C_{i,j,k}$ be the cube of all (x,y,z) such that $(i-\frac{1}{2})h < x < (i+\frac{1}{2})h, (j-\frac{1}{2})h < y < (j+\frac{1}{2})h, (k-\frac{1}{2})h < z < (k+\frac{1}{2})h$. Let

$$V(a,b,c) = L(a,1+a).L(b,1+b).L(c,1+c) \tag{2.35}$$

be the volume of the intersection of the cube of vertices $(i,j,k),\ i,j,k = 0$ or 1, with the cube of vertices $(a+i, b+j, c+k),\ i,j,k = 0$ or 1. If $\omega = \rho, \rho u, \rho v, \rho w$ successively, one sets

$$\overline{\omega}_{i,j,k} =$$

$$\sum_{-1\leq\lambda,\mu,\nu\leq 1} \omega_{i+\lambda,j+\mu,k+\nu}^{n} V(\lambda + ru_{i+\lambda,j+\mu,k+\nu}^{n},\ \mu + rv_{i+\lambda,j+\mu,k+\nu}^{n},\ \nu + rw_{i+\lambda,j+\mu,k+\nu}^{n}). \tag{2.36}$$

We extend (2.30)-(2.31) by taking an average over the cell $C_{i,j,k}$ and its 26 neighbors in order that Taylor's formula in ψ annihilates the first order terms.

2.3 Statement of the consistency theorem.

The constant values on $C_{i,j,k}$ of the approximate solutions $\omega_h(x,y,z,t)$ (usually denoted by ω to simplify the notation) are denoted $\omega_{i,j,k}^n$ for $t_n < t < t_{n+1}$, where $\omega = \rho, u, v, w, p, ...$ We assume that the initial density ρ^0 is a positive L^1 function and the initial velocities u^0, v^0, w^0 are L^∞ functions. We note $\vec{\nabla\Phi} = (\Phi_x, \Phi_y, \Phi_z)$ and $|\vec{\nabla\Phi}| = \sqrt{(\Phi_x)^2 + (\Phi_y)^2 + (\Phi_z)^2}$. For simplification, boundary problems are eliminated by assuming that the physical variables we are interested in tend to 0 at infinity.

Theorem 2.3.1. consistency of the scheme. *Assume that during some time interval* $[0,T]$ *(i.e.* $\forall (i,j,k) \in \mathbb{Z}^3$ *and* n / $t_n \leq T$*) one has*

$$|u_{i,j,k}^n|\frac{\Delta t}{\Delta x} \leq 1,\ |v_{i,j,k}^n|\frac{\Delta t}{\Delta x} \leq 1,\ |w_{i,j,k}^n|\frac{\Delta t}{\Delta x} \leq 1, \tag{2.37}$$

and the following condition (that always holds in 1-D)

$$\exists M > 0 \ / \ \forall i,j,k,n \ \ |(\vec{\nabla}\Phi)^n_{i,j,k}| \leq M \tag{2.38}$$

for all $h > 0$. Then the scheme is consistent in the sense of distributions when $h \to 0$.

More precisely we obtain : $\forall \psi \in C_c^\infty(\mathbb{R}^3 \times]0,T[)$

$$\int (\rho_h \psi_t + \rho_h u_h \psi_x + \rho_h v_h \psi_y + \rho_h w_h \psi_z) dxdydzdt = O(h), \tag{2.39}$$

$$\int \{\rho_h u_h \psi_t + \rho_h (u_h)^2 \psi_x + \rho_h u_h v_h \psi_y + \rho_h u_h w_h \psi_z + p_h \psi_x - \rho_h . (\Phi_x)_h \psi\} dxdydzdt = O(h), \tag{2.40}$$

and similar limits for the two other components of the Euler equation in $(\rho_h v_h), (\rho_h w_h)$,

$$\int (p_h - K\rho_h)\psi dxdydzdt = O(h), \tag{2.41}$$

$$\int \{\Phi_h \Delta\psi - 4\pi G \rho_h \psi\} dxdydzdt = O(h), \tag{2.42}$$

when $h \to 0$.

• *Presence of pressure.* System (2.1)-(2.4) models Jeans' gravitational collapse : when a medium has pressure, a perturbation bigger than a critical length can collapse under its own gravity, see figures 2.5.5 and 2.5.6 below. The presence of pressure does not allow the perturbation to collapse to a mathematical point, as shown in figure 2.5.6 below, and $|\vec{\nabla}\Phi|$ remains bounded. Assumptions (2.37)-(2.38) cover gravitational collapse in the presence of pressure.

• *Absence of pressure.* In the absence of pressure in 2-D, assumption (2.38) can no longer hold in the case of gravitational collapse to one single cell : figure 2.5.1 ; it is well known that the gradient of the gravitation potential can be unbounded in 2-D in the presence of point accumulation of matter and in 3-D in the presence of accumulation of matter on submanifolds of dimension 0 or 1 (points or strings). Nevertheless, it has been observed that the scheme works : figure 2.5.1. In the absence of pressure (i.e. $K = 0$) and if $\vec{\nabla}\Phi$ is bounded, then, for all values $T > 0$, the proof of Theorem 2.3.2 below proves that assumption (2.37) is satisfied as soon as $\frac{\Delta t}{\Delta x}$ is small enough.

Theorem 2.3.2. *In the absence of pressure and in one space dimension one can choose $\frac{\Delta t}{\Delta x}$ small enough such that the scheme applies and is consistent.*

Proof. The proof is given in 1-D to shorten the formulation. In the absence of pressure the scheme is simplified by dropping the averaging step (choice $\alpha = 0$) due to the absence of centered discretization in pressure. First, notice that if $a_n \leq u_i^n \leq b_n \ \forall i$, then if $\rho_i^{n+1} \neq 0$ one has

$$a_n \leq \frac{(\rho u)^n_{i-1} L(-1+ru^n_{i-1}, ru^n_{i-1}) + (\rho u)^n_i L(ru^n_i, 1+ru^n_i) + (\rho u)^n_{i+1} L(1+ru^n_{i+1}, 2+ru^n_{i+1})}{\rho_i^{n+1}} \leq b_n, \tag{2.43}$$

since, from (2.11) and (2.14) with $\alpha = 0$, numerator and denominator are same convex combinations. If $\rho_i^{n+1} = 0$ it follows from the proof of Proposition 2.3.1 in the pressureless case

that the quotient is undeterminate and its value is useless for the next step. Set

$$K = 4\pi G(\|\rho^0\|_{L^1} + |\beta|). \tag{2.44}$$

Now, let us prove that $\forall n$ such that $t_n \leq T$ one has

$$\min(u_0) - TK \leq u_i^n \leq \max(u_0) + TK \quad \forall i. \tag{2.45}$$

To this end, let $a_n \leq u_i^n \leq b_n \ \forall i$. Formulas (2.18), (2.17) without pressure, (2.15) with $\alpha = 0$, and (2.12) imply that

$$u_i^{n+1} = \frac{(\rho u)_{i-1}^n L(-1+ru_{i-1}^n, ru_{i-1}^n) + (\rho u)_i^n L(ru_i^n, 1+ru_i^n) + (\rho u)_{i+1}^n L(1+ru_{i+1}^n, 2+ru_{i+1}^n)}{\rho_i^{n+1}} - rh(\Phi_x)_i^{n+1}.$$

Therefore, from (2.43), (2.16) and (2.44)

$$a_n - rhK \leq u_i^{n+1} \leq b_n + rhK. \tag{2.46}$$

We obtain (2.45) by induction on n since $t_n = nrh \leq T$.

Now fix a value r such that

$$r(\|u_0\|_{L^\infty} + TK) < 1. \tag{2.47}$$

Then, as long as $t_n = nrh < T$ the scheme satisfies $r\|u\|_{L^\infty} < 1$. When the CFL condition (2.47) is satisfied time T can be attained. One has the stability results : ρ is positive and L^1 stable on $\mathbb{R} \times [0, T]$ since it is ruled by a transport, u and Φ_x are L^∞ stable from (2.45) and (2.16) respectively, ρu and ρu^2 are L^1 stable, since ρ is L^1 stable and u is L^∞ stable.□

2.4 Proof of Theorem 2.3.1.

We first give the proof in one space dimension.

• Set

$$I := \int (\rho\psi_t + \rho u\psi_x)dxdt. \tag{2.48}$$

Using repeatedly the L^1-stability in ρ and ρu one has : $I = \sum_{i,n} rh^2[\rho_i^n(\psi_t)_i^n + (\rho u)_i^n(\psi_x)_i^n] + O(h) = \sum_{i,n} rh^2[\rho_i^n \frac{\psi_i^{n+1} - \psi_i^n}{rh} + (\rho u)_i^n \frac{\psi_{i+1}^n - \psi_i^n}{h}] + O(h)$. Then

$$I = -h\sum_{i,n}[\rho_i^{n+1} - \rho_i^n + r((\rho u)_i^n - (\rho u)_{i-1}^n)]\psi_i^n + O(h) \tag{2.49}$$

from a change in indices.

From (2.14), $\rho_i^{n+1} = \overline{\rho}_i + \alpha(\overline{\rho}_{i-1} - 2\overline{\rho}_i + \overline{\rho}_{i+1})$. Therefore $I = I_1 + I_2 + O(h)$, where

$$I_1 = -h\sum_{i,n}[\overline{\rho}_i - \rho_i^n + r((\rho u)_i^n - (\rho u)_{i-1}^n)]\psi_i^n, \tag{2.50}$$

$$I_2 = -h\alpha\sum_{i,n}(\overline{\rho}_{i-1} - 2\overline{\rho}_i + \overline{\rho}_{i+1})\psi_i^n = -h\alpha\sum_{i,n}\overline{\rho}_i(\psi_{i+1}^n - 2\psi_i^n + \psi_{i-1}^n) = O(h) \tag{2.51}$$

since, from (2.11), the L^1 stability in ρ implies L^1-stablility in $\overline{\rho}$, and from Taylor's formula in ψ. Distinguishing two cases in the signs of velocities it has been proved in chapter 1 that $I_1 = O(h)$

(with a change in notation : here $\overline{\rho}_i$ replaces ρ_i^{n+1} in formula (1.37)). Then $I = O(h)$ which proves (2.39) in one space dimension.□

• Set

$$J := \int [(\rho u)\psi_t + (\rho u^2)\psi_x + p\psi_x - \rho\Phi_x\psi]dxdt. \tag{2.52}$$

Since $\rho u, \rho u^2, p, \rho$ are L^1 stable, the proof of formula (1.7) with $\omega = \rho u$ (and in presence of p) gives, as (2.48)-(2.49),

$$\int [(\rho u)\psi_t + (\rho u^2)\psi_x + p\psi_x]dxdt =$$

$$-h\sum_{i,n}[(\rho u)_i^{n+1} - (\rho u)_i^n + r((\rho u^2)_i^n - (\rho u^2)_{i-1}^n) + r(p_i^n - p_{i-1}^n)]\psi_i^n + O(h). \tag{2.53}$$

A direct evaluation gives

$$\int \rho\Phi_x\psi dxdt = \sum_{i,n}\rho_i^n(\Phi_x)_i^n \int_{(i-\frac{1}{2})h<x<(i+\frac{1}{2})h, nrh<t<(n+1)rh} \psi dxdt =$$

$$\sum_{i,n}\rho_i^n(\Phi_x)_i^n rh^2\psi_i^n + O(h) = \sum_{i,n}\rho_i^{n+1}(\Phi_x)_i^{n+1}rh^2\psi_i^n + O(h) \tag{2.54}$$

from the L^1 stability of $\rho\Phi_x$. Therefore, from (2.52)-(2.54)

$$J = -h\sum_{i,n}[(\rho u)_i^{n+1} - (\rho u)_i^n + r((\rho u^2)_i^n - (\rho u^2)_{i-1}^n) + r(p_i^n - p_{i-1}^n) + rh\rho_i^{n+1}(\Phi_x)_i^{n+1}]\psi_i^n + O(h). \tag{2.55}$$

Developping $(\rho u)_i^{n+1}$ from (2.15)-(2.17), one obtains the simplification of the terms in Φ_x and the decomposition $J = J_1 + J_2 + J_3 + O(h)$ where

$$J_1 = -h\sum_{i,n}[\overline{(\rho u)}_i - (\rho u)_i^n + r((\rho u^2)_i^n - (\rho u^2)_{i-1}^n)]\psi_i^n, \tag{2.56}$$

$$J_2 = -h\alpha\sum_{i,n}(\overline{\rho u}_{i-1} - 2\overline{\rho u}_i + \overline{\rho u}_{i+1})\psi_i^n = -h\alpha\sum_{i,n}\overline{\rho u}_i(\psi_{i+1}^n - 2\psi_i^n + \psi_{i-1}^n), \tag{2.57}$$

$$J_3 = \frac{rh}{2}\sum_{i,n}(p_{i+1}^{n+1} - p_{i-1}^{n+1} - 2(p_i^n - p_{i-1}^n))\psi_i^n = \frac{rh}{2}\sum_{i,n}p_i^n(\psi_{i-1}^{n-1} - \psi_{i+1}^{n-1} - 2\psi_i^n + 2\psi_{i+1}^n). \tag{2.58}$$

As for I_1 above it has been proved in chapter 1 that $J_1 = O(h)$; as (2.51) $J_2 = O(h)$, and $J_3 = O(h)$ from Taylor's formula in ψ and L^1 stability in p from (2.13)-(2.11). Therefore $J = O(h)$. This proves (2.40) in one space dimension.□

• Similarly, as (2.54),

$I' := \int (p - K\rho)\psi = rh^2\sum_{i,n}(p_i^n - K\rho_i^n)\psi_i^n + O(h) = rh^2\sum_{i,n}(p_i^{n+1} - K\rho_i^{n+1})\psi_i^n + O(h)$

from the L^1 stability in p and ρ. From (2.13)

$$I' = Krh^2 \sum_{i,n} (\overline{\rho}_i - \rho_i^{n+1})\psi_i^n + O(h).$$

Then, from (2.14)

$$I' = -Krh^2\alpha \sum_{i,n} (\overline{\rho}_{i-1} - 2\overline{\rho}_i + \overline{\rho}_{i+1})\psi_i^n + O(h) = O(h),$$

which proves (2.41) in one space dimension.□

• Now we check the consistency (2.42) for the Poisson equation. Using the boundedness of $|\Phi_x|$ and the L^1-stability of ρ, one obtains

$\int(-(\Phi_x)\psi_x - 4\pi G\rho\psi)dxdt = \sum_{i,n}[-(\Phi_x)_i^n rh^2(\psi_x)_i^n - 4\pi G\rho_i^n rh^2\psi_i^n] + O(h) = \sum_{i,n}[-(\Phi_x)_i^n rh^2 \frac{1}{h}(\psi_{i+1}^n - \psi_i^n) - 4\pi G\rho_i^n rh^2\psi_i^n] + O(h) = \sum_{i,n} rh^2[\frac{1}{h}((\Phi_x)_i^n - (\Phi_x)_{i-1}^n) - 4\pi G\rho_i^n]\psi_i^n + O(h) = O(h)$ from (2.16). □

Proofs of Theorem 2.3.1 in two and three space dimensions. They are direct adaptations of the 1-D proof concerning the above calculations. The extension of the 1-D results stating that $I_1 = O(h), J_1 = O(h)$ is difficult since we must consider all neighboring cells in the transport step. A full proof is given in section 1.11.□

2.5 Numerical simulations.

All numerical calculations below were done on a standard PC in a few minutes. We first give four simulations in the pressureless case ($K = 0$).

Velocity increases in a gravitational collapse. With a fixed value of r given a priori it is difficult to produce a simulation, which is explained by the theorem : in a gravitational collapse, r depends very much on time, see $\|u\|_\infty$ in the bottom left panel. If at each iteration one adapts the value of r at the maximum value to respect the CFL condition $r\|\vec{u}\|_{L^\infty} = 1$, then one easily observes a gravitational collapse to a point. In figure 2.5.1 one has a cloud of cosmic fluid in the form of a disk surrounded by a void (top left panel). The values of ρ and (u, v) inside the disk are at random between 0.9 and 1.1 and between -0.1 and 0.1 respectively. One performs 80 iterations in a 200×200 window, $G = \frac{1}{4\pi}$, in the absence of expansion. One observes collapse to a point located in the center of the window (top right panel). We show the evolution of the sup. of velocity ($\max(|u|, |v|)$, bottom left panel) and the sup. of density (bottom right panel). The maximum of $|\Phi_x|, |\Phi_y|$ follows the growth of $\max(\rho)$ in the bottom right panel and reaches a value 150 but only on the cells close to the point concentration of matter. In two dimensions, the gradient of the gravitation potential is unbounded in a point concentration of matter. Nevertheless the scheme works very well provided one follows the above described adaptation of the value of r that enables to ensure the CFL condition in the most efficient way. This suggests that consistency of the scheme still holds even in 2-D point concentrations of matter.

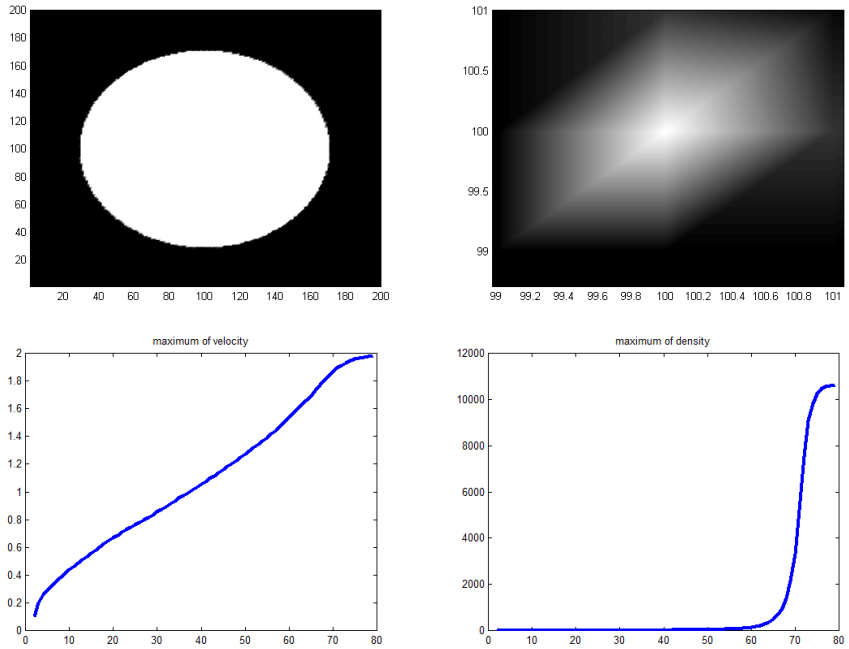


Figure 2.5.1. Simulation of a gravitational collapse in two dimensions in static background.

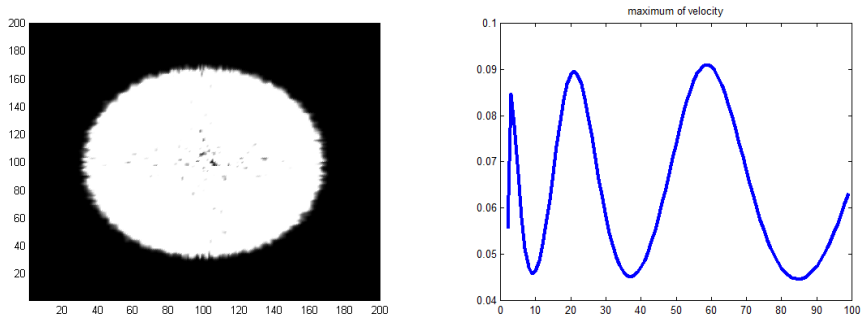


Figure 2.5.2. Gravitational collapse is frozen by fast expansion.

In figure 2.5.2, one considers the same cloud as in figure 2.5.1 but the background expands by a scale factor $a(t) = 1 + 10t$. After 100 iterations which raised the scale factor to the value 158 one observes that the cloud has not significantly changed (left panel). This shows that, even in the presence of gravitation, structures are frozen by fast expansion (Meszaros effect, [8] pp. 225-226), as observed in the absence of gravitation in figure 1.9.5. The presence of oscillations in sup. of velocity (right panel) shows that nevertheless the cloud is submitted to some stress due to the conflict between gravitation and expansion.

In figure 2.5.3 the cloud is rotating. Instead of a simple gravitational collapse to a central point as in figure 2.5.1, one observes creation of a "simili solar system". There is accumulation of the larger amount of matter in the center. First one observes formation of an irregular ring (top figure, 50 iterations). Sometimes a perfectly circular ring has been obtained, such as pictures in

[13], or some matter is ejected from the window. After 100 iterations (bottom figure) the ring has split into a few local accumulations of matter that reminds us of planets before accretion and a few dilute clouds of gas. This set is bound by gravitation and evolves slowly. Usually the "planets" rotate endlessly around the "sun" with slight modifications of the general configuration. The consistency of the scheme has been proved outside the central point as long as the "planets" are not pointlike, which gives confidence in the results. One observes the results are very sensitive to the initial data and that they evolve slowly with time as we could expect. The initial values of ρ are at random between 0 and 4, the initial velocities are all directed in a direction tangential to circles centered in the center of the window, with values $(0.1.rand.\frac{i}{\sqrt{(i^2+j^2)}}, 0.1.rand.\frac{j}{\sqrt{(i^2+j^2)}})$ where each *rand* denotes a random value between 0 and 1 ; the velocity chosen is equal to 0 in a neighborhood of the center ; as in figures 2.9.1 and 2.9.2 the values of r are adapted at each iteration so as to have $r\|u\|_{L^\infty} = 1$, absence of expansion, 100×100 window, G=0.0004. This problem is being intensively studied in computational physics by heuristic algorithms using a large number of pointmasses bound by gravitation, see [37] and references therein.

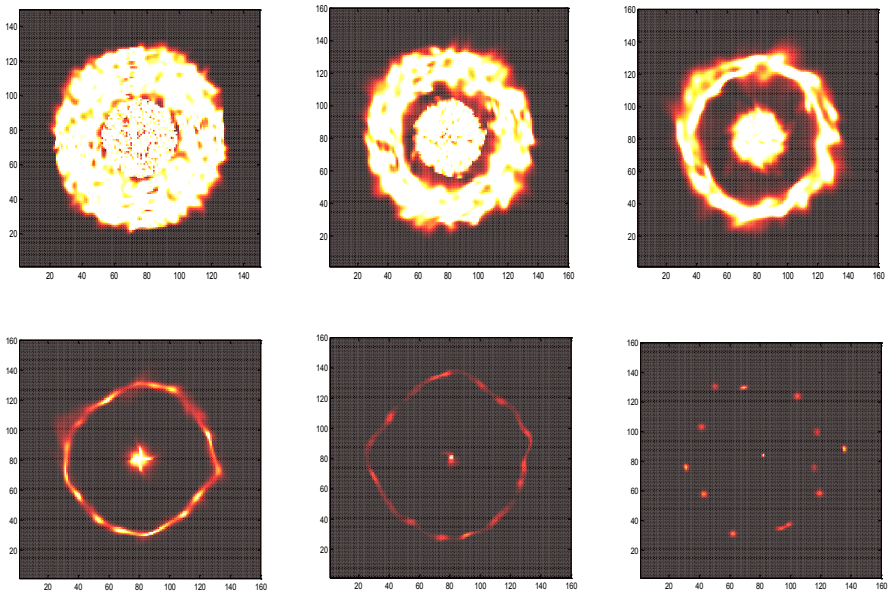


Figure 2.5.3. Formation of a ὰimili solar system from the gravitational collapse of a 2-D rotating dust cloud.

The 1-D simulation in figure 2.5.4 shows agglomeration of baryonic matter on the previously existing structures of dark matter when baryonic matter became decoupled with radiation. The top figure shows the initial conditions : dark matter (80 per cent, black continuous line) has formed structures when the universe was radiation dominated, while baryonic matter (20 per cent, red continuous line, scale multiplied by 10 for visualization, coupled to radiation before decoupling), is at random around the value 0.4 in density and between -0.1 and 0.1 in velocity. In the bottom figure, after a few iterations, one observes agglomeration of baryonic matter on the structures of dark matter, as expected : the baryonic material follows the behavior of dark matter [8] p. 260, [30] p. 473. This is modelled (with change of variable to take into account

the expansion) by two continuity equations (2.1) and two Euler equations (2.2) (dark matter and baryonic matter) with the same potential Φ ruled by the Poisson equation (2.3) where ρ is replaced by the sum of the two densities. For this bifluid system of five equations, stability and consistency hold exactly as in this chapter. 2-D simulations have given similar results.

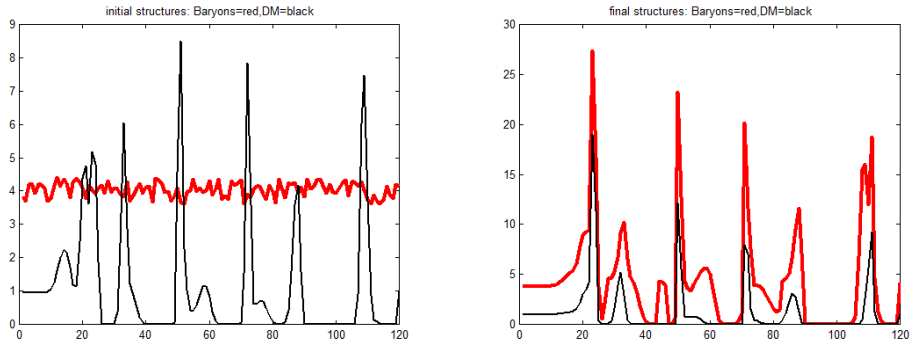


Figure 2.5.2. Potential wells of dark matter at decoupling : a system of five equations.

Now let us give simulations taking into account the role of pressure. Gravity tends to make small density perturbations in a static pressureless medium grow with time. In case the medium has pressure, pressure opposes this collapse. The classical theory of Jeans [8] (chapter 10) shows that only perturbations on a scale larger than the Jeans length can grow and shrink under their own gravity, thus producing structure formation. If the perturbation has a size smaller than the Jeans length the overdensity is smeared and dissipated by pressure. We numerically reproduce these facts from the fully nonlinear equations (2.1)-(2.4). We give the results in the one dimensional case since it allows a clearer visualization.

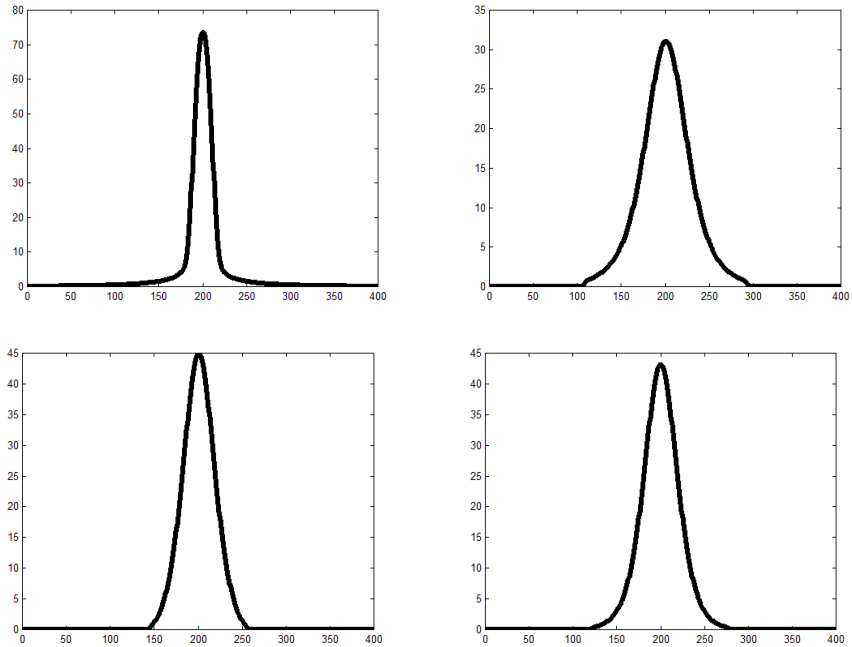


Figure 2.5.3. The basics of Jeans theory : a large enough cloud of gas reaches an equilibrium between the opposite actions of pressure and gravitation.

In figure 2.5.5 the initial conditions are two clouds of gas at rest centered at $x = 200$, of density=10, surrounded by a void : a large cloud of size 200 in the top panels, located on the interval $[100, 300]$, a small cloud of size 40 in the bottom panels, located on $[180, 220]$). The value of G is equal to 0.48, K=0.5 in the left panels and K=0 (i.e. absence of pressure) in the right panels. One performs 900 iterations with $r = 0.1$ and $\alpha = 0.1$. In the right panels one observes a gravitational collapse in the absence of pressure. In the top-left panel one observes a gravitational collapse of the large gas cloud even in the presence of pressure. In the bottom-left panel there is no gravitational collapse of the small gas cloud : one observes that the sup. of density passes from 10 to 2 : the overdensity is dissipated by pressure.

In figure 2.5.6 the data are the same as those of the top panels in figure 2.5.5 (large gas cloud). The initial size of the gas cloud is 200 on the interval [100,300], and its density is taken equal to 10. In the top-left panel the gas cloud has started a gravitational collapse : after 900 iterations, the sup. value of density has reached 75 with a support of size ≤ 100 : gravitation has dominated pressure. Then after 2000 iterations (top-right panel) the top value has decreased to 30 and the size of the support has increased to 200. This is due to the velocity (created by gravitation in the previous step) and to the pressure, whose influences have dominated the influence of gravitation in these iterations. Then, in the bottom-left panel, the size of the support diminishes again untill 100 while the top value reaches 45 : in these iterations gravitation has dominated pressure. After some small oscillations the cloud reaches an equilibrium (bottom-right panel) obtained after 18000 iterations.

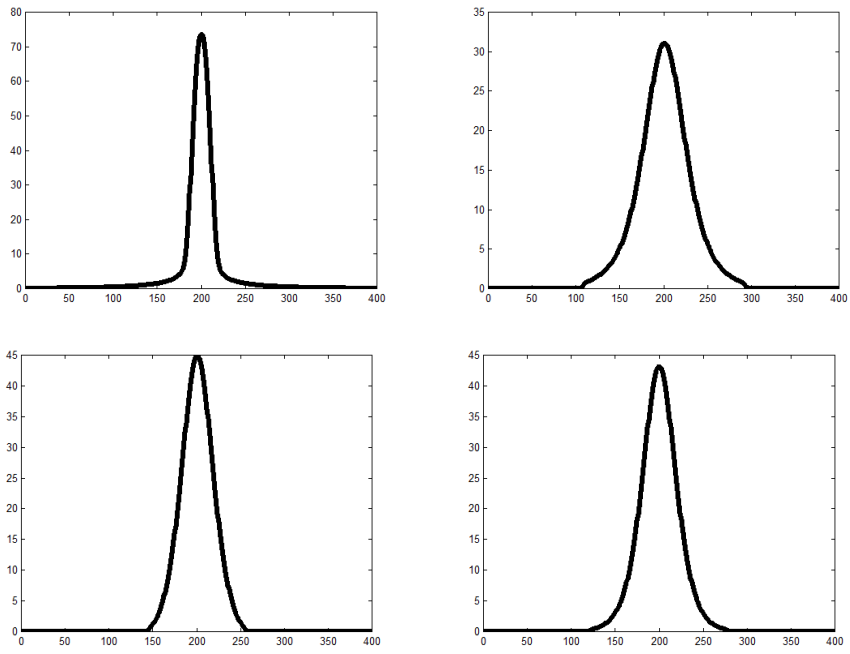


Figure 2.5.4. A cloud reaches an equilibrium between the opposite actions of pressure and gravitation.

2.6 Conclusion.

We have presented a numerical scheme for the system of collisional as well as non-collisional self-gravitating fluid dynamics. In 1-D and absence of pressure consistency of order one has been proved. In the other cases consistency holds under the assumptions of boundedness of the velocity field in the CFL condition and boundedness of the gradient of the gravitation potential. These two numerical properties have been checked in all numerical tests up to very small values of h in the presence of pressure ($K \neq 0$). In the absence of pressure ($K = 0$) and in three space dimension they have been rigorously proved whenever there is no point or string accumulation of matter. Even in these cases it has been observed that the scheme works well. If one does not accept the extrapolations of these properties for values of h smaller than those tested, the proof of the theorem shows that the approximate solutions from the scheme satisfy the equations up to a deviation of order one in h. As an application we have numerically reproduced various events in cosmology and astrophysics.

Chapitre 3

The system of Ideal Gas dynamics

In this chapter we present a 3-D numerical scheme for the approximation of the system of gas dynamics. Consistency in the sense of distributions is studied. We prove that, as long as the boundedness of the velocity field (in the CFL condition) and the positiveness of the energy are numerically verified when the space step tends to 0, the scheme provides a numerical solution which satisfies the equations in the sense of distributions with an approximation of order one in the space step. Numerical verifications of convergence are done from classical 1-D tests (Sod, Woodward-Colella, Toro). These verifications provide numerical evidence that the scheme produces the exact solution with arbitrary precision. This scheme gives back the numerical results on the six 2-D Riemann problems presented by P. D. Lax in [25] and [26], up to the smallest details. This simple order one low-cost 3-D scheme is obtained from the convection-pressure correction method proposed by Le Roux et al [2].

3.1 Introduction.

The system of ideal gases has been studied by many authors, see for instance [17], [28] and [34]. In [25] and [26] the author points out the need of a mathematical justification of the numerical solutions of the 2-D Riemann problems presented in these articles. In this chapter we introduce a simple numerical scheme which permits to provide a 3-D consistency proof in the sense of distributions under the numerical assumptions of boundedness of the velocity field (in the CFL condition) and positiveness of the energy when the space step $h \to 0$. Of course, from a theoretical point of view one cannot be sure that these numerical assumptions always hold for every h when $h \to 0$, however we point out that, for the scheme presented in this chapter, these assumptions have always been satisfied in the 1-D Sod, Woodward-Colella, Toro tests and the 2-D Riemann problems in [17], [25], [26], [28] and [34] for all tested values of h, some of them very small. The proof in this chapter shows that, for any given family of test functions with uniformly bounded support and uniformly bounded first and second derivatives, then the numerical solution satisfies the equations in the sense of distributions within a small deviation of order one in h, whenever the numerical velocity remains bounded (in the CFL condition) and the energy density remains positive, which is presumably verified in physical cases for all h when $h \to 0$ and can easily be checked numerically up to very small values of h. As far as the author knows the proof is new and relies on the specific form of the scheme. This proof is an extension to the system of ideal gases in 3-D of the consistency proof given in [2] for the far simpler system of pressureless fluid dynamics in 1-D. In addition to the consistency proof, numerical convergence and low-cost efficiency of the scheme are checked by classical 1-D tests (Sod, Woodward-Colella,

Toro). The numerical results in [17], [25], [26], [28] and [34] on 2-D Riemann problems are also obtained, even up to the smallest details, which suggests that the numerical schemes in [17], [28], [34] could be convergent, as this is conjectured by P. D. Lax in [25] and [26]. Now let us recall the system of ideal gases :

$$\frac{\partial \rho}{\partial t} + \vec{\nabla}.(\rho \vec{u}) = 0, \tag{3.1}$$

$$\frac{\partial}{\partial t}(\rho \vec{u}) + \vec{\nabla}.(\rho \vec{u} \otimes \vec{u}) + \vec{\nabla} p = \vec{0}, \tag{3.2}$$

$$\frac{\partial}{\partial t}(\rho e) + \vec{\nabla}.[(\rho e + p)\vec{u}] = 0, \tag{3.3}$$

$$p = (\gamma - 1)(\rho e - \rho \frac{\vec{u}^2}{2}), \tag{3.4}$$

where $\rho, \vec{u} = (u, v, w), p, e$ denote respectively the density, the velocity vector, the pressure and the density of total energy ; γ is a constant.

The scheme and its consistency proof adapt easily to systems of fluid dynamics involving the continuity equation (3.1), such as the Saint-Venant equations or the compressible Navier-Stokes equations.

