Sur la théorie et l'approximation numérique des problèmes hyperboliques non linéaires :

Saad Benharbit

THESE

présentée par

BENHARBIT Saad

Pour obtenir le titre de docteur de
l'université Joseph Fourier - Grenoble 1

(Arrêtés ministériels du 5 Juillet 1984 et du 23 Novembre 1988)

Spécialité : **ANALYSE NUMERIQUE**

Sur la théorie et l'approximation numérique des problèmes hyperboliques non linéaires. Application à la dynamique des gaz compressibles

---

Date de soutenance : 6 Juillet 1992

Composition du jury :

M. P. Barras	Président
M. G. Brugnot	
M. T. Gallouët	Rapporteur
M. M. Rascle	Rapporteur
M. J.P. Vila	

Thèse préparée au sein du laboratoire : L.M.C.

à mes parents

à Hbiba

à ma famille

REMERCIEMENTS

Tout d'abord je voudrais remercier le seigneur pour avoir guidé mes pas et m'avoir mis sur le droit chemin.

Mes remerciements iront ensuite aux membres du jury de cette thèse.

Pierre Barras me fait l'honneur d'assurer la présidence de ce jury et je lui en sais gré.

Thierry Gallouët et Michel Rascle ont accepté, malgré leurs charges de travail et leurs responsabilités, d'être rapporteurs de cette thèse. Je les remercie aussi pour le temps qu'ils ont bien voulu y consacrer ainsi que pour les conseils qu'ils m'ont donnés.

C'est avec un très grand plaisir que j'exprime ici toute ma gratitude à Gérard Brugnot pour m'avoir accueilli à la division Nivologie du Gemagref de Grenoble, et pour m'avoir à tout moment permis d'effectuer ce travail de thèse dans les meilleurs conditions.

Je tiens à exprimer ma plus vive reconnaissance à Jean-Paul Vila pour la confiance qu'il m'a témoignée en acceptant de diriger cette thèse. J'ai trouvé auprès de lui toute l'aide scientifique et toute la chaleur humaine dont j'ai eu besoin pour mener à bien ce travail. Il n'a jamais ménagé son temps - même pendant ses vacances - pour se mettre à ma disposition et pour me faire profiter de son savoir et de sa trés large culture scientifique. Je souhaite pour tout futur thésard d'avoir la chance de travailler avec un professeur de sa "qualité".

Abdallah Chalabi a été un merveilleux collaborateur et un ami précieux. Sa gentillesse et sa compétence m'ont apporté un réconfort et une aide d'une valeur inestimable. Les discussions que nous avons eues ont toujours été trés enrichissantes. Sa contribution à ce travail est à l'origine même de son aboutissement. Je ne sais trouver les mots justes pour lui exprimer toute ma reconnaissance. (Grenoble a beaucoup perdu avec son départ, tant mieux pour les Toulousains...).

J'ai pu apprécier tout au long de mon séjour à la division Novologie les rares qualités d'amitié et de bonne humeur qui y règnent. Je suis particulièrement heureux à travers ces remerciements d'exprimer mon amitié à tous les membres de la division sans oublier Christian Deymier, et plus particulièrement mon "pote" et ami de toujours Fred : La Nivo est un chouette endroit pour faire une thèse, ... ou autre chose.

Résumé

Dans ce travail de thèse sont étudiés des problèmes mathématiques (théorie et approximation) issus de la théorie de la dynamique des gaz compressibles et modélisés par les équations d'Euler. L'approximation numérique de ces problèmes physiques a nécessité une étude détaillée de quelques problèmes de conditions aux limites.

Les approximations numériques obtenues sont basées sur la méthode des volumes finis, qui nous a semblée la mieux adaptée pour la discrétisation des systèmes hyperboliques de lois de conservation en général.

La convergence de la méthode des volumes finis est obtenue pour les problèmes de lois de conservation scalaires avec des conditions aux limites, à l'aide d'un résultat d'unicité dans l'espace des solutions mesures.

Abstract

We study here some mathematical problems (theory and approximation) stemming from the theory of the gas dynamics. The mathematical model used to represent these physical phenomena is the system of Euler conservation laws equations. To obtain numerical approximations of the solutions we were induced to deal with some boundary value problems.

The results of the numerical simulations are all based on the finite volumes method which seems to be the most useful wheb treating the conservation laws equations in general.

A convergence result has been obtained for the finite volume method for a boundary value problem, by using a uniqueness result of measure valued solutions for a scalar conservation law equation.

PLAN

TROISIEME PARTIE : CONVERGENCE DE QUELQUES METHODES D'APPROXIMATION

INTRODUCTION

Ce travail est motivé par l'étude de problèmes mathématiques (théorie et approximation) issus de la théorie de la dynamique des gaz compressibles et modélisés par les équations d'Euler :

- L'étude de la propagation dans l'atmosphère, de l'onde de choc (de pression) provoquée par la détonation d'un mélange de gaz dans un canon : Application au déclenchement préventif d'avalanches.

- Ecoulement compressible de fluide parfait en situation transonique instationnaire : cas d'un jet à travers une brèche dans une des parois de protection de la cuve d'une centrale nucléaire.

Les ondes de choc apparaissant dans ces phénomènes sont en géneral caractéristiques des systèmes hyperboliques de lois de conservation. Les problèmes posés sont à la fois de type théorique (difficulté d'établir des résultats d'existence et d'unicité) et numérique (conception des méthodes d'approximation); sur ces deux points de nombreuses questions sont encore ouvertes.

Cette thèse comporte trois parties.

-1- La première partie est consacrée à la présentation des problèmes sur le plan mathématique et sur le plan physique, et à la modélisation et la résolution de quelques problèmes de conditions aux limites pour le système de la dynamique des gaz compressibles.

Dans le **premier chapitre** est étudiée la solution du problème de Riemann pour les équations d'Euler pour les gaz parfaits polytropiques. Les résultats de ce chapitre vont nous servir tout au long de cette thèse, aussi bien pour modéliser des problèmes de conditions aux limites au deuxième chapitre, que pour établir des méthodes d'approximation au quatrième chapitre.

Dans le **second chapitre** sont présentés les problèmes physiques décrits précédemment. Dans le cas de la propagation dans l'atmosphère, de l'onde de choc provoquée par la détonation, on a eu recours à la théorie de la détonation des gaz pour étudier les problèmes des conditions aux limites que l'on a modélisé par des demi-problèmes de Riemann.

Le problème de l'existence et de l'unicité de la solution du problème de la condition à la limite dans le cas du jet à travers une brèche dans une des parois de protection du "puit de cuve" a été traité dans le **troisième chapitre**. L'existence de la solution est prouvée par résolution de problèmes de Riemann, et un critère d'entropie a été utilisé pour choisir la "solution physique" du problème.

Pour chacune de ces deux modélisations, les résultats de la simulation numérique sont présentés à la fin des chapitres correspondants.

-2- On s'intéresse dans la **deuxième partie** à l'approximation numérique des problèmes bidimensionnels.

Dans **le chapitre III**, on commence par introduire quelques généralités sur la méthode des volumes finis, qui nous a semblée la mieux adaptée pour la discrétisation des systèmes hyperboliques de lois de conservation en général. Ensuite, dans le même chapitre, cette méthode est décrite pour les équation d'Euler, en utilisant les propriétés importantes d'hyperbolicité et d'invariance par rotation de ces équations.

Le **quatrième chapitre** est consacré au calcul des flux pour les schémas aux volumes finis. On a étudié et appliqué essentiellement le solveur approché de Roe qui consiste à linéariser la jacobienne du système de lois de conservation. Quelques résultats théoriques et leurs démonstrations, dus à Vila [Vil.1], concernant l'obtention de matrices de Roe, sont également présentés dans ce chapitre. Pour clore ce chapitre, on a présenté un algorithme original simplifié de calcul de flux à l'aide du solveur de Roe.

Les résultats numériques du deuxième chapitre ont tous été obtenus à l'aide d'un schéma aux volumes finis précis à l'ordre deux en espace et en temps. Le passage au second ordre en espace sur la méthode des volumes finis a été effectué par la méthode de "prédiction-correction", qui consiste à calculer des gradients par mailles et de les corriger dans le but d'éviter les oscillations dans la solution approchée. Plusieurs critères de correction (ou limitation) peuvent être choisis, l'algorithme que nous avons adopté consiste en une limitation "globale par maille" (annulation du gradient dès qu'un extrêmum est crée sur l'une des arêtes de l'élément). Un autre critère, moins sévère, introduit par Fezoui [Fez] consiste à limiter les gradients "localement", en définissant un gradient par arête puis en n'annulant que les gradients des arêtes où apparaît un extrêmum.

Pour les schémas monodimensionnels obtenus à l'aide de ce type de limitation, on démontre au **chapitre V**, qu'ils sont à variation totale décroissante (TVD), sous une certaine condition de CFL, et par la suite qu'ils sont convergents.

-3- On aborde dans la **troisième partie** l'étude théorique de la convergence des méthodes d'approximation pour les lois de conservation scalaires avec des conditions aux limites. Ces résultats de convergence sont démontrés sans estimations de type TVB (total variation bounded) ou TVD (total variation decreasing). Les estimations de la variation totale sont remplacées par des estimations "faibles" faisant intervenir la viscosité numérique des schémas. La convergence est obtenue dans l'espace des solutions mesures [DP.1]. Grace à un résultat d'unicité des solutions mesures pour les lois de conservation avec conditions aux limites de Szepessy [Sze], on peut alors déduire la convergence forte des solutions approchées dans L^1.

Nous avons adapté au cas du problème aux conditions aux limites, une formulation faible, due à Champier, Gallouët et Herbin [CGH], de la définition des solutions mesures permettant de s'affranchir de la condition de consistance forte avec les conditions initiales utilisée dans [Sze].

Le **chapitre VI** est consacré à la convergence de la méthode des différences finies pour les lois de conservation scalaires à flux non convexe, avec des conditions aux limites. On démontre également l'estimation faible avec la viscosité numérique, nécessaire à la convergence dans l'espace des solutions mesures. Cette estimation servira par ailleurs dans le chapitre suivant. Le point clé de ce chapitre est l'étude de la production locale d'entropie pour le schéma de Godounov dans le cas d'un flux non convexe, qui généralise celle obtenue par Coquel et Le Floch dans [CL].

Dans le **chapitre VII** on démontre la convergence de la méthode des volumes finis pour les lois de conservation scalaires pour un problème de conditions aux limites. Pour cela on écrit le schéma aux volumes finis comme combinaison convexe de schémas aux différences finies, ce qui nous permet d'utiliser les résultats et estimations du chapitre VI. Le résultat de convergence est ensuite étendu à la version précise au second ordre en espace décrite dans le chapitre IV.

Les deux chapitres de cette partie sont destinés à publication, en collaboration avec A. Chalabi* et J.P. Vila**.

---

* Laboratoire d'analyse numérique, Dept. mathématique, Univ. Paul Sabatier, Toulouse.

** Laboratoire d'analyse numérique, Dept. mathématique, Parc Valrose, Univ. Nice.

CHAPITRE I

PROBLEME DE RIEMANN POUR LA DYNAMIQUE DES GAZ COMPRESSIBLES

1. Le problème de Riemann pour la dynamique des gaz compressibles

1.1. Equations et propriétés générales

Le gaz est supposé parfait, il est donc caractérisé par la loi d'état suivante :

$$p = R\rho T \tag{1.1}$$

où ρ, p et T sont respectivement la densité, pression et température du gaz. Il est polytropique, son énergie interne spécifique et sa pression sont données par

$$e = C_v T \tag{1.2}$$

$$p = k e^{S/C_v} \rho^{\gamma} \tag{1.3}$$

où R, k, C_v et γ sont des constantes positives, avec: $\gamma < 1$. S est l'entropie du gaz.

Les équations d'Euler pour la dynamique des gaz s'écrivent

$$\begin{cases} \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) = 0 \\ \frac{\partial}{\partial t}(\rho u) + \frac{\partial}{\partial x}(\rho u^2 + p) = 0 \\ \frac{\partial}{\partial t}\left[\rho\left(\frac{u^2}{2} + e\right)\right] + \frac{\partial}{\partial x}\left[\rho u\left(\frac{u^2}{2} + e\right) + up\right] = 0 \end{cases} \tag{1.4}$$

Le problème de Riemann (voir [Smo]), associé au système (1.4) est caractérisé par une donnée initiale du type

$$U_0(x) = \begin{cases} U_l & \text{si} \quad x<0 \\ U_r & \text{si} \quad x>0 \end{cases} \tag{1.5}$$

La solution du problème (1.4)-(1.5) est constituée, en général, de la succession de trois type d'ondes. Chacune de ces ondes est associée à un champ défini par une des valeurs propres λ_1, λ_2, λ_3 de la Jacobienne du système (1.4). Les données du problème sont résumées dans le tableau suivant, les détails de calcul sont omis.

	i = 1	i = 2	i = 3
λ_i	u - c	u	u + c
vecteur propre R_i	$(\rho, -c, 0)^t$	$(h_s, -c, -h_\rho)^t$	$(\rho, c, 0)^t$
$R_i.\nabla\lambda_i$	$c - p c_p$	0	$c + p c_p$
Invariants de Riemann	$\left\{s, u + \frac{2}{\gamma-1}c\right\}$	$\{u, p\}$	$\left\{s, u - \frac{2}{\gamma-1}c\right\}$

où c est la célérité du son dans le gaz, définie par: $c^2 = \gamma p/\rho$. Les j-invariants de Riemann sont des fonctions W(U) de R^n dans R^n, telles que : $\langle W'(U), r_j(U)\rangle = 0$, pour tout U dans R^n (voir [Smo]). $\langle , \rangle$ désigne le produit scalaire dans R^n, et W'(U) = grad(W).

L'étude des j-invariants de Riemann montre que le premier et le dernier champ sont des champs vraiment non linéaires (VNL), i.e.

$$\text{grad } \lambda_k(U)\, r_k(U) \neq 0 \qquad \text{pour tout U dans } R^n, \quad (k=1 \text{ ou } 3)$$

Les ondes associées à ces champs sont des ondes de choc ou de détente, tandis que le deuxième champ λ_2 est un champ linéairement dégénéré (LD), i.e.

$$\text{grad } \lambda_k(U)\, r_k(U) = 0 \qquad \text{pour tout U dans } R^n, \quad (k=2)$$

la seule onde que l'on peut associer à ce champs est une onde de type discontinuité de contact.

Nous allons étudier explicitement les ondes de choc, de détente et les discontinuités de contact qui constituent la solution du problème de Riemann. En ce qui concerne les courbes de choc, on écrit les relations de Rankine-Hugoniot et les inégalités d'entropie

$$\sigma [U] = [F(U)] \tag{1.6.a}$$
$$\sigma [\eta(U)] \leq [\Pi(U)] \tag{1.6.b}$$

où U et F(U) sont donnés par

$$U = \left(\rho, \rho u, \rho\left(\frac{u^2}{2} + e\right)\right)^T \tag{1.7.a}$$
$$F(U) = \left(\rho u, p + \rho u^2, pu + \rho u\left(\frac{u^2}{2} + e\right)\right)^T \tag{1.7.b}$$

et (η,Π) est une entropie du système (1.4) : $\eta' F' = \Pi'$. σ étant la vitesse du choc. Ces relations peuvent s'écrire de façon plus simple en introduisant les variables

$$v = u - \sigma \qquad m = \rho v$$

On obtient :

$$\begin{cases} [m] = 0 \\ [p + mv] = 0 \\ m\left[\dfrac{2}{\gamma - 1} c^2 + v^2\right] = 0 \end{cases} \tag{1.8}$$

Les conditions de choc de Lax (voir [Smo] par exemple),donnent pour le 1-choc, les inégalités suivantes, qui sont équivalentes à (1.6.b)

$$\sigma < u_l - c_l \qquad u_r - c_r < \sigma < u_r,$$

et pour le 3-choc

$$u_l < \sigma < u_l + c_l \qquad u_r + c_r < \sigma$$

Si on désigne par l'indice 1 l'état d'une particule juste avant qu'elle n'atteigne le choc, et par l'indice 2 l'état de cette même particule juste après la traversée du choc, alors on a:

pour un 1-choc $\quad l=1$ et $r=2$

et pour un 3-choc $\quad l=2$ et $r=1$

autrement dit, pour un 1-choc, les particules traversent le choc de gauche à droite ($\sigma<u_l$ et $\sigma<u_r$). Tandis que pour un 3-choc, les particules traversent le choc de droite à gauche ($\sigma>u_l$ et $\sigma>u_r$). Si on pose :

$$z=\frac{\rho_2}{\rho_1} \qquad \beta=\frac{\gamma+1}{\gamma-1} \qquad \mu^2=\beta \tag{1.9}$$

on obtient l'expression suivante pour la vitesse de choc σ (voir [Smo] pour les détails de calculs)

$$\sigma=u_1\pm c_1\left[\frac{(\beta-1)\,z}{\beta-z}\right]^{1/2} \tag{1.10}$$

avec le signe + (resp. -) pour les 3-chocs (resp. 1-chocs).

1.2. Résolution du problème de Riemann

L'étude des différents champs associés au système d'équations (1.4) a montré que les champs λ_1 et λ_3 sont des champs vraiment non linéaires correspondants à des ondes de détente ou de choc. Le champ λ_2 est linéairement dégénéré, associé à une onde de type discontinuité de contact. D'autre part, les 2-invariants de Riemann sont la vitesse et la pression (u,p), ils sont donc préservés le long de la discontinuité de contact.

Cette remarque fournit un moyen simple de résolution du problème de Riemann (en deux étapes). En effet, deux états du gaz séparés par une discontinuité de contact ont la même vitesse et la même pression. Ils sont représentés par le même point dans le plan (u,p). L'idée est alors d'étudier les ondes de choc et de détente dans le plan (u,p), dans lequel la discontinuité de contact peut être ignorée.

1.2.1. Courbes de choc dans le plan (u,p)

Désignons par l'indice 0 l'état du gaz à gauche du choc de vitesse σ. La relation de Rankine-Hugoniot peut s'écrire

$$e - e_o + \frac{1}{2}(p+p_o)(\tau-\tau_o)=0 \tag{1.11}$$

où e désigne l'energie interne spécifique, donnée par : $e = p\,\tau/(\gamma-1)$, où on a posé : $\tau = 1/\rho$. On obtient la relation suivante reliant la vitesse et la pression

$$u_1 - u_o = \pm\,\Phi_o(p_1), \tag{1.12.a}$$

avec

$$\Phi_o(p) = (p - p_o)\sqrt{\frac{(1 - \mu^2)\tau_o}{p + \mu^2 p_o}} \qquad (1.12.b)$$

Dans (1.12.a), le signe + désigne l'onde de 3-choc, tandis que le signe - indique le 1-choc.

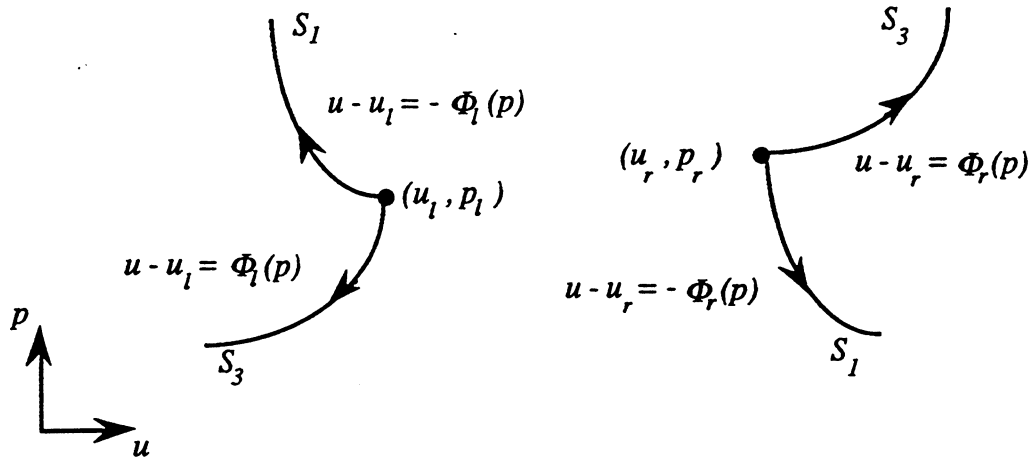
La fonction Φ_o vérifie les propriétés suivantes:

i) $\Phi_o' > 0$

ii) $\Phi_o(p_1) = \Phi_1(p_o)$

iii) $\Phi_o(p) \to \infty$ quand $p \to \infty$

iv) $\Phi_o'(p) \to 0$ quand $p \to \infty$



(a) **Figure 1.1.** (b)

Les figures (1.1) représentent, dans le plan (u,p), l'allure des courbes de chocs. La figure (1.1.a) représente l'ensemble des états qui peuvent être reliés à l'état (u_l,p_l,τ_l) à droite par une onde de choc de type S_i, i=1 ou 3. Autrement dit, dans ce cas, on dispose de l'état gauche. La figure (1.1.b) représente l'ensemble des états qui peuvent être reliés à l'état (u_r,p_r,τ_r), à gauche par une onde de choc de type S_i, i=1 ou 3. Autrement dit, dans ce cas, on dispose de l'état droit.

1.2.2. Courbes de détente dans le plan (u,p)

De la même façon que pour les ondes de choc, les équations des ondes de détente peuvent être obtenues explicitement. On écrit pour cela, qu'un i-invariant de Riemann est constant le long d'une i-détente (i = 1 ou 3). La fonction "u + 2c/(γ - 1)" (resp. "u - 2c/(γ - 1)") est constante le long d'une 1-détente (resp. 3-détente).

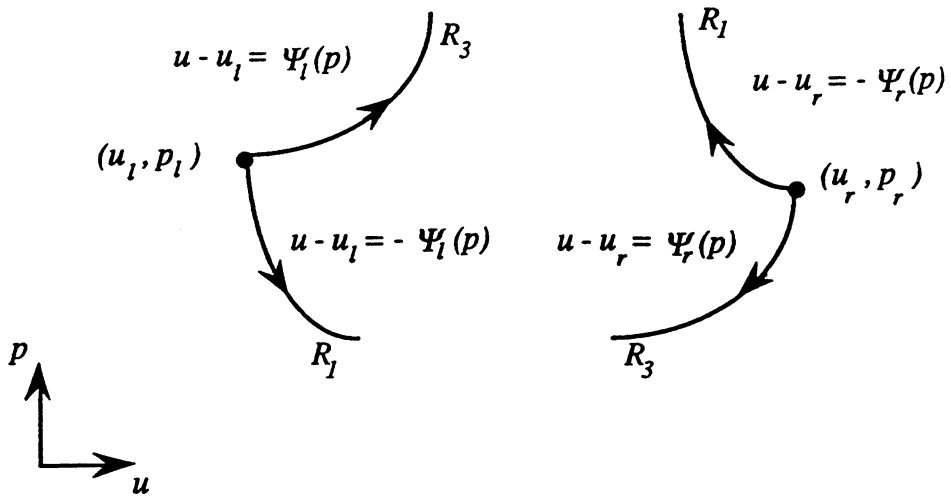
On obtient la relation suivante reliant la vitesse et la pression

$$u_1 - u_o = \pm \Psi_o(p_1), \qquad (1.13.a)$$

avec

$$\Psi_o(p) = \frac{\sqrt{1 - \mu^4}}{\mu^2} \sqrt{\tau_o}\ p_0^{1/2\gamma} \left(p^{(\gamma-1)/2} - p_0^{(\gamma-1)/2}\right) \qquad (1.13.b)$$

Dans (1.13.a), le signe + désigne la 3-détente, tandis que le signe - indique le 1-détente.



(a) **Figure 1.2.** (b)

La fonction Ψ_o vérifie les propriétés suivantes:

i) $\Psi_o' > 0$

ii) $\Psi_o(p_1) = -\Psi_1(p_o)$

iii) $\Psi_o(p) \to \infty$ quand $p \to \infty$

iv) $\Psi_o'(p) \to 0$ quand $p \to \infty$

Les figures (1.2.a)-(1.2.b) représentent, dans le plan (u,p), l'allure des courbes de détente, suivant le même principe que dans la figure 1.1.

CHAPITRE II

ETUDE DE QUELQUES PROBLEMES AUX LIMITES POUR LES EQUATIONS D'EULER

0. Introduction

Les applications physiques qui ont motivé ce travail relèvent toutes du domaine de la dynamique des gaz compressibles et plus précisemment :

- L'étude de la propagation dans l'atmosphère, de l'onde de choc (de pression) provoquée par la détonation d'un mélange de gaz dans un canon. (application "pacifiste !" au déclenchement préventif d'avalanche).
- Ecoulement compressible de fluide parfait en situation transonique instationnaire. Cas d'un jet à travers une brèche dans une des parois de protection de la cuve d'une centrale nucléaire.

Ces phénomènes sont modélisés par les équations d'Euler pour la dynamique des gaz compressibles.

1. Un système de déclenchement préventif d'avalanches : Le Gazex

Dans le domaine de la protection paravalanche par déclenchement artificiel, existe principalement deux techniques utilisant des procédés trés différents :

- Le câble transporteur d'explosif (CATEX), connu en France depuis près de 15 ans. Il permet de placer quelques kilogrammes d'explosifs au dessus du manteau neigeux d'un couloir avalancheux.
- L'exploseur à gaz (GAZEX), apparu il'y a 3 ans. Son principe de fonctionnement repose sur l'explosion d'un mélange gazeux détonant, en l'occurrence oxygène-propane, dans un tube (canon) orienté vers l'aval. La surpression ainsi créée sur le manteau neigeux provoque le départ de l'avalanche. Ces deux techniques peuvent paraître concurrentielles.

Le CEMAGREF de Grenoble a mené, pour le compte de la société SCHIPPERS S.A., fabriquant du GAZEX, une étude consistant à évaluer l'efficacité de ce nouveau procédé et à comparer ses performances avec celles du CATEX. Cette étude comporte essentiellement deux phases :

- Mesures sur le terrain (tirs d'essais) : d'une part des mesures de la surpression créée sur le manteau neigeux, et des mesures acoustiques pour évaluer l'intensité des ondes sonores produites par la détonation.
- Etude numérique comportant une modélisation mathématique du phénomène physique, complétée d'une simulation numérique de l'onde de pression qui suit la détonation. Les mesures effectuées sur le terrain dans ce sens, servent à initialiser partiellement le modèle numérique.

1.1. Le GAZEX

1.1.1. Description

Un tube est installé sur le site, relié par l'intermédiaire de tuyaux à un abri contenant deux cuves tampons. Des réserves d'oxygène et de propane permettent d'alimenter les cuves (voir schéma d'installation). Les pages suivantes montrent le canon, son alimentation, l'abri et les cuves.

1.1.2. Fonctionnement

Un système de commande permet par ouverture d'électrovannes, l'injection des gaz des cuves tampons vers le canon (pour que le mélange oxygène-propane respecte les proportions stochiométriques (détonation optimale) : 82% d'oxygène et 18% de propane). Deux allumeurs piezo-électriques permettent de déclencher l'explosion soit au fond, soit au milieu du canon (double effet).

Remarque 1.1.

La puissance du canon est fonction du volume de gaz injecté dans le canon. La société SCHIPPERS commercialise des modèles de canon de capacités variant entre 0.5 m^3 et 5m^3.

1.2. Modélisation mathématique

On rappelle que le but de l'étude numérique est de simuler la propagation dans l'atmosphère de l'onde de pression declenchée par la détonation. Ce phénomène est modélisé par le système d'équations de la dynamique des gaz polytropiques. Le problème est tridimensionnel, les équations d'Euler en trois dimensions s'écrivent :

$$\begin{cases} \partial\rho/\partial t + \mathrm{div}\,(\rho \vec{U}) = 0 \\ \partial(\rho\vec{U})/\partial t + \mathrm{div}\,(\rho \vec{U}\otimes\vec{U}) + \vec{\mathrm{grad}}(p) = 0 \\ \partial E/\partial t + \mathrm{div}\,(E+p)\rho\vec{U} = 0 \end{cases} \qquad (1.1)$$

où $\otimes$ désigne le produit tensoriel, ρ est la densité du gaz. $\vec{U} = (u;v;w)^T$ est le vecteur vitesse du gaz. E et p désignent respectivement l'energie totale et la pression du gaz. Les gaz étudiés seront tous assimilés à des gaz parfaits de rapport de chaleurs spécifiques $\gamma = 1{,}4$ (cette hypothèse sera justifiée plus loin). Pour fermer le système (1.1), on dispose de la relation :

$$p = (\gamma-1)\left[E - \frac{1}{2}\rho\,\|\vec{U}\|^2\right] \qquad (1.2)$$

où : $\|\vec{U}\|^2 = (u^2+v^2+w^2)$.

1.2.1. Symétrie du système

L'onde de pression qui suit la détonation, se propage dans toutes les directions de l'espace, vers l'avant du canon. Le problème est donc tridimensionnel. Si on se place sous l'hypothèse simplificatrice qui consiste à négliger les effets de reflexion de l'onde sur le manteau neigeux et les parois rocheuses situées dans les environs immédiats du canon, on peut alors supposer que le problème admet une symétrie autour de l'axe du canon.

Le fait de négliger ainsi les effets de reflexion, entraîne une légère sous estimation, par la simulation numérique, de l'intensité de l'onde. Le GAZEX est avant tout un système de protection paravalanche, il n'est donc pas gênant que la simulation fournisse un résultat légèrement en dessous de son efficacité réelle, surtout si on le sait à priori. Ceci justifie donc l'approximation ci-dessus.

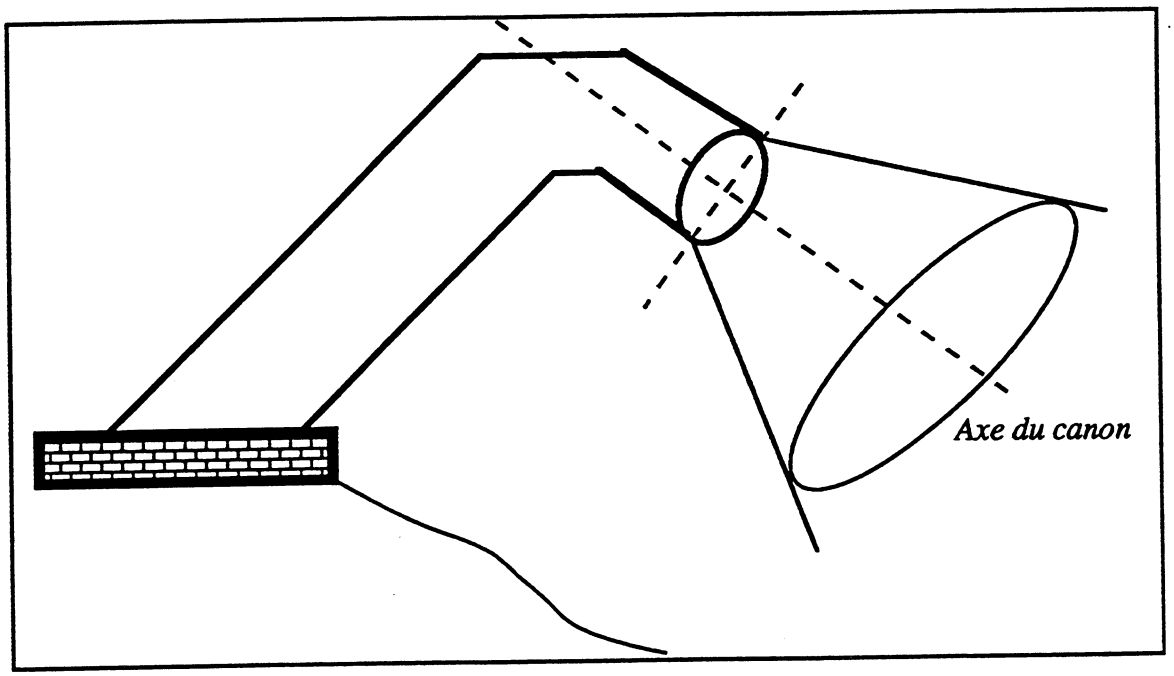


Figure 1.1. : Axisymétrie du problème.

En s'appuyant sur l'hypothèse de symétrie axiale, le système des équations d'Euler, une fois réecrit en coordonnées cylindriques (r,θ,z) dans le repère d'origine le centre de la bouche du canon, peut se ramener à un système bidimensionnel, la contribution suivant θ étant alors nulle.

1.2.2. Equations d'Euler en coordonnées cylindriques

Soit $\vec{U} = (u_r, u_\theta, u_z)^T$, les composantes du vecteur vitesse en coordonnées cylindriques. Les opérateurs divergence et gradient s'écrivent, en coordonnées cylindriques :

$$\operatorname{div}(\vec{U}) = \frac{\partial u_z}{\partial z} + \frac{1}{r}\frac{\partial (r u_r)}{\partial r} + \frac{1}{r}\frac{\partial u_\theta}{\partial \theta} \qquad \vec{\operatorname{grad}}(p) = \left(\frac{\partial p}{\partial z}, \frac{\partial p}{\partial r}, \frac{1}{r}\frac{\partial p}{\partial \theta}\right)^T$$

En réecrivant le système (1.1) en coordonnées cylindriques, après élimination des composantes suivant θ, et en remarquant que : $\frac{\partial p}{\partial r} = \frac{\partial rp}{\partial r} - p$, on obtient :

$$\frac{\partial}{\partial t}\begin{pmatrix}\rho\\ \rho u_z\\ \rho u_r\\ E\end{pmatrix} + \frac{\partial}{\partial z}\begin{pmatrix}\rho u_z\\ \rho u_z^2+p\\ \rho u_r u_z\\ u_z(E+p)\end{pmatrix} + \frac{1}{r}\frac{\partial}{\partial r}\begin{pmatrix}\rho r u_r\\ \rho r u_r u_z\\ \rho r u_r^2+rp\\ r u_r(E+p)\end{pmatrix} = \begin{pmatrix}0\\ 0\\ p/r\\ 0\end{pmatrix}$$

En multipliant partout par r et en posant : $\rho' = r\rho$, $E' = rE$ et $p' = rp$, on obtient le système suivant :

$$\frac{\partial}{\partial t}\begin{pmatrix}\rho'\\ \rho' u_z\\ \rho' u_r\\ E'\end{pmatrix} + \frac{\partial}{\partial z}\begin{pmatrix}\rho' u_z\\ \rho' u_z^2+p'\\ \rho' u_r u_z\\ u_z(E'+p')\end{pmatrix} + \frac{\partial}{\partial r}\begin{pmatrix}\rho' u_r\\ \rho' u_r u_z\\ \rho' u_r^2+p'\\ u_r(E'+p')\end{pmatrix} = \begin{pmatrix}0\\ 0\\ p\\ 0\end{pmatrix} \qquad (1.3)$$

On remarque qu'il s'agit exactement des équations d'Euler bidimensionnelles, avec second membre, où on a remplacé les coordonnées (x,y) par (z,r), et où le vecteur des variables conservatives est donné par : $U' = (\rho', \rho' u_z, \rho' u_r, E')^T$. On peut donc appliquer les mêmes schémas numériques que ceux utilisés pour discrétiser les équations d'Euler en 2D, en particulier les schémas basés sur la résolution du problème de Riemann pour les équations d'Euler. La seule modification à ajouter est le traitement du second membre. Dans ce cas particulier, le traitement du second membre ne présente pas de difficultés particulières, vu que le second membre ne contient que des termes d'ordre zero.

1.3. Modélisation numérique, Problème des conditions aux limites

L'hypothèse d'axisymétrie nous ramène à résoudre un problème bidimensionnel avec un terme source. Le domaine Ω choisi pour la discrétisation est représenté par la figure 1.2. La longueur de Ω est de 120m (Des estimations empiriques de la zone de surpression à 20 mbar; pression d'un skieur d'un poids moyen de 70 kg; provoquée par la détonation, l'évaluent à environ une centaine de métres du canon.

La modélisation numérique nécessite, une fois une méthode de résolution et un schéma numérique choisis, une analyse spécifique du problème des conditions aux limites. Dans ce cas précis, nous avons à traiter plusieurs types de conditions aux limites (voir figure 1.2.) :

- Conditions aux limites de type reflexion sur les parois latérales du canon : zone (1).
- Conditions aux limites de type sortie libre : zones (2), (3) et (4).
- Condition à la limite "particulière" en sortie du canon (pendant la durée de l'explosion).

La qualité des résultats obtenus dépendra fortement de la façon de calculer la condition à la limite en sortie du canon, autrement dit :

$$U(x=0, t>t_o) = \begin{pmatrix}\rho(x=0, t>t_o)\\ u(x=0, t>t_o)\\ p(x=0, t>t_o)\end{pmatrix}$$

où l'origine (x=0) est prise au centre de la "bouche" du tube, et l'origine du temps t_o est l'instant où le gaz commence à sortir du tube.

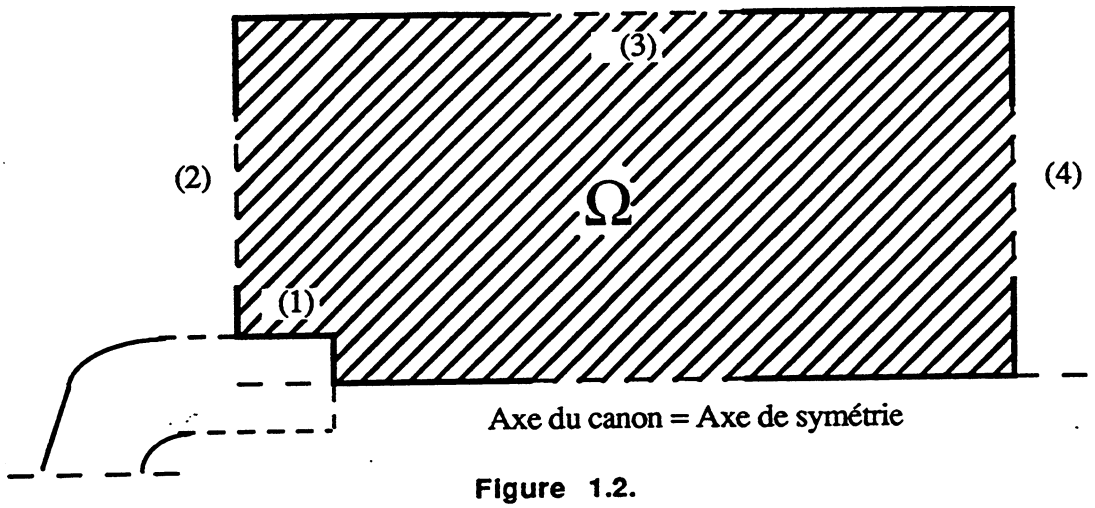


Figure 1.2.

En effet, le phénomène étudié est trés bref, la durée d'une détonation est de quelques ms, son intensité est trés forte (pression et vitesse du gaz en sortie de l'ordre de 60 bar et 1100 ms^{-1} respectivement). Il faut donc connaître de manière précise les courbes de densité, vitesse et pression en sortie du canon pendant l'explosion. Il est par ailleurs trés délicat d'obtenir des mesures précises de ces variables, à cause de la puissance de l'explosion (arrachement des fils des capteurs) et de l'état accidenté du terrain (instabilité des perches de sustension des capteurs).

Il s'avère donc nécessaire de faire une étude spécifique pour les conditions aux limites en sortie du canon. Pour cela, on a commencé par assimiller les mélanges gazeux (oxygène, propane et air) à un seul gaz de même rapport de chaleurs spécifiques (γ=1.4), cela permet d'éviter d'étudier les équations de la dynamique des gaz pour un mélange multi-espèces. Cette approximation a une influence trés faible sur les résultats obtenus, en effet cette constante est en géneral comprise entre 1.25 et 1.3 pour des gaz brûlés [Med]. Ensuite, on a étudié le problème de Riemann "généralisé" dont la donnée initiale à un instant t, est constituée de l'état du gaz à l'intérieur du canon d'une part, et de l'état du gaz à l'extérieur (atmosphère) d'autre part. Pour cela, il a été nécessaire de passer par l'étude de la théorie de la détonation d'un mélange gazeux dans un tube [Med].

Cette étude par le biais du problème de Riemann généralisé (cf §1.3.1.-1.3.4.), nous a donné une idée sur l'évolution dans le temps de la condition à la limite en sortie du tube. Dans le but de vérifier la validité des approximations qu'il a été nécessaire de faire, nous avons effectué une simulation numérique de la phase de détonation (voir résultats §.1.3.5).

1.3.1. Détonation d'un gaz dans un cylindre, avec amorçage au fond fermé, ouvert à l'autre extrémité

La détonation implique une onde de choc, le calcul des valeurs de pression que ces ondes produisent, se fait en général, d'une part à l'aide des équations de la détonation, et d'autre part à l'aide des équations de choc.

Berthelot et Vieille [Med] qui observèrent ce phénomène pour plusieurs mélanges gazeux, reconnurent que l'onde se propage d'un mouvement uniforme, autrement dit, sa vitesse est indépendante de la longueur sur laquelle on l'observe. Ils la trouvèrent également indépendante de la matière et du diamètre du tube; la vitesse était encore la même que le tube soit droit ou enroulé sur un tambour ou qu'il présente des coudes. Enfin, quand on amorçe la détonation à une extrémité du tube, la vitesse observée est la même quand l'extrêmité est close ou quand elle est ouverte à l'air libre. On peut donc assimiler le canon à un tube cylindrique ouvert à une extrémité. L'explosion étant amorcée au fond fermé.

1.3.2. Théorie de Chapman-Jouguet

Jouguet a mis en place en 1906, une théorie appelée théorie hydrodynamique de la détonation. Sous sa forme la plus simplifiée, le raisonnement considère une onde de choc réduite à un plan, en supposant que la réaction chimique s'accomplit instantanément à la traversée de ce plan, qui est véritablement une surface de discontinuité pour la composition chimique.

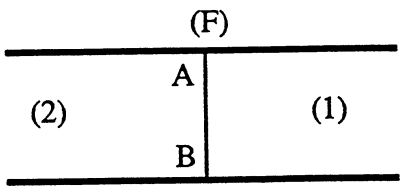


Figure 1.3.

L'onde de choc à front plan, de célérité notée U_1 ou U_2 selon qu'il s'agit de la célérité par rapport au milieu (1) ou au milieu (2); avance dans l'état (1). La vitesse σ du front AB de l'onde de choc, par rapport à des axes liés à l'observateur pour qui le fluide a les vitesses u_1 et u_2 vaut U_1+u_1 ou U_2+u_2 :

$$\sigma = U_1 + u_1 = U_2 + u_2 \tag{1.4}$$

Les relations de choc de Rankine-Hugoniot s'écrivent :

$$[F(V)] = \sigma [V] = \sigma (V_2 - V_1) \tag{1.5}$$

avec : $V = (\rho, \rho u, E)^T$ $\qquad F(V) = (\rho u, \rho u^2 + p, u(E+p))^T$

Si on pose : $m = \rho U$, les relations de Rankine-Hugoniot deviennent

$$\begin{cases} [m] = 0 \\ [p - mU] = 0 \\ m\left[\frac{2}{\gamma-1} c^2 + U^2\right] = 0 \end{cases} \tag{1.6}$$

On en déduit les équations suivantes :

$$\frac{U_1}{\tau_1} = \frac{U_2}{\tau_2} = \frac{p_2-p_1}{u_2-u_1} \tag{1.7.a}$$

$$\frac{U_1^2}{\tau_1^2} = \frac{U_2^2}{\tau_2^2} = \frac{p_2-p_1}{\tau_1-\tau_2} \tag{1.7.b}$$

$$(p_1+p_2)(\tau_2-\tau_1) + 2(E_2-E_1) = 0 \tag{1.7.c}$$

où E_i désigne l'énergie interne spécifique du gaz, et τ_i son volume spécifique ($\tau_i = \frac{1}{\rho_i}$). Si on note c_i la célérité du son par rapport au milieu (i), Jouguet proposa de considérer que la **détonation stable** est celle qui vérifie :

$$U_2 = c_2 \tag{1.8}$$

ce qui donne :

$$U = c_2 + W - u_1 \tag{1.9}$$

Cette onde critique, ou onde explosive, est la moins rapide, mais est la seule stable parmi les ondes de détonation possibles, c'est ce qu'on appelle onde de "détonation régulière".

Dans notre cas de figure, le milieu (1) est au repos (u_1=0), de sorte que U_1 est la célérité de l'onde par rapport à l'observateur, célérité que l'on peut sans inconvénient désigner par U (sans indice); la vitesse u_2 est la vitesse matérielle dans le milieu qui vient d'être traversé par l'onde de choc (vitesse des gaz brûlés), vitesse que nous désignerons par W, les équations (1.8) et (1.9) deviennent alors

$$U = \tau_1 \sqrt{\frac{p_2-p_1}{\tau_1-\tau_2}} \tag{1.10}$$

$$W = (\tau_1-\tau_2) \sqrt{\frac{p_2-p_1}{\tau_1-\tau_2}} \tag{1.11}$$

ce qui donne, par division membre à membre, les relations suivantes :

$$\frac{W}{U} = \frac{\tau_1-\tau_2}{\tau_1} \tag{1.12}$$

$$W.U = p_1(\tau_1 - \tau_2) \tag{1.13}$$

qui relient la vitesse matérielle W et la vitesse de l'onde U à des grandeurs simples.

En avant du plan (F) (figure 1.3.), l'explosif n'ayant pas encore réagit et au repos est caractérisé par les valeurs ρ_1, p_1 et T_1 des grandeures densité, pression et température. De l'autre côté du plan, le gaz a subi une réaction chimique irréversible, l'ayant amené à un état caractérisé par : ρ_2, p_2 et T_2.

L'explosion entraîne aussi une mise en mouvement du gaz brûlé, à des vitesses de l'ordre de 10^3 $m.s^{-1}$, à ne pas confonde avec la célérité de l'onde, pouvant atteindre des valeurs de l'ordre de 2500 à 3000 $m.s^{-1}$. D'autre part, la célérité de l'onde sonore est donnée, en fonction de la densité et de la pression, par

$$c_2 = \sqrt{\frac{\gamma_2 p_2}{\rho_2}} \tag{1.14}$$

où γ_2, rapport des chaleurs spécifiques des gaz brûlés, est voisin de 1.25 quand la majorité des gaz brûlés sont diatomiques, en général il est compris entre 1.25 et 1.3 [Med].

En combinant les relations (1.7), (1.8) et (1.14), on obtient

$$\frac{p_2-p_1}{\tau_1-\tau_2} = \frac{c_2^2}{\tau_2^2} \tag{1.15}$$

Si on considère que le gaz à l'état (1) est à une pression p_1 = 1 atm, ce qui est le cas ici, p_2 dans l'onde de détonation vaut de 40 à 50 fois p_1. On peut, d'une manière approchée, remplacer $(p_2 - p_1)$ par p_2 dans (1.15). Il en résulte, d'aprés (1.14), que :

$$\frac{\tau_1}{\tau_2} \approx \frac{\gamma_2+1}{\gamma_2} \tag{1.16}$$

ce qui fournit une bonne approximation du rapport :

$$\frac{\tau_1}{\tau_2} \approx 1.8 \tag{1.17}$$

L'équation (1.12) donne alors l'approximation suivante du rapport $\frac{W}{U}$

$$\frac{W}{U} \approx 0.44 \tag{1.18}$$

1.3.3. Propagation de l'onde de choc dans le canon

Le canon est assimilé à un tube ouvert à une extrémité et fermé à l'autre, où est déclenchée la détonation. La pression initiale du gaz au repos, est prise égale à la pression extérieure p_1=1atm. Le déclenchement ayant eu lieu à l'instant t=0, au fond fermé (F) (figure 1.4.). A l'instant t, l'onde sera en A. La distance FA étant égale au produit D.t (D = célérité de l'onde). La pression immédiatement en amont de A est la pression de détonation p_2 (environ 50 atm).

Au droit de l'onde A, la pression est donc soumise à un saut de pression de 1atm à p_2 au moment de passage de l'onde, mais la pression appliquée baisse rapidement. Le profil de pression entre F et A a la forme représentée par la figure 1.5.

Entre F et B, le gaz qui se retrouve au repos est à une pression appelée pression statique : p_{s2}, de valeur à peu prés égale à la moitié de la pression de détonation p_2. Le rapport de distance FB et FA reste constant, et égal à 0.55 environ, pendant la propagation de la détonation.

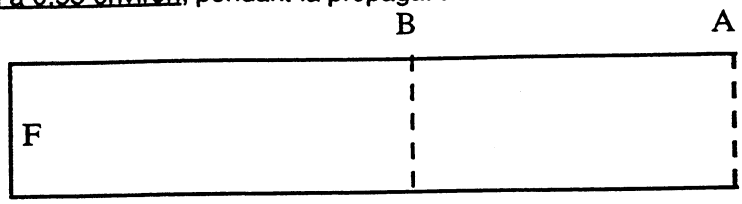


Figure 1.4.

Quand l'onde parvient au fond ouvert, les gaz situés derrière l'onde s'échappent à l'air, et il se produit une onde de raréfaction qui remonte en amont et ramène la pression à 1 atm.

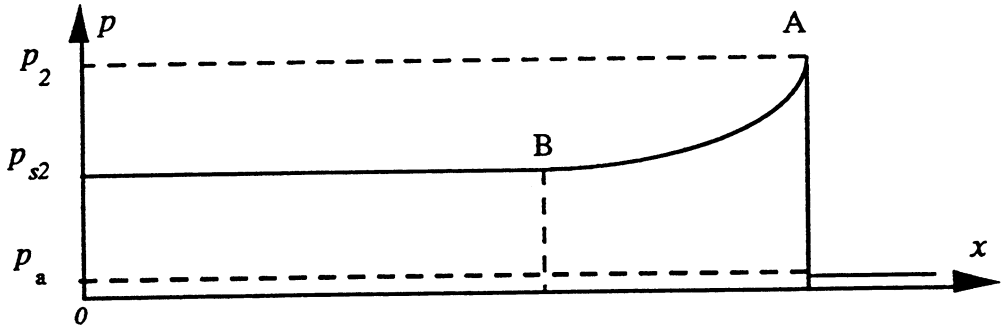


Figure 1.5.

1.3.4. Courbes de densité, vitesse et pression en sortie du tube

Soit L la longueur du canon, à l'instant : $t_o = L/D$, on se retrouve à l'intérieur du canon, avec la configuration suivante (voir figures 1.6 & 1.7).

- L'onde a atteint le bout du canon, tout le gaz ayant brûlé. Sa densité est, au vu de la relation (1.17), donnée par : $\rho_2 = 1.8\ \rho_1$, ρ_1 désigne la densité du gaz avant l'explosion
- Tout le gaz situé entre F et B est au repos, à la pression p_{S2}.
- Entre B et A, la vitesse et la pression ont des distributions qu'on peut assimiler à des segments de droites, avec :

en B :	v = 0	$p = p_{S2}$
en A :	v = W	$p = p_2$.