3.2 Statement of the scheme.

The real line is divided into intervals $I_i =]ih - \frac{1}{2}h, ih + \frac{1}{2}h[, i \in \mathbb{Z}$. We set $t_n = nrh$ for r small enough. We will construct step functions $\rho(x,t)$, $u(x,t)$, $p(x,t)$, ... depending on h, which are constant on the rectangles $I_i \times]t_n, t_{n+1}[$, whose step values are denoted $\rho_i^n, u_i^n, p_i^n, \ldots$, respectively. The indices h are skipped to simplify the notation : ρ stands for $\rho_h, \ldots$. From these step functions ρ and u, we define the step functions ρu, $\rho u^2, \ldots$ by $(\rho u)_i^n = \rho_i^n u_i^n$ and $(\rho u^2)_i^n = \rho_i^n (u_i^n)^2, \ldots$. The initial condition (ρ^0, u^0) is discretized on the intervals I_i by taking mean values on these intervals. We always assume that u^0 is a L^∞ function and that ρ^0, e^0 are positive L^1 functions null at infinity.

Statement of the scheme in one dimension. In one space dimension the equations (3.1)-(3.4) reduce to

$$\rho_t + (\rho u)_x = 0, \tag{3.5}$$

$$(\rho u)_t + (\rho u^2)_x + p_x = 0, \tag{3.6}$$

$$(\rho e)_t + (\rho e u)_x + (pu)_x = 0, \tag{3.7}$$

$$p = (\gamma - 1)(\rho e - \frac{\rho u^2}{2}). \tag{3.8}$$

We assume the set $\{\rho_i^n, (\rho u)_i^n, (\rho e)_i^n, u_i^n, p_i^n\}_{i \in Z}$ is given. The set $\{\rho_i^{n+1}, (\rho u)_i^{n+1}, (\rho e)_i^{n+1}, u_i^{n+1}, p_i^{n+1}\}_{i \in Z}$ is defined as follows.

If $a < b$ one sets

$$L(a,b) := length\ of\ [0,1] \cap [a,b], \tag{3.9}$$

i.e.

$$L(a,b) = max(0, min(1,b) - max(0,a)). \tag{3.10}$$

• *Transport step.* In order to use the convergence proof in [2] we set

$$\overline{\rho}_i := \rho_{i-1}^n L(-1 + ru_{i-1}^n, ru_{i-1}^n) + \rho_i^n L(ru_i^n, 1 + ru_i^n) + \rho_{i+1}^n L(1 + ru_{i+1}^n, 2 + ru_{i+1}^n). \tag{3.11}$$

When the CFL condition $r\|u\|_\infty \leq 1$ is satisfied the first term represents the matter issued from the cell I_{i-1} between times t_n and t_{n+1} that lies in the cell I_i at time t_{n+1}. The second term represents the matter from the cell I_i that remains in I_i at time t_{n+1}. The third term is similar to the first one : it concerns matter issued from the cell I_{i+1} that lies in the cell I_i at time t_{n+1}. Note that $\overline{\rho}_i$ depends on n, which is not explicitely stated to shorten the notation. The same discretization as the one in (3.11) gives

$$\overline{(\rho u)}_i := (\rho u)_{i-1}^n L(-1 + ru_{i-1}^n, ru_{i-1}^n) +$$
$$(\rho u)_i^n L(ru_i^n, 1 + ru_i^n) + (\rho u)_{i+1}^n L(1 + ru_{i+1}^n, 2 + ru_{i+1}^n), \tag{3.12}$$

$$\overline{(\rho e)}_i := (\rho e)_{i-1}^n L(-1 + ru_{i-1}^n, ru_{i-1}^n) +$$
$$(\rho e)_i^n L(ru_i^n, 1 + ru_i^n) + (\rho e)_{i+1}^n L(1 + ru_{i+1}^n, 2 + ru_{i+1}^n). \tag{3.13}$$

The state law is set in the form

$$p_i^{n+1} := (\gamma - 1)[\overline{(\rho e)}_i - \frac{(\overline{(\rho u)}_i)^2}{2\overline{\rho}_i}] \tag{3.14}$$

if $\overline{\rho}_i \neq 0$,

$$p_i^{n+1} = 0 \tag{3.15}$$

if $\overline{\rho}_i = 0$.

• *Averaging step.* The averaging step is needed to avoid oscillations caused by the centered discretization in the next step. From numerical tests we choose a value $\alpha \in]0, \frac{1}{2}[$ and we set

$$\rho_i^{n+1} := \alpha\overline{\rho}_{i-1} + (1 - 2\alpha)\overline{\rho}_i + \alpha\overline{\rho}_{i+1}, \tag{3.16}$$

$$\widetilde{(\rho u)}_i := \alpha\overline{(\rho u)}_{i-1} + (1 - 2\alpha)\overline{(\rho u)}_i + \alpha\overline{(\rho u)}_{i+1}, \tag{3.17}$$

$$\widetilde{(\rho e)}_i := \alpha\overline{(\rho e)}_{i-1} + (1 - 2\alpha)\overline{(\rho e)}_i + \alpha\overline{(\rho e)}_{i+1}. \tag{3.18}$$

• *Pressure correction step.* We set

$$(\rho u)_i^{n+1} := \widetilde{(\rho u)}_i - \frac{r}{2}(p_{i+1}^{n+1} - p_{i-1}^{n+1}), \tag{3.19}$$

$$u_i^{n+1} := \frac{(\rho u)_i^{n+1}}{\rho_i^{n+1}} \tag{3.20}$$

if $\rho_i^{n+1} \neq 0$, and any value if $\rho_i^{n+1} = 0$. We set

$$(\rho e)_i^{n+1} := \widetilde{(\rho e)}_i - \frac{r}{2}(p_{i+1}^{n+1}u_{i+1}^{n+1} - p_{i-1}^{n+1}u_{i-1}^{n+1}). \tag{3.21}$$

Proposition 3.2.1. *$\rho_i^n = 0$ implies $(\rho u)_i^n = 0, (\rho e)_i^n = 0$ and $p_i^n = 0$.*

proof. The proof is an induction on n. For $n = 0$ it holds by construction. Assume the property holds for n. Then, if $\rho_i^{n+1} = 0$, (3.16) implies

$$\overline{\rho}_{i-1},\ \overline{\rho}_i,\ \overline{\rho}_{i+1} = 0. \tag{3.22}$$

From (3.11)-(3.13), $\overline{\rho}_i = 0$ implies $\overline{(\rho u)}_i = 0$ and $\overline{(\rho e)}_i = 0$, see proposition 1.6.1 in chapter 1 for details, the induction assumption is used here. Therefore, (3.22) implies

$$\overline{(\rho u)}_{i-1} = 0,\ \overline{(\rho u)}_i = 0,\ \overline{(\rho u)}_{i+1} = 0,\ \ \overline{(\rho e)}_{i-1} = 0,\ \overline{(\rho e)}_i = 0,\ \overline{(\rho e)}_{i+1} = 0.$$

Then, from (3.17)-(3.18), one obtains $\widetilde{(\rho u)}_i = 0$ and $\widetilde{(\rho e)}_i = 0$. Formula (3.15) implies $p_{i-1}^{n+1} = 0, p_i^{n+1} = 0, p_{i+1}^{n+1} = 0$. Finally all terms in (3.19), (3.21) are null.□

It follows from (3.11), (3.16) that ρ is positive. Since the coefficients L in (3.10) represent transport i.e. a new repartition of matter at time t_{n+1}, as in Theorem 1.5.1 and section 1.5, one has $\sum_i \rho_i^n h = \sum_i \rho_i^0 h$. From the positiveness of ρ one has the L^1 stability in ρ.

Statement of the scheme in two and three dimensions. The equations in the two dimensional case are

$$\rho_t + (\rho u)_x + (\rho v)_y = 0, \tag{3.23}$$

$$(\rho u)_t + (\rho u^2)_x + (\rho uv)_y + p_x = 0, \tag{3.24}$$

$$(\rho v)_t + (\rho uv)_x + (\rho v^2)_y + p_y = 0, \tag{3.25}$$

$$(\rho e)_t + (\rho eu)_x + (\rho ev)_y + (pu)_x + (pv)_y = 0, \tag{3.26}$$

$$p = (\gamma - 1)(\rho e - \rho\frac{u^2 + v^2}{2}). \tag{3.27}$$

The two dimensional space (x, y) is divided into square cells $C_{i,j}$ of side h and centers $(ih, jh)_{i,j\in\mathbb{Z}}$: $C_{i,j}$ is the set of all (x, y) such that $ih - \frac{h}{2} < x < ih + \frac{h}{2}$ and $jh - \frac{h}{2} < y < jh + \frac{h}{2}$. We assume the set $\{\rho_{i,j}^n, (\rho u)_{i,j}^n, (\rho v)_{i,j}^n, (\rho e)_{i,j}^n, u_{i,j}^n, v_{i,j}^n, p_{i,j}^n\}_{i,j\in Z}$ is given. The set $\{\rho_{i,j}^{n+1}, (\rho u)_{i,j}^{n+1}, (\rho v)_{i,j}^{n+1}, (\rho e)_{i,j}^{n+1}, u_{i,j}^{n+1}, v_{i,j}^{n+1}, p_{i,j}^{n+1}\}_{i,j\in Z}$ is defined as follows. We set

$$A(a, b) := L(a, 1 + a).L(b, 1 + b) \tag{3.28}$$

which is the area of the intersection of the square of vertices $(0, 0), (0, 1), (1, 0), (1, 1)$ with the square of vertices $(a, b), (1 + a, b), (a, 1 + b), (1 + a, 1 + b)$. Then we set

• *Transport step.* As in the 1D case, when the CFL condition $r\|u\|_\infty$ holds, we set

$$\overline{\rho}_{i,j} := \sum_{-1\le\lambda,\mu\le 1} \rho_{i+\lambda,j+\mu}^n A(\lambda + ru_{i+\lambda,j+\mu}^n,\ \mu + rv_{i+\lambda,j+\mu}^n), \tag{3.29}$$

$$(\overline{\rho u})_{i,j} := \sum_{-1\le\lambda,\mu\le 1} (\rho u)_{i+\lambda,j+\mu}^n A(\lambda + ru_{i+\lambda,j+\mu}^n,\ \mu + rv_{i+\lambda,j+\mu}^n), \tag{3.30}$$

$$(\overline{\rho v})_{i,j} := \sum_{-1\le\lambda,\mu\le 1} (\rho v)^n_{i+\lambda,j+\mu} A(\lambda + ru^n_{i+\lambda,j+\mu},\ \mu + rv^n_{i+\lambda,j+\mu}), \tag{3.31}$$

$$(\overline{\rho e})_{i,j} := \sum_{-1\le\lambda,\mu\le 1} (\rho e)^n_{i+\lambda,j+\mu} A(\lambda + ru^n_{i+\lambda,j+\mu},\ \mu + rv^n_{i+\lambda,j+\mu}), \tag{3.32}$$

$$p^{n+1}_{i,j} := (\gamma-1)((\overline{\rho e})_{i,j} - \frac{((\overline{\rho u})_{i,j})^2 + ((\overline{\rho v})_{i,j})^2}{2\overline{\rho}_{i,j}}). \tag{3.33}$$

Interpretation of (3.29)-(3.32) is a transport in 2-D, see section 1.6, similarly to (3.11), (3.13) in 1-D.

- *Averaging step.* Let α, $0 < \alpha < \frac{1}{20}$, be given in the scheme. Set

$$\begin{aligned}\rho^{n+1}_{i,j} := \alpha(2\overline{\rho}_{i-1,j-1} + 2\overline{\rho}_{i-1,j+1} + 2\overline{\rho}_{i+1,j-1} + 2\overline{\rho}_{i+1,j+1} + 3\overline{\rho}_{i-1,j} + \\ 3\overline{\rho}_{i,j-1} + 3\overline{\rho}_{i,j+1} + 3\overline{\rho}_{i+1,j}) + (1-20\alpha)\overline{\rho}_{i,j},\end{aligned} \tag{3.34}$$

$$\begin{aligned}\widetilde{(\rho u)}_{i,j} := \alpha(2\overline{(\rho u)}_{i-1,j-1} + 2\overline{(\rho u)}_{i-1,j+1} + 2\overline{(\rho u)}_{i+1,j-1} + 2\overline{(\rho u)}_{i+1,j+1} + 3\overline{(\rho u)}_{i-1,j} + \\ 3\overline{(\rho u)}_{i,j-1} + 3\overline{(\rho u)}_{i,j+1} + 3\overline{(\rho u)}_{i+1,j}) + (1-20\alpha)\overline{(\rho u)}_{i,j}.\end{aligned} \tag{3.35}$$

We set the same formula for $\widetilde{(\rho v)}_{i,j}, \widetilde{(\rho e)}_{i,j}$, replacing u by v, e respectively.

Remark. The scheme adapts to the shallow water equations. Then, it has been noticed in the cylindrical dam break test of Toro [41], pp. 245-260, that the averaging step does not work in some regions thus producing strong oscillations and an uncorrect result. To make the averaging efficient in these regions it suffices to change (u, v) into $(u + rand, v - rand)$ in each iteration, where *rand* is a random value between 0 and 4. Then one obtains the correct solution. Therefore, in certain geometrical situations the averaging (3.34)-(3.35) should be modified to make it efficient.

- *Pressure correction step.* A centered discretization gives

$$(\rho u)^{n+1}_{i,j} := \widetilde{\rho u}_{i,j} - \frac{r}{2}(p^{n+1}_{i+1,j} - p^{n+1}_{i-1,j}). \tag{3.36}$$

A similar formula is given for $(\rho v)^{n+1}_{i,j}$, using y-derivatives. If $\rho^{n+1}_{i,j} \neq 0$, then

$$u^{n+1}_{i,j} := \frac{(\rho u)^{n+1}_{i,j}}{\rho^{n+1}_{i,j}},\ v^{n+1}_{i,j} := \frac{(\rho v)^{n+1}_{i,j}}{\rho^{n+1}_{i,j}}, \tag{3.37}$$

if $\rho^{n+1}_{i,j} = 0$ then $u^{n+1}_{i,j}$ can be given any value as in 1-D (Proposition 3.2.1 holds with the same proof).

$$(\rho e)^{n+1}_{i,j} := \widetilde{\rho e}_{i,j} - \frac{r}{2}(p^{n+1}_{i+1,j}u^{n+1}_{i+1,j} - p^{n+1}_{i-1,j}u^{n+1}_{i-1,j}) - \frac{r}{2}(p^{n+1}_{i,j+1}v^{n+1}_{i,j+1} - p^{n+1}_{i,j-1}v^{n+1}_{i,j-1}). \tag{3.38}$$

The scheme in 2-D has the same properties as those in 1-D.

The scheme in three space dimension is very similar to the scheme in two space dimension (29)-(38). Let $C_{i,j,k}$ be the cube of all (x, y, z) such that $(i-\frac{1}{2})h < x < (i+\frac{1}{2})h, (j-\frac{1}{2})h < y < (j+\frac{1}{2})h, (k-\frac{1}{2})h < z < (k+\frac{1}{2})h$. Let

$$V(a,b,c) = L(a,1+a).L(b,1+b).L(c,1+c) \tag{3.39}$$

be the volume of the intersection of the cube of vertices (i,j,k), $i,j,k = 0$ or 1, with the cube of vertices $(a+i, b+j, c+k)$, $i,j,k = 0$ or 1. If $\omega = \rho, \rho u, \rho v, \rho w, \rho e$ successively, one sets

$$\overline{\omega}_{i,j,k} =$$

$$\sum_{-1\leq\lambda,\mu,\nu\leq 1} \omega^n_{i+\lambda,j+\mu,k+\nu} V(\lambda + ru^n_{i+\lambda,j+\mu,k+\nu},\ \mu + rv^n_{i+\lambda,j+\mu,k+\nu},\ \nu + rw^n_{i+\lambda,j+\mu,k+\nu}). \tag{3.40}$$

We extend (3.34)-(3.35) by taking an average over the cell $C_{i,j,k}$ and its 26 neighbors in order that Taylor's formula in ψ annihilates the first order terms.

3.3 Statement of the consistency theorem.

The approximate solutions $\omega_h(x,y,z,t)$ (denoted here ω to simplify the notation) are constant equal to $\omega^n_{i,j,k}$ (depending on h) on $C_{i,j,k}$ for $nrh < t < (n+1)rh$ where $\omega = \rho, u, v, w, p, ...$ We assume ρ^0 and e^0 are positive L^1 functions and u^0, v^0, w^0 are L^∞ functions. For simplification, boundary problems are eliminated by assuming that the physical variables under concern tend to 0 at infinity.

Theorem 3.3.1. Consistency under numerical assumptions. *Assume that on some time interval $[0,T]$ (i.e. $\forall (i,j,k) \in \mathbb{Z}^3$ and $\forall n \leq \frac{T}{rh}$) one has $\forall h > 0$ small enough the CFL condition*

$$r|u^n_{i,j,k}| \leq 1,\ r|v^n_{i,j,k}| \leq 1,\ r|w^n_{i,j,k}| \leq 1, \tag{3.41}$$

and the positiveness of the energy

$$e^n_{i,j,k} \geq 0. \tag{3.42}$$

Then concerning the conservation laws (3.1)-(3.3) the scheme is consistent in the sense of distributions. The consistency in the sense of distributions of the state law (3.4) takes place in regions in which ρ is strictly positive and in which the approximate solution has the familiar aspect of piecewise $\mathcal{C}^1$ functions having limits on both sides of the surfaces of discontinuity : shock waves, contact discontinuities, rarefaction waves, for instance.

This means that $\forall \psi \in C^\infty_c(\mathbb{R}^3 \times]0,T[)$,

$$\int (\rho_h \psi_t + \rho_h u_h \psi_x + \rho_h v_h \psi_y + \rho_h w_h \psi_z) dxdydzdt \to 0, \tag{3.43}$$

$$\int \{\rho_h u_h \psi_t + \rho_h (u_h)^2 \psi_x + \rho_h u_h v_h \psi_y + \rho_h u_h w_h \psi_z + p_h \psi_x\} dxdydzdt \to 0, \tag{3.44}$$

and similar limits for the two other components of the Euler equation in $(\rho_h v_h), (\rho_h w_h)$, and the energy equation. Further

$$\int [p_h - (\gamma - 1)((\rho e)_h - \frac{(\rho u)_h . u_h}{2})] \psi dxdydzdt \to 0, \tag{3.45}$$

when the support of ψ is contained in one of the regions mentioned above.

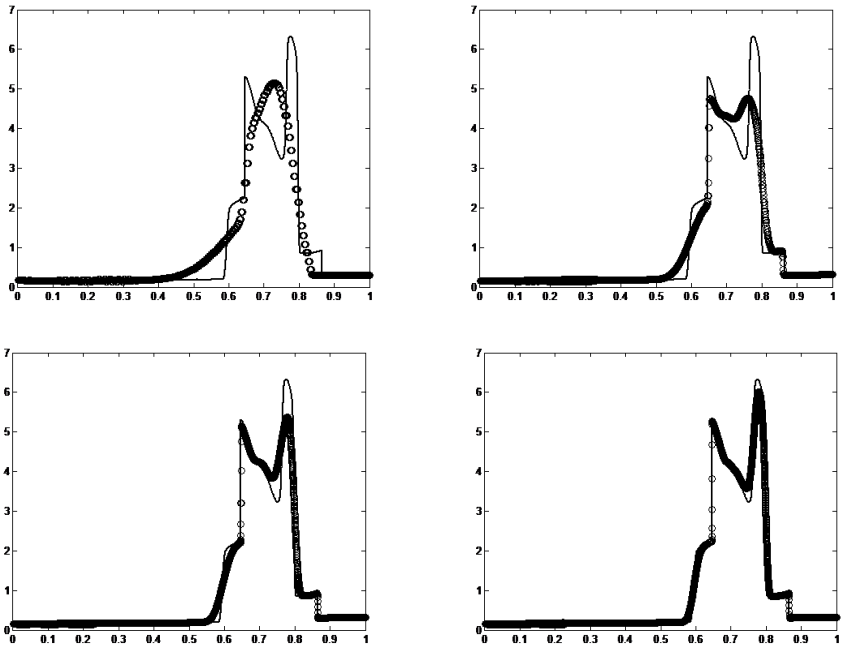
In all numerical tests one has observed for all considered values of h, some of them extremely small, that (3.41) holds provided $r > 0$ small enough and α chosen not too small, and that (3.42) holds. One has no physically relevant example in which (3.42) would not be satisfied as soon as (3.41) holds. The state law is obtained from experiments in space-time regions in which the flow is not too irregular, therefore the above limitation on the validity of the state law together with assumptions (3.41)-(3.42) cover the domain of physical relevance of the system of perfect gases.

In practice it is impossible to verify (3.41)-(3.42) for an infinite set of values of h : the proof of the theorem gives an approximation result.

Remark : the theorem as a rigorous approximation result with computer aided proof. One can notice from the proof that the bounds that prove the consistency are of order one in h and do not depend on the test function ψ itself, but on the size of its support and on L^∞ bounds of its first and second derivatives. Therefore, for small h and for a family of test functions with uniformly bounded size of support as well as first and second derivatives, the theorem can be transformed into an approximation result on the smallness of the left members of (3.43)-(3.45), which has the advantage to be for sure if one considers only values of h for which one has checked the properties (3.41) and (3.42).

First, we show the numerical evidence of convergence on six 1-D standard tests [40], [43]. Section 6 provides six 2-D tests from [17], [28] and [34].

3.4 Numerical evidence of convergence of the scheme.



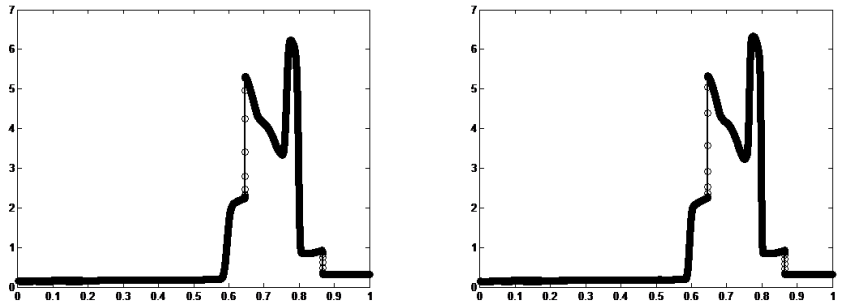
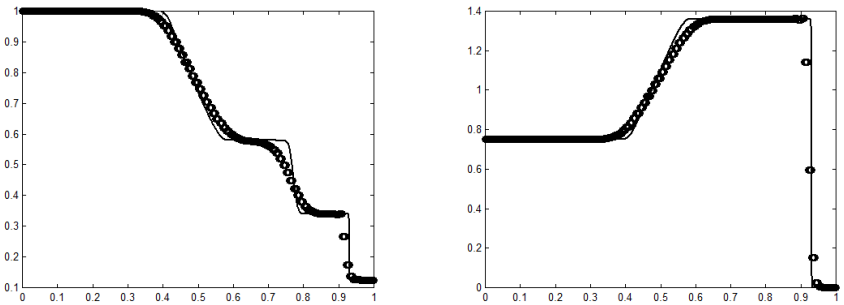


Figure 3.4.1. Evolution of the final result in the !woodward-Colella test and comparison with the exact solution.

The initial conditions are : if $0 < x < 0.1$ then $p = 1000$, if $0.1 < x < 0.9$ then $p = 0.01$, if $0.9 < x < 1$ then $p = 100$; $\rho = 1$ and $u = 0$ everywhere, $\gamma = 1.4$. Reflecting boundary conditions are used (strictly speaking this test does not enter into the theorem where boundary influence is eliminated). The value of α has been fixed $= 0.10$. From the top-left panel to the bottom-right panel the values of h are $200^{-1}, 1200^{-1}, 4000^{-1}, 10000^{-1}, 40000^{-1}, 160000^{-1}$ and the numbers of iterations are $196, 1196, 3996, 9995, 52400, 217000$ respectively. The CFL number r was chosen as large as possible, according to the value of α and to the number of iterations ($r = 0.038, 0.035, 0.032, 0.032, 0.029, 0.028$). One observes the numerical convergence to the exact solution which is nearly obtained in the middle-right panel. In the two top panels the results are roughly identical to the results from the Godunov scheme reported in [43] p. 144 with the same values of h (although the scheme in this chapter is considerably simpler). The exact solution is attained rather quickly and there is no degeneracy at all after a very large number of iterations that were stopped at 217 000 in the right-bottom panel : the averaging step has no undesirable smoothing effect.

In the Sod test the exact solution is reached after a small number of iterations with a very good accuracy. The Sod test was also continued during several hundred thousand iterations without any indication of degeneracy. Now we present the results of the four Toro tests for ideal gases [40] pp. 214-220.



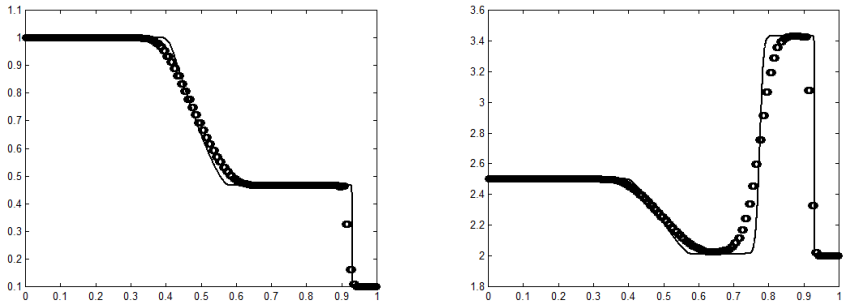


Figure 3.4.2. Toro test 1 : èntropy satisfaction:

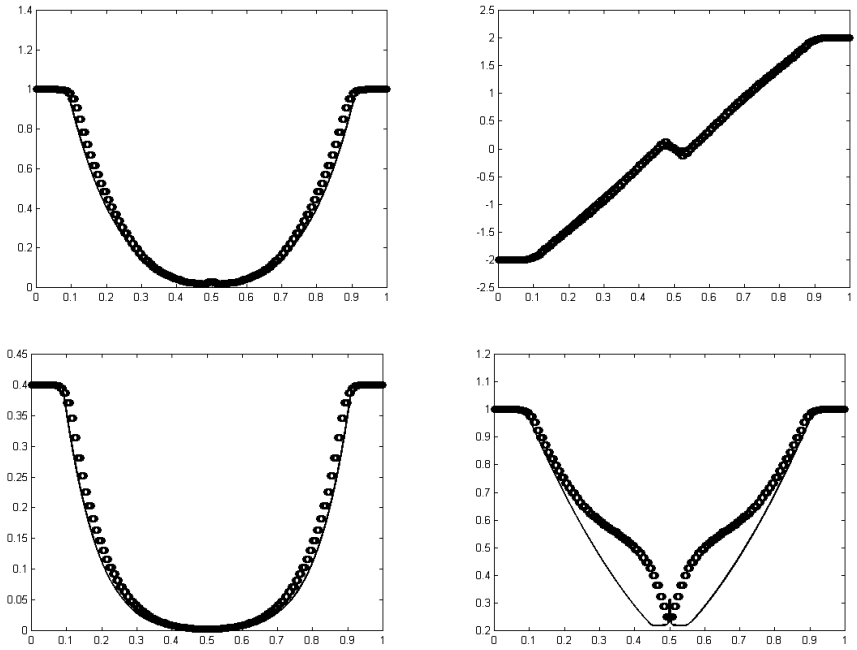
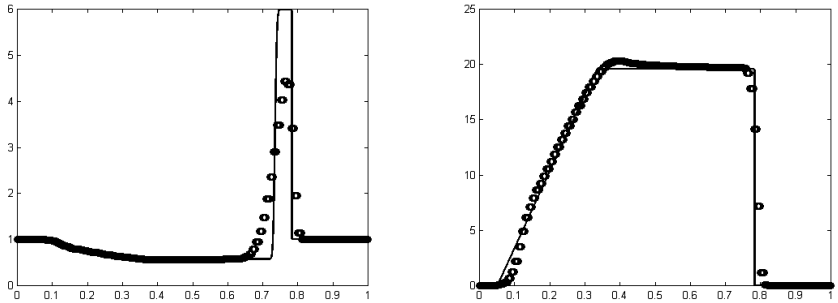


Figure 3.4.3. Toro test 2 : performance for low density flows:



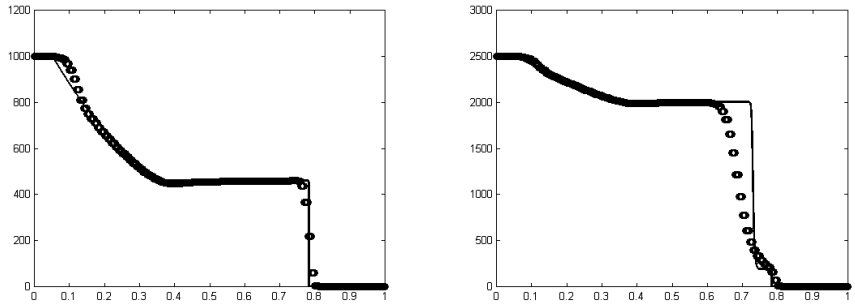


Figure 3.4.4. Toro test 3 : robustness and accuracy.

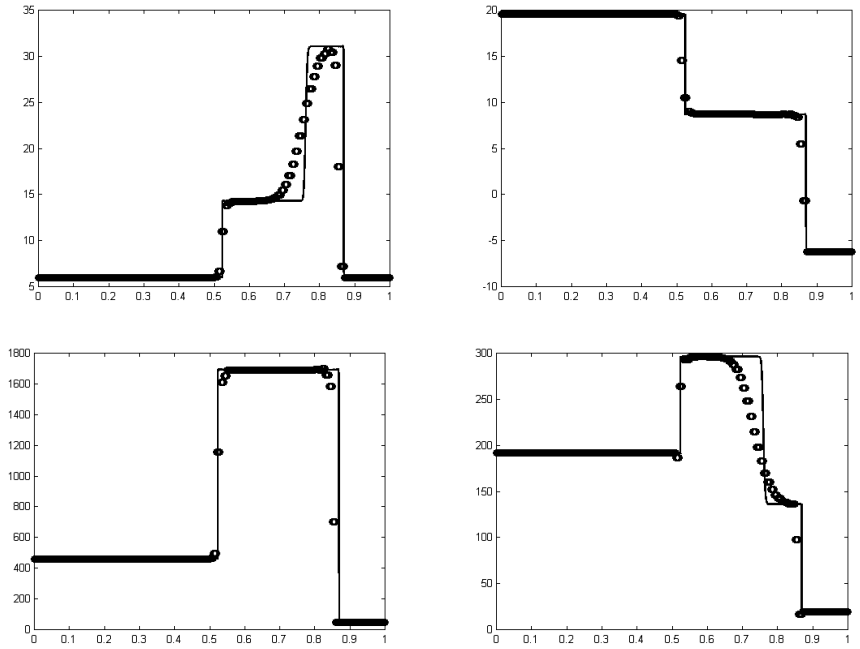


Figure 3.4.5. Toro test 4 : robustness.

Figures 3.4.2 to 3.4.5 present results of the scheme on the four Toro tests [40] pp. 214-220 in the conditions of these tests : 100 cells and same output time in order to permit comparison with the Godunov, Lax-Friedrichs and Richtmyer schemes presented in [40]. The scheme in this chapter gives numerical results comparable with those of the Godunov scheme for the same value of h (sometimes better results : see the internal energy in figure 3.4.3). From left-top to right-bottom panel, one represents the density, the velocity, the pressure and the internal energy. In all tests the elapsed time was about 0.03 seconds on a standard PC. The solution represented by the continuous line is the solution from the scheme in this chapter for small enough values of h. We observed it coincides with the exact solution given in Toro [40]. Therefore these four figures provide four numerical verifications of the theoretical proof of consistence. We notice the presence of a spike in internal energy in figure 3.4.3 : it was already observed in chapter 1 (figure 1.9.1 and numerous papers quoted there), where the scheme is limited to its first step. It shows

that convergence cannot be obtained in sup norm even in regions where the exact solution is continuous. Now we give the details of the tests in the following array. The Riemann problem $(\rho_l, u_l, p_l, \rho_r, u_r, p_r)$ of figure 3.4.5 is $(5.99924, 19.5975, 460.894, 5.99242, -6.19633, 46.0950)$. On the left we give the values r, α and the number of iterations used in the conditions of the Toro tests ($h = 0.01$) ; on the right we give the values of h, r, α used to superpose exactly the numerical solution on the exact solution given in [40].

$fig.$	$\rho_l, u_l, p_l, \rho_r, u_r, p_r$	r	α	$iter$	h	r	α
2	1,0.75,1,0.125,0,0.1	0.5	0.05	40	$6\ 10^{-4}$	0.5	0.1
3	1,-2,0.4,1,2,0.4	0.48	0.03	31	$7\ 10^{-5}$	0.45	0.05
4	1,0,1000,1,0,0.01	0.02	0.01	60	$5\ 10^{-5}$	0.012	0.1
5	see above	0.08	0.1	37	10^{-4}	0.08	0.1

3.5 Consistency proofs : first part.

Proof of the theorem in one dimension.

• Set

$$I := \int (\rho\psi_t + \rho u\psi_x)dxdt. \tag{3.46}$$

Since ρ and ρu are L^1-stable, it is proved, formula (1.37), that

$$I = -h\sum_{i,n}[\rho_i^{n+1} - \rho_i^n + r((\rho u)_i^n - (\rho u)_{i-1}^n)]\psi_i^n + O(h) \tag{3.47}$$

(in [2] one has ψ_i^{n+1} in place of ψ_i^n ; it does not matter : $h\sum_{i,n}(\rho_i^{n+1} - \rho_i^n)(\psi_i^{n+1} - \psi_i^n) = h\sum_{i,n}\rho_i^n(\psi_i^n - \psi_i^{n-1} - \psi_i^{n+1} + \psi_i^n) = h\sum_{i,n}\rho_i^n O(h^2) = O(h)$, same for the term in ρu). From (3.16), $\rho_i^{n+1} = \overline{\rho}_i + \alpha(\overline{\rho}_{i-1} - 2\overline{\rho}_i + \overline{\rho}_{i+1})$. Therefore $I = I_1 + I_2 + O(h)$, where

$$I_1 = -h\sum_{i,n}[\overline{\rho}_i - \rho_i^n + r((\rho u)_i^n - (\rho u)_{i-1}^n)]\psi_i^n, \tag{3.48}$$

$$I_2 = -h\alpha\sum_{i,n}(\overline{\rho}_{i-1} - 2\overline{\rho}_i + \overline{\rho}_{i+1})\psi_i^n = -h\alpha\sum_{i,n}\overline{\rho}_i(\psi_{i+1}^n - 2\psi_i^n + \psi_{i-1}^n) = O(h) \tag{3.49}$$

since, from (3.11), the L^1 stability in ρ implies the L^1-stablility in $\overline{\rho}$. It has been proved in the end of the proof of Theorem1.8.1 that $I_1 = O(h)$ (with a change in notation : here $\overline{\rho}_i$ replaces ρ_i^{n+1} in formula (37) in [2]). Then $I = O(h)$, which proves (3.43) in one space dimension.□

• Set

$$J := \int [(\rho u)\psi_t + (\rho u^2)\psi_x + p\psi_x]dxdt. \tag{3.50}$$

Since $\rho u, \rho u^2, p, \rho$ are L^1 stable the proof in chapter 1, (formula (1.37) with $\omega = \rho u$ and in presence of p), gives, as (3.47),

$$J = -h\sum_{i,n}[(\rho u)_i^{n+1} - (\rho u)_i^n + r((\rho u^2)_i^n - (\rho u^2)_{i-1}^n) + r(p_i^n - p_{i-1}^n)]\psi_i^n + O(h). \tag{3.51}$$

Developping $(\rho u)_i^{n+1}$ from (3.17)-(3.19) one obtains the decomposition $J = J_1 + J_2 + J_3 + O(h)$ where

$$J_1 = -h\sum_{i,n}[\overline{(\rho u)}_i - (\rho u)_i^n + r((\rho u^2)_i^n - (\rho u^2)_{i-1}^n)]\psi_i^n = O(h) \tag{3.52}$$

as I_1, from the end of proof of Theorem 1.8.1,

$$J_2 = -h\alpha\sum_{i,n}(\overline{\rho u}_{i-1} - 2\overline{\rho u}_i + \overline{\rho u}_{i+1})\psi_i^n = -h\alpha\sum_{i,n}\overline{\rho u}_i(\psi_{i+1}^n - 2\psi_i^n + \psi_{i-1}^n) = O(h) \tag{3.53}$$

as (3.49),

$$J_3 = \frac{rh}{2}\sum_{i,n}(p_{i+1}^{n+1} - p_{i-1}^{n+1} - 2(p_i^n - p_{i-1}^n))\psi_i^n = \frac{rh}{2}\sum_{i,n}p_i^n(\psi_{i-1}^{n-1} - \psi_{i+1}^{n-1} - 2\psi_i^n + 2\psi_{i+1}^n) = O(h) \tag{3.54}$$

from Taylor's formula in ψ and the L^1 stability in p (from (3.14) the L^1 stability in p follows from the L^1 stability in $\overline{\rho e}$, implied from (3.13) by the L^1 stability in ρe, and from the bound $|\frac{\overline{\rho u}_i}{\overline{\rho}_i}| \leq max_i|u_i^n|$, from (3.11), (3.12) using the proof of (1.32). Therefore $J = O(h)$. This proves (3.44) in one space dimension.□

• Assuming ρ null at infinity, the formula $\sum_i \rho_i^n e_i^n h = \sum_i \rho_i^0 e_i^0 h$ holds from (3.21), (3.18), (3.13). Then the assumed positiveness of e, (3.42), implies the L^1 stability in ρe. The energy equation is treated similarly as J since pu is L^1 stable.