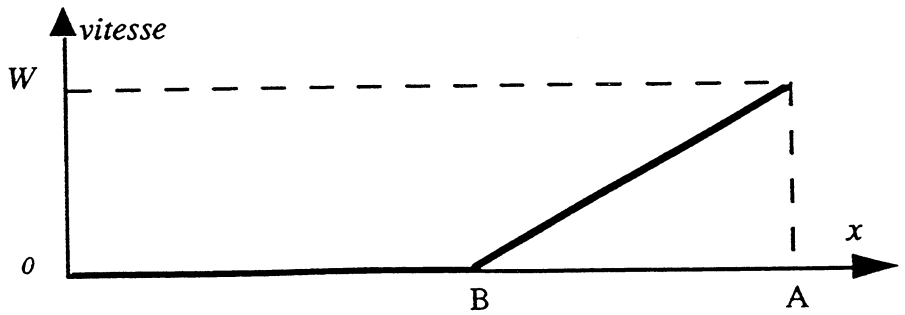


Figure 1.6.

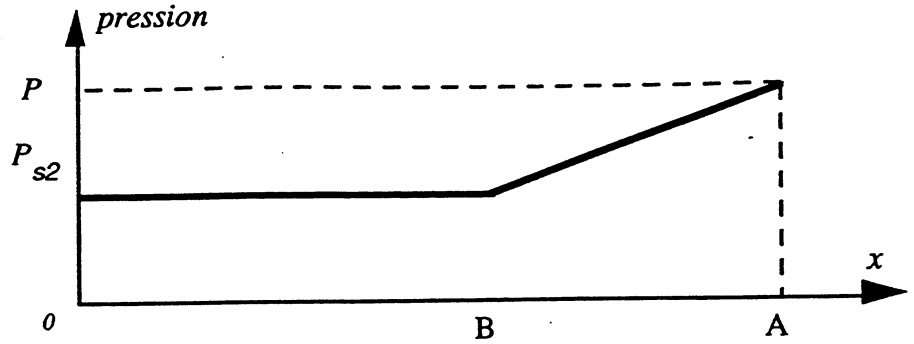


Figure 1.7.

Question :

Que se passe t-il pour $t>t_o$, t suffisamment voisin de t_o?

Au vu des rappels sur la théorie de la détonation, le calcul de la condition à la limite en sortie du canon, i.e. le calcul de : $U(x=0, t>t_o) = (\rho(x=0, t>t_o), u(x=0, t>t_o), p(x=0, t>t_o))^T$, où l'origine (x=0) est prise à l'extrémité libre du canon; se ramène donc à résoudre le **problème de Riemann généralisé** suivant :

$$\begin{cases} \partial_t U + \partial_x f(U) = 0 \\ U(0,x) = U_o(x) = \begin{cases} U_L(x) & \text{si} \quad x<0 \\ U_R(x) & \text{si} \quad x>0 \end{cases} \end{cases} \qquad (1.19)$$

On rappelle qu'un problème de Riemann généralisé est un problème de Riemann pour lequel les valeurs de U à droite et à gauche du choc de la condition initiale $U_o(x)$, ne sont pas constantes et sont fonctions de x. Ici, $U_L(x)$ est l'état du gaz à l'intérieur du canon à l'instant t_o (pour x variant entre (-L) et 0, où L est la longueur du canon), et $U_R(x)$ est l'état de l'air à l'extérieur du canon :

$$U_R(x) = \begin{pmatrix} \rho_{air} \\ u_{air} \\ p_{air} \end{pmatrix} = \begin{pmatrix} 1.29\ kgm^{-3} \\ 0\ ms^{-1} \\ 1\ atm \end{pmatrix}$$

Nos besoins se limitant au calcul de la valeur de la solution du problème (1.19) à l'origine (x=0). La résolution exacte du problème de Riemann généralisé, même dans le cas de la dynamique des gaz parfaits, étant relativement complexe [Lit]), nous avons donc préferé décomposer le problème en plusieurs étapes et supposer, dans un premier temps, que l'on a à résoudre le problème de Riemann classique : $PR(U_{LO},U_{RO})$, où on a posé :

$$U_{LO} = U_L(0) \qquad \text{et} \qquad U_{RO} = U_R(0)$$

$$U_{LO} = (\rho_2, W;p_2)^T \qquad \text{et} \qquad U_{RO} = (\rho_{air}, u_{air}, p_{air})^T$$

Comme $p_2>p_{air}$, la solution de $PR(U_{LO},U_{RO})$, contient (voir Chap.1.§.1) une 1-détente sonique associée à la valeur propre $\lambda_1=u_{LO}-c_{LO}$ ($\lambda_1\leq 0$), avec : $c_{LO}^2 = (\gamma p_{LO}/\rho_{LO})$. Dans le plan (u,p), la solution est la suivante :

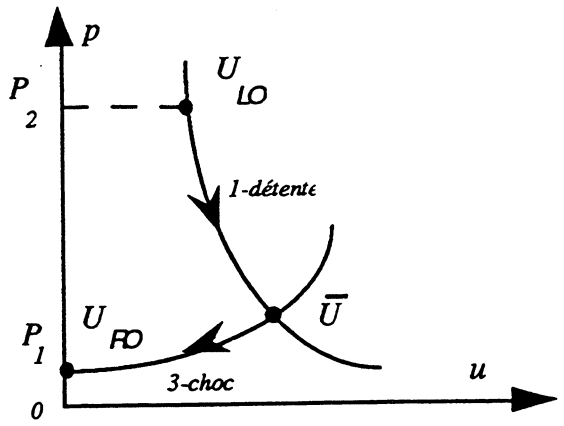


Figure 1.8.

i.e. U_{LO} est relié à un état intermédiaire $\bar{U}$, par une 1-détente, ensuite $\bar{U}$ est relié à U_{LO} par un 3-choc.

Le calcul de la valeur stationnaire : $U_*(x/t = 0) = (\rho_*, u_*, p_*)$ s'effectue à l'aide des relations définissant les états qu'on peut relier à U_{LO} par une 1-détente :

$$u_* + \frac{2c_*}{\gamma - 1} = u_{LO} + \frac{2c_{LO}}{\gamma - 1} \tag{1.20}$$

et

$$\frac{p_*}{\rho_*^\gamma} = \frac{p_{LO}}{\rho_{LO}^\gamma} \tag{1.21}$$

qui traduisent que les 1-invariants de Riemann sont constants le long d'une 1-détente. c_* est donné par : $c_* = \left(\frac{\gamma p_*}{\rho_*}\right)^{1/2}$. D'autre part, u_* et c_* vérifient la relation suivante :

$$u_* - c_* = x/t = 0 \tag{1.22}$$

En combinant les relations (1.20)-(1.22), on tire les expressions suivantes pour ρ_*, u_* et p_* :

$$u_* = c_* = \frac{\gamma-1}{\gamma+1}\left[u_{LO} + \frac{2c_{LO}}{\gamma - 1}\right] \tag{1.23}$$

$$p_* = p_2 \left(\frac{c_*}{c_{LO}}\right)^{2\gamma/(\gamma-1)} \tag{1.24}$$

$$\rho_* = \frac{\gamma p_*}{c_*^2} \tag{1.25}$$

Pendant une durée trés courte Δt_o, le gaz brûlé va s'échapper du tube avec une densité ρ_*, une vitesse u_*, et une pression p_*, qui vont en diminuant, jusqu'à ce que dans tout le tube, le gaz restant se retrouve au repos, avec une densité ρ_2 et une pression égale à la pression statique p_{s2}.

On récupère, cette fois ci, un problème de Riemann classique $PR(U_{L1}, U_{R1})$, où les états initiaux sont donnés par :

$$U_{L1} = \begin{pmatrix} \rho_2 \\ 0 \\ p_{s2} \end{pmatrix} \qquad \text{et} \qquad U_{R1} = \begin{pmatrix} \rho_{air} \\ u_{air} \\ p_{air} \end{pmatrix}$$

qu'on peut résoudre de la même façon. La solution comporte une onde de raréfaction qui remonte en amont et ramène la pression à 1 atm. La valeur stationnaire du nouveau problème de Riemann $PR(U_{L1},U_{R1})$ est obtenue en résolvant les mêmes équations (1.20) et (1.21) que précèdemment. On obtient :

$$u_* = c_* = \frac{\gamma-1}{\gamma+1}\left[u_{L1} + \frac{2c_{L1}}{\gamma - 1}\right] \tag{1.26}$$

$$p_* = p_2 \left(\frac{c_*}{c_{L1}}\right)^{2\gamma/(\gamma-1)} \tag{1.27}$$

$$\rho_* = \frac{\gamma\, p_*}{c_*^2} \tag{1.28}$$

où on a posé $c_{L1} = \left(\frac{\gamma\, p_{L1}}{\rho_{L1}}\right)^{1/2}$.

Aprés la courte phase transitoire entre t_o et $t_o+\Delta t_o$, suit une phase d'une durée Δt_o, pendant laquelle le gaz restant dans le canon sort à une densité, une vitesse et une pression constantes, et égales respectivement à ρ_*, u_* et p_*. Les fonctions $\rho(x=0;t)$; $u(x=0;t)$; et $p(x=0;t)$ présentent ainsi un palier entre les instants : $t_1 = t_o+\Delta t_o$, et $t_2 = t_1+\Delta t_1$.

Δt_1 étant le temps nécessaire à l'onde de raréfaction (de vitesse négative), pour atteindre le fond du tube. Il est donc facile à estimer, la vitesse de la 1-détente étant donnée par : $\lambda_1 = u_{L1} - c_{L1} = -c_{L1}$, car u_{L1} est nulle. On obtient donc, pour Δt_1 :

$$\Delta t_1 = \frac{L}{c_{L1}}$$

Finalement, les courbes de pression, vitesse et densité en fonction du temps, en sortie du canon, ont les allures présentées par les figures 1.9.

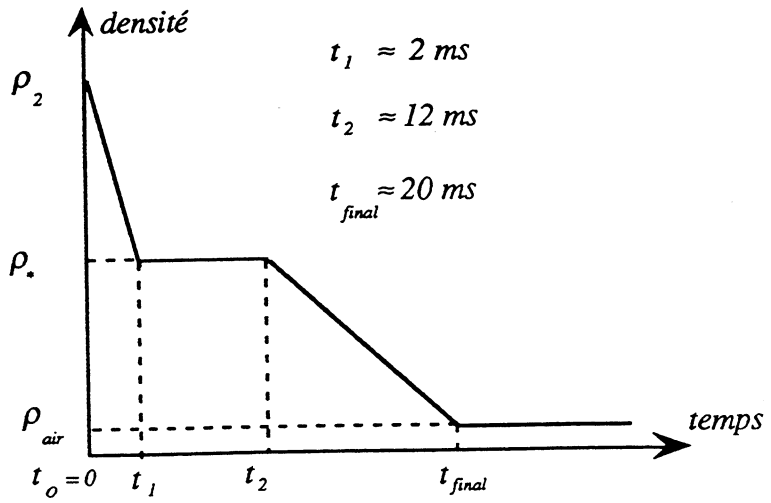


Figure 1.9.a. Densité en sortie du tube pendant la détonation

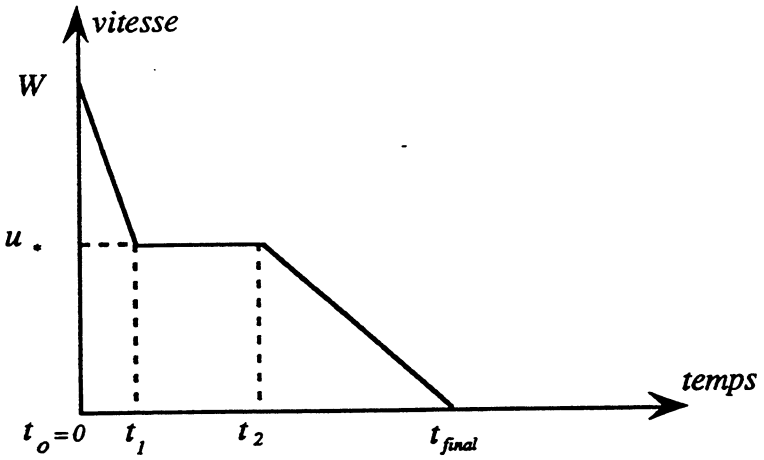


Figure 1.9.b. Vitesse en sortie du tube pendant la détonation

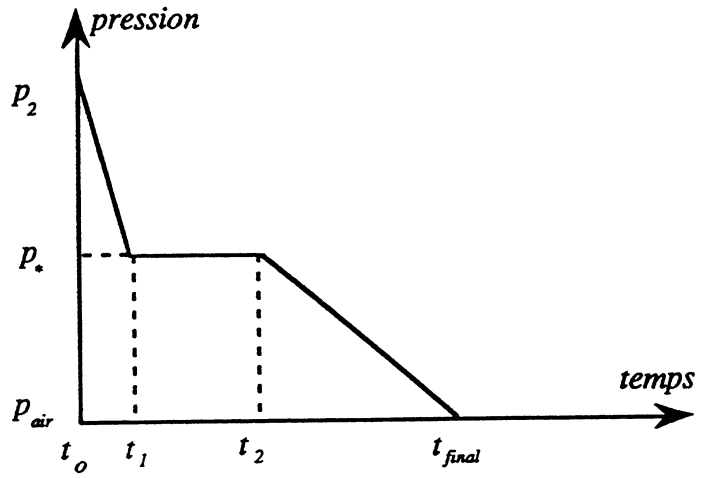


Figure 1.9.c. Pression en sortie du tube pendant la détonation

1.3.5. Validation numérique

Pour vérifier la cohérence des résultats concernants la condition à la limite en sortie du canon, on a effectué une simulation numérique du problème. On a résolu numériquement le problème de Cauchy associé au système de la dynamique des gaz compressibles, avec comme condition initiale, l'état du gaz dans le canon à l'instant t_o, i.e. à l'instant où l'onde de détonation parvient au bout ouvert du tube.

Le domaine de discrétisation est représenté par la figure 1.8. Il s'agit de la juxtaposition de deux parties, l'une interne et l'autre externe au tube. La partie interne est un rectangle représentant une coupe longitudinale de l'intérieur du canon, où on impose des conditions aux limites de type reflexion sur les parois latérales (L1)-(L2), et sur la face du fond (F). La partie externe est une portion rectangulaire de l'atmosphère avoisinant le tube. Sur les bords (L3)-(L7) de la zone externe, on a imposé une condition à la limite de type sortie libre (pour les détails du traitement des conditions aux limites, voir [Vil.1] par exemple), tandis qu'à l'intérieur de cette zone, la condition initiale est prise égale aux conditions atmosphériques "normales", i.e.

$p = 1$ atm

$u = 0.0$ m.s^{-1}

$\rho = 1.29$ kg/m^3

Cette modélisation est parfaitement justifiée, étant donné qu'à partir de l'instant t_0, plus aucune réaction chimique dégageant de l'énergie ne se produit à l'intérieur du canon. Les équations qui modélisent le phénomène physique sont bien les équations d'Euler complètes.

Les résultats numériques sont représentés par les courbes de la figure 1.9.

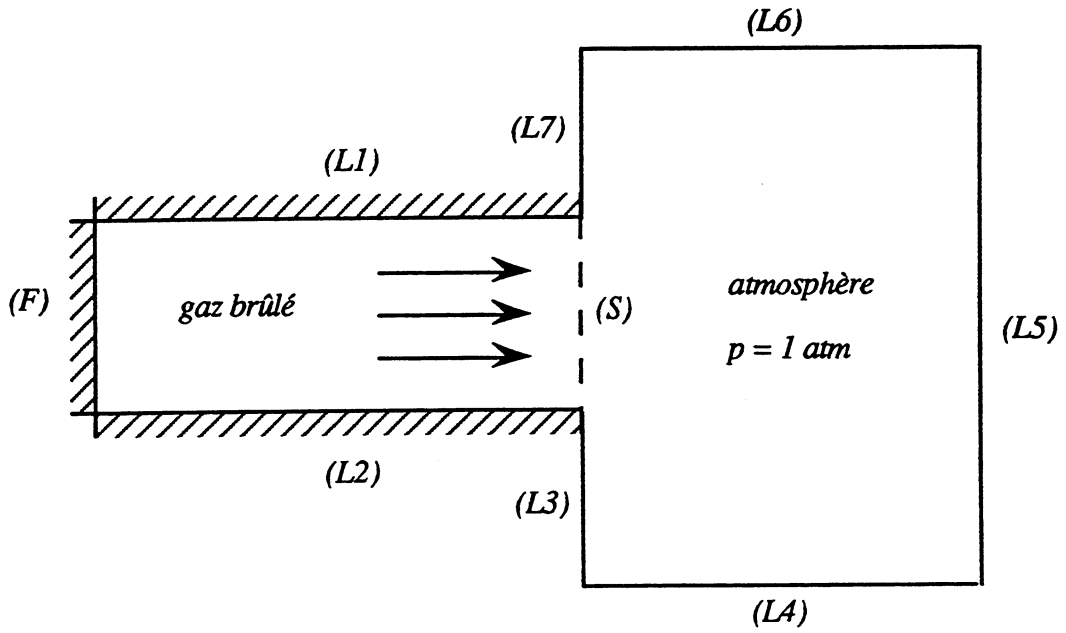


Figure 1.8.

1.4. Résultats numériques

Cette partie est consacrée à la présentation des résultats de la simulation numérique. On a utilisé un maillage rectangulaire irrégulier pour mailler le domaine Ω de la figure 1.2. Les mailles sont plus fines au voisinage de la bouche du canon, de taille minimale: $\Delta x_{min} = \Delta y_{min} = r$ (rayon du canon). A fur et à mesure que l'on s'éloigne du canon, la taille des mailles augmente, en suivant une progression géometrique, jusqu'à atteindre la valeur maximale de $\Delta x_{max} = \Delta y_{max} = 1.00$ m.

On s'est intéressé à deux types de résultats:

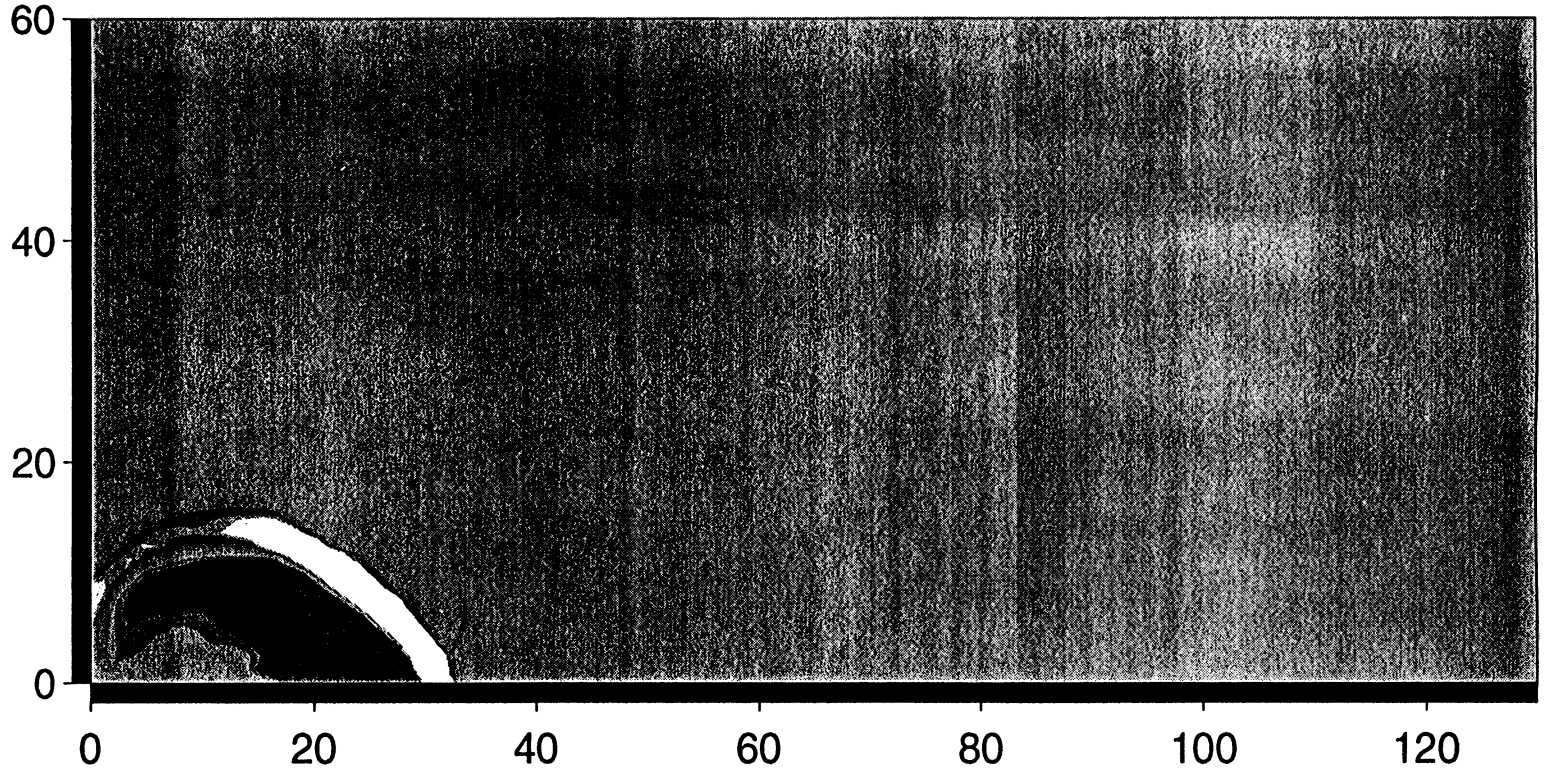
- Les courbes de "pressions instantanées", correspondants à la distribution de pression, à un instant donné, dans tout le domaine Ω.
- Les "enveloppes de pressions", correspondants à la valeur maximale de pression atteinte en chaque point du domaine Ω pendant toute la durée de l'expérience.

Le schéma utilisé pour discrétiser les équations d'Euler est un schéma aux volumes finis du second ordre en espace et en temps. Les flux numériques sont calculés à l'aide du solveur de Roe (Cf Chap.IV). La CFL utilisée est de 0.45. Le pas de temps est variable, calculé à l'aide de la CFL et de la relation suivante

$$\frac{Dt}{\min(Dx)} \underset{U=(u,v)}{Sup} \left(\sqrt{u^2+v^2} + c\right) = CFL = 0.45$$

Les figures des pages suivantes représentent les résultats obtenus à des instants différents (tout les 30 ms). Les valeurs de surpression ou de depression qui aparaissent dans les légendes des différentes courbes sont exprimées en millibars.

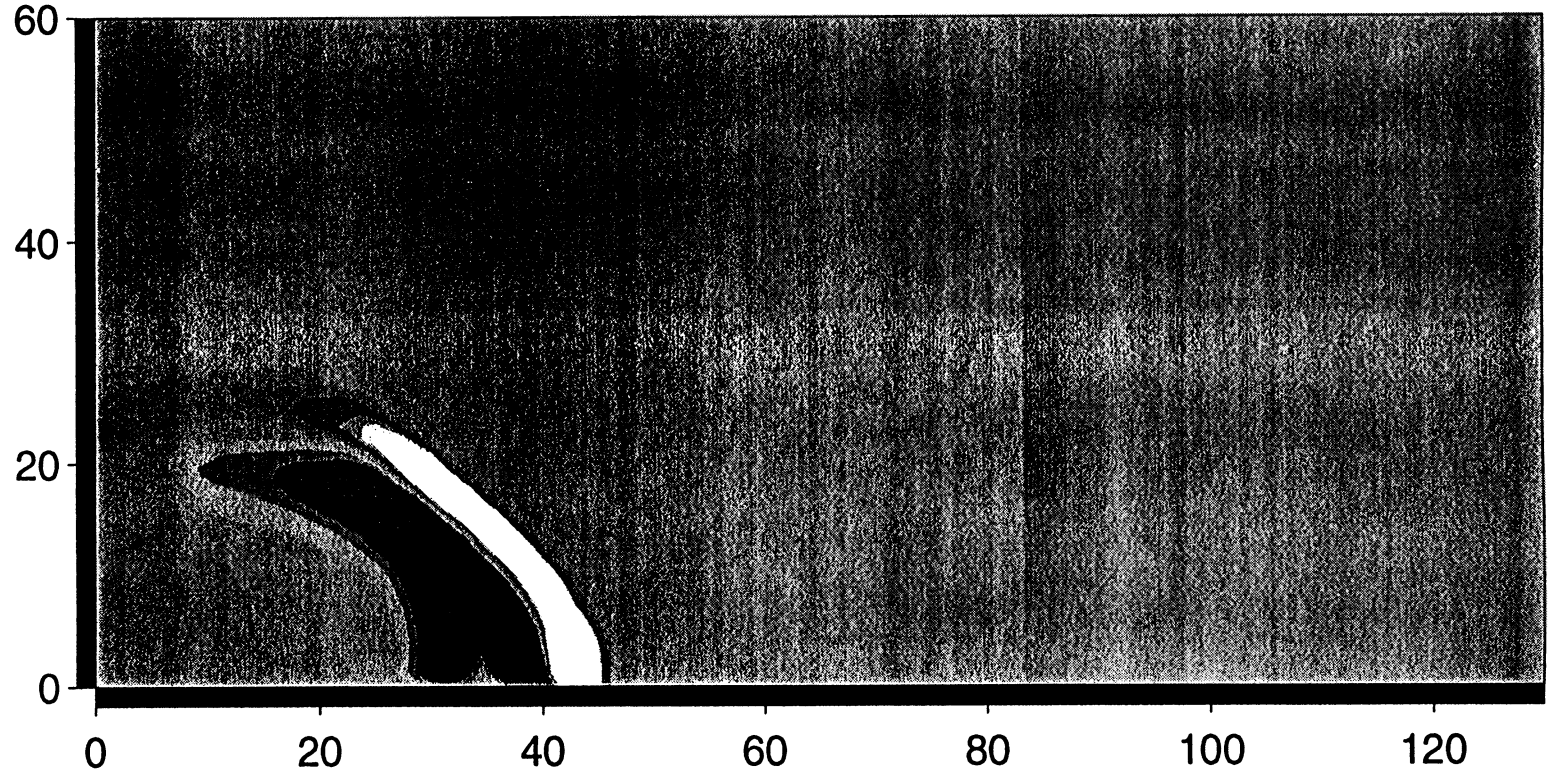
- PRESSIONS INSTANTANEES APRES 30 ms -



ABOVE 40
35 - 40
30 - 35
25 - 30
20 - 25
15 - 20
10 - 15
5 - 10
0 - 5
-5 - 0
-10 - -5
-15 - -10
BELOW -15

- TUBE : 1.5 M3 -

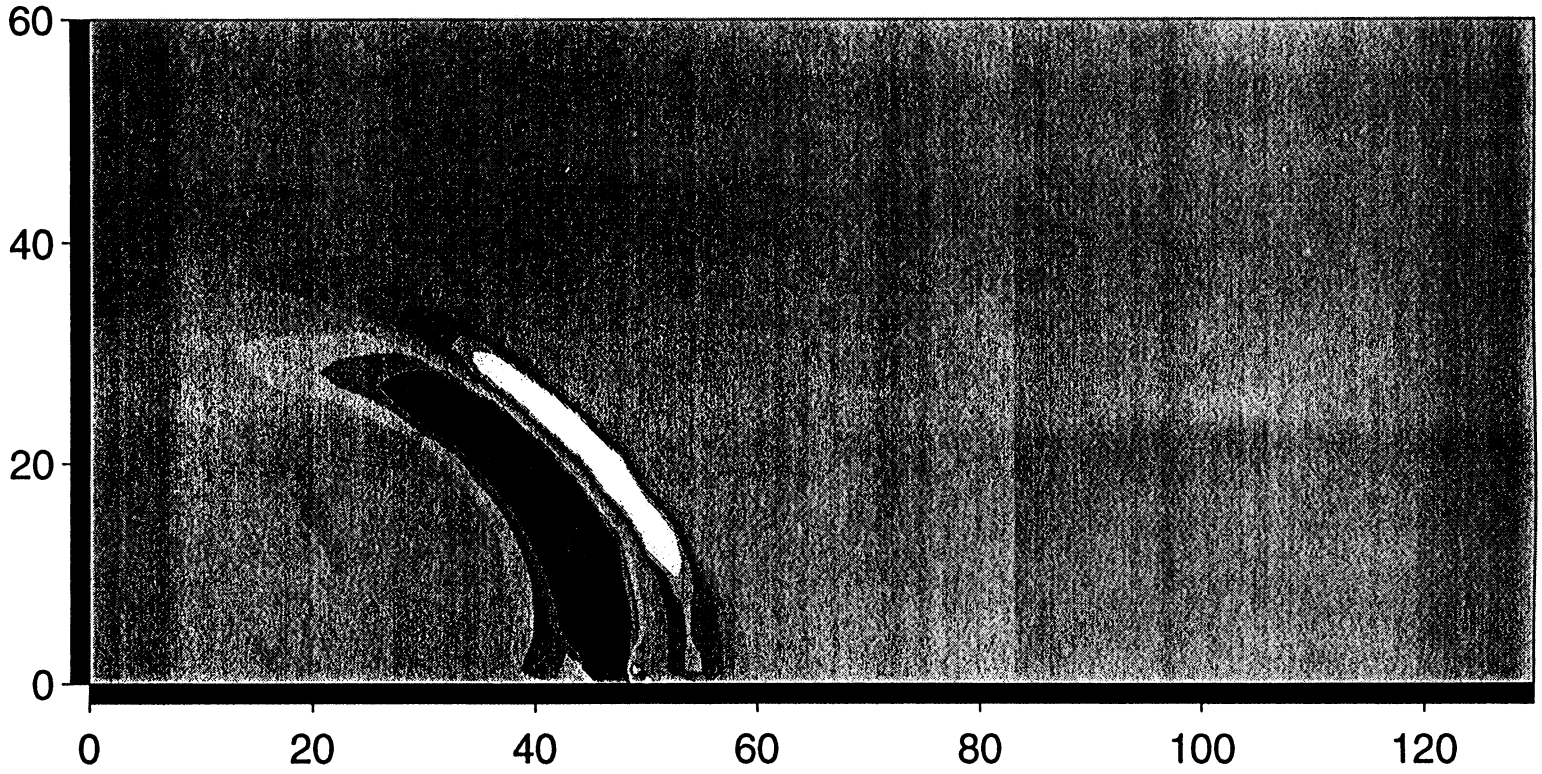
- PRESSIONS INSTANTANEES APRES 60 ms -



ABOVE 40
35 - 40
30 - 35
25 - 30
20 - 25
15 - 20
10 - 15
5 - 10
0 - 5
-5 - 0
-10 - -5
-15 - -10
BELOW -15

- TUBE : 1.5 M3 -

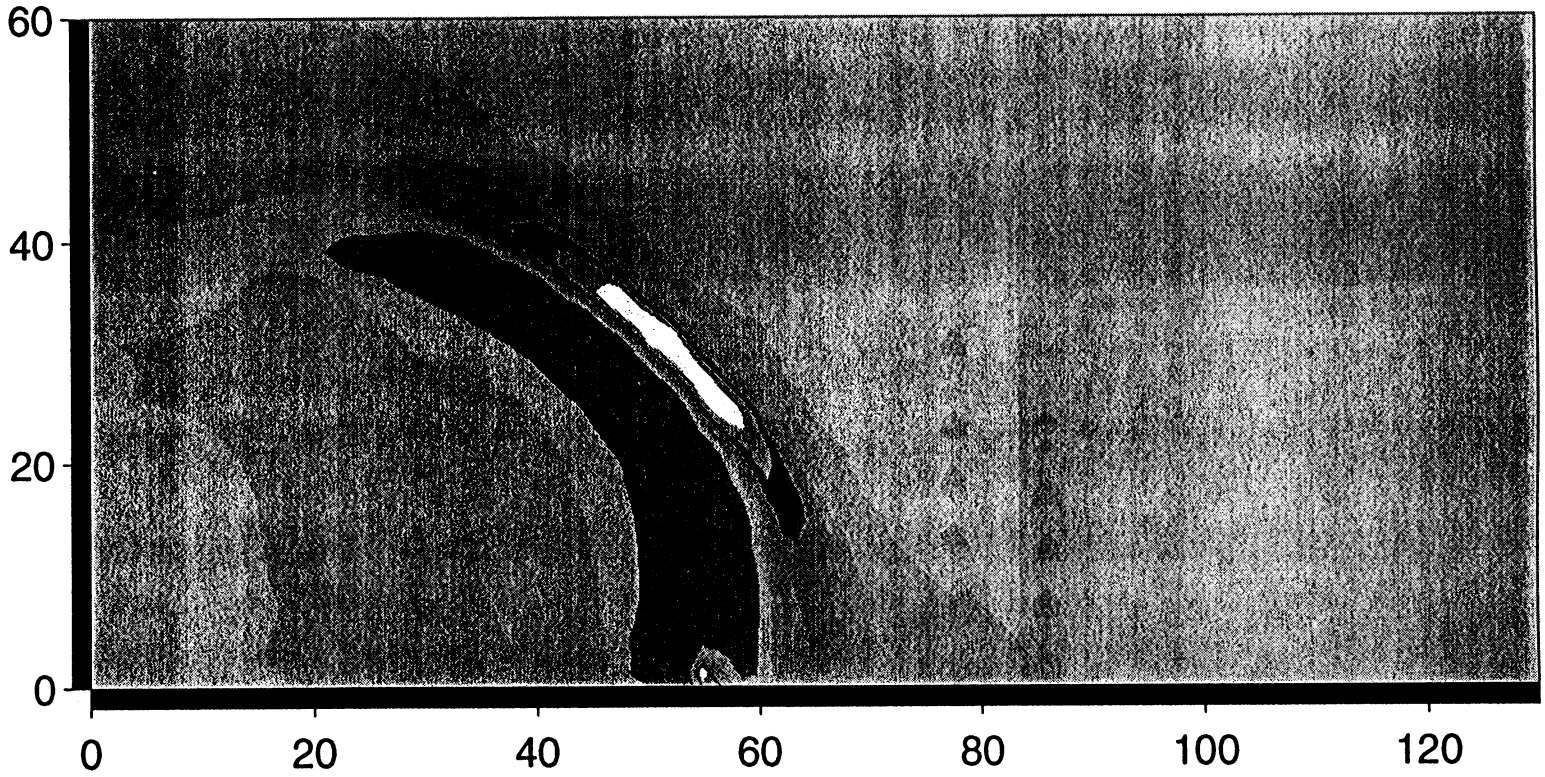
- PRESSIONS INSTANTANEES APRES 90 ms -



ABOVE 40
35 - 40
30 - 35
25 - 30
20 - 25
15 - 20
10 - 15
5 - 10
0 - 5
-5 - 0
-10 - -5
-15 - -10
BELOW -15

- TUBE : 1.5 M3 -

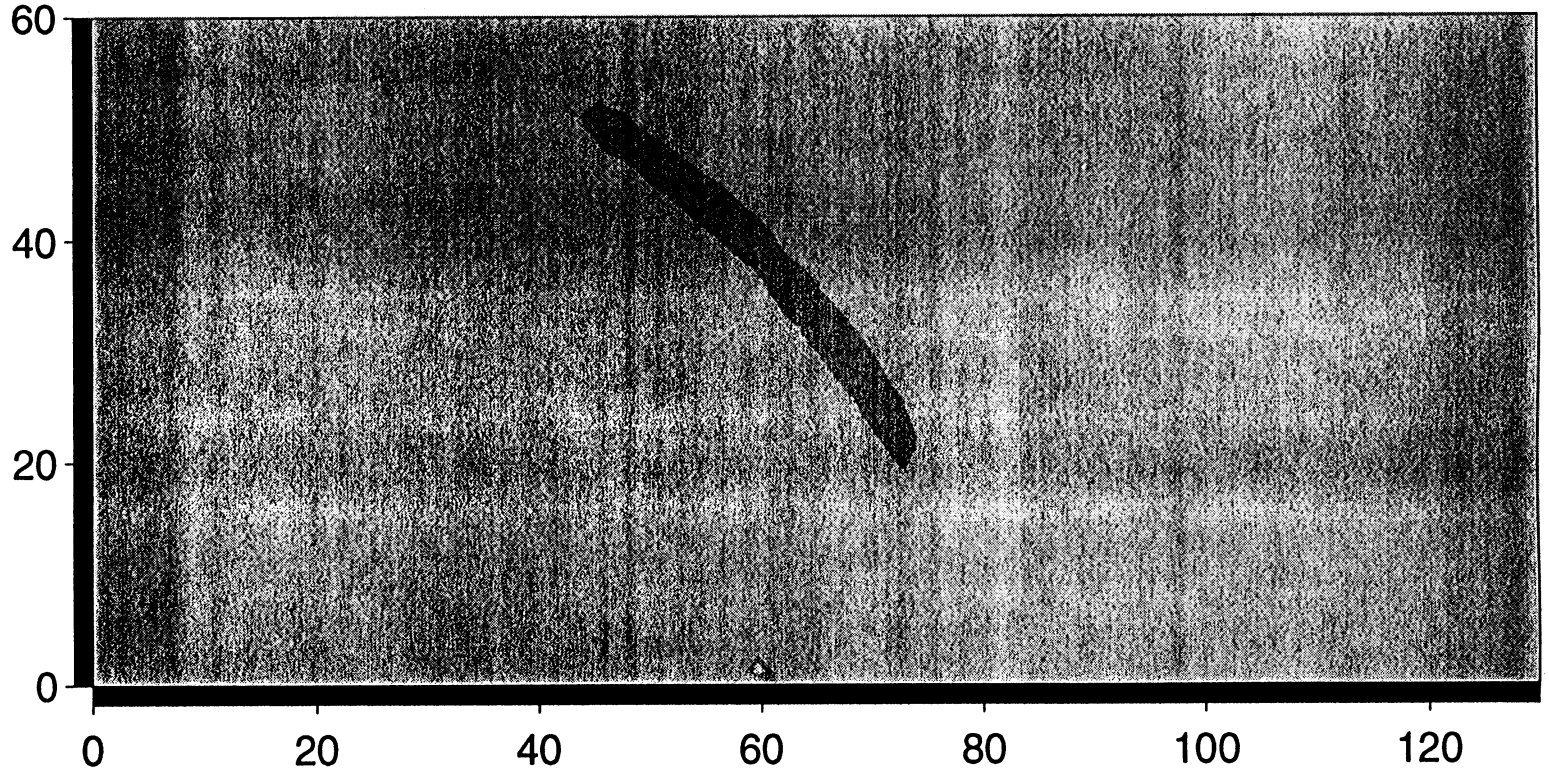
- PRESSIONS INSTANTANEES APRES 120 ms -



ABOVE 30
25 - 30
20 - 25
15 - 20
10 - 15
5 - 10
0 - 5
-5 - 0
BELOW -5

- TUBE : 1.5 M3 -

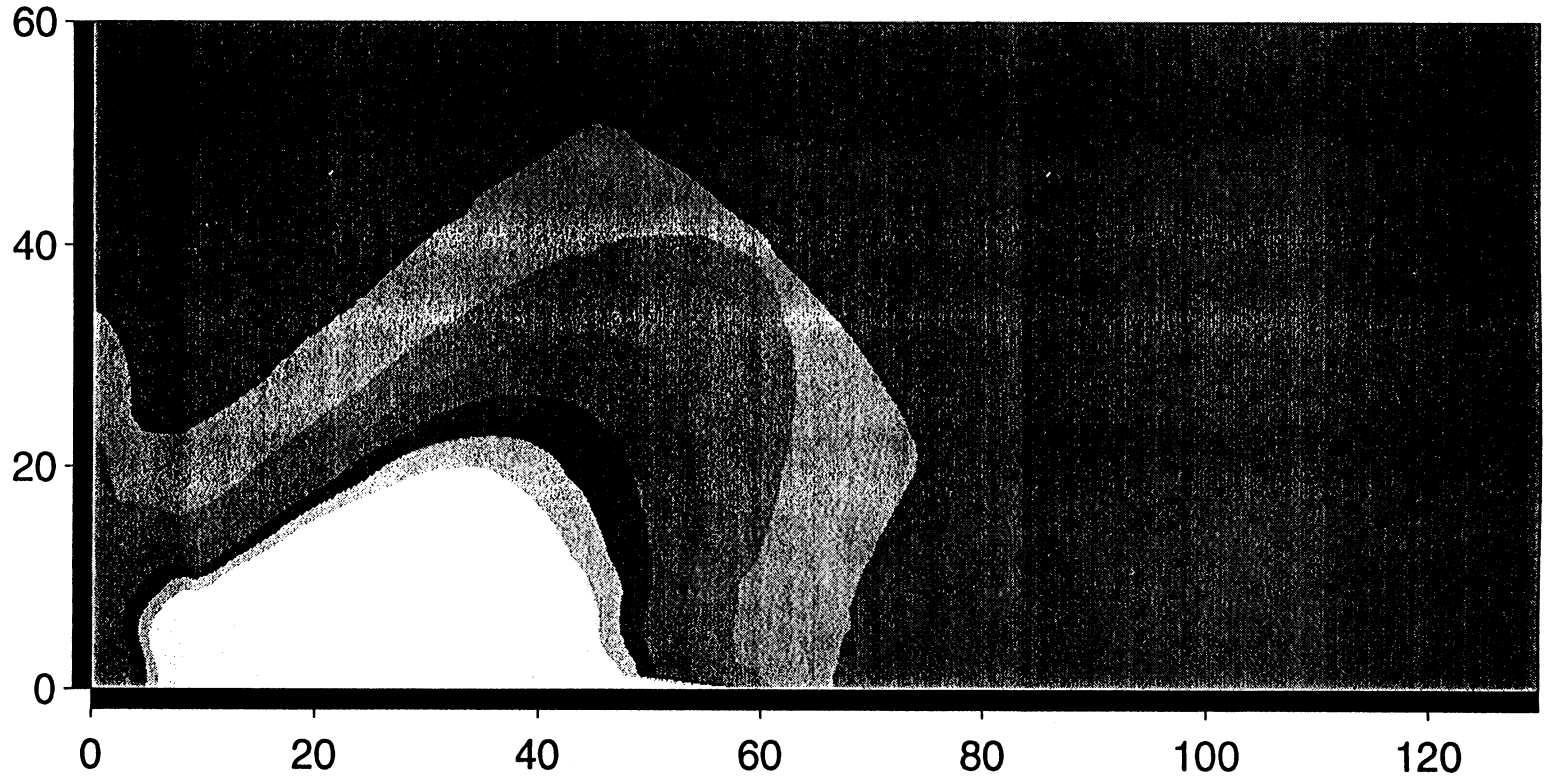
- PRESSIONS INSTANTANEES APRES 150 ms -



- TUBE : 1.5 M3 -

ABOVE 20
10 - 20
BELOW 10

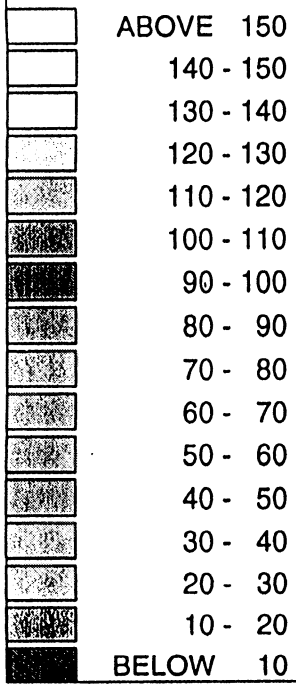
- ENVELOPPE DE PRESSION APRES 150 ms (1.5 M3) -

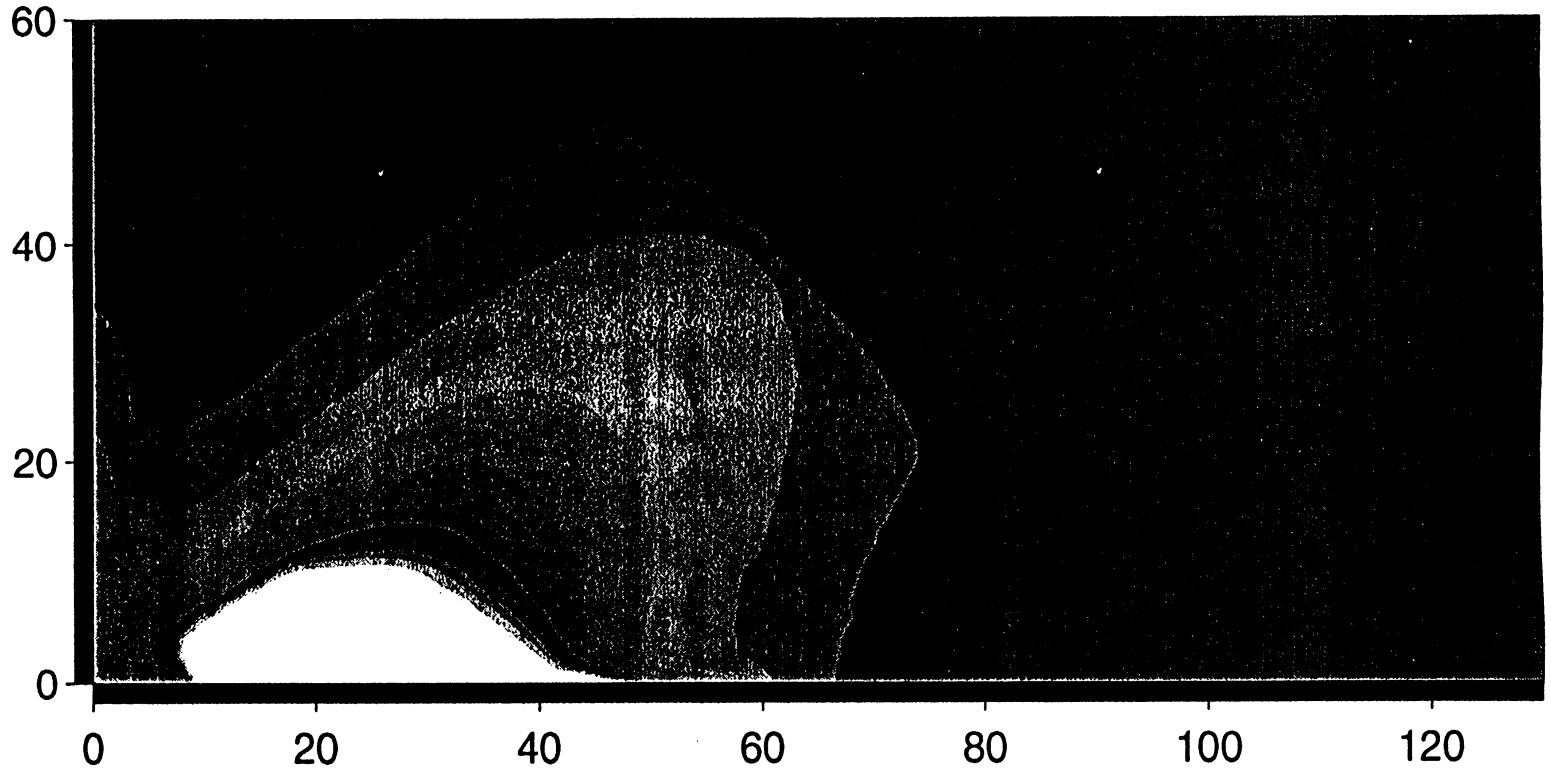


ABOVE 60
50 - 60
40 - 50
30 - 40
20 - 30
10 - 20
BELOW 10

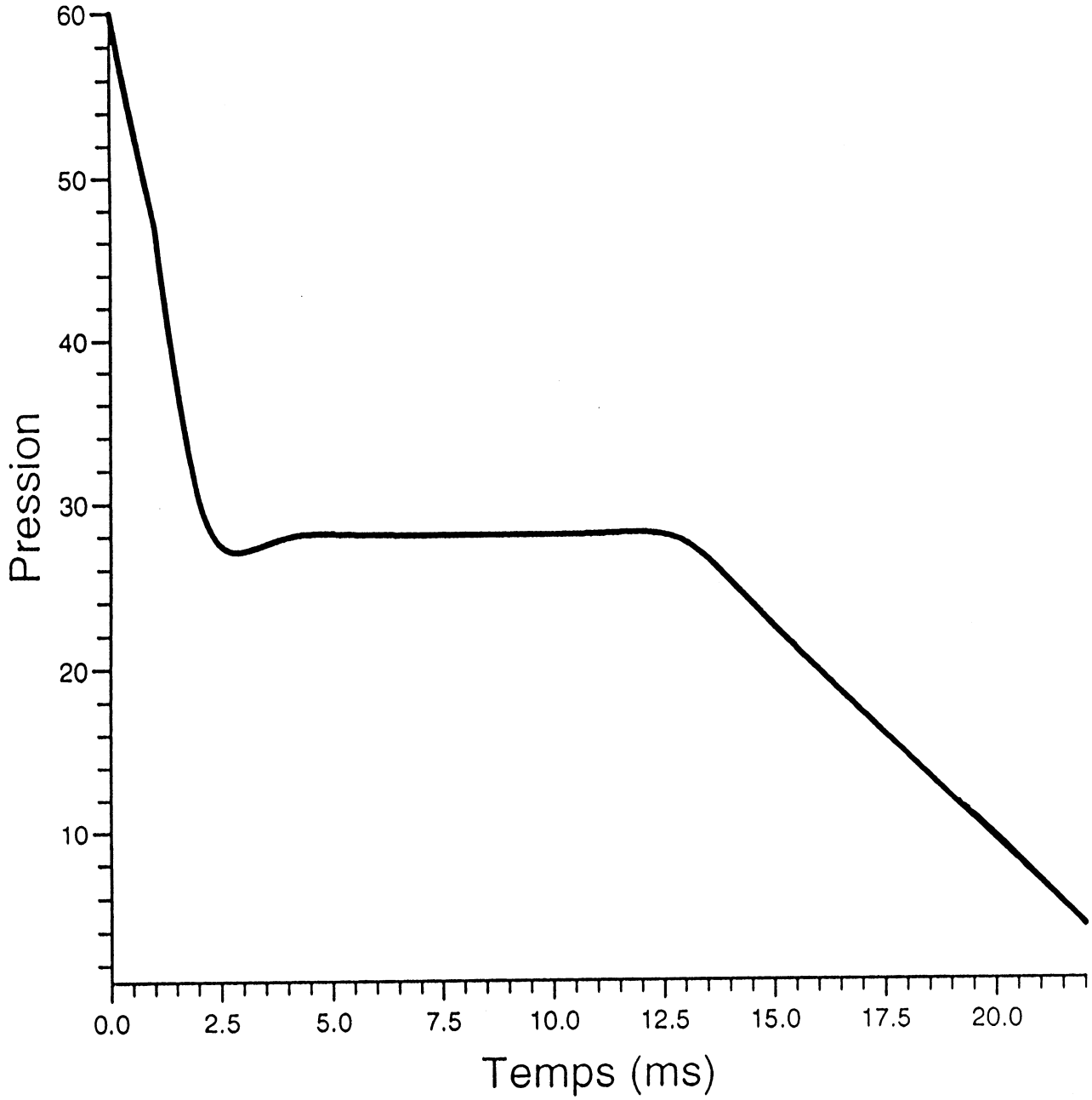
- TUBE : 1.5 M3 -

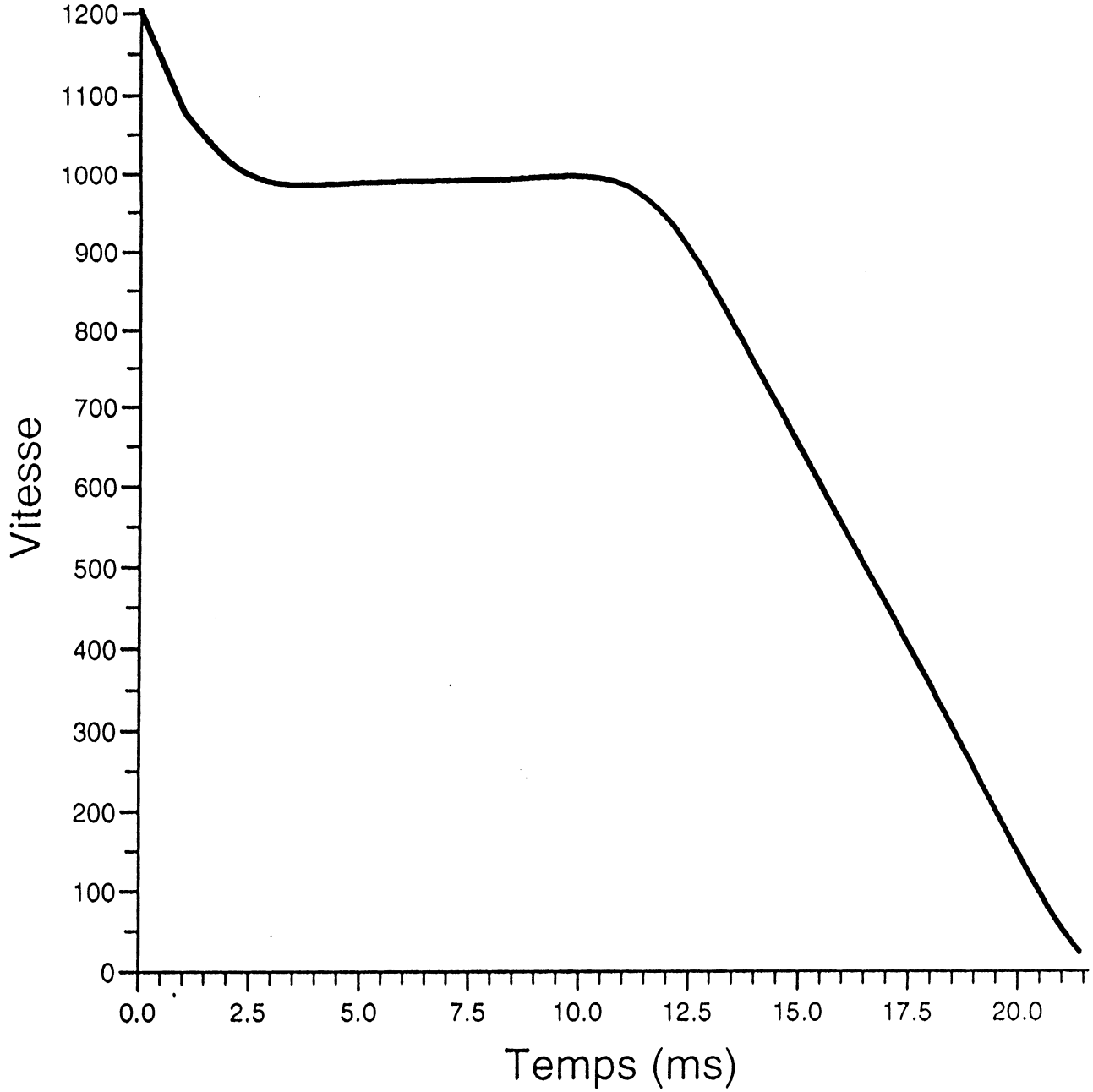
- ENVELOPPE DE PRESSION APRES 150 ms (1.5 M3) -

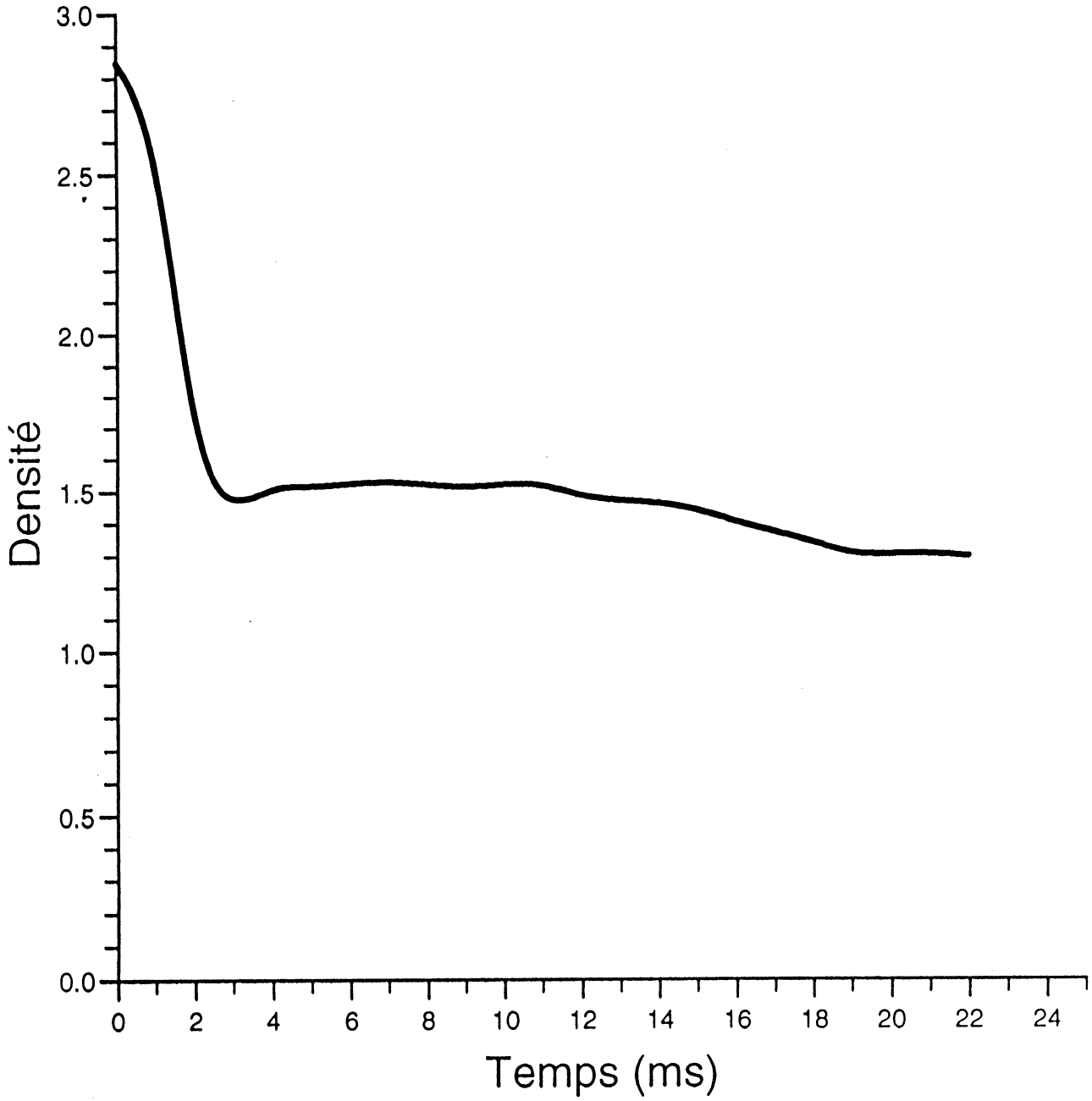




- TUBE : 1.5 M3 -


60
50
40
30
20
10
Pression
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Temps (ms)


1200
1100
1000
900
800
700
600
500
400
300
200
100
0
Vitesse
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Temps (ms)


3.0
2.5
2.0
1.5
1.0
0.5
0.0
Densité
0
2
4
6
8
10
12
14
16
18
20
22
24
Temps (ms)

2. Etude d'un écoulement transitoire compressible dans un puit de cuve

2.0. Introduction

Dans le cadre de l'étude de quelques problèmes liés à la sécurité dans les centrales nucléaires, Le SEPTEN, en collaboration avec EDF, a envisagé le cas de la rupture des parois de protection de la cuve centrale, ce qui entraîne la formation d'un jet à trés grand débit, à travers la brèche, qui percute la cuve (voir figure 2.1). La configuration schématique du puit de cuve retenue pour la modélisation est celle représentée par le schéma ci-dessous (choisie en concertation avec le SEPTEN).

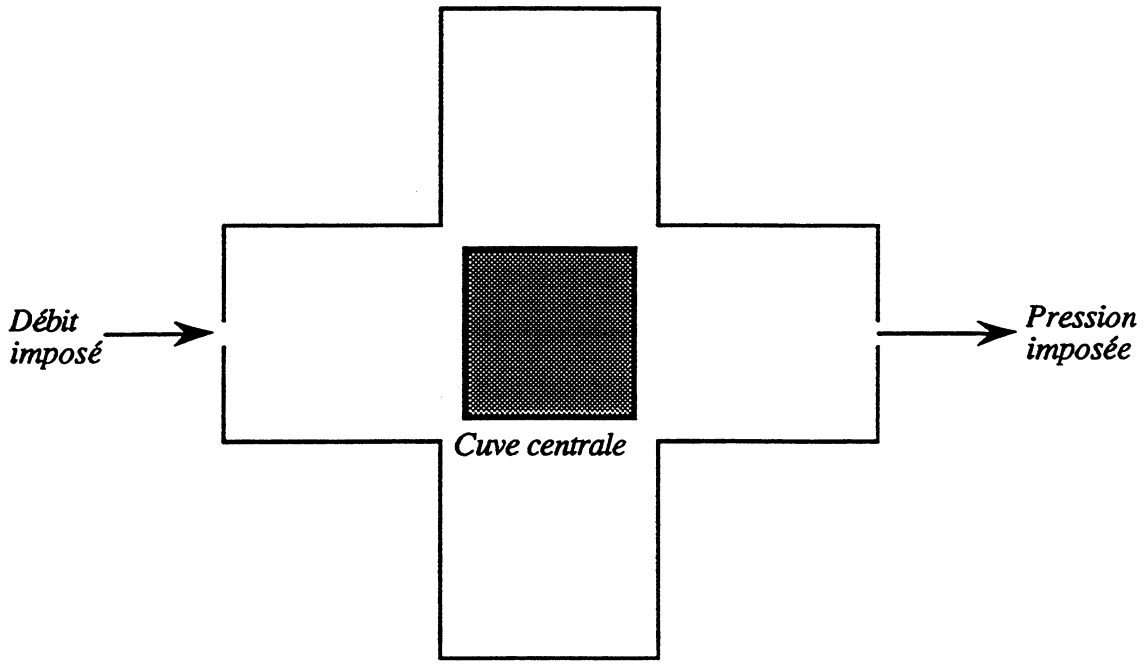


Figure 2.1. Configuration schématique du puit de cuve

C'est dans le cadre d'un groupe de reflexion sur la modélisation des écoulements compressibles, que nous avons entrepris, pour le compte d'EDF, l'étude du cas test du puit de cuve.

2.1. Position du problème

Le fluide injecté est de l'air. Les données du problème sont les suivantes :

- Débit massique imposé à la brèche d'entrée : $Q_0 = 450.00\ kg.s^{-1}.m^{-2}$,
- Enthalpie imposée à la brèche d'entrée : $H_0 = 450.00\ kJ.kg^{-1}$,
- Pression en sortie constante et égale à 1 bar.

Les principaux résultats que l'on désire obtenir à l'aide de la modélisation numérique sont :

- Le champ de pression à l'intérieur de la cuve,
- Les efforts de pression sur la cuve centrale.

Toute la difficulté ici est de transcrire ce problème en termes mathématiques de problème aux conditions aux limites. Ce sera le but de l'étude qui va suivre.

2.2. Modélisation mathématique

Le fluide étudié (air) est assimilé à un gaz parfait de rapport de chaleurs spécifiques 1,4. L'écoulement est alors modélisé par les équations d'Euler bidimensionnelles :

$$\begin{cases} \frac{\partial \rho}{\partial t} + \frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial y} = 0 \\ \frac{\partial \rho u}{\partial t} + \frac{\partial (\rho u^2 + p)}{\partial x} + \frac{\partial \rho u v}{\partial y} = 0 \\ \frac{\partial \rho v}{\partial t} + \frac{\partial \rho u v}{\partial x} + \frac{\partial (\rho v^2 + p)}{\partial y} = 0 \\ \frac{\partial E}{\partial t} + \frac{\partial u(E+p)}{\partial x} + \frac{\partial v(E+p)}{\partial y} = 0 \end{cases} \qquad (2.1)$$

où ρ désigne la densité du gaz, $\vec{U} = (u, v)^T$ est le vecteur vitesse en coordonnées cartésiennes, E et p désignent respectivement l'énergie totale et la pression du gaz.

2.3. Condition à la limite amont : brèche d'entrée

2.3.1. Problème modèle : le demi-problème de Riemann

La technique généralement adoptée pour traiter ce type de condition à la limite, est de considérer un problème modèle, celui du demi-problème de Riemann. Il s'agit du problème suivant :

$$\begin{cases} U_t + f(U)_x = 0 \\ U(x,0) = U_R \qquad (x,t) \in \mathbb{R}_+ \times \mathbb{R}_+ \\ \varphi(U(0,t)) = 0 \end{cases} \qquad (2.2)$$

$\varphi(U(0,t))$ est la condition à la limite en x=0. Ici φ est la fonction régulière, à valeur dans $\mathbb{R}^2$, donnée par :

$$\varphi(U) = \begin{pmatrix} Q - Q_o \\ H - H_o \end{pmatrix} \qquad (2.3)$$

Ce problème est l'analogue d'un problème de Riemann, mais il est posé dans le quadrant droit ($x \geq 0, t \geq 0$) (voir figure 2.2). Comme le problème de Riemann, ce problème est invariant par homothétie de l'espace et du temps. Il est donc naturel de chercher la solution sous la forme U(x/t) et d'étudier l'existence et l'unicité des solutions entropiques de ce problème dans l'espace des fonctions C^1 par

morceaux (où l'on sait que le problème de Riemann a une solution unique) et en imposant :

$\lim_{x/t \to 0^+} U(x/t) = U_L$, avec U_L vérifiant la condition à la limite : $\varphi(U_L)=0$.

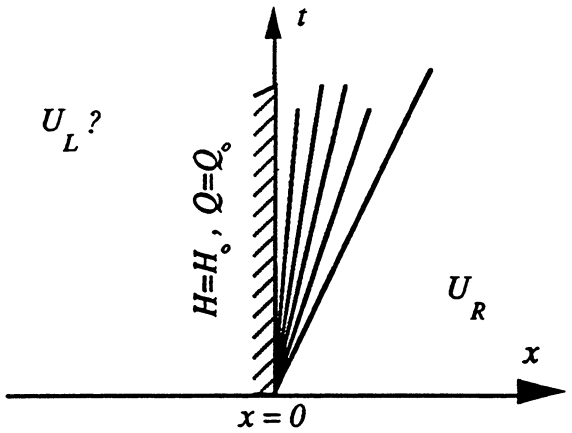


Figure 2.2. Problème modèle : Demi-problème de Riemann

L'idée de résolution est la suivante : il suffit de trouver un état gauche U_L tel que la restriction de la solution $w(x/t; U_L,U_R)$ du problème de Riemann $PR(U_L,U_R)$ au quadrant ($x\geq 0$, $t\geq 0$) soit solution de (2.2)-(2.3), autrement dit tel que : $\varphi(w(0; U_L,U_R))=0$. Pour cela, il suffit que toutes les ondes du problème de Riemann $PR(U_L,U_R)$ soient de vitesse positive et que $\varphi(U_L)=0$.

Nous avons démontré le théorème suivant :

Théorème 2.1.

Le problème (2.2)-(2.3) a une infinité de solutions entropiques dans la classe des fonctions C^1 par morceaux de x/t et continue en x=0, si

$$u_R - \frac{2c_R}{\gamma - 1} < A\sqrt{2H_o} \qquad (2.4)$$

avec : $A = \sqrt{(\gamma+1)/(\gamma-1)}$.

Parmi toutes ces solutions, il en existe une seule d'entropie (physique) maximale en x/t=0. Elle est constituée de la restriction de la solution du problème de Riemann $PR(U_L,U_R)$ au quadrant ($x\geq 0, t\geq 0$), où l'état U_L est l'état sonique donné par :

$$\begin{cases} u_L = \sqrt{\alpha H_o} & \text{avec} \quad \alpha = 2\,\dfrac{\gamma-1}{\gamma+1} \\ \varphi(U_L)=0 \end{cases} \qquad (2.5)$$

Si (2.4) n'est pas vérifiée, toutes les solutions du problème (2.2)-(2.3) contiennent le vide.

Remarque 2.1.

En pratique, la vitesse du gaz à l'intérieur du puit de cuve ne dépasse jamais la borne $A\sqrt{2H_o}$. Par conséquence, la condition (2.4) est, à fortiori, toujours vérifiée.

La démonstration de ce théorème comporte plusieurs étapes, dans un premier temps (voir §.2.3.2.) sera étudiée dans le plan (u,p) (afin d'ignorer les discontinuités de contact des différents problèmes de Riemann que l'on sera amenés à résoudre) la courbe paramètrée par la relation (2.3), ensuite dans le §.2.3.3. seront recherchés sur cette cette courbe ceux qui sont "admissibles" pour le problème (2.2-2.3), et finalement sera sélectionné au §.2.3.4., suivant le critère enoncé dans le théorème 2.1., l'état d'entropie maximale.

La première étape de la résolution du problème de Riemann $PR(U_L,U_R)$, pour la dynamique des gaz parfaits polytropiques (Cf. Chap.I.§.4), est l'étude dans le plan (u,p), des courbes $\Sigma_1(U_L)$ (resp. $\Sigma_3(U_R)$) des états qu'on peut relier à U_L (resp. U_R) par un 1-choc ou une 1-détente (resp. 3-choc ou une 3-détente). La discontinuité de contact n'apparaissant pas dans ce plan, puisque u et p sont des 2-invariants de Riemann. Il est donc intéressant d'étudier, dans le plan (u,p), la courbe Γ_o définie par la paramètrisation : $\varphi(U)=0$.

2.3.2. Etude dans le plan (u,p) de la courbe Γ_o

L'enthalpie H, et l'energie totale E du gaz sont données par :

$$H=\frac{E+p}{\rho} \qquad\qquad E=\frac{p}{\gamma-1}+\frac{1}{2}\rho u^2$$

Un point $U=(\rho,u,p)^T$ appartient à la courbe Γ_o s'il vérifie :

$$\begin{cases}\rho = Q_o/u \\ H_o = (E+p)/\rho\end{cases} \tag{2.6}$$

aprés élimination de ρ et E dans (2.6), on obtient la relation suivante entre u et p :

$$p = \Psi(u) = \frac{\gamma-1}{\gamma}\left[H_o - \frac{u^2}{2}\right]\frac{Q_o}{u} \tag{2.7}$$

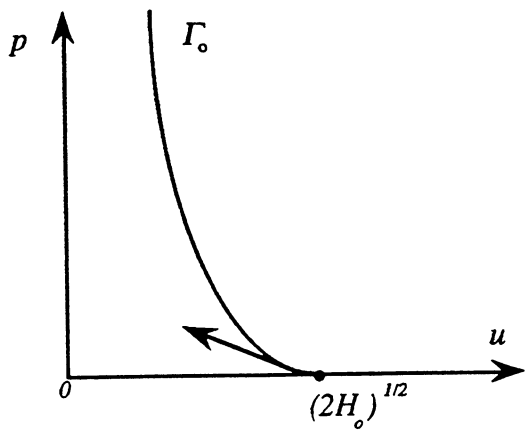


Figure 2.3. Courbe Γ_o dans le plan (u,p)

L'étude de la fonction Ψ montre que l'on a les résultats suivants :

$$\begin{cases}\text{(i) } \lim\limits_{u\to 0}\Psi(u) = \infty & \text{(iii) } \Psi'\left[\sqrt{2H_o}\right] = -\frac{\gamma-1}{\gamma}Q_o < 0 \\ \text{(ii) } \Psi\left[\sqrt{2H_o}\right] = 0 & \text{(iv) } \left(\Psi(u)\geq 0\right) <=> \left(0\leq u\leq\sqrt{2H_o}\right)\end{cases} \tag{2.8}$$

Dans le cas où la solution du problème $PR(U_L,U_R)$ contient une 1-onde (où U_L est un état vérifiant $\varphi(U_L)=0$), une condition nécessaire (voire nécessaire et suffisante) pour avoir : $w(0; U_L,U_R) = U_L$, est :

$$u_L - c_L \geq 0 \tag{2.9}$$

car la vitesse σ_1 d'une 1-onde de choc issue de U_L vérifie : $\sigma_1 \leq u_L - c_L$. Commençons donc par déterminer l'ensemble des état U de la courbe Γ_o vérifiant la condition (2.9).

Soit $U = (\rho,u,p)^T$, un état tel que $\varphi(U)=0$. En utilisant (2.6), la condition (2.9) peut s'écrire :

$$u - \sqrt{\frac{\gamma-1}{2}\left[2H_o - u^2\right]} \geq 0$$

ou encore, comme $u \geq 0$

$$u^2 \geq \alpha H_o \tag{2.10}$$

avec : $\alpha = 2\dfrac{\gamma-1}{\gamma+1} < 2.$

Les relations (2.8)-(iv) et (2.10) nous permettent de délimiter une zone dite "zone admissible" sur la courbe Γ_o (voir figure 2.4), définie par la relation :

$$\sqrt{\alpha H_o} \leq u \leq \sqrt{2H_o} \tag{2.11}$$

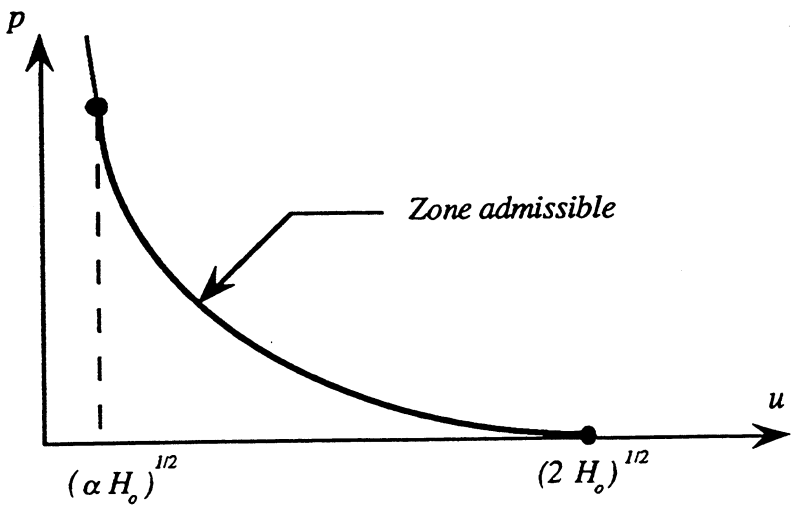


Figure 2.4.

2.3.3. Résolution du problème de Riemann

Dans cette partie sera étudiée, pour un état U_L choisi sur la zone admissible de Γ_o (voir figure 2.8), la solution du problème de Riemann $PR(U_L,U_R)$. Cette solution dépendera de la position dans le plan (u,p), de l'état U_R par-rapport à la courbe Γ_o.

Commençons par rappeler quelques propriétés des courbes $\Sigma_3(U_R)$ des états que l'on peut relier à U_R par une 3-onde (Cf. Chap.1.). La 3-courbe de choc issue de U_R, a pour équation dans le plan (u,p) :

$$u = u_R + (p - p_R)\left(\frac{(1-\mu^2)}{\rho_R\,[p + \mu^2 p_R]}\right)^{1/2} \qquad \text{avec : } \mu^2 = \frac{\gamma-1}{\gamma+1}.$$

La 3-courbe de détente passant par U_R, est représentée par les relations :

$$u - 2c/(\gamma-1) = u_R - 2c_R/(\gamma-1) \qquad p/\rho^\gamma = p_R/\rho_R^\gamma$$

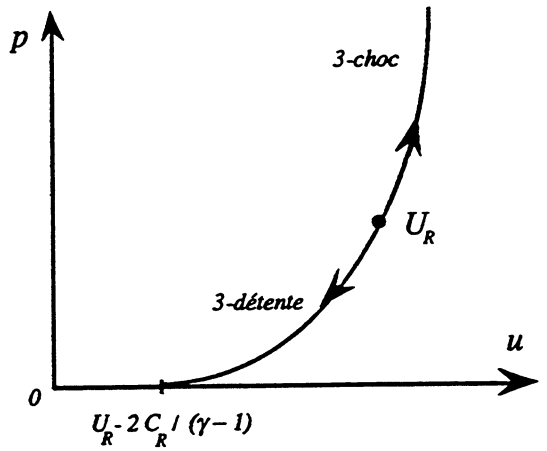


Figure 2.5. Courbe $\Sigma_3(U_R)$ dans le plan (u,p)

Au vu des allures de la courbe Γ_o et des 3-courbes dans le plan (u,p), on a le résultat suivant dont la démonstration est évidente et donc omise.

Proposition 2.1.

La courbe $\Sigma_3(U_R)$ a une intersection non vide avec la courbe Γ_o si et seulement si l'état U_R vérifie la condition suivante :

$$u_R - 2c_R/(\gamma-1) \le \sqrt{2H_o} \qquad (2.12)$$

Considérons maintenant un état U_L sur la courbe Γ_o . La courbe $\Sigma_1(U_L)$ des états que l'on peut relier à U_L par une 1-onde, est constituée de la courbe de détente issue de U_L et donnée par les relations :

$$u + 2c/(\gamma-1) = u_L + 2c_L/(\gamma-1) \qquad p/\rho^\gamma = p_L/\rho_L^\gamma$$

et de la 1-courbe de choc donnée par la relation suivante, (Cf. Chap.1.§.1) :

$$u = u_L - (p - p_L)\left(\frac{(1-\mu^2)}{\rho_L\,[p + \mu^2 p_L]}\right)^{1/2} \qquad (2.13)$$

Proposition 2.2.

Pour tout état U_L tel que $\varphi(U_L)=0$, on a :

$$\sqrt{2H_o} \le u_L + \frac{2}{\gamma-1}\,c_L \le A\sqrt{2H_o} \qquad (2.14)$$

avec : $A = \sqrt{(\gamma+1)/(\gamma-1)}$. La 1-courbe $\Sigma_1(U_L)$, passant par U_L, est au dessus (resp. au dessous) de la courbe Γ_o dans la zone $\{u \ge u_L\}$ (resp. $\{0 \le u \le u_L\}$).

Démonstration

Considérons un état U tel que $\varphi(U)=0$, alors il vérifie

$$p = \Psi(u) \qquad \text{pour } 0 \le u \le \sqrt{2H_o}$$

et donc $$u + \frac{2}{\gamma-1} c = f(u)$$

avec $$f(u) = u + \frac{2}{\gamma-1}\sqrt{\frac{\gamma-1}{2}\left[2H_o - u^2\right]} \qquad \text{pour } 0 \le u \le \sqrt{2H_o}$$

L' étude de la fonction f" montre que f' est décroissante sur l'intervalle $[0,\sqrt{2H_o}]$, et que $f'(\sqrt{\alpha H_o})=0$. Par conséquence, $f'(u)\le 0$ sur toute la zone admissible : $I_a = [\sqrt{\alpha H_o},\sqrt{2H_o}]$. La fonction f est donc décroissante sur I_a. Il est par ailleurs facil de vérifier que : $f(\sqrt{2H_o}) = \sqrt{2H_o}$, et $f(\sqrt{\alpha H_o}) = A\sqrt{2H_o}$, d'où le résultat annoncé.

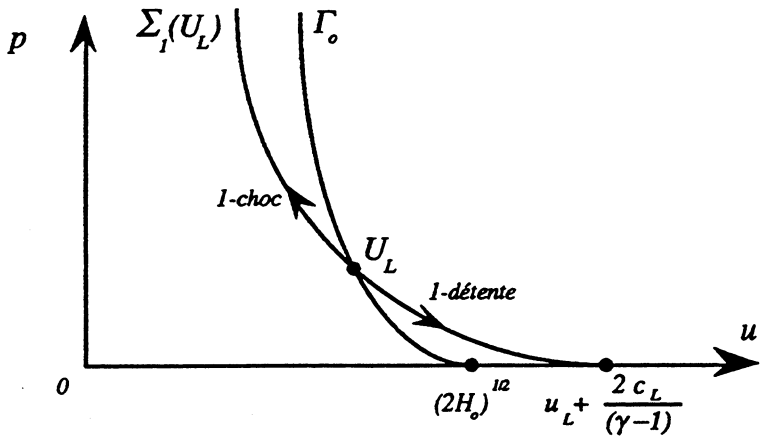


Figure 2.6. 1-courbe $\Sigma_1(U_L)$ dans le plan (u,p)

Or on sait (voir [Smo]) que la solution du problème de Riemann $PR(U_g,U_d)$ contient du vide si :

$$u_g + 2c_g/(\gamma-1) \le u_d - 2c_d/(\gamma-1),$$

par conséquence, si l'état U_R dans (2.2), vérifie : $A\sqrt{2H_o} \le u_R - 2c_R/(\gamma-1)$, alors la solution de (2.2)-(2.3) contiendra nécessairement le vide. Nous supposerons pour la suite que ce n'est pas le cas.

Posons pour toute la suite : $U_{R0} = \Gamma_o \cap \Sigma_3(U_R)$, $U_\bullet = \Sigma_1(U_L) \cap \Sigma_3(U_R)$

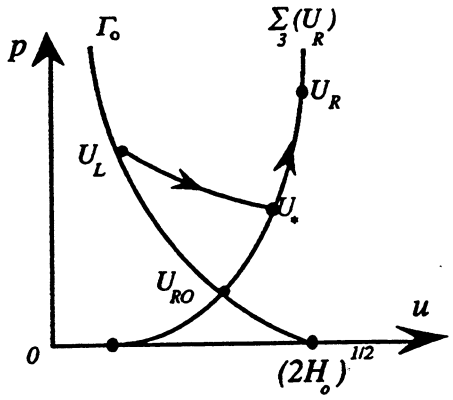
Avec ces notations, la solution de $PR(U_L,U_R)$, est constituée d'une 1-onde reliant U_L à $U_\bullet$ suivie d'une 3-onde reliant $U_\bullet$ à U_R, séparées par une discontinuité de contact.

Supposons dans un premier temps que l'état U_R est au dessus de la courbe Γ_o. Nous pouvons alors distinguer quatres cas de figures (voir figures (2.7.a)-(2.7.d)), suivant la position de l'état U_L dans la zone admissible définie par (2.11), par-rapport à U_{R0}.

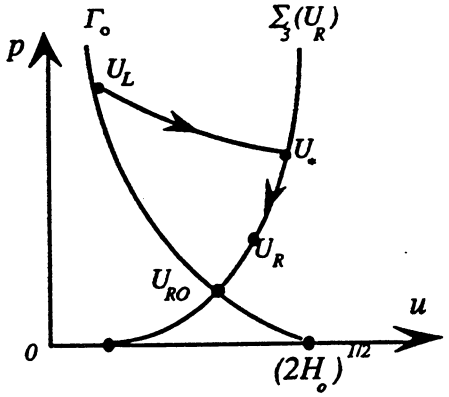
Dans le cas (2.7.a), où U_L est au dessus de l'état U_{R0}, $w(x/t; U_L,U_R)$ est constituée d'une 1-détente reliant U_L à U_*, suivie d'une 3-détente reliant U_* à U_R. Cette solution est "admissible" au sens que sa restriction au quadrant ($x\geq 0, t\geq 0$) est solution du problème (2.2)-(2.3).

Il en est de même pour les solutions de la figure (2.7.b) (resp. (2.7.c)), qui sont constituées d'une 1-détente (resp. 1-choc) entre U_L et U_*, suivie d'une 3-choc (resp. 3-détente) entre U_* et U_R.

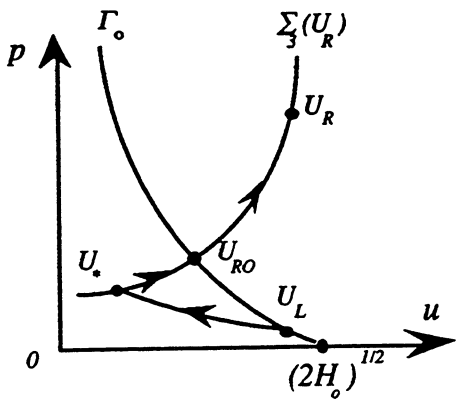
La figure (2.7.d) représente le cas limite où on choisit : $U_L=U_{R0}$. Dans ce cas particulier, $w(x/t; U_L,U_R)$ est constituée d'une 3-détente reliant U_L à U_R. Il s'agit là encore d'une solution "admissible".



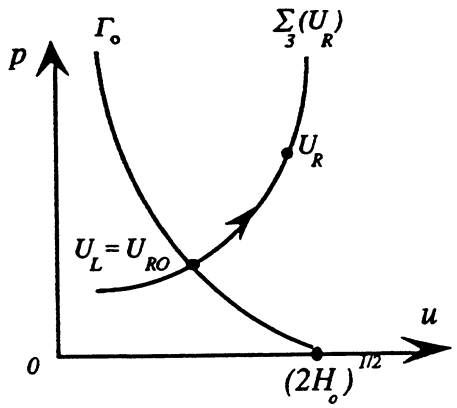
(2.7.a)
1-détente suivie d'une 3-détente



(2.7.b)
1-détente suivie d'un 3-choc



(2.7.c)
1-choc suivi d'une 3-détente



(2.7.d)
Cas limite de 2.7.a - 2.7.c

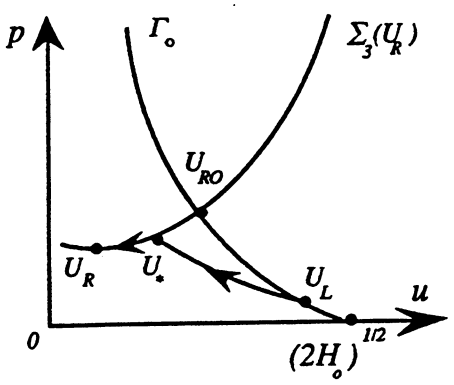
Plaçons nous maintenant dans le cas où l'état U_R est au dessous de la courbe Γ_o. Nous pouvons ici aussi distinguer quatres cas possibles (voir figures (2.8.a)-(2.8.d)), suivant la position de l'état U_L par-rapport à U_{R0}, dans la zone admissible.

Dans le cas (2.8.a), où U_L est au dessus de l'état U_{R0}, $w(x/t; U_L,U_R)$ est constituée d'une 1-choc reliant U_L à U_*, suivie d'une 3-choc reliant U_* à U_R. Cette solution est "admissible" au sens que sa restriction au quadrant ($x\geq 0, t\geq 0$) est solution du problème (2.2)-(2.3).

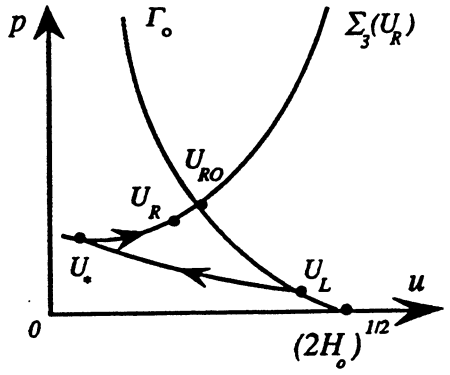
Il en est de même pour les solutions de la figure (2.8.b) (resp. (2.8.c)), qui sont constituées d'une 1-choc (resp. 1-détente) entre U_L et U_*, suivie d'une 3-détente (resp. 3-choc) entre U_* et U_R.

Pour la figure (2.8.d) où on a : ($U_L=U_{R0}$), $w(x/t; U_L,U_R)$ est constituée d'une 3-choc reliant U_L à U_R. Il s'agit là encore d'une solution "admissible".

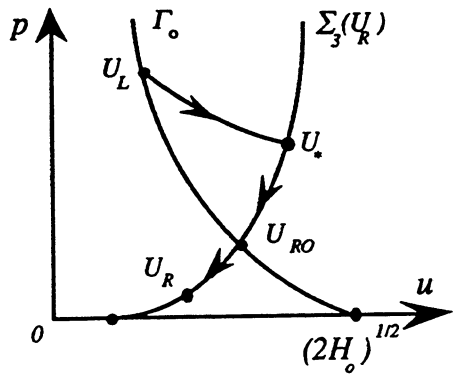
On peut conclure à la suite de cette analyse, que le problème (2.2)-(2.3) admet une infinité de solutions, et ceci quelque soit la position de U_R par-rapport à la courbe Γ_o. Parmi toutes ces solutions, il en existe une seule d'état stationnaire **d'entropie S maximale**. Le but de l'étude qui va suivre est de déterminer cette solution. Cela revient à déterminer sur la courbe Γ_o, l'état U_L d'entropie maximale et pour lequel $w(x/t; U_L, U_R)$ soit admissible pour le problème (2.2)-(2.3).



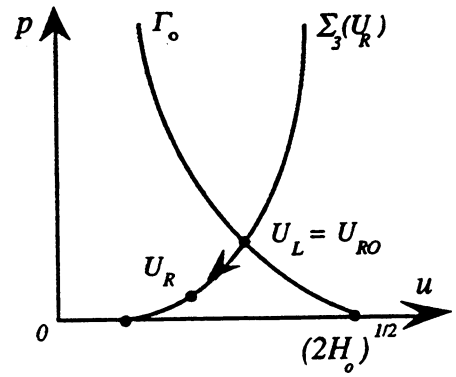
(2.8.a)
1-choc suivi d'un 3-choc



(2.8.b)
1-choc suivi d'une 3-détente



(2.8.c)
1-détente suivie d'un 3-choc



(2.8.d)
Cas limite de 2.8.a - 2.8.c

2.3.4. Sélection de la solution "entropique"

Etudions les variations de S(U) sur la courbe Γ_o. On sait que l'entropie S est proportionnelle à la quantité $\frac{p}{\rho^\gamma}$ (voir Chap.I.§1). Il suffira donc d'étudier les variations de : $\frac{p}{\rho^\gamma}$. Considérons pour cela un état $U=(\rho,u,p)^T$ sur Γ_o, on écrit :

$$\frac{p}{\rho^\gamma} = \frac{\Psi(u)}{\rho^\gamma} \qquad \rho = Q_o/u$$

$$\frac{\Psi(u)}{\rho^\gamma} = \frac{\gamma-1}{\gamma}\left[H_o - \frac{u^2}{2}\right]\frac{u^{\gamma-1}}{(Q_o)^{\gamma-1}}$$

Les variations de S sont celles de : $g(u) = \left[H_o - \frac{u^2}{2}\right]u^{\gamma-1}$. L'étude de la fonction g montre que sa dérivée $g'(u) = \frac{2}{\gamma+1}u^{\gamma-2}\left[\alpha H_o - u^2\right]$, où on a posé : $\alpha = 2(\gamma-1)/(\gamma+1)$, s'annulle pour $(u = \sqrt{\alpha H_o})$ ou $(u=0)$. g' est

positive sur $[0,\sqrt{\alpha H_o}]$ et négative sur $[\sqrt{\alpha H_o},\sqrt{2H_o}]$. L'entropie admet donc un maximum sur $[0,\sqrt{2H_o}]$. Ce maximum correspond à l'état sonique sur la courbe Γ_o.

Remarque 2.1.

On peut considérer un autre critère pour le choix de la solution admissible du problème (2.2)-(2.3), celui de la maximisation du **flux d'entropie** et non pas de l'entropie elle même (apport maximum d'entropie dans la cuve). Le flux d'entropie est obtenu en écrivant le système de la dynamique des gaz compressibles, en coordonnées Euleriennes, avec les variables (ρ,u,S) [Smo]:

$$\begin{cases} \rho_t + (\rho u)_x = 0 \\ u_t + u u_x + p_x/\rho = 0 \\ s_t + u s_x = 0 \end{cases} \qquad (2.15)$$

s désigne l'entropie spécifique du gaz : S = ρs. En multipliant la première et la troisième équation de (2.15) respectivement par s et par ρ, et en les sommant on obtient l'équation conservative suivante :

$$S_t + (uS)_x = 0 \qquad (2.16)$$

Pour tout état U, le flux d'entropie est donc donné par : Π(U) = uS. Etudions les variations de Π(U) sur la courbe Γ_o. On trouve cette fois que les variations de Π(U) sont celles de : $g(u) = \frac{up}{\rho^\gamma} = [H_o - \frac{u^2}{2}]u^\gamma$. L'étude de la fonction g montre que sa dérivée $g'(u) = u^{\gamma-1}[\gamma H_o - \frac{(\gamma+1)}{2}u^2]$, s'annulle pour (u = 0) ou ($u = \sqrt{\beta H_o}$), où β est donné par : $\beta = \frac{2\gamma}{(\gamma+1)}$. g' est positive sur $[0,\sqrt{\beta H_o}]$ et négative sur $[\sqrt{\beta H_o},\sqrt{2H_o}]$. Le flux d'entropie admet donc un maximum sur $[0,\sqrt{2H_o}]$. Ce maximum se trouve en plus sur la zone adimissible définie par la relation (2.11), car β vérifie : $\alpha \le \beta \le 2$.

2.3.5. Application numérique

Dans ce paragraphe, on s'intéresse à ce qui ce passe en pratique, et en particulier au cas correspondant aux valeurs numériques du débit et de l'enthalpie qui sont imposés dans notre cas test (Q_o = 450.00 kg.s^{-1}.m^2, H_o = 450.00 kJ/kg).

A l'instant initial, le gaz à l'intérieur de la cuve est au repos, sa pression est prise égale à la pression atmosphérique et sa densité égale à celle de l'air,

$$U_R = (\rho_R, u_R, p_R)^T = (1.29\ \text{kg.m}^{-3}, 0.0\ \text{ms}^{-1}, 1.0\ \text{atm})^T$$

Le calcul montre que la condition (2.4) du théorème 2.1. est vérifiée puisque (u_R = 0) :

$$u_R - \frac{2c_R}{\gamma-1} < 0 \le A\sqrt{2H_o}$$

D'autre part, pour violer (2.4), il faut que la vitesse du gaz à l'intérieur de la cuve dépasse la valeur

$$u_R \ge u_{min} = \sqrt{2H_o} + 2c_R/(\gamma-1)$$

avec : $\sqrt{2H_o}$ = 948.8 $m.s^{-1}$, ce qui équivaut à approximativement trois fois la célérité de l'onde sonore. le second terme vaut, pour la valeur la plus petite de c_R (c=340 $m.s^{-1}$) : $2c_R/(\gamma-1) \equiv 1700.\ m.s^{-1}$. Ce qui nous donne une valeur minimale de la vitesse de l'ordre de : $u_{min} \equiv 2600.\ m.s^{-1}$.

Cette valeur n'est évidemment jamais atteinte dans notre cas de figure, le problème (2.2)-(2.3), de la condition à la limite amont (brèche d'entrée), admet toujours une solution "entropique" unique.

2.4. Condition à la limite aval : brèche de sortie

Rappelons que la condition à la limite avale (brèche de sortie), est une condition de type pression imposée ($p=p_a$=1 atm).

2.4.1. Le demi-problème de Riemann

Nous considérons également un demi-problème de Riemann comme problème modèle

$$\begin{cases} U_t + f(U)_x = 0 \\ U(x,0) = U_L \\ \varphi(U(0,t)) = 0 \end{cases} \qquad (x,t) \in R_- \times R_+ \qquad (2.17)$$

La fonction φ du problème (2.17), est ici donnée par :

$$\varphi(U) = p - p_a \qquad (2.18)$$

Comme dans la partie 2.3., on connait l'état à l'intérieur de la cuve, c'est l'état gauche dans le demi-problème de Riemann (2.17), et on cherche un état U_R tel que la restriction de la solution du problème de Riemann $PR(U_L,U_R)$ au quadrant ($x\leq 0, t\geq 0$) soit solution de (2.17)-(2.18).

On va démontrer le théorème suivant

Théorème 2.2.

Supposons que l'état U_L soit subsonique($u_L - c_L \leq 0$) et de pression toujours supérieur ou égale à la pression atmosphérique p_a. Si la vitesse et la pression de U_L vérifie la relation suivante

$$u_L - c_L + \frac{\gamma+1}{\gamma-1} c_L \left[1 - \left(\frac{p_L}{p_a}\right)^{(1-\gamma)/2\gamma}\right] \geq 0 \qquad (2.19)$$

Le problème (2.17)-(2.18) n'admet pas de solution, on ne peut alors pas imposer la pression $p=p_a$ en sortie. Sinon le problème (2.17)-(2.18) admet une infinité de solutions dans la classe des fonctions C^1 par morceaux de x/t et continue en $x=0$, définies pour $x/t\leq 0$. Parmi ces solutions, celle correspondant à l'état sonique est admissible pour (2.17)-(2.18), et la solution associée est composée dans le quadrant ($x/t\leq 0$), d'une seule onde.

Remarque 2.2.

Pour la simulation numérique, on a choisit comme état U_R l'état sonique, donné par :

$$u_R = u_a \tag{2.20.a}$$

$$p_R = p_a \tag{2.20.b}$$

$$\rho_R = (\gamma p_a) / u_a^2 \tag{2.20.c}$$

2.4.2. Démonstration du théorème 2.2.

Posons pour toute la suite : $U_a = (u_a,p_a)^T = \Sigma_1(U_L) \cap \{p=p_a\}$, intersection dans le plan (u,p), de la 1-courbe passant par U_L et de la droite $\{p=p_a\}$ (voir figure 2.9).

Soit U_R un état quelconque choisi de façon à ce que la 3-courbe : $\Sigma_3(U_R)$ passe par U_a. La solution de $PR(U_L,U_R)$ est constituée d'une 1-onde reliant U_L à un état $U_{a,1}$ donné par $U_{a,1} = (\rho_{a,1}, u_a, p_a)^T$, suivie d'une discontinuité de contact entre $U_{a,1}$ et un état $U_{a,2}$ donné par $U_{a,2} = (\rho_{a,2}, u_a, p_a)^T$, et d'une 3-onde reliant $U_{a,2}$ à U_R. Analysons cette solution.

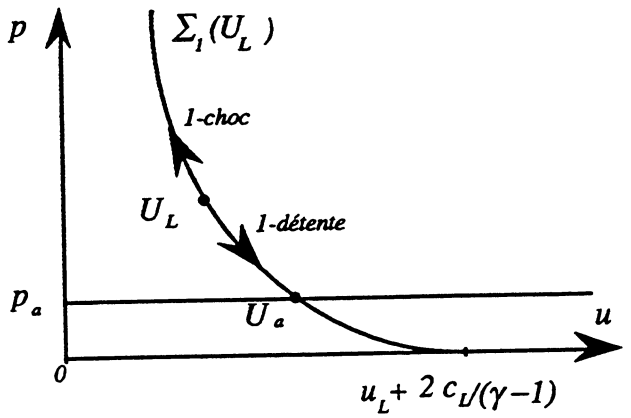


Figure 2.9.

Comme on a supposé que $p_L \geq p_a$, alors $w(x/t;U_L,U_R)$ est constituée d'une 1-détente reliant U_L à $U_{a,1}$ suivie d'une discontinuité de contact entre $U_{a,1}$ et un état $U_{a,2}$, et d'une 3-onde reliant $U_{a,2}$ à U_R. On a alors les relations suivantes

$$p_L/p_{a,1} = (\rho_L/\rho_{a,1})^\gamma \tag{2.21}$$

$$u_L + 2c_L/(\gamma-1) = u_a + 2c_{a,1}/(\gamma-1) \tag{2.22}$$

qui traduisent que les 1-invariants de Riemann sont constants le long de la 1-détente reliant U_L à U_a.

Il suffit pour que l'état stationnaire $w(0;U_L,U_R)$ vérifie $\varphi(w(0;U_L,U_R))=0$, que : $(u_a-c_{a,1}\leq 0)$ et que $(u_a \geq 0)$, puisque pour tout (x,t) tel que : $u_a-c_{a,1} \leq x/t \leq u_a$, on a : $w(x/t;U_L,U_R)=U_{a,1}$ et $p(U_{a,1})=p_a$. D'après les relations (2.20) et (2.21) la condition $(u_a-c_{a,1}\leq 0)$ est équivalente à

$$u_a-c_{a,1} = u_L - c_L + \frac{\gamma+1}{\gamma-1} c_L \left[1 - \left(\frac{p_L}{p_a}\right)^{(1-\gamma)/2\gamma}\right] \leq 0$$

On en déduit que si (2.19) n'est pas vérifiée alors le problème (2.17-2.18) n'admet pas de solution, autrement dit la pression ne peut être imposée en sortie à p_a pour la simple raison que : la pression de l'état stationnaire du problème de Riemann $PR(U_L,U_R)$ sera comprise (éventuellement strictement) entre p_L et p_a, sauf si l'état U_a est sonique.

Si par contre (2.19) est vérifiée alors on a $u_a-c_{a,1} \leq 0$ et par la suite $\varphi(w(0;U_L,U_R))=0$. La solution par la figure (2.10) dans le plan (x,t)

De cette remarque et des relations (2.21)-(2.22), on déduit que (2.17)-(2.18) admet dans ce cas, une infinité de solutions correspondants aux états U_R dont la 3-courbe $\Sigma_3(U_R)$ passe par U_a.

On choisit alors comme état U_R, l'état suivant : $U_R = U_a$, et on choisit la densite de U_R de façon à ce que cet état soit sonique. On obtient alors :

$$U_R = (\rho_R, u_R, p_R)^T = \left(\frac{\gamma p_a}{u_a^2}, u_R, p_R\right)^T$$

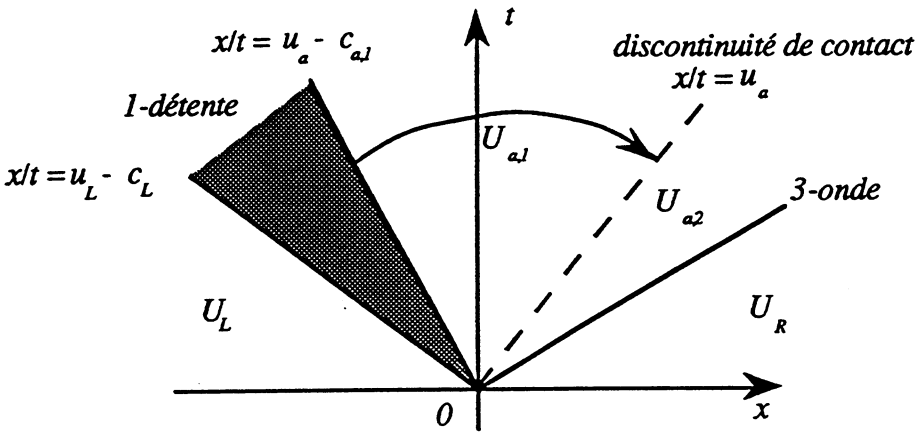


Figure 2.10.