• Concerning the state law, let

$$K := \int [p - (\gamma - 1)(\rho e - \frac{(\rho u).u}{2})]\psi dxdt. \tag{3.55}$$

Since $K = \sum_{i,n}\{p_i^n - (\gamma-1)[(\rho e)_{i,n} - \frac{(\rho u)_i^n u_i^n}{2}]\int_{cell(i,n)}\psi dxdt\}$, Taylor's formula in ψ and the L^1 stability in $p, \rho e, \rho u^2$ imply $K = rh^2\sum_{i,n}\{p_i^n - (\gamma-1)[(\rho e)_i^n - \frac{(\rho u)_i^n u_i^n}{2}]\}\psi_i^n + O(h) = rh^2\sum_{i,n}\{p_i^{n+1} - (\gamma-1)[(\rho e)_i^n - \frac{(\rho u)_i^n u_i^n}{2}]\}\psi_i^n + O(h)$ (the change of upper index in p enters into $O(h)$ as after (47)). From (14), (20)

$$K = rh^2(\gamma-1)\sum_{i,n}[(\overline{\rho e}_i - (\rho e)_i^n) - \frac{1}{2}(\frac{(\overline{(\rho u)}_i)^2}{\overline{\rho}_i} - \frac{((\rho u)_i^n)^2}{\rho_i^n})]\psi_i^n + O(h). \tag{3.56}$$

If ρ, e, u are continuously differentiable functions on the support of ψ, except possibly on a finite number of curves in the (x,t) space, in which they have limits on both sides (shock waves, contact discontinuities, ...) and if ρ is strictly positive, then from (3.11)-(3.13) the quantity $[...]$ in (3.56) tends to 0 "almost everywhere" on the support of ψ, therefore $K \to 0$ when $h \to 0$. □

Proofs of the theorem in two and three space dimensions. They are practically identical to the proof in the one dimensional case except the proofs of lemmas 3.5.1, 3.5.2 below, which are given in section 1.11.

• Set

$$I := \int(\rho\psi_t + \rho u\psi_x + \rho v\psi_y)dxdydt. \tag{3.57}$$

Since ρ, ρu and ρv are L^1-stable, an immediate 2-D extension of the one dimensional proof of formula (1.37), gives the 2-D analog of (3.47) :

$$I = -h^2 \sum_{i,j,n} [\rho_{i,j}^{n+1} - \rho_{i,j}^n + r((\rho u)_{i,j}^n - (\rho u)_{i-1,j}^n) + r((\rho v)_{i,j}^n - (\rho v)_{i,j-1}^n)]\psi_{i,j}^n + O(h). \quad (3.58)$$

From (34), $\rho_{i,j}^{n+1} = \overline{\rho}_{i,j} + \alpha(2\overline{\rho}_{i-1,j-1} + \ldots + 3\overline{\rho}_{i-1,j} + \ldots - 20\overline{\rho}_{i,j})$. Therefore $I = I_1 + I_2 + O(h)$, where

$$I_1 := -h^2 \sum_{i,j,n} [\overline{\rho}_{i,j} - \rho_{i,j}^n + r((\rho u)_{i,j}^n - (\rho u)_{i-1,j}^n) + r((\rho v)_{i,j}^n - (\rho v)_{i,j-1}^n)]\psi_{i,j}^n, \quad (3.59)$$

$$I_2 := -h^2\alpha \sum_{i,j,n} [2\overline{\rho}_{i-1,j-1} + \ldots + 3\overline{\rho}_{i-1,j} + \ldots - 20\overline{\rho}_{i,j}]\psi_{i,j}^n = -h^2\alpha \sum_{i,j,n} \overline{\rho}_{i,j}(2\psi_{i+1,j+1}^n +$$

$$2\psi_{i+1,j-1}^n + 2\psi_{i-1,j+1}^n + 2\psi_{i-1,j-1}^n + 3\psi_{i+1,j}^n + 3\psi_{i-1,j}^n + 3\psi_{i,j+1}^n + 3\psi_{i,j-1}^n - 20\psi_{i,j}^n) = O(h) \quad (3.60)$$

from Taylor's formula in ψ and the L^1 stability of $\overline{\rho}$.

lemma 3.5.1. $I_1 = O(h)$.

The lemma is proved in section 1.11.

Therefore $I = O(h)$, which proves (3.43) in two space dimension.□

- Set

$$J := \int [(\rho u)\psi_t + (\rho u^2)\psi_x + (\rho uv)\psi_y + p\psi_x]dxdydt. \quad (3.61)$$

Since $\rho u, \rho u^2, \rho uv, p$ are L^1 stable one can prove as formula (1.37), see (3.58), that

$$J = -h^2 \sum_{i,j,n} [(\rho u)_{i,j}^{n+1} - (\rho u)_{i,j}^n + r((\rho u^2)_{i,j}^n -$$

$$(\rho u^2)_{i-1,j}^n) + r((\rho uv)_{i,j}^n - (\rho uv)_{i,j-1}^n) + r(p_{i,j}^n - p_{i-1,j}^n)]\psi_{i,j}^n + O(h). \quad (3.62)$$

Therefore, from (3.35), (3.36), $J = J_1 + J_2 + J_3 + O(h)$ where

$$J_1 := -h^2 \sum_{i,j,n} [\overline{(\rho u)}_{i,j} - (\rho u)_{i,j}^n + r((\rho u^2)_{i,j}^n - (\rho u^2)_{i-1,j}^n) + r((\rho uv)_{i,j}^n - (\rho uv)_{i,j-1}^n)]\psi_{i,j}^n, \quad (3.63)$$

$$J_2 := -h^2\alpha \sum_{i,j,n} (2(\overline{\rho u})_{i-1,j-1} + \ldots + 3(\overline{\rho u})_{i-1,j} + \ldots - 20(\overline{\rho u})_{i,j})\psi_{i,j}^n = O(h), \quad (3.64)$$

as (3.60),

$$J_3 := \frac{rh^2}{2} \sum_{i,j,n} [p_{i+1,j}^{n+1} - p_{i-1,j}^{n+1} - 2(p_{i,j}^n - p_{i-1,j}^n)]\psi_{i,j}^n = O(h), \quad (3.65)$$

as (3.54).

lemma 3.5.2. $J_1 = O(h)$.

The lemma is proved in section 1.11.

Therefore $J = O(h)$.□

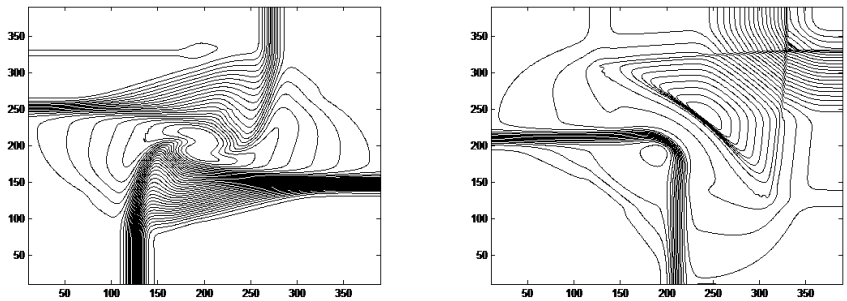
• The proofs for the energy equation and for the state law are similar to those in 1-D. □

• In the three dimensional case the scheme and proofs are similar. The difficulty in the proofs lies in the 3-D extension of Lemma 3.5.1 and Lemma 3.5.2. This extension is done in section 1.11.□

3.6 2-D Riemann problems in gas dynamics.

In figures 3.6.1, 3.6.2 we give the numerical results from the scheme in this paper on the six 2-D Riemann problems considered by P. D. Lax in [25], [26], and calculated in [17], [28] and [34]. The initial data are not recalled for brevity; they can be found in [28], configurations 6, 7, 13, 12, 19, 11, from top-left to bottom-right. The respective values of T are 0.25, 0.25, 0.30, 0.25, 0.30, 0.365 as in [25] and [26].

In figure 3.6.1 one has chosen $h = \frac{1}{400}$ as in [28]. The computation times are extremely short at the price of a poor quality : they range from 1.5 minutes to 3 minutes on a standard PC with an average of 2 minutes. One recognizes the results in [17], [25], [26], [28] and [34] even if they are far less accurate. The respective values of r are 0.65, 0.65, 0.575, 0.425, 0.5, 0.375, and the respective values of α are 0.02, 0.02, 0.02, 0.03, 0.02, 0.02. The convergence result asserts that these figures represent an approximate solution of these 2-D Riemann problems in the sense of distributions. This shows that very short calculations on a standard PC with the convergent scheme in this chapter can put faith on numerical results obtained by far more elaborate schemes from computational fluid dynamics.



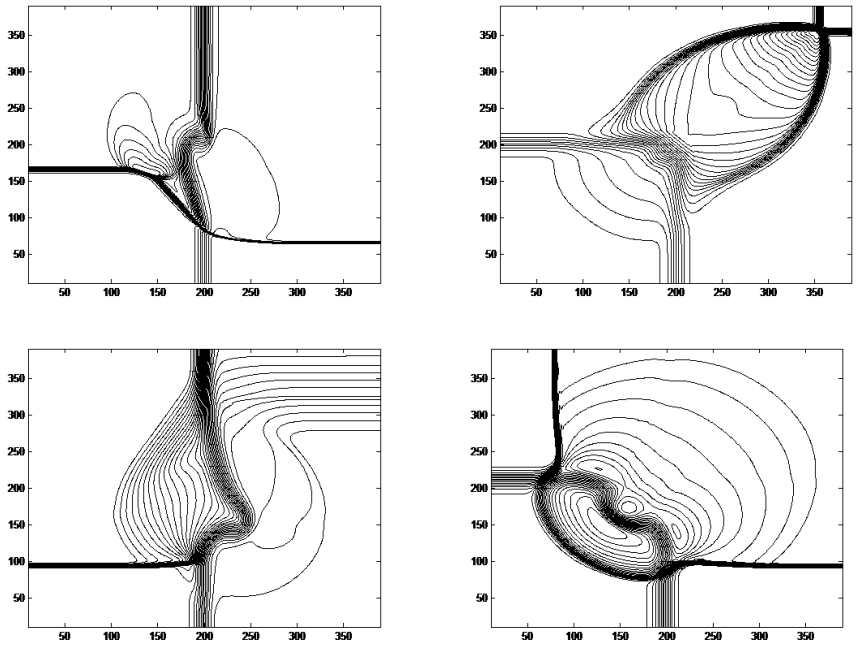
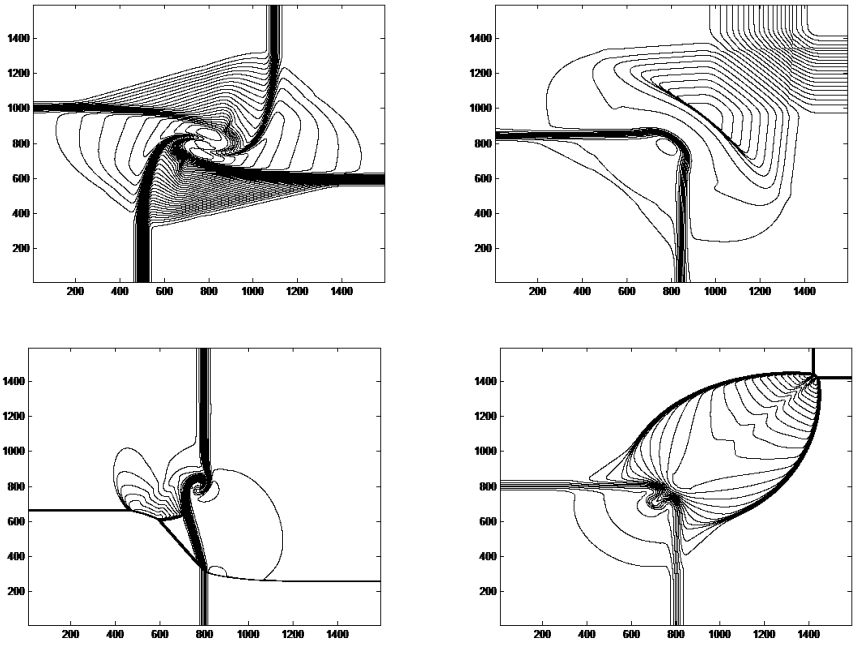


Figure 3.6.1. The numerical simulations of the 2-D Riemann problems after 2 minutes calculations on a standard PC.



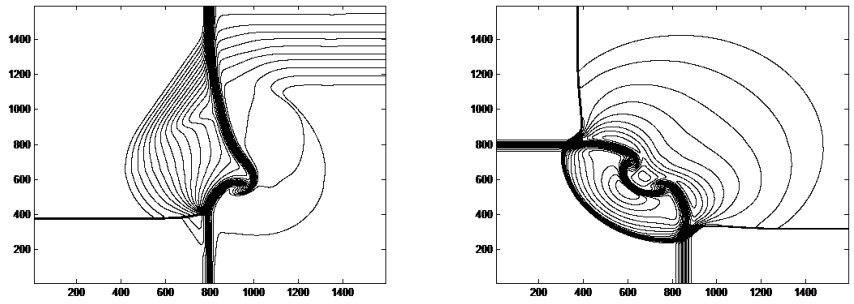


Figure 3.6.2. More precise calculations from a smaller value of the space step.

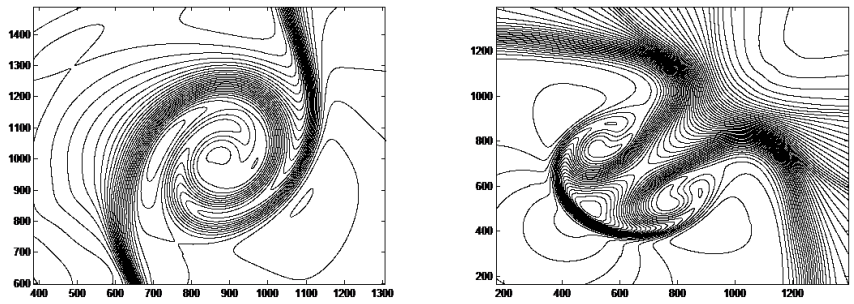


Figure 3.6.3. Two details from long calculations with the same PC.

In figure 3.6.2 one has chosen $h = \frac{1}{1600}$, with a duration time from 1.5 to 3 hours for each problem on a standard PC. The respective values of r are $0.625, 0.550, 0.550, 0.400, 0.475, 0.350$ and the respective values of α are 0.02, 0.03, 0.03, 0.02, 0.02. One observes numerous details from [17], [25], [26], [28] and [34], where the discontinuities are much thinner. In figure 3.6.3 two details from the center of the two middle-panels are computed : since they are located in the center of the window one has increased the number of iterations untill the details appear as large as possible. One observes details that can be guessed from [17], [25], [26], [28] and [34]. The consistence proof in this chapter ensures that these details approximate a solution of the equations, which gives a strong presumption that the schemes in [17], [28] and [34] could be convergent. One checks again that the scheme in this chapter can give arbitrary precision provided h small enough.

3.7 Conclusion.

In this chapter we have presented a simple numerical scheme for the system of ideal gases in 3-D. We have studied its consistence in the sense of distributions and shown, under the numerical assumptions of boundedness of the velocity field (so as to satisfy the CFL condition (3.41)) and positiveness of the energy, that the numerical solution from the scheme tends to satisfy the equations in the sense of distributions. In concrete terms, for any given bounded family of test functions (concerning supports and values of first and second derivatives), the proof provides a bound of order one in the space step for the left members of the equations, as long as these numerical assumptions are verified. Accuracy and low-cost efficiency have been checked numerically from the 1-D Woodward-Colella and Toro tests in [43] and [40], as well as from the 2-D

Riemann problems in [25] and [26]. All calculations have been done on a standard PC. Since it is immediate to check the CFL condition (3.41) and positiveness of the energy, the simplicity and efficiency of the scheme in several space dimension could make it useful in scientific computing where one is often confronted with the problem of confidence in the validity of numerical calculations. Indeed comparisons with the numerical solutions of the 2-D Riemann problems from the schemes presented in [17], [28] and [34] show that we have obtained again the same figures up to the smallest details, which could contribute to be confident in far more efficient schemes from computational fluid dynamics for which consistence proofs are lacking. Our consistency study suggests that the schemes in [17], [28] and [34] could actually be convergent in some suitable weak sense, as this will be considered mathematically in chapter 5 where a suitable functional space will be introduced for this purpose.

Deuxième partie

Weak limits of the approximate solutions as boundary values of holomorphic functions.

Chapitre 4

Introduction of the holomorphic tool

For some nonlinear equations of hydrodynamics used in cosmology to model radiation dominated universes we propose a method which permits transformations of the equations and calculations of discontinuous solutions. These formulas permit to select numerical schemes for these equations. As an application, we present a numerical simulation for the coupled system modeling evolution of densities of a mixture of a Newtonian fluid and a relativistic fluid.

4.1 Introduction.

Nonlinear calculations are usually unavoidable in derivation of the equations from physical postulates. In case of nonsmooth solutions, "formal" nonlinear calculations on equations of fluid dynamics can lead to wrong results : indeed these calculations can strongly modify the nonsmooth solutions. Therefore it is important to know the calculations that are permitted and those that are forbidden. This chapter focusses on two systems modeling radiation dominated universes, [8] p. 221, [30] pp. 35-38 and p. 465, when the linear regime breaks down. In particular we study discontinuous solutions of these equations in the fully nonlinear regime in order to obtain explicit formulas for the jump conditions.

The linearized equations of motion provide an excellent description of gravitational instability when density fluctuations are small. But the linear regime breaks down as soon as the density fluctuations cease to be small, which makes perturbation theory no longer valid. Therefore it is indispensible to solve the equations in the fully nonlinear regime [8] pp. 304-332, [30] pp. 482-493. To this end, in case of discontinuous solutions, we propose a method of calculation that consists in the introduction of a "small" parameter to regularize the problem so as to permit calculations. After the calculations the regularization is removed by letting the parameter tend to 0. This method uses (implicitely or explicitely) functions of complex variables to perform explicit calculations and obtain solutions.

In [8] p. 221 the motion of a relativistic fluid in cosmology is modelled by the system (continuity equation, Euler equation, Poisson equation)

$$\frac{\partial \rho}{\partial t} + \vec{\nabla}.((\rho + \frac{p}{c^2})\vec{v}) = 0, \tag{4.1}$$

$$(\rho + \frac{p}{c^2})(\frac{\partial \vec{v}}{\partial t} + (\vec{v}.\vec{\nabla})\vec{v}) + \vec{\nabla}p + (\rho + \frac{p}{c^2})\vec{\nabla}\Phi = \vec{0}, \tag{4.2}$$

$$\Delta\Phi = 4\pi G(\rho + 3\frac{p}{c^2}), \tag{4.3}$$

where c is the velocity of light, ρ the energy density, $\vec{v}$ the velocity vector, p the pressure and Φ the gravitation potential. These equations are completed by a state law of the form $p = \mathcal{P}(\rho)$ where $\mathcal{P}$ is a function. A usual equation of state is

$$p = K\rho c^2, \tag{4.4}$$

where K is a constant value, [8] p. 222, with $K = \frac{1}{3}$ in the case of a radiation dominated fluid [8] p. 221, [30] p. 37, p. 465. A more complete system is given in [30] p. 465 (7 lines after formula 15.25 to take into account the omitted term in formula 15.24) : equation (4.2) is replaced by

$$\frac{\partial\vec{v}}{\partial t} + (\vec{v}.\vec{\nabla})\vec{v} + \frac{\vec{\nabla}p + \frac{\partial p}{\partial t}\vec{v}}{\rho + \frac{p}{c^2}} + \vec{\nabla}\Phi = \vec{0}, \tag{4.5}$$

which differs from (4.2) by division by $\rho + \frac{p}{c^2}$ and the presence of the supplementary term $\frac{\partial p}{\partial t}\vec{v}$. This equation is a simplification of the equations in [30] p. 36, [42] p. 49.

Since the fields are considered as relatively weak, there is no need to use general relativity : these equations of special-relativistic hydrodynamics are formally derived from special-relativity fluid mechanics and Newtonian gravity with a relativistic source term, see [42] pp. 47-51, [30] pp. 18-25, pp. 35-37, pp 464-465. They can be considered as issued from the general expression of the energy-momentum tensor of a perfect fluid [30] p. 19 at a limit for small velocities and weak fields. Equation 4.1 is a generalization of the conservation of energy. Equations 4.2 and 4.5 are relativistic generalizations of Euler's equation for momentum conservation in fluid dynamics. Since one cannot assume $p << \rho c^2$, gravitation is modelled by equation (4.3), see [8] p. 221, [30] pp. 24-25. pp. 35-37, pp. 50-51.

We show in section 4.2 that formal calculations on discontinuous solutions of system (4.1)-(4.4) lead to inconsistencies. Further, the first term in (4.2) and the third term in (4.5) do not make sense in case of discontinuous solutions since they appear in form of a product of a discontinuous function and a Dirac delta function. This last fact is at the origin of specific trouble in numerical schemes since, for these nonconservative equations, one does not have a priori well defined Rankine-Hugoniot jump conditions.

In the third section of this chapter, one states precisely a mathematical context for this method, so as to use it in the study of Cauchy problems for these equations of special-relativistic fluid dynamics. Then, one calculates explicit solutions for these two systems in the case of a solution made of two constant states separated by a discontinuity. Existence of solutions from this method is shown below from explicit calculations in physically significant cases. A numerical scheme is presented and tested in section 6 relatively to the explicit jump conditions obtained in sections 4.4 and 4.5.

4.2 Inconsistencies from formal calculations.

Formal calculations consist in using the classical rules of mathematical calculations (valid on smooth functions) even on nonsmooth functions without a mathematical justification of the

validity of these calculations. In this section we show that formal calculations on system (4.1)-(4.4) lead to inconsistencies, i.e. contradictory results. In one space dimension and absence of gravitation, immediate formal calculations transform system (4.1), (4.2), (4.4) into

$$\rho_t + (K+1)(\rho u)_x = 0, \tag{4.6}$$

$$u_t + [\frac{1}{2}u^2 + \frac{Kc^2}{1+K}\log\rho]_x = 0. \tag{4.7}$$

Setting

$$q := (1+K)\frac{u^2}{2} - \frac{Kc^2}{1+K}\log\rho, \tag{4.8}$$

system (4.6)-(4.7) is transformed into

$$u_t + (\frac{K+2}{2}u^2)_x = q_x, \tag{4.9}$$

$$q_t + [(1+K)\frac{u^3}{3}]_x = Kc^2 u_x. \tag{4.10}$$

The proof is a mere formal verification from formulas (4.6)-(4.8) : insert (4.8) into (4.9)-(4.10) and use (4.6)-(4.7). We seek shock wave solutions in the form of two constant states separated by a discontinuity moving with constant speed denoted V. According to the usual formula $V = \frac{\Delta(f(u))}{\Delta u}$ that gives the velocity of shock waves of the equation $u_t + f(u)_x = 0$, the jump conditions of the conservative equations (4.9)-(4.10) are

$$V = \frac{\frac{K+2}{2}\Delta(u^2) - \Delta q}{\Delta u}, \quad V = \frac{(K+1)\Delta(\frac{u^3}{3}) - Kc^2\Delta u}{\Delta q}.$$

Elimination of Δq gives that the velocity V of the shock wave is solution of the second degree equation

$$V^2 - \{\frac{2+K}{2}(u_r+u_l)\}V + \{\frac{1+K}{3}(u_r^2 + u_r u_l + u_l^2) - Kc^2\} = 0. \tag{4.11}$$

An algebraic inconsistency is put in evidence as follows. Formula (4.23) below has been calculated inside the proof of Theorem 4.1.1. This formula follows from suitably chosen formal calculations on physical ground, for which it has been checked in the proof of Theorem 4.1.1 that this choice ensures existence of shock wave solutions with well defined jump conditions. Formula (4.23) and the state law (4.4) imply $\frac{\rho_r}{\rho_l} = 1 + \frac{1+K}{K}\frac{V-u_l}{V}(\exp\frac{V(u_r-u_l)}{c^2} - 1)$. Since it is in conservative form, equation (4.6) gives the classical jump condition $V = (1+K)\frac{\Delta(\rho u)}{\Delta\rho}$, which gives $\frac{\rho_r}{\rho_l} = \frac{V-(1+K)u_l}{V-(1+K)u_r}$. Finally, we obtain

$$\frac{V-(1+K)u_l}{V-(1+K)u_r} = 1 + \frac{1+K}{K}\frac{V-u_l}{V}(\exp\frac{V(u_r-u_l)}{c^2} - 1). \tag{4.12}$$

Given u_r, u_l, K, c^2, one can compute two values V_1, V_2 from (4.11). Then insertion of these values into both members of (4.12) shows that (4.12) does not hold in general (take for instance $K = c = u_r = 1, u_l = 0$). This is an algebraic contradiction.

How to avoid the inconsistencies ? Let us consider the way these equations are obtained. Equations (4.1)-(4.2) are issued from special-relativistic hydrodynamics since the fields are still

weak. They are an extension of the classical laws of mass and momentum conservation. They have already been subject to formal nonlinear calculations [42] pp.47-49, [30] pp.35-36. The state law (4.4) is directly obtained from a physical reasoning or observation. One should be allowed to perform nonlinear calculations on the equations (4.1)-(4.2) and (4.1), (4.5), since nonlinear calculations have already been done to obtain them, but not necessarily on the state law (4.4). Since the inconsistencies in section 4.2 are obtained from formal calculations involving both (4.1), (4.2) and (4.4), does the idea to calculate freely on (4.1)-(4.2) or (4.1), (4.5) only permit to avoid inconsistencies in presence of shock waves calculations for systems (4.1)-(4.4) and (4.1), (4.5), (4.3), (4.4) ?

Another problem under concern here is the presence of the product of a discontinuous function and a Dirac delta function in the first term of equation 4.2 as well as in the third term of equation 4.5. Such a product does not make sense classically. To remedy for these problems (the inconsistencies and the above undefined products) we introduce a method of regularization directly inspired from classical calculations of physics and mathematics, using a small regularizing parameter, that will permit to give a positive answer : in the space of the regularized objects one can compute freely on (4.1)-(4.2) and (4.1), (4.5) concerning shock waves, and (4.4) (on which nonlinear calculations are forbidden to avoid inconsistencies) gives a needed supplementary piece of information. These explicit calculations permit to put in evidence numerical schemes in agreement with the jump conditions obtained from them.

4.3 Mathematical context.

It is usual in physics and mathematics to regularize an irregular function f, denoted here $f(x), x \in \mathbb{R}^n$, by introducing a small parameter $\epsilon > 0$, so as to replace the irregular function $f(x)$ by smooth functions $f_\epsilon(x)$ denoted here $f(x,\epsilon)$, such that $f(x,\epsilon)$ tend weakly to $f(x)$ when $\epsilon \to 0$.

The method we use consists in transfering the physical problem under consideration to a larger space made of the regularized objects $f(x,\epsilon)$. One considers in this new space functions that play the role of representatives of the Heaviside function and of the Dirac delta function. In order to benefit from the property of uniqueness of analytic continuation, so as to identify a function and its restriction to a smaller strip, the functions $f(x,\epsilon)$ are analytic functions in the variables x and ϵ, which amounts to consider holomorphic functions of complex variables $f(z,\zeta), z = x + iy, \zeta = \epsilon + i\eta, x, y \in \mathbb{R}^n, \epsilon, \eta \in \mathbb{R}$, defined in complex neighborhoods of the real space. Convenient neighborhoods are defined as follows.

The letters r, θ, μ will always denote real numbers such that

$$0 < r < 1,\ 0 < \theta < \frac{\pi}{6},\ 0 < \mu < \frac{1}{2}. \tag{4.13}$$

The values r, θ, μ can be as small as needed. One considers the open strip in $\mathbb{R}^{2n+2}$ parallel to the real space $\mathbb{R}^n$ of variable x defined by

$S(r,\theta,\mu) = \{(z,\zeta) \in \mathbb{C}^n \times \mathbb{C}$ such that

$$x \in \mathbb{R}^n,\ 0 < |\zeta| < r,\ -\theta < arg\zeta < \theta, |y_i| < \mu\epsilon\ \forall i = 1, ..., n\}. \tag{4.14}$$

The real space $\mathbb{R}^n$ lies on the boundary of $S(r,\theta,\mu)$ by letting ζ tend to 0 (and therefore, since

$\epsilon < \frac{|\zeta|}{2}$ from (4.13), $y \to 0$). Let $\mathcal{F}$ be the set of all strips $S(r, \theta, \mu)$. The set $\mathcal{F}$ is a net for the inclusion :

$$\forall\ S_1, S_2 \in \mathcal{F}\ \exists S_3 \in \mathcal{F}\ /\ S_3 \subset S_1 \cap S_2.$$

We denote by *const* a positive real number which may not be the same from an expression to the following. If $S \in \mathcal{F}$, one defines

$$H_S := \{\text{holomorphic functions F}: \quad S \longmapsto \mathbb{C}, \quad (z, \zeta) \longmapsto F(z, \zeta)\}.$$

Note that $\frac{\partial}{\partial x_i} H_S \subset H_S$. If $S' \subset S$, with $S, S' \in \mathcal{F}$, the restriction map

$$H_S \longmapsto H_{S'},$$

$$F \longmapsto F|_{S'},$$

is injective from the uniqueness of analytic continuation since $F|_{S'} = 0 \Rightarrow F = 0$ in the connected strip S. For convenience we note $H_S \subset H_{S'}$. Now we identify a function and its analytic continuation.

Definition. In the reunion of the sets H_S one considers the equivalence relation

$$(F_1, S_1) \equiv (F_2, S_2)$$

$$\Leftrightarrow$$

$$\exists S_3 \subset S_1 \cap S_2 \quad / \quad F_1|_{S_3} = F_2|_{S_3}.$$

The set of all equivalence classes is by definition a space of germs of holomorphic functions on $\mathbb{R}^n$ in the x-variable. Since this space is also classically refered to as an inductive limit we denote it by $\lim H_S$.

These "germs" can also be refered to as "functions" provided one retains that a function and any of its analytic extensions on a connected open set are identified.

In other words this means that one considers the reunion of the sets H_S, and then that $F \in H_{S_1}, G \in H_{S_2}$ are identified iff there is $S_3 \in \mathcal{F}$ such that $S_3 \subset S_1 \cap S_2$ and $F|_{S_3} = G|_{S_3}$. This definition consists precisely in defining on the reunion of the sets $H_S, S \in \mathcal{F}$, the above equivalence relation. $LimH_S$ is stable by differentiation and multiplication

$$\frac{\partial}{\partial x_i}(LimH_S) \subset LimH_S \quad \forall i,$$

$$LimH_S \times LimH_S \subset LimH_S.$$

Now let us check that $LimH_S$ contains objects that we shall need in the sequel, more precisely Heaviside and Dirac functions. To this end notice that to any function $f \in L^\infty(\mathbb{R}^n)$, we can associate several elements $F \in LimH_S$ that "give back" f on $\mathbb{R}^n$ considered on the boundary of $S(r, \theta, \mu)$ as the following weak limit

$$\forall \psi \in \mathcal{C}_c^\infty(\mathbb{R}^n) \quad \lim_{\epsilon \to 0} \int F(x, \epsilon)\psi(x)dx = \int f(x)\psi(x)dx \tag{4.15}$$

where $\mathcal{C}_c^\infty(\mathbb{R}^n)$ is the space of infinitely differentiable functions on $\mathbb{R}^n$ with compact support. We say that f is the *real interpretation* of F. This can be done by convolution : set for instance the mollifier

$$\rho(z) = const\frac{1}{((z_1)^2+1)^s \dots ((z_n)^2+1)^s},\ s \in \mathbb{N},\ s \geq 1 \tag{4.16}$$

with $\int \rho(x)dx = 1$ in order that the function $\lambda \longmapsto \frac{1}{\epsilon^n}\rho(\frac{\lambda}{\epsilon})$ provides an approximation of the identity by convolution when $\epsilon \to 0$. Then, we set

$$F(z,\zeta) := \int_{\mathbb{R}^n} f(\lambda)\frac{1}{\zeta^n}\rho(\frac{\lambda - z}{\zeta})d\lambda. \tag{4.17}$$

Lemma 4.3.1. $\forall f \in L^\infty(\mathbb{R}^n)$ *the function* F *defined in (4.17) is in* $LimH_S$ *and has* f *as real interpretation. Further, if* f *is continuous at a point* x_0*, then* $F(x,\epsilon) \to f(x_0)$ *when* $\epsilon \to 0$ *and* $x \to x_0$.

Proof. For simplicity the proof is given in the case $n = 1$. Then

$$F(z,\zeta) = const.\zeta^{2s-1}\int \frac{f(\lambda)}{[(\lambda - z)^2 + \zeta^2]^s}d\lambda. \tag{4.18}$$

Auxiliary calculation : $|(\lambda - z)^2 + \zeta^2| \geq |Real((\lambda - z)^2 + \zeta^2)| = (\lambda - x)^2 + \epsilon^2 - y^2 - \eta^2 \geq \epsilon^2(1-\mu^2-tan^2\theta) > \frac{\epsilon^2}{2}$ from (4.13). Therefore the denominator in (4.18) does not take the value 0 when $(z,\zeta) \in S(r,\theta,\mu)$. In the sequel we will use that $|(\lambda-z)^2+\zeta^2| \geq (\lambda-x)^2+\alpha^2\epsilon^2$ with $\alpha = \frac{1}{\sqrt{2}}$.

If $f \in L^\infty$, (4.18) gives $|F(z,\zeta)| \leq const|\zeta|^{2s-1}\|f\|_\infty \int \frac{d\lambda}{[(\lambda-x)^2+\alpha^2\epsilon^2]^s} \leq const\|f\|_\infty$, since $\int \frac{d\lambda}{[(\lambda-x)^2+\alpha^2\epsilon^2]^s} = \int \frac{d\lambda}{[(\lambda)^2+\alpha^2\epsilon^2]^s} = (\alpha\epsilon)^{1-2s}\int \frac{d\mu}{(\mu^2+1)^s}$.

The last assertion follows from the formula $F(x,\epsilon) = \int f(x+k\epsilon)\rho(k)dk$.□

As a consequence, if f is the classical Heaviside function, when $\epsilon \to 0$ one has $F(x,\epsilon) \to 0$ if $x < 0$ and $F(x,\epsilon) \to 1$ if $x > 0$. Since the space $LimH_S$ is stable by differentiation, $\frac{\partial F}{\partial x}$ has the Dirac δ function as real interpretation.

These results can be easily extended to $\mathbb{R}^n \times]0,T[$, considering f null out of $\mathbb{R}^n \times]0,T[$. A Heaviside function in H is an element of $LimH_S$ whose real interpretation is the Heaviside function. A Dirac function is an element of $LimH_S$ whose real interpretation is the Dirac delta distribution.

Besides the concept of solution of equations in the sense of equality in the space $LimH_S$, we are forced to consider also solutions in a weak sense, for which a natural definition is as follows.

Definition of a concept of weak solution. The "function" $U = (U^j)_{j=1,\dots,m} \in (LimH_S)^m$ relative to $\mathbb{R}^{n+1}$ is a weak solution of the system

$$U_t + \sum_{i=1}^n A_i(U)\frac{\partial U}{\partial x_i}\ \text{“}=''\ 0 \tag{4.19}$$

of m scalar equations iff each component of $U_t + \sum_{i=1}^{n} A_i(U)\frac{\partial U}{\partial x_i}$ has the null function as real interpretation, i.e.

$$\forall j = 1, ..., m,\ \forall \psi \in C_c^\infty(\mathbb{R}^{n+1}),\quad \int_{\mathbb{R}^{n+1}} [(U^j)_t + \sum_{i=1}^{n}(A_i(U)\frac{\partial U}{\partial x_i})^j(x,t,\epsilon)]\psi(x,t)dxdt \to 0 \quad (4.20)$$

when $\epsilon \to 0^+$. This is denoted by $U_t + \sum_{i=1}^{n} A_i(U)\frac{\partial U}{\partial x_i} \overset{weak}{=} 0$.

As the usual concept of a weak solution this concept of weak solution suffers from nonuniqueness and classical examples show that free manipulation of equations can change the solution in the case of discontinuous solutions.

4.4 Calculation of a jump condition I.

In section 4.2 it was shown that formal calculations can be wrong in case of nonsmooth solutions. In this section we test in absence of gravitation the idea presented in section 4.2 in case of nonsmooth solutions and we explicit the jump formulas so obtained. We shall calculate on discontinuous solutions in one space dimension because this is simple and representative of the general situation in the case of shock waves. We recall that the equations stated with the strong equality in our context can be manipulated freely and that the weak equality in our context does not allow free manipulation of the equations in the case nonsmooth solutions are concerned.

The small parameter ζ is not apparent in the calculations : since they need the context of this chapter in order to make sense this small parameter is implicit. For the solutions under concern the equations are reduced to algebraic equations (4.22), (4.23), (4.25), (4.28) that can be satisfied at once, thus proving the existence of strong solutions of the first two equations in (4.21) from explicit calculations.

Theorem 4.4.1. *The system (4.1)-(4.4) of special-relativistic fluids, with G=0 in one space dimension admits step functions solutions when stated in the context of this chapter in the following form, where the symbol = in the first two equations means one has strong solutions while the third equation (state law) is satisfied only in the weak sense*

$$\rho_t + ((\rho + \frac{p}{c^2})u)_x = 0,\ (\rho + \frac{p}{c^2})(u_t + uu_x) + p_x = 0,\ p \overset{weak}{=} \mathcal{P}(\rho) \quad (4.21)$$

with $\mathcal{P}$ an algebraic function, (4.4) for instance.
The jump conditions are

$$V = \frac{\Delta(\rho u) + \frac{\Delta(pu)}{c^2}}{\Delta\rho}, \quad (4.22)$$

which is the classical jump condition of the conservative first equation in (4.21), and, further, the nonclassical jump condition

$$V\Delta p = c^2(\rho_l + \frac{p_l}{c^2})(V - u_l)(exp\frac{V\Delta u}{c^2} - 1) \quad (4.23)$$

which follows from the nonconservative second equation. As a consequence the second equation in (4.21) can equivalently be stated in the form $u_t + uu_x + \frac{p_x}{\rho + \frac{p}{\gamma^2}} = 0$ (these two formulations are

found in texts of cosmology).

The statement (4.21) is physically sound since the state law has a far weaker meaning than the equations of special relativity from which the first two equations in (4.21) follow : these equations correspond to conservation of mass and momentum in the Newtonian version, as relativistic extensions in the domain of weak fields [8] p. 221, [30] pp. 18-19, 24-25, 35-38, 50-51, 464-465.

Role of the state law. In the proof one considers solutions of the form

$$\omega(x,t) = \omega_l + (\omega_r - \omega_r)H_\omega(x - Vt), \tag{4.24}$$

$\omega = u, p, \rho$ and H_ω a Heaviside function depending on the physical variable ω. The role of the state law $p \overset{weak}{=} \mathcal{P}(\rho)$ is simply to state $p_l = \mathcal{P}(\rho_l)$ and $p_r = \mathcal{P}(\rho_r)$, without any information on the Heaviside functions of p and ρ involved in the jump. From the definition of $Y = \rho + \frac{p}{c^2}$ one has $\Delta Y H_Y = \Delta\rho H_\rho + \frac{\Delta p}{c^2}H_p$ and (4.25) gives the relation $\Delta\rho H_\rho = -\frac{\Delta p}{c^2}H_p + \frac{V\frac{\Delta p}{c^2}H_p + Y_l\Delta u H_u}{V - u_l - \Delta u H_u}$. Formula (4.28) gives H_p as a function of H_u and (4.25) gives $H_\rho$as a function of H_p, H_u. Therefore both H_ρ and H_p are well defined functions of H_u. The statement of the state law in the strong form would impose another relation between H_p and H_ρ, for instance $H_p = H_\rho$ in case of the state law $p = const.\rho$, thus giving a contradiction which is at the origin of the absurd result shown in section 4.2 from formal calculations.

proof. The proof consists in plugging (4.24) with $\omega = u, p, \rho$ and respective Heaviside functions H_u, H_p and H_ρ into the left members of the first two equations in (4.21), and seek under what conditions the results are null. One finds that this amounts to formulas between V$and the left - right values$u$_l, u_r, p_l, p_r, \rho_l, \rho_r$ (= the jump conditions (4.22)-(4.23)), plus explicit relations between H_u, H_p and H_ρ that amount to express two of them in function of the third one (4.25)-(4.28). We give the calculations in full detail although they are a reproduction in this context of elementary calculations. For convenience one sets $Y = \rho + \frac{p}{c^2}$ and $Y(x,t) = Y_l + \Delta Y H_Y(x - Vt)$. From the formulas $\rho = Y - \frac{p}{c^2}$ and $(\rho + \frac{p}{c^2})u = Y_l u_l + Y_l\Delta u H_u + u_l\Delta Y H_Y + \Delta u\Delta Y H_u H_Y$, the first equation gives

$$-V\Delta Y H_Y' + V\frac{\Delta p}{c^2}H_p' + Y_l\Delta u H_u' + u_l\Delta Y H_Y' + \Delta u\Delta Y(H_u H_Y)' = 0.$$

By a mere integration, using $H_u(-\infty) = H_Y(-\infty) = H_p(-\infty) = 0$ to fix the integration constant, one obtains the formula

$$H_Y = \frac{V\frac{\Delta p}{c^2}H_p + Y_l\Delta u H_u}{\Delta Y(V - u_l - \Delta u H_u)}. \tag{4.25}$$

Since $H_u(+\infty) = H_Y(+\infty) = H_p(+\infty) = 1$ one obtains the formula

$$V\frac{\Delta p}{c^2} + Y_l\Delta u = \Delta Y(V - u_l - \Delta u)$$

from which easy calculations give the jump condition (4.22) (which classically follows from the first equation in (4.21) which is in conservative form : this is the reason why a mere integration has so easily given the result, as done classically to obtain jump conditions for systems in conservative form).

The second equation in (4.21) is in nonconservative form. It will be more difficult to obtain the jump condition. Plugging (4.24) into it with $\omega = Y, u, p$ gives

$$(Y_l + \Delta Y H_Y)(-V\Delta u H'_u + u_l \Delta u H'_u + (\Delta u)^2 H_u H'_u + \Delta p H'_p = 0$$

i.e.

$$\Delta p H'_p = \Delta u H'_u (Y_l + \Delta Y H_Y)(V - u_l - \Delta u H_u). \tag{4.26}$$

Note that H'_u is a Dirac delta function and H_Y a Heaviside function, therefore one observes classically undefined products $H'_u H_Y, H'_u H_Y H_u$ which make sense here as elements of $Lim H_S$. From (4.25) one obtains

$$Y_l + \Delta Y H_Y = Y_l + \frac{V\frac{\Delta p}{c^2} H_p + Y_l \Delta u H_u}{V - u_l - \Delta u H_u} = \frac{Y_l(V - u_l) + V\frac{\Delta p}{c^2} H_p}{V - u_l - \Delta u H_u}. \tag{4.27}$$

Therefore, from (4.26)

$$\Delta p H'_p = \Delta u H'_u (Y_l(V - u_l) + V\frac{\Delta p}{c^2} H_p)$$

that can be written in the form

$$H'_p - (V\frac{\Delta u}{c^2} H'_u) H_p - \frac{\Delta u}{\Delta p} Y_l (V - u_l) H'_u = 0.$$

Explicit integration is done by considering that this is an ODE in the unknown function H_p, following the classical method for the linear ODEs $a(x)y' + b(x)y + c(x) = 0$. It makes sense since the coefficients a, b, c are classical functions defined in some strip $S(r, \theta, \mu)$ (one chooses Heaviside functions defined and bounded in this strip like those exposed in section 4.3).

First step : homogeneous equation.

$H'_p = (V\frac{\Delta u}{c^2} H'_u) H_p$, which implies $H_p = const. \exp(V\frac{\Delta u}{c^2} H_u)$.

Second step : variation of the constant. The full equation becomes

$const'. \exp(V\frac{\Delta u}{c^2} H_u) + const. V\frac{\Delta u}{c^2} H'_u \exp(V\frac{\Delta u}{c^2} H_u) - V\frac{\Delta u}{c^2} H'_u const. \exp(V\frac{\Delta u}{c^2} H_u) = \frac{\Delta u}{\Delta p} Y_l (V - u_l) H'_u$.

Therefore

$const' = \frac{\Delta u}{\Delta p} Y_l (V - u_l) \exp(-V\frac{\Delta u}{c^2} H_u) H'_u$, i.e.

$const = \frac{\Delta u}{\Delta p} Y_l (V - u_l) \frac{-c^2}{V \Delta u} \exp(-V\frac{\Delta u}{c^2} H_u) + other\ const.$

Finally the formula for the solutions of the ODE is

$$H_p = \frac{1}{\Delta p} Y_l (V - u_l) \frac{-c^2}{V} + const. \exp(V\frac{\Delta u}{c^2} H_u).$$

Using $H_u(-\infty) = H_p(-\infty) = 0$ to fix the integration constant, one obtains

$$H_p = -Y_l(V-u_l)\frac{c^2}{V\Delta p}(1-exp(V\frac{\Delta u}{c^2}H_u)). \tag{4.28}$$

The nonclassical jump condition (4.23) follows from setting $H_p(+\infty) = 1 = H_u(+\infty)$ as boundary conditions. Formulas (4.28), (4.25) amount to state H_Y and H_p as functions of H_u and (4.22)-(4.23) ensure that H_Y and H_p from these formulas are Heaviside functions provided H_u is. Therefore these formulas (4.22), (4.23), (4.25), (4.28) are equivalent to the existence of a strong solution in the requested form (4.24) of a shock wave. □

Comments. Theorem 4.4.1 amounts to a choice of "formal" calculations that are proved to be permitted even in case of step function solutions (those on expressions stated with = in (4.21)) and "formal" calculations that are (unless exception such as linear calculations) forbidden (those stated with " $\overset{weak}{=}$ "), such as the third equation in (4.21).

Consistency with Newtonian mechanics. At the limit $c \to +\infty$ one obtains easily from (4.23) the jump condition in the Newtonian case : indeed (4.23) gives $\Delta p = \rho_l(V-u_l)\Delta u$. Inserting $V = \frac{\Delta(\rho u)}{\Delta\rho}$ (from (4.22) with $c \to +\infty$) one obtains the formula

$$\Delta p \Delta\rho = \rho_r\rho_l(\Delta u)^2. \tag{4.29}$$

The Newtonian system classically stated (weak classical solutions)

$$\rho_t + (\rho u)_x = 0,\ (\rho u)_t + (\rho u^2)_x + p_x = 0 \tag{4.30}$$

has the classical jump conditions $V = \frac{\Delta(\rho u)}{\Delta\rho}$ and $V = \frac{\Delta(\rho u^2)+\Delta p}{\Delta(\rho u)}$. Elimination of V gives (4.29).□

4.5 Calculation of a jump condition II.

Theorem 4.5.1. *The system (4.1), (4.5), (4.3), (4.4) of special-relativistic fluids in one space dimension in absence of gravitation admits step function solutions when stated in the context of this chapter as*

$$\rho_t + ((\rho + \frac{p}{c^2})u)_x = 0,\ u_t + uu_x + \frac{p_x + up_t}{\rho + \frac{p}{c^2}} = 0,\ p \overset{weak}{=} \mathcal{P}(\rho). \tag{4.31}$$

As a consequence, formal nonlinear calculations on the first two equations are justified. The jump condition for the second equation is

$$(\rho_r + \frac{p_r}{c^2})^{c^2}(V-u_r)^{c^2}(1-Vu_r) = (\rho_l + \frac{p_l}{c^2})^{c^2}(V-u_l)^{c^2}(1-Vu_l). \tag{4.32}$$

proof. The calculations given below are similar to those in the proof of Theorem 4.4.1. The first equation has been studied in the proof of Theorem 4.4.1. It gives (4.25) for H_Y and the jump condition (4.22). With the above notations the second equation in (4.31) can be stated

$$Y(u_t + uu_x) + p_x + up_t = 0, \tag{4.33}$$

Inserting (4.24) with $\omega = Y, u, p$ into equation (4.33) gives

$$(Y_l + \Delta Y H_Y)[(-V + u_l)\Delta u H'_u + (\Delta u)^2 H_u H'_u] + \Delta p H'_p + (u_l + \Delta u H_u)(-V \Delta p H'_p) = 0,$$

i.e.

$$\Delta p H'_p (1 - V(u_l + \Delta u H_u)) = (Y_l + \Delta Y H_Y)(V - u_l - \Delta u H_u)\Delta u H'_u.$$

From (4.27) (which follows from (4.25) i.e. from the first equation in (4.31))

$$\Delta p H'_p (1 - V(u_l + \Delta u H_u)) = (Y_l (V - u_l)\Delta u H'_u + V \tfrac{\Delta p}{c^2} H_p)\Delta u H'_u.$$

Then, the differential equation satisfied by the Heaviside function H_p is :

$$\Delta p (1 - V(u_l + \Delta u H_u)) H'_p - V \tfrac{\Delta p}{c^2} \Delta u H'_u H_p - Y_l (V - u_l)\Delta u H'_u = 0.$$

First step in solution of this ODE. Homogeneous equation

$$(1 - V(u_l + \Delta u H_u)) H'_p = \tfrac{V}{c^2} \Delta u H'_u H_p,$$

whose solution is

$$H_p = const(-V \Delta u H_u + (1 - V u_l))^{-\frac{1}{c^2}}.$$

Second step : Variation of the constant. One finds

$$\Delta p (1 - V u_l - V \Delta u H_u)^{1 - \frac{1}{c^2}} const' = Y_l (V - u_l)\Delta u H'_u,$$

$const' = \frac{Y_l (V - u_l)\Delta u H'_u}{\Delta p (1 - V u_l - V \Delta u H_u)^{1 - \frac{1}{c^2}}}$, i.e. by integration

$$const = \tfrac{-Y_l (V - u_l)}{V \Delta p} c^2 (-V \Delta u H_u + 1 - V u_l)^{\frac{1}{c^2}} + const.$$

Finally one finds the solution

$$H_p = -\tfrac{Y_l (V - u_l) c^2}{V \Delta p} + const.(-V \Delta u H_u + 1 - V u_l)^{-\frac{1}{c^2}}.$$

Using $H_u(-\infty) = H_p(-\infty) = 0$ to fix the integration constant, one obtains

$const = \frac{Y_l (V - u_l) c^2}{V \Delta p} (1 - V u_l)^{\frac{1}{c^2}}$. The solution is

$$H_p = -\tfrac{Y_l (V - u_l) c^2}{V \Delta p} + \tfrac{Y_l (V - u_l) c^2}{V \Delta p} (\tfrac{1 - V u_l}{-V \Delta u H_u + 1 - V u_l})^{\frac{1}{c^2}}.$$

Setting $H_p(+\infty) = 1 = H_u(+\infty)$ gives the jump condition

$$\tfrac{V \Delta p}{Y_l (V - u_l) c^2} = -1 + (\tfrac{1 - V u_l}{1 - V u_r})^{\frac{1}{c^2}}, i.e.$$

$$\frac{V \Delta p + Y_l (V - u_l) c^2}{Y_l (V - u_l) c^2} = (\frac{1 - V u_l}{1 - V u_r})^{\frac{1}{c^2}}. \tag{4.34}$$

The formula following (4.25), i.e. $V \frac{\Delta p}{c^2} + Y_l \Delta u = \Delta Y (V - u_r)$, can be stated as
$V \Delta p + Y_l u_r c^2 - Y_l u_l c^2 = Y_r (V - u_r) c^2 - Y_l (V - u_r) c^2, i.e.\ V \Delta p + Y_l (V - u_l) c^2 = Y_r (V - u_r) c^2.$

Inserting this formula into the formula (4.34) gives

$$\frac{Y_r (V - u_r) c^2}{Y_l (V - u_l) c^2} = (\frac{1 - V u_l}{1 - V u_r})^{\frac{1}{c^2}}.$$

Finally, the jump condition for the second equation is

$(Y_r)^{c^2} (V - u_r)^{c^2} (1 - V u_r) = (Y_l)^{c^2} (V - u_l)^{c^2} (1 - V u_l)$. i.e. (4.32).□

4.6 Numerical approximations of relativistic fluid models.

In this section, we propose a numerical scheme for the solution of the two systems presented in introduction. It extends at once to two and three space dimension without dimensional splitting. The systems (4.1)-(4.4) and (4.1), (4.5), (4.3), (4.4) of relativistic fluid dynamics are in nonconservative form : close numerical schemes can give different discontinuous solutions, so one cannot be confident in the results given by the schemes unless they are validated. Schemes are given in the genuine physical situation : presence of gravitation, expanding background, two and three space dimension, in which they give the qualitatively expected results. But there are no very precise observational data that could validate them from a quantitative viewpoint. Therefore validation of the schemes is a problematic task. We will use formula (4.22)-(4.23), (4.22) and (4.32) to validate the respective schemes. Then we will compare them and evaluate their domain of validity.

In one dimension the space-time cells are rectangles $[(i-\frac{1}{2})h,(i+\frac{1}{2})h]\times[nrh,(n+1)rh]$, $h=$ the space step, $i\in\mathbb{Z}, n\in\mathbb{N}, r>0$ small enough.

Numerical scheme for system (4.1)-(4.4). Multiply by u equation (4.1), multiply by $\frac{\rho}{\rho+\frac{p}{c^2}}$ equation (4.2) and add the two equations thus obtained :

$$(\rho u)_t+[(\rho+\frac{p}{c^2})u^2]_x=\frac{p}{c^2}uu_x-\frac{\rho}{\rho+\frac{p}{c^2}}p_x-\rho\Phi_x. \tag{4.35}$$

This transformation is mathematically allowed from Theorem 4.4.1. Then the state law is inserted into the equations : this insertion is not permitted since it leads to the inconsistencies found in section 4.2, but we will test a posteriori from the formulas (4.22)-(4.23) the validity of the scheme. If one suppresses gravitation as in section 4.4, insertion of the state law (4.4) gives the system

$$\rho_t+[(1+K)\rho u]_x=0, \tag{4.36}$$

$$(\rho u)_t+[(1+K)\rho u^2]_x=K\rho uu_x-\frac{c^2K}{1+K}\rho_x. \tag{4.37}$$

We apply a splitting of equations to this system. Let $\rho_i^n,(\rho u)_i^n,u_i^n{}_{i\in\mathbb{Z}}$ be given. If $a,b\in\mathbb{R}$ we set, formula (1.16),

$$L(a,b)=length\ of\ [0,1]\cap[a,b]=max(0,min(1,b)-max(0,a)). \tag{4.38}$$

- *Convection step with the field of velocity* w_i^n

$$w_i^n=(1+K)u_i^n, \tag{4.39}$$

$$\overline{\rho}_i:=\rho_{i-1}^nL(-1+rw_{i-1}^n,rw_{i-1}^n)+\rho_i^nL(rw_i^n,1+rw_i^n)+\rho_{i+1}^nL(1+rw_{i+1}^n,2+rw_{i+1}^n). \tag{4.40}$$

When the CFL condition $r|w_i^n|\leq 1\ \ \forall i,n$ is satisfied the first term multiplied by h represents the quantity ρ issued from the cell I_{i-1} between times t_n and t_{n+1} that lies in the cell I_i at time t_{n+1}. Indeed the cell $[(i-\frac{3}{2})h,(i-\frac{1}{2})h]$ has been transported according to the vector rw_{i-1}^nh, since w_{i-1}^n is the numerical velocity and the duration time is rh. The overlap with the fixed cell $[(i-\frac{1}{2})h,(i+\frac{1}{2})h]$ has length rw_{i-1}^nh if $w_{i-1}^n\geq 0$, 0 if $w_{i-1}^n\leq 0$ (taking into account the CFL condition $|w_{i-1}^n|\leq 1$). From (4.38) one finds $L(-1+rw_{i-1}^n,rw_{i-1}^n)=rw_{i-1}^n$ if $w_{i-1}^n\geq 0$, 0 if $w_{i-1}^n\leq 0$. Division by h is due to the fact $\overline{u}_i,u_j^n$ are mean values on cells of length h.
The second term in (4.40) multiplied by h represents the quantity ρ issued from the cell I_i that

remains in I_i at time t_{n+1}. Indeed the cell $[(i-\frac{1}{2})h,(i+\frac{1}{2})h]$ has been transported by the vector $rw_i^n h$. The overlap with the fixed cell $[(i-\frac{1}{2})h,(i+\frac{1}{2})h]$ is $h-rw_i^n h$ if $w_i^n \geq 0$, $h+rw_i^n h$ if $w_i^n \leq 0$. From (4.38) one finds $L(rw_i^n, 1+rw_i^n) = 1-rw_i^n$ if $w_i^n \geq 0$, $1+rw_i^n$ if $w_i^n \leq 0$.
The third term in (4.40) is similar to the first one : it concerns the quantity ρ issued from the cell I_{i+1} that lies in the cell I_i at time t_{n+1}. Note that $\overline{\rho}_i$ depend on n, which is not explicitely stated to shorten the notation. The same discretization as the one in (4.40) gives

$$\overline{(\rho u)}_i := (\rho u)_{i-1}^n L(-1+rw_{i-1}^n, rw_{i-1}^n) + (\rho u)_i^n L(rw_i^n, 1+rw_i^n) + (\rho u)_{i+1}^n L(1+rw_{i+1}^n, 2+rw_{i+1}^n), \tag{4.41}$$

$$\overline{u}_i = \frac{\overline{\rho u}_i}{\overline{\rho}_i}. \tag{4.42}$$

• *Averaging step.* This step is needed in general to avoid oscillations that can occur in its absence due to the centered discretization in the third step. Let $\alpha \in [0, \frac{1}{2}[$. We set

$$\rho_i^{n+1} := \alpha\overline{\rho}_{i-1} + (1-2\alpha)\overline{\rho}_i + \alpha\overline{\rho}_{i+1}, \tag{4.43}$$

$$\widetilde{(\rho u)}_i := \alpha\overline{(\rho u)}_{i-1} + (1-2\alpha)\overline{(\rho u)}_i + \alpha\overline{(\rho u)}_{i+1}. \tag{4.44}$$

The case $\alpha = 0$ corresponds to the absence of averaging. The presence of $\alpha \neq 0$ is often needed to compensate possible oscillations due to the centered discretization in the correction step.

• *Correction step (dropping provisionally the gravitation potential).*

$$(\rho u)_i^{n+1} = \widetilde{\rho u}_i + \frac{Kr}{2}\overline{(\rho u)}_i(\overline{u}_{i+1} - \overline{u}_{i-1}) - \frac{c^2 Kr}{2(1+K)}(\overline{\rho}_{i+1} - \overline{\rho}_{i-1}), \tag{4.45}$$

$$u_i^{n+1} = \frac{(\rho u)_i^{n+1}}{\rho_i^{n+1}}. \tag{4.46}$$

The CFL condition is

$$r(1+K)\|u\|_\infty < 1. \tag{4.47}$$

If $\rho_i^0 > 0\ \forall i$ then the middle L in (4.40) is nonzero and, by induction on n, one can easily verify that $\overline{\rho}_i > 0$ and $\rho_i^{n+1} > 0\ \forall n$. From formulas (4.40) and (4.43), $\sum_i \rho_i^n h = \sum_i \rho_i^0 h \quad \forall n$, which is the L^1 stability in ρ.

Numerical scheme for system (4.1), (4.5), (4.3), (4.4). One mutiplies equations (4.1) by u and equation (4.5) by ρ, and add the equations so obtained. This gives

$$(\rho u)_t + (1+K)(\rho u^2)_x = K\rho u u_x - \frac{Kc^2}{1+K}\rho_x + Kc^2(\rho u)_x u - \rho\Phi_x. \tag{4.48}$$

The system thus obtained is the same as the one obtained from (4.1)-(4.4) with the additional term $Kc^2(\rho u)_x u$ in the Euler equation. We adopt an identical numerical scheme except that formula (4.45) is replaced by

$$(\rho u)_i^{n+1} = \widetilde{\rho u}_i + \frac{Kr}{2}\overline{(\rho u)}_i(\overline{u}_{i+1} - \overline{u}_{i-1}) - \frac{c^2 Kr}{2(1+K)}(\overline{\rho}_{i+1} - \overline{\rho}_{i-1}) + \frac{c^2 Kr}{2}\overline{u}_i(\overline{\rho u}_{i+1} - \overline{\rho u}_{i-1}), \tag{4.49}$$

in which this additional term is treated by a centered discretization. For both systems gravitation

is treated by a centered discretization.

To validate the schemes one compares their results with the jump formulas (4.22)-(4.23), (4.22) and (4.32) respectively. One chooses $c = 1$. Then for each shock in an interval $]a, b[,\ a < b$ with constant values on both sides one computes the quantities

$$V := (1 + K)\frac{(\rho u)(b) - (\rho u)(a)}{\rho(b) - \rho(a)}, \tag{4.50}$$

$$r_1 := VK(\rho(b) - \rho(a)), \tag{4.51}$$

$$r_2 = (1 + K)\rho(a)(V - u(a))[exp(V(u(b) - u(a)) - 1], \tag{4.52}$$

for system(4.1)-(4.4). For system (4.1), (4.5), (4.3), (4.4) one replaces r_1 and r_2 by

$$r_1 = (1 + K)\rho(a)(V - u(a))(1 - Vu(a)), \tag{4.53}$$

$$r_2 = (1 + K)\rho(b)(V - u(b))(1 - Vu(b)). \tag{4.54}$$

From the jump formulas (4.22)-(4.23) or (4.22) and (4.32) respectively, one should have $r_1 = r_2$. For each shock wave one computes the relative errors E_l, E_r on the left, right discontinuities respectively, from the formula

$$E = \frac{2|r_1 - r_2|}{|r_1| + |r_2|}. \tag{4.55}$$

One chooses $K = \frac{1}{3},\ G = 0,\ h = \frac{1}{2000},\ r = 0.1,\ 6000$ iterations, $\alpha = 0.05$. For each Riemann problem in the array below one has computed the relative error on each shock wave, first on system (4.1)-(4.4), then on system (4.1), (4.5), (4.3), (4.4) with $\alpha = 0.05$.

ρ_l	u_l	ρ_r	u_r	$E_l, (1,2,3,4)$	$E_r, (1,2,3,4)$	$E_l, (1,5,3,4)$	$E_r, (1,5,3,4)$
2	0.6	3	0.4	0.002	0.000	0.000	0.000
8	0.5	6	0.3	0.000	0.008	0.000	0.000
5	0.7	3	0.5	0.004	0.013	0.000	0.010
8	0.7	6	0.4	0.001	0.012	0.000	0.019
2	0.9	4	0.1	0.025	0.011	0.000	0.000
2	0.9	9	0.1	0.031	0.000	0.000	0.000
1	0.9	9	0.1	0.031	no jump	0.000	0.000
5	0.9	2	0.1	0.002	0.058	0.000	0.000
9	0.9	1	0.1	no jump	0.124	0.000	0.008

The velocity of light has been chosen =1, so jumps from 0.9 to 0.1 represent extremely strong variations in the initial data which lie outside of the physical domain of validity of the equations from the factor $(1 - v^2)$ (case $c = 1$) in [42] (2.10.16) p. 49. The conclusion is that the scheme for system (4.1)-(4.4) works well for relatively moderate jumps such as the first four ones (relative error no more than 1 per cent), but sometimes works very poorly for large jumps such as the last two ones. For Riemann problems such as $(\rho_l = 9, u_l = 0.9, \rho_r = 1, u_r = 0.1)$ the intermediate value of the velocity step from the scheme for system (4.1)-(4.4) bypasses the velocity of light which puts in evidence the inadequacy of the model (4.1)-(4.4) in this case : indeed important simplifications have been done relatively to the Euler equation in [42] p. 49. The scheme for the

model (4.1), (4.5), (4.3), (4.4) gives better results.

Now we compare the two systems and their schemes by means of figures in three situations. The numerical solution of system (4.1)-(4.4) is represented by a continuous line and the numerical solution of system (4.1), (4.5), (4.3), (4.4) is represented by o.

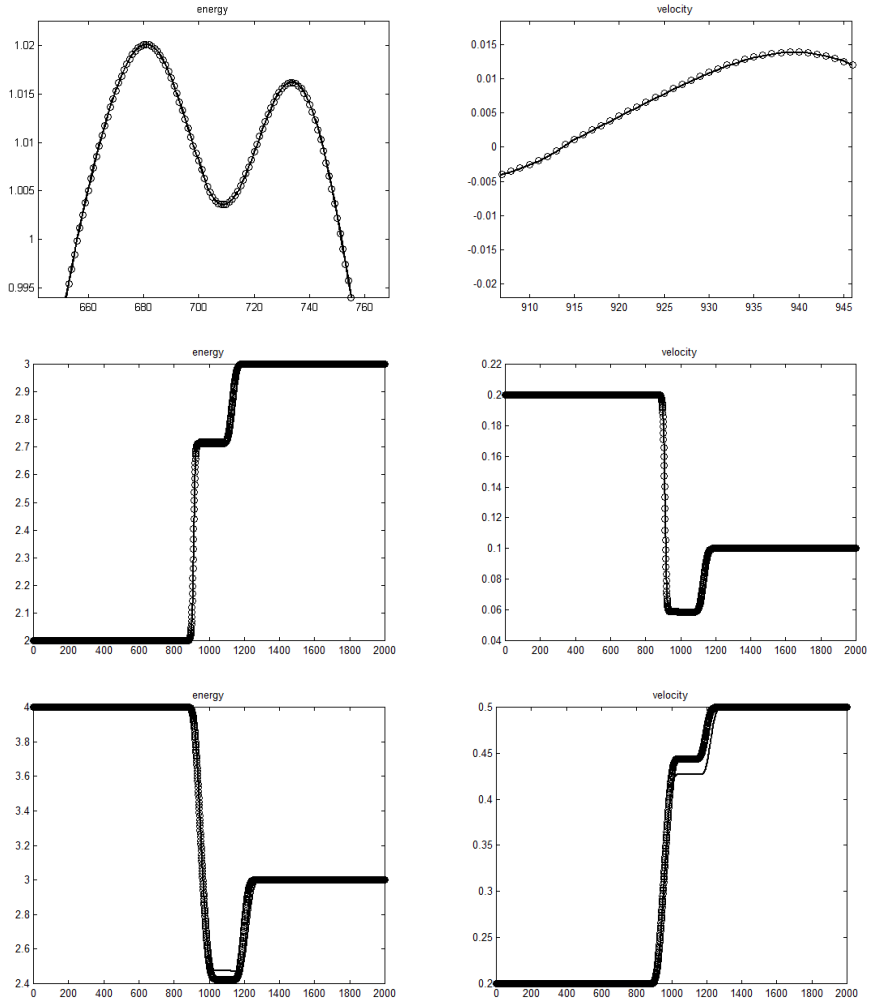


Figure 4.6.1. Comparison of the solutions from the two different systems in three different situations involving velocities < 0.5 c.

In figure 4.6.1 we compare the numerical solutions from the two different systems in three different situations involving velocities $< \frac{c}{2}$. In the top panels one considers initial conditions at random between 0.9 and 1.1 in energy density and between -0.1 and 0.1 in velocity. The panels represent zooms showing that the two systems give exactly the same result. This case is close to the domain of validity of perturbation theory. In the middle panels $(\rho_l, \rho_r, u_r, u_l) = (2, 3, 0.2, 0.1)$ and $(4, 3, 0.2, 0.5)$ in the bottom panels. One observes the two systems give close results.

The scheme can be extended at once to two and three space dimension. For the first step this is done in formulas (1.21-1.27). For the second step the averaging is simply done by considering the 8 neighbors of a cell in 2-D and the 26 neighbors in 3-D. In 2-D the averaging analog to (4.43)-(4.44) can be as follows. Let α, $0 < \alpha < \frac{1}{20}$, be given in the scheme. Set

$$\rho_{i,j}^{n+1} := \alpha(2\overline{\rho}_{i-1,j-1} + 2\overline{\rho}_{i-1,j+1} + 2\overline{\rho}_{i+1,j-1} + 2\overline{\rho}_{i+1,j+1} + 3\overline{\rho}_{i-1,j} + 3\overline{\rho}_{i,j-1} + 3\overline{\rho}_{i,j+1} + 3\overline{\rho}_{i+1,j}) + (1 - 20\alpha)\overline{\rho}_{i,j}. \tag{4.56}$$

Finally the third step is a mere centered discretization. Let us give the formulas in 2-D. Formula (1.37) is replaced by

$$(\rho u)_t + (1+K)(\rho u^2)_x = K\rho u u_x - \rho v u_y - \frac{Kc^2}{1+K}\rho_x, \tag{4.57}$$

and formula (4.45) is replaced by

$$(\rho u)_{i,j}^{n+1} = \widetilde{\rho u}_{i,j} + \frac{Kr}{2}\overline{\rho u}_{i,j}(\overline{u}_{i+1,j} - \overline{u}_{i-1,j}) - \frac{r}{2}\overline{\rho v}_{i,j}(\overline{u}_{i,j+1} - \overline{u}_{i,j-1}) - \frac{c^2 Kr}{2(1+K)}(\overline{\rho}_{i+1,j} - \overline{\rho}_{i-1,j}), \tag{4.58}$$

and similar formulas for ρv. Formulas (4.48) and (4.49) are respectively replaced by

$$(\rho u)_t + (1+K)(\rho u^2)_x = K\rho u u_x - \rho v u_y - \frac{Kc^2}{1+K}\rho_x + Kc^2(\rho u)_x u, \tag{4.59}$$

$$(\rho u)_{i,j}^{n+1} = \widetilde{\rho u}_{i,j} + \tfrac{Kr}{2}\overline{\rho u}_{i,j}(\overline{u}_{i+1,j} - \overline{u}_{i-1,j}) -$$

$$\frac{r}{2}\overline{\rho v}_{i,j}(\overline{u}_{i,j+1} - \overline{u}_{i,j-1}) - \frac{c^2 Kr}{2(1+K)}(\overline{\rho}_{i+1,j} - \overline{\rho}_{i-1,j}) + \frac{c^2 Kr}{2}\overline{u}_{i,j}(\overline{\rho u}_{i+1,j} - \overline{\rho u}_{i-1,j}). \tag{4.60}$$

Due to the absence of dimensional splitting, the 2-D and 3-D scheme retain the numerical accuracy of the 1-D scheme as observed in Part I.

The numerical scheme can include the Poisson equation (4.3) from a direct integration in 1-D and from any classical numerical resolution in 2-D and 3-D. Then the partial derivatives of the gravitation potential are calculated by a centered discretization and inserted into (4.49) and (4.60).

The applications to cosmology request expanding background. Since this is classical, see Proposition 1.10.1 in chapter 1, [8] p. 216, p.294, p. 312, [30] p. 462-463, [31] p. 233, we do not state it explicitely.

4.7 Coexistence of a Newtonian fluid and a relativistic fluid.

Between the epoch of equivalence of matter and radiation and the epoch of decoupling of baryons and radiation a classical scenario consists in the coexistence of a Newtonian component (cold dark matter) and a relativistic component (baryons tightly coupled to radiation). It is considered dark matter perturbations were growing and therefore were creating a gravitational attraction of baryons counterbalanced by the huge internal pressure from the coupling of baryons with photons.

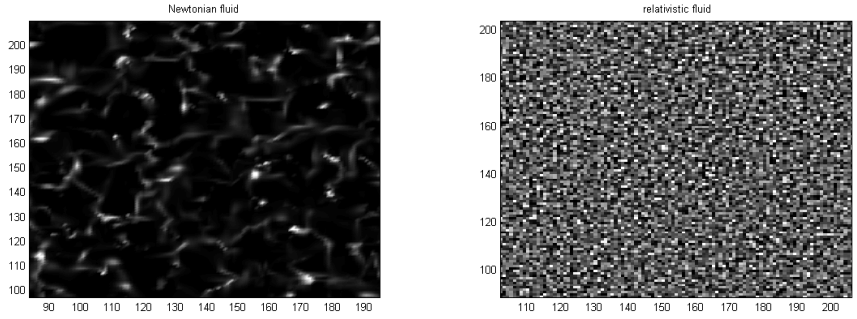


Figure 4.7.1. A mixture of a Newtonian fluid and a relativistic fluid in a slowly expanding universe.