2.5. Résultats numériques

Les pas d'espace utilisés sont les suivants :

dx = 0.01 ou 0.02 m

dy = 0.01 ou 0.02 m

Le maillage comporte ainsi 1900 éléments (voir figure 2.11).

On s'est intéressé à deux types de résultats:

- Les courbes de "pressions instantanées", correspondants à la distribution de pression, à un instant donné, dans tout le domaine d'étude.
- La résultante des efforts de pression sur les parois de la cuve centrale.

Le schéma numérique, le pas de temps et la CFL utilisés pour cette simulation sont les mêmes que dans le cas du Gazex (Cf. Chap.I.§.1). La premières famille de courbes d'isovaleurs, obtenue à l'aide d'un schéma du second ordre en espace et en temps (schéma de Van Leer), représente les

distributions de pressions instantannées dans le puit tous les 0.75 ms. Les valeurs de surpression qui aparaissent dans les légendes de ces différentes courbes sont exprimées en millibars.

Les trois dernières courbes représentent les variations dans le temps des efforts de pression sur la cuve centrale. Ces trois courbes sont obtenues, dans l'ordre, pour la première par le schéma aux volumes finis du premier ordre en espace et en temps, décrit au §.2.4.1. du chapitre III, il s'agit du schéma de Godounov. La deuxième courbe est obtenue à l'aide d'un schéma du second ordre en espace et en temps (cf. §.2.4.2. du chapitre III) qui généralise le schéma de Van Leer [Van] aux volumes finis.

En ce qui concerne la troisième et dernière courbe, elle correspond au résultat d'un schéma aux volumes finis semi-implicit. Ce résultat a été obtenu par Herard de l'équipe LNH-EDF* dans [Her]. Elle est présentée à titre comparatif. On remarque que ce schéma semi-implicit et le schéma de Godounov amortissent entièrement les oscillations liées aux ondes soniques qui se propagent dans la cuve, tandis que le schéma de Van Leer les met bien en évidence.

---

* Laboratoire National d'Hydraulique (EDF), Groupe recherches, Chatou.

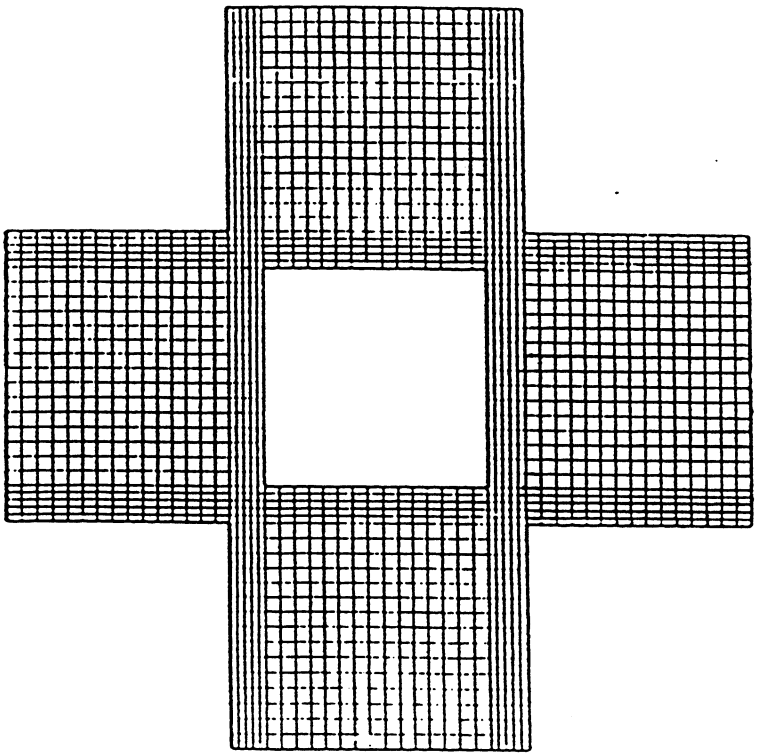
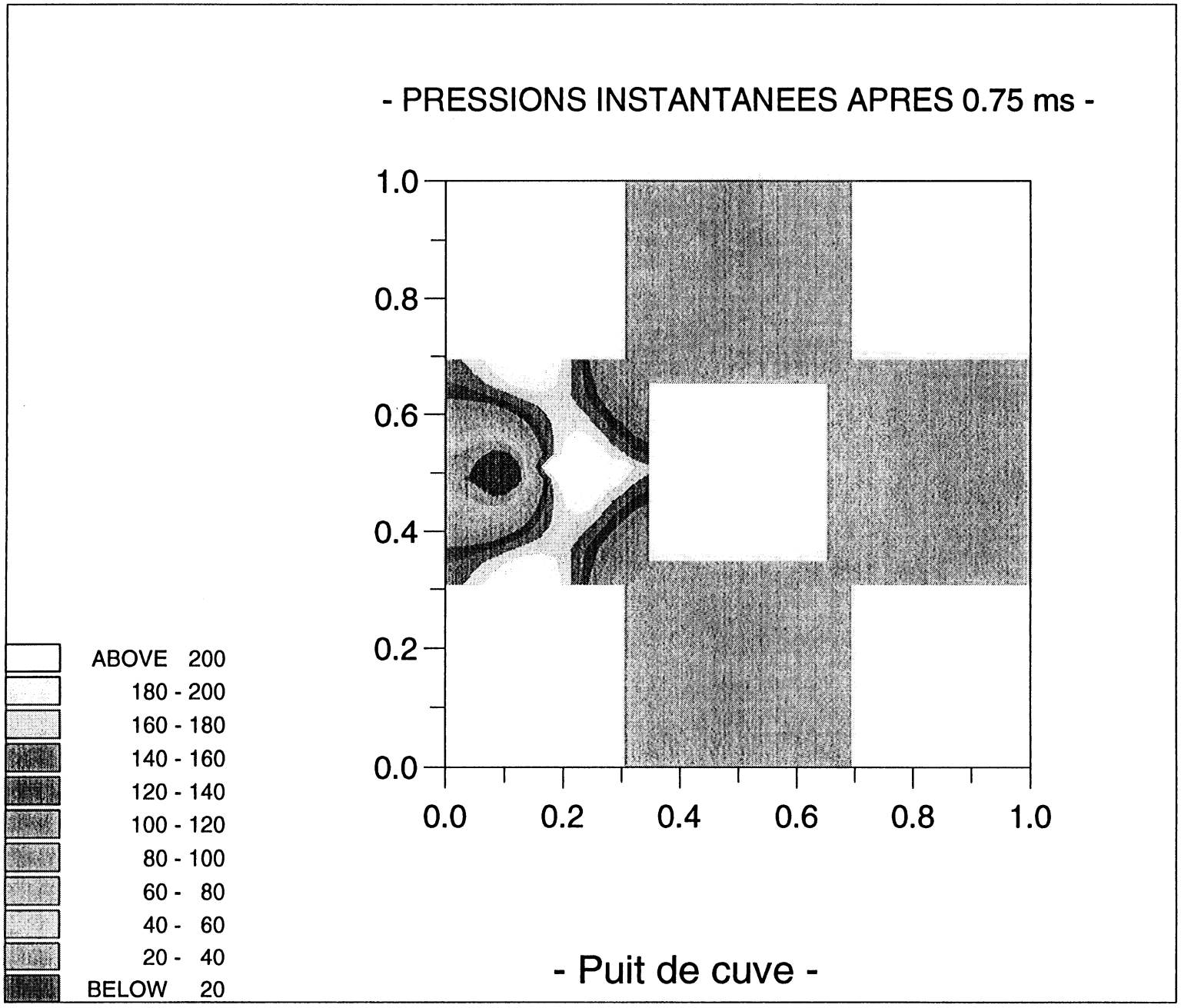
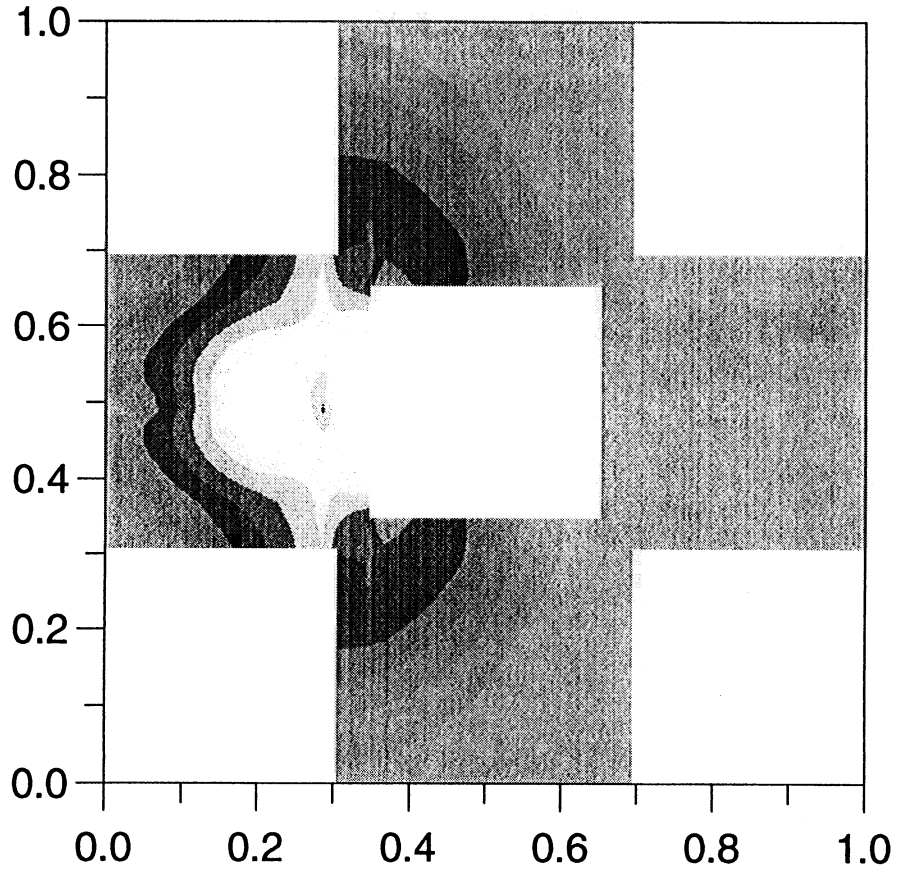


Figure 2.11. : Maillage du puit de cuve

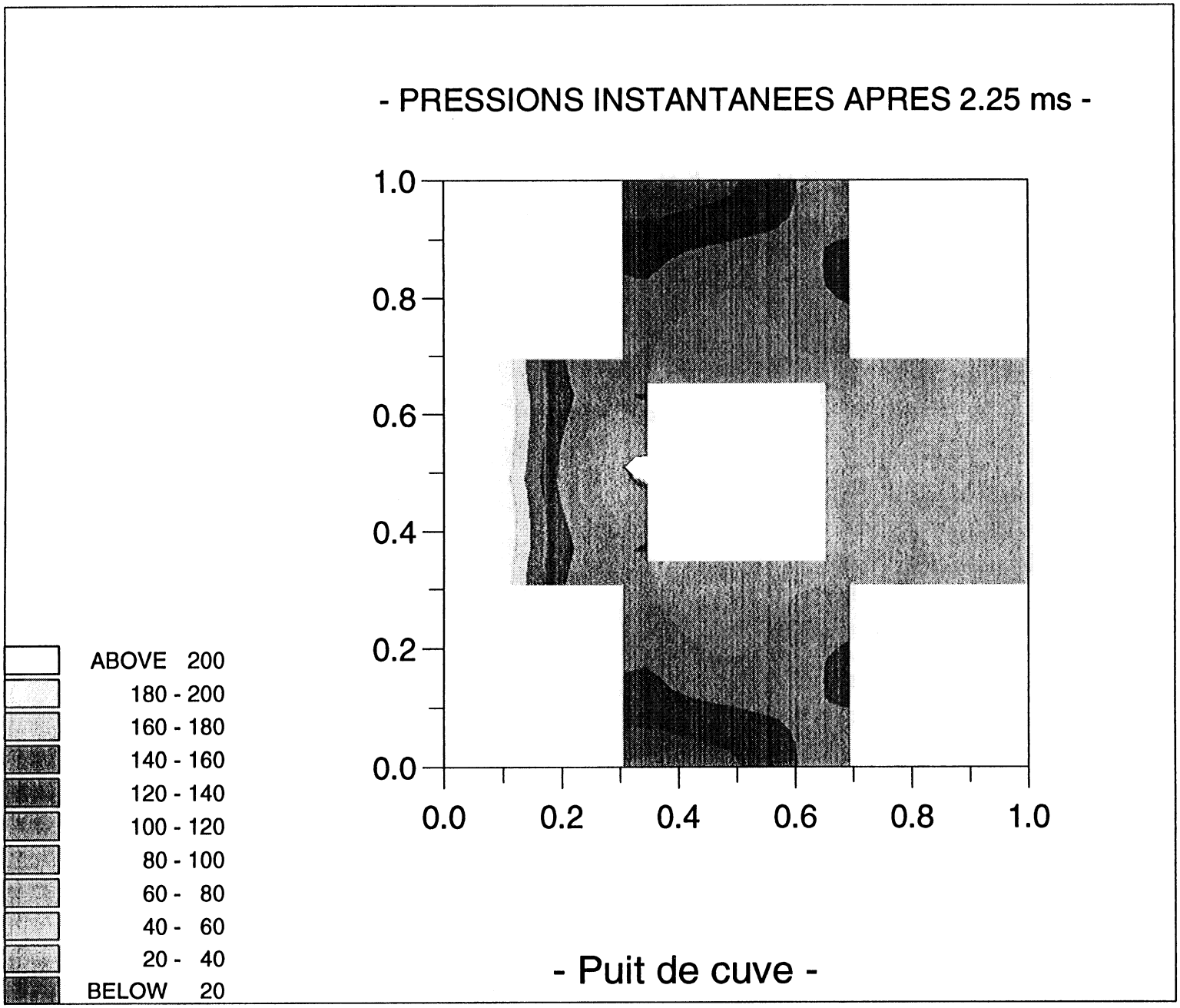

- PRESSIONS INSTANTANEES APRES 0.75 ms -
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
ABOVE 200
180 - 200
160 - 180
140 - 160
120 - 140
100 - 120
80 - 100
60 - 80
40 - 60
20 - 40
BELOW 20
- Puit de cuve -

- PRESSIONS INSTANTANEES APRES 1.5 ms -

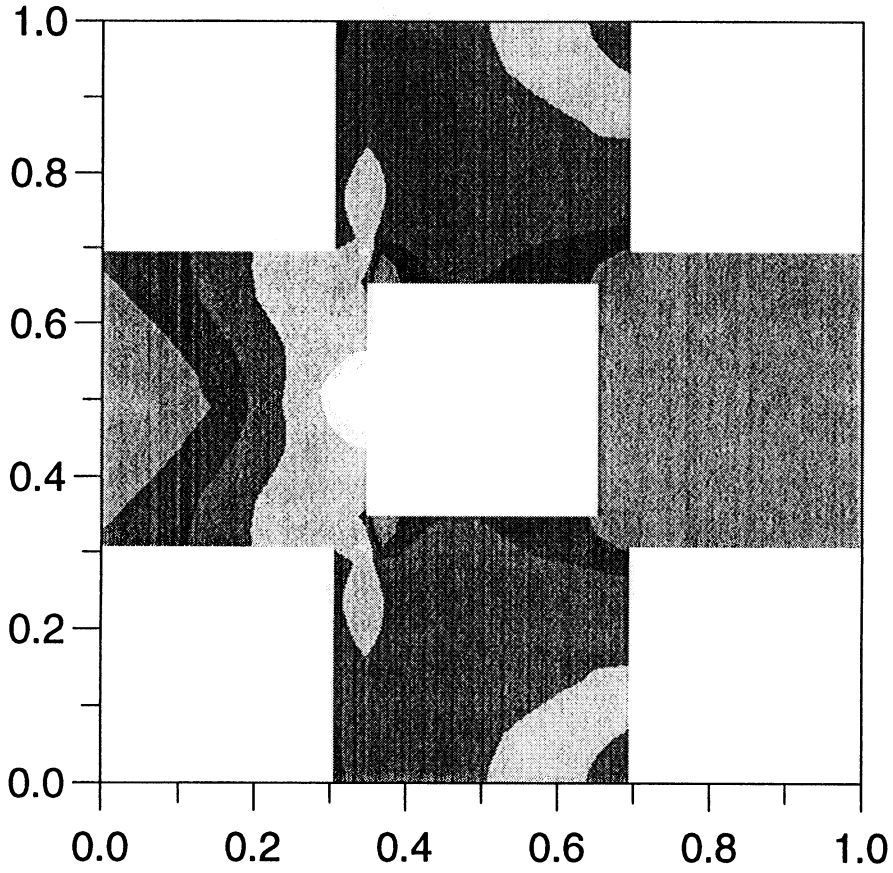


ABOVE 200
180 - 200
160 - 180
140 - 160
120 - 140
100 - 120
80 - 100
60 - 80
40 - 60
20 - 40
BELOW 20

- Puit de cuve -

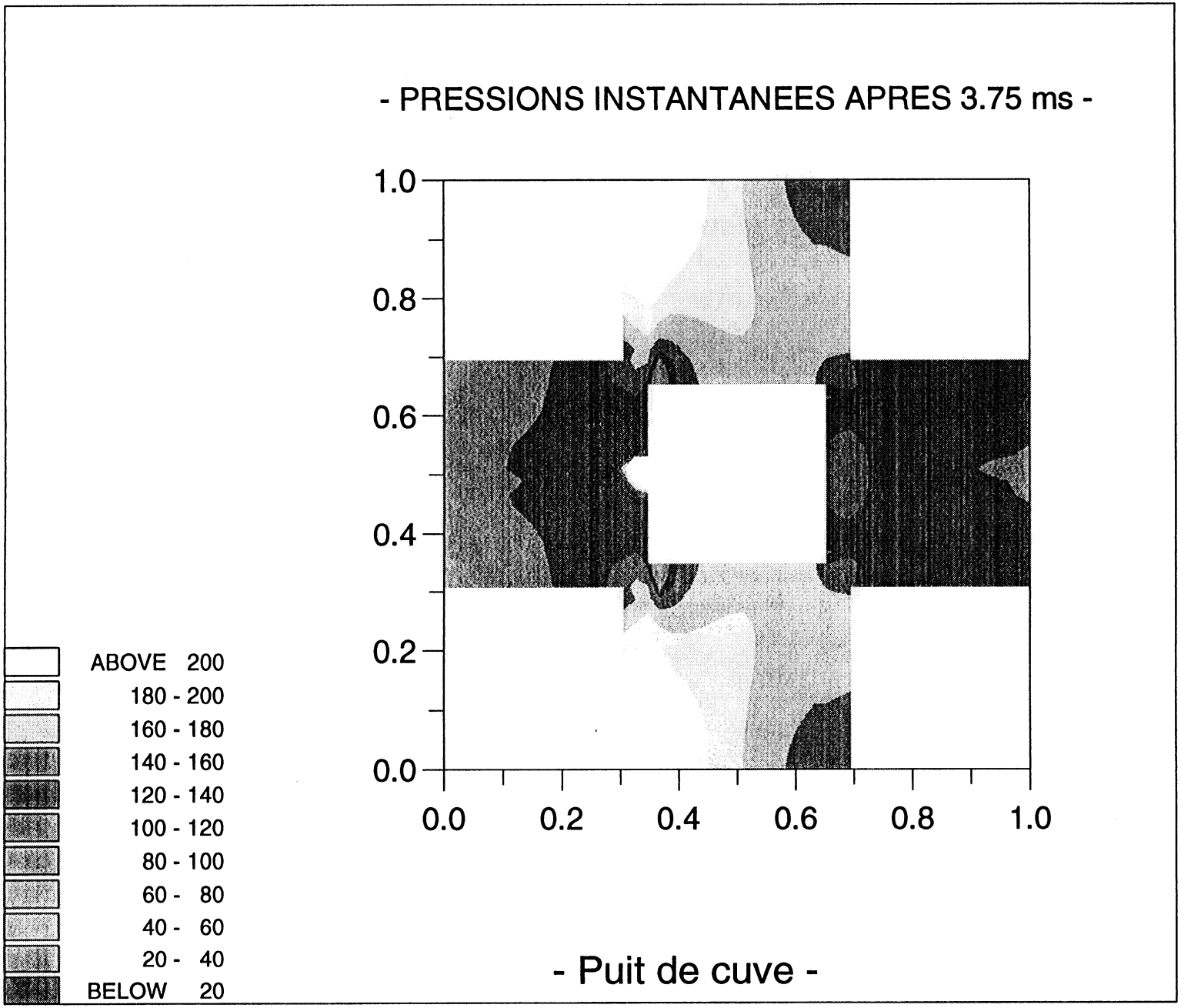

- PRESSIONS INSTANTANEES APRES 2.25 ms -
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
ABOVE 200
180 - 200
160 - 180
140 - 160
120 - 140
100 - 120
80 - 100
60 - 80
40 - 60
20 - 40
BELOW 20
- Puit de cuve -

- PRESSIONS INSTANTANEES APRES 3.00 ms -

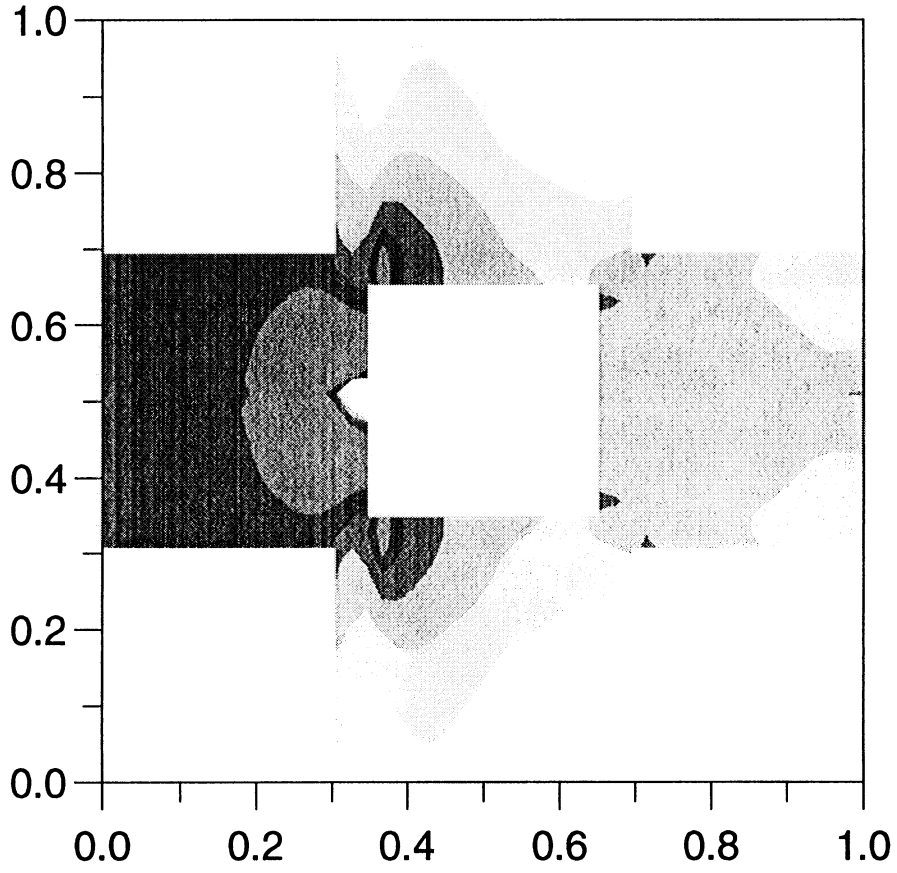


ABOVE 200
180 - 200
160 - 180
140 - 160
120 - 140
100 - 120
80 - 100
60 - 80
40 - 60
20 - 40
BELOW 20

- Puit de cuve -

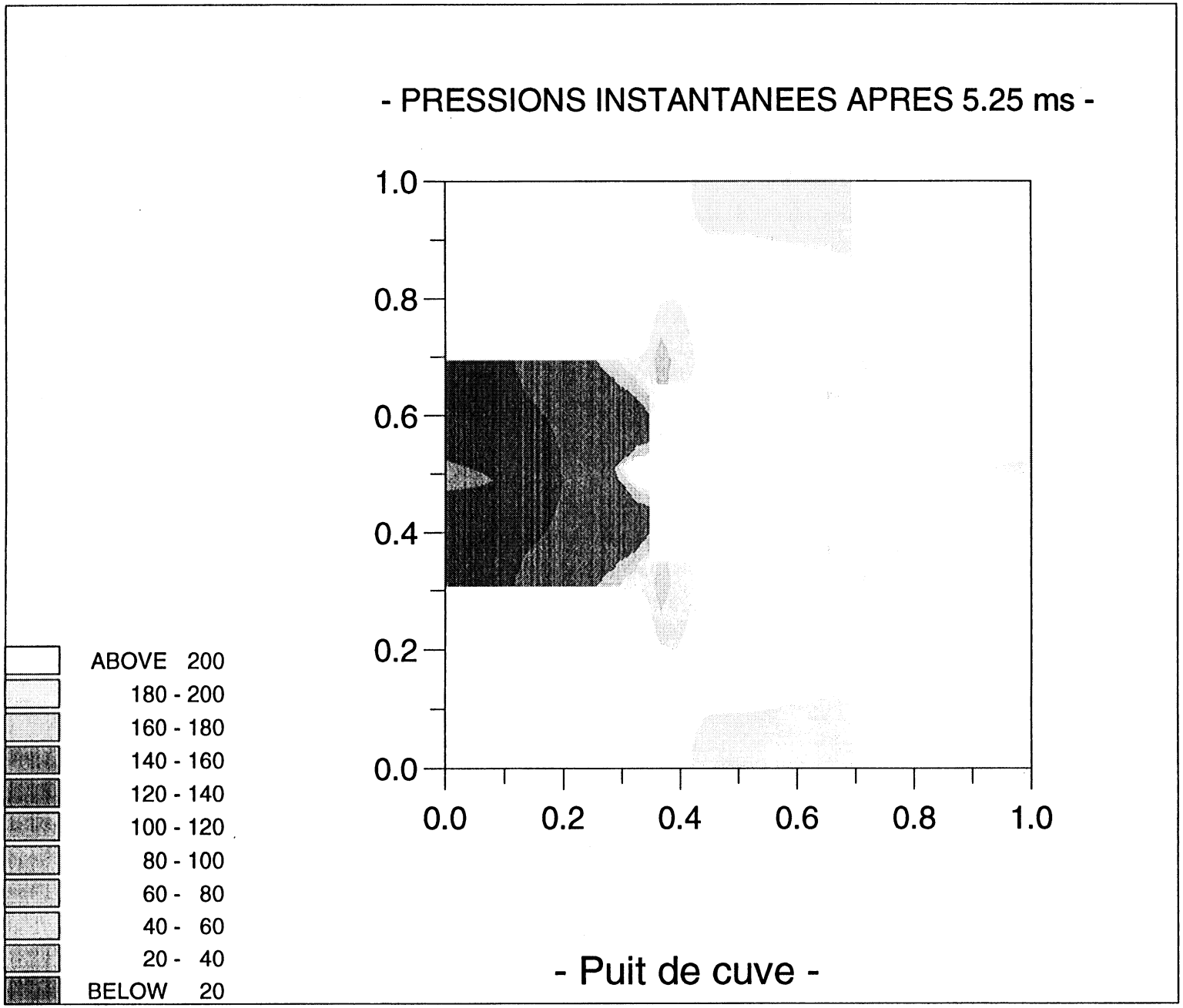

- PRESSIONS INSTANTANEES APRES 3.75 ms -
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
ABOVE 200
180 - 200
160 - 180
140 - 160
120 - 140
100 - 120
80 - 100
60 - 80
40 - 60
20 - 40
BELOW 20
- Puit de cuve -

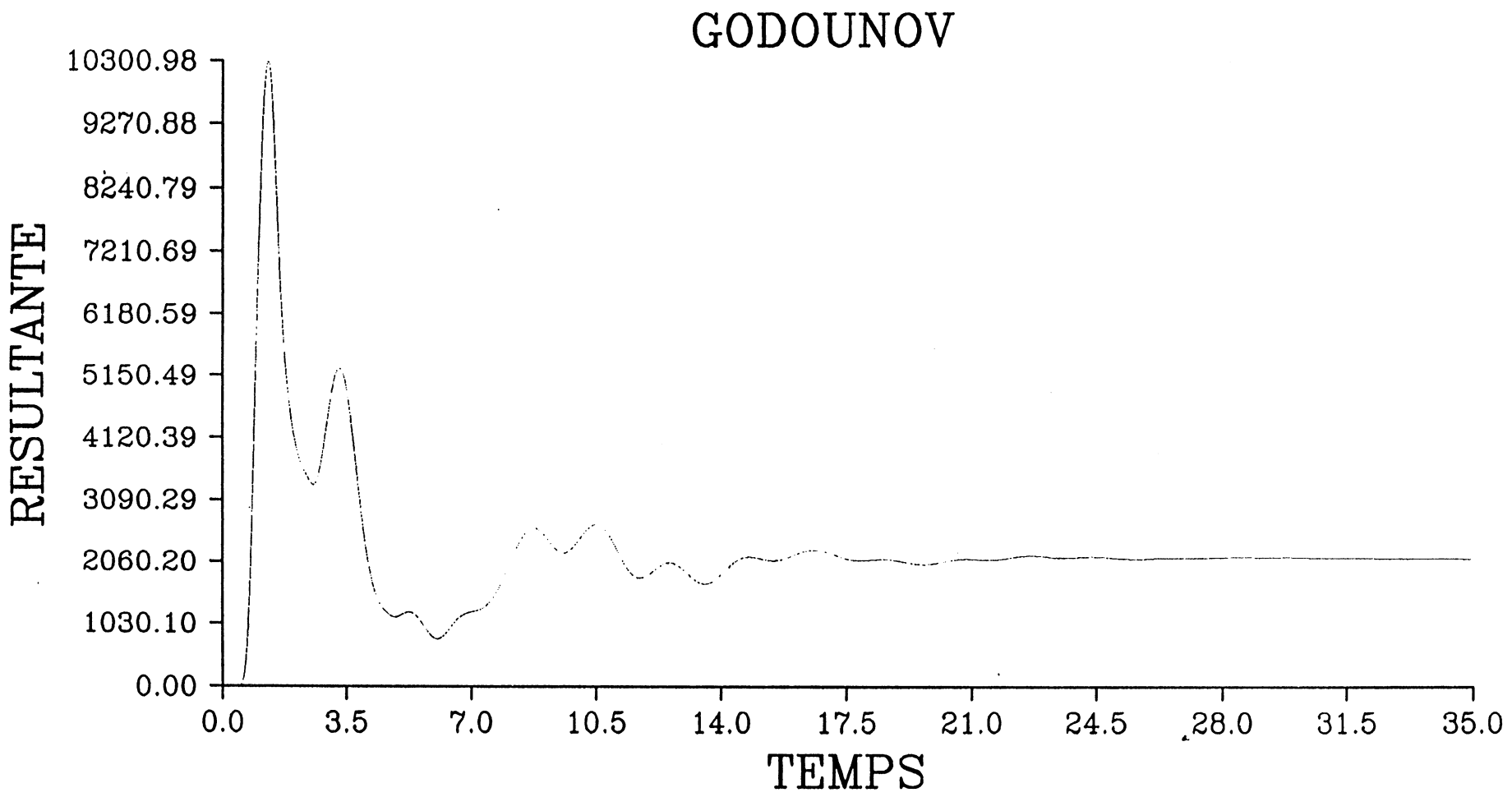
- PRESSIONS INSTANTANEES APRES 4.5 ms -

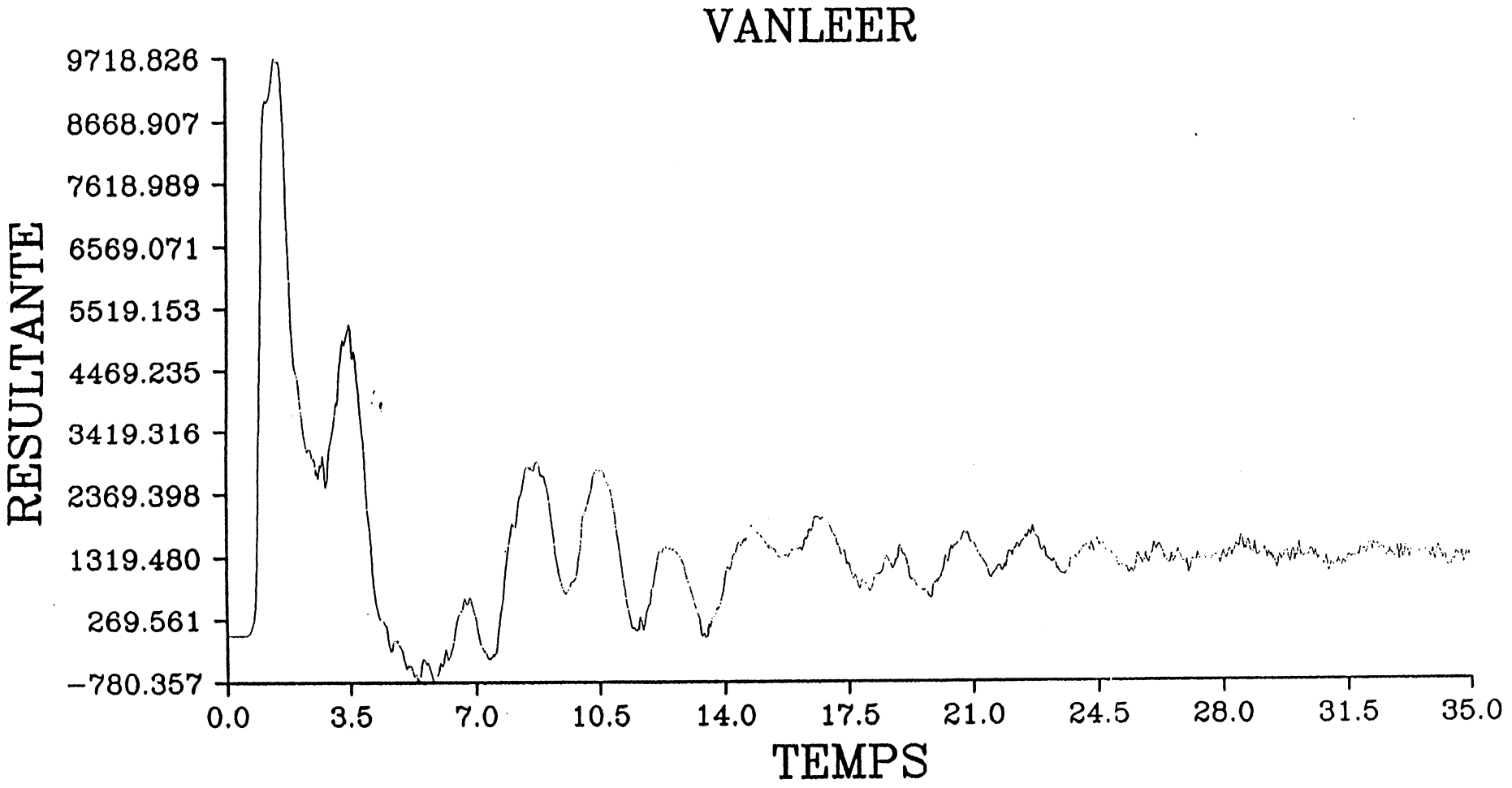


ABOVE 200
180 - 200
160 - 180
140 - 160
120 - 140
100 - 120
80 - 100
60 - 80
40 - 60
20 - 40
BELOW 20

- Puit de cuve -


- PRESSIONS INSTANTANEES APRES 5.25 ms -
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
ABOVE 200
180 - 200
160 - 180
140 - 160
120 - 140
100 - 120
80 - 100
60 - 80
40 - 60
20 - 40
BELOW 20
- Puit de cuve -


GODOUNOV
RESULTANTE
10300.98
9270.88
8240.79
7210.69
6180.59
5150.49
4120.39
3090.29
2060.20
1030.10
0.00
0.0
3.5
7.0
10.5
14.0
17.5
21.0
24.5
28.0
31.5
35.0
TEMPS


VANLEER
RESULTANTE
9718.826
8668.907
7618.989
6569.071
5519.153
4469.235
3419.316
2369.398
1319.480
269.561
−780.357
0.0
3.5
7.0
10.5
14.0
17.5
21.0
24.5
28.0
31.5
35.0
TEMPS

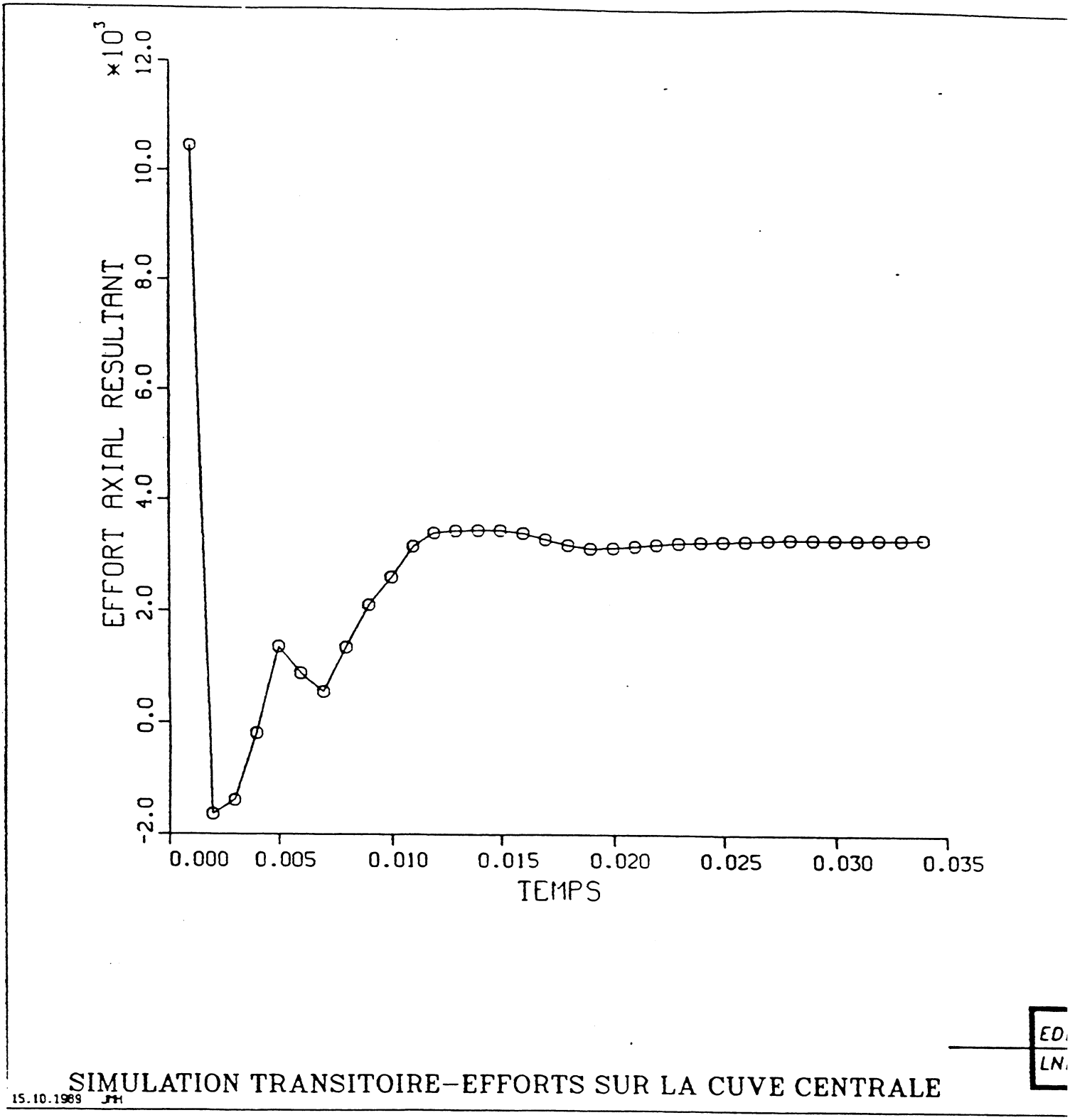

EFFORT AXIAL RESULTANT
*10^3
12.0
10.0
8.0
6.0
4.0
2.0
0.0
-2.0
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.035
TEMPS
SIMULATION TRANSITOIRE-EFFORTS SUR LA CUVE CENTRALE
15.10.1989

Figure 5 :

CHAPITRE III

SCHEMAS AUX VOLUMES FINIS

GENERALITES

APPLICATION AUX EQUATIONS D'EULER POUR LA DYNAMIQUE DES GAZ COMPRESSIBLES

0. Introduction

Les principales applications physiques abordées dans ce travail sont, comme on l'a vu précedemment (Cf. Première Partie) modélisées par les équations de la dynamique des gaz compressibles. Il s'agit d'un système de lois de conservations hyperbolique non linéaire (S.H.N.L).

Nous avons utilisé essentiellement la méthode des volumes finis pour discrétiser ce système, d'une part parce que cette méthode présente une grande cohérence avec la physique, surtout dans le cas des systèmes qui sont sous forme conservative, ce qui est le cas pour les équations d'Euler. D'autre part, les schémas type volumes finis sont de par leur définition, conservatifs, ce qui leur garantie une certaine robustesse, qualité essentielle lorsque dans la solution des équations discrétisées peuvent apparaître des discontinuités, ce qui est là encore le cas des S.H.N.L.

Rappelons brièvement que la méthode des volumes finis (une description plus détaillée sera fournie plus loin), consiste à faire un bilan sur des volumes de contrôle (ce sont dans la plus part des cas les cellules du maillage). Ces bilans nécessitent un calcul de flux à travers les faces des volumes de contrôle. La technique adoptée ici pour le calcul de ces flux est la résolution de problèmes de Riemann locaux à l'aide d'un solveur approché: solveur de Roe qui consiste à linéariser la Jacobienne du flux du S.H.N.L., et donc à trouver une matrice, appelée matrice de Roe, qui approche cette Jacobienne (Cf. Chap.III.§.5).

Le système de la dynamique des gaz parfaits polytropiques, de par la forme explicite de la loi d'état,

$$P = P(\rho, I) \tag{1.1}$$

se prête facilement aux calculs. Il est en effet relativement aisé de déterminer une matrice de Roe dans ce cas. Par contre, il est plus compliqué de trouver une "bonne" linéarisation (au sens de Roe), pour la dynamique des gaz réels, caractérisés par une loi d'état quelconque. Cette difficulté est essentiellement due à la perte du caractère hyperbolique de la matrice de Roe déterminée dans le cas des gaz parfaits.

1. Position du problème - Notations

Nous cherchons à résoudre le problème de Cauchy suivant:

$$\begin{cases} \partial_t U + \partial_x f(U) = 0 \\ U(0,x,y) = U_o(x,y) \end{cases} \qquad (x,t) \in \mathbb{R}^n \times \mathbb{R}^+, \quad U \in \mathbb{R}^n \tag{1.2}$$

où : $f = (f_1,\ldots,f_n)$ est le flux numérique hyperbolique du système, i.e.: pour tout ω dans $\mathbb{R}^n$ tel que: $\Sigma\omega_i=1$, la fonction f_ω définie par:

$$f_\omega = \Sigma\, \omega_i f_i$$

est hyperbolique. Dans toute la suite, et par simple souci de simplification de la présentation, seul le cas n=2 sera traité, la généralisation à n quelconque se fait sans difficulté.

Soit T_h une polyhédrisation quelconque de $\mathbb{R}^2$: $\mathbb{R}^2 = \bigcup_{K \in T_h} K$. Introduisons quelques notations relatives au maillage.

1.1. notations

Soit K un polyhèdre de T_h, e une arête de K. On note :

- $|e|$: la longueure de l'arête e.
- $n_{e,K}$: la normale à e sortante de K.
- K_e : l'élèment de T_h tel que : $K \cap K_e = e$
- $|K|$: la surface de K.
- $P_K = \sum_{e \in \partial K} |e|$: le périmètre de K, ∂K désigne le contour de K.
- h_K = diamètre de K.
- $h = \sup_{K \in T_h} (h_K)$.
- τ : le pas de temps.

On désigne par $U_h(x,t)$, l'approximation numérique du problème (1.2), et pour chaque élément K de T_h, on pose:

$$U_K^n = U_h(x_K, \tau.n)$$

où x_K désigne les coordonnés du barycentre de l'élément K.

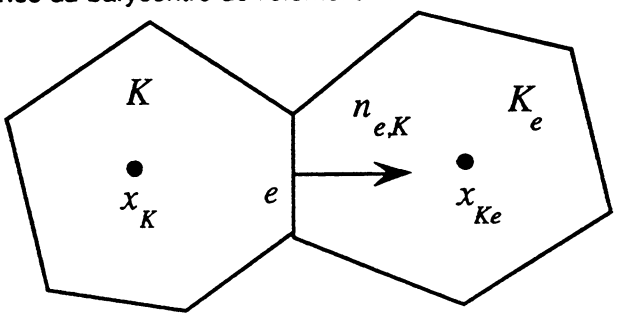


Figure 1.3. : Maillage, méthode des volumes finis

On définit pour chaque élément K de T_h, et pour chaque arête e de K, un flux numérique, $g_{K,e}$: $\mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$, auquel on demandera de vérifier, dans la plus part des cas, des propriétés de conservativité et de consistance:

1) conservativité: $g_{K,e}(u,v) = -g_{K_e,e}(v,u)$

2) consistance avec: $f(u).n_{e,K}$ où $f(u) = (f_1(u);f_2(u))^T$

i.e. $g_{K,e}(u,u) = f(u).n_{e,K}$

1.2. Schéma numérique

Le calcul des valeurs approchées U_K^{n+1} de U_E à l'instant t_{n+1}, à partir de celles obtenues à t_n: U_K^n, se fait de la façon suivante, si on note:

$$U_{e,K}^n = \lim_{\substack{x \to x_e \\ x \in K}} U_h(x,\tau.n) = U_h(x_e^-,\tau.n)$$

$$U_{e,K_e}^n = \lim_{\substack{x \to x_e \\ x \in K_e}} U_h(x,\tau.n) = U_h(x_e^+,\tau.n)$$

où x_e désigne les coordonnées du centre de l'arête e (voir Figure 1.4.). Autrement dit, $U_{e,K}^n$ (resp. U_{e,K_e}^n) représente la valeur de la restriction de U_h à K (resp. K_e) au centre x_e de l'arête e.

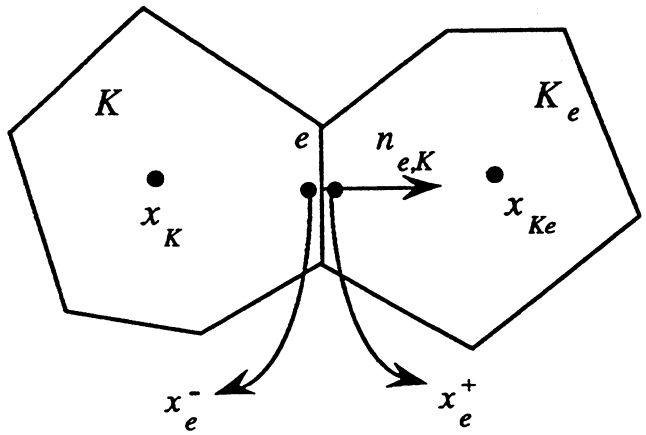


Figure 1.4.

Au vu de ces notations, le schéma aux volumes finis explicite, basé sur la flux numérique $g_{K,e}$, s'écrit:

$$U_K^{n+1} = U_K^n - \frac{\tau}{|K|} \sum_{e \in \partial K} |e|\, g_{K,e}(U_{e,K}^n, U_{e,K_e}^n) \qquad (1.3)$$

Dans le cas particulier où l'approximation U_h est constante par maille (schéma d'ordre un en espace), on a:

$$U_{e,K}^n = U_K^n \qquad \text{et} \qquad U_{e,K_e}^n = U_{K_e}^n$$

et le schéma (1.3) s'écrit:

$$u_K^{n+1} = u_K^n - \frac{\tau}{|K|} \sum_{e \in \partial K} |e|\, g_{K,e}(u_K^n, u_{K_e}^n) \qquad (1.4)$$

2. Méthode des volumes finis pour les équations d'Euler

2.1. Les équations

Rappelons que les équations d'Euler pour la dynamique des gaz compressibles s'écrivent

$$\begin{cases} \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) + \frac{\partial}{\partial y}(\rho v) = 0 \\ \frac{\partial}{\partial t}(\rho u) + \frac{\partial}{\partial x}(\rho u^2+p) + \frac{\partial}{\partial y}(\rho uv) = 0 \\ \frac{\partial}{\partial t}(\rho v) + \frac{\partial}{\partial x}(\rho uv) + \frac{\partial}{\partial y}(\rho v^2+p) = 0 \\ \frac{\partial E}{\partial t} + \frac{\partial}{\partial x}[u(E+p)] + \frac{\partial}{\partial y}[v(E+p)] = 0 \end{cases} \qquad (2.1)$$

où E désigne l'énergie totale du gaz, donnée en fonction de l'énergie interne spécifique e et de l'énergie cinétique par

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

La forme conservative des équations d'Euler est donnée par :

$$\frac{\partial \underset{\sim}{U}}{\partial t} + \underset{\sim}{\text{div}}\,\underset{\approx}{F}(\underset{\sim}{U}) = 0 \qquad (2.2)$$

$$\underset{\sim}{U} = (\rho, \rho u, \rho v, E)^T \qquad (2.3.a)$$

$$\underset{\approx}{F}(\underset{\sim}{U}) = (\underset{\sim}{F}_1(\underset{\sim}{U}), \underset{\sim}{F}_2(\underset{\sim}{U}))^T \qquad (2.3.b)$$

$$\underset{\sim}{F}_1(\underset{\sim}{U}) = (\rho u, \rho u^2+p, \rho uv, u(E+p))^T \qquad (2.3.c)$$

$$\underset{\sim}{F}_2(\underset{\sim}{U}) = (\rho v, \rho uv, \rho v^2+p, v(E+p)))^T \qquad (2.3.d)$$

ρ est la densité du gaz, $\underset{\sim}{u} = (u,v)^T$ est le vecteur vitesse.