The simulation in figure 4.7.1 is a numerical solution from the continuity and Euler equations for a Newtonian fluid ([8] p. 207, [30] p. 460, [31] p. 233) and the equations (4.1)-(4.3) or (4.1), (4.5), (4.3), coupled by a Poisson equation $\Delta\Phi = 4\pi G\rho_N + 4\pi G(\rho_R + \frac{p_R}{c^2})$, where ρ_N and ρ_R, p_R concern respectively the Newtonian and the relativistic fluid. The initial conditions are energy densities at random between 0.9 and 1.1, and x and y-velocities between between -0.1 and 0.1, for each fluid; $G = 1, r = 10,\ 200$ iterations, $\alpha = 0.05; K = \frac{1}{3}, c = 300000$. The scale factor has grown from 1 to 2.36 and the Newtonian fluid is supposed to be collisionless (no pressure). The Newtonian fluid is treated by the scheme in [C] with further a centered discretization of the gravitation potential. The relativistic fluid is indifferently treated by the schemes for systems (4.1), (4.2), (4.4) or (4.1), (4.5, 4.4). We observe creation of structures for the Newtonian fluid (left panel) and complete absence of structure for the relativistic fluid (right panel). In case of very fast expansion it has been observed absence of creation of structures (or an initial structure is frozen) : this is the Meszaros effect or stagnation. After decoupling the baryons are also treated by the Newtonian system and one has observed agglomeration of baryons on the existing structures of dark matter.

These facts are well known in perturbation theory. The interest of the method and results in this chapter lies in that they permit to investigate the fully nonlinear regime as needed [8] pp. 330-334, [30] pp. 485-493, [31] pp.285-288.

4.8 Conclusion.

The method presented in this chapter has permitted to perform calculations on basic nonlinear systems of equations in the theory of large structure formation in cosmology [8] p. 221, [30] p. 465 and calculate explicit irregular solutions. We have obtained jump conditions and explicit solutions for these systems. Finally one has produced a 1-D, 2-D, 3-D numerical scheme corresponding to these formulas in a physically significant fully nonlinear domain.

Chapitre 5

A holomorphic functional space.

An analysis of singular shock solutions of the Keyfitz-Kranzer system suggests a regularization of singular shocks in a functional space of classical germs of holomorphic functions. In this functional space a sequence of approximate solutions can converge to a well defined limit which can be a singular shock solution of the equations in a natural sense similar to the classical concept of a weak solution. In this context we obtain compactness and an analog of the classical result "consistency and stability imply convergence".

5.1 Introduction.

Singular shocks have been put in evidence by Keyfitz-Kranzer in a study of their model system [22], [21]. M. Sever [36] has shown families of equations that admit singular shock waves as solutions. The singular shocks have been observed from different numerical techniques : Dafermos-Di Perna viscosity in [22], [21], usual viscosity in [33], [35], and a unique solution to the Riemann problem has been obtained in [22], [21]. The work in this chapter has been motivated by the Cauchy problem.

We introduce a functional space in which a L^1-stable sequence of approximate step-function solutions can converge to a solution of the equations, even when such a solution involves singular shocks or delta shocks. This convergence is obtained from a compactness argument in a functional space of holomorphic functions having the usual space $\mathbb{R} \times \mathbb{R}^+$ on the boundary of their domain. The singular shocks or delta shocks appear as "boundary values" of holomorphic functions. These boundary values have properties close to those of the classical weak solutions. Such sequences of approximate solutions are provided by a numerical scheme in chapter 6, valid in particular for the Keyfitz-Kranzer system

$$u_t + (u^2 - v)_x = 0, \tag{5.1}$$

$$v_t + (\frac{1}{3}u^3 - u)_x = 0, \tag{5.2}$$

and for the system

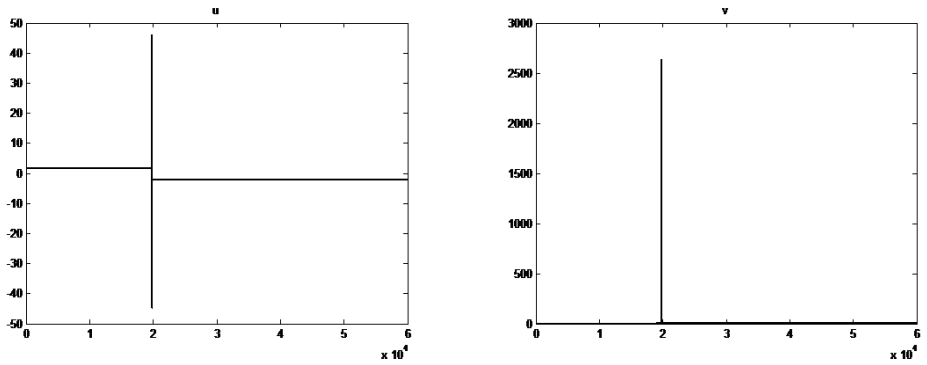
$$u_t + (u^2)_x = 0, \tag{5.3}$$

$$v_t + (uv)_x = 0, \tag{5.4}$$

originally considered by Korchinski [24], who put in evidence delta shocks in the solution of Riemann problems.

The mathematical context in use is a context of holomorphic functions defined on strips having the real axis on their boundary. The genuine solutions are the germs of holomorphic functions. The numerical results, which are always those already observed by all authors, are their weak limits on the real axis, called here their "real interpretation". In chapter 4 the same context and method have permitted to do explicit nonlinear calculations on classical special-relativistic equations widely used in cosmology, which have classical shock solutions that do not make sense within the distributions.

Justification and origin of the holomorphic regularization in this chapter. Here is a depiction of a typical singular shock (u, v) solution of the Keyfitz-Kranzer equations.



.Figure 5.1..1. A typical singular shock (u, v)

.

One observes that the function u in the singular shock presents an apparent contradiction which will be resolved by distinguishing the "genuine solution" from its aspect in the sense of distributions. The singularity which is observed on the discontinuity of u is insignificant from the viewpoint of distribution theory when $h \to 0$ since the area under each peak tends to 0 when $h \to 0$, even not taking into account that the two peaks tend to compensate each other in the integral $\int u(x,t)\psi(x)dx$. Therefore u can be viewed, with the "filter" of distribution theory, as a simple travelling discontinuity. But the two peaks in u are a basic ingredient in the solution of the equations. Indeed, if the function u in (5.1) were a mere discontinuity, then u_t would be in the form of a Dirac delta function located on the discontinuity (i.e. a delta wave) ; from (5.1), $(u^2 \quad v)_x$ would also be in the form of a delta wave. Therefore $u^2 \quad v$ would have the form of a Heaviside function. Then it would be impossible for u^2 to compensate the delta peak in v. Therefore u cannot be a mere discontinuity although it is interpreted as a mere discontinuity in distribution theory. The small peaks in the function u, which are insignificant in the sense of distributions, do play a basic role in the solution of the equation : they permit that u^2 could compensate the delta function in v. The same reasoning holds in (5.2) : v_t shows a derivative δ' of the Dirac δ function which, if u were a mere discontinuity, could not be compensated by $(\frac{u^3}{3} \quad u)_x$ that would be in the form of a Dirac delta function only.

The explanation proposed in this chapter consists in a distinction between the "genuine solution", denoted by U, which is not a distribution and carries the "small, but basically important" singularities observed in figure 5.1.1, and the aspect of U in the sense of distributions : a simple discontinuity which is not solution and is only the interpretation of the solution in the sense of distributions.

5.2 Mathematical context.

This context originated in the introduction of a regularizing small parameter in chapter 4 section 4.3 for calculations on equations of cosmology. At first a function $f = f(x)$ was regularized as a function $f(x, \xi)$, where ξ is a > 0 regularizing parameter, such that $f(x, \xi) \to f(x)$ in the sense of distributions when $\xi \to 0$, i.e.

$$\forall \psi \in \mathcal{C}_c^\infty(\mathbb{R}^n) \ \int f(x,\xi)\psi(x)dx \to \int f(x)\psi(x)dx$$

when $\xi \to 0$. Then we intended to use the property that the functions $(x, \xi) \to f(x, \xi)$ are analytic in (x, ξ), which amounts, using their extension to the complex domain, to transform them into $f(z, \zeta), z = x + iy \in \mathbb{C}^n, \zeta = \xi + i\eta \in \mathbb{C}$.

Since ξ, η, y are arbitrarily small our functional space is a kind of space of germs of holomorphic functions located on the space $\mathbb{R}^n$ (variable x), i.e. these functions are defined in variable open sets in $\mathbb{C}^n \times \mathbb{C}$ (variables $z = x + iy, \zeta = \xi + i\eta$) having the real space $\mathbb{R}^n$ on their boundary. Although these holomorphic germs are defined in a slightly original way concerning the domains of the functions $(z, \zeta) \longmapsto f(z, \zeta)$, nevertheless they are very classical mathematical objects. The classical theory of normal families of holomorphic functions provides the needed compactness property, even in case of singular shock waves. Now let us give details.

The letters r, θ, μ will always denote real numbers such that

$$0 < r < 1, \ 0 < \theta < \frac{\pi}{6}, \ 0 < \mu < \frac{1}{2}. \tag{5.5}$$

The values r, θ, μ will be as small as needed. One considers the open strip in $\mathbb{R}^{2n+2}$ parallel to the real space $\mathbb{R}^n$ of variable x defined by

$$S(r,\theta,\mu) =$$

$$\{(z,\zeta) \in \mathbb{C}^n \times \mathbb{C} \ / \ x \in \mathbb{R}^n, \ 0 < |\zeta| < r, \ -\theta < arg\zeta < \theta, |y_i| < \mu\xi \ \forall i = 1, ..., n\}. \tag{5.6}$$

The real space $\mathbb{R}^n$ lies on the boundary of $S(r, \theta, \mu)$ by letting $\zeta = \xi + i\eta$ tend to 0 (therefore from (5.6) $y \to 0$). Let $\mathcal{F}$ be the set of all strips $S(r, \theta, \mu)$, when $r, \theta, \mu \to 0$. The set $\mathcal{F}$ is a net for the inclusion :

$$\forall \ S_1, S_2 \in \mathcal{F} \ \exists S_3 \in \mathcal{F} \ / \ S_3 \subset S_1 \cap S_2.$$

We denote by *const* a positive real number which may not be the same from an expression to the following. If $S \in \mathcal{F}$, i.e. $S = S(r, \theta, \mu)$ for some r, θ, μ, one defines

$\mathcal{H}_S := \{\text{holomorphic functions } F : S \longmapsto \mathbb{C}, (z,\zeta) \longmapsto F(z,\zeta)\}$.

If $S' \subset S$ the restriction map $\mathcal{H}_S \longmapsto \mathcal{H}_{S'}$, $F \longmapsto F|_{S'}$, is injective from the uniqueness of analytic continuation. In the reunion of the sets $\mathcal{H}_S$ one considers the equivalence relation

$$(F_1, S_1) \equiv (F_2, S_2) \Leftrightarrow \ \exists S_3 \subset S_1 \cap S_2 \ \ / \ \ F_1|_{S_3} = F_2|_{S_3}.$$

The set of all equivalence classes is by definition our space of germs of holomorphic functions on $\mathbb{R}^n$ in the x-variable. Since this space is also classically refered to as an inductive limit we denote it by $Lim\mathcal{H}_S$. We introduce normed spaces contained in $Lim\mathcal{H}_S$.

$\mathcal{H}_{r,\theta,\mu,N} := \{$holomorphic germs that have a representative which is a holomorphic function $F : S(r,\theta,\mu) \longmapsto \mathbb{C}$ such that

$$|F(z,\zeta)| = O(\frac{1}{|\zeta|^N}) \quad \forall (z,\zeta) \in S(r,\theta,\mu)\}$$

with the norm

$$\|F\|_{r,\theta,\mu,N} = sup_{(z,\zeta)\in S(r,\theta,\mu)} |\zeta|^N |F(z,\zeta)|. \tag{5.7}$$

Lemma 5.2.1. $(\mathcal{H}_{r,\theta,\mu,N}, \|\|_{r,\theta,\mu,N})$ *is a Banach space.*

Proof. Since a Cauchy sequence (F_n) is bounded it satisfies the inequality $|F(z,\zeta)| \leq \frac{const}{|\zeta|^N}$ uniformly in n. Let K be a compact subset of $S(r,\theta,\mu)$. Since K is at a strictly positive distance from the boundary of $S(r,\theta,\mu)$ there exists $\epsilon > 0$ such that $(z,\zeta) \in K \Rightarrow |\zeta| > \epsilon$, since from (5.6) $(z, \zeta = 0) \notin S(r,\theta,\mu)$. Therefore a Cauchy sequence is a normal family of holomorphic functions [32]. Therefore the pointwise limit is a holomorphic function. Then the standard proof works. □

Lemma 5.2.2. If $r' \leq r,\ \theta' \leq \theta,\ \mu' < \mu,\ N' > N$, *then any partial derivative in the* x*-variable is a continuous linear map from* $\mathcal{H}_{r,\theta,\mu,N}$ into $\mathcal{H}_{r',\theta',\mu',N'}$.

Proof. It follows at once from Cauchy's integral formula. Indeed if $(z_0,\zeta) \in S(r,\theta,\mu-\epsilon), \epsilon > 0$ small enough, then, $|z - z_0| < \epsilon\xi \Rightarrow (z,\zeta) \in S(r,\theta,\mu)$. Indeed $|y_{0,i}| < (\mu - \epsilon)\xi$ and $|y_i - y_{0,i}| < \epsilon\xi \Rightarrow |y_i| < \mu\xi$. Cauchy's inequality then gives $|\frac{\partial}{\partial x_i} f(z_0,\zeta)| \leq \frac{const}{|\zeta|^N}\frac{1}{\epsilon\xi} = \frac{const}{|\zeta|^{N+1}}$.□

Lemma 5.2.3. If $(F_n)_n$ *is a bounded sequence in the normed space* $\mathcal{H}_{r,\theta,\mu,N}$ *then there is a subsequence* $(F_{n(p)})_p$ *and a germ of holomorphic function* $F \in \mathcal{H}_{r,\theta,\mu,N}$ *such that* $F_{n(p)} \to F$ *when* $p \to +\infty$ *uniformly on the compact subsets of the strip* $S(r,\theta,\mu)$.

Proof. From the proof of lemma 5.2.1 the family (F_n) is a normal family of holomorphic functions on the open set $S(r,\theta,\mu)$ [32].□

We denote by $\mathcal{H}(\mathbb{R}^n)$ the inductive limit of the spaces $\mathcal{H}_{r,\theta,\mu,N}$ directed by inclusions, when $r,\theta,\mu \to 0$ and $N \to \infty$. Now let us check that $\mathcal{H}(\mathbb{R}^n)$ contains many objects that can represent the usual irregular functions and distributions. To this end notice that to any function $f \in L^p(\mathbb{R}^n),\ 1 \leq p \leq \infty$, we can associate several elements $F \in \mathcal{H}(\mathbb{R}^n)$ that "give back" f on $\mathbb{R}^n$ considered on the boundary of $S(r,\theta,\mu)$ in the following way

$$\forall \psi \in \mathcal{C}_c^\infty(\mathbb{R}^n) \quad \lim_{\xi\to 0} \int F(x,\xi)\psi(x)dx = \int f(x)\psi(x)dx. \tag{5.8}$$

When (5.8) holds we say that f is the *real interpretation* of F. This can be done by convolution : set for instance the mollifier

$$\rho(z) = const \frac{1}{((z_1)^2+1)^s \dots ((z_n)^2+1)^s},\ s \in \mathbb{N},\ s \geq 1 \tag{5.9}$$

and define

$$F(z,\zeta) := \int_{\mathbb{R}^n} f(\lambda) \frac{1}{\zeta^n} \rho(\frac{\lambda - z}{\zeta}) d\lambda. \tag{5.10}$$

Lemma 5.2.4. $\forall f \in L^p(\mathbb{R}^n),\ \ 1 \leq p \leq \infty$, *the function* F *defined in (5.10) is in* $\mathcal{H}(\mathbb{R}^n)$ *and it has* f *as real interpretation. Further, if* f *is continuous at a point* x_0 *then* $F(x,\xi) \to f(x_0)$ *when* $\xi \to 0$ *and* $x \to x_0$.

Proof. For simplicity the proof is given in the case $n = 1$. Let r, θ, μ satisfying (5.5) be given. From (5.10)

$$F(z,\zeta) = const.\zeta^{2s-1} \int_{\mathbb{R}} \frac{f(\lambda)}{[(\lambda - z)^2 + \zeta^2]^s} d\lambda. \tag{5.11}$$

Auxiliary calculation : $|(\lambda - z)^2 + \zeta^2| \geq |Real((\lambda - z)^2 + \zeta^2)| = (\lambda - x)^2 + \xi^2 - y^2 - \eta^2 \geq (\lambda - x)^2 + \xi^2(1 - \mu^2 - tan^2\theta) > (\lambda - x)^2 + \frac{\xi^2}{2}$ from (5.5)-(5.6). Therefore the denominator in (5.11) does not take the value 0 when $(z,\zeta) \in S(r,\theta,\mu)$; this is the motivation for the last inequalities in (5.5)-(5.6). We will use that $|(\lambda - z)^2 + \zeta^2| \geq (\lambda - x)^2 + \alpha^2\xi^2$ with $\alpha = 2^{-\frac{1}{2}} > 0$.□

• If $f \in L^\infty$, (5.11) gives $|F(z,\zeta)| \leq const|\zeta|^{2s-1}\|f\|_\infty \int \frac{d\lambda}{[(\lambda-x)^2+\alpha^2\xi^2]^s} \leq const\|f\|_\infty$, since $\int \frac{d\lambda}{[(\lambda-x)^2+\alpha^2\xi^2]^s} = \int \frac{d\lambda}{[(\lambda)^2+\alpha^2\xi^2]^s} = (\alpha\xi)^{1-2s} \int \frac{d\mu}{(\mu^2+1)^s}$.

• If $f \in L^1$ the auxiliary calculation gives

$$|F(z,\zeta)| \leq const \frac{1}{\xi} \|f\|_{L^1}. \tag{5.12}$$

More generally if $f \in L^p, 1 \leq p < \infty$, one obtains
$|F(z,\zeta)| \leq const|\zeta|^{2s-1}\|f\|_{L^p}(\int \frac{d\lambda}{[(\lambda-x)^2+\alpha^2\xi^2]^{qs}})^{\frac{1}{q}} \leq const|\zeta|^{-\frac{1}{p}}\|f\|_{L^p},\ \frac{1}{p} + \frac{1}{q} = 1.$

The last assertion is classical from the formula $F(x,\xi) = \int f(x + k\xi)\rho(k)dk$ and the fast decrease of ρ at ∞.□

These results can be easily extended to $\mathbb{R}^n \times]0,T[$, considering f null out of $\mathbb{R}^n \times]0,T[$. A Heaviside function $H \in \mathcal{H}(\mathbb{R})$ is a germ whose real interpretation is the Heaviside function ; it suffices to take as f in (5.10) the Heaviside function. A Dirac function δ in $\mathcal{H}(\mathbb{R})$ is a germ whose real interpretation is the Dirac delta distribution. To obtain a Dirac function it suffices to take the derivative of a Heaviside function $H \in \mathcal{H}(\mathbb{R})$.

Besides the concept of solution of equations in the sense of equality in $\mathcal{H}(\mathbb{R} \times]0,T[)$, we consider also solutions in a weak sense, for which a natural definition (in the case $n = 1$ for simplification) is as follows. Let $\Phi : \mathbb{R}^m \to \mathbb{R}^m$ be a set of m polynomials in m variables, for instance $U_t + \frac{\partial}{\partial x}\Phi(U) = 0$ can be (5.1)-(5.2) or (5.3)-(5.4).

Definition of a concept of weak solution. $U = (U^j)_{j=1,...,m}$, where each $U^j \in \mathcal{H}(\mathbb{R}\times]0,T[)$, is a weak solution of the system $U_t + \frac{\partial}{\partial x}\Phi(U)$ " $=$" 0 of m scalar equations iff each component of $U_t + \frac{\partial}{\partial x}\Phi(U)$ has the null function as real interpretation i.e.

$$\forall j = 1, ..., m,\ \forall \psi \in C_c^\infty(\mathbb{R}\times]0,T[),\quad \int_{\mathbb{R}\times]0,T[} [(U^j)_t + \frac{\partial}{\partial x}(\Phi(U))^j](x,t,\xi)\psi(x,t)dxdt \to 0 \quad (5.13)$$

when $\xi \to 0^+$. This is denoted by $U_t + \frac{\partial}{\partial x}(\Phi(U)) \overset{weak}{=} 0$.

As the usual concept of a weak solution this concept of weak solution suffers from nonuniqueness and classical examples show it does not allow free manipulation of equations.

5.3 Consistency and stability imply convergence.

One assumes the existence of sequences $(u_n), (v_n)$ of step functions $\mathbb{R}\times]0,T[) \longmapsto \mathbb{R}$, constant on rectangles $](i-\frac{1}{2})h_n, (i+\frac{1}{2})h_n[\times](j-\frac{1}{2})k_n, (j+\frac{1}{2})k_n[$, where $h_n \to 0, k_n \leq h_n$ when $n \to \infty$. We assume the sequences $(u_n), (v_n)$ satisfy the following properties :

(i) *Consistency in the sense of distributions :* $\forall \psi \in \mathcal{C}_c^\infty(\mathbb{R}\times]0,T[)$

$$\int [u_n\psi_t + (u_n^2 - v_n)\psi_x]dxdt \to 0, \quad (5.14)$$

$$\int [v_n\psi_t + (\frac{(u_n)^3}{3} - u_n)\psi_x]dxdt \to 0, \quad (5.15)$$

when $n \to +\infty$, in the case (1,2), and similar properties in the case (5.3)-(5.4).

(ii) *Stability :* there exists a real number $const > 0$, independent on n and t, such that

$$\int_{\mathbb{R}} |u_n(x,t)|dx \leq const, \int_{\mathbb{R}} |v_n(x,t)|dx \leq const \quad (5.16)$$

for almost all $t \in]0,T[$. Of course this implies

$$\int_{\mathbb{R}\times]0,T[} |u_n(x,t)|dxdt \leq const, \int_{\mathbb{R}\times]0,T[} |v_n(x,t)|dxdt \leq const \quad (5.17)$$

and

$$\|u_n\|_\infty \leq \frac{const}{h_n}, \|v_n\|_\infty \leq \frac{const}{h_n} \quad (5.18)$$

in the interior of the rectangles of sides h_n, k_n where these functions are constant (consider the extreme case in which these functions are null except in one rectangle only, and apply (5.16)).

It follows from (5.17) that the sequences $(u_n), (v_n)$ are bounded in $L^1(\mathbb{R}\times]0,T[)$. Therefore by *weak compactness one can extract subsequences that converge * weakly in the space $\mathcal{M}_b(\mathbb{R}\times]0,T[)$ of bounded Radon measures to some elements $u, v \in \mathcal{M}_b(\mathbb{R}\times]0,T[)$. From now on we simplify the notation by considering that the whole sequences $(u_n), (v_n)$ are convergent.

Theorem 5.3.1. *Under the above assumptions (i) and (ii) of consistency and stability the scheme converges in the sense :*

there exists a subsequence of the sequence (u_n, v_n), still denoted (u_n, v_n) to simplify the notation, two sequences $(U_n), (V_n)$ of elements of $\mathcal{H}(\mathbb{R}\times]0,T[)$ and a pair U, V of elements of $\mathcal{H}(\mathbb{R}\times]0,T[)$ such that
i)$\forall n$, *U_n, V_n have the real interpretations u_n, v_n respectively,*
ii) U, V have the real interpretation u, v respectively,
iii) $U_n \to U, V_n \to V$ uniformly on any compact set of a strip $S(r,\theta,\mu)$,
iv) the pair (U, V) is a weak solution in the sense (5.13) of the equations (5.1)-(5.2) (respectively (5.3)-(5.4) if (5.14)-(5.15) are adapted to (5.3)-(5.4)).

Proof. The letter t can represent a complex number when coupled with z or a real number when coupled with x. This will not create any confusion. We use a holomorphic mollifier

$$\rho(z,t) := \frac{const}{(1+z^2)^s(1+t^2)^s}, \tag{5.19}$$

where $z = x+iy, t = \tau+i\tau' \in \mathbb{C}, x, y, \tau, \tau' \in \mathbb{R}$. The real value *const* is such that $\int \rho(x,\tau)dxd\tau = 1$, for some $s \in \mathbb{N}, s > 1$ to be fixed later. We set $\rho_{\epsilon_1,\epsilon_2}(z,t) := \frac{1}{\epsilon_1.\epsilon_2}\rho(\frac{z}{\epsilon_1}, \frac{t}{\epsilon_2})$ where $\epsilon_1, \epsilon_2 \in \mathbb{C}$.

We set

$$U_n(z,t,\zeta) := [u_n * \rho_{\epsilon_1,\epsilon_2}](z,t), \quad \epsilon_1 = \zeta.(h_n)^\alpha, \epsilon_2 = \zeta.(k_n)^\alpha, \tag{5.20}$$

for some $\alpha > 0$ to be fixed later. We use the same formula for V_n, replacing u_n by v_n.

It follows from (5.17) and lemma 5.2.4, that U_n, V_n are defined on the strip $S(r,\theta,\mu)$ in (x,t) variable in $\mathbb{R}\times]0,T[\ \forall(r,\theta,\mu)$ satisfying (5.5), and that they admit u_n, v_n as real interpretations respectively. The families $\{U_n\}, \{V_n\}$ are bounded in the normed space $\mathcal{H}_{r,\theta,\mu,1}$ from (5.12), which permits to apply lemma 5.2.3. We denote again by $(U_n), (V_n)$ the convergent sequences thus obtained and by U, V their respective limits in $\mathcal{H}(\mathbb{R}\times]0,T[)$. The main part of the proof consists in proving that for s and α large enough (independent on ψ) one has

$$\int [U_n(x,t,\xi)\psi_t(x,t) + (U_n^2 - V_n)(x,t,\xi)\psi_x(x,t)]dxdt \to$$
$$\int [u_n(x,t)\psi_t(x,t) + (u_n^2(x,t) - v_n(x,t))\psi_x(x,t)]dxdt \tag{5.21}$$

uniformly in n when $\xi \to 0$, as well as

$$\int [V_n(x,t,\xi)\psi_t(x,t) + (\frac{1}{3}U_n^3 - U_n)(x,t,\xi)\psi_x(x,t)]dxdt \to \int [v_n\psi_t + (\frac{1}{3}u_n^3 - u_n)\psi_x]dxdt$$

for (5.2). Similar formulas are proved for the two equations (5.3)-(5.4). This convergence is obtained at once in linear terms such as $\int (U_n - u_n)\psi_t$: indeed in one dimension to simplify the notation, one has $|\int (u_n * \rho_\epsilon - u_n)\psi| = |\int u_n(x)\rho(\mu)[\psi(x+\epsilon\mu) - \psi(x)]d\mu dx| \leq const.\epsilon\|u_n\|_{L^1}$. In the case of nonlinear terms this will be proved in the next section. Assume (5.21) holds. Then consider the following diagram

$$\int [U_n\psi_t + (U_n^2 - V_n)\psi_x]dxdt \overset{\xi\to 0, fixed\ n}{\longrightarrow} \int [u_n\psi_t + (u_n^2 - v_n)\psi_x]dxdt$$
$$n \to \infty \downarrow \text{ fixed } \xi \qquad\qquad n \to \infty \downarrow$$

$$\int [U\psi_t + (U^2 - V)\psi_x]dxdt \qquad \stackrel{\xi\to 0}{\longrightarrow} \qquad 0.$$

From (5.21) the limit in the top horizontal arrow is uniform in n. The left vertical arrow is a simple limit for fixed ξ from the definition of U as limit of the U_n's uniformly on compact subsets of $S(r,\theta,\mu)$. The right vertical arrow is the limit (5.14). Therefore since the top horizontal arrow is uniform in n then the bottom horizontal arrow holds as a limit when $\xi \to 0$, the double limit holds and the diagram is commutative. □

5.4 Proof of the uniform convergence.

In this section we prove the uniform convergence in the top horizontal line of the diagram, i.e. (5.21). In the proof we intend to use compactness of the support of the mollifier, which is impossible since the mollifier ρ is analytic. Therefore the proof is based on a cut-off of the (positive for real variables) mollifier into a "main part of integral close to 1" which is compactly supported in $[-\xi^{-\beta}h^{-1}, \xi^{-\beta}h^{-1}] \times [-\xi^{-\beta}k^{-1}, \xi^{-\beta}k^{-1}]$, $\beta \in]0,1[$ given and a "minor part", of integral close to 0, supported in the complement of this rectangle.

To simplify the formulation the quantity $\int f(u_n, v_n)\psi_x dxdt$ is replaced by a quantity $\int f(u_n)\psi dxdt$ where f is a function of one variable ($f(u) = u^2, u^3$ for (5.1)-(5.3), $f(u,v) = uv$ in (5.4) is treated in the same way) and where we use (5.16)-(5.18) on u_n.

We aim at proving that

$$\int f(U_n(x,t,\xi))\psi(x,t)dxdt \to \int f(u_n(x,t))\psi(x,t)dxdt \tag{5.22}$$

uniformly in n when $\xi \to 0$. We set

$$\rho(z,t) = \frac{const}{(1+z^2)^s(1+t^2)^s},\ \rho_{\epsilon_1,\epsilon_2}(z,t) = \frac{1}{\epsilon_1.\epsilon_2}\rho(\frac{z}{\epsilon_1}, \frac{t}{\epsilon_2}), \tag{5.23}$$

$s \in \mathbb{N}$ to be chosen later. For given n we replace h_n, k_n by h, k respectively to shorten the notation. We set

$$U_n(z,t,\zeta) = (u_n * \rho_{\zeta h^\alpha, \zeta k^\alpha})(z,t), \tag{5.24}$$

$\alpha > 0$ to be chosen later. Then, it follows from (5.12) and (5.16) that $U_n \in \mathcal{H}_{r,\theta,\mu,1}$ for any r, θ, μ satisfying (5.5). In systems (5.1)-(5.2) and (5.2)-(5.3) there exists N such that

$$|f(u)| \le const.|u|^N,\ |f'(u)| \le const.|u|^{N-1}, \tag{5.25}$$

for $|u|$ large enough. Let $\beta \in]0,1[$ be given. As explained above the function ρ is cut-off into

$$\rho = \rho\chi_{\xi^{-\beta}h^{-1}, \xi^{-\beta}k^{-1}} + (\rho - \rho\chi_{\xi^{-\beta}h^{-1}, \xi^{-\beta}k^{-1}}) \tag{5.26}$$

where $\chi_{\mu,\nu}$ denotes the characteristic function of the rectangle $]-\mu,\mu[\times]-\nu,\nu[$, $\mu > 0, \nu > 0$. For large μ, ν we will use the following bound from (5.23) :

$\int_\mu^{+\infty} \frac{1}{(1+x^2)^s}dx \le const \int_\mu^{+\infty} \frac{dx}{x^{2s}} = const.\mu^{-2s+1}$, with $const$ independent on s since $s \ge 1$. Therefore

$$\int_{\xi^{-\beta}h^{-1}}^{+\infty} \frac{1}{(1+x^2)^s}dx \leq const.\xi^{\beta(2s-1)}h^{2s-1}. \tag{5.27}$$

Proposition 5.4.1. *For $s \geq \frac{1+N}{2}$ and $\alpha \geq 2+N$ then $|\int [f(U_n(x,t,\xi))-f(u_n(x,t))]\psi(x,t)dxdt| \to 0$ uniformly in n when $\xi \to 0$.*

Proof. First decompose $\int [f(U_n(x,t,\xi)) - f(u_n(x,t))]\psi(x,t)dxdt = I + I_1 + I_2 + I_3$ where

$$I = \int \{f[u_n * (\rho\chi_{\xi^{-\beta}h^{-1},\xi^{-\beta}k^{-1},})_{\xi h^\alpha,\xi k^\alpha}] - f(u_n) * (\rho\chi_{\xi^{-\beta}h^{-1}\xi^{-\beta}k^{-1}})_{\xi h^\alpha,\xi k^\alpha}\}(x,t)\psi(x,t)dxdt, \tag{5.28}$$

$$I_1 = -\int [f(u_n) * (\rho - \rho\chi_{\xi^{-\beta}h^{-1},\xi^{-\beta}k^{-1}})_{\xi h^\alpha,\xi k^\alpha}](x,t)\psi(x,t)dxdt, \tag{5.29}$$

$$I_2 = \int \{f[u_n * (\rho\chi_{\xi^{-\beta}h^{-1},\xi^{-\beta}k^{-1}})_{\xi h^\alpha,\xi k^\alpha} + u_n * (\rho - \rho\chi_{\xi^{-\beta}h^{-1},\xi^{-\beta}k^{-1}})_{\xi h^\alpha,\xi k^\alpha}] - f[u_n * (\rho\chi_{\xi^{-\beta}h^{-1},\xi^{-\beta}k^{-1}})_{\xi h^\alpha,\xi k^\alpha}]\}(x,t)\psi(x,t)dxdt, \tag{5.30}$$

$$I_3 = \int [(f(u_n) * \rho_{\xi h^\alpha,\xi k^\alpha})(x,t) - f(u_n)(x,t))]\psi(x,t)dxdt. \tag{5.31}$$

Functions $f(U_n(x,t,\xi))$ and $f(u_n(x,t))$ are respectively the first term in I_2, see (5.24) and (5.26), and the second term in I_3. Simplifications occur between the first term in I and the second term in I_2, the second term in I and the second term from the parenthesis in I_1, the first term inside the ρ parenthesis in I_1 and the first term in I_3. We will give separate bounds for I, I_1, I_2 and I_3.

• *Bound of I.* The u_n's are step functions constant on the rectangles $R_{i,j} :=](i-\frac{1}{2})h, (i+\frac{1}{2})h[\times](j-\frac{1}{2})k, (j+\frac{1}{2})k[$. Let us state

$$l := \int \rho\chi_{\xi^{-\beta}h^{-1},\xi^{-\beta}k^{-1}}dxdt = \int_{-\xi^{-\beta}h^{-1}}^{\xi^{-\beta}h^{-1}} \int_{-\xi^{-\beta}k^{-1}}^{\xi^{-\beta}k^{-1}} \rho(x,t)dxdt. \tag{5.32}$$

From (27) and $\int \rho(x,t)dxdt = 1$,

$$l = (1 - const.\xi^{\beta(2s-1)}h^{(2s-1)})(1 - const.\xi^{\beta(2s-1)}k^{(2s-1)}) = 1 - const.\xi^{\beta(2s-1)}h^{(2s-1)}. \tag{5.33}$$

Since ξ and h are small and since $-\beta+1 > 0, \alpha-1 > 0$, the support of $(\rho\chi_{\xi^{-\beta}h^{-1},\xi^{-\beta}k^{-1}})_{\xi h^\alpha,\xi k^\alpha}$, namely $[-\xi^{-\beta+1}h^{\alpha-1}, \xi^{-\beta+1}h^{\alpha-1}] \times [-\xi^{-\beta+1}k^{\alpha-1}, \xi^{-\beta+1}k^{\alpha-1}]$, is small for ξ, h, k small, so it is contained in $[-\frac{h}{2}, \frac{h}{2}] \times [-\frac{k}{2}, \frac{k}{2}]$. In the central parts

$$[(i-\frac{1}{2})h+\xi^{-\beta+1}h^{\alpha-1}, (i+\frac{1}{2})h-\xi^{-\beta+1}h^{\alpha-1}]\times[(j-\frac{1}{2})k+\xi^{-\beta+1}k^{\alpha-1}, (j+\frac{1}{2})k-\xi^{-\beta+1}k^{\alpha-1}] \tag{5.34}$$

of the rectangles $R_{i,j}$ the functions

$$f[u_n * (\rho\chi_{\xi^{-\beta}h^{-1},\xi^{-\beta}k^{-1}})_{\xi h^\alpha,\xi k^\alpha}] \tag{5.35}$$

and

$$f(u_n) * (\rho\chi_{\xi^{-\beta}h^{-1},\xi^{-\beta}k^{-1}})_{\xi h^\alpha,\xi k^\alpha} \tag{5.36}$$

are respectively equal to $f(lu_n)$ and $lf(u_n)$, since u_n is constant on the rectangles $R_{i,j}$ and from the small size of the support $[-\xi^{-\beta+1}h^{\alpha-1}, \xi^{-\beta+1}h^{\alpha-1}] \times [-\xi^{-\beta+1}k^{\alpha-1}, \xi^{-\beta+1}k^{\alpha-1}]$ of the mollifier in (5.35) and (5.36).