2.2. Propriétés d'invariance par rotation et Hyperbolicité

Ce système possède en outre des propriétés remarquables de symétrie et d'invariance par rotation. Définissons la transformation $R(n)$ sur $\underset{\sim}{U}$ associée à la rotation:

$$(x,y) \rightarrow (\bar{x},\bar{y})$$

$$\underset{\sim}{\bar{U}} = (h, \rho\bar{u}, \rho\bar{v})^T$$

$$\underset{\sim}{\bar{U}} \equiv R(n).\underset{\sim}{U} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & n_x & n_y \\ 0 & -n_y & n_x \end{pmatrix}.\underset{\sim}{U} \qquad (2.4)$$

$\underset{\approx}{F}$ possède la propriété suivante :

$$\forall\ \underset{\sim}{\mu} \in R^2; \qquad \underset{\approx}{F}(\underset{\sim}{U}).\underset{\sim}{\mu} = |\mu|\ .R(\underset{\sim}{\mu})^{-1}.\underset{\sim}{F}_1(\overline{\underset{\sim}{U}}) = |\mu|\ .R(\underset{\sim}{\mu})^{-1}.\underset{\sim}{F}_1(R(\underset{\sim}{\mu}).\underset{\sim}{U}) \tag{2.5}$$

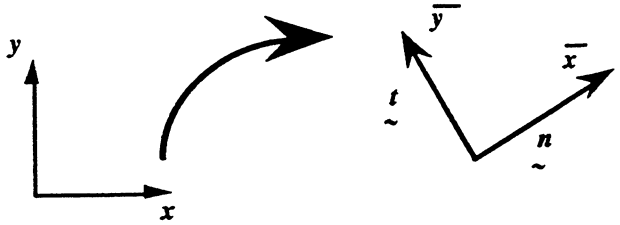


Figure 2.1. : Rotation

Le système en $\overline{\underset{\sim}{U}}$ s'écrit alors dans le repère $(\bar{x},\bar{y})$:

$$\frac{\partial \overline{\underset{\sim}{U}}}{\partial t} + \overline{\underset{\sim}{div}}\,\underset{\approx}{F}(\overline{\underset{\sim}{U}}) = 0 \tag{2.6}$$

$\overline{\underset{\sim}{div}}$ désignant l'opérateur divergence dans le repère $(\bar{x},\bar{y})$. La relation (2.6) traduit l'invariance par rotation des équations d'Euler. Le système se met sous forme quasilinéaire:

$$\frac{\partial \underset{\sim}{U}}{\partial t} + \underset{\approx}{A}(\underset{\sim}{U}).\mathrm{grad}(\underset{\sim}{U}) = 0 \tag{2.7}$$

Il est hyperbolique :

$$\forall\ \underset{\sim}{\mu} \in R^2\ ,\ \forall\ \underset{\sim}{U} \qquad \underset{\approx}{A}(\underset{\sim}{U}).\underset{\sim}{\mu} = (\underset{\sim}{F}_1)'(R(\underset{\sim}{\mu})^{-1}\underset{\sim}{U}) \tag{2.8}$$

est diagonalisable à valeurs propres réelles λ_1, λ_2 et λ_3, données par:

$$\lambda_1 = \underset{\sim}{u}.\underset{\sim}{\mu} - c \qquad \lambda_2 = \underset{\sim}{u}.\underset{\sim}{\mu} \qquad \lambda_3 = \underset{\sim}{u}.\underset{\sim}{\mu} + c \tag{2.9}$$

où c est la célérité du son dans le faz, donnée par: $c = \left(\frac{\chi p}{\rho}\right)^{1/2}$

2.3. Maillage et Volume de contrôle

Soit T_h un maillage de Ω réalisé à l'aide de quadrangles K. On a $\Omega = \bigcup_{K \in T_h} K$. Pour un élément K_i de T_h on définit ∂K_i le bord de l'élément K_i. On note K(i) l'ensemble des indices des éléments voisins de K_i, pour $j \in K(i)$, ∂K_{ij} l' arête commune à K_i et K_j, et A_j l'aire de K_j.

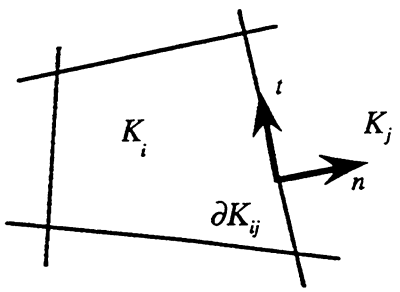


Figure 2.2. : Elément quadrangulaire

2.4. Approximation et Interpollation

On cherche à approcher $\underset{\sim}{U}$ à l'aide d'une interpolation de type Eléments Finis discontinus. On utilise deux familles d'éléments:

$$-V_0^h = \left\{ \underset{\sim}{U} \in \underset{\sim}{L}^2(\Omega) \; ; \; \forall K \in T_h \; \underset{\sim}{U} \in Q_o(K) \right\}$$

$$-V_1^h = \left\{ \underset{\sim}{U} \in \underset{\sim}{L}^2(\Omega) \; ; \; \forall K \in T_h \; \underset{\approx}{\text{grad}}(\underset{\sim}{U}) \in Q_o(K) \right\}$$

où $Q_o(K)$ désigne l'ensemble des polynômes constants sur K.

2.4.1. Approximation et Interpollation dans V_0^h

$\underset{\sim}{U} \in V_0^h$ est caractérisé par la donnée $\forall K \in T_h$ de la valeur de $\underset{\sim}{U}$ (constante) sur K. La projection $\underset{\sim}{L}^2$ sur chaque K (volume de contrôle) permet de construire un opérateur d'interpollation T_o précis à l'ordre 1 (en norme $\underset{\sim}{L}^2$ et $\underset{\sim}{L}^\infty$) de $\underset{\sim}{L}^2(\Omega)$ dans V_0^h, le schéma numérique associé sera précis au premier ordre en h (au mieux). Pour $\underset{\sim}{U} \in \underset{\sim}{L}^2(\Omega)$, $T_0(\underset{\sim}{U})$ est défini pour $\underset{\sim}{x} \in K_i$ par :

$$T_0(\underset{\sim}{U})(\underset{\sim}{x}) = \frac{1}{A_i} \int_{K_i} \underset{\sim}{U}(\underset{\sim}{x}) \, d\underset{\sim}{x} \qquad (2.10)$$

2.4.2. Approximation et Interpollation dans V_1^h

$\underset{\sim}{U} \in V_1^h$ est caractérisé par la donnée $\forall K \in T_h$ de la valeur moyenne de $\underset{\sim}{U}$ sur K, et de la valeur de $\underset{\approx}{\text{grad}}(\underset{\sim}{U})$ (constante) sur K.

On construit un opérateur d'interpollation T_1 précis à l'ordre 2 (en norme $\underset{\sim}{L}^2$ et $\underset{\sim}{L}^\infty$) de $\underset{\sim}{L}^2(\Omega)$ dans V_1^h. Cette interpollation comporte 3 étapes :

i - **une étape de <u>projection</u> $\underset{\sim}{L}^2$ sur chaque K:**

elle permet de définir les valeurs moyennes par élément $\underset{\sim}{U}_i$:

$$\underset{\sim}{U}_i = \int_{K_i} \frac{1}{A_i} \underset{\sim}{U}(\underset{\sim}{x}) \, d\underset{\sim}{x} \qquad (2.11)$$

ii - **une étape de <u>prédiction</u>:**

Il s'agit de construire, à partir des valeurs moyennes $\underset{\sim}{U}_i$, des gradients par éléments $\underset{\approx}{\text{grad}}(\underset{\sim}{U})_i$, assurant une approximation de $\underset{\sim}{U}$ précise au second ordre (en norme $\underset{\sim}{L}^2$ et $\underset{\sim}{L}^\infty$). Plusieurs techniques sont possibles. Le cas le plus simple est celui d'une approximation sur un maillage cartésien. Il est alors

possible d'estimer ces gradients en découplant les calculs selon les directions x et y, on obtient avec des notations différences finies standard:

$$\underset{\sim}{\text{grad}}_x(\underset{\sim}{U})_{ij} = \frac{\underset{\sim}{U}_{i+1,j} - \underset{\sim}{U}_{i-1,j}}{2\Delta x} \tag{2.12.a}$$

$$\underset{\sim}{\text{grad}}_y(\underset{\sim}{U})_{ij} = \frac{\underset{\sim}{U}_{i,j+1} - \underset{\sim}{U}_{i,j-1}}{2\Delta x} \tag{2.12.b}$$

Dans le cas d'un maillage non structuré, les deux valeurs pourront être obtenues par minimisation de la fonctionnelle:

$$\psi_i = \sum_{j \in K(i)} \left| \underset{\sim}{U}_i + (x_j - x_i)\, \underset{\sim}{\text{grad}}_x(\underset{\sim}{U})_i + (y_j - y_i)\, \underset{\sim}{\text{grad}}_y(\underset{\sim}{U})_i - \underset{\sim}{U}_j \right|^2 \tag{2.13}$$

La somme peut éventuellement être étendue aux éléments qui n'ont qu'un sommet commun avec K_i.

iii - une étape de correction

Cette étape est nécessaire pour préserver la stabilité du schéma. Elle consiste à limiter le gradient de $\underset{\sim}{U}$ sur chaque élément afin de ne pas créer de nouvels extremums de $\underset{\sim}{U}$ en dehors des valeurs au centre de gravité de chaque maille. Cette correction est réalisée localement, en calculant les valeurs de $\underset{\sim}{U}$ au milieu de chaque arête ∂K_{ij} de K_i, pour $j \in K(i)$, et en limitant alors le module de grad($\underset{\sim}{U}$) sur l'élément K_i de façon à ce que la valeur sur l'arête reste dans l'intervalle engendré par les $\underset{\sim}{U}(K_j)$ ($j \in K(i)$).

On définit ainsi un opérateur d'interpollation T_1 précis à l'ordre 2 (en norme $\underset{\sim}{L}^2$ et $\underset{\sim}{L}^\infty$) de $\underset{\sim}{L}^2(\Omega)$ dans V_1^h. Les schémas numériques associés seront précis au second ordre en h.

Cette technique généralise les corrections 1D décrites dans Van Leer [Van]. Il est cependant possible d'utiliser une correction de type 1D indépendamment dans chaque direction d'espace lorsque le maillage est constitué de rectangles (où si l'on se ramène localement, à l'aide d'un changement de variable à une structure rectangulaire), on obtient alors les formules suivantes :

$$\underset{\sim}{\text{grad}}_x(\underset{\sim}{U})_{ij} = s_{xij} \max\left\{0, \min\left(a\frac{\underset{\sim}{U}_{i,j} - \underset{\sim}{U}_{i-1,j}}{\Delta x}, \frac{\underset{\sim}{U}_{i+1,j} - \underset{\sim}{U}_{i-1,j}}{2\Delta x}, a\frac{\underset{\sim}{U}_{i+1,j} - \underset{\sim}{U}_{i,j}}{\Delta x}\right)\right\} \tag{2.14.a}$$

$$\underset{\sim}{\text{grad}}_y(\underset{\sim}{U})_{ij} = s_{yij} \max\left\{0, \min\left(a\frac{\underset{\sim}{U}_{i,j} - \underset{\sim}{U}_{i,j-1}}{\Delta y}, \frac{\underset{\sim}{U}_{i,j+1} - \underset{\sim}{U}_{i,j-1}}{2\Delta y}, a\frac{\underset{\sim}{U}_{i,j+1} - \underset{\sim}{U}_{i,j}}{\Delta y}\right)\right\} \tag{2.14.b}$$

$$s_{xij} = \mathrm{sgn}\,\frac{\underset{\sim}{U}_{i+1,j} - \underset{\sim}{U}_{i-1,j}}{2\Delta x} \qquad s_{yij} = \mathrm{sgn}\left(\frac{\underset{\sim}{U}_{i,j+1} - \underset{\sim}{U}_{i,j-1}}{2\Delta y}\right)$$

a constante: $1 \le a \le 2$

La limitation (2.14) peut être qualifiée de "limitation globale", dans la mesure où la pente de chaque élément est systématiquement mise à zéro dès qu'un extrêmum apparait à l'une des deux interfaces. Une technique de limitation moins astreignante introduite par Fezoui [Fez] consiste à définir deux pentes par élément, une pente par demi-maille, et de n'annuller que les pentes des interfaces où apparait un extremûms. Le but du chapitre V est de démontrer que le schéma ainsi obtenus sont convergents, pour cela on démontrera qu'il est à variation totale décroissante.

2.5. Principe des schémas

2.5.1. Cadre général

On cherche à approcher une solution faible du problème de Cauchy sur $\Omega x[0,T]$ associé à (1.1) avec une donnée initiale (à t=0): $\forall\ \varphi$ fonction test

$$-\int_{\Omega x[0,T]} \frac{\partial \varphi(\underset{\sim}{x},t)}{\partial t}\ \underset{\sim}{U}(\underset{\sim}{x},t)\ d\underset{\sim}{x}\,dt - \int_{\Omega x[0,T]} \underset{\approx}{F}(\underset{\sim}{U}(\underset{\sim}{x},t)) : \underset{\approx}{\mathrm{grad}}(\varphi(\underset{\sim}{x},t))\ d\underset{\sim}{x}\,dt = 0 \qquad (2.15)$$

$$\underset{\sim}{U}(\underset{\sim}{x},0) = \underset{\sim}{U}_o(\underset{\sim}{x}) \qquad (2.16)$$

On discrétise [0,T] en une suite d'intervalles $I^n = [t^n, t^{n+1}[$ de longueur $\Delta t = t^{n+1} - t^n$. La solution approchée recherchée dans $V_0^h x[0,T]$ (resp. $V_1^h x[0,T]$) sera indépendante de t sur chaque intervalle I^n.

Discrétisation de la condition initiale:

$$\underset{\sim}{U}_0^h = T_o(\underset{\sim}{U}_0) \quad (\text{resp. } T_1(\underset{\sim}{U}_0)) \qquad (2.17)$$

Etape de projection :

On suppose que pour $t = t^n$, $\underset{\sim}{U}_h^n \in V_0^h$ (resp. V_1^h), on considère alors la solution exacte $\underset{\sim}{U}_E(\underset{\sim}{x},t)$ du problème (1.1) avec pour donnée initiale $\underset{\sim}{U}_h^n$ à $t = t^n$. On définit alors la solution approchée à t^{n+1} par:

$$\underset{\sim}{U}_h^{n+1} = T_o(\underset{\sim}{U}_E(\underset{\sim}{x}, t^{n+1})) \quad (\text{resp. } T_1(\underset{\sim}{U}_E(\underset{\sim}{x}, t^{n+1}))) \qquad (2.18)$$

En prenant pour fonction test dans (2.15) la fonction caractéristique de $K_i x[0,T]$, on obtient après intégration par partie:

$$\underset{\sim}{U}_i^{n+1} = \underset{\sim}{U}_i^n - \frac{\Delta t}{A_i} \sum_{j \in K(i)} \Phi_{ij}^{n+1/2} \qquad (2.19)$$

$\Phi_{ij}^{n+1/2}$ est le flux associé à l'arête ∂K_{ij}

$$\Phi_{ij}^{n+1/2} = \frac{1}{\Delta t} \int_{\partial K_{ij} x [0,T]} \underset{\approx}{F}(\underset{\sim}{U}_E(x,t)).\ \underset{\sim}{n}\, ds\, dt \qquad (2.20)$$

Ceci permet de définir $\underset{\sim}{U}_h^{n+1}$ dans V_0^h ainsi que les valeurs moyennes de $\underset{\sim}{U}_h^{n+1}$ sur chaque élément K_i si l'on cherche la solution approchée dans V_1^h. Il faut néanmoins chercher des approximations utilisables numériquement, des intégrales dans (2.20).

Dans ce qui suit on se limitera à des schémas explicites en temps, plusieurs choix sont possibles suivant l'ordre de précision recherché. Il est inutile de rechercher une discrétisation précise au second ordre en temps si par ailleurs la précision en espace est limitée à l'ordre 1 (ce qui est le cas si l'on cherche $\underset{\sim}{U}$ dans V_0^h).

Si l'on cherche la solution approchée dans V_1^h on pourra obtenir un schéma du premier ordre en temps ou un schéma du second ordre en temps .

Calcul des flux $\Phi_{ij}^{n+1/2}$

Dans le cas unidimensionnel, un calcul exact de ces flux est possible en considérant le problème de Riemann posé à chaque interface, on obtient alors le schéma de Godunov. Dans le cas multidimensionnel on approche Φ_{ij} de la manière suivante:

$$\Phi_{ij}^{n+1/2} = \Phi_{ij}(t^{n+1/2})_{app} \qquad (2.21.a)$$

avec $t^{n+1/2} = \frac{t^{n+1}+t^n}{2}$,

et $$\Phi_{ij}(t)_{app} = \int_{\partial K_{ij}} \underset{\approx}{F}(\underset{\sim}{V}(x,t)).\ \underset{\sim}{n}\, ds \qquad (2.21.b)$$

où $\underset{\sim}{V}(x,t)$ est la solution du problème de Riemann :

$$\frac{\partial \underset{\sim}{\overline{V}}}{\partial \tau} + \overline{\operatorname{div}}\, \underset{\approx}{\overline{F}}(\underset{\sim}{\overline{V}}) = 0 \qquad (2.22.a)$$

avec pour donnée initiale en $\tau = t$

$$\begin{cases} \underset{\sim}{\overline{V}}(\overline{x},t) = \underset{\sim}{\overline{U}}\,(K_i)(\underset{\sim}{x}_{ij},t) & \text{pour} \quad \overline{x} \le 0 \\ \underset{\sim}{\overline{V}}(\overline{x},t) = \underset{\sim}{\overline{U}}\,(K_j)(\underset{\sim}{x}_{ij},t) & \text{pour} \quad \overline{x} > 0 \end{cases} \qquad (2.22.b)$$

où $\underset{\sim}{x}_{ij}$ désigne le milieu de l'arête ∂K_{ij}. On prends alors $t = t^n$ pour un schéma précis au premier ordre en temps, et $t = t^{n+1/2}$ pour un schéma précis au second ordre en temps.

Dans ce dernier cas la solution au temps $t = t^{n+1/2}$ peut être obtenue de manière assez simple à l'aide d'une linéarisation assez sommaire des équations au niveau de l'arête, par exemple :

$$\underset{\sim}{U}(K_i)(\underset{\sim}{x}_{ij},t^{n+1/2})) = \underset{\sim}{U}(K_i)(\underset{\sim}{x}_{ij},t^n)) - \frac{\Delta t}{2}\,\underset{\approx}{F'}(\underset{\sim}{U}(K_i)(\underset{\sim}{x}_i,t^n)).\underset{\approx}{\text{grad}}(\underset{\sim}{U}(K_i)(\underset{\sim}{x}_i,t^n)) \tag{2.23}$$

Plus généralement on pourra utiliser le flux numérique g(u,v) d'un schéma 1D à 3 points associé à l'équation :

$$\frac{\partial \overline{\underset{\sim}{U}}}{\partial t} + \frac{\partial(R(\overline{\underset{\sim}{\mu}}).\underset{\sim}{F}_1(\overline{\underset{\sim}{U}}))}{\partial x} = \underset{\sim}{0} \tag{2.24}$$

et l'on pourra alors remplacer (2.21) par :

$$\Phi_{ij}(t)_{app} = L_{ij}\, g(\overline{\underset{\sim}{U}}(K_i)(\overline{\underset{\sim}{x}}_{ij},t),\overline{\underset{\sim}{U}}(K_j)(\overline{\underset{\sim}{x}}_{ij},t)) \tag{2.25}$$

où L_{ij} est la longueur de l'arête ∂K_{ij}

Remarque 3.1.

En pratique, nous avons adopté la méthode décrite par la relation (2.25), pour le calcul du flux numérique. Plus précisemment, nous avons utilisé le solveur de Roe décrit dans le Chapitre IV. §.3.2.

Chapitre IV : Solveur de Roe - Calcul des flux

1. Schémas type Godounov

Dans ce paragraphe, on se placera dans le cas monodimensionnel. Le problème étudié est le suivant :

$$\begin{cases} \partial_t U + \partial_x f(U) = 0 & (x,t) \in \mathbb{R}x\mathbb{R}^+ \\ U(x,0) = U_o(x) & U \in \mathbb{R}^n \end{cases} \tag{1.1}$$

Nous nous intéressons dans cette partie à une famille de schémas aux différences finies, introduite par Harten-Lax et Van Leer [Van]. Ces derniers se sont inspirés des techniques développées par Godounov [God] pour établir une classe plus générale de schémas : les *schémas type Godounov.* Ces schémas utilisent un solveur approché du problème de Riemann.

1.1. Solveur approché du problème de Riemann

Soit $PR(U_g,U_d)$ le problème de Riemann suivant :

$$\begin{cases} \partial_t U + \partial_x f(U) = 0 & (x,t) \in \mathbb{R}x\mathbb{R}^+ \\ U(x,0) = U_o(x) & U \in \mathbb{R}^n \end{cases} \qquad U_o(x) = \begin{cases} U_g & \text{si } x \le 0 \\ U_d & \text{si } x \ge 0 \end{cases} \tag{1.2}$$

Définition 1.1.

On appelle solveur approché de Riemann, une approximation notée U_a, de la solution exacte, qui vérifie :

i) la consistance avec la forme intégrale du système :

$$\int_{-h_x/2}^{+h_x/2} U_a(x/\tau;\, U_g,U_d)\, dx = \frac{1}{2}(U_g + U_d) - \frac{\tau}{h}[f(U_d) - f(U_g)] \tag{1.3}$$

ii) La consistance avec la forme intégrale de la condition d'entropie : $\eta_{,t}+F_{,x}\le 0$ associée à (1.1) :

$$\int_{-h_x/2}^{+h_x/2} \eta(U_a(x/\tau;U_g,U_d))\, dx \le \frac{1}{2}(\eta(U_g) + \eta(U_d)) - \frac{\tau}{h}[F((U_d) - F(U_g)] \tag{1.4}$$

Définition 1.2. *Schéma type Godounov*

Soit $U_a(x/\tau;\, U_g,U_d)$ un solveur approché de (1.2), le schéma type Godounov associé est définit par :

$$U_j^{n+1} = \frac{1}{h_x}\int_0^{+h_x/2} U_a(x/\tau;\, U_{j-1}^n, U_j^n)\, dx + \frac{1}{h_x}\int_{-h_x/2}^{0} U_a(x/\tau;\, U_j^n, U_{j+1}^n)\, dx \tag{1.5}$$

2. Linéarisation du problème de Riemann - Solveur de Roe

2.1. Problème de Riemann linéaire

On étudie dans cette partie la solution du problème de Riemann linéaire suivant :

$$\begin{cases} \partial_t U + A.\partial_x U = 0 & (x,t) \in \mathbb{R}\times\mathbb{R}^+ \\ U(x,0) = U_o(x) & U \in \mathbb{R}^n \end{cases} \qquad U_o(x) = \begin{cases} U_g & \text{si } x \le 0 \\ U_d & \text{si } x \ge 0 \end{cases} \tag{2.1}$$

où A est une matrice (nxn) strictement hyperbolique, $\lambda_1 < \lambda_2 < \ldots < \lambda_n$ ses n valeurs propres associées aux n vecteurs propres $r_1, r_2, \ldots, r_n$.

Le problème (2.1) admet une solution analytique autoinvariante : $U(x,t; U_g,U_d) = U(x/t; U_g,U_d)$, constituée de (n-1) états intermédiaires, $U_1, U_2, \ldots, U_n$, reliant U_g à U_d, et séparés par n ondes de type discontinuités de contact de vitesses respectives λ_i. Les états intermédiaires U_i sont obtenus à l'aide de la décomposition sur la base des vecteurs propres $(r_i)_{1\le i\le n}$, de $(U_d - U_g)$, i.e. si on pose :

$$U_d - U_g = \sum_{i=1}^{n} \alpha_i r_i \tag{2.2}$$

les U_l sont donnés par :

$$U_l = U_g + \sum_{i=1}^{n} \alpha_i r_i \tag{2.3}$$

Avec ces notations, la solution $U(x/t; U_g,U_d)$ du problème de Riemann est donnée par :

$$U(x/t; U_g,U_d) = \begin{cases} U_g & \text{si} \quad x/t < \lambda_1 \\ U_l & \text{si} \quad x/t \in]\lambda_l,\lambda_{l+1}[\\ U_d & \text{si} \quad x/t > \lambda_n \end{cases} \tag{2.4}$$

2.2. Solveur de Roe

Le problème de Riemann $PR(U_g,U_d)$ peut également être écrit sous la forme suivante :

$$\begin{cases} \partial_t U + Df(U).\partial_x U = 0 & (x,t) \in \mathbb{R}\times\mathbb{R}^+ \\ U(x,0) = U_o(x) & U \in \mathbb{R}^n \end{cases} \qquad U_o(x) = \begin{cases} U_g & \text{si } x \le 0 \\ U_d & \text{si } x \ge 0 \end{cases} \tag{2.5}$$

où Df représente la Jacobienne de f.

L'idée de Roe est de linéariser localement cette Jacobienne, pour résoudre un problème de Riemann linéaire et construire ainsi un schéma type Godounov. Dans [Roe], P.L. Roe a introduit une matrice $A(U_g,U_d)$, appelée depuis lors matrice de Roe, permettant de remplacer avantageusement (i.e. à moindre coût) la résolution d'un problème de Riemann non linéaire par la résolution d'un problème

de Riemann linéaire de matrice $A(U_g,U_d)$. La matrice $A(U_g,U_d)$ doit vérifier les conditions suivantes, appelées conditions de Roe :

i) $A(U_g,U_d)$ est diagonalisable et a toutes ses valeurs propres réelles (2.6.a)

ii) $\forall U_g$ et U_d, $f(U_d) - f(U_g) = A(U_g,U_d).(U_d - U_g)$ (2.6.b)

iii) $A(U,U) = f'(U)$ (2.6.c)

Cette matrice permet, en utilisant la solution du problème de Riemann linéaire, de calculer facilement le flux numérique du schéma type Godunov associé, on obtient les expressions suivantes :

$$G(U_g,U_d) = \frac{1}{\Delta t}\int_0^{\Delta t} A(U_g,U_d)\, U_R(\frac{x}{t};U_g,U_d)\, dt$$

$$G(U_g,U_d) = \frac{1}{2}\left[f(U_g) + f(U_d) - |A(U_g,U_d)|.(U_d-U_g)\right] \quad (2.7.a)$$

$$= f(U_g) + A^-(U_g,U_d).(U_d-U_g) \quad (2.7.b)$$

$$= f(U_d) - A^+(U_g,U_d).(U_d-U_g) \quad (2.7.c)$$

Si on note $(\lambda_i)_{1\le i\le n}$ les valeurs propres de $A(U_g,U_d)$, $(r_i)_{1\le i\le n}$ (resp. $(l_i)_{1\le i\le n}$) ses vecteurs propres à droite (resp. à gauche), les matrices $|A|$, A^+ et A^- sont définies aisément par passage dans la base des vecteurs propres de A. Le flux numérique $G(U_g,U_d)$ s'écrit alors sous la forme simplifiée suivante :

$$G(U_g,U_d) = \frac{1}{2}[f(U_g) + f(U_d)] + \sum_{i=1,n} |\lambda_i|\, \alpha_i\, r_i \quad (2.8.a)$$

$$= f(U_g) + \sum_{i=1,n} \lambda_i^-\, \alpha_i\, r_i \quad (2.8.b)$$

$$= f(U_d) - \sum_{i=1,n} \lambda_i^+ \alpha_i\, r_i \quad (2.8.c)$$

avec $\alpha_i = \langle l_i, U_d - U_g\rangle$ (2.9)

où on a posé : $\lambda_i^+ = \dfrac{\lambda_i + |\lambda_i|}{2}$ $\qquad$ $\lambda_i^- = \dfrac{\lambda_i - |\lambda_i|}{2}$

3. Matrices de Roe et vecteurs paramètres

Dans ses premiers travaux sur la linéarisation de la Jacobienne du système hyperbolique (2.5), Roe a proposé un outil pratique pour déterminer des matrices vérifiant les conditions (2.6) : les matrices de Roe, et ceci pour les équations d'Euler pour la dynamique des gaz parfaits polytropiques.

Cette technique est basée sur un vecteur paramètre W, tel que les composantes du vecteur des variables conservatives U et celles de f(U) soient des fonctions quadratiques des composantes de W.

Malgré la simplicité du résultat final proposé par Roe dans son article original [Roe] et dans ceux qui ont suivis, la démarche qu'il a adopté pour l'établir comporte quelques étapes calculatoires relativement compliquées. Vila [Vil.3] en s'inspirant des idées de Roe, propose un résultat qui permet de s'affranchir de tous ces calculs, en établissant, dans le cas général, un lien simple entre la matrice de Roe et la Jacobienne Df du système hyperbolique étudié. Cette relation, de par sa simplicité, permet d'effectuer relativement aisément, une extension des résultats de la dynamique des gaz parfaits au cadre plus complexe de la dynamique des gaz réels.

3.1. Théorie générale

On cherche une matrice de Roe pour le système :

$$\frac{\partial U}{\partial t}+\frac{\partial f(U)}{\partial x}=0 \qquad\qquad U\in \mathbb{R}^n \tag{3.1}$$

En pratique U, de composantes $(u^l)_{1\le l\le n}$, n'est défini que sur un domaine de $\mathbb{R}^n$, que nous noterons dom(U), on notera également $(f^l)_{1\le l\le n}$ les composantes de f.

Définition 3.1.

On dit que le vecteur $W \in \mathbb{R}^n$, de composantes $(w^l)_{1\le l\le n}$, est un vecteur paramètre admissible pour (f,U), si et seulement si : u^i

i) U(W) est un difféomorphisme de dom(W) dans dom(U) (3.2.a)

ii) $u^k(W)$ est une forme quadratique symétrique des $(w^i)_{1\le i\le n}$

$$u^k(W) = \sum_{1\le i,j\le n} b_{ij}^k w^i w^j + \sum_{1\le i\le n} b_i^k w^i \tag{3.2.b}$$

iii) $f^k(U(W))$ est une forme quadratique symétrique des $(w^i)_{1\le i\le n}$

$$f^k(U(W)) = \sum_{1\le i,j\le n} c_{ij}^k w^i w^j + \sum_{1\le i\le n} c_i^k w^i \tag{3.2.c}$$

Théorème 3.1. [Vil.3]

On suppose que W est un vecteur paramètre admissible pour (f,U). Alors, la matrice $A(U_g,U_d)$ définie par :

$$A(U_g,U_d) = Df\,(U(\bar{W})) \tag{3.3}$$

avec

$$\bar{W} = \frac{1}{2}(W_g+W_d) \tag{3.4}$$

est une matrice de Roe pour f. De plus, on a les relations suivantes : ($\Delta x = x_d - x_g$)

$$\Delta f = Df(U(\bar{W})).\Delta U \tag{3.5.a}$$

$$\Delta U = B.\Delta W \tag{3.5.b}$$

$$\Delta f = C.\Delta W \tag{3.5.c}$$

avec :

$$B = \frac{\partial U}{\partial W}(\bar{W}) \qquad C = \frac{\partial f}{\partial W}(\bar{W}) \tag{3.6}$$

En résumé :

La matrice de Roe fournie par ce théorème n'est rien d'autre que la valeur de la Jacobienne du flux Df appliquée à un vecteur paramètre W, admissible pour (f,U) au sens de la définition 3.1. (i.e. W vérifie les conditions (3.2)).

Cette caractérisation simple des matrices de Roe à l'aide des vecteurs paramètres admissibles, permet d'une part, de simplifier les calculs du flux numérique du schéma type Godounov associé au solveur de Roe, du moins pour le système des équations d'Euler pour la dynamique des gaz parfaits polytropiques, d'autre part, d'étendre relativement aisément ce solveur aux gaz réels.

3.2. Application à la dynamique des gaz parfaits polytropiques

Les équations d'Euler s'écrivent :

$$\frac{\partial U}{\partial t} + \frac{\partial f(U)}{\partial x} = 0 \qquad U \in \mathbb{R}^n \tag{3.7.a}$$

avec

$$U = (\rho, \rho u, E)^T \qquad f(U) = (\rho u, \rho u^2+p, \rho u H)^T \tag{3.7.b}$$

et

$$H = \frac{E + p}{\rho} \qquad E = I + \frac{1}{2}\rho u^2 \tag{3.7.c}$$

où : ρ, u, E, I, H, et p désignent respectivement la densité, vitesse, énergie totale, énergie interne, enthalpie et pression du gaz.

3.2.1. Quelques notations

Soit : $i = \frac{I}{\rho}$, $h = \frac{I + p}{\rho}$, $c^2 = \chi + \kappa h$

où i, h et c désignent respectivement l'énergie interne spécifique, l'enthalpie spécifique et la célérité du son dans le gaz, avec :

$$\chi = \left(\frac{\partial p}{\partial \rho}\right)_I \qquad \kappa = \left(\frac{\partial p}{\partial I}\right)_\rho$$

La pression p est une fonction de ρ et de I. Cette fonction vérifie les relations thermodynamiques suivantes : Si on pose $\tau = \frac{1}{\rho}$, il existe une fonction entropie spécifique s telle que :

$$p(\rho;I) = g(\tau;s) \qquad i = i(\tau;s) \tag{3.8}$$

avec : $i_\tau = -p \qquad T = i_s$

T désigne la température du gaz. Le domaine physique dom(U) est défini par

$$\rho > 0 \qquad \text{et} \qquad i > 0 \tag{3.9}$$

Les fonction g et i doivent vérifier les relations suivantes :

$$\text{i) } g > 0 \qquad \text{ii) } g_\tau < 0 \qquad \text{iii) } g_{\tau\tau} > 0$$
$$\text{iv) } g_s > 0 \qquad \text{v) } T = i_s > 0 \tag{3.10}$$

On renvoie à Smith [Smi] pour une étude détaillée des fonctions i et g, où il démontre, en particulier, que le problème de Riemann admet une solution unique si et seulement si la pression vérifie la condition suivante : $\frac{\partial P(\tau,i)}{\partial \tau} \leq \frac{p^2}{2i}$.

La matrice A(U)=DF(U) est donnée par :

$$A(U) = \begin{pmatrix} 0 & 1 & 0 \\ \chi-(2-\kappa)u^2/2 & (2-\kappa)\,u & \kappa \\ u\left(\chi+\kappa u^2/2-H\right) & H-\kappa u^2 & (1+\kappa)u \end{pmatrix} \tag{3.11}$$

ses valeurs propres sont :

$$\lambda_1 = u-c \qquad \lambda_2 = u \qquad \lambda_3 = u+c \tag{3.12}$$

La matrice des vecteurs propres à droite de A(U) est :

$$R = \begin{pmatrix} 1 & 1 & 1 \\ u-c & u & u+c \\ H-uc & (u^2/2)-(\chi/\kappa) & H+uc \end{pmatrix} \tag{3.13}$$

Remarque 3.1.

La condition $g_\tau < 0$ est équivalente à $c^2 > 0$. C'est la condition d'hyperbolicité du système.

Dans le cas des gaz parfaits pour lesquels la loi d'état est donnée par :

$$p(\rho,I) = (\gamma-1)\, I = (\gamma-1)\,\rho\, i \tag{3.14}$$

où γ est le rapport des chaleurs spécifiques, Roe propose le vecteur paramètre admissible W suivant :

$$W = \left(\sqrt{\rho},\ \sqrt{\rho}\,u,\ \frac{E+p}{\sqrt{\rho}}\right)^T \tag{3.15}$$

Vila dans [Vil.3] établit le résultat suivant qui généralise celui énnocé ci-dessus à toutes les lois d'état pour lesquelles les relations (3.2.a) et (3.2.b) soient vérifiées.

Proposition 3.1.

Si W est un vecteur paramètre admissible pour (U,f) donnés par (3.7.b) et (3.7.c), alors on a :

$$\begin{cases} w_1 = a_1\sqrt{\rho} \\ w_2 = a_2\sqrt{\rho}\ u \\ w_3 = a_3\,\dfrac{E+p}{\sqrt{\rho}} + c_1\sqrt{\rho} + c_2\sqrt{\rho}\ u \end{cases} \tag{3.16}$$

où a_1, a_2, a_3, c_1 et c_2, sont des constantes réelles, telles que $(a_1 a_2 a_3 \neq 0)$. Réciproquement, un vecteur W vérifiant (3.16) est vecteur paramètre admissible pour (f,U) donnés par (3.7.b) et (3.7.c) si et seulement si :

$$p(\rho,i) = \kappa i + \chi_1\rho + \chi_2\sqrt{\rho} + \delta \tag{3.17}$$

Dans toute la suite, le vecteur W, défini par (3.15), sera utilisé comme vecteur paramètre admissible. Ce choix permet de simplifier considérablement les calculs.

Roe et Pike ont présenté dans [RP], une nouvelle méthode algèbrique leur permettant de calculer aisément les paramètres α_i en fonction de ΔV, où V est donné par

$$V = (\rho, u, p)^T \tag{3.18}$$

En utilisant le résultat du théorème 3.1, Vila [Vil.3] a proposé une démonstration directe de ce résultat (sans calculs algèbriques), qui reste valable dans le cas plus général de l'équation d'état donnée par (3.17).

Posons : $V = (\rho, u, p)^T$, $\tilde{U} = \left(\tilde{\rho}, \tilde{\rho}\tilde{u}, E(\tilde{\rho},\tilde{u},\tilde{H})\right)^T$, $\tilde{\rho} = \sqrt{\rho_g \rho_d}$

$\tilde{c}_* = c(\tilde{U})$

$\tilde{c}_*^2 = \kappa \tilde{h}_* + \chi_1 + \chi_2/(\tilde{\rho})^{1/2}$ $\qquad \tilde{h}_* = h(\tilde{U}) = \tilde{H} - \frac{1}{2}\tilde{u}^2$

Proposition 3.2

Soit W le vecteur paramètre admissible pour les équation d'Euler avec la loi d'état (3.17), alors les coefficients $\alpha_i = \langle l_i(\bar{W}), \Delta U \rangle$ (i=1,2,3), pour le calcul du flux numérique défini par (2.8), sont tels que :

$$\alpha_i = \langle l_i(\tilde{U}), \frac{\partial U}{\partial V}(\tilde{U})).\Delta V \rangle \qquad (3.19.a)$$

$$A(U(\bar{W})) = A(\tilde{U}) \qquad (3.19.b)$$

et

$$\alpha_1 = (\Delta p - \tilde{c}_* \tilde{\rho} \Delta u) / (2\tilde{c}_*^2) \qquad (3.20.a)$$

$$\alpha_2 = \Delta \rho - (\Delta p) / \tilde{c}_*^2 \qquad (3.20.b)$$

$$\alpha_3 = (\Delta p + \tilde{c}_* \tilde{\rho} \Delta u) / (2\tilde{c}_*^2) \qquad (3.20.c)$$

Démonstration

Voir [Vil.3].

Ce résultat trés simple de calcul des α_i fournit un algorithme de calcul du flux numérique, d'une grande simplicité de mise en oeuvre, trés performant et d'une grande rapidité d'execution.

3.2.2. Algorithme simplifié de calcul des flux

Pour simplifier le calcul du flux $g(u_g,u_d)$ donné par les formules (2.8) on remarque que comme les formules (2.8.a)-(2.8.c) sont équivalentes, il suffit de choisir celle qui comporte le moins de termes à calculer (dans la sommation). Ainsi la formule (2.8.a) est éliminée d'office car elle comporte toujours le nombre maximal d'opérations. Par contre les formules (2.8.b) et (2.8.c) vont comporter, au vu des expressions des valeurs propres données par (3.12), vont contenir un nombre minimal de termes suivant le signe de la valeur propre $\lambda_2 = u$.

En effet cette remarque fournit l'algorithme simplifié suivant :

SI $\lambda_2 < 0$ **ALORS** (on choisit la formule (2.8.c) qui donne ici :)

$g(u_g,u_d) = f(u_d)$

SI $\lambda_3 > 0$ **ALORS**

$g(u_g,u_d) = g(u_g,u_d) - \lambda_3 \alpha_3 r_3$

SINON (on choisit la formule (2.8.b) qui donne ici:)

$g(u_g,u_d) = f(u_g)$

SI $\lambda_1 > 0$ **ALORS**

$g(u_g,u_d) = g(u_g,u_d) + \lambda_1 \alpha_1 r_1$

FIN ALGORITHME

Ainsi la somme intervenant dans le calcul du flux contient à chaque fois au maximum un terme, au lieu de trois dans le cas géneral. On n'a donc pas besoin du terme α_2. Cet algorithme reste encore valable pour les équations d'Euler bidimensionnelles ou tridimensionnelles parce que les valeurs propres dans ce cas sont encore données par des expressions analogues au cas monodimensionnel, en effet dans le cas bidimensionnel les valeurs propres sont au nombre de quatre et sont données par

$$\lambda_1 = u\text{-}c \qquad \lambda_2 = \lambda_3 = u \qquad \lambda_4 = u\text{+}c$$

et dans le cas tridimensionnel, il y'en a cinq et elles sont données par

$$\lambda_1 = u\text{-}c \qquad \lambda_2 = \lambda_3 = \lambda_4 = u \qquad \lambda_5 = u\text{+}c$$

et dans tous les cas on n'a pas besoin de calculer les coefficients α_i correspondants à la valeur propre centrale $\lambda = u$.

Remarque 3.2

Si la pression est une fonction affine de ρ et I, autrement dit si le coefficient χ_2 dans (3.17) est nul (χ_2=0, et $p(\rho,i) = \kappa i + \chi_2\rho + \delta$), alors on a : $\tilde{c}_* = c(U(\bar{W}))$.

Remarque 3.3

La pression p, définie par (3.17), vérifie (3.10) sur tout le domaine admissible dom(U) défini par (3.9), si et seulement si :

$$\kappa > 0, \qquad \chi_1 > 0, \qquad \chi_2 > 0, \qquad \delta > 0 \tag{3.21}$$

Etant donnés (U_g,U_d) dans dom(U), le théorème 3.1 nous fournit une classe de matrices de Roe pour le système hyperbolique non linéaire (3.7.a), pourvu qu'on trouve un vecteur W admissible pour (f,U), au sens de la définition 3.1. Cette matrice est donnée par

$$A(U_g,U_d) = Df(U(\bar{W}))$$

La question maintenant, est de savoir sous quelles conditions, le vecteur $U(\bar{W})$ est lui aussi dans dom(U). On considère toujours les équations d'Euler pour la dynamique des gaz, avec une loi d'état affine, i.e. ($\chi_2 = 0$, et $\chi = \chi_1$)

$$p = p(\rho,i) = \kappa i + \chi\rho + \delta \tag{3.22}$$

posons

$$\bar{\rho} = \rho(U(\bar{W})) = \frac{\sqrt{\rho_g}+\sqrt{\rho_d}}{2} \qquad \tilde{i} = \frac{\sqrt{\rho_g}\, i_g + \sqrt{\rho_d}\, i_d}{\sqrt{\rho_g} + \sqrt{\rho_d}} \tag{3.23}$$

$$\tilde{h} = \frac{\sqrt{\rho_g}\, h_g + \sqrt{\rho_d}\, h_d}{\sqrt{\rho_g} + \sqrt{\rho_d}} \qquad \tilde{c}^2 = \frac{\sqrt{\rho_g}\, c_g^2 + \sqrt{\rho_d}\, c_d^2}{\sqrt{\rho_g} + \sqrt{\rho_d}}$$

$$\bar{c} = c(U(\bar{W})) \qquad \tilde{c} = c(\tilde{U})$$

$$\bar{i} = i(U(\bar{W})) \qquad \tilde{i} = i(\tilde{U})$$

alors on a $$\left(\bar{c}\right)^2 = \left(\tilde{c}\right)^2 = \left(\tilde{c}\right)^2 + \frac{\kappa}{8}\frac{\tilde{\rho}}{2\bar{\rho}_*}(\Delta u)^2 = \kappa + \chi\tilde{h} + \frac{\kappa}{8}\frac{\tilde{\rho}}{2\bar{\rho}_*}(\Delta u)^2$$

$$\bar{i} = \tilde{i} + \frac{1}{4(\kappa+1)}\left\{\frac{\delta\left(\Delta\sqrt{\rho}\right)^2}{\tilde{\rho}\,\bar{\rho}^*} + \frac{\tilde{\rho}}{2\bar{\rho}_*}(\Delta u)^2\right\}$$

$$\tilde{i} = \tilde{i} + \frac{\tilde{\rho}(\Delta u)^2}{8(\kappa+1)\bar{\rho}_*}$$

Au vu de ces relations, nous pouvons prouver le résultat suivant :

Proposition 3.3

Etant donnés (U_g, U_d) dans dom(U), si le paramètre κ de la loi d'état (3.22) vérifie : ($\kappa>0$), alors, les matrices de Roe associées à $U(\bar{W})$ et $\tilde{U}$ est hyperbolique ($\bar{c}\geq 0$), et $\tilde{U}$ est dans dom(U). En plus, si :

$\delta>0$, alors $U(\bar{W})$ est également dans dom(U).

En résumé, les résultat présentés dans cette section permettent de généraliser la construction de Roe et de "Roe et Pike" à la famille des lois d'état définie par (3.23), et surtout de définir un cadre théorique pour étudier la construction des matrices de Roe.

Remarque 3.4.

Même si la reconstruction décrite dans la partie 3.2, n'est pas directement applicables dans le cas où la pression n'est pas de la forme (3.17), autrement dit, lorsque p n'est pas une fonction quadratique des composantes du vecteur W donné par (3.15), elle constitue néanmoins, un excellent outil de base pour étudier ce type de lois d'état, et essentiellement le cas général de la dynamique des gaz réels. En effet, dans ce domaine, différents auteurs ont proposé des méthodes pour étendre la technique de construction de Roe ([Gl], [To]).

Vila [Vil.3] a établit un résultat généralisant le théorème 3.1, et permettant de définir un cadre pour la construction de la matrice de Roe pour la dynamique des gaz réels. Ce résultat consiste essentiellement en une technique de "linéarisation locale" de la pression. Cette linéarisation assure la dépendance quadratique de U et du flux f(U), en fonction des composantes du vecteur paramètre W, et permet par la suite, d'utiliser le résultat du théorème 3.1 et la paramétrisation de la matrice de Roe par le vecteur W défini par (3.15).

CHAPITRE V

ETUDE D'UN SCHEMA TVD D'ORDRE 2

0. Introduction

On étudie dans cette partie un schéma type Godounov d'ordre deux, basé sur la méthode de prédiction-correction introduite par Van Leer [Van] dans la construction des schémas type MUSCL.

La construction de ce schéma est inspirée d'un papier de Vila [Vil.2], où il effectue une analyse générale d'une classe de schémas d'ordre deux de type Godounov, pour lesquels il prouve un résultat de convergence vers la solution faible entropique du problème continu, en établissant la stabilité TVD (sous une condition de CFL plus restrictive que celle utilisée pour les schémas type Godounov, qui sont comme on le sait, d'ordre un parce que monotones).

Le schéma présenté dans cette partie est légèrement différent de la famille de schémas étudiée dans [Vil.2], qui sont également construits par la méthode de prédiction-correction. La différence, comme on le verra plus en détail dans la suite, est située essentiellement au niveau de la phase de correction (ou limitation). La technique de prédiction-correction consiste, dans un premier temps, à calculer un gradient par élément pour chercher une approximation dans l'espace des fonctions affines par morceaux. Ensuite, ces gradients sont limités de façon à préserver la monotonie du schéma. La limitation introduite par Van Leer et utilisée par Vila, est une limitation de type "limitation globale", dans ce sens que le gradient de la fonction approchée est annulé en un point, dès que le gradient dans la direction d'un élément voisin est nul. Cette technique de limitation, même si elle permet d'obtenir un schéma d'ordre deux en espace, produit une trop grande viscosité numérique et une dégradation importante de la précision.

Un critère de limitation moins sévère a été introduit par Fezoui [Fez], appelé critère de "limitation ponctuelle". Ce critère consiste non pas à limiter les gradients eux même, mais à limiter les valeurs d'intégration (valeurs de la variable approchée aux interfaces des mailles).

Dans [Fez], des tests numériques effectuées pour discrétiser les équations d'Euler bidimensionnelles, pour le problème de Riemann, ont montré l'efficacité de la technique de limitation ponctuelle. Cependant, aucun résultat théorique concernant le schéma n'a été prouvé.

Le but de cette partie est de prouver que ce schéma vérifie une propriété de stabilité TVD, sous une certaine condition de CFL qu'on précisera plus loin. Ensuite, en utilisant les résultats de Vila [Vil.2], concernant la dérivation de l'inégalité d'entropie continue à partir d'une inégalité d'entropie discrète, par l'utilisation d'une correction comportant un terme additif en Ch^{α}, on peut prouver un résultat de convergence vers la solution faible entropique du problème continu.

1. Schémas du second ordre : Méthode de prédiction-correction

1.1. Notations

On cherche à approximer la solution du problème de Cauchy suivant :

$$\begin{cases} \partial_t U + \partial_x f(U) = 0 & (x,t) \in \mathbb{R}\times\mathbb{R}^+ \\ U(x,0) = U_o(x) & U \in \mathbb{R} \end{cases} \tag{1.1}$$

où f est une fonction régulière de $\mathbb{R}\to\mathbb{R}$. U_o est une fonction dans $L^1(\mathbb{R})\cap L^\infty(\mathbb{R})$. Par souci de simplicité, nous utiliserons un mallaige régulier. Soit τ le pas de temps, h_x le pas d'espace. Le rapport :

$$\lambda_x = \frac{\tau}{h_x} \tag{1.2}$$

sera considéré constant. On définit un maillage régulier à l'aide des notations suivantes :

$$t_n = n\tau \quad (n \in \mathbb{N}), \quad x_i = ih_x, \quad (i \in \mathbb{Z} \text{ ou } i \in \mathbb{Z}+\frac{1}{2})$$

$$I_i =]x_{i-1/2}, x_{i+1/2}[$$

$$\Delta_+ a_i = a_{i+1} - a_i.$$

On note $U^h(x,t)$ l'approximation numérique de la solution faible entropique du problème (1.1). On cherche U^h dans les espaces V_h^0 et V_h^1 définis par :

$$V_h^0 = \{f(x);\ \forall\, i \in \mathbb{Z},\ f \text{ est constante sur } I_i\} \tag{1.3.a}$$

$$V_h^1 = \{f(x);\ \forall\, i \in \mathbb{Z},\ f \text{ est affine sur } I_i\} \tag{1.3.b}$$

On suppose que $U^h(x,t_n)$ est dans V_h^0 (resp. V_h^1) et on cherche à calculer $U^h(x,t_{n+1})$ dans V_h^0 (resp. V_h^1). Si $U^h(x,t_n)$ est dans V_h^0, les valeurs moyennes sur les intervalles I_i sont utilisées comme degré de liberté dans V_h^0, et on pose (pour simplifier les notations)

$$U_i = U^h(x_i,t_n), \qquad U^i = U^h(x_i,t_{n+1}) \tag{1.4}$$

Si $U^h(x,t_n)$ est dans V_h^1, on pose

$$\begin{cases} U_i = U^h(x_i, t_n), \\ U_{i+1/2,-} = U^h(x_{i+1/2}-0, t_n), \\ U_{i+1/2,+} = U^h(x_{i+1/2}+0, t_n), \\ \delta_i = (U_{i+1/2,-} - U_{i-1/2,+})/2. \end{cases} \tag{1.5}$$

De la même façon que dans (1.4), les variables à l'instant t_{n+1} seront représentées par un indice placé en exposant. Dans ce cas, les degrés de liberté choisis pour représenter U^h dans V_h^1, sont les valeurs moyennes sur les intervalles I_i et les "pentes" δ_i.

Dans le paragraphe suivant, la technique de prédiction-correction est présentée à l'aide du schéma de Godounov (schéma du premier ordre) pour obtenir un schéma du second ordre.

1.2. Schéma de Godounov du second ordre

La solution approchée $U^h(x,n\tau)$ du problème (1.1), calculée à l'aide du schéma de Godounov, vérifie

$$U^h(x, n\tau) \in V_h^o, \qquad U^h(x, (n+1)\tau) \in V_h^o,$$

$$U^i = \frac{1}{h_x}\int_{I_i} U_E(x, (n+1)\tau)\, dx \tag{1.6}$$

où U_E est la solution exacte du problème (1.1), avec comme condition initiale : $U_o(x)=U^h(x,n\tau)$. En utilisant la formule de Green, on obtient

$$U^i = U_i - \lambda_x \Delta_+ f(U_{i-1/2}) \tag{1.7}$$

où $U_{i-1/2}$ désigne la solution du problème de Riemann $PR(U_{i-1},U_i)$ en $(x=x_{i-1/2})$, c'est-à-dire (1.1) avec

$$U_o(x) = \begin{cases} U_{i-1} & \text{si} \quad x<x_{i-1/2} \\ U_i & \text{si} \quad x>x_{i-1/2} \end{cases}$$

L'idée de Van Leer est de généraliser cette méthode, en utilisant V_h^1 comme espace fonctionnel pour la solution approchée. Autrement dit, on suppose $U^h(x,n\tau)$ dans V_h^1 et on cherche $U^h(x,(n+1)\tau)$ dans V_h^1. Le calcul de U^i et de δ_i se fait en deux étapes :

Etape (i) : Calcul des valeurs moyennes, données comme dans (1.6) par

$$U^i = \frac{1}{h_x}\int_{I_i} U_E(x, (n+1)\tau)\, dx \tag{1.8}$$

où U_E est la solution exacte du problème (1.1), avec comme condition initiale : $U_o(x)=U^h(x,n\tau)$. En utilisant la formule de Green, on obtient

$$U^i = U_i - \lambda_x \Delta_+ g(U_{i-1/2}) \tag{1.9}$$

où $U_{i-1/2}$ désigne la solution du problème de Riemann $PR(U_{i-1},U_i)$ en $(x=x_{i-1/2})$, c'est-à-dire (1.1) avec

$$U_o(x) = \begin{cases} U_{i-1} - \dfrac{h_x}{2}\delta_{i-1} & \text{si} \quad x<x_{i-1/2} \\ U_i + \dfrac{h_x}{2}\delta_i & \text{si} \quad x>x_{i-1/2} \end{cases} \tag{1.10}$$

Etape (ii) : Calcul des pentes δ_i. Une première estimation est donnée par la formule suivante

$$\tilde{\delta}^i = (U^{i+1} - U^{i-1})/2 \tag{1.11}$$

Le schéma construit à l'aide de cette famille de pentes n'est pas TVD. Il est nécessaire, pour éviter les oscillations, d'utiliser une correction. Dans [Vil.2], la la limitation utilisée est la suivante

$$\delta_i = \sigma^i \max\left\{0, \min\left\{\sigma^i \Delta_+ U^i, \left|\tilde{\delta}^i\right|, \sigma^i \Delta_+ U^{i-1}\right\}\right\} \tag{1.12}$$

avec

$$\sigma^i = \mathrm{sgn}(\tilde{\delta}^i) \tag{1.13}$$

Il est clair que cette limitation entraîne l'annulation de la pente de l'élément x_i (et le retour au premier ordre) dès qu'il y'a création d'un extrêmun local à l'une des deux interfaces $x_{i\pm1/2}$. Autrement dit, dès que

$$\begin{cases} \min(U_{i-1}, U_i) \geq U_{i-1/2} \text{ ou } U_{i-1/2} \geq \max(U_{i-1}, U_i) \\ \text{ou} \\ \min(U_i, U_{i+1}) \geq U_{i+1/2} \text{ ou } U_{i+1/2} \geq \max(U_i, U_{i+1}) \end{cases} \quad \text{alors} \quad \delta^i = 0$$

Cette correction du schéma antidiffusif sert essentiellement à éviter les oscillations et à <u>préserver la monotonie</u>, en ce sens que si $U^h(x,n\tau)$ est une fonction monotone en x, alors $U^h(x,(n+1)\tau)$ l'est aussi.

Il n'est pas nécessaire d'utiliser une correction au niveau des pentes (trop pénalisante). Une limitation locale, au niveau des valeurs arêtes, est suffisante comme (on le verra plus loin), pour obtenir un schéma d'ordre deux, de faibles amplitudes d'oscillations et de surcroît TVD.

1.3. Schéma du second ordre à limitation locale

Dans ce paragraphe, sera construite une classe de schémas du second ordre, par la méthode de prédiction-correction, en utilisant une limitation "locale" (moins sévère que la limitation globale décrite précèdemment). Cette famille de schémas précis à l'ordre deux sera construite à l'aide de flux numériques associés à des schémas à trois points TVD.

Le principe de la limitation locale est de corriger les valeurs d'intégration ou valeurs arêtres et non les pentes des éléments elles mêmes. Pour pouvoir se ramener au cadre de la limitation globale, on a choisi de définir deux pentes par élément (une pente par demi-maille), δ_i^- et δ_i^+ par

$$\begin{cases} \delta_i^- = \sigma^i \max\left\{0, \min\left\{a\, \sigma^i \Delta_+ U_{i-1}, \left|\tilde{\delta}^i\right|, \mathrm{sgn}(\Delta_+ U_i)\left|\tilde{\delta}^i\right|\right\}\right\} & \text{(1.14.a)} \\ \delta_i^+ = \sigma^i \max\left\{0, \min\left\{a\, \sigma^i \Delta_+ U_i, \left|\tilde{\delta}^i\right|, \mathrm{sgn}(\Delta_+ U_{i-1})\left|\tilde{\delta}^i\right|\right\}\right\} & \text{(1.14.b)} \end{cases}$$

ou

$$\begin{cases} \delta_i^- = \sigma^i \max\{0, \min\{a\,\sigma^i \Delta_+ U_{i-1}, |\tilde{\delta}^i|, b\,\sigma^i \Delta_+ U_i, Ch^\alpha\}\} & (1.15.a) \\ \delta_i^+ = \sigma^i \max\{0, \min\{a\,\sigma^i \Delta_+ U_i, |\tilde{\delta}^i|, b\,\sigma^i \Delta_+ U_{i-1}, Ch^\alpha\}\} & (1.15.b) \end{cases}$$

où C est une constante réelle positive, α est un paramètre réel tel que : $0 < \alpha < 1$, et

$$\sigma^i = \mathrm{sgn}(\tilde{\delta}^i)$$

$$\tilde{\delta}^i = \frac{U^{i+1} - U^{i-1}}{2} = \frac{\Delta_+ U_{i-1} + \Delta_+ U_i}{2}$$

et "a" est une constante réelle à déterminer, et b est une constante rélle que l'on choisira suffisement grande pour ne pas tenir compte dans la limitation du terme correspondant.

Les termes $(\mathrm{sgn}(\Delta_+ U_i)\,|\,\tilde{\delta}^i\,|)$ et $(\mathrm{sgn}(\Delta_+ U_{i-1})\,|\,\tilde{\delta}^i\,|)$ sont introduits dans la limitation (1.15), pour annuler la pente en x_i lorsque $U^h(x,n\tau)$ a un extrêmum local en x_i (i.e. $\mathrm{sgn}(\Delta_+ U_{i-1}) \neq \mathrm{sgn}(\Delta_+ U_i)$). Les valeurs arêtes sont définies par

$$\begin{cases} U_{i+1/2,-} = U_i + \frac{1}{2}\delta_i^+ \\ U_{i-1/2,+} = U_i - \frac{1}{2}\delta_i^- \end{cases} \qquad (1.16)$$

Le schéma à 5 points s'écrit

$$U^i = U_i - \lambda_x \Delta_+ g(U_{i-1/2,-}, U_{i-1/2,+}) \qquad (1.17)$$

où g est le flux numérique d'un schéma à trois points TVD, autrement dit, les coefficients incrémentaux définis par

$$\begin{cases} C(u,v) = \lambda_x \dfrac{f(u) - g(u,v)}{v - u} \\ D(u,v) = \lambda_x \dfrac{f(v) - g(u,v)}{v - u} \end{cases} \qquad (1.18)$$

vérifient

$$\forall\, u,v \in \mathbb{R} \qquad C(u,v) \geq 0,\quad D(u,v) \geq 0,\quad \text{et} \quad Q(u,v) = C(u,v) + D(u,v) \leq 1 \qquad (1.19)$$

La propriété (1.19) est équivalente à

$$\forall\, u,v \in \mathbb{R} \qquad |w(u,v)| \leq Q(u,v) \leq 1 \qquad (1.20)$$

où Q(u,v) est la viscosité numérique du schéma à trois points de flux numérique g, défini par

$$Q(u,v) = C(u,v) + D(u,v) = \lambda_x \frac{f(u) + f(v) - 2g(u,v)}{v - u} \qquad (1.21)$$

et

$$w(u,v) = \lambda_x \frac{f(v) - f(u)}{v - u} \qquad (1.22)$$

Pour alléger la présentation, nous allons introduire (voir [Vil.2]), la famille de viscosités numérique $H(\nu,\mu)$, définie pour chaque couple de réels (ν,μ), par

Définition 1.1

Un schéma à trois points appartient à la classe de schémas $H(\nu_o,\mu_o)$, si et seulement si sa viscosité numérique vérifie la propriété suivante

$H(\nu_o,\mu_o)$ Soit M un nombre réel tel que $\max_{|u|\leq M} |f'(u)| \leq \nu_o$

alors $|w(u,v)| \leq Q(u,v) \leq \mu_o$

Remarque 1.1.

La classe $H(\nu,\mu)$ n'est pas vide, elle contient par exemple, la classe des schémas à trois points TVD qui est dans $H(1,1)$, le schéma de Godounov est dans $H(\nu,\mu)$ tel que : $\nu \leq 1$ et $\nu \leq \mu$.

Nous avons le résultat suivant, concernant la stabilité TVD et la convergence du schéma (1.16)-(1.17).

Théorème 1.1

Si le flux numérique g du schéma (1.17) appartient à la classe $H(\nu,\mu)$ telle que

$$\mu + \frac{a+b}{2}(\nu+\mu) \leq 1 \qquad (1.23)$$

où "a et b" sont définies dans (1.14) et (1.15) et vérifie : $0 \leq a \leq 1$ et $a < b$, alors le schéma (1.16)-(1.17) est TVD. En plus, si on utilise la limitation (1.15.a)-(1.15.b), le schéma correspondant converge vers la solution faible entropique de (1.1).

Remarque 1.2.

Pour que le schéma (1.16)-(1.17) soit consistant avec le second ordre en espace, il faut prendre la constante "a" égale à 1 (a=1). En pratique, pour réduire la condition de CFL, on peut prendre une valeur inférieure à 1 pour a (a<1), le schéma obtenu n'est plus consistant avec le second ordre, mais il est néanmoins meilleur que celui du premier ordre (moins diffusif). Le schéma du premier ordre est obtenu pour a=2.

Nous nous contenterons de démontrer la stabilité TVD du schéma (1.16)-(1.17), la convergence vers la solution faible entropique s'en déduit aisément par utilisation de la limitation (1.15.a)-(1.15.b) (voir Vila [Vil.2]). La principale difficulté dans la démonstration de convergence de ce type de schémas étant l'obtention de la stabilité en norme BV ou la stabilité TVD, la convergence en est déduite, via la stabilité L^∞, par application du théorème de compacité de Helly.

1.4. Forme incrémentale du schéma

Il est en général commode, pour établir la stabilité TVD d'un schéma, de l'écrire sous forme incrémentale. Pour le schéma (1.17) on obtient

$$U^i = U_i - \lambda_x \Delta_+ g(U_{i-1/2,-}, U_{i-1/2,+})$$

$$U^i = U_i + C_{i+1/2}\ \Delta_+ U_i - D_{i-1/2}\ \Delta_+ U_{i-1} \qquad (1.24)$$

avec

$$C_{i+1/2} = \frac{1}{\Delta_+ U_i}\left[C(U_{i+1/2,-}, U_{i+1/2,+})\ \Delta_2 U_i + C(U_{i-1/2,+}, U_{i+1/2,-})\ \Delta_1 U_i\right] \qquad (1.25.a)$$

$$D_{i+1/2} = \frac{1}{\Delta_+ U_i}\left[D(U_{i+3/2,-}, U_{i+1/2,+})\ \Delta_3 U_i + DU_{i+1/2,-}, U_{i+1/2,+})\ \Delta_2 U_i\right] \qquad (1.25.b)$$

et les $\Delta_k U_i$ (k=1,..3), sont donnés par (voir figure 1.1)

$$\begin{cases} \Delta_1 U_i = U_{i+1/2,-} - U_{i-1/2,+} \\ \Delta_2 U_i = U_{i+1/2,+} - U_{i+1/2,-} \\ \Delta_3 U_i = U_{i+3/2,-} - U_{i+1/2,+} \end{cases}$$

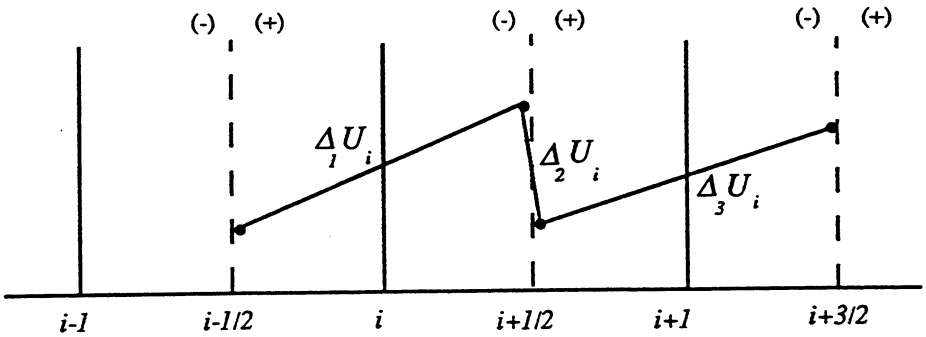


figure 1.1

Harten a montré qu'un schéma écrit sous forme incrémentale, est TVD, autrement dit

$$\sum_{i\in\mathbb{Z}} |\Delta_+ U_i| \le \sum_{i\in\mathbb{Z}} |\Delta_+ U^i| \qquad (1.26)$$

si pour tout i dans $\mathbb{Z}$,

$$C_{i+1/2} \ge 0, \quad D_{i+1/2} \ge 0, \qquad (1.27.a)$$

et

$$(C_{i+1/2} + D_{i+1/2}) \le 1 \qquad (1.27.b)$$

Lemme 1.1

Si le flux numérique g du schéma (1.17) est le flux numérique d'un schéma à trois points TVD, et si la constante "a" vérifie : $0 \le a \le 1$, alors on a

$$C_{i+1/2} \ge 0 \quad \text{et} \quad D_{i+1/2} \ge 0 \qquad (1.28)$$

Démonstration du lemme 1.1

Les coefficients incrémentaux C(u,v) et D(u,v) définis par (1.18) sont positifs car g est le flux numérique d'un schéma à trois points TVD. Une condition suffisante pour obtenir (1.28), est de montrer que lorsque $\Delta_+U_i \neq 0$, les rapports $\frac{\Delta_k U_i}{\Delta_+ U_i}$ sont positifs pour tout i dans $\mathbb{Z}$ et k dans {1,2,3}.

Si on pose $\alpha_i^{\pm} = \frac{\delta_i^{\pm}}{\Delta_+ U_i}$ $\qquad$ $\beta_i^{\pm} = \frac{\delta_{i+1}^{\pm}}{\Delta_+ U_i}$

on obtient

$$\begin{cases} \frac{\Delta_1 U_i}{\Delta_+ U_i} = \frac{1}{2}(\alpha_i^+ + \alpha_i^-) \\ \frac{\Delta_2 U_i}{\Delta_+ U_i} = 1 - \frac{1}{2}(\beta_i^- + \alpha_i^+) \\ \frac{\Delta_3 U_i}{\Delta_+ U_i} = \frac{1}{2}(\beta_i^+ + \beta_i^-) \end{cases}$$

d'après construction des demi-pentes $\delta_i^{\pm}$, on a

$$\begin{cases} 0 \le \delta_i^-/\Delta_+U_{i-1} \le a \\ 0 \le \delta_i^+/\Delta_+U_i \le a \end{cases} \quad \text{et} \quad \begin{cases} 0 \le \delta_{i+1}^-/\Delta_+U_i \le a \\ 0 \le \delta_{i+1}^+/\Delta_+U_{i+1} \le a \end{cases}$$

ce qui entraîne

$$\begin{cases} 0 \le \alpha_i^+ \le a \\ 0 \le \beta_i^- \le a \end{cases} \qquad (1.29)$$

d'après (1.29), on a

$$1 - \frac{1}{2}(\beta_i^- + \alpha_i^+) \ge 1-a$$

Si : $a \le 1$, alors : $\frac{\Delta_2 U_i}{\Delta_+ U_i} \ge 0$.

D'autre part, les α_i^+ sont positifs, il suffirait donc de montrer que $\alpha_i^- \ge 0$, cela entraînerait

$$\frac{\Delta_2 U_i}{\Delta_+ U_i} = \alpha_i^+ + \alpha_i^- \ge 0$$

Or α_i^- peut être écrit de façon équivalente, sous la forme suivante

$$\alpha_i^- = \frac{\delta_i^-}{\Delta_+ U_i} = \frac{\delta_i^-}{\Delta_+ U_{i-1}} \frac{\Delta_+ U_{i-1}}{\Delta_+ U_i}$$

Le terme $\frac{\delta_i^-}{\Delta_+ U_i}$ est positif ou nul, par construction de δ_i^- (voir (1.14.a)). D'autre part, une condition nécessaire pour que δ_i^- soit non nul, est que : $\Delta_+U_{i-1}/\Delta_+U_i \ge 0$. En effet, lorsque Δ_+U_{i-1} et Δ_+U_i ne

sont pas de même signe, la demi-pente δ_i^- est mise à zero (extrêmum locale en x_i). Donc : $\alpha_i^- \geq 0$, et

$(\alpha_i^- + \alpha_i^+) \geq 0$.

On démontre de la même façon que $(\beta_i^- + \beta_i^+) \geq 0$. Ceci termine la démonstration du lemme 1.1.

Posons pour la suite

$$v_{i+1/2} = \lambda_x \frac{F(U_{i+1/2,+}) - F(U_{i+1/2,-})}{U_{i+1/2,+} - U_{i+1/2,-}}$$

$$v_i = \lambda_x \frac{F(U_{i+1/2,-}) - F(U_{i-1/2,+})}{U_{i+1/2,-} - U_{i-1/2,+}}$$

$$Q_{i+1/2} = Q(U_{i+1/2,-}, U_{i+1/2,+})$$

$$Q_i = Q(U_{i-1/2,+}, U_{i+1/2,-})$$

$$\nu = \max_j (|v_i|, |v_{i+1/2}|) \tag{1.30}$$

$$\mu = \max_j (|Q_i|, |Q_{i+1/2}|) \tag{1.31}$$

Lemme 1.2.

Si ν et μ définis par (1.30) et (1.31) vérifient

$$\mu + (a+b)(\nu+\mu) \leq 1 \tag{1.32}$$

alors

$$E_{i+1/2} = C_{i+1/2} + D_{i+1/2} \leq 1 \tag{1.33}$$

Démonstration du lemme 1.2

En utilisant la définition des termes $C_{i+1/2}$ et $D_{i+1/2}$, on obtient

$$E_{i+1/2} = Q_{i+1/2}\left(1 - \frac{1}{2}(\beta_i^- + \alpha_i^+)\right) + \frac{1}{4}(Q_i - v_i)(\alpha_i^- + \alpha_i^+)$$

$$+ \frac{1}{4}(Q_{i+1} + v_{i+1})(\beta_i^- + \beta_i^+)$$

or d'après le lemme 1.1, on a

$$\left(1 - \frac{1}{2}(\beta_i^- + \alpha_i^+)\right) \geq 0, \qquad (\alpha_i^- + \alpha_i^+) \geq 0 \qquad \text{et} \qquad (\beta_i^- + \beta_i^+) \geq 0.$$

ce qui permet d'obtenir la majoration suivante pour $E_{i+1/2}$:

$$E_{i+1/2} \leq \mu\left[1 - \frac{1}{2}(\beta_i^- + \alpha_i^+)\right] + \frac{1}{4}(\nu+\mu)\left[(\alpha_i^- + \alpha_i^+) + (\beta_i^- + \beta_i^+)\right] \tag{1.34}$$

i) Majoration de : $1 - \frac{1}{2}(\beta_i^- + \alpha_i^+)$

On a : $0 \le \alpha_i^+$, et $0 \le \beta_i^-$, et donc

$$1 - \frac{1}{2}(\beta_i^- + \alpha_i^+) \le 1$$

ii) Majoration de : $(\alpha_i^- + \alpha_i^+)$

Le cas le plus défavorable est celui où $\frac{\Delta_+ U_i}{\Delta_+ U_{i-1}} > 0$ (même signe). On peut supposer, dans ce cas, sans perte de généralité, que $\Delta_+ U_i > 0$ et $\Delta_+ U_{i-1} > 0$. On distingue alors deux cas :

a) cas n°1 : $\Delta_+ U_i / \Delta_+ U_{i-1} > 1$: (i.e. $\Delta_+ U_i > \Delta_+ U_{i-1} > 0$)

alors, nécessairement : $\tilde{\delta}^i = \frac{U^{i+1} - U^{i-1}}{2} = \frac{\Delta_+ U_{i-1} + \Delta_+ U_i}{2} > 0$

ce qui donne pour δ_i^- et δ_i^+

$$\delta_i^- = a\,\Delta_+ U_{i-1}, \qquad \delta_i^+ = \tilde{\delta}^i$$

et alors $$(\alpha_i^- + \alpha_i^+) = \frac{1}{2}\left(1 + \frac{\Delta_+ U_{i-1}}{\Delta_+ U_i}\right) + a\,\frac{\Delta_+ U_{i-1}}{\Delta_+ U_i} \le 1+a$$

b) cas n°2 : $\Delta_+ U_i / \Delta_+ U_{i-1} < 1$: (i.e. $\Delta_+ U_i < \tilde{\delta}^i < \Delta_+ U_{i-1}$)

$$\alpha_i^+ = a \qquad \alpha_i^- \le b$$

et donc

$$(\alpha_i^- + \alpha_i^+) \le a+b$$

On démontre de la même façon que $(\beta_i^- + \beta_i^+) \le a+b$.

En utilisant ces majorations dans (1.34), on obtient

$$E_{i+1/2} \le \mu + \frac{1}{2}(\nu+\mu)(a+b)$$

Ceci termine la démonstration du lemme 1.2.

Remarque 1.3.

La limitation (1.15) pour laquelle l'estimation TVD a été démontrée n'est pas exactement celle présentée dans [Fez], et n'est donc pas une vraie "limitation locale". On peut montrer facilement à l'aide d'un contre exemple (cas linéaire) que le schéma obtenu à l'aide de la limitation de [Fez] n'est L^∞-stable. C'est pour cette raison que l'on a introduit le terme pondéré par le coefficient b dans la définition (1.15) pour assurer la stabilité TVD et par la suite la stabilité L^∞ et la convergence du schéma.

La limitation (1.15) permet néanmoins d'améliorer le résultat par-rapport à la limitation classique si l'on choisit le coefficient b suffisemment grand (a<b).

CHAPITRE VI

PRODUCTION D'ENTROPIE DANS UN E-SCHEMA

APPLICATION A LA CONVERGENCE DE LA METHODE DES DIFFERENCES FINIES

1. Introduction

Let Ω be a bounded open set of $\mathbb{R}^2$ with smooth boundary $\Gamma = \partial\Omega$, and outward unit normal n. Consider for $u : \Omega x\mathbb{R}_+ \to \mathbb{R}$, the conservation law

$$u_t + f_1(u)_{x_1} + f_2(u)_{x_2} = 0 \tag{1.1}$$

with initial condition

$$u(0,x) = u_o(x) \tag{1.2}$$

and boundary condition $\forall\, k \in \mathbb{R}$, $(x,t) \in \Gamma x\mathbb{R}_+$

$$(\mathrm{sgn}(u(x,t) - k) - \mathrm{sgn}(a(x,t) - k))\,(f(u(x,t) - f(k))\,.\,n(x) \geq 0 \tag{1.3}$$

where $f = (f_1,f_2)$ is smooth function : $\mathbb{R} \to \mathbb{R}^2$, and u_o is in $L^\infty(\Omega)$. $a : \Gamma x\mathbb{R}_+ \to \mathbb{R}$ is a given smooth function and the function sgn : $\mathbb{R} \to \mathbb{R}$ is defined by

$$\mathrm{sgn}(x) = \begin{cases} x/|x|, & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Existence and uniqueness are proved for BV-solutions of (1.1)-(1.3) in [BLN]. We are dealing in this paper with the convergence of the approximation of the weak entopy solution of this problem by explicite finite difference **E-schemes** (see [Tad]).

Leroux [Ler] has proved the first result of convergence of approximate solutions to (1.1-1.3) constructed by the Godounov's and Lax-Friedrichs's finite difference scheme; his approch is based on the so-called BV estimate, which provides the strong L1-convergence, thanks to the Helly's compactness theorem.

A new technique based on the theory of measure valued solution for conservation laws introduced by Di Perna [DP.2], is used for proving the convergence of finite difference schemes in several space variables, without assuming a BV-estimate.

This theory has been used by Szepessy [Sze] to prove the convergence of a scheme based on finite element method (Stream-Line diffusion).Using the same technique, Coquel and Lefloch proved the convergence of a finite difference scheme for a conservation law with a strictly convex continuous flux.

This paper is organized as follows : Section two is devoted to some preliminaries on the theory of measure valued solutions and to a uniqueness theorem of the measure valued solution of (1.1-1.3). The result of convergence is established in the third section.

2. A uniqueness result for measure valued solutions

2.0. Introduction

In this section, we study the solutions of the scalar conservation laws with initial and boundary conditions (1.1-1.3). A uniqueness theorem for such solutions is proved which is analogous to a uniqueness theorem for measure-valued solutions to the pure initial value problem by Gallouët and Herbin [GH].

2.1. Preliminaries on measure-valued solutions to conservation laws

In this section, we recall some notions related to the measure valued solutions for hyperbolic conservation laws. A Young measure ν defined on $\Omega x\mathbb{R}_+$ is an application $\Omega x\mathbb{R}_+ \to \text{Prob}(\mathbb{R})$, such that for any function a in $C(\mathbb{R})$, the application $(x,t)\to\langle\nu_{x,t},a\rangle$ is in $L^\infty(\Omega x\mathbb{R}_+)$,

with $$\langle\nu_{x,t},a(\lambda)\rangle = \int_{\mathbb{R}} a(\lambda)\, d\nu_{x,t}(\lambda) = \mu_a(x,t) \qquad \forall\, (x,t) \in \Omega x\mathbb{R}_+$$

Let $\{u^h\}$, $u^h \in L^\infty(\Omega x\mathbb{R}_+)$ be a uniformly bounded sequence, i.e.

$$\| u^h \|_{L^\infty(\Omega x\mathbb{R}+)} \le K \tag{2.1}$$

(in the applications the u^h will be approximate solutions to (1.1-1.3)). Then there exists according to Young's theorem, cf [DP.2], [Tar], a subsequence which we still label $\{u^h\}$ and an associated Young measure $\nu_{(x,t)} : \Omega x\mathbb{R}_+ \to \text{Prob}(\mathbb{R})$ such that

$$\{\lambda : |\lambda| \le K\} \supset \text{supp}(\nu_{(x,t)}) \qquad \text{for a.e. } (x,t) \text{ in } \Omega x\mathbb{R}_+ \tag{2.2}$$

and for all g in $C(\mathbb{R})$ the $L^\infty(\Omega x\mathbb{R}_+)$ weak star limit

$$g(u^h(.)) \to \mu_g(.) \qquad \text{as} \quad h \to 0 \tag{2.3}$$

exists. Here $\text{Prob}(\mathbb{R})$ is the space of all positive Borel measures on $\mathbb{R}$ with unit total mass.

To a given Young measure $\nu : \Omega x\mathbb{R}_+ \to \text{Prob}(\mathbb{R})$ satisfying (2.2) we shall associate (see [Sze]), in general in a non-unique way, a Young measure $\gamma\nu_{(.)} : \Gamma x\mathbb{R}_+ \to \text{Prob}(\mathbb{R})$ which we consider to be a "trace" of ν on $\Gamma x\mathbb{R}_+$. For this purpose we introduce, like in [Sze], the change of coordinates $x \to (x_b,y)$ for x in a neighborhood of Γ : $x_b = x - y n(x_b)$, where $(x_b,y) \in \Gamma x(0,\varepsilon)$, for some $\varepsilon>0$.

lemma 2.1. [Sze]

If $\nu : \Omega x R_+ \rightarrow$ Prob(R) is a Young measure associated to the sequence $\{u^h\}$, then there is a sequence $\{y_h \in (0,\varepsilon)\}$ where $y_h \rightarrow 0$ and there is a measurable Young measure $\gamma\nu : \Gamma x R_+ \rightarrow$ Prob(R) such that

$$\{\lambda : |\lambda| \leq K\} \supset \text{supp}(\gamma\nu_{(x_b,t)}) \qquad \text{for a.e. } (x,t) \text{ in } \Gamma x R_+$$

and for every g in C(R) the $L^\infty(\Gamma x R_+)$ weak star limit

$$\langle \nu_{(x(.,y_h),.)}, g(\lambda)\rangle \rightarrow \mu_g(.) \qquad \text{as } h \rightarrow 0$$

exists, i.e.

$$\lim_{h\rightarrow 0} \int_{\Gamma x R_+} \langle \nu_{(x(.,y_h),.)}, g(\lambda)\rangle \varphi \, ds \, dt = \int_{\Gamma x R_+} \gamma\mu_g(x_b,t) \varphi \, ds \, dt \qquad (2.4)$$

$\forall \varphi \in L_1(\Gamma x R_+)$, where ds is the Lebesgue measure on Γ and

$$\gamma\mu_g(x_b,t) = \int_R g(\lambda) \, d\gamma\nu_{x_b,t}(\lambda) = \langle \gamma\nu_{x_b,t}, g(\lambda)\rangle \qquad (2.5)$$

for a.e. (x_b,t) in $(\Gamma x R_+)$.

We may now introduce the definition of a measure-valued solution to (1.1-1.3).

Définition. 2.1.

A Young measure $\nu : \Omega x R_+ \rightarrow$ Prob(R) is a mv-solution to problem (1.1-1.3) if $\forall \phi \in C_o^1(\bar{\Omega} x [0,+\infty[)$, $\phi \geq 0$, $\forall k \in R$,

$$\int_{\Omega x R_+} \left\{ \langle \nu_{(x,t)}, |\lambda - k| \rangle \phi_t + \langle \nu_{(x,t)}, [f(\lambda) - f(k)] \, \text{sgn}(\lambda - k) \rangle \nabla\phi \right\} dx \, dt$$

$$- \int_{\Gamma x R_+} \langle \gamma\nu_{x_b,t}, f(\lambda) - f(k) \rangle . n(x_b) \, \phi \, \text{sgn}(a-k) \, ds \, dt$$

$$+ \int_\Omega \langle \nu_{(x,0)}, |\lambda - k| \rangle \phi(x,0) dx \geq 0 \qquad (2.6)$$

where $\gamma\nu$ is associated to ν through lemma 2.1.

Remark 2.1.

In general lemma 2.1. associates to ν a trace $\gamma\nu$ in a non-unique way. However, in the proof of lemma 2.1. Szepessy shows that the expected value $\langle \gamma\nu_{x,t}, f(\lambda)\rangle$ in uniquely defined.

In [BLN], existence and uniqueness are proved for BV-solutions. For u in $BV^{loc}(\Omega x R_+)$ to be a BV-solution of (1.1-1.3) means precisely that the Dirac measure concentrated at u is a mv-solution, i.e. that $\nu_{x,t} = \delta_{u(x,t)}$ and $\gamma\nu_{x,t} = \delta_{\gamma u(x,t)}$ satisfies (2.6) , where γu denotes the trace of u on $\Gamma x R_+$. This proves also the existence of a mv-solution to (1.1-1.3).

We have the following uniqueness result for mv-solutions whose proof is essentially based on a uniqueness theorem for measure-valued solutions to the pure initial value problem [GH].

Theorem. 2.1.

Let the initial data u_o be in $L^\infty(\Omega)$. Assume that $\nu : \Omega x R_+ \to Prob(R)$ is a mv-solution to (1.1-1.3). Then if u is the unique measure valued solution to the boundary value problem (1.1-1.3), we have :

$$\nu_{(x,t)} = \delta_{u(x,t)}, \quad \text{for every} \quad (x,t) \in \Omega x R_+$$

Remark 2.2.

Note that Szepessy [Sze] had already proved a uniqueness result on the entropy measure valued solution to (1.1-1.3), with stronger assumption on ν, in particular with ν satisfying a "strong consistency" with the initial condition :

$$\lim_{t \to 0} = \int_\Omega \langle \nu_{(x,t)}, |\lambda - u_o| \rangle \, dx = 0 \tag{2.7}$$

proof of theorem 2.1.

We just have to notice that if in (2.6), we take the test function ϕ in $C_o^1(\bar{\Omega} x]0,+\infty[)$, we get the definition of a mv-solution given by Szepessy in [Sze] without the strong consistency with the initial condition (2.7). In [Sze], it is proved that if a Young measure ν satisfies (2.6) with ϕ in $C_o^1(\bar{\Omega} x]0,+\infty[)$ and relation (2.7) then $\nu_{(x,t)} = \delta_{u(x,t)}$, where u is the unique BV-solution of (1.1-1.3). Thus it remains to prove that relation (2.6) implies (2.7). For this purpose we write (2.6) with ϕ in $C_o^1(\Omega x [0,+\infty[)$, this leads to

$$\int_{\Omega x R_+} \left\{ \langle \nu_{(x,t)}, |\lambda - k| \rangle \phi_t + \langle \nu_{(x,t)}, [f(\lambda) - f(k)] \, sgn(\lambda - k) \rangle \nabla \phi \right\} dx \, dt$$

$$+ \int_\Omega \langle \nu_{(x,0)}, |\lambda - k| \rangle \phi(x,0) dx \geq 0 \tag{2.8}$$

which is the definition of a mv-solution given by Gallouët and Herbin in [GH], to prove the existence and uniqueness of a mv-solution of problem (1.1)-(1.2). Using the theorem in [GH] we get (2.7). This ends the proof of theorem 2.1.

Theorem. 2.2.

Let $\{u^h\}$, $u^h \in L^\infty(\Omega x R_+)$ be a uniformly bounded sequence of approximate solutions to (1.1-1.3)). If the Young measure ν associated with $\{u^h\}$ and defined on $\Omega x R_+$ and its trace $\gamma\nu$ defined on $\Gamma x R_+$ satisfie relation (2.6) then $\{u^h\}$ tends to u in $L^1(\Omega x R_+)$, where u is the unique entropy weak solution to (1.1-1.3).

Proof of theorem. 2.2.

The proof of this theorem is essentially based on the result of uniqueness of the measure valued solution of problem (1.1-1.2) and which is due to Di Perna [DP.1]. In the case of the pure initial value problem (1.1-1.2), the proof is detailed in [GH].

3. Convergence of the finite difference method

3.1. Description of the scheme

In this section, we describe the explicite finite difference scheme. For simplicity we consider a one dimensional problem, i.e. let $\Omega =]0,1[$ and $\Gamma = \{0,1\}$. We approximate the following problem

$$u_t + f(u)_x = 0 \tag{3.1}$$

with initial condition

$$u(0,x) = u_o(x) \qquad \text{for} \quad x \text{ in } \Omega \tag{3.2}$$

and boundary condition

$$\sup_{k \in I(\gamma u(0,t),\, a(t))} \left\{ \operatorname{sgn}\left(\gamma u(0,t) - a(t)\right) \left(f(\gamma u(0,t)) - f(k)\right) \right\} = 0 \tag{3.3.a}$$

$$\inf_{k \in I(\gamma u(1,t),\, b(t))} \left\{ \operatorname{sgn}\left(\gamma u(1,t) - b(t)\right) \left(f(\gamma u(1,t)) - f(k)\right) \right\} = 0 \tag{3.3.b}$$

where γu denotes the trace of u on $\Gamma \times \mathbb{R}_+$, f is a smooth function : $\mathbb{R} \to \mathbb{R}$, and u_o is in $L^\infty(\Omega)$. a, b : $\mathbb{R}_+ \to \mathbb{R}$ are given smooth functions.

Remark 3.1.

in the one dimensional case, the boundary condition (1.3) is equivalent to the one given by (3.3).

3.1.1. Notations

We introduce some notations related to the mesh. Let τ, h be the time, x-space increment. The ratio $\lambda = (\tau/h)$ will be kept constant and should satisfy a Courant-Friedrichs-Levy (CFL) condition. Let N be an integer such that : $N = 1/h$, and set : $I_N = \{1,2, ..., N\}$. We define a regular grid by setting :

$$t = n\tau \quad (n \in \mathbb{N}), \quad x_i = i\,h, \quad (i \in I_N \text{ or } I_{N+1/2})$$

$$I_{n,i} = [t_n, t_{n+1}[\times [x_{i-1/2}, x_{i+1/2}[$$

The approximate solution $u^h : \Omega \times \mathbb{R}_+ \longrightarrow \mathbb{R}$, to problem (3.1-3.3), is a piecewise constant function satisfying :

$$u^h(x,t) = u_i^n, \quad \text{for} \quad t \in [t_n, t_{n+1}[, \; x \in [x_{i-1/2}, x_{i+1/2}[\qquad \text{for } i > 1$$

For $t = 0$, (u_i^o) is defined from the initial data u_o by :

$$u_i^o = \frac{1}{h} \int_{x_{i-1/2}}^{x_{i+1/2}} u_o(x,y)\, dxdy \tag{3.4.a}$$

Then, for each integer n, the sequence u_i^n is given by the following 3-points explicit finite difference scheme :

$$u_i^{n+1} = u_i^n - \lambda(f_{i+1/2}^n - f_{i-1/2}^n) \tag{3.4.b}$$

$$u_0^{n+1} = \frac{1}{\tau}\int_{J_n} a(t)\,dt \tag{3.4.c}$$

$$u_{N+1}^{n+1} = \frac{1}{\tau}\int_{J_n} b(t)\,dt \tag{3.4.d}$$

with $$f_{i+1/2}^n = f(u_i^n, u_{i+1}^n) \tag{3.5}$$

where f is a numerical flux associated with an E-scheme [Tad], consistent with the exact flux f, i.e.

$$f(u,u) = f(u) \tag{3.6}$$

It is well known that the class of E-schemes are entropy satisfying, under a Courant-Friedrichs-Levy (CFL) stability condition of the form, (see [Tad]) :

$$\lambda \operatorname*{Sup}_u \left|\frac{df(u)}{du}\right| \le \frac{1}{2} \tag{3.7}$$

That is, for all k in R, we have the following discrete entropy inequality

$$|u_i^{n+1} - k| - |u_i^n - k| + \lambda\left[\Pi^E(u_i^n, u_{i+1}^n) - \Pi^E(u_{i-1}^n, u_i^n)\right] \le 0 \tag{3.8}$$

where Π^E is the numerical entropy flux associated to the entropy (η_k, F_k),

$$\eta_k(u) = |u-k|, \qquad F_k(u) = (f(u) - (f(k))\,\mathrm{sgn}(u-k) \tag{3.9}$$

consistent with F_k. Following Tadmor [Tad], Π^E is a local convex combination of the numerical entropy fluxes associated respectively to the Godounov and the Modified-Lax-Friedrichs schemes : Π^G and Π^M, given by

$$\Pi^G(u,v) = F_k(w(0; u,v)) \tag{3.10.a}$$

$$\Pi^M(u,v) = \frac{F_k(v)+F_k(u)}{2} - \frac{1}{4\lambda}(\eta_k(v) - \eta_k(u)) \tag{3.10.b}$$

where w(0; u,v) is the solution at x=0, of the following Riemann Problem, PR(u,v) :

$$\begin{cases} w_t + f(w)_x = 0 \\ w_o(x) = \begin{cases} u & \text{if} \quad x<0 \\ v & \text{if} \quad x>0 \end{cases} \end{cases}$$

w(0; u,v) has also the following characterization : $w(0; u,v) \in I(u,v)$ such that

$$\mathrm{sgn}(v-u)\, f(w(0; u,v)) = \min_{k \in I(u,v)} \left\{\mathrm{sgn}(v-u)\, f(k)\right\} \tag{3.11}$$

With theses notations $\Pi^E(u,v)$ is given by

$$\Pi^E(u,v) = \theta(u,v)\,\Pi^M(u,v) + (1-\theta(u,v))\,\Pi^G(u,v) \tag{3.12}$$

where $\theta(u,v) \in [0,1]$.

3.2. Convergence

Let us assume that the family of approximate solutions u^h is in $L^\infty(\Omega \times \mathbb{R}_+)$, i.e.

$$\| u^h \|_{L^\infty(\Omega \times \mathbb{R}_+)} \leq K_1 \tag{3.13}$$

we may construct (see [Tar]) a non-unique Young measure ν associated with the family $\{u^h\}$, as follows : for each continuous function $g : \mathbb{R} \to \mathbb{R}$, we have :

$$\langle \nu, g \rangle = \lim_{h \to 0} g(u^h) \qquad \text{the limit is taken in } L^\infty \text{ weak-star sense} \tag{3.14}$$

Let us now present a theorem of convergence, which allows us to apply Theorem 2.1. and deduce the strong L1-convergence of $\{u^h\}$ to the unique weak entropy satisfying solution of (3.1-3.3).

Theorem 3.1.

Assume that the numerical flux f is consistent with equation (3.1), and the family of approximate solutions $\{u^h\}$ defined by the finite difference scheme (3.4)-(3.5), satisfies the following properties :

i) the uniform L^∞ bound (3.13),

ii) the following "weak estimate of the space derivatives" depending on the **numerical viscosity** of the scheme :

$$h \sum_{n \leq T} \sum_{1 \leq i \leq N-1} Q^2(u_i^n, u_{i+1}^n) \cdot |u_{i+1}^n - u_i^n|^2 \leq K \tag{3.15}$$

where T and K are positive constants independent of h, and Q(u,v) is the numerical viscosity of the scheme defined by :

$$Q(u,v) = \frac{f(u)+f(v) - 2f(u,v)}{v - u} \tag{3.16}$$

Then, the Young-measure ν associated with $\{u^h\}$ is a mv-solution to (3.1-3.3).

Let us introduce the following notations

$$C(u,v) = \frac{f(u) - f(u,v)}{v - u} \tag{3.17}$$

$$D(u,v) = \frac{f(v) - f(u,v)}{v - u} \tag{3.18}$$

and

$$\Delta_+ u_i^n = u_{i+1}^n - u_i^n \tag{3.19}$$

where C and D denote the incremental coefficients of the scheme (3.4.a).

To prove theorem 3.1., we need the following lemmas :

Lemma 3.1.

Assume that the numerical fluxes f is consistent with equation (3.1). If the family $\{u^h\}$ defined by the finite difference scheme (3.4)-(3.5), satisfies the uniform L^∞ bound (3.13). Then we have :

$$\Pi(u_i^n,u_{i+1}^n) - F_k(u_i^n) \le - \mathrm{sgn}(u_i^n-k)\, C(u_i^n,u_{i+1}^n)\, \Delta_+ u_i^n \tag{3.20}$$

and

$$F_k(u_i^n) - \Pi(u_{i-1}^n,u_i^n) \le \mathrm{sgn}(u_i^n-k)\, D(u_{i-1}^n,u_i^n)\, \Delta_+ u_{i-1}^n \tag{3.21}$$

Proof of lemma 3.1.

Let us consider the following one dimensional scheme :

$$u_{i+1/2}^{n+1} = u_i^n - \lambda \left(f(u_i^n,u_{i+1}^n) - f(u_i^n,u_i^n)\right) \tag{3.22}$$

which, thanks to the consistency of the numerical flux f with the continuous flux f, may be written as :

$$u_{i+1/2}^{n+1} = u_i^n - \lambda \left(f(u_i^n,u_{i+1}^n) - f(u_i^n)\right) \tag{3.23}$$

It is well known that if the numerical flux f is associated with an E-scheme (see Tadmor [Tad]), then, under a Courant-Friedrichs-Levy (CFL) stability condition of the form (3.7), it is entropy satisfying. That is, we have the following discret entropy inequality satisfied by the numerical entropy flux Π^E :

$$\eta_\kappa(u_{i+1/2}^{n+1}) - \eta_\kappa(u_i^n) + \lambda\left(\Pi^E(u_i^n,u_{i+1}^n) - \Pi^E(u_i^n,u_i^n)\right) \le 0$$

which, thanks to the consistency of the numerical entropy flux Π^E with the continuous flux F_k, may be written as :

$$\lambda\left(\Pi^E(u_i^n,u_{i+1}^n) - F_k(u_i^n)\right) \le \eta_\kappa(u_i^n) - \eta_\kappa(u_{i+1/2}^{n+1}) \tag{3.24}$$

We use a classical result whose proof is very easy

$$\eta_\kappa(v) - \eta_\kappa(u) \le \mathrm{sgn}(v-k)\,(v-u) \qquad \forall\, u,v \in \mathbb{R}, \tag{3.25}$$

to obtain the following inequality

$$\lambda\left(\Pi^E(u_i^n,u_{i+1}^n) - F_k(u_i^n)\right) \le \mathrm{sgn}(u_i^n-k)\,(u_i^n-u_{i+1/2}^{n+1}) \tag{3.26}$$

By using relation (3.23), we may replace $(u_i^n-u_{i+1/2}^{n+1})$ in (3.26) by $\lambda\left(f(u_i^n,u_{i+1}^n) - f(u_i^n)\right)$ and thus obtain the desired relation (3.20). We similarly prove (3.21). This ends the proof of lemma 3.1.

lemma 3.2.

We suppose that $\{u^h\}$ satisfies the stability condition (3.13). Let $\{W_{1/2}^h\}$ be a sequences of functions associated to $\{u^h\}$ and defined by

$$W_{1/2}^h(t) = \theta(u_o^n,u_1^n)\, f(u_{1/2}^n) + (1-\theta(u_o^n,u_1^n))\left\{f(u_o^n) + f(u_1^n) - \frac{1}{2\lambda}(u_1^n - u_o^n)\right\} \qquad \text{if } t \in [t_n, t_{n+1}[\tag{3.27}$$

where $\theta : \mathbb{R}^2 \to [0,1]$. Then the $L^\infty(\mathbb{R}_+)$ weak star limit of $W_{1/2}^h(t)$ exists and

$$\lim_{h\to 0} \int_{\mathbb{R}_+} W_{1/2}^h(t)\, \psi(t)\, dt = \int_{\mathbb{R}_+} \langle \nu_{(0,t)}, f(\lambda)\rangle\, \psi(t)\, dt \qquad \forall\, \psi \in L_1(\mathbb{R}_+)$$

proof of lemma 3.2.