In the vertical strip $S_i := [(i+\frac{1}{2})h - \xi^{-\beta+1}h^{\alpha-1}, (i+\frac{1}{2})h + \xi^{-\beta+1}h^{\alpha-1}] \times \mathbb{R}$ and in the horizontal strips $\mathbb{R} \times [(j+\frac{1}{2})k - \xi^{-\beta+1}k^{\alpha-1}, (j+\frac{1}{2})k + \xi^{-\beta+1}k^{\alpha-1}]$ centered at the interfaces of the rectangles $R_{i,j}$ the two functions $u_n * (\rho\chi_{\xi^{-\beta}h^{-1},\xi^{-\beta}k^{-1}})_{\xi h^\alpha,\xi k^\alpha}$ and (5.36) both present a mere junction due to the convolution by the positive function $(\rho\chi_{\xi^{-\beta}h^{-1},\xi^{-\beta}k^{-1}})_{\xi h^\alpha,\xi k^\alpha}$, between the constant values u_n and $lf(u_n)$ respectively considered above in the central parts of the rectangles. Therefore from (5.18) and (5.25) each of the two functions (5.35) and (5.36) has absolute values less than $const.h^{-N}$ (recall $h_n = h$) on these strips.

Taking into account these two kinds of domains : the union of the rectangles in the centers of the cells and the union of the strips, formula (5.28) gives

$$|I| \leq \int |f(lu_n(x,t)) - lf(u_n(x,t))||\psi(x,t)|dxdt + \int_{\bigcup strips} const.h^{-N}|\psi(x,t)|dxdt. \tag{5.37}$$

From (5.33) setting $\epsilon = const.\xi^{\beta.(2s-1)}h^{2s-1}$, then $l = 1 - \epsilon$. Therefore
$f(l.u_n(x)) - lf(u_n(x)) = f[(1-\epsilon)u_n(x)] - (1-\epsilon)f(u_n(x)) = -\epsilon f'(\ldots)u_n(x) + \epsilon f(u_n(x))$. From (5.18) and (5.25)

$$|f(lu_n(x)) - lf(u_n(x))| \leq const.\epsilon.h^{-(N-1)}h^{-1} + const.\epsilon h^{-N} \leq const.\xi^{\beta.(2s-1)}h^{2s-1}h^{-N}.$$

The number of horizontal strips S_i is less than $\frac{const}{h}$ from the compactness of the support of ψ, and each one has width $2\xi^{-\beta+1}h^{\alpha-1}$. Therefore the whole area of the domain of integration of the union of the vertical strips $\bigcup S_i$ is less than $\frac{const}{h}.2\xi^{-\beta+1}h^{\alpha-1}$. The same bound with k in place of h holds for the union of the horizontal strips. Therefore the second integral in (37) is less than $\frac{const}{h}\xi^{-\beta+1}h^{\alpha-1}h^{-N}$ since $k \leq h$ and since we will choose $\alpha \geq 2 + N$.

One obtains

$$|I| \leq const.\xi^{\beta(2s-1)}h^{2s-1-N} + \frac{const}{h}\xi^{-\beta+1}h^{\alpha-1}h^{-N},$$

which implies

$$|I| \leq const.max(\xi^{\beta(2s-1)}, \xi^{1-\beta}).max(h^{2s-1-N}, h^{\alpha-2-N}). \tag{5.38}$$

Since β has been chosen in $]0,1[, 0 < \xi < 1, 0 < h < 1$ the choices

$$s \geq \frac{1+N}{2}, \quad \alpha \geq 2 + N. \tag{5.39}$$

imply that $I \to 0$ uniformly in h when $\xi \to 0$.

• *Bound of* I_1. From (5.29), $I_1 = -\int (f(u_n))(x,t).(\rho - \rho\chi_{\xi^{-\beta}h^{-1},\xi^{-\beta}k^{-1}})_{\xi h^\alpha,\xi k^\alpha}(y,\tau).\psi(x+y,t+\tau)dxdydtd\tau$, which, from (5.25) and (5.23), implies

$$|I_1| \leq const(1 + \|u_n\|_\infty)^N \frac{1}{\xi h^\alpha.\xi k^\alpha} \int |(\rho - \rho\chi_{\xi^{-\beta}h^{-1},\xi^{-\beta}k^{-1}})(\tfrac{y}{\xi h^\alpha}, \tfrac{\tau}{\xi k^\alpha})|dyd\tau.$$

From (5.18),
$|I_1| \leq const.h^{-N}\int(\rho - \rho\chi_{\xi^{-\beta}h^{-1},\xi^{-\beta}k^{-1}})(\lambda)d\lambda \leq const.h^{-N}\xi^{\beta(2s-1)}h^{2s-1}$ from (5.33), i.e.

$$|I_1| \leq const.h^{2s-1-N}\xi^{\beta(2s-1)}. \tag{5.40}$$

Therefore in order that $I_1 \to 0$ uniformly in h, i.e. uniformly in h, when $\xi \to 0$ we choose

$$s \geq \frac{1+N}{2}. \tag{5.41}$$

• *Bound of* I_2. From (5.30), the mean value theorem gives

$|I_2| \leq \int |f'(\ldots)|.|u_n * (\rho - \rho\chi_{\xi^{-\beta}h^{-1},\xi^{-\beta}k^{-1}})_{\xi h^\alpha,\xi k^\alpha}(x,t)\psi(x,t)|dxdt.$

From (5.18) and (5.25),

$|I_2| \leq const.h^{-(N-1)} \int |u_n(x,t)(\rho - \rho\chi_{\xi^{-\beta}h^{-1},\xi^{-\beta}k^{-1}})_{\xi h^\alpha,\xi k^\alpha}(y,\tau)\psi(x+y,t+\tau)|dxdydtd\tau.$

Therefore, using the bound obtained above for I_1 with u_n instead of $f(u_n)$, i.e. one line before (5.40) with a bound h^{-1} instead of h^{-N}, we obtain

$$|I_2| \leq h^{-(N-1)}const.h^{-1}\xi^{\beta(2s-1)}h^{2s-1}, \tag{5.42}$$

which is same as (5.40). Finally, $I_2 \to 0$ uniformly in h when $\xi \to 0$ provided $s \geq \frac{N+1}{2}$.

• *Bound of* I_3. We have $\int[(f(u_n) * \rho_{\xi h^\alpha,\xi k^\alpha})(x,t)]\psi(x,t)dxdt = \int f(u_n)(x,t)\rho(\lambda,\mu)\psi(x+\xi h^\alpha\lambda, t+\xi k^\alpha\mu)dxd\lambda dtd\mu$.
Therefore, since $\int \rho(\lambda,\mu)d\lambda d\mu = 1$, it follows from (5.31) that
$I_3 = \int f(u_n(x,t))\rho(\lambda,\mu)[\psi(x+\xi h^\alpha\lambda, t+\xi k^\alpha\mu) - \psi(x,t)]dxd\lambda dtd\mu$

$$\leq const.h^{-N}\xi h^\alpha \int |\lambda\mu|\rho(\lambda,\mu)d\lambda d\mu$$

from (5.18)-(5.25) and since $k \leq h$. Then $I_3 \leq const.\xi.h^{\alpha-N}$. It suffices to have $\alpha \geq N$.□

5.5 Applications.

The consistency in the sense of distributions of the numerical scheme in Part I and chapter 6 provides examples of sequences of approximate solutions for which one can apply the theorem : a solution of the equations is exhibited by compactness as limit of a sequence of approximate solutions. This permits to put in evidence a solution of the Cauchy problems involving singular shocks if one admits that the properties to be checked to apply the theorem in chapter 6 for the Keyfitz-Kranzer system go on to hold indefinitely when $h \to 0$. Concerning delta shocks solutions of system (5.3)-(5.4) a full proof of consistency is given in chapter 6 ; then one can apply the theorem : $\forall u^0 \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R}), \forall v^0 \in L^1(\mathbb{R})$ $\exists U, V \in \mathcal{H}(\mathbb{R} \times \mathbb{R}^+)$ which are solutions of the equations in the sense (5.13) and are limits of the numerical scheme in chapter 6. The problem of finding criteria for uniqueness of these solutions remains open : one can only argue that they correspond to the limit of the scheme and it has always been observed that this limit is the correct known solution. The singular shocks show clearly that the classical functional spaces are inadequate in general to provide solutions of equations. This has justified the introduction of a new functional space. The results in this chapter as well as the numerical scheme in chapter 6 and its consistency proof extend clearly to 2-D and 3-D. These results show that in the context of the functional space of holomorphic germs weak asymptotic methods [12] can give rise to a

solution of the equations by compactness as limit of a subsequence extracted from the family of approximate solutions and can be applied to the syustems of fluid dynamics in Part I. In order to get closer to uniqueness the use of a stronger concept of weak solution could be useful : we could state (5.13) in the stronger form

$$\forall j = 1, ..., m,\ \forall \psi \in C_c^\infty(\mathbb{R}\times]0,T[),\quad \int_{\mathbb{R}\times]0,T[} [(U^j)_t + \frac{\partial}{\partial x}(\Phi(U))^j](x,t,\zeta)\psi(x,t)dxdt \to 0 \quad (5.43)$$

when $\zeta \to 0$ in the sector $|arg(\zeta)| < \theta$. We do not know if the theorem holds with this stronger formulation.

In the section below we give two examples in which some uniqueness holds at the level of explicit calculations (from chapter 4). In chapter 7 we give examples of existence-uniqueness of a different nature.

5.6 Examples from explicit calculations.

These examples are simplified versions of the contents of chapter 4 in order to make clear that some results of uniqueness could be possible in the context of holomorphic germs presented in this chapter. We did not succeed to extend them to the Cauchy problem in absence of explicit calculations.

• *First example : shock waves for nonconservative systems.* Consider the nonconservative system

$$u_t + (u^2)_x = v_x, \quad (5.44)$$

$$v_t + uv_x \overset{weak}{=} u_x, \quad (5.45)$$

in which the first equation is stated with the equality in $Lim\mathcal{H}_S$ while the second one is stated with the weak equality. Let us seek a solution in the form of a discontinuity moving with constant speed V, i.e. of the form

$$u(x,t) = u_l + \Delta u H_u(x - Vt), \quad (5.46)$$

$$v(x,t) = v_l + \Delta v H_v(x - Vt), \quad (5.47)$$

where $H_u, H_v \in Lim\mathcal{H}_S$ are Heaviside functions. Inserting (5.46)-(5.47) into (5.44) gives

$$\Delta u H_v = (-V + 2u_l)\Delta u H_u + (\Delta u)^2 (H_u)^2,$$

i.e.

$$H_v(z,\zeta) = -(V + 2u_l)\frac{\Delta u}{\Delta v} H_u(z,\zeta) + \frac{(\Delta u)^2}{\Delta v}(H_u)^2(z,\zeta) \quad (5.48)$$

as well as the classical relation obtained by setting $H_u(x,\xi) = 1, H_v(x,\xi) = 1$ in the formula above, which is nothing else than the classical jump condition of the conservative system (5.44). Equation (5.45) gives

$$-V\Delta v H_v' + (u_l + \Delta u H_u)\Delta v H_v' \overset{weak}{=} \Delta u H_u',$$

i.e.

$$\forall \psi \in \mathcal{C}_c^\infty(\mathbb{R}) \int \{-V\Delta v H_v'(x,\xi) + (u_l + \Delta u H_u(x,\xi))\Delta v H_v'(x,\xi) - \Delta u H_u'(x,\xi)\}\psi(x)dx \to 0 \tag{5.49}$$

when $\xi \to 0$. We recall that inserting (5.48) into (5.49), integrating in x and letting $\xi \to 0$ gives the second jump condition for system (5.44)-(5.45) which is in nonconservative form (see chapter 4).

The formula (5.48) and the two jump conditions imply that (5.44)-(5.45) is satisfied in its mixed strong-weak form. The interesting point is that this mixed strong-weak form fix the jump conditions which is some kind of existence-uniqueness limited to solutions of the form (5.46)-(5.47), presumably because of the limitations inherent to explicit calculations.

• *Second example : shock waves for the system of isothermal fluid dynamics.* Consider the system of isothermal fluid dynamics stated in the form (see chapter 4 for a justification)

$$\rho_t + (\rho u)_x = 0, \tag{5.50}$$

$$(\rho_u)_t + (\rho u^2)_x + p_x = 0, \tag{5.51}$$

$$p \overset{weak}{=} K\rho. \tag{5.52}$$

where K is a constant. We seek shock waves solutions of the usual form

$$\rho = \rho_l + \Delta\rho H_\rho(x - Vt), \tag{5.53}$$

$$\rho u = (\rho u)_l + \Delta(\rho u) H_{\rho u}(x - Vt), \tag{5.54}$$

$$p = p_l + \Delta p H_p(x - Vt). \tag{5.55}$$

Insertion of (5.53)-(5.54) into the continuity equation (5.50) gives

$$H_{\rho u} = H_\rho \tag{5.56}$$

and the classical jump condition for (5.50). Insertion of into the Euler equation (5.51) gives

$$H_p(z,\zeta) = V\Delta(\rho u)H_\rho(z,\zeta) - \frac{(\rho u)_l^2 + 2(\rho u)_l\Delta(\rho u)H_\rho(z,\zeta)(\Delta(\rho u))^2(H_\rho(z,\zeta))^2}{\rho_l + \Delta\rho H_\rho(z,\zeta)} + const. \tag{5.57}$$

Setting that the Heaviside functions are 0 for $x < 0$ and 1 for $x > 0$ gives the value of *constant* and the classical jump condition for (5.51). The last equation is stated in the weak sense as explained in chapter 4 since its statement in the strong sense would have led to inconsistencies. Then one has obtained that the two classical jump formulas, plus the relations (5.56)-(5.57) between the three Heaviside functions (that fix H_p and $H_{\rho u}$ as a function of H_ρ), plus the two formulas $p_l = K\rho_l, p_r = K\rho_r$ from (5.52) finally provide a solution of the system (5.50)-(5.52) where the first two equations are stated in the strong sense similarly as the result obtained in the first example.

Remark on the elimination of unstable discontinuities. The Heaviside functions $H(x,\xi)$ are analytic functions of the real variable x for each $\xi > 0$. Therefore they do not have an "infinite" slope at $x = 0$ as usual when H is considered in the space L^∞ (think at the function $\frac{1}{\xi}arctan(\frac{x}{\xi})$ which can be used to create Heaviside functions). Therefore unstable discontinuities from Heaviside functions are automatically eliminated at least concerning solutions in the (strong) sense

with $=$ in the space of holomorphic germs since their slope is already prepared in the initial condition with $H(x,\xi)$, $\xi \neq 0$. However we have been unable to transfer this remark into a general uniqueness result, perhaps because of the nonuniqeness of viscous solutions [1].

5.7 Conclusion.

These two examples give the impression that strong solutions do exist to some extent provided the system of N equations is well behaved (as this is the case for the equations of fluid dynamics considered above), with the statement of $N-1$ equations with the strong equality. Therefore results far stronger than the general existence of weak solutions shown in this section could presumably be obtained in particular cases.

Chapitre 6

Construction of approximate solutions.

In this chapter we present a numerical scheme for the approximation of singular shock solutions of the Keyfitz-Kranzer model system and many other systems of conservation laws. Consistency in the sense of distributions is studied. As long as some numerical properties are verified when the space step tends to 0, we prove that the scheme provides a numerical solution that satisfies the equations in the sense of distributions with an approximation that tends to 0 when $h \to 0$. We also show that this scheme adapts to degenerate systems. This is illustrated by two examples : the system presenting delta wave solutions originally studied by Korchinski and another system studied by Keyfitz-Kranzer that models elasticity. Consistence of the scheme in the sense of distributions is fully proved in the case of the Korchinski model.

6.1 Introduction.

Singular shocks have been discovered and investigated by different authors, see [22], [21], [33], [36], [35]. They have been observed from various viscosity techniques : Dafermos-Di Perna viscosity in [22], [21], usual viscosity in [33], [35]. In the case of singular shocks, viscosity solutions converge so weakly that their pointwise limits do not satisfy the classical Rankine-Hugoniot conditions. Besides this fact a unique entropic solution to the Riemann problem has been obtained in [22] for arbitrarily large data. In this chapter we propose a numerical scheme based on a splitting technique that captures the singular shocks. We observe results exactly similar to those obtained in [22], [33] with their respective viscosity techniques. Studies have shown the relevance of this scheme for other systems presenting irregular solutions. In our study of irregular shocks we consider two standard first order model systems of two equations whose solutions of the Riemann problem involve singular shocks and delta shocks. We also notice that this scheme provides neat results for the Keyfitz-Kranzer system of elasticity [23] for which the intrinsic difficulty is different from those in the two systems above.

This chapter focusses on the Keyfitz-Kranzer system

$$u_t + (u^2 - v)_x = 0, \tag{6.1}$$

$$v_t + (\frac{1}{3}u^3 - u)_x = 0, \tag{6.2}$$

which produces singular shocks, and the system

$$u_t + (u^2)_x = 0, \tag{6.3}$$

$$v_t + (uv)_x = 0, \tag{6.4}$$

originally considered by Korchinski [24] who discovered and investigated delta shocks in the solution of the Riemann problem.

Let u_h, v_h be the sequence of approximate solutions from the scheme. Under simple numerical properties to be rigorously proved, or to be admitted from numerical tests, we prove that the scheme is consistent in the sense of distributions in the following sense : $\forall(\phi,\psi) \in (\mathcal{C}_c^\infty(\mathbb{R}\times\mathbb{R}^+))^2$,

$$\int [u_h\phi_t + ((u_h)^2 - v_h)\phi_x]dxdt \to 0, \quad \int [v_h\psi_t + (\frac{1}{3}(u_h)^3 - u_h)\psi_x]dxdt \to 0, \tag{6.5}$$

respectively $\int [u_h\phi_t + ((u_h)^2)\phi_x]dxdt \to 0, \quad \int [v_h\psi_t + (u_hv_h)\psi_x]dxdt \to 0,$

when the space step $h \to 0$. This means that the functions u_h, v_h tend to satisfy the equations when $h \to 0$.

For system (6.1)-(6.2) we check numerically that the needed assumptions are satisfied for values of h as small as possible. We rigorously prove that, in the case of system (6.3)-(6.4), for any initial condition $u^0 \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ and $v^0 \in L^1(\mathbb{R})$, these assumptions are satisfied. Therefore the scheme is consistent in the above sense. Of course, in the first case, from a rigorous point of view, one cannot be sure that these numerical assumptions always hold for every h when $h \to 0$. The proof in this chapter shows that, for any given family of test functions with uniformly bounded support and uniformly bounded first and second derivatives, then a numerical solution satisfies the equations in the sense of distributions within a small deviation depending on h whenever these assumptions remain valid.

6.2 A numerical scheme.

The singular shocks of the Keyfitz-Kranzer equations are unbounded which makes the elaboration of numerical schemes difficult : in the scheme below the numerical velocity u in system (6.1)-(6.2) can be unbounded when the space step h tends to 0 which forces us to accept that the CFL coefficient r tends to 0 when $h \to 0$ in order to preserve the CFL condition $r\|u\|_{L^\infty} \leq 1$. Therefore $r = r_h$ depends on h and also on time so that $r_h\|u_h\|_{L^\infty} \leq 1$.

If r_h tends to 0 (i.e. if $\|u_h\|_{L^\infty}$ tends to ∞) slowly enough, then one can nevertheless obtain a convenient numerical scheme, although of an order less than one, on condition that for each iteration the assumptions are verified when $h \to 0$. This ensures consistence of the scheme in the sense of distributions, although the limit is not a distribution in general : it can be a singular shock in the case of the Keyfitz-Kranzer equations. Numerical results are given to prove that the set of assumptions is satisfied in representative situations of singular shocks. In the case of the Keyfitz-Kranzer equations the scheme consists in a splitting of equations into the two subsystems

$$u_t + (u^2)_x = 0, \tag{6.6}$$

$$v_t + (vu)_x = 0, \tag{6.7}$$

which is treated by transport with velocity u, and

$$u_t = v_x, \tag{6.8}$$

$$v_t = (vu - \frac{u^3}{3} + u)_x, \tag{6.9}$$

which is treated by a centered discretization. In between, we introduce an average step in u, v which is needed in general to avoid oscillations due to the centered discretization. More generally the method applies to systems

$$u_t + [u\Phi(u,v)]_x = [A(u,v)]_x, \tag{6.10}$$

$$v_t + [v\Phi(u,v)]_x = [B(u,v)]_x, \tag{6.11}$$

which are split into the two subsystems

$$u_t + [u\Phi(u,v)]_x = 0, \tag{6.12}$$

$$v_t + [v\Phi(u,v)]_x = 0, \tag{6.13}$$

where $\Phi(u,v)$ plays the role of numerical velocity and

$$u_t = [A(u,v)]_x, \tag{6.14}$$

$$v_t = [B(u,v)]_x. \tag{6.15}$$

Systems (6.12)-(6.13) is a family of degenerate systems. In particular the scheme in this chapter gives neat results for the system (4) in [23] which models an elastic string problem.

The numerical scheme. The real line is divided into intervals $I_i =]ih - \frac{1}{2}h, ih + \frac{1}{2}h[, i \in \mathbb{Z}$. We set $t_n = nrh$ for r small enough. We will construct step functions $u(x,t)$, $v(x,t)$ depending on h, which are constant on the rectangles $I_i \times]t_n, t_{n+1}[$, whose step values are denoted u_i^n, v_i^n respectively. The indices h are often skipped to simplify the notation : u stands for $u_h, ...$ If $a < b$ one sets

$$L(a,b) := length\ of\ [0,1] \cap [a,b], \tag{6.16}$$

i.e.

$$L(a,b) = max(0, min(1,b) - max(0,a)). \tag{6.17}$$

The notation L allows a synthetic formulation of the transport, without being forced to distinguish several cases depending on the signs of the numerical velocities. By induction we assume that the set of values $\{u_i^n, v_i^n\}_{i\in\mathbb{Z}}$ is known. We obtain the set of values $\{u_i^{n+1}, v_i^{n+1}\}_{i\in\mathbb{Z}}$ as follows.

- *First step : transport with velocity Φ during time rh*

$$\Phi_i^n := \Phi(u_i^n, v_i^n), \tag{6.18}$$

$$\overline{u}_i := u_{i-1}^n L(-1 + r\Phi_{i-1}^n, r\Phi_{i-1}^n) + u_i^n L(r\Phi_i^n, 1 + r\Phi_i^n) + u_{i+1}^n L(1 + r\Phi_{i+1}^n, 2 + r\Phi_{i+1}^n), \tag{6.19}$$

$$\overline{v}_i := v_{i-1}^n L(-1 + r\Phi_{i-1}^n, r\Phi_{i-1}^n) + v_i^n L(r\Phi_i^n, 1 + r\Phi_i^n) + v_{i+1}^n L(1 + r\Phi_{i+1}^n, 2 + r\Phi_{i+1}^n). \tag{6.20}$$

When the CFL condition $r|\Phi_i^n| \leq 1 \quad \forall i, \ \forall n$ is satisfied, the first terms in (6.19)-(6.20), when multiplied by h, represent the quantities u, v issued from the cell I_{i-1} between times t_n and t_{n+1} that lie in the cell I_i at time t_{n+1}. Indeed, the cell $I_{i-1} = [(i-\frac{3}{2})h, (i-\frac{1}{2})h]$ has been transported according to the vector $r\Phi_{i-1}^n h$, since Φ_{i-1}^n is the numerical velocity and the duration time is rh. The overlap with the fixed cell $I_i = [(i-\frac{1}{2})h, (i+\frac{1}{2})h]$ has a length of $r\Phi_{i-1}^n h$ if $\Phi_{i-1}^n \geq 0$, 0 if $\Phi_{i-1}^n \leq 0$, taking into account the CFL condition $r|\Phi_{i-1}^n| \leq 1$. From (6.16) one finds $L(-1 + r\Phi_{i-1}^n, r\Phi_{i-1}^n) = r\Phi_{i-1}^n$ if $\Phi_{i-1}^n \geq 0$, 0 if $\Phi_{i-1}^n \leq 0$. Division by h is due to the fact that $\overline{u}_i, u_j^n$ are mean values on cells of length h.

The second terms in (6.19)-(6.20), when multiplied by h, represent the quantities u, v issued from the cell I_i that remain in I_i at time t_{n+1}. Indeed, the cell $[(i-\frac{1}{2})h, (i+\frac{1}{2})h]$ has been transported by the vector $r\Phi_i^n h$. The overlap with the fixed cell $[(i-\frac{1}{2})h, (i+\frac{1}{2})h]$ is $h - r\Phi_i^n h$ if $\Phi_i^n \geq 0$, $h + r\Phi_i^n h$ if $\Phi_i^n \leq 0$. From (6.16) one finds $L(r\Phi_i^n, 1 + r\Phi_i^n) = 1 - r\Phi_i^n$ if $\Phi_i^n \geq 0$, $1 + r\Phi_i^n$ if $\Phi_i^n \leq 0$.

The third terms are similar to the first ones : they concern the quantities u, v issued from the cell I_{i+1} that lie in the cell I_i at time t_{n+1}, with the same verification as above. Note that $\overline{u}_i, \overline{v}_i$ depend on n, which is not explicitely stated to shorten the notation.

- *Averaging step.* For a value $\alpha, 0 \leq \alpha < 0.5$, to be chosen, we set

$$\widetilde{u}_i := \alpha \overline{u}_{i-1} + (1-2\alpha)\overline{u}_i + \alpha \overline{u}_{i+1}, \tag{6.21}$$

$$\widetilde{v}_i := \alpha \overline{v}_{i-1} + (1-2\alpha)\overline{v}_i + \alpha \overline{v}_{i+1}. \tag{6.22}$$

In the case $A = 0, B = 0$ the averaging step is useless. Indeed, the idea underlying the elaboration of the scheme is that the first step works well without averaging, and that the numerical defects of the centered discretization in the last step should be compensated by the averaging step performed before it. The splitting should be chosen so as to minimize the importance of the terms involved in the last step.

- *Last step : centered discretization*

$$u_i^{n+1} := \widetilde{u}_i + \frac{r}{2}[A(u_{i+1}^n, v_{i+1}^n) - A(u_{i-1}^n, v_{i-1}^n)], \tag{6.23}$$

$$v_i^{n+1} := \widetilde{v}_i + \frac{r}{2}[B(u_{i+1}^n, v_{i+1}^n) - B(u_{i-1}^n, v_{i-1}^n)]. \tag{6.24}$$

The scheme works well for singular shocks and delta shocks. The theorem below shows that it gives an approximate solution of the equations.

Statement of the theorem. Let $T > 0$ be given. Let us seek a solution on $\mathbb{R} \times [0, T]$. The initial conditions u^0, v^0 are discretized as usual by mean values in the cells since they are supposed to be L^1 functions. Let us apply the scheme under the assumptions (6.25)-(6.29) below : there exists a sequence of values h, $h \to 0$, a corresponding sequence of values r, $r > 0$, and real numbers $\beta, \gamma \in [0, 1[$ such that when $h \to 0$

$$\frac{h}{r} \to 0 \tag{6.25}$$

$$\forall n \leq \frac{T}{rh} \ \forall i \quad r|\Phi_i^n| \leq 1, \tag{6.26}$$

which is the CFL condition,

$$\forall n \leq \frac{T}{rh} \ \forall i \quad h^\beta|\Phi_i^n| = O(1), \tag{6.27}$$

which is a constraint on the numerical velocity allowing it to tend to infinity,

$$\forall n \leq \frac{T}{rh} \ \forall i \quad \sum_i |u_i^n|h = O(1), \ \sum_i |v_i^n|h = O(1), \tag{6.28}$$

which is the L^1-stability in u, v,

$$\forall n \leq \frac{T}{rh} \ \forall i \quad \sum_i |A(u_i^n, v_i^n)|h^{1+\gamma} = O(1), \sum_i |B(u_i^n, v_i^n)|h^{1+\gamma} = O(1). \tag{6.29}$$

Theorem 6.3.1. Consistency of the scheme. *As long as (6.25)-(6.29) are satisfied then the scheme is consistent on* $\mathbb{R}\times]0,T[$ *in the sense of distributions, i.e. if* u_h, v_h*, are the step functions from the scheme, then,* $\forall \psi \in \mathcal{C}_c^\infty(\mathbb{R}\times]0,T[)$,

$$\int [u_h\psi_t + u_h\Phi(u_h, v_h)\psi_x - A(u_h, v_h)\psi_x]dxdt \to 0, \tag{6.30}$$

$$\int [v_h\psi_t + v_h\Phi(u_h, v_h)\psi_x - B(u_h, v_h)\psi_x]dxdt \to 0, \tag{6.31}$$

when $h \to 0$. *More precisely the integrals in (30,31) are equal to*

$$O(\frac{h}{r}) + O(h^{1-\beta}) + O(h^{1-\gamma}). \tag{6.32}$$

The scheme will be of order one in the usual cases in which r is constant, $\beta = \gamma = 0$, but of an order strictly less than one for singular shocks from the fact that the values of the numerical velocity increase when $h \to 0$, which forces $r \to 0$ and $\beta > 0$.

6.3 Proof of the theorem.

One has $\int u\psi_t dxdt = \sum_{i,n} u_i^n \int_{cell_{i,n}} \psi_t dxdt = \sum_{i,n} u_i^n[(\psi_t)_i^n + O(h)]rh^2 = \sum_{i,n} u_i^n \frac{\psi_i^n - \psi_i^{n-1}}{rh} rh^2 + \sum_{i,n} u_i^n O(rh)rh^2 + \sum_{i,n} u_i^n O(h)rh^2$.
Since $|\sum_{i,n} u_i^n O(h)rh^2| \leq \sum_n rh \sum_i |u_i^n||O(h)|h \leq const.T|O(h)|$ from (6.28), one obtains

$$\int u\psi_t dxdt = \sum_{i,n}(u_i^n - u_i^{n+1})h\psi_i^n + O(h). \tag{6.33}$$

Similarly

$\int \Phi(u,v)u\psi_x dxdt = \sum_{i,n} \Phi_i^n u_i^n \int_{cell_{i,n}} \psi_x dxdt = \sum_{i,n} \Phi_i^n u_i^n (\psi_x)_i^n rh^2 + \sum_{i,n} \Phi_i^n u_i^n O(h)rh^2 = \sum_{i,n} \Phi_i^n u_i^n \frac{\psi_{i+1}^n - \psi_i^n}{h} rh^2 + \sum_{i,n} \Phi_i^n u_i^n O(h)rh^2 + \sum_{i,n} \Phi_i^n u_i^n O(h)rh^2$.

From (6.27)-(6.28) $|\sum_{i,n}\Phi_i^n u_i^n O(h) r h^2| \leq \sum_n rh \sum_i |\Phi_i^n||u_i^n||O(h)|h \leq const.Th^{-\beta}h \leq const.h^{1-\beta}$. Finally

$$\int \Phi(u,v)u\psi_x dxdt = -h\sum_{i,n}(\Phi_i^n u_i^n - \Phi_{i-1}^n u_{i-1}^n)r\psi_i^n + O(h^{1-\beta}). \tag{6.34}$$

Similarly

$\int A(u,v)\psi_x dxdt = \sum_{i,n} A(u_i^n, v_i^n)\int_{cell_{i,n}} \psi_x dxdt = \sum_{i,n} A(u_i^n, v_i^n)(\psi_x)_i^n rh^2 + \sum_{i,n} A(u_i^n, v_i^n)O(h)rh^2 = \sum_{i,n} A(u_i^n, v_i^n)\frac{\psi_{i+1}^n - \psi_i^n}{h} rh^2 + \sum_{i,n} A(u_i^n, v_i^n)O(h)rh^2$.

From (6.29) $|\sum_{i,n} A(u_i^n, v_i^n)O(h)rh^2| \leq \sum_n rh\sum_i |A(u_i^n, v_i^n)||O(h)|h \leq const.T.h^{-\gamma}h \leq const.h^{1-\gamma}$. Therefore

$$\int A(u,v)\psi_x dxdt = \sum_{i,n} rh[A(u_{i-1}^n, v_{i-1}^n) - A(u_i^n, v_i^n)]\psi_i^n + O(h^{1-\gamma}). \tag{6.35}$$

Setting

$$I := \int [u\psi_t + u\Phi(u,v)\psi_x - A(u,v)\psi_x]dxdt, \tag{6.36}$$

one finally obtains from (6.33)-(6.36)

$$I = -h\sum_{i,n}[u_i^{n+1} - u_i^n + r(u_i^n\Phi_i^n - u_{i-1}^n\Phi_{i-1}^n) -$$

$$r(A(u_i^n, v_i^n) - A(u_{i-1}^n, v_{i-1}^n))]\psi_i^n + O(h) + O(h^{1-\beta}) + O(h^{1-\gamma}). \tag{6.37}$$

Up to this point the formulas of the scheme have not yet been used. From (6.23) and (6.21)

$u_i^{n+1} = \overline{u}_i + \alpha(\overline{u}_{i-1} - 2\overline{u}_i + \overline{u}_{i+1}) + \frac{r}{2}[A(u_{i+1}^n, v_{i+1}^n) - A(u_{i-1}^n, v_{i-1}^n)]$.

Therefore, from (6.37)

$$I = I_1 + I_2 + I_3 + O(h) + O(h^{1-\beta}) + O(h^{1-\gamma}), \tag{6.38}$$

where

$$I_1 = -h\sum_{i,n}[\overline{u}_i - u_i^n + r(u_i^n\Phi_i^n - u_{i-1}^n\Phi_{i-1}^n)]\psi_i^n, \tag{6.39}$$

$$I_2 = -h\alpha\sum_{i,n}(\overline{u}_{i-1} - 2\overline{u}_i + \overline{u}_{i+1})\psi_i^n, \tag{6.40}$$

$$I_3 = -\frac{1}{2}\sum_{i,n} hr\{A(u_{i+1}^n, v_{i+1}^n) - A(u_{i-1}^n, v_{i-1}^n) - 2[A(u_i^n, v_i^n) - A(u_{i-1}^n, v_{i-1}^n)]\}\psi_i^n. \tag{6.41}$$

We are going to prove successively bounds for I_1, I_2, I_3.

• *Bound for* I_1. In I_1 fix an index i_0 and consider successively the two cases $\Phi_{i_0}^n \leq 0$ and $\Phi_{i_0}^n \geq 0$.

If $\Phi_{i_0}^n \leq 0$ then, from (6.16) and the CFL condition (6.26), $L(r\Phi_{i_0}^n, 1 + r\Phi_{i_0}^n) = 1 + r\Phi_{i_0}^n$, $L(1 + r\Phi_{i_0}^n, 2 + r\Phi_{i_0}^n) = -r\Phi_{i_0}^n$ and $L(-1 + r\Phi_{i_0}^n, r\Phi_{i_0}^n) = 0$. Therefore from (6.19)

$$\overline{u}_{i_0} = u^n_{i_0}(1 + r\Phi^n_{i_0}) + \text{terms not involving } u^n_{i_0},$$

$$\overline{u}_{i_0-1} = -u^n_{i_0} r\Phi^n_{i_0} + \text{terms not involving } u^n_{i_0},$$

$$\overline{u}_{i_0+1} \text{ does not involve } u^n_{i_0}.$$

From the CFL condition the other terms $\overline{u}_i$ do not involve $u^n_{i_0}$. Therefore, in the sum $\sum_i \overline{u}_i\psi^n_i$ the term $u^n_{i_0}$ occurs in (and only in)

$$u^n_{i_0}(1 + r\Phi^n_{i_0})\psi^n_{i_0} - u^n_{i_0} r\Phi^n_{i_0}\psi^n_{i_0-1}.$$

Consequently in the sum $\sum_i[\overline{u}_i - u^n_i + r(u^n_i\Phi^n_i - u^n_{i-1}\Phi^n_{i-1})]\psi^n_i$, the term involving $u^n_{i_0}$ is

$$u^n_{i_0}(1 + r\Phi^n_{i_0})\psi^n_{i_0} - u^n_{i_0} r\Phi^n_{i_0}\psi^n_{i_0-1} - u^n_{i_0}\psi^n_{i_0} + ru^n_{i_0}\Phi^n_{i_0}\psi^n_{i_0} - ru^n_{i_0}\Phi^n_{i_0}\psi^n_{i_0+1} \tag{6.42}$$

where the first two terms come from $\overline{u}_{i_0}$ and $\overline{u}_{i_0-1}$. The sum (6.42) is equal to $ru^n_{i_0}(\Phi^n_{i_0})[\psi^n_{i_0} - \psi^n_{i_0-1} + \psi^n_{i_0} - \psi^n_{i_0+1}] = ru^n_{i_0}\Phi^n_{i_0}O(h^2)$ from Taylor's formula applied to ψ.

If $\Phi^n_{i_0} \geq 0$ then, an analogous reasoning involving $\overline{u}_{i_0}$ and $\overline{u}_{i_0+1}$ instead of $\overline{u}_{i_0}$ and $\overline{u}_{i_0-1}$ gives the value 0. Therefore from (6.39)

$|I_1| \leq h\sum_{i_0,n} u^n_{i_0} r\Phi^n_{i_0} O(h^2) = \sum_n rh\sum_i \Phi^n_i u^n_i hO(h)$, i.e. from (6.27)-(6.28)

$$I_1 = O(h^{1-\beta}). \tag{6.43}$$

• *Bound for* I_2. From (6.40) $I_2 = -h\alpha\sum_{i,n}\overline{u}_i(\psi^n_{i+1} - 2\psi^n_i + \psi^n_{i-1}) = \alpha\sum_n rh\frac{1}{r}\sum_i \overline{u}_i O(h^2) = \alpha T\frac{h}{r}O(1)$ since one has $\sum_i |\overline{u}_i|h \leq \sum_i |u^n_i|h = O(1)$. Indeed, (6.19) implies the formula

$$|\overline{u}_i| \leq |u^n_{i-1}|L(-1 + r\Phi^n_{i-1}, r\Phi^n_{i-1}) + |u^n_i|L(r\Phi^n_i, 1 + r\Phi^n_i) + |u^n_{i+1}|L(1 + r\Phi^n_{i+1}, 2 + r\Phi^n_{i+1}). \tag{6.44}$$

The definition (6.16) of L implies $L(-1 + a, a) + L(a, 1 + a) + L(1 + a, 1 + 2a) = 1$. Therefore from (6.44) $\sum_i |\overline{u}_i| \leq \sum_i |u^n_i|$. This implies

$$I_2 = O(\frac{h}{r}). \tag{6.45}$$

• *Bound for* I_3. $I_3 = -\frac{hr}{2}\sum_{i,n}\{A(u^n_i, v^n_i)\psi^n_{i-1} - A(u^n_i, v^n_i)\psi^n_{i+1} - 2A(u^n_i, v^n_i)\psi^n_i + 2A(u^n_i, v^n_i)\psi^n_{i+1}\} = -\frac{1}{2}\sum_n rh\sum_i A(u^n_i, v^n_i)[\psi^n_{i-1} - 2\psi^n_i + \psi^n_{i+1}] = const.Th^{-\gamma}O(h)$ from Taylor's formula in ψ and . Therefore

$$I_3 = O(h^{1-\gamma}). \tag{6.46}$$

Finally from (6.38), (6.43), (6.45), (6.46)

$$I = O(h^{1-\beta}) + O(h^{1-\gamma}) + O(\frac{h}{r}), \tag{6.47}$$

which ends the proof.□

6.4 Approximation of the Keyfitz-Kranzer system.

We consider successively the three different typical solutions of Riemann problems in figures 8, 7, 6 in [33] : singular shock, intermediate overcompressive shock and usual shocks. The numerical solutions obtained from the scheme are identical to those shown in [33] even in absence of additional viscosity. We first consider the Riemann problem in figure 8 in [4], which shows a singular shock. The initial data is $(u_l, v_l, u_r, v_r) = (1.5, 0, -2.065426, 1.410639)$. We adopt the values $\alpha = 0.2, \beta = 0.5, \gamma = 0.4$. One chooses the value of r_h close to the maximum value of r that satisfies the CFL condition (6.26). For simplicity we denote

$$\text{"(27)"} := h^{\beta} max_{i,n} |u_i^n|,$$

$$\text{"(28)"} := max_n (\sum_i |u_i^n| h, \sum_i |v_i^n| h),$$

$$\text{"(29)"} := max_n (\sum_i |A(u_i^n, v_i^n)| h^{1+\gamma}, \sum_i |B(u_i^n, v_i^n)| h^{1+\gamma})$$

for the values in the assumptions of the theorem.

In order to check the consistence theorem, we present the values of $\frac{h}{r}$ that must tend to 0 from (6.25), and the values "27","28", "29" that must be bounded. Results of a test for $T = 5$ with the interval $[-4, 4]$ are given in the table below.

h	r	$\frac{h}{r}$	"(27)"	"(28)"	"(29)"
0.0400	0.300	0.1333	0.6289	14.97	3.62
0.0200	0.240	0.0833	0.5830	14.97	2.84
0.0100	0.170	0.0588	0.5309	14.96	2.26
0.0050	0.132	0.0379	0.5271	14.93	1.84
0.0025	0.095	0.0263	0.5178	14.90	1.53
0.00125	0.065	0.0192	0.5021	14.87	1.29
0.00062	0.040	0.0156	0.4326	14.85	1.10
0.00031	0.025	0.0125	0.4024	14.83	0.96

Now we choose $T = 1$ and the interval $[-0.5, 0.5]$ in order to reach smaller values of h. The values of the parameters are again $\alpha = 0.2, \beta = 0.5, \gamma = 0.4$

h	r	$\frac{h}{r}$	"(27)"	"(28)"	"(29)"
0.0020	0.18	0.0111	0.2444	1.9232	0.1791
0.0010	0.13	0.0077	0.2337	1.9170	0.1480
0.0005	0.09	0.0056	0.2225	1.9109	0.1252
0.00025	0.06	0.0042	0.2090	1.9054	0.1081
0.000125	0.043	0.0029	0.2070	1.8999	0.0973
0.0000833	0.035	0.0024	0.2051	1.8972	0.0926
0.0000625	0.030	0.0021	0.2028	1.8955	0.0898
0.0000500	0.026	0.0019	0.1979	1.8944	0.0874
0.0000333	0.021	0.0016	0.1955	1.8923	0.0848
0.0000250	0.019	0.0013	0.2010	1.8907	0.0847
0.0000166	0.015	0.0011	0.1957	1.8891	0.0829
0.0000125	0.012	0.0010	0.1847	1.8884	0.0803

In figure 6.4.1 one can observe that the scheme reproduces exactly the aspect of the singular shock in figure 8, in reference [33].

The values of r are chosen close to the maximum values for which the scheme satisfies the CFL condition $r\|u\|_\infty \leq 1$. One observes that the quantity $\frac{h}{r} \to 0$ as $\sqrt{h}$ and that the three quantities in the columns "27", "28","29" are bounded (since quantity "27" is proportional to the sup. of $|u|$ it is very sensitive to the chosen value of r close to the sup. of values of r that satisfy the CFL condition). Therefore, since $\beta = 0.5, \gamma = 0.4$, the scheme is of order 0.5 in h from (6.32). This is not a good result in general from a numerical viewpoint; however, the presence of singular shocks gives a numerical velocity which is of the order $\frac{1}{\sqrt{h}}$ instead of a constant in the usual situations in which the scheme is always of order 1. We can also see that the bounds in the proof of the theorem are not optimal since one has used a bound involving the factor $\|\Phi_i^n\|_\infty$ while $\Phi_i^n = u_i^n$ is uniformly bounded independently of time except on the singular shock. Indeed, one can see that the scheme gives acceptable results. On a standard PC top values of the peak in v in the above tests have reached the value 3700 for the Riemann problem under consideration while they have reached values 10^6 in the case of system (6.3)-(6.4). For the system (6.3)-(6.4) we will rigorously prove in section 6 that the scheme is of order 1. The set of results in these two tables gives a reasonable presumption that the decrease of $\frac{h}{r}$ and the boundedness of the three quantities "27","28","29" continue to hold when $h \to 0$, which would allow the theorem to be applied with confidence. If we only consider values of h for which (6.25)-(6.29) have been tested then the proof of the theorem gives a bound (depending on the sup norm of the derivatives of order two of ψ and its support) for the integrals in (6.30)-(6.31) according to (6.32).

Then we consider the Riemann problem $(u_l, v_l, u_r, v_r) = (1.5, 0, \quad 1.895644, 1.343466)$ represented in figure 7 in [33]. In this case we choose $\alpha = 0.2, \beta = 0, \gamma = 0$. We obtain the following table

h	r	$\frac{h}{r}$	"(27)"	"(28)"	"(29)"
0.0050	0.45	0.0111	1.9205	1.6913	1.2831
0.0010	0.45	0.0022	1.9205	1.6965	1.2753
0.0005	0.45	0.0011	1.9205	1.6972	1.2743
0.00025	0.45	0.0006	1.9205	1.6975	1.2739
0.000125	0.45	0.0003	1.9205	1.6977	1.2736
0.0000625	0.45	0.0001	1.9205	1.6977	1.2735

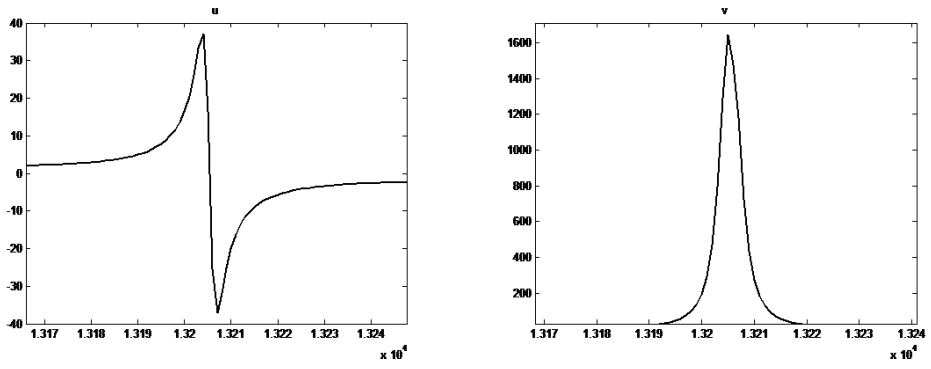


Figure 6.4.1. The numerical solution from the last test in the second table.

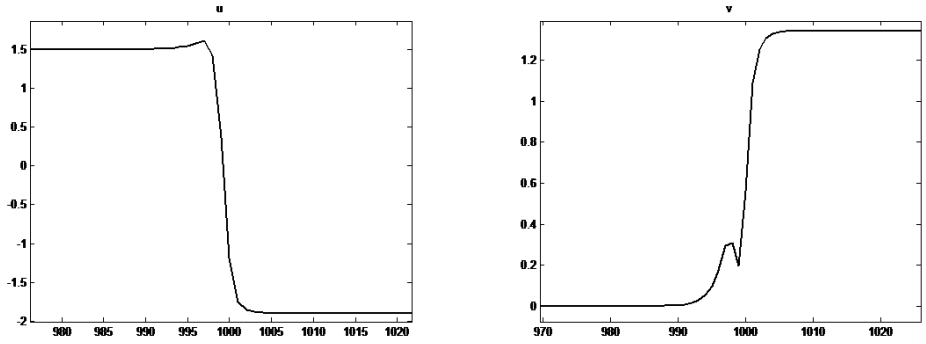


Figure 6.4.2. The numerical solution from the Riemann problem considered in the third table (h=0.04, r=0.45).

In figure 6.4.2 one can observe that the scheme in this chapter reproduces exactly the aspect of the limit overcompressive shock in figure 7 in [33]. An enlargement has been done in the horizontal direction to observe the detailed structure of the shock.

The results are very clear due to the boundedness of u in this case. There is a very natural presumption that these results continue to hold when $h \to 0$. One can see that the scheme is of order one in h as this follows from the theorem.

For the third Riemann problem, $(u_l, v_l, u_r, v_r) = (1.5, 0, \quad 1.725862, 1.276293)$ in figure 6 in [33], in which there is no singular shock, the results are very clear, exactly the same as those in the above table. We have always observed results as good in the case of bounded numerical velocity.

We now present a system for which a full proof of consistency in the sense of distributions has been obtained.

6.5 Application to the Korchinski system.

• One considers the 2×2 system (6.3)-(6.4) which produces delta-waves in the variable v, see [24]. Here $\Phi(u, v) = u$, $A = B = 0$. In this case one can choose $\alpha = 0$ in (6.21)-(6.22) since the last step (6.23)-(6.24) is absent. Then $u_i^{n+1} = \overline{u_i}, v_i^{n+1} = \overline{v_i}$; the choice $\alpha > 0$ works as well with the same proofs. It follows from (6.44) that $\sum_i |\overline{u}_i| \leq \sum_i |u_i^n|$. Therefore by induction on n $\sum_i |u_i^{n+1}| \leq \sum_i |u_i^0|$. The same proof applies for v. Choosing the initial condition u^0, v^0 in L^1 this proves (6.28). To prove (6.25)-(6.27) we will prove the maximum principle in the numerical velocity u.

Lemma 6.5.1. If $rmax_i |u_i^0| \leq \frac{1}{2}$ *then* u *satisfies the maximum principle.*

Proof. Let the index i be fixed. Consider the various possible combinations of signs in the three values $u_{i\ 1}^n, u_i^n, u_{i+1}^n$. In each case one will check that

$$min(u_{i\ 1}^n, u_i^n, u_{i+1}^n) \leq \overline{u}_i = u_i^{n+1} \leq max(u_{i\ 1}^n, u_i^n, u_{i+1}^n)$$

which proves the maximum principle by induction on n. By induction up to order n the condition $rmax_i |u_i^0| \leq \frac{1}{2}$ implies $rmax_i |u_i^n| \leq \frac{1}{2}$. Now we pass to order $n+1$.

• Case (+,+,+). Formula (6.19) with $\Phi = u$ gives

$$\overline{u}_i = u_{i-1}^n r u_{i-1}^n + u_i^n(1 - r u_i^n) = u_i^n + r(u_{i-1}^n - u_i^n)(u_{i-1}^n + u_i^n). \tag{6.48}$$

First note that $\overline{u}_i \geq 0$ because $1 - r u_i^n \geq 0$ from the property $r max_i |u_i^n| \leq \frac{1}{2}$. We consider successively the two cases $u_i^n \geq u_{i-1}^n$ and $u_i^n \leq u_{i-1}^n$. If $u_i^n \geq u_{i-1}^n$ then (6.48) gives $\overline{u}_i \leq u_i^n$. If $u_i^n \leq u_{i-1}^n$ then $\overline{u}_i - u_{i-1}^n = (u_i^n - u_{i-1}^n)[1 - r(u_i^n + u_{i-1}^n)] \leq 0$ since the last factor is ≥ 0 by induction. We have checked that

$$0 \leq \overline{u}_i \leq max(u_{i-1}^n, u_i^n).$$

• Case (+,+,-). Formula (6.19) gives

$$\overline{u}_i = u_{i-1}^n r u_{i-1}^n + u_i^n(1 - r u_i^n) + u_{i+1}^n(-r u_{i+1}^n). \tag{6.49}$$

First, let us prove that $\overline{u}_i \geq u_{i+1}^n$. The properties $u_{i-1}^n \geq 0, u_i^n \geq 0, r u_i^n \leq \frac{1}{2}$ imply that

$\overline{u}_i \geq u_{i+1}^n(-r u_{i+1}^n) \geq u_{i+1}^n$ since $0 \leq -r u_{i+1}^n \leq \frac{1}{2}$ and $u_{i+1}^n \leq 0$.

Now let us check that $\overline{u}_i \leq max(u_{i-1}^n, u_i^n)$. Formula (6.49) and $u_{i+1}^n \leq 0$ imply $\overline{u}_i \leq u_{i-1}^n r u_{i-1}^n + u_i^n(1 - r u_i^n)$. From this inequality the proof is the same as in the case (+,+,+).

• Case (-,+,+). Formula (6.19) gives $\overline{u}_i = u_i^n(1 - r u_i^n)$ which implies $\overline{u}_i \leq u_i^n$ since $0 \leq r u_i^n \leq \frac{1}{2}$ and, $\overline{u}_i \geq 0$.

• Case (-,+,-). Formula (6.19) gives

$$\overline{u}_i = u_i^n(1 - r u_i^n) + u_{i+1}^n(-r u_{i+1}^n) = u_i^n + r[-(u_{i+1}^n)^2 - (u_i^n)^2] \leq u_i^n.$$

Now $\overline{u}_i - u_{i+1}^n = u_i^n - u_{i+1}^n - r[(u_{i+1}^n)^2 + (u_i^n)^2]$. Since $u_i^n u_{i+1}^n \leq 0, (u_i^n)^2 + (u_{i+1}^n)^2 \leq (u_i^n)^2 + (u_{i+1}^n)^2 - 2u_i^n u_{i+1}^n = (u_i^n - u_{i+1}^n)^2$. Therefore $\overline{u}_i - u_{i+1}^n \geq u_i^n - u_{i+1}^n - r(u_i^n - u_{i+1}^n)^2 = (u_i^n - u_{i+1}^n)[1 - r(u_i^n - u_{i+1}^n)] \geq 0$ since the second factor is positive, which implies $\overline{u}_i \geq u_{i+1}^n$.

In the four cases in which $u_i^n \leq 0$ the verifications are similar.

Finally, we have proved properties (6.25)-(6.28), with r independent of h, $\beta = 0$ and $\gamma = 0$ since $A = B = 0$. Therefore from the theorem the scheme converges in the sense of distributions and is of order one in h. It has been checked numerically that its real interpretation is the well known solution.

6.6 Conclusion.

We have presented a numerical scheme which captures the singular shock solutions of the Keyfitz-Kranzer model without recourse to a vanishing viscosity method. We have observed numerically exactly the same results previously observed by the various authors. The consistency of the scheme for this system has been checked numerically up to very small values of h. The theorem states that the approximate solutions from the scheme tend to satisfy the equations in the sense of distributions. This scheme adapts to degenerate systems such as the Korchinski model system and the Keyfitz-Kranzer system of elasticity. In the case of the Korchinski system consistency in the sense of distributions has been fully proved.

Chapitre 7

Extension of Sobolev spaces.

Motivated by the need to find strong solutions in the setting of holomorphic germs for the Poisson equations involved in the systems of self-gravitating fluids, we introduce L^p spaces and Sobolev spaces in this setting. In contrast with their classical analogs, they contain very irregular distributions but as their classical analogs they permit to apply the Lax-Milgram theorem. This section is only sketched to show that the setting of germs is compatible with the classical linear results needed for a deeper future study. For simplification the results are given in one space dimension. It is clear that a large part of the classical theory can extend to the setting of holomorphic germs even in several space dimension, although this extension is not studied here.

7.1 The Dirichlet problem in the whole space.

The letters r, θ, μ will always denote real numbers such that

$$0 < r < 1,\ 0 < \theta < \frac{\pi}{2},\ 0 < \mu < 1. \tag{7.1}$$

If $\zeta \in \mathbb{C}, z \in \mathbb{C}$, one sets $\zeta = \xi + i\eta, z = x + iy,\ \xi, \eta \in \mathbb{R},\ x, y \in \mathbb{R}$. One sets

$$S(r, \theta, \mu) = \{(z, \zeta) \in \mathbb{C} \times \mathbb{C}; 0 < |\zeta| < r,\ -\theta < arg\zeta < \theta, |y| < \mu Re(\zeta)\}. \tag{7.2}$$

Lemma 7.1.1. Let $1 \leq p \leq \infty, N \in \mathbb{N}$ and let $\{F_\alpha\}$ be a family of holomorphic functions on an open set $\Omega \subset \mathbb{C} \times \mathbb{C}$ such that

$$\int_\Omega |\zeta^N F_\alpha(z, \zeta)|^p dxdyd\xi d\eta \leq C_1 \tag{7.3}$$

where C_1 is a constant independent on α. Let K be a compact subset of Ω. Then there exists a constant C_2 independent on α such that

$$\forall (z, \zeta) \in K \quad |\zeta^N F_\alpha(z, \zeta)| \leq C_2. \tag{7.4}$$

proof. Let us start by proving that (7.3) for some p implies (7.3) for $p = 1$ when integration is restricted to a compact set. Let K' be a compact set in Ω containing K in its interior. If $\frac{1}{p} + \frac{1}{q} = 1$, Hölder's formula gives :

$$\int_{K'} |\zeta^N F(z, \zeta)| \leq (\int_{K'} 1)^{\frac{1}{q}} (\int_\Omega |\zeta^N F(z, \zeta)|^p)^{\frac{1}{p}} \leq const.$$

The bound (7.4) will follow from the mean formula

$$|f(z_0,\zeta_0)| \leq \frac{1}{\pi^2\epsilon^4}\int_{|z-z_0|\leq\epsilon,|\zeta-\zeta_0|\leq\epsilon} |f(z,\zeta)|dxdyd\xi d\eta$$

for a holomorphic function f in a neighborhood of the polydisc $|z-z_0| \leq \epsilon, |\zeta-\zeta_0| \leq \epsilon$. We apply this formula with $(z_0,\zeta_0) \in K$, $\epsilon <$ the distance from K to the complement of K' and $f(z,\zeta) = \zeta^N F(z,\zeta)$.□

If $S := S(r,\theta,\mu)$ and if $N \in \mathbb{N}$ one defines $\mathcal{S}_{S,N}$ as the set of all $F : S \longmapsto \mathbb{C}$, holomorphic such that

$$\forall n \in \mathbb{N}\ \exists const > 0 \mid \forall\ (z,\zeta) \in S,\ |F(z,\zeta)| \leq \frac{const}{|\zeta|^N}\frac{1}{(1+|x|^n)}. \tag{7.5}$$

$\mathcal{S}_{S,N}$ is an infinite dimensional vector space. Indeed consider functions $F(z)$ which are in the image of $\mathcal{C}_c^\infty(\mathbb{R})$ through the Fourier transform. One can also consider $\rho(z) = \frac{exp(-z^2)}{\sqrt{\pi}}$ and $F(z,\zeta) = \frac{1}{\zeta}\rho(\frac{z}{\zeta})$: for r,θ,μ small enough one can check that $F \in \mathcal{S}_{S,1}$ and is a Dirac delta function. In $\mathcal{S}_{S,N}$ one considers the sesquilinear form

$$< F,G >_{\mathcal{L}^2_{S,N}} := \int_S |\zeta|^{2N} F(z,\zeta)\overline{G}(z,\zeta)dxdyd\xi d\eta \tag{7.6}$$

which is a Hermitian scalar product on $\mathcal{S}_{S,N}$. Let $\mathcal{L}^2_{S,N}$ be the completion of $\mathcal{S}_{S,N}$ for this scalar product. Therefore $\mathcal{L}^2_{S,N}$ is a (complex) Hilbert space and $\mathcal{S}_{S,N}$ is dense in $\mathcal{L}^2_{S,N}$.

Proposition 7.1.1. $\mathcal{L}^2_{S,N}$ is made with holomorphic functions on S.

proof. If $F \in \mathcal{L}^2_{S,N}$, from a property of the completion, there is a sequence $(F_n)_n$ in $\mathcal{S}_{S,N}$ which converges to F for the norm $\|\ \|_{\mathcal{L}^2_{S,N}}$ in the Hilbert space $\mathcal{L}^2_{S,N}$. Therefore $\|F_n\|_{\mathcal{L}^2_{S,N}}$ is bounded uniformly in n. Applying Lemma 7.1.1 to transform an integral bound into a sup. bound one obtains that $\{F_n\}$ is a normal family [32]. Therefore there is a subsequence that converges uniformly on compact subsets of S. Therefore F is holomorphic in S.□

One defines an analog of the classical Sobolev space $\mathcal{H}^1$ by stating : $\mathcal{H}^1_{S,N}$ is the set of all maps $F : S \longmapsto \mathbb{C}$ such that $F \in \mathcal{L}^2_{S,N}$ and $\frac{dF}{dx} \in \mathcal{L}^2_{S,N}$. $\mathcal{H}^1_{S,N}$ is equipped with the scalar product

$$< F,G >_{\mathcal{H}^1_{S,N}} := < F,G >_{\mathcal{L}^2_{S,N}} + < \frac{dF}{dx}, \frac{dG}{dx} >_{\mathcal{L}^2_{S,N}}\ . \tag{7.7}$$

Proposition 7.1.2. $\mathcal{H}^1_{S,N}$ is a Hilbert space.

proof. Let $(F_n)_n$ be a Cauchy sequence in $\mathcal{H}^1_{S,N}$. Since $\mathcal{L}^2_{S,N}$ is complete there are $F,G \in \mathcal{L}^2_{S,N}$ such that $F_n \to F$, $\frac{\partial F_n}{\partial x} \to G$ in $\mathcal{L}^2_{S,N}$. Using the same proof as in Proposition 7.1.1 $(F_n)_n$ is a normal family, therefore, from [32], $G = \frac{\partial F}{\partial x}$.□

One defines the subspace $\mathcal{H}^{0,1}_{S,N}$ of $\mathcal{H}^1_{S,N}$ as the closure of $\mathcal{S}_{S,N}$ in $\mathcal{H}^1_{S,N}$. $\mathcal{H}^{0,1}_{S,N}$ is a Hilbert space for the scalar product induced by $\mathcal{H}^1_{S,N}$. If $u,v \in \mathcal{H}^{0,1}_{S,N}$ the classical integration by parts

formula $< \frac{\partial u}{\partial x}, v >_{\mathcal{L}^2_{S,N}} = - < u, \frac{\partial v}{\partial x} >_{\mathcal{L}^2_{S,N}}$ holds by continuation of the same formula in $\mathcal{S}_{S,N}$.

Lemma 7.1.2. The map $j : \mathcal{L}^2_{S,N} \longmapsto (\mathcal{H}^{0,1}_{S,N})', f \longmapsto (u \mapsto < f, \overline{u} >_{\mathcal{L}^2_{S,N}})$ *is linear continuous and injective.*

proof. $| < f, \overline{u} >_{\mathcal{L}^2_{S,N}} | \leq \|f\|_{\mathcal{L}^2_{S,N}} \|u\|_{\mathcal{L}^2_{S,N}}$ therefore $j(f) \in (\mathcal{H}^{0,1}_{S,N})'$ with operator norm $\leq \|f\|_{\mathcal{L}^2_{S,N}}$. Therefore the linear map j has norm ≤ 1. It is injective from the density of $\mathcal{S}_{S,N}$ into $\mathcal{L}^2_{S,N}$.□

Denoting this map as an inclusion one has the sequence of continuous inclusions :

$$\mathcal{S}_{S,N} \subset \mathcal{H}^{0,1}_{S,N} \subset \mathcal{L}^2_{S,N} = (\mathcal{L}^2_{S,N})' \subset (\mathcal{H}^{0,1}_{S,N})'. \tag{7.8}$$

The following spaces are defined as inductive limits in the category of vector spaces

$$\mathcal{L}^2(\mathbb{R}) := \lim_{\rightarrow} \mathcal{L}^2_{S,N},\ \mathcal{H}^{0,1}(\mathbb{R}) := \lim_{\rightarrow} \mathcal{H}^{0,1}_{S,N} \tag{7.9}$$

when $r \to 0,\ \theta \to 0,\ \mu \to 0,\ N \to +\infty$. One has

$$\mathcal{H}^{0,1}(\mathbb{R}) \subset \mathcal{L}^2(\mathbb{R}). \tag{7.10}$$

As in the classical setting one can use the Lax-Milgram theorem (here in the complex case [16]).

A standard model equation. Consider the model equation

$$-u''(x) + c(x)u(x) = f(x),\ u(-\infty) = 0 = u(+\infty) \tag{7.11}$$

where $f \in \mathcal{L}^2(\mathbb{R})$. For instance f can be a Dirac delta function : this space $\mathcal{L}^2(\mathbb{R})$ is very different from the classical space of square integrable functions since it contains objects such as the Dirac delta function and its derivatives. The function c is assumed to be holomorphic and bounded on some set S and have some positiveness property ; it can be a step function representing Heaviside functions).

Lemma 7.1.3. cu is in $\mathcal{L}^2_{S,N}$.

proof. There exists a sequence (u_n) in $\mathcal{S}_{S,N}$ such that $\|u_n - u\|_{\mathcal{L}^2_{S,N}} \to 0$ since $u \in \mathcal{L}^2_{S,N}$ and $\mathcal{S}_{S,N}$ is dense in $\mathcal{L}^2_{S,N}$. $cu_n \in \mathcal{S}_{S,N}$ from the assumption that c is bounded on S. $\|cu_n - cu\|^2_{\mathcal{L}^2_{S,N}} = \int_S |\zeta|^{2N} |cu_n - cu|^2 \leq \|c\|^2_\infty \|u_n - u\|^2_{\mathcal{L}^2_{S,N}} \to 0$, therefore $cu \in \mathcal{L}^2_{S,N}$ from the definition of $\mathcal{L}^2_{S,N}$.□

Let $V := \mathcal{H}^{0,1}_{S,N}$ equipped with the scalar product of $\mathcal{H}^1_{S,N}$. If $u, v \in V$ set

$$a(u,v) = < u', v' >_{\mathcal{L}^2_{S,N}} + < cu, v >_{\mathcal{L}^2_{S,N}},\ L(v) = < f, v >_{\mathcal{L}^2_{S,N}} . \tag{7.12}$$

One has the familiar bounds

$$|a(u,v)| \leq const \|u\|_V \|v\|_V,\ |L(v)| \leq const \|v\|_V. \tag{7.13}$$

It suffices to use the assumption that c is bounded on S.

From (7.12) and (7.6)

$a(u,u) = \int_S |\zeta|^{2N} |u'(z,\zeta)|^2 dxdyd\xi d\eta + \int_S |\zeta|^{2N} c(z,\zeta)|u(z,\zeta)|^2 dxdyd\xi d\eta.$

One adopts a positiveness assumption on c :

$$\exists \gamma > 0, \quad Re(c(z,\zeta)) \geq \gamma \ \forall (z,\zeta) \in S. \tag{7.14}$$

This implies that

$$|a(u,u)| \geq const \|u\|_V^2. \tag{7.15}$$

Therefore one can apply the Lax Milgram theorem in the complex case [16] :

$$\exists ! u \in V; \ a(u,v) = L(v) \ \forall v \in V \tag{7.16}$$

i.e. from the validity of the integration by parts in $\mathcal{L}^2_{S,N}$, $< -u'' + cu - f, v >_{\mathcal{L}^2_{S,N}} = 0 \ \forall v \in \mathcal{H}^{0,1}_{S,N}$, which is equivalent to $u'' = cu - f$ in $(\mathcal{H}^{0,1}_{S,N})'$. Since $cu - f \in \mathcal{L}^2_{S,N} = (\mathcal{L}^2_{S,N})' \subset (\mathcal{H}^{0,1}_{S,N})'$, then $u'' = cu - f$ in $\mathcal{L}^2_{S,N}$. Is such a u unique ?

Let $u_1, u_2 \in \mathcal{H}^{0,1}(\mathbb{R})$ be such that $u_i'' = cu_i - f$ in $\mathcal{L}^2(\mathbb{R}), i = 1,2$. From the definitions (7.9) as inductive limits $\exists \ S, N$ such that $u_1, u_2 \in \mathcal{H}^{0,1}_{S,N}$. Choosing S small enough and N large enough c satisfies (7.14) on S and is bounded on S, $f \in \mathcal{L}^2_{S,N}$. Setting $V := \mathcal{H}^{0,1}_{S,N}$, both u_1, u_2 satisfy (7.16), therefore they are equal in $\mathcal{H}^{0,1}_{S,N}$, therefore in $\mathcal{L}^2(\mathbb{R})$. One has proved :

Theorem 7.1.1. *Under the above boundedness and positiveness assumptions on c, for any $f \in \mathcal{L}^2(\mathbb{R})$ there is a unique $u \in \mathcal{H}^{0,1}(\mathbb{R})$ such that $-u'' + cu = f$ in $\mathcal{L}^2(\mathbb{R})$.*

7.2 Periodic problems.

Let the period T be given. Here we set $\mathcal{S}^{per}_{S,N} := \{F : S \longmapsto \mathbb{C}$, holomorphic, periodic of period T in x, such that

$$\exists const > 0 \ ; \ \ |F(z,\zeta)| \leq \frac{const}{|\zeta|^N} \ \forall \ (z,\zeta) \in S\}. \tag{7.17}$$

In $\mathcal{S}^{per}_{S,N}$ one considers the sesquilinear form

$$< F, G >_{\mathcal{L}^{2,per}_{S,N}} := \int_{S \cap \{x \in [0,T]\}} |\zeta|^{2N} F(z,\zeta) \overline{G}(z,\zeta) dxdyd\xi d\eta \tag{7.18}$$

which is an Hermitian scalar product on $\mathcal{S}^{per}_{S,N}$.

Let $\mathcal{L}^{2,per}_{S,N}$ be the completion of $\mathcal{S}^{per}_{S,N}$ for this scalar product. Therefore $\mathcal{L}^{2,per}_{S,N}$ is a (complex) Hilbert space and $\mathcal{S}^{per}_{S,N}$ is dense in $\mathcal{L}^{2,per}_{S,N}$.

Proposition 7.2.1. $\mathcal{L}^{2,per}_{S,N}$ is made of holomorphic functions on S.

proof. It is the same as the proof of Proposition 7.1.1.□

One defines $\mathcal{H}^{1,per}_{S,N}$ as the closure of $\mathcal{S}^{per}_{S,N}$ in $\mathcal{L}^{2,per}_{S,N}$ for the scalar product

$$< F, G >_{\mathcal{H}^{1,per}_{S,N}} := < F, G >_{\mathcal{L}^{2,per}_{S,N}} + < \frac{dF}{dx}, \frac{dG}{dx} >_{\mathcal{L}^{2,per}_{S,N}} .$$

Therefore $\mathcal{H}^{1,per}_{S,N}$ is a Hilbert space. One has the inclusions

$$\mathcal{S}^{per}_{S,N} \subset \mathcal{H}^{1,per}_{S,N} \subset \mathcal{L}^{2,per}_{S,N}$$

with dense inclusions. Therefore as usual one has the inclusions :

$$\mathcal{S}^{per}_{S,N} \subset \mathcal{H}^{1,per}_{S,N} \subset \mathcal{L}^{2,per}_{S,N} = (\mathcal{L}^{2,per}_{S,N})' \subset (\mathcal{H}^{1,per}_{S,N})'. \tag{7.19}$$

The following spaces are defined as inductive limits in the category of vector spaces

$$\mathcal{L}^{2,per}(\mathbb{R}) := \lim_{\rightarrow} \mathcal{L}^{2,per}_{S,N}, \ \mathcal{H}^{1,per}(\mathbb{R}) := \lim_{\rightarrow} \mathcal{H}^{1,per}_{S,N} \tag{7.20}$$

when $r \to 0,\ \theta \to 0,\ \mu \to 0,\ N \to +\infty$. One has

$$\mathcal{H}^{1,per}(\mathbb{R}) \subset \mathcal{L}^{2,per}(\mathbb{R}). \tag{7.21}$$

As in the classical setting one can use the Lax-Milgram theorem in the complex case [16]. This yields results of existence and uniqueness of solutions of equations without classical solutions.

A standard model equation. Consider the model equation

$$-u''(x) + c(x)u(x) = f(x), \quad u(-\infty) = 0 = u(+\infty) \tag{7.22}$$

where $f \in \mathcal{L}^{2,per}(\mathbb{R})$ is periodic with period T. The statement of the periodic problem is : find v holomorphic on some strip S, periodic with period T, such that

$$-v'' + cv = f. \tag{7.23}$$

The variational formulation

$$\int_{S\cap\{x\in[0,T]\}} |\zeta|^{2N}(v'w' + cvw) = \int_{S\cap\{x\in[0,T]\}} |\zeta|^{2N} fw \ \forall w \in \mathcal{L}^{2,per}_{S,N} \tag{7.24}$$

follows from the integration by parts formula $\int_0^T v''w = [v'w]_0^T - \int_0^T v'w' = \int_0^T v'w'$ from the periodicity. We set

$$a(u,v) = \int_{S\cap\{x\in[0,T]\}} |\zeta|^{2N}(u'v' + cuv), \quad L(u,v) = \int_{S\cap\{x\in[0,T]\}} fv. \tag{7.25}$$

Under the boundedness and positiveness assumptions on c (7.14) one has again existence and uniqueness of a solution of (7.23) similarly as in Theorem 7.1.1.