Let φ be a positive function in $C^1(\mathbb{R}^2 x \mathbb{R}_+)$ with compact support in $[0,1[x]0,+\infty[$, defined by

$$\varphi(x,t) = \rho_\delta(x)\, \psi(t)$$

where ρ_δ is given by (see [Ler])

$$\rho_\delta(x) = \begin{cases} 0 & \text{if} \quad x \geq \delta \\ \leq 1 & \text{if} \quad 0 \leq x < \delta \\ 1 & \text{if} \quad x = 0 \end{cases}$$

and : $\sup_{0\leq x\leq\delta} |\rho_\delta'(x)| \leq C/\delta$, where C is a constant independent of h and δ.

we set $\varphi_i^n = \varphi(x_i, t_n)$

After multiplication by "$\tau h\, \varphi_i^n$" and summation on all n in $\mathbb{N}$ and i in I_N, the entropy inequality (3.8) :

$$\frac{1}{\tau}\sum_{n,i} \left[|u_i^{n+1} - k| - |u_i^n - k|\right] \varphi_i^n\, \tau h + \frac{1}{h}\sum_{n,i} \left[\Pi^E(u_i^n, u_{i+1}^n) - \Pi^E(u_{i-1}^n, u_i^n)\right] \varphi_i^n\, \tau h \leq 0 \qquad (3.28)$$

We define the function $\varphi^h : \mathbb{R}^2 x \mathbb{R}_+ \to \mathbb{R}$, by

$$\varphi^h(t,x) = \varphi_i^n \quad , \quad \text{for} \quad (x,t) \in I_{n,i}$$

and for the sake of simplicity, we introduce the functions : $\Delta_\tau \varphi^h$, $\Delta_x \varphi^h$, defined by :

$$\Delta_\tau \varphi^h = \frac{1}{\tau}\left(\varphi^h(x,t+\tau) - \varphi^h(x,t)\right)$$

and

$$\Delta_x \varphi^h = \frac{1}{h}\left(\varphi^h(x+h/2,t) - \varphi^h(x-h/2,t)\right)$$

Integrating by parts, the first term of (3.28) may be equivalently written as :

$$\frac{1}{\tau}\sum_{n,i} \left[|u_i^{n+1} - k| - |u_i^n - k|\right] \varphi_i^n\, \tau h = \frac{1}{\tau}\sum_{n,i} [\varphi_i^{n-1} - \varphi_i^n] \left|u_i^n - k\right| \tau h$$

$$= -\int_{\Omega x \mathbb{R}_+} |u^h(x,t) - k|\, \Delta_\tau \varphi^h \, dt\, dx \qquad (3.29)$$

We now treat of the second sum in (3.28). Let us first introduce, the following notations :

$$\Pi_{i+1/2}^n = \Pi^E(u_i^n, u_{i+1}^n), \qquad n \in \mathbb{N},\ 0 \leq i \leq N,$$

$$\varphi_{i+1/2}^n = \varphi^h(x_{i+1/2}, t_n), \qquad n \in \mathbb{N},\ 0 \leq i \leq N$$

and

$$S_1 = \frac{1}{h}\sum_{n,i} \left[\Pi^E(u_i^n, u_{i+1}^n) - \Pi^E(u_{i-1}^n, u_i^n)\right] \varphi_i^n\, \tau h$$

Integrating by parts, we find

$$S_1 = \frac{1}{h}\sum_{n,i} \left\{[\varphi_i^n - \varphi_{i+1/2}^n]\, \Pi_{i+1/2}^n - [\varphi_i^n - \varphi_{i-1/2}^n]\, \Pi_{i-1/2}^n\right\} \tau h - \frac{1}{h}\sum_n \Pi_{1/2}^n\, \varphi_{1/2}^n$$

then, writing

$$\Pi_{i+1/2}^n = F_k(u_i^n) + [\Pi_{i+1/2}^n - F_k(u_i^n)] \qquad \text{and} \qquad \Pi_{i-1/2}^n = F_k(u_i^n) + [\Pi_{i-1/2}^n - F_k(u_i^n)]$$

we get for S_1

$$S_1 = \frac{1}{h}\sum_{n,i} F_k(u_i^n)\,[\varphi_{i-1/2}^n - \varphi_{i+1/2}^n]\,\tau\,h$$

$$+\frac{1}{h}\sum_{n,i}[\Pi_{i+1/2}^n - F_k(u_i^n)][\varphi_i^n - \varphi_{i+1/2}^n]\,\tau\,h + \frac{1}{h}\sum_{n,i}[\Pi_{i-1/2}^n - F_k(u_i^n)][\varphi_{i-1/2}^n - \varphi_i^n]\,\tau\,h$$

$$-\frac{1}{h}\sum_{n}\Pi_{1/2}^n\,\varphi_{1/2}^n\,\tau\,h$$

the first term in S_1 may be written as follows

$$\frac{1}{h}\sum_{n,i} F_k(u_i^n)\,[\varphi_{i-1/2}^n - \varphi_{i+1/2}^n]\,\tau\,h = -\int_{\Omega\times R_+} F_k(u^h(x,t)\,\Delta_x\varphi^h(x,t)\,dtdx \qquad (3.30)$$

we set

$$T_1 = -\frac{1}{h}\sum_{n}\Pi_{1/2}^n\,\varphi_{1/2}^n\,\tau\,h,$$

Let us transform T_1. For this purpose we know that

$$\Pi_{1/2}^{n,E} = \theta_{1/2}^n\,\Pi_{1/2}^{n,G} + (1-\theta_{1/2}^n)\,\Pi_{1/2}^{n,M}$$

where we have set $\theta_{1/2}^n = \theta(u_o^n, u_1^n) \in [0,1]$, $(F_k(u),\eta_k(u))$ are given by formulae (3.9). Thus we get for T_1 :

$$T_1 = -\frac{1}{h}\sum_{n}\left\{\theta_{1/2}^n\,\Pi_{1/2}^{n,G} + (1-\theta_{1/2}^n)\,\Pi_{1/2}^{n,M}\right\}\varphi_{1/2}^n\,\tau\,h$$

To estimate T_1, we use the following inequalities, whose proof is easy ([Ler]) by using characterization (3.11) of $u_{1/2}^n$ for the first one and thanks to the stability condition (3.7) for the second one

$$\Pi_{1/2}^{n,G} \le \mathrm{sgn}(u_o^n-k)\,\{f(u_{1/2}^n) - f(k)\} \qquad \forall k \in R \qquad (3.31.a)$$

$$\Pi_{1/2}^{n,M} \le \frac{1}{2}\,\mathrm{sgn}(u_o^n - k)\,\{f(u_o^n) - 2f(k) + f(u_1^n) - \frac{1}{2\lambda}(u_1^n - u_o^n)\} \qquad \forall k \in R \qquad (3.31.b)$$

We deduce from (3.31) the following inequality

$$T_1 \ge -\int_{R_+}\mathrm{sgn}(u^h(0,t)-k)\,(W_{1/2}^h(t) - f(k))\,\varphi^h(0,t)dt \qquad (3.32)$$

Combining relations (3.29), (3.30) and (3.32) we may write inequality (3.28) as follows

$$\int_{\Omega\times R_+}\{\eta_k(u^h)\,\Delta_t\varphi^h + F_k(u^h)\,\Delta_x\varphi^h\}\,dtdx + \int_{R_+}\mathrm{sgn}(u^h(0,t)-k)\,(W_{1/2}^h(t) - f(k))\,\varphi^h(0,t)dt \ge R_1^h \qquad (3.33)$$

with

$$R_1^h = -\frac{1}{h}\sum_{n,i}[\Pi_{i+1/2}^n - F_k(u_i^n)][\varphi_{i+1/2}^n - \varphi_i^n]\,\tau\,h + \frac{1}{h}\sum_{n,i}[F_k(u_i^n) - \Pi_{i-1/2}^n][\varphi_{i-1/2}^n - \varphi_i^n]\,\tau\,h$$

If we pass to the limit on h in (3.33), the first term of the left hand side tends to

$$\int_{\Omega\times R_+}\left\{<\nu_{(x,t)},\eta_k(\lambda)>\,\psi'(t)\rho_\delta(x) + <\nu_{(x,t)},F_k(\lambda)>\,\psi(t)\rho_\delta'(x)\right\}dx\,dt$$

$$+\int_{R_+}\mathrm{sgn}(a(t)-k)\,(W_{1/2}(t) - f(k))\,\psi(t)\,dt \qquad (3.34)$$

where $W_{1/2}(t)$ denotes the L^∞ weak star limit of $W^h_{1/2}(t)$. This limit exists since we have $u^n_{1/2} \in I(u^n_o,u^n_1)$ and that the sequence $\{u^h\}$ satisfies (3.13). (with : $I(a,b) = [\min(a,b), \max(a,b)]$).

Using the weak estimate (3.15) of theorem 3.1. we get that the the term R^h_1 tends to zero with h (see proof of theorem 3.1). If we pass to the limit on δ in (3.34), we get

$$\lim_{\delta\to 0} \int_{\Omega\times R_+} <\nu_{(x,t)},\eta_k(\lambda)>\ \psi'(t)\rho_\delta(x)\,dx\,dt = 0$$

on the other hand, we have

$$\int_{\Omega\times R_+} <\nu_{(x,t)},F_k(\lambda)>\psi(t)\rho'_\delta(x)\,dx\,dt = \int_0^\delta \rho'_\delta(x)\,G_k(x)dx \qquad (3.35)$$

where $$G_k(x) = \int_{R_+} <\nu_{(x,t)},F_k(\lambda)>\psi(t)\,dt$$

set : $L_k = \lim_{x\to 0} G_k(x)$. By definition of the "trace" $\gamma\nu$ (see lemma 2.1.), we know that

$$L_k = \int_{R_+} <\gamma\nu_{(0,t)},F_k(\lambda)>\psi(t)\,dt$$

to determin the limit of (3.35) we write

$$\int_0^\delta \rho'_\delta(x)\,G_k(x)dx = \int_0^\delta \rho'_\delta(x)\,[G_k(x)-L_k]dx + \int_0^\delta \rho'_\delta(x)\,L_k dx$$

It is evident that the first term of the left hand side tends to 0 with δ. On the other hand we have

$$\int_0^\delta \rho'_\delta(x)\,dx = \rho_\delta(\delta) - \rho_\delta(0) = -1$$

and then

$$\lim_{\delta\to 0}\int_0^\delta \rho'_\delta(x)\,G_k(x)dx = -\int_{R_+} <\gamma\nu_{(0,t)},F_k(\lambda)>\psi(t)\,dt \qquad (3.36)$$

We finally obtain

$$\int_{R_+} <\gamma\nu_{(0,t)},\ \mathrm{sgn}(\lambda-k)(f(\lambda)-f(k))>\psi(t)\,dt \le \int_{R_+} \mathrm{sgn}(a(t)-k)\,(W_{1/2}(t) - f(k))\,\psi(t)\,dt \qquad \forall\, k\in\mathbb{R}$$

in particular for : $k \ge \mathrm{Max}(K_1, \sup_{k\in\mathrm{supp}(\psi)} a(t))$, and $k \le \mathrm{Min}(K_1, \min_{k\in\mathrm{supp}(\psi)} a(t))$, which implies

$$\int_{R_+} <\gamma\nu_{(t)},\ f(\lambda)>\psi(t)\,dt = \int_{R_+} W_{1/2}(t)\,\psi(t)\,dt$$

where $\mathrm{supp}(\psi)$ is the support of ψ. This ends the proof of lemma 3.2.

Proof of theorem 3.1.

Let φ be a positive function in $C^1(\mathbb{R}^2\times\mathbb{R}_+)$ with compact support in $\bar{\Omega}\times[0,+\infty[$. Set :

$$\varphi^n_i = \inf_{(x,t)\in I_{n,i}} \varphi(x_i,t_n) \qquad (3.37)$$

After multiplication by "$\tau h\, \varphi_i^n$" and summation on all $n \le [T/\tau]$ and i in I_N, inequality (3.8) gives :

$$\frac{1}{\tau}\sum_{n,i}\left[|u_i^{n+1}-k|-|u_i^n-k|\right]\varphi_i^n\,\tau h+\frac{1}{h}\sum_{n,i}\left[\Pi^E(u_i^n,u_{i+1}^n)-\Pi^E(u_{i-1}^n,u_i^n)\right]\varphi_i^n\,\tau h\le 0 \qquad (3.38)$$

We will use the same notations that in the proof of lemma 3.2.

Integrating by parts, the first term of (3.38) may be equivalently written as :

$$\frac{1}{\tau}\sum_{n,i}\left[|u_i^{n+1}-k|-|u_i^n-k|\right]\varphi_i^n\,\tau h=\frac{1}{\tau}\sum_{n,i}\left[\varphi_i^{n-1}-\varphi_i^n\right]|u_i^n-k|\,\tau h-\frac{1}{\tau}\sum_i\varphi_i^0\,|u_i^0-k|\,\tau h$$

$$=-\int_{\Omega\times R_+}|u^h(x,t)-k|\,\Delta_\tau\varphi^h\,dt\,dx-\int_R|u^h(x,0)-k|\,\varphi^h(x,0)dx \qquad (3.39)$$

We now treat of the second sum in (3.38). Integrating by parts, we find

$$S_1=\frac{1}{h}\sum_{n,i}\left\{\left[\varphi_i^n-\varphi_{i+1/2}^n\right]\Pi_{i+1/2}^{n,E}-\left[\varphi_i^n-\varphi_{i-1/2}^n\right]\Pi_{i-1/2}^{n,E}\right\}\tau h-\frac{1}{h}\sum_n\left\{\Pi_{1/2}^{n,E}\,\varphi_{1/2}^n-\Pi_{N+1/2}^{n,E}\,\varphi_{N+1/2}^n\right\}\tau h$$

then, writing

$$\Pi_{i+1/2}^n=F_k(u_i^n)+\left[\Pi_{i+1/2}^n-F_k(u_i^n)\right] \qquad \text{and} \qquad \Pi_{i-1/2}^n=F_k(u_i^n)+\left[\Pi_{i-1/2}^n-F_k(u_i^n)\right]$$

we get for S_1

$$S_1=\frac{1}{h}\sum_{n,i}F_k(u_i^n)\left[\varphi_{i-1/2}^n-\varphi_{i+1/2}^n\right]\tau h$$

$$+\frac{1}{h}\sum_{n,i}\left[\Pi_{i+1/2}^{n,E}-F_k(u_i^n)\right]\left[\varphi_i^n-\varphi_{i+1/2}^n\right]\tau h+\frac{1}{h}\sum_{n,i}\left[\Pi_{i-1/2}^{n,E}-F_k(u_i^n)\right]\left[\varphi_{i-1/2}^n-\varphi_i^n\right]\tau h$$

$$-\frac{1}{h}\sum_n\Pi_{1/2}^{n,E}\,\varphi_{1/2}^n\,\tau h+\frac{1}{h}\sum_{n,i}\Pi_{N+1/2}^{n,E}\,\varphi_{N+1/2}^n\,\tau h \qquad (3.40)$$

the first term of the right hand side of (3.40) may be written as follows

$$\frac{1}{h}\sum_{n,i}F_k(u_i^n)\left[\varphi_{i-1/2}^n-\varphi_{i+1/2}^n\right]\tau h=-\int_{\Omega\times R_+}F_k(u^h(x,t)\,\Delta_x\varphi^h(x,t)\,dtdx. \qquad (3.41)$$

Set

$$T_1=-\frac{1}{h}\sum_n\Pi_{1/2}^n\,\varphi_{1/2}^n\,\tau h, \qquad \text{and} \qquad T_2=\frac{1}{h}\sum_n\Pi_{N+/2}^n\,\varphi_{N+1/2}^n\,\tau h$$

Let us transform T_1. Using the same notations that in the proof of lemma 3.2., we get

$$T_1=-\frac{1}{h}\sum_n\left\{\theta_{1/2}^n\,\Pi_{1/2}^{n,G}+(1-\theta_{1/2}^n)\,\Pi_{1/2}^{n,M}\right\}\varphi_{1/2}^n\,\tau h$$

To estimate T_1, we use inequalities (3.31) and get

$$T_1\ge-\int_{R_+}\mathrm{sgn}(u^h(0,t)-k)\,\left(W_{1/2}^h(t)-f(k)\right)\varphi^h(0,t)dt \qquad (3.42)$$

where $W^h_{1/2}(t) : R_+ \to R$, is given by (3.27). We get similarly for T_2

$$T_2 \geq \int_{R_+} \operatorname{sgn}(u^h(1,t)-k)\ (W^h_{N+1/2}(t) - f(k))\ \varphi^h(1,t)dt \tag{3.43}$$

where $W^h_{N+1/2}(t) : R_+ \to R$, is given by

$$W^h_{N+1/2}(t) = \theta^n_{N+1/2}\, f(u^n_{N+1/2}) + (1-\theta^n_{N+1/2}) \left\{ f(u^n_{N+1}) + f(u^n_N) + \frac{1}{2\lambda}\,(u^n_{N+1} - u^n_N) \right\} \qquad \text{if } t \in [t_n, t_{n+1}[$$

Combining relations (3.39), (3.41), (3.42) and (3.43) we may write inequality (3.38) as follows

$$\begin{aligned} &\int_{\Omega x R_+} \left\{ \eta_k(u^h(x,t))\ \Delta_\tau \varphi^h + F_k(u^h(x,t)\ \Delta_x \varphi^h(x,t) \right\} dtdx \\ &+ \int_R \eta_k(u^h(x,0))\ \varphi^h(x,0)dx + \int_{R_+} \operatorname{sgn}(u^h(0,t)-k)\ (W^h_{1/2}(t) - f(k))\ \varphi^h(0,t)dt \\ &- \int_{R_+} \operatorname{sgn}(u^h(1,t)-k)\ (W^h_{N+1/2}(t) - f(k))\ \varphi^h(1,t)dt \quad \geq \quad R^h_1 \end{aligned} \tag{3.44}$$

with

$$R^h_1 = -\frac{1}{h} \sum_{n,i} \left[\Pi^n_{i+1/2} - F_k(u^n_i) \right] \left[\varphi^n_{i+1/2} - \varphi^n_i \right] \tau\, h + \frac{1}{h} \sum_{n,i} \left[F_k(u^n_i) - \Pi^n_{i-1/2} \right] \left[\varphi^n_{i-1/2} - \varphi^n_i \right] \tau\, h$$

We study separately each term of the left hand side of (3.44). We first write

$$\int_{\Omega x R_+} |u^h(x,t)-k|\ \Delta_\tau \varphi^h\ dtdx \ = \iint \eta_k(u^h) \frac{\partial \varphi}{\partial t}\ dt\, dx + \iint \eta_k(u^h) \left(\frac{\partial \varphi}{\partial t} - \Delta_\tau \varphi^h \right) dt\, dx$$

so that :

$$\begin{aligned} \left| \iint \left(\eta_k(u^h)\ \Delta_\tau \varphi^h - \langle \nu, \eta_k \rangle \frac{\partial \varphi}{\partial t} \right) dt\, dxdy \right| &\leq \left| \iint \left(\eta_k(u^h) - \langle \nu, \eta_k \rangle \right) \frac{\partial \varphi}{\partial t}\ dt\, dxdy \right| \\ &+ \left| \iint \eta_k(u^h) \left(\frac{\partial \varphi}{\partial t} - \Delta_\tau \varphi^h \right) dt\, dxdy \right| \end{aligned} \tag{3.45}$$

When τ tends to zero, the first term of the right hand side of (3.45) tends to zero by definition of the Young measure ν, and the second one tends also to zero since the function φ is in $C^1(R^2 x R^+)$,

$$\lim_{h \to 0} \left\| \Delta_\tau \varphi^h - \frac{\partial \varphi}{\partial t} \right\|_{L^1(R^2 x R^+)} = 0$$

and thanks to the uniform L^∞ estimate (3.12). Thus, the limit of the first term in (3.25) is

$$\lim_{h \to 0} \int_{\Omega x R_+} \eta_k(u^h)\ \Delta_\tau \varphi^h\ dtdx \ = \int_{\Omega x R_+} \langle \nu_{(x,t)}, \eta_k(\lambda) \rangle \ \frac{\partial \varphi}{\partial t}\ dx\, dt \tag{3.46.a}$$

Using the same arguments, we get the limit of the other terms of the left hand side in (3.25) :

$$\lim_{h \to 0} \int_{\Omega x R_+} F_k(u^h)\ \Delta_x \varphi^h\ dt\, dxdy = \int_{\Omega x R_+} \langle \nu_{(x,t)}, F_k(\lambda) \rangle \ \frac{\partial \varphi}{\partial x}\ dx\, dt \tag{3.46.b}$$

$$\lim_{h \to 0} \int_R \eta_k(u^h(x,0))\ \varphi^h(x,0)dx = \int_\Omega \langle \nu_{(x,0)}, \eta_k(\lambda) \rangle \ \varphi(x,0)\ dx \tag{3.46.c}$$

Using lemma 3.2. we get the limit of the two last terms in the left hand side of (3.44)

$$\lim_{h\to 0} \int_{R_+} \operatorname{sgn}(u^h(0,t)-k)\ (W^h_{1/2}(t) - f(k))\ \varphi^h(0,t)dt =$$

$$\int_{R_+} \operatorname{sgn}(a(t)-k)\ \langle \nu_{(0,t)},\ f(\lambda) - f(k)\rangle\ \varphi(0,t)dt \qquad (3.46.d)$$

$$\lim_{h\to 0} \int_{R_+} \operatorname{sgn}(u^h(1,t)-k)\ (W^h_{N+1/2}(t) - f(k))\ \varphi^h(1,t)dt =$$

$$\int_{R_+} \operatorname{sgn}(b(t)-k)\ \langle \nu_{(1,t)},\ f(\lambda) - f(k)\rangle\ \varphi(1,t)dt \qquad (3.46.e)$$

We now consider the right hand side of relation (3.44). since, by definition of φ_i^n (3.45), we have :

$$\varphi^n_{i+1/2} - \varphi^n_i \geq 0 \qquad \text{and} \qquad \varphi^n_{i-1/2} - \varphi^n_i \geq 0 \qquad (3.47)$$

and in view of formulae (3.20) and (3.21) of lemma 3.1., we have :

$$R_1^h \leq C_1^h + D_1^h$$

with $$C_1^h = -\tau \sum_n \sum_{1\leq i\leq N} \operatorname{sgn}(u_i^n-k)\ [\varphi^n_{i+1/2} - \varphi_i^n]\ C(u_i^n,u_{i+1}^n)\ \Delta_+u_i^n$$

and $$D_1^h = \tau \sum_n \sum_{1\leq i\leq N} \operatorname{sgn}(u_i^n-k)\ [\varphi^n_{i-1/2} - \varphi_i^n]\ D(u_{i-1}^n,u_i^n)\ \Delta_+u_{i-1}^n$$

Let us now recall that, under the stability condition (3.7), the incremental coefficients C and D of an E-scheme satisfy the following properties

$$\forall\, u,v \in \mathbb{R},\quad C(u,v)\geq 0,\quad D(u,v)\geq 0 \quad \text{and} \quad C(u,v)+D(u,v)=Q(u,v) \leq 1$$

We have, since : $|\operatorname{sgn}(u_i^n-k)| \leq 1$

$$|C_1^h + D_1^h| \leq \tau \sum_n \sum_{1\leq i\leq N-1} |\varphi^n_{i+1/2} - \varphi^n_i|\ [C(u_i^n,u_{i+1}^n) + D(u_i^n,u_{i+1}^n)]\ |\Delta_+u_i^n|$$

$$+\tau \sum_n \left\{ |\varphi^n_{1/2} - \varphi^n_i|\ D(u_0^n,u_1^n)\ |\Delta_+u_0^n| + |\varphi^n_{N+1/2} - \varphi^n_N|\ C(u_N^n,u_{N+1}^n)\ |\Delta_+u_N^n| \right\}$$

since φ is in $C_o^1(\Omega x\mathbb{R}_+)$, we can write

$$|C_1^h + D_1^h| \leq \frac{1}{2}\ \tau h\ \left\|\frac{\partial\varphi}{\partial x}\right\|_\infty \left\{ \sum_n \sum_{1\leq i\leq N-1} Q(u_i^n,u_{i+1}^n)\ |\Delta_+u_i^n| \right.$$

$$\left. + \sum_n D(u_o^n,u_1^n)\ |\Delta_+u_o^n| + \sum_n C(u_N^n,u_{N+1}^n)\ |\Delta_+u_N^n| \right\} \qquad (3.48)$$

By the Hölder inequality, and using relation (3.14) of theorem 3.1. we find :

$$\tau h \sum_n \sum_{0\le i\le N} Q(u_i^n,u_{i+1}^n)\ |\Delta_+u_i^n| \le \Big(\sum_n \sum_{1\le i\le N-1} h\, Q^2(u_i^n,u_{i+1}^n)\ |\Delta_+u_i^n|^2\Big)^{1/2} \Big(\sum_n \sum_{1\le i\le N-1} \tau^2 h\Big)^{1/2}$$

$$\le K_2\ \mathrm{supp}(\varphi)\ \tau^{1/2} \qquad (3.49.a)$$

Similarly (since $0\le C,D\le 1$) we get for the two last sums in (3.48) the following inequalities

$$\tau h \sum_n D(u_0^n,u_1^n)\ |\Delta_+u_0^n| \le 2\,K_1\,\mathrm{supp}(\varphi)\,h \qquad (3.49.b)$$

$$\tau h \sum_n C(u_N^n,u_{N+1}^n)\ |\Delta_+u_N^n| \le 2\,K_1\,\mathrm{supp}(\varphi)\,h \qquad (3.49.c)$$

From (3.49) we get

$$|C_1^h + D_1^h| \le K'\tau^{1/2} + 2\,K_1\,\mathrm{supp}(\varphi)\,h$$

where K' is independent of h and just depend of φ, finally

$$\lim_{h\to 0} R_1^h = 0$$

This ends the proof of theorem 3.2.

In order to apply the new method of proof of convergence based on the measure-valued solution theory, it is clear that the main difficulty is to derive the weak estimate (3.14) of theorem 3.1. That is what next section is devoted to.

4. A weak estimate of the space derivatives for the E-schemes

4.1. Introduction

In this section, we derive a sharp entropy inequality to the first order accurate E-schemes applied to an equation with one space variable. For this purpose we will first consider, in paragraph 4.2., the Godounov scheme and then we will recall, in paragraph 4.3., a result related to the Lax-Friedrichs scheme and due to Di Perna [DP.2] which provides a quadratic estimate of the local entropy dissipation of the Lax-Friedrichs scheme. The extension to a general E-scheme is obtained in paragraph 4.4, by writing it as a convex local combination of the Godounov scheme and the Lax-Friedrichs scheme [Tad].

Let us recall that such an inequality has been derived by Coquel and Lefloch [CL.2] for the Godounov scheme applied to a monodimensional conservation law equation with a strictly convex flux. The strict convexity implies that the solution of the Riemann problem in the Godounov scheme is composed of only a shock or a rarefaction wave. This makes the analysis of the solution of the Riemann problem easier than in the case of a non necessarely convex flux (see paragraph 4.2.2.).

4.2. An estimate of the entropy dissipation in the Godounov scheme

4.2.1 Godounov numerical flux

The Godounov scheme uses the solution of the one dimensional Riemann problem, which we state : For all (u,v) in $\mathbb{R}^2$ and for each regular function $g : \mathbb{R} \longrightarrow \mathbb{R}$, find $w : \mathbb{R}_+ x \mathbb{R} \to \mathbb{R}$ solution of the following initial-value problem :

$$\begin{cases} \partial_t w + \partial_x g(w) = 0 \\ w(0,x) = \begin{cases} u & \text{if} \quad x<0 \\ v & \text{if} \quad x>0 \end{cases} \end{cases} \tag{4.1}$$

The solution w of this problem satisfies the following property : $w(x, t; u, v, g) = w(x/t ; u, v, g)$. The resolution of this problem will be detailed in **§ 4.2.2.** The Godounov numerical flux g^G, is given by :

$$g^G(u,v) = g(w(0; u, v, g)) \qquad \forall\ (u,v) \in \mathbb{R}^2 \tag{4.2}$$

With the notations of section 3, the Godounov scheme is given by :

$$u_i^{n+1} = u_i^n - \lambda_x \left[g\left(w(0; u_i^n, u_{i+1}^n, g)\right) - g\left(w(0; u_{i-1}^n, u_i^n, g)\right) \right] \tag{4.4}$$

where λ_x satisfy a Courant-Friedrichs-Levy (CFL) stability condition of the form :

$$\lambda_x \operatorname{Sup}_u \left|\frac{dg}{du}(u)\right| \leq \left(\frac{1}{2} - \varepsilon\right) \tag{4.5}$$

where the supremum is taken on all the values of u under consideration, and ε is a positive real constant contained in the interval]0,1/2[, and independent of h.

Following Tadmor [Tad], the Godounov scheme may be viewed as an L^2 projection of the solution of the local Riemann problems on the interfaces of the mesh. Thus, the entropy dissipation in this scheme is composed of two terms. One term is generated by the L^2 projection error. The other corresponds to the entropy dissipation inside a shock wave or inside a rarefaction wave, which appears in the solution of the Riemann problem.

That is why it is necessary to recall the resolution of the Riemann problem for a scalar non linear conservation law (with a non necessarly convex flux).

4.2.2. The Riemann problem

Let us recall the Riemann problem : for each (u,v) in R^2, and each regular function $g : R \rightarrow R$, find the unique entropy weak solution $w : R_+ x\ R \rightarrow R$ of the following initial-value problem :

PR(u,v; g)

$$\begin{cases} \partial_t w + \partial_x g(w) = 0 \\ w(0,x) = \begin{cases} u & \text{if } x<0 \\ v & \text{if } x>0 \end{cases} \end{cases}$$

An example of flux g is represented in the figure 4.1. below.

Idea : Leroux [Ler]

If $g_c(u,v)$ (resp. $g^c(u,v)$) represents the convex (resp. concave) envelope of g over the interval (u,v), then we have this remarkable property :

$$\begin{cases} \text{if} & u \leq v & \text{then} & w(.;u,v,g) = w(.;u,v,g_c(u,v)) \\ \text{if} & u \geq v & \text{then} & w(.;u,v,g) = w(.;u,v,g^c(u,v)) \end{cases}$$

We will only treat of the case ($u \leq v$), the other one is similar.

Solution :

Let $\{u_i\}$ be the family in [u,v], with <u>finite</u> cardinal, where the concavity of g changes. Let us also suppose that these values are arranged with the following convention :

$$\forall\, w \in]u_{2i}, u_{2i+1}[\qquad g(w) = g_c(u,v)\ (w)$$
$$\forall\, w \in]u_{2i+1}, u_{2i+2}[\qquad g(w) > g_c(u,v)\ (w).$$

If we set $\xi = (x/t)$, the solution of the Riemann problem is given by :

i) $w(\xi; u, v, g) = (g')^{-1}(\xi)$ if $g'(u_{2i}) < \xi < g'(u_{2i+1})$,

ii) $w(\xi; u, v, g)$ contains a shock in : $\xi = g'(u_{2i+1})$, with speed σ_{2i+1} given by :

$$\sigma_{2i+1} = \frac{g(u_{2i+2}) - g(u_{2i+1})}{u_{2i+2} - u_{2i+1}} = g'(u_{2i+1}) = g'(u_{2i+2}) \qquad (4.6)$$

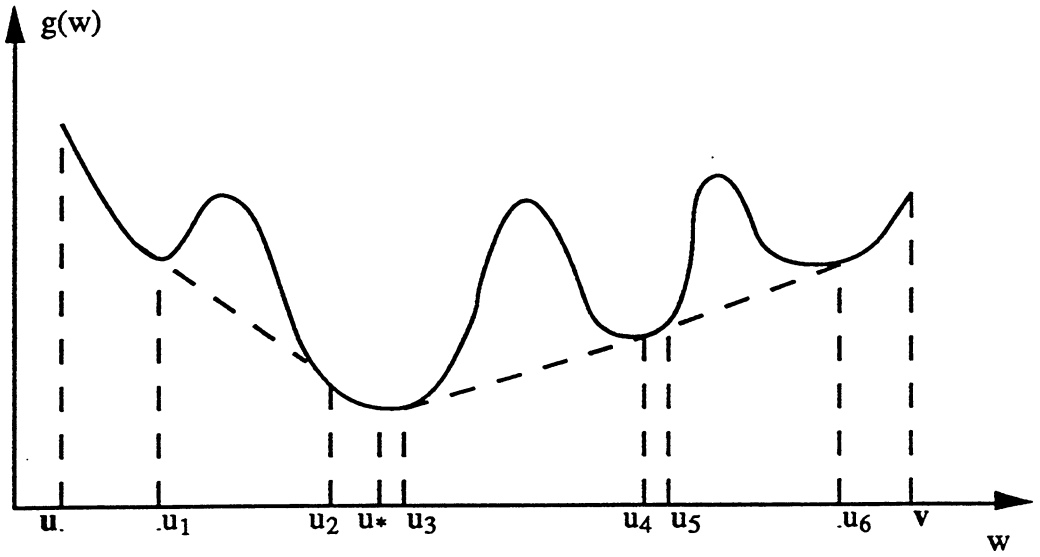


figure 4.1.

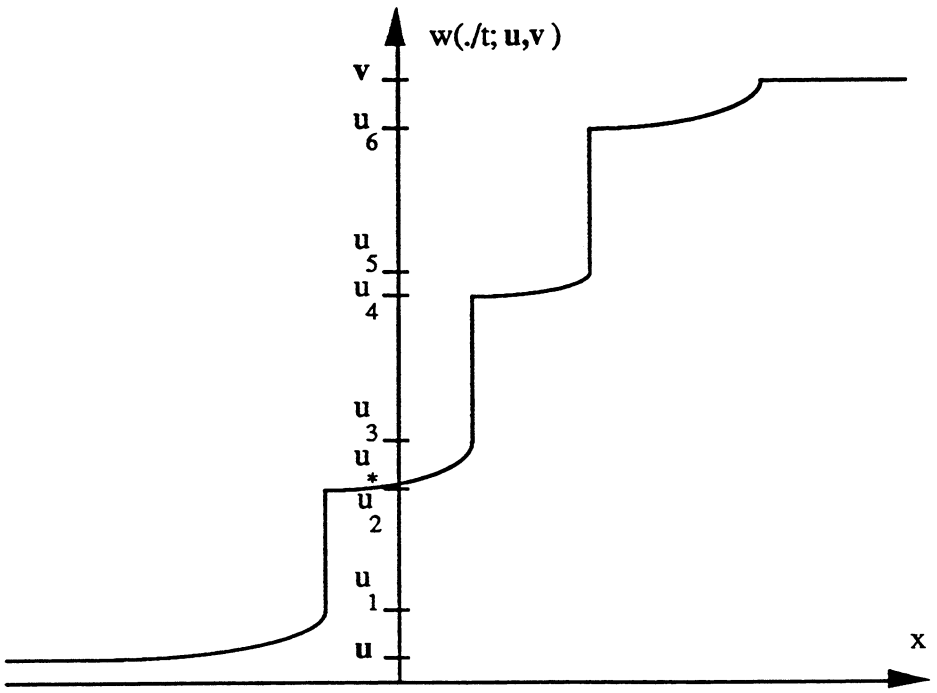


figure 4.2.

If we want to describe the solution $w(.\,; u, v, g)$ at a given $t>0$, we find :

i) $w(\frac{\cdot}{t}; u,v,g)$ consist of a rarefaction wave which lies between u_{2i} and u_{2i+1} if: $tg'(u_{2i})<x<tg'(u_{2i+1})$,

ii) $w(\frac{\cdot}{t}; u,v,g)$ is a shock between u_{2i+1} and u_{2i+2} at : $x = t\,g'(u_{2i+1})$, with speed given by (4.6).

We finally notice, thanks to this analysis, that the solution of the Riemann problem for a scalar conservation law with a non necessarly convex flux, is composed of either a shock only, or a finite

number of rarefaction waves separated "directly" by discontinuities. The graph of the solution of the Riemann problem PR(u,v;g), where the flux g is represented by figure 4.1., is depicted in figure 4.2.

4.2.3. Entropy dissipation in the Godounov scheme

We first recall that the Godounov scheme (4.4), may be seen as an L^2 projection of the solution of the local Riemann problems on the interfaces of the mesh. If we set

$$u_{i+1/2}^{n+1} = \frac{2}{h_x} \int_{-h_x/2}^{0} w(\frac{x}{\tau}; u_i^n, u_{i+1}^n, g)\, dx \qquad (4.7.a)$$

$$u_{i+1/2}^{n+1} = \frac{2}{h_x} \int_{0}^{h_x/2} w(\frac{x}{\tau}; u_{i-1}^n, u_i^n, g)\, dx \qquad (4.7.b)$$

and following Tadmor [Tad], the Godounov finite difference scheme has also an averaged form (see figure 4.3.) :

$$u_i^{n+1} = \frac{1}{2} (u_{i+1/2}^{n+1} + u_{i-1/2}^{n+1}) \qquad (4.8)$$

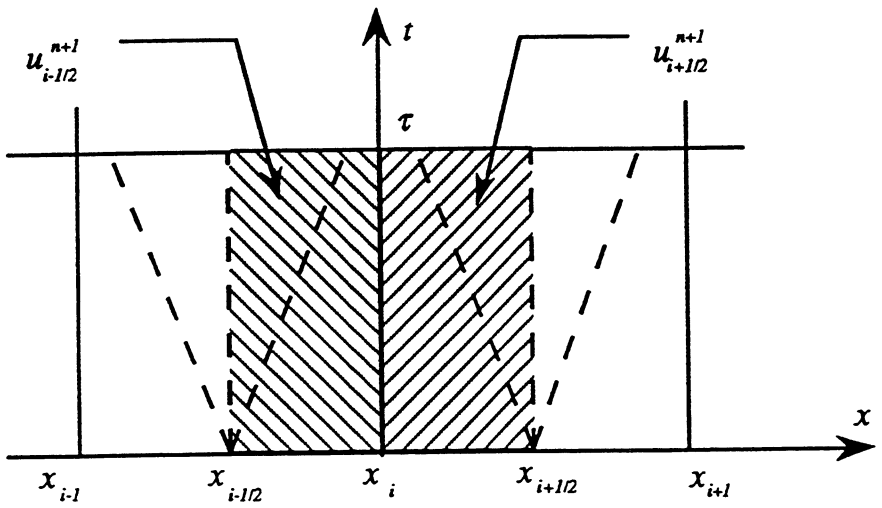


Figure 4.3.

Let us consider the entropy (U_o, F_o) of the scalar conservation law of problem (4.1) defined by :

$$U_o(u) = \frac{u^2}{2} \qquad F_o(u) = \int_0^u v\, g'(v)\, dv.$$

For the sake of simplicity, we introduce the following notations :

$$w_{i+1/2}^{n+1} = w(0; u_i^n, u_{i+1}^n, g)$$

$$w_{i+1/2}^{n+1}(\xi) = w(\xi; u_i^{n+1}, u_{i+1}^n, g) \qquad \text{with} \qquad \xi = \frac{x}{t}$$

Following an idea introduced by Tadmor, and used by Coquel-Lefloch in [CL.2], we get the following result concerning the entropy dissipation in the Godounov scheme.

Lemma 4.1.

Consider the Godounov scheme (4.4) under the CFL condition (4.5), then for each n in $\mathbb{N}$, each i in $\mathbb{Z}$, we have :

$$U_o(u_{i+1/2}^{n+1}) - U_o(u_i^n) - 2\lambda_x\left[F_o(u_i^n) - F_o(w_{i+1/2}^{n+1})\right] = 2\lambda_x J_{i,d}^n - \frac{1}{h_x}\int_{-h_x/2}^{0} \left| w_{i+1/2}^{n+1}\left(\frac{x}{\tau}\right) - u_{i+1/2}^{n+1}\right|^2 dx \qquad (4.9.a)$$

$$U_o(u_{i-1/2}^{n+1}) - U_o(u_i^n) + 2\lambda_x\left[F_o(u_i^n) - F_o(w_{i-1/2}^{n+1})\right] = 2\lambda_x J_{i,g}^n - \frac{1}{h_x}\int_{0}^{h_x/2} \left| w_{i-1/2}^{n+1}\left(\frac{x}{\tau}\right) - u_{i-1/2}^{n+1}\right|^2 dx \qquad (4.9.b)$$

where the terms $J_{i,g}^n$, $J_{i,d}^n$ equal to zero except when the Riemann solution contains shocks, and in this case they are given by :

$$J_{i,d}^n = \sum_{k=1}^{n_c^-}\left\{F_o(u_{i,k}^{n,+}) - F_o(u_{i,k}^{n,-}) - \sigma_{i,k}^{n,-}\left[U_o(u_{i,k}^{n,+}) - U_o(u_{i,k}^{n,-})\right]\right\} \qquad (4.10.a)$$

$$J_{i,g}^n = \sum_{k=1}^{n_c^+}\left\{F_o(u_{i,k}^{n,+}) - F_o(u_{i,k}^{n,-}) - \sigma_{i,k}^{n,+}\left[U_o(u_{i,k}^{n,+}) - U_o(u_{i,k}^{n,-})\right]\right\} \qquad (4.10.b)$$

and satisfy $\quad J_{i,g}^n \le 0 \qquad J_{i,d}^n \le 0 \qquad (4.11)$

where $(u_{i,k}^{n,+})$ and $(u_{i,k}^{n,-})$ denote the values of u in $[u_i^n, u_{i+1}^n]$ respectively on the left side and the right side of the shocks with negative (resp. positive) speed $\sigma_{i,k}^{n,-}$ (resp. $\sigma_{i,k}^{n,+}$). n_c^- (resp. n_c^+) denotes the number of shocks waves with negative (resp. positive) speed in the Riemann problem : $PR(u_i^n, u_{i+1}^n; g)$. ∎

Proof of Lemma 4.1.

We only prove formulae (4.9.a) and (4.11), the proof of (4.9.b) is similar. The CFL condition (4.5) ensures that all the waves of the solution of the Riemann problem $PR(u_i^n, u_{i+1}^n; g)$, are contained in the cone $((0,0); ((h_x/2) - \varepsilon, \tau); ((-h_x/2) + \varepsilon, \tau))$. In any area of smoothness of the function $w_{i+1/2}^{n+1}(\xi)$, we have the conservation law of entropy :

$$\partial_t U_o(w_{i+1/2}^{n+1}(\xi)) + \partial_x F_o(w_{i+1/2}^{n+1}(\xi)) = 0$$

which, integrated by parts over each domain of regularity of the subcell $]0,\tau[\times](-h_x/2),0[$, gives under the CFL condition (4.5) :

$$\int_{-h_x/2}^{0} U_o(w_{i+1/2}^{n+1}\left(\frac{x}{\tau}\right))\,dx - U_o(u_i^n);+ \tau\left\{F_o(u_i^n) - F_o(w(0; u_i^n, u_{i+1}^n; g))\right\} = \tau J_{i,d}^n \qquad (4.13)$$

where $J_{i,d}^n$ is defined by (4.10.a). On the other hand, since $U_o(u) = \frac{u^2}{2}$, we have the following identity :

$$U_o(w_{i+1/2}^{n+1}\left(\frac{x}{\tau}\right)) - U_o(u_{i+1/2}^{n+1}) - u_{i+1/2}^{n+1}(w_{i+1/2}^{n+1})) - u_{i+1/2}^{n+1}) = \frac{1}{2}\left| w_{i+1/2}^{n+1}(\xi) - u_{i+1/2}^{n+1}\right|^2$$

which, averaged over the inteval $]-\frac{h_x}{2}; 0[$, and in view of the definition of $u_{i+1/2}^{n+1}$, gives :

$$\frac{2}{h_x}\int_{-h_x/2}^{0} U_o(w_{i+1/2}^{n+1}(\tfrac{x}{\tau}))\,dx = U_o(u_{i+1/2}^{n+1}) + \frac{1}{h_x}\int_{-h_x/2}^{0} |w_{i+1/2}^{n+1}(\tfrac{x}{\tau}) - u_{i+1/2}^{n+1}|^2 dx$$

which used in (4.13), yields the required equality (4.9.a).

On the other hand, we have an entropy inequality satisfied by all the entropies of the scalar conservation law equation of problem (4.1) : $[F_o] \le \sigma\,[U_o]$, where [u] denotes the jump of u across the shock curve of speed σ, which gives in our case :

$$F_o(u_{i,k}^{n,+}) - F_o(u_{i,k}^{n,-}) - \sigma_{i,k}^{n}\,[U_o(u_{i,k}^{n,+}) - U_o(u_{i,k}^{n,-})] \le 0$$

and by summation on all the shocks of negative speed, we finally get the first inequality in (4.11). This ends the proof of lemma 4.1.

Remark 4.1.

Combining relations (4.9.a) and (4.11.a), we get an estimate of the entropy dissipation in the Godounov scheme only due to the L^2 projection :

$$U_o(u_{i+1/2}^{n+1}) - U_o(u_i^n) - 2\lambda_x[F_o(u_i^n) - F_o(w_{i+1/2}^{n+1})] \le -\frac{1}{h_x}\int_{-h_x/2}^{0} |w_{i+1/2}^{n+1}(\tfrac{x}{\tau}) - u_{i+1/2}^{n+1}|^2 dx \qquad (4.14.a)$$

$$U_o(u_{i-1/2}^{n+1}) - U_o(u_i^n) + 2\lambda_x[F_o(u_i^n) - F_o(w_{i-1/2}^{n+1})] \le -\frac{1}{h_x}\int_{0}^{h_x/2} |w_{i-1/2}^{n+1}(\tfrac{x}{\tau}) - u_{i-1/2}^{n+1}|^2 dx \qquad (4.14.b)$$

It remains thus to make a "quadratic" estimate of the entropy dissipation of the L^2 projection.

Theorem 4.1.

If the sequence (u_i^n) is given by the Godounov scheme (4.4), under the CFL condition (4.5), then for each n, for each i, we have :

$$U_o(u_i^{n+1}) - U_o(u_i^n) + \lambda_x[F_o(w_{i+1/2}^{n+1}) - F_o(w_{i-1/2}^{n+1})] \le$$
$$-2\varepsilon\left\{|u_{i+1}^n - u_i^n|^2\,|C_x^G(u_i^n,u_{i+1}^n)|^2 + |u_i^n - u_{i-1}^n|^2\,|D_x^G(u_{i-1}^n,u_i^n)|^2\right\} \qquad (4.15)$$

where C_x^G and D_x^G represent the incremental coefficients of the Godounov scheme (4.4), defined by :

$$\begin{cases} C_x^G(u,v) = \lambda_x \dfrac{g(u) - g(w(0;u,v,g))}{v-u} \\ D_x^G(u,v) = \lambda_x \dfrac{g(v) - g(w(0;u,v,g))}{v-u} \end{cases} \qquad (4.16)$$

The constant ε is defined by the CFL condition (4.5), and is independent of τ, h_x and g.

Proof of theorem 4.1.

Inequality (4.14.a) may be equivalently written as :

$$U_o(u_{i+1/2}^{n+1}) - U_o(u_i^n) - 2\lambda_x[F_o(u_i^n) - F_o(w_{i+1/2}^{n+1})] \le -PL2(i+1/2,n)$$

where PL2(i+1/2,n) denotes the entropy dissipation due to the L^2 projection, i.e.

$$PL2(i+1/2,n) = \frac{1}{h_x}\int_{-h_x/2}^{0} |w_{i+1/2}^{n+1}(\frac{x}{\tau}) - u_{i+1/2}^{n+1}|^2 dx$$

In order to evaluate $u_{i+1/2}^{n+1}$ defined by (4.7.a), let us analyse the solution of the Riemann problem $PR(u_i^n, u_{i+1}^n; g)$, over the subcell $]-h_x/2,0[$ at τ. We will assume in the following that $u_i^n \le u_{i+1}^n$; the other case is similar and yields the same expression for $u_{i+1/2}^{n+1}$.

Using the same notations as in paragraph §4.2.2. Let $\{u_i\}$ (resp. $\{x_i\}$) the values of u (resp. of x) in $[u_i^n, u_{i+1}^n]$ (resp. $[-h_x/2,0]$) such that :

i) $x_i = \tau\, g'(u_i)$,

ii) $w_{i+1/2}^{n+1}(\frac{\cdot}{\tau})$ is a rarefaction between u_{2i} and u_{2i+1}, if : $x_{2i} \le x \le x_{2i+1}$

iii) $w_{i+1/2}^{n+1}(\frac{\cdot}{\tau})$ contains a shock between u_{2i+1} and u_{2i+2} at x_{2i+1} of speed σ_{2i+1} given, by :

$\sigma_{2i+1} = g'(u_{2i+1})$. And with the Rankine-Hugoniot relation, we have :

$$g(u_{2i+2}) - g(u_{2i+1}) = g'(u_{2i+1})\,[u_{2i+2} - u_{2i+1}]. \qquad (4.17)$$

Note that : $x_{2i+1} = x_{2i+2}$. Finally

$$u_{i+1/2}^{n+1} = \frac{2}{h_x}\int_{-h_x/2}^{x_o} w_{i+1/2}^{n+1}(\frac{x}{\tau})\,dx + \frac{2}{h_x}\sum_{x_{2i}\le 0}\int_{x_{2i}}^{x_{2i+1,-}} w_{i+1/2}^{n+1}(\frac{x}{\tau})\,dx \qquad (4.18)$$

where we have set : $x_{2i+1,-} = \min(0, x_{2i+1})$ and $x_o = \min(0, \tau g'(u_i^n))$. The first term of (4.18) may be evaluated as follows :

$$\frac{2}{h_x}\int_{-h_x/2}^{x_o} w_{i+1/2}^{n+1}(\frac{x}{\tau})\,dx = \frac{2}{h_x}(x_o + \frac{h_x}{2})\,u_i^n$$

because $w_{i+1/2}^{n+1}(\frac{x}{\tau})$ is constant for : $x \le x_o$ and we have : $w_{i+1/2}^{n+1}(\frac{x}{\tau}) = u_i^n$.

Over each interval of the form $]x_{2i}, x_{2i+1}[$, the solution of problem PR(u,v;g) is a rarefaction wave, and it is given explicitly by solving the equation : $w(\xi; u, v, g) = (g')^{-1}(\xi)$. For each i, we have :

$$\frac{2}{h_x}\int_{x_{2i}}^{x_{2i+1}} w_{i+1/2}^{n+1}(\frac{x}{\tau})\,dx = \frac{2}{h_x}\int_{x_{2i}}^{x_{2i+1}} (g')^{-1}(\frac{x}{\tau})\,dx$$

$$= \frac{2\tau}{h_x}\int_{u_{2i}}^{u_{2i+1}} u\, g''(u)\,du$$

$$= \frac{2\tau}{h_x}\left\{u_{2i+1}\, g'(u_{2i+1}) - u_{2i}\, g'(u_{2i}) - g(u_{2i+1}) + g(u_{2i})\right\}$$

and

$$u_{i+1/2}^{n+1} = \frac{2}{h_x}(x_o + \frac{h_x}{2})\,u_i^n + \sum_{x_{2i}\le 0}\frac{2\tau}{h_x}\left\{u_{2i+1}\, g'(u_{2i+1}) - u_{2i}\, g'(u_{2i}) - g(u_{2i+1}) + g(u_{2i})\right\}$$

By the Rankine-hugoniot relation (4.17)

$$g(u_{2i+2}) - g(u_{2i+1}) = \sigma_{2i+1}\,[u_{2i+2} - u_{2i+1}]$$

all the terms of the last summation eliminate except the first and the last ones. We are now going to distinguishe three cases, depending on the values taken by the sonic point $w_{i+1/2}^{n+1}$ of the solution of problem PR(u,v;g), we have :

i) if $w_{i+1/2}^{n+1} = u_i^n$, i.e. $g'(u_i^n) \geq 0$: and : $x_o = 0$, then :

$$u_{i+1/2}^{n+1} = u_i^n$$

ii) if $w_{i+1/2}^{n+1} = u_{i+1}^n$, i.e. $g'(u_{i+1}^n) \leq 0$ and : $x_o = \tau g'(u_i^n)$ then :

$$u_{i+1/2}^{n+1} = u_i^n + \frac{2\tau}{h_x}\left\{w_{i+1/2}^{n+1}\, g'(w_{i+1/2}^{n+1}) - g(w_{i+1/2}^{n+1}) + g(u_i^n)\right\} - \frac{2\tau}{h_x} w_{i+1/2}^{n+1}\, g'(w_{i+1/2}^{n+1})$$

$$= u_i^n + \frac{2\tau}{h_x}\left\{g(u_i^n) - g(w_{i+1/2}^{n+1})\right\}$$

iii) if $u_i^n < w_{i+1/2}^{n+1} < u_{i+1}^n$, which implies : $g'(w_{i+1/2}^{n+1}) = 0$: (sonic rarefaction), then

$$u_{i+1/2}^{n+1} = \frac{2}{h_x}(\tau\, g'(u_i^n) + \frac{h_x}{2})\, u_i^n + \frac{2\tau}{h_x}\left\{w_{i+1/2}^{n+1}\, g'(w_{i+1/2}^{n+1}) - g(w_{i+1/2}^{n+1}) + g(u_i^n) - u_i^n\, g'(u_i^n)\right\}$$

$$= u_i^n + \frac{2\tau}{h_x}\left\{g(u_i^n) - g(w_{i+1/2}^{n+1})\right\}$$

We can deduce, thanks to this analysis, that we always have :

$$u_{i+1/2}^{n+1} = u_i^n + \frac{2\tau}{h_x}\left\{g(u_i^n) - g(w_{i+1/2}^{n+1})\right\}. \tag{4.19}$$

Using (4.19), and the CFL condition (3.2), which yields : $x_o + \frac{h_x}{2} \geq \varepsilon\, h_x$, we get

$$PL2(i+1/2,n) \geq \frac{1}{h_x}\int_{-h_x/2}^{x_o} \left| w(\frac{x}{\tau}; u_i^n, u_{i+1}^n; g) - u_{i+1/2}^{n+1}\right|^2 dx$$

$$\geq \frac{1}{h_x}\,(x_o + \frac{h_x}{2})\,\left|u_i^n - u_{i+1/2}^{n+1}\right|^2$$

$$\geq \frac{1}{h_x}\,(x_o + \frac{h_x}{2})\,(\frac{2\tau}{h_x})^2\,\left|g(u_i^n - g(u_*)\right|^2$$

$$PL2(i+1/2,n) \geq \varepsilon\,(\frac{2\tau}{h_x})^2\,\left|g(u_i^n) - g(u_*)\right|^2 \tag{4.20}$$

using (4.16), we get :

$$PL2(i+1/2,n) \geq \varepsilon\,(\frac{2\tau}{h_x})^2\,(\frac{h_x}{\tau})^2\,\left|u_{i+1}^n - u_i^n\right|^2\,\left|C_x^G(u_i^n, u_{i+1}^n)\right|^2$$

$$PL2(i+1/2,n) \geq 4\,\varepsilon\,\left|u_{i+1}^n - u_i^n\right|^2\,\left|C_x^G(u_i^n, u_{i+1}^n)\right|^2 \tag{4.21}$$

Finally, we get the following inequality :

$$U_o(u_{i+1/2}^{n+1}) - U_o(u_i^n) - 2\lambda_x\left[F_o(u_i^n) - F_o(w_{i+1/2}^{n+1})\right] \leq -4\,\varepsilon\,\left|u_{i+1}^n - u_i^n\right|^2\,\left|C_x^G(u_i^n, u_{i+1}^n)\right|^2 \tag{4.22}$$

Strating from (4.14.b), we get an inequality similar to (4.22), which estimates the local entropy dissipation with the D_x^G incremental coefficient :

$$U_o(u_{i-1/2}^{n+1}) - U_o(u_i^n) + 2\lambda_x \left[F_o(u_i^n) - F_o(w_{i+1/2}^{n+1})\right] \le - 4\,\varepsilon\, |u_i^n - u_{i-1}^n|^2\, |D_x^G(u_{i-1}^n, u_i^n)|^2 \qquad (4.23)$$

By the convexity of U_o and, averaging inequalities (4.22) and (4.23), we get the required inequality :

$$U_o(u_i^{n+1}) - U_o(u_i^n) + \lambda_x \left[F_o(w_{i+1/2}^{n+1}) - F_o(w_{i-1/2}^{n+1})\right] \le - 2\,\varepsilon\, |u_{i+1}^n - u_i^n|^2\, |C_x^G(u_i^n, u_{i+1}^n)|^2$$
$$- 2\,\varepsilon\, |u_i^n - u_{i-1}^n|^2\, |D_x^G(u_{i-1}^n, u_i^n)|^2$$

which ends the proof of theorem 4.1.