Theorem 7.2.1 : periodic problem. *Under the boundedness and positiveness above assumptions on c, for any $f \in \mathcal{L}^{2,per}(\mathbb{R})$ which is periodic of period T there is a unique $u \in \mathcal{H}^{1,per}(\mathbb{R})$, periodic with period T solution of -u"+cu=f.*

7.3 Dirichlet problem on a finite interval [a,b].

Let $a, b \in \mathbb{R}, a < b, N \in \mathbb{N}$ and

$S(r,\theta,\mu) = \{(z,\zeta) \in \mathbb{C} \times \mathbb{C}$ such that

$$a < x < b,\ 0 < |\zeta| < r,\ -\theta < arg\zeta < \theta, |y| < \mu Re(\zeta)\}. \tag{7.26}$$

Let $\Sigma = \{(y,\zeta)\ 0 < |\zeta| < r,\ -\theta < arg\zeta < \theta, |y| < \mu Re(\zeta)\}$. We set (with abusively same notation as in section 7.1 since a, b are omitted there).

$\mathcal{S}_S = \{F : S \longrightarrow \mathbb{C},\ (z,\zeta) \to F(z,\zeta)$ holomorphic, continuous on the set $\{(z,\zeta); a \leq x \leq b,\ 0 < |\zeta| < r,\ -\theta < arg\zeta < \theta, |y| < \mu Re(\zeta)\}$, such that i), ii) and iii) below hold$\}$:

$$i)\ sup_{(z,\zeta)\in S}|F(z,\zeta)| < +\infty \tag{7.27}$$

$$ii)\ F(a+iy,\zeta) = F(b+iy,\zeta)\ \forall (y,\zeta) \in \Sigma \tag{7.28}$$

$$iii)\ lim_{r\to 0,\theta\to 0,\mu\to 0}(\frac{1}{mes(\Sigma)}\int_{(y,\zeta)\in\Sigma} F(a+iy,\zeta)dyd\xi d\eta) = 0, \tag{7.29}$$

where $mes(\Sigma)$ denotes the Lebesgue measure of the set Σ; the same formula with b in the place of a follows from ii). With a=0,b=π polynomials in $sin(z)$ with null constant term show that this space is infinite dimensional.

The properties (7.28)-(7.29) replace here the classical statement of null values at the points a and b. Let $N \in \mathbb{N}$ be given. On $\mathcal{S}_S$ we define the sesquilinear form

$$< F, G >_{\mathcal{L}^2_{S,N}} := \int_S |\zeta|^{2N} F(z,\zeta)\overline{G(z,\zeta)}dxdyd\xi d\eta. \tag{7.30}$$

This sesquilinear form is a Hermitian scalar product : $< F, F >_{\mathcal{L}^2_{S,N}} = 0 \Rightarrow F = 0$. We note $\mathcal{S}_{S,N}$ the preHilbert space thus obtained.

Definition 7.3.1. $\mathcal{L}^2_{S,N}$:=*the completion of the preHilbert space* $(\mathcal{S}_S, < F, G >_{\mathcal{L}^2_{S,N}})$.

Proposition 7.3.1. $\mathcal{L}^2_{S,N}$ *is made of holomorphic functions on* S.

The proof follows from the lemma 1 as the proof of Proposition 7.1.1.□

Definition 7.3.2. $\mathcal{H}^1_{S,N} := \{F : S \to \mathbb{C}$ *such that* F *and* $\frac{\partial F}{\partial x} \in \mathcal{L}^2_{S,N}\}$ *equipped with the scalar product*

$$< F, G >_{\mathcal{H}^1_{S,N}} := < F, G >_{\mathcal{L}^2_{S,N}} + < \frac{\partial F}{\partial x}, \frac{\partial G}{\partial x} >_{\mathcal{L}^2_{S,N}}. \tag{7.31}$$

Proposition 7.3.2. $\mathcal{H}^1_{S,N}$ *is complete, i.e. it is a Hilbert space.*

proof. It is the same as the proof of Proposition 7.1.2.□

Proposition 7.3.3. If $v \in \mathcal{H}^1_{S,N}$ the map $V :]a,b[\to \mathbb{C}$,

$$x \to V(x) = \int_{|\zeta|<r, |arg\zeta|<\theta, |y|<\mu\xi} \zeta^N v(x+iy,\zeta) dy d\xi d\eta \tag{7.32}$$

can be continuously extended to $[a,b]$. Further,

$$sup_{x\in[a,b]} \int_{|\zeta|<r, |arg\zeta|<\theta, |y|<\mu\xi} |\zeta^N v(x+iy,\zeta)| dy d\xi d\eta \leq const \|v\|_{\mathcal{H}^1_{S,N}}. \tag{7.33}$$

proof. Let $\overline{V}(x) = \int_{y,\zeta} \int_{t=\frac{a+b}{2}}^{x} \zeta^N \frac{\partial v}{\partial x}(t+iy,\zeta) dt dy d\xi d\eta$.
The map $\overline{V}$ is defined if $x \in [a,b]$ since $\int_{y,\zeta} \int_{t=\frac{a+b}{2}}^{x} |\zeta^N \frac{\partial v}{\partial x}(t+iy,\zeta)| dt dy d\xi d\eta \leq (\int_S 1^2)^{\frac{1}{2}} (\int_S |\zeta^N v|^2)^{\frac{1}{2}} \leq const. \|v\|_{\mathcal{L}^2_{S,N}}$.
Then, again from the Cauchy-Schwarz inequality,
$|\overline{V}(x_1) - \overline{V}(x_2)| = |\int_{y,\zeta} \int_{t=x_2}^{x_1} \zeta^N \frac{\partial v}{\partial x}(t+iy,\zeta) dt dy d\xi d\eta| \leq const |x_1 - x_2|^{\frac{1}{2}} (\int_{y,\zeta} \int_{t=x_2}^{x_1} |\zeta^N \frac{\partial v}{\partial x}(t+iy,\zeta)|^2 dt dy d\xi d\eta)^{\frac{1}{2}} \leq const |x_1 - x_2|^{\frac{1}{2}} \|\frac{\partial v}{\partial x}\|_{\mathcal{L}^2_{S,N}}$.

Therefore $\overline{V}$ is (uniformly) continuous on $[a,b]$. The functions V and $\overline{V}$ are defined on $]a,b[$ and have the same x-derivatives. Therefore their difference is constant. Since $\overline{V}$ is defined and continuous on $[a,b]$, so is V. Now let us prove the bound (7.32). The formula $v(x_1+iy,\zeta) = v(x_2+iy,\zeta) + \int_{x_1}^{x_2} \frac{\partial v}{\partial x}(t+iy,\zeta) dt$ implies that

$$|v(x_1+iy,\zeta)| \leq |v(x_2+iy,\zeta)| + \int_{x_1}^{x_2} |\frac{\partial v}{\partial x}(t+iy,\zeta)| dt.$$

Integration gives :

$$\int_{y,\zeta} |\zeta^N v(x_1+iy,\zeta)| dy d\xi d\eta \leq \int_{y,\zeta} |\zeta^N v(x_2+iy,\zeta)| dy d\xi d\eta + \int_S |\zeta^N \frac{\partial v}{\partial x}(t+iy,\zeta)| dt dy d\xi d\eta.$$

Applying the Cauchy-Schwarz inequality in the last integral we obtain

$$\int_{y,\zeta} |\zeta^N v(x_1+iy,\zeta)| dy d\xi d\eta \leq \int_{y,\zeta} |\zeta^N v(x_2+iy,\zeta)| dy d\xi d\eta + (\int_S 1 dt dy d\xi d\eta)^{\frac{1}{2}} (\int_S |\zeta^N \frac{\partial v}{\partial x}(t+iy,\zeta)|^2 dt dy d\xi d\eta)^{\frac{1}{2}}.$$

The last integral is equal to $const \|\frac{\partial v}{\partial x}\|_{\mathcal{L}^2_{S,N}}$. Integration in x_2 on $[a,b]$ of the inequality so obtained and use of the Cauchy-Schwarz inequality in the first integral in second member give

$$(b-a) \int_{y,\zeta} |\zeta^N v(x_1+iy,\zeta)| dy d\xi d\eta \leq const \|v\|_{\mathcal{L}^2_{S,N}} + const \|\frac{\partial v}{\partial x}\|_{\mathcal{L}^2_{S,N}} \leq const \|v\|_{\mathcal{H}^1_{S,N}}. \square$$

Definition 7.3.3. $\mathcal{H}^{0,1}_{S,N}$ denotes the closure of $\mathcal{S}_S$ in $\mathcal{H}^1_{S,N}$.

Therefore $\mathcal{H}^{0,1}_{S,N}$ is a Hilbert space for the scalar product $<,>_{\mathcal{H}^1_{S,N}}$ and $\mathcal{S}_S$ is dense in $\mathcal{H}^{0,1}_{S,N}$.

Proposition 7.3.3. If $F \in \mathcal{H}^{0,1}_{S,N}$ then $\frac{1}{mes(\Sigma)} \int_{(y,\zeta)\in\Sigma} \zeta^N F(a+iy,\zeta) dy d\xi d\eta \to 0$ when $r,\theta,\mu \to 0$, and same result with b in place of a.

proof. Let $F \in \mathcal{H}^{0,1}_{S,N}$ and let $(F_n)_n$ be a sequence of functions in $\mathcal{S}_S$ such that $F_n \to F$ in $\mathcal{H}^1_{S,N}$. Obviously

$\frac{1}{mes(\Sigma)}|\int_{(y,\zeta)\in\Sigma}\zeta^N F(a+iy,\zeta)dyd\xi d\eta| \leq \frac{1}{mes(\Sigma)}\int_{(y,\zeta)\in\Sigma}|\zeta^N F| \leq \frac{1}{mes(\Sigma)}\int_{(y,\zeta)\in\Sigma}|\zeta^N(F_n - F)| + \frac{1}{mes(\Sigma)}\int_{(y,\zeta)\in\Sigma}|\zeta^N F_n| \leq \frac{1}{mes(\Sigma)} const.\|F_n - F\|_{\mathcal{H}^1_{S,N}} + \frac{1}{mes(\Sigma)}\int_{(y,\zeta)\in\Sigma}|\zeta^N F_n|$

from Proposition 7.3.3. If $\epsilon > 0$ is given $\exists n_0\ n \geq n_0 \Rightarrow \|F_n - F\|_{\mathcal{H}^1_{S,N}} \leq \frac{\epsilon}{2}$. Choose $n \geq n_0$ and r, θ, μ small enough such that such that the second term is $\leq \frac{\epsilon}{2}$ from (7.29).□

Proposition 7.3.4. If $F \in \mathcal{H}^{0,1}_{S,N}$ then $F(a+iy,\zeta) = F(b+iy,\zeta)$ if $0 < |\zeta < r, |arg\zeta| < \theta, |y| < \mu\xi$.

proof. If $v \in \mathcal{H}^1_{S,N}$ from Proposition 7.3.3,

$\int_{(y,\zeta)\in\Sigma}|\zeta^N[v(\frac{a+b}{2} + x_1 + iy,\zeta) - v(\frac{a+b}{2} - x_1 + iy,\zeta)]| \leq const.\|v\|_{\mathcal{H}^1_{S,N}}$. Since by definition $\mathcal{S}_S$ is dense in $\mathcal{H}^{0,1}_{S,N}$ there is a sequence (F_n) in $\mathcal{S}_S$ such that $\|F_n - F\|_{\mathcal{H}^1_{S,N}} \to 0$. Using $F = (F - F_n) + F_n$,

$\int_{(y,\zeta)\in\Sigma}|\zeta^N[F(\frac{a+b}{2} + x_1 + iy,\zeta) - F(\frac{a+b}{2} - x_1 + iy,\zeta)]| \leq \int_{(y,\zeta)\in\Sigma}|\zeta^N[(F - F_n)(\frac{a+b}{2} + x_1 + iy,\zeta) - (F - F_n)(\frac{a+b}{2} - x_1 + iy,\zeta)]| + \int_{(y,\zeta)\in\Sigma}|\zeta^N[F_n(\frac{a+b}{2} + x_1 + iy,\zeta) - F_n(\frac{a+b}{2} - x_1 + iy,\zeta)]|.$

Set $x_1 := \frac{b-a}{2}$; then $\frac{a+b}{2} + x_1 = b, \frac{a+b}{2} - x_1 = a$. Therefore

$\int_{(y,\zeta)\in\Sigma}|\zeta^N[F(b+iy,\zeta) - F(a+iy,\zeta)]| \leq \int_{(y,\zeta)\in\Sigma}|\zeta^N[(F-F_n)(b+iy,\zeta) - (F-F_n)(a+iy,\zeta)]| + \int_{(y,\zeta)\in\Sigma}|\zeta^N[F_n(b+iy,\zeta) - F_n(a+iy,\zeta)]|.$

The last integral is nul from (7.28). From Proposition 7.3.3,

$$\int_{(y,\zeta)\in\Sigma}|\zeta^N[F(b+iy,\zeta) - F(a+iy,\zeta)]| \leq const.\|F - F_n\|_{\mathcal{H}^1_{S,N}}.$$

Since $\|F - F_n\|_{\mathcal{H}^1_{S,N}} \to 0$ it follows that

$$\int_{(y,\zeta)\in\Sigma}|\zeta^N[F(b+iy,\zeta) - F(a+iy,\zeta)]| = 0$$

therefore $F(b+iy,\zeta) = F(a+iy,\zeta)$.□

As a consequence the integration by parts formula

$$\int_a^b\int_{y,\zeta}\zeta^N\frac{\partial u}{\partial x}.v = -\int_a^b\int_{y,\zeta}\zeta^N u.\frac{\partial v}{\partial x}$$

is valid since the two boundary terms simplify.

In the sequel $\mathcal{H}^{0,1}_{S,N}$ will be equipped with the scalar product from $\mathcal{H}^1_{S,N}$. Let $j : \mathcal{L}^2_{S,N} \longmapsto (\mathcal{H}^{0,1}_{S,N})'$ defined by $j(F) : \mathcal{H}^{0,1}_{S,N} \longmapsto \mathbb{C}, u \longmapsto j(F)(u) :=< F, u >_{\mathcal{L}^2_{S,N}}$. The map j is linear continuous :

$$|j(F)(u)| \leq | < F, u >_{\mathcal{L}^2_{S,N}} | \leq \|F\|_{\mathcal{L}^2_{S,N}}\|u\|_{\mathcal{H}^1_{S,N}}.$$

The map j is injective : $j(F)=0 \Rightarrow < F,u >_{\mathcal{L}^2_{S,N}}=0 \quad \forall u \in \mathcal{H}^{0,1}_{S,N}$, i.e. $\forall u \in \mathcal{S}_S$ and $\mathcal{S}_S$ is dense in $\mathcal{L}^2_{S,N}$. From now on we note $\mathcal{L}^2_{S,N} \subset (\mathcal{H}^{0,1}_{S,N})'$ through the map j. Therefore we have the sequence of inclusions

$$\mathcal{S}_S \subset \mathcal{H}^{0,1}_{S,N} \subset \mathcal{L}^2_{S,N} = (\mathcal{L}^2_{S,N})' \subset (\mathcal{H}^{0,1}_{S,N})'. \tag{7.34}$$

The spaces $\mathcal{H}^{0,1}_{S,N}, \mathcal{H}^1_{S,N}, \mathcal{L}^2_{S,N}$ are made of holomorphic functions on S. To pass to the setting of germs we consider the inductive limits

$$\mathcal{L}^2(a,b) = \lim_{\rightarrow} \mathcal{L}^2_{S,N}, \mathcal{H}^1(a,b) = \lim_{\rightarrow} \mathcal{H}^1_{S,N}, \mathcal{H}^{0,1}(a,b) = \lim_{\rightarrow} \mathcal{H}^{0,1}_{S,N},$$

when $r,\theta,\mu \to 0$ and $N \to +\infty$. We have

$$\mathcal{H}^{0,1}(a,b) \subset \mathcal{H}^1(a,b) \subset \mathcal{L}^2(a,b).$$

Now let us give an example of use of the Lax-Milgram theorem in this context. We consider the model equation

$$-u''(x) + c(x)u(x) = f(x),\ u(a)=0, u(b)=0, \tag{7.35}$$

where $f \in \mathcal{L}^2(a,b)$ and c is a bounded holomorphic function on S (c can be a germ whose real interpretation is a Heaviside function or f can be a Dirac delta measure). We choose S small enough such that $f \in \mathcal{L}^2_{S,N}$, for some N large enough. Let $V := \mathcal{H}^{0,1}_{S,N}$ with $\|\|_V := \|\|_{\mathcal{H}^1_{S,N}}$. If $u,v \in V$ we set

$$a(u,v) = < \frac{\partial u}{\partial x}, \frac{\partial v}{\partial x} >_{\mathcal{L}^2_{S,N}} + < cu, v >_{\mathcal{L}^2_{S,N}}, \quad L(v) := < f, v >_{\mathcal{L}^2_{S,N}}. \tag{7.36}$$

lemma 7.3.1. $u \in \mathcal{L}^2_{S,N} \Rightarrow\ cu \in \mathcal{L}^2_{S,N}$.

proof. Since $\mathcal{S}_S$ is dense in $\mathcal{L}^2_{S,N}$ there exists a sequence (u_n) of elements of $\mathcal{S}_S$ that tends to u, i.e. $\|u-u_n\|_{\mathcal{L}^2_{S,N}} \to 0$. The product cu_n is an element of $\mathcal{S}_S$ since c is bounded, and $\|cu_n - cu\|_{\mathcal{L}^2_{S,N}} \to 0$. Apply the definition of $\mathcal{L}^2_{S,N}$.□

We have :

$|a(u,v)| \leq \|\frac{\partial u}{\partial x}\|_{\mathcal{L}^2_{S,N}}.\|\frac{\partial v}{\partial x}\|_{\mathcal{L}^2_{S,N}} + \|c\|_\infty \|u\|_{\mathcal{L}^2_{S,N}} \|v\|_{\mathcal{L}^2_{S,N}} \leq const(\|\frac{\partial u}{\partial x}\|_{\mathcal{L}^2_{S,N}} + \|u\|_{\mathcal{L}^2_{S,N}})(\|\frac{\partial v}{\partial x}\|_{\mathcal{L}^2_{S,N}} + \|v\|_{\mathcal{L}^2_{S,N}}) \leq const\|u\|_V\|v\|_V,$

$|L(v)| \leq \|f\|_{\mathcal{L}^2_{S,N}}\|v\|_{\mathcal{L}^2_{S,N}} \leq \|f\|_{\mathcal{L}^2_{S,N}}\|v\|_V,$

$a(u,u) = \|\frac{\partial u}{\partial x}\|^2_{\mathcal{L}^2_{S,N}} + \int_S |\zeta|^{2N} c(z,\zeta)|u(z,\zeta)|^2 dxdyd\xi d\eta,$

We assume that

$$\forall (z,\zeta) \in S \quad Re(c(z,\zeta)) \geq \alpha > 0. \tag{7.37}$$

Then

$Re(\int_S |\zeta|^{2N} c(z,\zeta)|u(z,\zeta)|^2 dxdyd\xi d\eta) \geq \alpha \|u\|^2_{\mathcal{L}^2_{S,N}}$. This implies that

$$a(u,u) \geq \|\frac{\partial u}{\partial x}\|^2_{\mathcal{L}^2_{S,N}} + \alpha\|u\|^2_{\mathcal{L}^2_{S,N}} \geq const\|u\|^2_V.$$

One can apply the Lax-Milgram theorem in the complex case : there exists a unique $u \in V$ such that

$$a(u,v) = L(v) \quad \forall v \in V, \tag{7.38}$$

i.e.

$$\int_S |\zeta|^{2N} \frac{\partial u}{\partial x}\frac{\partial \overline{v}}{\partial x} + \int_S |\zeta|^{2N} cu\overline{v} = \int_S |\zeta|^{2N} f\overline{v},$$

which implies $< -\frac{\partial^2 u}{\partial x^2} + cu - f, v >_{\mathcal{L}^2_{S,N}} = 0 \quad \forall v \in \mathcal{H}^{0,1}_{S,N}$. Therefore $-\frac{\partial^2 u}{\partial x^2} + cu - f = 0$ in $(\mathcal{H}^{0,1}_{S,N})'$. Since $cu - f \in \mathcal{L}^2_{S,N}$ and $\mathcal{L}^2_{S,N} \subset (\mathcal{H}^{0,1}_{S,N})'$, $\frac{\partial^2 u}{\partial x^2} \in \mathcal{L}^2_{S,N}$.

We have found $u \in \mathcal{H}^{0,1}(a,b)$ such that $\frac{\partial^2 u}{\partial x^2} = cu - f$ in $\mathcal{L}^2(a,b)$. Such a u is unique : let $u_1, u_2 \in \mathcal{H}^{0,1}(a,b)$ such that $\frac{\partial^2 u_i}{\partial x^2} = cu - f$, $i = 1,2$. There exists S, N such that $u_1, u_2 \in \mathcal{H}^{0,1}_{S,N}$ and $f \in \mathcal{L}^2_{S,N}$. Both u_1 and u_2 are solutions of (7.38). From the uniqueness in the Lax-Milgram theorem one has $u_1|_S = u_2|_S$ i.e. $u_1 = u_2$ as germs. The following has been proved

Theorem 7.3.1 : Dirichlet problem in a finite interval. *For all $f \in \mathcal{L}^2(a,b)$ and all c satisfying (7.36) there exists a unique $u \in \mathcal{H}^{0,1}(a,b)$ such that $-\frac{\partial^2 u}{\partial x^2} + cu = f$.*

7.4 An example of minimization of a nonlinear functional.

Let us consider the functional $F : \mathcal{H}^{0,1}_{S,N} \longmapsto \mathbb{R}$,

$$v \longmapsto F(v) = \frac{1}{2}\int_S |\zeta|^{2N}(|\frac{\partial v}{\partial x}|^2 + |v|^2)dxdyd\xi d\eta - Re\int_S |\zeta|^{2N} f\overline{v}dxdyd\xi d\eta,$$

with given $f \in \mathcal{L}^2_{S,N}$. Let $V := \mathcal{H}^{0,1}_{S,N}$. Then

$$F(v) = \frac{1}{2}\|v\|^2_V - Re < f, v >_{\mathcal{L}^2_{S,N}} .$$

The map F is continuous on V, and strictly convex since the quadratic form is issued from a scalar product. The map F is coercive since $\|v\|^2_V$ dominates the linear term when $\|v\|_V \to +\infty$. As a consequence there exists a unique $u_0 \in V$ such that $F(u_0) = min_{v\in V} F(v)$.

Further F is Gateaux-differentiable on V and

$$dF(v).w = Re(< -\frac{d^2 v}{dx^2} + v - f, w >_{\mathcal{L}^2_{S,N}}).$$

Since u_0 minimizes F, $dF(u_0).w = 0 \quad \forall w \in V$. Therefore $Re(< -\frac{\partial^2 v}{\partial x^2} + v - f, w >_{\mathcal{L}^2_{S,N}}) = 0 \quad \forall w \in \mathcal{H}^{0,1}_{S,N}$. The change of w into iw gives the nullness of the imaginary part and therefore $< -\frac{\partial^2 u_0}{\partial x^2} + u_0 - f, w >_{\mathcal{L}^2_{S,N}} = 0 \quad \forall w \in \mathcal{H}^{0,1}_{S,N}$.

Therefore

$$-\frac{\partial^2 u_0}{\partial x^2} + u_0 - f = 0$$

in $(\mathcal{H}^{0,1}_{S,N})'$. Since $u_0 - f \in \mathcal{L}^2_{S,N} \subset (\mathcal{H}^{0,1}_{S,N})'$,

$$\frac{\partial^2 u_0}{\partial x^2} = u_0 - f \tag{7.39}$$

in $\mathcal{L}^2_{S,N}$. We have found $u_0 \in \mathcal{H}^{0,1}(a,b)$ such that (7.37) holds. Uniqueness is proved as for Theorem 7.3.1.

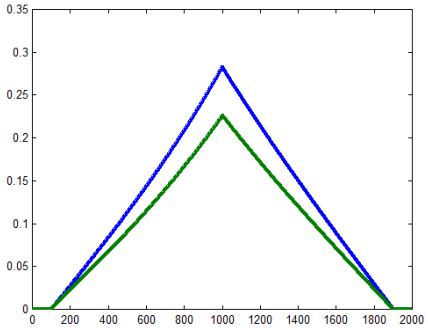
7.5 A physical example and numerical evidence.

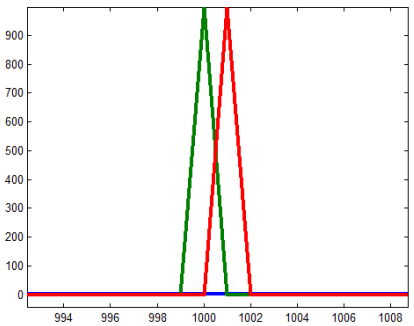
As an example of an equation of physics without classical solution that motivates these techniques let us quote the beam-column equation [14]. The out of plane transverse displacement w of a beam subject to in-plane loads is governed by the equation

$$\frac{\partial^2}{\partial x^2}(EI\frac{\partial^2 w}{\partial x^2}) \quad \frac{\partial}{\partial x}(N\frac{\partial w}{\partial x}) = p \tag{7.40}$$

where p represents loads in the transverse y-direction, N is the tensile resultant, E is the Young modulus of the beam and I is the area moment of inertia of the beam's cross section. Numerical simulations and physical evidence show that when the transverse force p acts on a point x_0 then the right and left hand side derivatives of the displacement w are different. Then $\frac{\partial w}{\partial x}$ is discontinuous at x_0 and $\frac{\partial^2 w}{\partial x^2}$ is a Dirac delta function located at x_0. Then if EI is discontinuous at x_0 the product $EI\frac{\partial^2 w}{\partial x^2}$ does not make sense classically. This equation can be solved in the same way as the model above by introducing in a similar way Sobolev spaces of order 2 in the present context, in the real line for the Dirichlet problem , in a segment for periodic or Neumann conditions. In this situation it can model pipe-lines,...

We present one of the simplest examples in which one observes a nonclassical solution. It concerns the equation $a(x)u''(x) + u(x) = f(x)$ where the function f is a Dirac delta function and where the function a is discontinuous on the support of f. The discretization in use is the standard one. One observes that the same numerical solution u is obtained independently of the meshsize. Its derivative u' is discontinuous on the support of the Dirac delta function therefore the product $a(x)u''(x)$ has the form of a Heaviside function multiplied by a Dirac delta function. One observes that a translation of the Dirac delta function by one cell suffices to change the solution u as expected in presence of this kind of product.





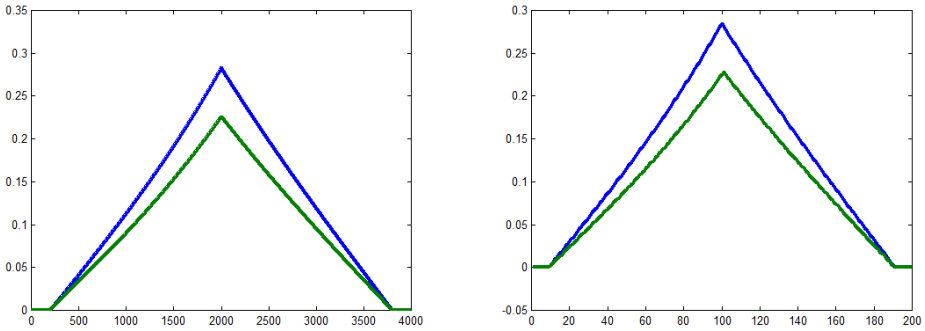


Figure 7.5.1. A numerical solution of a Dirichlet problem on [0,1] with irregular coefficient and Dirac second member with different precisions : 50, 100, 1000, 5000 space steps.

In the figure we show two solutions of the Dirichlet problem $a(x)u''(x) + u(x) = f(x)$ on the segment $[0,1]$; $a(x) = 1$ if $0 < x < 0.5, a(0.5) = 1.5, a(x) = 2$ if $0.5 < x < 1$. The second member f consists successively of two Dirac delta functions that have slightly different supports around the discontinuity of a at $x = 0.5$. One observes neatly different numerical solutions u that do not depend on the meshsize $h = 50^{-1}, 100^{-1}, 1000^{-1}, 5000^{-1}$ from top-left to bottom-right. Therefore the results, which are obtained by the standard discretization of the second order derivative, are not an artefact of calculations. This numerically puts in evidence a non classical problem since an infinitesimal translation of the Dirac delta function f changes significantly the solution. Indeed one observes that the solutions are not differentiable at $x = 0.5$ therefore the term au'' is in form of a product which is classically meaningless.

Bibliographie

[1] A.V. Azevedo, D. Marchesin. Multiple Viscous Solutions for Systems of Conservation Laws, Trans. AMS 347, 1995, pp.3061-3077.

[2] R. Baraille, G. Bourdin, F. Dubois, A.Y. Le Roux. Une version à pas fractionnaire du schéma de Godunov pour l'hydrodynamique. Comptes Rendus Acad Sci. Paris 314, (1992), pp. 147-152.

[3] C. Berthon, M. Breuss, M.-O. Titeux. A relaxation scheme for the approximation of the pressureless Euler equations. Numerical Methods for PDEs, 22 (2006), pp. 484-505.

[4] F. Bouchut, S. Jin, X. Li. Numerical approximation of pressureless and isothermal gas dynamics. SIAM J. Numer. Anal. 41,1, 2003, pp.135-158.

[5] F. Charru. Instabilités hydrodynamiques. EDP Sciences, CNRS Editions, Paris, 2007.

[6] G-Q Chen, D. Wang. The Cauchy problem for the Euler equations for compressible fluids. Handbook of Mathematical Fluid Dynamics, Vol 1, 2002, pp. 421-543.

[7] A. Chertock, A. Kurganov, Y. Rykov. A new sticky particle method for pressureless gas dynamics. SIAM J. Numer. Anal. 45, 2007, pp. 2408-2441.

[8] P. Coles, F.Lucchin. Cosmology. The Origin and Evolution of Cosmic Structure. 2002. Wiley, second edition.

[9] M. Colombeau. A method of projection of delta waves in a Godunov scheme and application to pressureless fluid dynamics. SIAM J. Numer. Anal. 48, 5, 2010, pp. 1900-1919.

[10] M. Colombeau. A consistent numerical scheme for self-gravitating fluid dynamics. Numerical Methods for PDEs, accepted for publication.

[11] M. Colombeau. Numerical approximation of systems modelling large structure formation in cosmology, arXiv : 0905.0230 (2009).

[12] V.G. Danilov, G.A. Omel'yanov, and V.M. Shelkovich. Weak Asymptotic Method and Interaction of Nonlinear Waves,A1MS Translations vol 208,2003, pp 33-164.

[13] W. E, Y. Rykov, Y. Sinai. Generalized variational principles, global weak solutions and behavior with random initial data for systems of conservation laws arising in adhesion particle dynamics. Comm. Math. Phys. 177, 1996, p. 349-380.

[14] Efunda.com/formulae/solid mechanics/columns/beam column (on the web)

[15] Y. Egorov. A contribution to the theory of generalized functions. Russian Math. Surveys 45,5,1990, p.1-49 .

[16] W.D. Evans. Spectral theory and differential operators. Oxford University Press, 1987.

[17] S. Hou, X.D. Liu. Solutions of multi-dimensional hyperbolic systems of conservation laws by square entropy condition satisfying discontinuous Galerkin method. J. Sci. Computing, 31, (2007) pp. 127-151.

[18] F. Huang. Weak solution to pressureless type system. Communications in Partial Differential Equations 30 (2005), pp. 283-304.

[19] F. Huang, Z. Wang. Well posedness for pressureless flow. Comm. Math. Physics 222 (2001), pp.117-146.

[20] A.E. Hurd, D.H. Sattinger. Questions of existence and uniqueness for hyperbolic equations with discontinuous coefficients. Trans. A.M.S. 132 (1968), pp.159-174.

[21] B. L. Keyfitz, H.C. Kranzer. Spaces of Weighted Measures for Conservation Laws with Singular Shock Solutions. J. Diff. Eq. 118,1995, pp. 420-451.

[22] B. L. Keyfitz, H.C. Kranzer. A strictly hyperbolic system of conservation laws admitting singular shocks. In Keyfitz B. and Shearer M. (eds). Nonlinear Evolution Equations that change type. IMA Math. Appl. vol 27, Springer verlag pp. 107-125, 1990.

[23] B. L. Keyfitz, H.C. Kranzer. A system of nonstrictly hyperbolic conservation laws arising in elasticity theory. Archive for Rat. Mech. Anal. 72,1980, pp. 219-241.

[24] D. Korchinski. Solution of the Riemann problem for a 2×2 system of conservation laws possessing no classical weak solution. Thesis. Adelphi University, 1977.

[25] P.D. Lax. Computational fluid dynamics. J. Scientific Computing, 31, 1-2, (2007), pp.185-193.

[26] P.D. Lax. Mathematics and Physics. Bulletin of the AMS, 45,1, (2008), pp. 135-152.

[27] R. J. LeVeque. The dynamics of pressureless dust clouds and delta waves. J. Hyperbolic Diff. Eq. 1,2004, pp. 315-327.

[28] X.D. Liu, P.D. Lax. Solution of two-dimensional Riemann problems of gas dynamics by positive schemes. SIAM J. Sci. Comp. 19, (1998) pp. 319-340.

[29] T. Nguyen, A. Tudorascu. Pressureless Euler/Euler-Poisson systems via adhesion dynamics and scalar conservation laws. SIAM J. Math. Ana. 40, 2, 2008, p. 754-775.

[30] J.A. Peacock. Cosmological Physics. 1999. Cambridge University Press.

[31] P. Peter, J-Ph. Uzan. Cosmologie primordiale. Belin, Paris, 2005.

[32] W.Rudin. Real and Complex Analysis, second edition, Mc Graw-Hill, New York, 1974.

[33] R. Sanders, M. Sever. The numerical study of singular shocks regularized by small viscosity. J. of Scientific Computing 19,1-3, pp. 385-404, 2003.

[34] C.W. Schultz-Rinne, J.P. Collins, H.M. Glaz. Numerical solution of the Riemann problem for two-dimensional gas dynamics. SIAM J. Sci. Comp. 14, (1993), pp. 1394-1414.

[35] M. Sever. Viscous structure of singular shocks. Nonlinearity 15, 2002, pp. 705-725.

[36] M. Sever. Distribution Solutions of Nonlinear Systems of Conservation Laws. Memoirs of the AMS 889, 2007.

[37] C.C. Stark, M.J. Kuchner. A new algorithm for self-consistent 3-D modeling of collisions in dusty debris disks. ArXiv.org, September 2009.

[38] D.C. Tan, T. Zhang, Y.X. Zheng. Delta shock waves as limits of vanishing viscosity for hyperbolic systems of conservation laws. J. Diff. Eq. 112 (1994), pp.1-32.

[39] D.C. Tan, T. Zhang. Two dimensional Riemann problems for a hyperbolic system of nonlinear conservation laws. J. Diff. Eq. 111 (1994), pp.255-283.

[40] E. Toro. Riemann solvers and numerical methods for fluid dynamics. Springer Verlag 1997, second edition 1999. Berlin-Heidelberg-New York.

[41] E. Toro. Shock-Capturing Methods for Free-Surface Flows. J. Wiley and Sons LTD, New York, 2001.

[42] S. Weinberg. Gravitation and Cosmology : Principles and Applications of the General Theory of Relativity. Wiley, New York, 1972.

[43] P. Woodward, Ph. Colella. The numerical simulation of two dimensional fluid flow with strong shocks. J. Comput. Phys. 54, 1984, pp. 115-173.

[44] A. Yao, W. Sheng. Initial boundary value problem of nonlinear hyperbolic system for conservation laws with delta-shock waves. J. Shanghai Univ., 12,4 (2008), pp.306-310.

[45] S. Yang, T. Zhang. The MmB difference solutions of 2-D Riemann problems for a 2×2 hyperbolic system of conservation laws. Impact of Computing in Science and Engineering 3,2 (1991), pp.146-180.