Now, starting from inequality (4.15) of theorem 4.1., we will derive a global estimation of the entropy dissipation of the Godounov scheme.

Theorem 4.2.

Let $\{u^h\}$ be a sequence given by the Godounov scheme (4.4), under the CFL condition (4.5). Assume that $\{u^h\}$ satisfies the L^∞-estimate (3.13), and that the initial data u_o is in $L^\infty(\Omega)$ and let T be a positive constante, then we have the following estimate :

$$\varepsilon \sum_{n\tau \le T} \sum_{1 \le i \le N-1} h_x |u_{i+1}^n - u_i^n|^2\, |Q_x^G(u_i^n, u_{i+1}^n)|^2 \le K_1(T) \qquad (4.24)$$

where $K_1(T)$ is a constant independent of h and g. Q_x^G is the numerical viscosity of the scheme :

$$Q_x^G(u,v) = \lambda_x \frac{g(u) + g(v) - 2g^G(u,v)}{v - u} = C_x^G(u,v) + D_x^G(u,v)$$

Proof of theorem 4.2.

Multiplying (4.15) of theorem 4.1. by h_x and summing on $n \le n_T$ ($n_T = [T/\tau]$) and $1 \le i \le N$, we get :

$$\sum_{n,i} h_x \left[U_o(u_i^{n+1}) - U_o(u_i^n)\right] + \lambda_x \sum_{n,i} h_x \left[F_o(w_{i+1/2}^{n+1}) - F_o(w_{i-1/2}^{n+1})\right] \le$$

$$- 2\varepsilon \sum_{n,i} h_x |u_{i+1}^n - u_i^n|^2\, |C_x^G(u_i^n, u_{i+1}^n)|^2 - 2\varepsilon \sum_{n,i} h_x |u_i^n - u_{i-1}^n|^2\, |D_x^G(u_{i-1}^n, u_i^n)|^2 \qquad (4.25)$$

The first sum of the left hand side of (4.25) gives :

$$\sum_{n,i} h_x \left[U_o(u_i^{n+1}) - U_o(u_i^n)\right] = \sum_i h_x U_o(u_i^{n_T}) - \sum_i h_x U_o(u_i^o)$$

$$= \frac{1}{2} \|u^h(.,n_T\tau)\|_{L^2(\Omega)} - \frac{1}{2} \|u^o\|_{L^2(\Omega)} \qquad (4.26)$$

The second sum of the left hand side of (4.25) gives :

$$\lambda_x \sum_{n,i} h_x \left[F_o(w_{i+1/2}^{n+1}) - F_o(w_{i-1/2}^{n+1})\right] = \sum_n \tau \left[F_o(w_{N+1/2}^{n+1}) - F_o(w_{1/2}^{n+1})\right] \qquad (4.27)$$

The right hand side of (4.25) may be equivalently written as :

$$2\varepsilon \sum_{n,i} h_x |u_{i+1}^n - u_i^n|^2 |C_x^G(u_i^n, u_{i+1}^n)|^2 + 2\varepsilon \sum_{n,i} h_x |u_i^n - u_{i-1}^n|^2 |D_x^G(u_{i-1}^n, u_i^n)|^2 =$$

$$+ 2\varepsilon \sum_{n\le T} \sum_{1\le i\le N-1} h_x |u_{i+1}^n - u_i^n|^2 \left\{ |C_x^G(u_i^n, u_{i+1}^n)|^2 + |D_x^G(u_i^n, u_{i+1}^n)|^2 \right\}$$

$$+ 2\varepsilon \sum_{n\le T} h_x |u_{N+1}^n - u_N^n|^2 |C_x^G(u_N^n, u_{N+1}^n)|^2 + 2\varepsilon \sum_{n\le T} h_x |u_1^n - u_0^n|^2 |D_x^G(u_0^n, u_1^n)|^2 \qquad (4.28)$$

Combining (4.26), (4.27) and (4.28), we get :

$$2\varepsilon \sum_{n\le T} \sum_{1\le i\le N-1} h_x |u_{i+1}^n - u_i^n|^2 \left\{ |C_x^G(u_i^n, u_{i+1}^n)|^2 + |D_x^G(u_i^n, u_{i+1}^n)|^2 \right\} \le$$

$$\frac{1}{2} ||u^o||_{L^2(\Omega)} - \sum_n \tau \left[F_o(w_{N+1/2}^{n+1}) - F_o(w_{1/2}^{n+1}) \right] \qquad (4.29)$$

where we have eliminate the two last sums of the right hand side of (4.28) and the first term of the right hand side of (4.26) since they are non negatives.

Let us estimate the right hand side of (4.29). For this purpose, since $\{u^h\}$ satisfies the L^∞-estimate (3.13), we deduce that the sequence $w_{i+1/2}^{n+1}(\xi)$ also does, then by using the definition of F_o we get

$$2\varepsilon \sum_{n\le T} \sum_{1\le i\le N-1} h_x |u_{i+1}^n - u_i^n|^2 \left\{ |C_x^G(u_i^n, u_{i+1}^n)|^2 + |D_x^G(u_i^n, u_{i+1}^n)|^2 \right\} \le C(F, u^o)$$

where : $C(F,u^o) = \frac{1}{2} ||u^o||_{L^2(\Omega)} + 2\,T\,||F(u^h)||_\infty$, is a real positive constant independent of h. Finally, using : $(a+b)^2 \le 2(a^2+b^2)$, gives the required inequality. This ends the proof of theorem 4.2.

4.3. An estimate of the entropy dissipation in the Lax-Friedrichs scheme

4.3.1 Lax-Friedrichs numerical flux

The Lax-Friedrichs numerical flux g^L is given by :

$$g^L(u,v) = \frac{1}{2}(g(u) + g(v)) + \frac{1}{2\lambda_x}(v-u) \qquad (4.30)$$

where the coefficient λ_x is independent of u,v and satisfies :

$$\lambda_x \sup_u |g'(u)| < 1/2 \qquad (4.31)$$

The Lax-Friedrichs scheme is given by :

$$u_i^{n+1} = u_i^n - \lambda_x \left(g^L(u_i^n, u_{i+1}^n) - g^L(u_{i-1}^n, u_i^n) \right) \qquad (4.32)$$

As it is well known, the numerical viscosity of the Lax-Friedrichs scheme is constant and equals :

$$Q_x^L(u,v) = 1/2 \qquad \text{for every u and v in } \mathbb{R} \qquad (4.33)$$

4.3.2 Entropy dissipation in the Modified Lax-Friedrichs scheme

In this section, we recall a result related to the Lax-Friedrichs scheme and due to Di Perna [DP.1], which provides a quadratic estimate of the local entropy dissipation of the Lax-Friedrichs scheme.

Theorem 4.3. Di Perna [DP.1]

If $\{u^h\}$ is given by the Lax-Friedrichs scheme (4.32), under the CFL condition (4.5), and T is a positive constante, then we have :

$$\sum_{n\Delta t\le T,\ 1\le i\le N-1} h_x \left| Q_x^L(u_i^n,u_{i+1}^n)\right|^2 \left| u_{i+1}^n-u_i^n\right|^2 \le K_2(T) \qquad (4.34)$$

where $K_2(T)$ is a constant independent of h and g.

4.4. An estimate of the entropy dissipation in the E-scheme

4.4.1. Entropy dissipation in the E-scheme

The E-scheme numerical flux g^E is a function $R^2 \to R$, such that the incremental coefficients :

$$C_x^E(u,v) = h_x \frac{g(u) - g^E(u,v)}{v - u} \qquad (4.36.a)$$

$$D_x^E(u,v) = h_x \frac{g(v) - g^E(u,v)}{v - u} \qquad (4.36.b)$$

satisfy $C_x^E(u,v) \ge 0 \qquad D_x^E(u,v) \ge 0 \qquad \forall (u,v) \in R^2. \qquad (4.37)$

The E-scheme is given by :

$$u_i^{n+1} = u_i^n - \lambda_x \left(g^E(u_i^n, u_{i+1}^n) - g^E(u_{i-1}^n, u_i^n) \right) \qquad (4.38)$$

Theorem 4.4.

If the sequence $\{u^h\}$ is given by the E-scheme (4.38), under the CFL condition (4.5), and T is a positive constante, then we have :

$$\sum_{n\Delta t\le T,\ 1\le i\le N-1} h_x \left| Q_x^E(u_i^n,u_{i+1}^n)\right|^2 \left| u_{i+1}^n-u_i^n\right|^2 \le K(T) \qquad (4.40)$$

where K(T) is a constant dependent of ε and independent of h and g.

Proof of theorem 4.4.

Following Tadmor [Tad], we can write the numerical viscosity of the E-scheme (3.48) as a local convex combination of the numerical viscosity of the Godounov scheme and the numerical viscosity of the Lax-Friedrichs scheme :

$$Q_x^E(u,v) = \chi(u,v).Q_x^L(u,v) + \big(1 - \chi(u,v)\big).Q_x^G(u,v) \qquad \forall (u,v) \in \mathbb{R}^2 \qquad (4.41)$$

where $\chi(u,v)$ is the coefficient of the convex combination, and is in $]0,1[$ for every u and v. Using (4.24) and (4.40), we have :

$$\sum_{n\le T,\ 1\le i\le N-1} h_x \,|u_{i+1}^n - u_i^n|^2 \left\{ |Q_x^L(u_i^n, u_{i+1}^n)|^2 + |Q_x^G(u_i^n, u_{i+1}^n)|^2 \right\} \le \frac{K_1(T)}{\varepsilon} + K_2(T) \qquad (4.42)$$

Using the convexity of the function u^2, (4.41) gives

$$\begin{aligned} |Q_x^E(u_i^n, u_{i+1}^n)|^2 &\le \chi(u_i^n, u_{i+1}^n).\,|Q_x^L(u_i^n, u_{i+1}^n)|^2 + \big(1 - \chi(u_i^n, u_{i+1}^n)\big).\,|Q_x^G(u_i^n, u_{i+1}^n)|^2 \\ &\le |Q_x^L(u_i^n, u_{i+1}^n)|^2 + |Q_x^G(u_i^n, u_{i+1}^n)|^2 \end{aligned}$$

which inserted in (4.42) gives the required inequality (4.40), with :

$$K(T) = \frac{K_1(T)}{\varepsilon} + K_2(T).$$

This ends the proof of theorem (4.4).

If the sequence $\{u^h\}$ given by the finite difference E-scheme (4.38), satisfies an uniform L^∞ bound, then by using estimate (4.40) of theorem 4.4. and the result of theorem 3.1. of section 3, we prove that the Young measure ν associated with $\{u^h\}$ is a mv-solution to equation (1.1-1.3).

CHAPITRE VII

CONVERGENCE DE LA METHODE DES VOLUMES FINIS

1. Introduction

Let Ω be a bounded open set of $\mathbb{R}^2$ with smooth boundary $\Gamma = \partial\Omega$, and outward unit normal n. Consider for $u : \Omega x\mathbb{R}_+ \to \mathbb{R}$, the conservation law

$$u_t + f_1(u)_{x_1} + f_2(u)_{x_2} = 0 \tag{1.1}$$

with initial condition

$$u(0,x) = u_o(x) \tag{1.2}$$

and boundary condition $\forall\, k \in \mathbb{R}$, $(x,t) \in \Gamma x\mathbb{R}_+$

$$(\mathrm{sgn}(u(x,t) - k) - \mathrm{sgn}(a(x,t) - k))\,(f(u(x,t) - f(k))\,.\,n(x) \geq 0 \tag{1.3}$$

where $f = (f_1,f_2)$ is smooth function : $\mathbb{R} \to \mathbb{R}^2$, and u_o is in $L^\infty(\Omega)$. $a : \Gamma x\mathbb{R}_+ \to \mathbb{R}$ is a given smooth function and the function sgn : $\mathbb{R} \to \mathbb{R}$ is defined by

$$\mathrm{sgn}(x) = \begin{cases} x/|x|, & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Existence and uniqueness are proved for measure valued solutions of (1.1-1.3) in [Sze]. We are dealing in this paper with the convergence of the approximation of the weak entopy solution of this problem by the explicite finite volume method.

A new technique based on the theory of measure valued solution for conservation laws introduced by Di Perna [DP.2], is used for proving the convergence of finite volume schemes in two space variables, without assuming a BV-estimate.

This theory has been used by Szepessy [Sze] to prove the convergence of a scheme based on finite element method (Stream-Line diffusion). Champier, Gallouët and Herbin [CGH] used also this technique to prove the convergence of an upstream finite volume scheme for a non linear hyperbolic equation on a triangular mesh. Using the same technique, Coquel and Lefloch proved the convergence of a finite volume scheme for a conservation law with a strictly convex continuous flux.

This paper is organized as follows : Section two is devoted to some preliminaries on the theory of measure valued solutions and to a uniqueness theorem of the measure valued solution of (1.1-1.3). The result of convergence is established in the third section.

2. A uniqueness result for measure valued solutions

2.0. Introduction

In this section, we study the solutions of the scalar conservation laws with initial and boundary conditions (1.1-1.3). A uniqueness theorem for such solutions is proved which is analogous to a uniqueness theorem for measure-valued solutions to the pure initial value problem by Gallouët and Herbin [GH].

2.1. Preliminaries on measure-valued solutions to conservation laws

In this section, we recall some notions related to the measure valued solutions for hyperbolic conservation laws. A Young measure ν defined on $\Omega x\mathbb{R}_+$ is an application $\Omega x\mathbb{R}_+ \to \text{Prob}(\mathbb{R})$, such that for any function a in $C(\mathbb{R})$, the application $(x,t)\to\langle\nu_{x,t},a\rangle$ is in $L^\infty(\Omega x\mathbb{R}_+)$,

with $$\langle\nu_{x,t},a(\lambda)\rangle = \int_{\mathbb{R}} a(\lambda)\, d\nu_{x,t}(\lambda) = \mu_a(x,t) \qquad \forall\ (x,t) \in \Omega x\mathbb{R}_+$$

Let $\{u^h\}$, $u^h \in L^\infty(\Omega x\mathbb{R}_+)$ be a uniformly bounded sequence, i.e.

$$\| u^h \|_{L^\infty(\Omega x\mathbb{R}_+)} \le K \tag{2.1}$$

(in the applications the u^h will be approximate solutions to (1.1-1.3)). Then there exists according to Young's theorem, cf [DP.2], [Tar], a subsequence which we still label $\{u^h\}$ and an associated Young measure $\nu_{(x,t)} : \Omega x\mathbb{R}_+ \to \text{Prob}(\mathbb{R})$ such that

$$\{\lambda : |\lambda| \le K\} \supset \text{supp}(\nu_{(x,t)}) \qquad \text{for a.e. } (x,t) \text{ in } \Omega x\mathbb{R}_+ \tag{2.2}$$

and for all g in $C(\mathbb{R})$ the $L^\infty(\Omega x\mathbb{R}_+)$ weak star limit

$$g(u^h(.)) \to \mu_g(.) \qquad \text{as} \quad h \to 0 \tag{2.3}$$

exists. Here $\text{Prob}(\mathbb{R})$ is the space of all positive Borel measures on $\mathbb{R}$ with unit total mass.

To a given Young measure $\nu : \Omega x\mathbb{R}_+ \to \text{Prob}(\mathbb{R})$ satisfying (2.2) we shall associate (see [Sze]), in general in a non-unique way, a Young measure $\gamma\nu_{(.)} : \Gamma x\mathbb{R}_+ \to \text{Prob}(\mathbb{R})$ which we consider to be a "trace" of ν on $\Gamma x\mathbb{R}_+$.

We may now introduce the definition of a measure-valued solution to (1.1-1.3).

Définition. 2.1.

A Young measure $\nu : \Omega x R_+ \to Prob(R)$ is a mv-solution to (1.1-1.3) if $\forall\, \phi \geq 0 \in C_o^1(\bar{\Omega} x R_+)$, $\forall\, k \in R$,

$$\int_{\Omega x R} \left\{ \langle \nu_{(x,t)}, |\lambda - k| \rangle \phi_t + \langle \nu_{(x,t)}, [f(\lambda) - f(k)]\, sgn(\lambda - k) \rangle \nabla\phi \right\} dx\, dt$$

$$- \int_{\Gamma x R_+} \langle \gamma\nu_{x_b,t}, f(\lambda) - f(k) \rangle . n(x_b)\, \phi\, sgn(a-k)\, ds\, dt$$

$$+ \int_{\Omega} \langle \nu_{(x,0)}, |\lambda - k| \rangle \phi(x,0) dx \geq 0 \qquad (2.6)$$

where : $f(u) = (f_1(u), f_2(u))^T$.

Remark 2.1.

In general the "trace" $\gamma\nu$ is associated to ν a in a non-unique way. However it is shown in [Sze] that the expected value $\langle \gamma\nu_{x,t}, f(\lambda) \rangle$ in uniquely defined.

In [BLN], existence and uniqueness are proved for BV-solutions. For u in $BV^{loc}(\Omega x R_+)$ to be a BV-solution of (1.1-1.3) means precisely that the Dirac measure concentrated at u is a mv-solution, i.e. that $\nu_{x,t} = \delta_{u(x,t)}$ and $\gamma\nu_{x,t} = \delta_{\gamma u(x,t)}$ satisfies (2.6) , where γu denotes the trace of u on $\Gamma x R_+$. This proves also the existence of a mv-solution to (1.1-1.3).

We have the following uniqueness result for mv-solutions whose proof is essentially based on a uniqueness theorem for measure-valued solutions to the pure initial value problem by Gallouët and Herbin [GH].

Theorem. 2.1.

Let the initial data u_o be in $L^\infty(\Omega)$. Assume that $\nu : \Omega x R_+ \to Prob(R)$ is a mv-solution to (1.1-1.3). Then if u is the unique measure valued solution to the boundary value problem (1.1-1.3), we have :

$$\nu_{(x,t)} = \delta_{u(x,t)}, \quad \text{for every} \quad (x,t) \in \Omega x R_+$$

Proof of theorem. 2.1. (see [BCV.1])

Theorem. 2.2.

Let $\{u^h\}$, $u^h \in L^\infty(\Omega x R_+)$ be a uniformly bounded sequence of approximate solutions to (1.1-1.3). If the Young measure ν associated to $\{u^h\}$ and defined on $\Omega x R_+$ and its trace $\gamma\nu$ defined on $\Gamma x R_+$ satisfie relation (2.6) then $\{u^h\}$ tends to u in $L^1(\Omega x R_+)$, where u is the entropy weak solution to (1.1-1.3).

Proof of theorem. 2.2.

The proof of this theorem is essentially based on the result of uniqueness of the measure valued solution of problem (1.1-1.2) and which is due to Di Perna [DP.1]. In the case of the pure initial value problem (1.1-1.2), the proof is detailed in [GH].

3. Convergence of finite volume schemes

3.1. Description of the scheme

In this section, we describe the finite volume method for a general mesh T_h of R^2 :

$$R^2 = \bigcup_{K \in T_h} K$$

3.1.1. Notations

We introduce some notations related to the mesh. Let K be a polyhedron of T_h, e an edge of K. We set :

- $|e|$: the length of e,
- $n_{e,K}$: the outward unit normal to e (from K),
- K_e : the polyhedron of T_h such that : $K \cap K_e = e$,
- $|K|$: the area of the polyhedron K,
- $P_K = \sum_{e \in \partial K} |e|$: the perimeter of K, ∂K denotes the contour of K,
- h_K = diameter of K,
- $h = \sup_{K \in T_h} (h_K)$,
- τ : the time increment,
- $\lambda = \frac{\tau P_K}{|K|}$.

Hypothesis on the mesh

Let us assume that $\tau \to 0$ when $h \to 0$, and that the ration $\frac{\tau}{h}$ is kept constant. We suppose also, without loss of generality, that :

$$\exists\, a,b \text{ in } R_+^* \text{ such that } \quad a.h < \tau < b.h \tag{3.1.a}$$

$$\exists\, c,d \text{ in } R_+^* \text{ such that } \quad c.h < |e| < d.h \tag{3.1.b}$$

inequality $\tau < bh$ is insured by the following CFL condition :

$$\lambda \sup_u |f'(u).n_{e,K}| \leq (\frac{1}{2} - \varepsilon) \tag{3.2}$$

where ε is a positive constant independent of h and satisfying : $\varepsilon \in]0,1/2[$, anf $f(u) = (f_1(u),f_2(u))^T$. The supremum is taken over the considered values of u.

3.1.2 Finite volume schemes

We define for each polyhedron K, and for each edge e of K, a numerical flux : $g_{K,e} : \mathbb{R}^2 \to \mathbb{R}$, which satisfies the following properties :

1) conservativity : $g_{K,e}(u,v) = - g_{K_e,e}(v,u)$ (3.3)

2) consistency with : $f(u).n_{e,K}$

i.e $g_{K,e}(u,u) = f(u).n_{e,K}$ (3.4)

With the former notations, the finite volume scheme may be written :

$$u_K^{n+1} = u_K^n - \frac{\tau}{|K|} \sum_{e \in \partial K} |e|\, g_{K,e}(u_K^n, u_{K_e}^n) \tag{3.5}$$

$$u_K^o = \frac{1}{|K|} \int_K u_o(x)\, dx \tag{3.6.a}$$

$$u_e^n = \frac{1}{\tau |e|} \int_e \int_{n\tau}^{\tau(n+1)} a(x,t)\, d\Gamma(x)\, dt \tag{3.6.b}$$

We define the family of functions u^h : $\Omega \times \mathbb{R}_+ \to \mathbb{R}$

$$u^h(x,t) = u_K^n \quad \text{if} \quad (x,t) \in K \times [n\tau,(n+1)\tau[\tag{3.7.a}$$

Let us define a piecewise constant unit normal function as follows

$$n^h(x,t) = n_{e,K} \quad \text{if} \quad (x,t) \in \{e\} \times [n\tau,(n+1)\tau[\tag{3.7.b}$$

and $$u_\Gamma^h(x,t) = u_e^n \quad \text{if} \quad (x,t) \in \{e\} \times [n\tau,(n+1)\tau[\tag{3.7.b}$$

If u^h is in $L^\infty(\Omega \times \mathbb{R}_+)$, we can define the measure ν as follows, for each function a in $C(\mathbb{R})$, we have :

$$\langle \nu, a \rangle = \lim_{h \to 0} a(u^h) \qquad \text{the limit is taken in } L^\infty \text{ weak-star sense} \tag{3.8}$$

3.2. L^∞-stability

We recall that to introduce the measure ν associated with the family of functions $\{u^h\}$ defined by (3.7.a), it is necessary to establish a L^∞ estimate :

Proposition 3.1.

If the numerical flux $g_{K,e}$ used in the finite volume scheme (3.5)-(3.6) is associated with an E-scheme, then under the following CFL stability condition

$$\lambda \sup_{u \le \|u_o\|_\infty} |f'(u).n_{e,K}| \le 1 \tag{3.9}$$

the familly $\{u^h\}$ defined by (3.7.a) satisfies the following local maximum principle :

$$\min\Big(u_K^n, \min_{e \in \partial K}(u_{K_e}^n)\Big) \le u_K^{n+1} \le \max\Big(u_K^n, \max_{e \in \partial K}(u_{K_e}^n)\Big) \qquad (\forall n \in \mathbb{N}) \tag{3.10}$$

Proof of proposition 3.1.

The scheme (3.5) may be written as :

$$u_K^{n+1} = \left\{1 - \frac{\tau}{|K|}\sum_{e\in\partial K} |e|\, C(u_K^n, u_{K_e}^n)\right\} u_K^n + \frac{\tau}{|K|}\sum_{e\in\partial K} \left\{|e|\, C(u_K^n, u_{K_e}^n)\right\} u_{K_e}^n \qquad (3.11)$$

where we have set

$$\forall\,(u,v)\in \mathbb{R}^2 \qquad \begin{cases} C(u,v) = \dfrac{g_{K,e}(u,u) - g_{K,e}(u,v)}{v-u} \\ D(u,v) = \dfrac{g_{K,e}(v,v) - g_{K,e}(u,v)}{v-u} \end{cases}$$

On the other hand, since $g_{K,e}(u,v)$ is a three points numerical flux associated with an E-scheme, then under the CFL stability condition (3.2), it satisfies the properties of the TVD numerical fluxes :

$$\forall\,(u,v)\in \mathbb{R}^2 \qquad \begin{cases} C(u,v) \geq 0,\quad D(u,v) \geq 0 \\ \lambda[C(u,v)+D(u,v)] \leq 1 \end{cases}$$

we have, in particular : $0 \leq \lambda C(u,v) \leq 1$. With theses notations, (3.11) becomes :

$$u_K^{n+1} = \alpha_K\, u_K^n + \sum_{e\in\partial K} \alpha_{K_e}\, u_{K_e}^n \qquad (3.12)$$

where : $$\alpha_K = 1 - \lambda \sum_{e\in\partial K} \frac{|e|}{P_K}\, C(u_K^n, u_{K_e}^n)$$

and $$\alpha_{K_e} = \lambda \frac{|e|}{P_K}\, C(u_K^n, u_{K_e}^n)$$

with : $$0 \leq \alpha_{K_e} \leq 1$$

and $$\alpha_K = \sum_{e\in\partial K} \frac{|e|}{P_K}\,[1 - \alpha_{K_e}]$$

This shows that α_K is positive, and that u_K^{n+1} is a convex combination of u_K^n and $(u_{K_e}^n)_{e\in\partial K}$, then we get :

$$\min\Big(u_K^n, \min_{e\in\partial K}(u_{K_e}^n)\Big) \leq u_K^{n+1} \leq \max\Big(u_K^n, \max_{e\in\partial K}(u_{K_e}^n)\Big)$$

Since u_o is in $L^\infty(\Omega)$, then the L^∞-estimate on u^h is an immediate consequence of (3.10). That is :

$$\|u^h\|_\infty \leq \|u_o\|_\infty \qquad (3.13)$$

This ends the proof of proposition 3.1.

Since the L^∞ estimate on u^h is obtained, one can define the Young measure ν by (3.8). In the following, we will prove that ν is mv-solution to problem (1.1-1.3).

For the sake of simplicity , we will introduce the following notations, we set :

* $\eta(u) = |u - k|$
* $F_i(u) = \mathrm{sgn}(u-k)(f_i(u) - f_i(k))$ $\qquad i = 1$ or 2

* $f_{e,K}(u) = \langle f(u), n_{e,K} \rangle$

* $F_{e,K}(u) = \langle F(u), n_{e,K} \rangle$

* $\Delta^n_{K,e}(f) = g_{K,e}(u^n_K, u^n_{K_e}) - f_{e,K}(u^n_K)$

* $\Delta^n_{K,e}(F) = \Pi_{K,e}(u^n_K, u^n_{K_e}) - F_{e,K}(u^n_K)$

* $\Delta^n_e(u) = u^n_{K_e} - u^n_K$

3.3. Convergence

In this paragraph we will prove the following theorem of convergence which allows us to apply theorem 2.2. and deduce the strong L^1 convergence of the finite volume method towards the unique weak entropy solution of (1.1-1.3).

Theorem 3.1.

Let $\{u^h\}$ be a sequence given by the finite volume method (3.5)-(3.6), where $g_{K,e}$ is an E-flux and u_o is in $L^\infty(\Omega)$. Assume that $\{u^h\}$ satisfies the L^∞-estimate (3.13), then under the CFL condition (3.2), if $\{u^h\}$ satisfies the following estimate for every positive constant T:

$$\sum_{n\tau \le T} \sum_{I(\Omega)\supset\{e\}} \tau\, |e|\, |\Delta^n_e(u)|^2\, Q^2_e(u^n_K, u^n_{K_e}) \le C(T,\Omega) \tag{3.14}$$

where $Q_e(u,v)$ is the numerical viscosity associated with $g_{K,e}$, or $g_{K_e,e}$ given by :

$$Q_e(u,v) = \left[f_{e,K}(u) + f_{e,K}(v) - 2g_{K,e}(u,v)\right] / (v-u) = \left[f_{e,K_e}(u) + f_{e,K_e}(v) - 2g_{K_e,e}(u,v)\right] / (v-u) \tag{3.15}$$

where : $I(\Omega) = \overset{\circ}{\Omega}$ and $C(T,\Omega)$ is a positive constant dependent of ε and Ω and independent of h, then the Young measure ν associated with $\{u^h\}$ is a mv-solution of (1.1-1.3)

Henceforward, $C_{K,e}$ and $D_{K,e}$ will denote the incremental coefficients associated to $g_{K,e}$,, defined by

$$C_{K,e}(u,v) = \left[f_{e,K}(u) - g_{K,e}(u,v)\right] / (v-u) \tag{3.16.a}$$

$$D_{K,e}(u,v) = \left[f_{e,K}(v) - g_{K,e}(u,v)\right] / (v-u) \tag{3.16.b}$$

The proof of theorem 3.1. is based on an estimate of a discrete entropy inequality satisfied by te finite volume scheme (3.5). In the following proposition, we establish this local discrete entropy inequality for the finite volume scheme, which is equivalent to the discrete entropy inequality satisfied by the finite difference E-schemes.

Proposition 3.2.

Consider the sequence u_K^n defined by the finite volume method (3.5)-(3.6), under the CFL stability condition (3.2), where $g_{K,e}$ is an E-flux. For every K in T_h, and for every edge e from K, there is a numerical entropy flux $\Pi_{K,e} : \mathbb{R}^2 \to \mathbb{R}$, which satisfies the following properties :

i) consistency with $F_{e,K}$: $\Pi_{K,e}(u,u) = F_{e,K}(u)$

ii) conservativity : $\Pi_{K,e}(u,v) = -\Pi_{K_e,e}(v,u)$

iii) discrete entropy inequality : $\eta(u_K^{n+1}) - \eta(u_K^n) + \frac{\tau}{|K|} \sum_{e \in \partial K} |e|\, \Pi_{K,e}(u_K^n, u_{K_e}^n) \le 0.$

To prove proposition (3.2), we will first establish the following result, which states that the finite volume scheme (3.5) may be written as a convex combination of finite difference schemes.

Lemma 3.1.

One has the following decomposition of the finite volume scheme (3.5) :

$$u_K^{n+1} = \sum_{e \in \partial K} \alpha_e\, u_{K,e}^{n+1} \tag{3.17}$$

with $\alpha_e = \frac{|e|}{P_K}$,

and $$u_{K,e}^{n+1} = u_K^n - \frac{\tau P_K}{|K|} \left[g_{K,e}(u_K^n, u_{K_e}^n) - g_{K,e}(u_K^n, u_K^n) \right] \tag{3.18}$$

Proof of lemma 3.1.

From the consistency of the numerical flux $g_{K,e}$ with $f_{e,K}$, we may write scheme (3.5) as follows :

$$u_K^{n+1} = u_K^n - \frac{\tau}{|K|} \sum_{e \in \partial K} |e| \left[g_{K,e}(u_K^n, u_{K_e}^n) - g_{K,e}(u_K^n, u_K^n) \right] \tag{3.19}$$

since : $$\sum_{e \in \partial K} |e|\, g_{K,e}(u_K^n, u_K^n) = f(u_K^n) \sum_{e \in \partial K} |e|\, n_{e,K} = 0$$

then $$u_K^{n+1} = \sum_{e \in \partial K} \frac{|e|}{P_K} u_K^n - \sum_{e \in \partial K} \frac{\tau |e|}{|K|} \left[g_{K,e}(u_K^n, u_{K_e}^n) - g_{K,e}(u_K^n, u_K^n) \right]$$

$$= \sum_{e \in \partial K} \frac{|e|}{P_K} \left\{ u_K^n - \frac{\tau P_K}{|K|} \left[g_{K,e}(u_K^n, u_{K_e}^n) - g_{K,e}(u_K^n, u_K^n) \right] \right\}$$

$$u_K^{n+1} = \sum_{e \in \partial K} \frac{|e|}{P_K} u_{K,e}^{n+1}.$$

Proof of proposition 3.2.

By (3.17) and (3.18), for every K in T_h, and every n in $\mathbb{N}$, we have :

$$u_K^{n+1} = \sum_{e \in \partial K} \frac{|e|}{P_K} u_{K,e}^{n+1}$$

with $$u_{K,e}^{n+1} = u_K^n - \frac{\tau P_K}{|K|} \left[g_{K,e}(u_K^n, u_{K_e}^n) - g_{K,e}(u_K^n, u_K^n) \right]$$

where $g_{K,e}$ is a numerical flux associated with an E-scheme. It is well known that E-schemes are, under the CFL stability condition (3.2), entropy satisfying, that is one has an entropy numerical flux $\Pi_{K,e}$: $\mathbb{R}^2 \to \mathbb{R}$, consistent with $F_{e,K}$, satisfying :

$$\eta(u_{K,e}^{n+1}) - \eta(u_K^n) + \frac{\tau P_K}{|K|}\left[\Pi_{K,e}(u_K^n, u_{K_e}^n) - \Pi_{K,e}(u_K^n, u_K^n)\right] \le 0 \tag{3.20}$$

After multiplication by $\frac{|e|}{P_K}$ and summation on all e in ∂K, inequality (3.20) becomes :

$$\sum_{e\in\partial K} \frac{|e|}{P_K}\eta(u_{K,e}^{n+1}) - \eta(u_K^n) + \sum_{e\in\partial K}\frac{\tau|e|}{|K|}\left[\Pi_{K,e}(u_K^n, u_{K_e}^n) - \Pi_{K,e}(u_K^n, u_K^n)\right] \le 0 \tag{3.21}$$

Since η is convex and $\Pi_{K,e}$ consistent with $F_{e,K}$ (3.21) gives

$$\eta(u_K^{n+1}) \le \sum_{e\in\partial K}\frac{|e|}{P_K}\eta(u_{K,e}^{n+1}) \tag{3.22}$$

and

$$\eta(u_K^{n+1}) - \eta(u_K^n) + \frac{\tau}{|K|}\sum_{e\in\partial K}|e|\,\Pi_{K,e}(u_K^n, u_{K_e}^n) \le 0.$$

this ends the proof of proposition 3.2.

To prove theorem 3.1, we also need the following technical lemmas whose proof is easy in view of assumptions (3.1.a)-(3.1.b) (for lemma 3.2.).

Lemma 3.2.

For any function φ in $C^2(\Omega\times\mathbb{R}_+)$, for each edge e of any $K\in T_h$, and any integer $n\in\mathbb{N}$, if we set :

$$\Phi_{K,e}^n = \inf_{(x,t)\in I_{K,n}} \varphi(t,x) - \frac{1}{\tau|e|}\int_{\tau n}^{\tau(n+1)}\int_e \varphi(t,x)\,d\Gamma dt \tag{3.23.a}$$

with : $I_{K,n} = K\times[\tau n, \tau(n+1)[$, then we have :

$$\Phi_{K,e}^n \le 0 \qquad \text{and} \qquad |\Phi_{K,e}^n| \le (\tau + h)\,||\varphi||_{C^1(I_{K,n})} \tag{3.23.b}$$

Lemma 3.3

If the numerical flux $g_{K,e}$ used in relation (3.18) is an E-flux, then we have, under the CFL stability condition (3.2), the following property :

$$(\forall K\in T_h)\,(\forall e\in K)\,(\forall n\in\mathbb{N}) \qquad \Delta_{K,e}^n(F) \le \mathrm{sgn}(u_K^n - k)\,\Delta_{K,e}^n(f) \tag{3.24}$$

Proof of lemma 3.3.

It is well known that, under the CFL stability condition (3.2), the E-scheme is entropy satisfying. According to this, the sequence $(u_{K,e}^{n+1})$ given by (3.18) satisfies the following entropy inequality :

$$\eta(u_{K,e}^{n+1}) - \eta(u_K^n) + \frac{\tau P_K}{|K|}\left[\Pi_{K,e}(u_K^n, u_{K_e}^n) - \Pi_{K,e}(u_K^n, u_K^n)\right] \le 0$$

which, thanks to the notations introduced above, may be written as :

$$\eta(u_{K,e}^{n+1}) - \eta(u_K^n) + \lambda\, \Delta_{K,e}^n(F) \le 0 \tag{3.25}$$

On the other hand, we have

$$|v-k| - |u-k| \le \mathrm{sgn}(v-k)\,(v-u) \qquad \forall\, u,v \in \mathbb{R},\ \forall\, k \in \mathbb{R}$$

which implies for (3.25)

$$\lambda\, \Delta_{K,e}^n(F) \le \mathrm{sgn}(u_K^n - k)\,(u_K^n - u_{K,e}^{n+1}) \tag{3.26}$$

and using the definition of $u_{K,e}^{n+1}$

$$u_K^n - u_{K,e}^{n+1} = \lambda\, [g_{K,e}(u_K^n, u_{K_e}^n) - g_{K,e}(u_K^n, u_K^n)] \tag{3.27}$$

Combining (3.26) and (3.27) we get the requiered inequality. This ends the proof of lemma 3.3.

To prove theorem 3.1., we also need the following technical lemma, whose proof is given at the end of this paragraph

lemma 3.4.

Let $W^h(x,t)$ the function defined by :

$$W^h(x,t) = \mathrm{sgn}(u_e^n - k)\,\theta_e^n(u_{K(e)}^n, u_e^n)\, f_{e,K}(w(0; u_{K(e)}^n, u_e^n))$$
$$+\frac{1}{2}\,\mathrm{sgn}(u_e^n - k)(1-\theta_e^n(u_{K(e)}^n, u_e^n))\left\{f_{e,K}(u_{K(e)}^n) + f_{e,K}(u_e^n) - \frac{1}{2\lambda}(u_e^n - u_{K(e)}^n)\right\} \quad \text{if } (x,t) \in \{e\}\times[t_n, t_{n+1}[$$

where $\theta_e^n : \mathbb{R}^2 \to [0,1]$.

Suppose that $\{u^h\}$ satisfies (3.2), then the L^∞-weak star limit of $W^h(x,t)$ exists and

$$\lim_{h\to 0} \int_{\Gamma\times\mathbb{R}_+} W^h(x,t)\,\psi(x,t)\,d\Gamma(x)\,dt = \int_{\Gamma\times\mathbb{R}_+} \langle \gamma\nu_{(x,t)}, f(\lambda)\rangle\, n\,\psi(x,t)\,d\Gamma(x)\,dt \qquad \forall\, \psi \in C_o^1(\Gamma\times\mathbb{R}_+)$$

Proof of theorem 3.1.

By proposition 3.2, we have the following entropy inequality

$$\eta\left(u_K^{n+1}\right) - \eta\left(u_K^n\right) + \frac{\tau}{|K|}\sum_{e\in\partial K} |e|\, \Pi_{K,e}(u_K^n, u_{K_e}^n) \le 0 \tag{3.28}$$

Thanks to the consistency of $\Pi_{K,e}$ with $F_{e,K}$, inequality (3.28) may be written as :

$$\eta\left(u_K^{n+1}\right) - \eta\left(u_K^n\right) + \frac{\tau}{|K|}\sum_{e\in\partial K} |e|\, \Delta_{K,e}^n(F) \le 0 \tag{3.29}$$

Let now φ be a positive function in $C^2(\bar{\Omega}\times\mathbb{R}_+)$ with compact support in $\bar{\Omega}\times\mathbb{R}_+$. Set :

$$\varphi_K^n = \inf_{(t,x)\,\in\, I_{K,n}} \varphi(t,x)$$

After multiplication by $|K|\varphi_K^n$ and summation on all n and K such that $I_{K,n}\cap Supp(\varphi)\neq\emptyset$ (where Supp(φ) denotes the support of φ), inequality (3.29) gives

$$\sum_{n,K}\left[\eta(u_K^{n+1})-\eta(u_K^n)\right]|K|\varphi_K^n+\tau\sum_{n,K}\sum_{e\in\partial K}\varphi_K^n\,|e|\,\Delta_{K,e}^n(F)\le 0 \tag{3.30}$$

The first term of (3.30) may be integrated by parts, i.e.

$$\sum_{n,K}\left[\eta(u_K^{n+1})-\eta(u_K^n)\right]|K|\varphi_K^n=-\sum_{n,K}|K|[\varphi_K^{n+1}-\varphi_K^n]\,\eta(u_K^{n+1})-\sum_K|K|\,\varphi_K^o\,\eta(u_K^o)$$

which provides the formula

$$\sum_{n,K}\left[\eta(u_K^{n+1})-\eta(u_K^n)\right]|K|\varphi_K^n=-\int_{\Omega\times R_+}\eta(u^h)\,\partial_t\varphi(t,x)\,dt\,dx-\int_{R_+}\eta(u^h(x,0))\,\varphi(x,0)\,dx$$

$$+\sum_{n,K}\eta(u_K^{n+1})\int_{\tau n}^{\tau(n+1)}\int_K\left\{\partial_t\varphi-\frac{1}{\tau}[\varphi_K^{n+1}-\varphi_K^n]\right\}dt\,dx \tag{3.31}$$

Let us now transform the second term of (3.30). Using the conservativity property of $\Pi_{K,e}$, we get :

$$\sum_{\substack{K\in T_h\\ e\in\partial K}}\int_e\Pi_{K,e}(u_K^n,u_{K_e}^n)\,\varphi(t,x)\,d\Gamma(x)\,dt=\sum_{e\in\partial\Omega}\int_e\Pi_{K,e}(u_{K(e)}^n,u_e^n)\,\varphi(x,t)\,d\Gamma(x)\,dt \tag{3.32}$$

where $\partial\Omega$ is the boundary of Ω, and K(e) is the polyhedron containing e (when e is on the boundary $\partial\Omega$). On the other hand, the Green formula gives :

$$\int_K F(u_K^n)\,\nabla_x\varphi\,dx=\sum_{e\in\partial K}\int_e F(u_K^n)\,n_{e,K}\,\varphi(x,t)\,d\Gamma(x). \tag{3.33}$$

Combining (3.32) and (3.33) yields the following formula for the second term of inequality (3.30) :

$$\tau\sum_{n,K}\sum_{e\in\partial K}\varphi_K^n\,|e|\,\Delta_{K,e}^n(F)=-\int_{\Omega\times R_+}F(u^h)\,\nabla_x\varphi\,dx\,dt$$

$$+\sum_{n,K}\sum_{e\in\partial K}\int_{\tau n}^{\tau(n+1)}\int_e\Delta_{K,e}^n(F)\,(\varphi_K^n-\varphi)\,d\Gamma(x)\,dt \tag{3.34}$$

$$+\sum_{n}\sum_{e\in\partial\Omega}\int_{\tau n}^{\tau(n+1)}\int_e\Pi_{K,e}(u_{K(e)}^n,u_e^n)\,\varphi(x,t)\,d\Gamma(x)\,dt$$

In view of formulae (3.31) and (3.34), inequality (3.30) may be equivalently written as :

$$\int_{\Omega\times R_+}\left[\eta(u^h)\,\partial_t\varphi(t,x)\,dt\,dx+F(u^h)\,\nabla_x\varphi\right]dx\,dt+\int_\Omega\eta(u^h(x,0))\,\varphi(x,0)\,dx-T_\Gamma\ge R_1^h+R_2^h \tag{3.35}$$

where

$$T_\Gamma=\sum_n\sum_{e\in\partial\Omega}\int_{\tau n}^{\tau(n+1)}\int_e\Pi_{K,e}(u_{K(e)}^n,u_e^n)\,\varphi(x,t)\,d\Gamma(x)\,dt$$

$$R_1^h=\sum_{n,K}\eta(u_K^{n+1})\int_{\tau n}^{\tau(n+1)}\int_K\left\{\partial_t\varphi-\frac{1}{\tau}[\varphi_K^{n+1}-\varphi_K^n]\right\}dx\,dt \tag{3.36}$$

and $$R_2^h = \sum_{n,K} \sum_{e \in \partial K} \int_{t^n}^{t^{(n+1)}} \int_e \Delta_{K,e}^n(F)\,(\varphi_K^n - \varphi)\, d\Gamma(x)\, dt \qquad (3.37)$$

Let us now notice that if we prove that the left hand side of (3.35) tend, with h, to the left hand side of (2.6), and that the limit of the right hand side of (3.35) is non negative this will end the proof of theorem 3.1.

By definition of the Young measure ν associated with u^h, we have :

$$\lim_{h\to 0} \int_{\Omega\times R_+} \left[\eta(u^h)\,\partial_t\varphi(t,x)\,dt\,dx + F(u^h)\,\nabla_x\varphi\right] dx\,dt + \int_\Omega \eta(u^h(x,0))\,\varphi(x,0)\,dx =$$

$$\int_{\Omega\times R_+} \left\{\langle\nu,\eta\rangle\,\partial_t\varphi + \langle\nu,F\rangle\,\nabla_x\varphi\right\} dx\,dt + \int_\Omega \eta(u_o(x))\,\varphi(x,0)\,dx \qquad (3.38)$$

Let us transform the last term of the left hand side of (3.35) : T_Γ. For this purpose we write the numerical entropy flux $\Pi_{K,e}$ as a local convex combination of the Godounov's and the Modified-Lax-Friedrichs ones (see [Tad]), which are respectively given by

$$\Pi^G(u,v) = F_{e,K}(w(0;u,v)) \qquad (3.39.a)$$

$$\Pi^M(u,v) = \frac{1}{2}\left[F_{e,K}(v)+F_{e,K}(u)\right] - \frac{1}{4\lambda}(\eta(v)-\eta(u)) \qquad (3.39.b)$$

where $w(0;u,v)$ is the solution at x=0, of the following Riemann Problem, PR(u,v) :

$$\begin{cases} w_t + f(w)_x = 0 \\ w_o(x) = \begin{cases} u & \text{if} \quad x<0 \\ v & \text{if} \quad x>0 \end{cases} \end{cases}$$

$w(0;u,v)$ has also the following characterization : $w(0;u,v) \in I(u,v)$ such that

$$\operatorname{sgn}(v-u)\, f(w(0;u,v)) = \min_{k \in I(u,v)} \left\{\operatorname{sgn}(v-u)\, f_{e,K}(k)\right\} \qquad (3.40)$$

where $I(u,v) = [\min(u,v), \max(u,v)]$. We may write $\Pi_{K,e}$ as follows

$$\Pi_{K,e}(u_{K(e)}^n, u_e^n) = \theta_e^n\, \Pi^G(u_{K(e)}^n, u_e^n) + (1-\theta_e^n)\, \Pi^M(u_{K(e)}^n, u_e^n)$$

where we have set : $\theta_e^n = \theta(u_{K(e)}^n, u_e^n) \in [0,1]$. Thus we get for T_Γ :

$$T_\Gamma = \sum_n \sum_{e\in\partial\Omega} \int_{t^n}^{t^{(n+1)}} \int_e \left\{\theta_e^n\, \Pi^G(u_{K(e)}^n, u_e^n) + (1-\theta_e^n)\, \Pi^M(u_{K(e)}^n, u_e^n)\right\} \varphi(x,t)\, d\Gamma(x)\, dt \qquad (3.41)$$

To estimate T_Γ, we use the following inequalities, whose proof is easy ([Ler]) by using characterization (3.40) of $w(0; u_{K(e)}^n, u_e^n)$ for the first one and thanks to the stability condition (3.2) for the second one

$$\Pi^G(u_{K(e)}^n, u_e^n) \geq \operatorname{sgn}(u_e^n - k)\left\{f_{e,K}(w(0; u_{K(e)}^n, u_e^n)) - f_{e,K}(k)\right\} \qquad \forall k \in \mathbb{R} \quad (3.42.a)$$

$$\Pi^M(u_{K(e)}^n, u_e^n) \geq \frac{1}{2}\operatorname{sgn}(u_e^n - k)\left\{f_{e,K}(u_{K(e)}^n) - 2f_{e,K}(k) + f_{e,K}(u_e^n) - \frac{1}{2\lambda}(u_e^n - u_{K(e)}^n)\right\} \qquad \forall k \in \mathbb{R} \quad (3.42.b)$$

We deduce from (3.42) the following inequality

$$T_\Gamma \geq \int_{\Gamma\times R+} \operatorname{sgn}(u_\Gamma^h(x,t) - k)\,(W^h(x,t) - f(k).n^h(x,t))\,\varphi(x,t)\, d\Gamma(x)dt \qquad (3.43.a)$$

where

$$W^h(x,t) = \theta_e^n \, f_{e,K}(w(0; u^n_{K(e)}, u^n_e))$$

$$+\frac{1}{2}(1-\theta_e^n)\left\{f_{e,K}(u^n_{K(e)}) + f_{e,K}(u^n_e) - \frac{1}{2\lambda}(u^n_e - u^n_{K(e)})\right\} \qquad \text{if } (x,t) \in \{e\}x[t_n, t_{n+1}[$$

Let $W(x,t)$ be the L^∞-weak star limit of $W^h(x,t)$. This limit exists since we have $w(0; u^n_{K(e)}, u^n_e) \in I(u^n_{K(e)}, u^n_e)$ and that the sequence $\{u^h\}$ satisfies (3.13). Thanks to lemma 3.4., we know that $W(x,t)$ is equal (weakly) to the "trace" $\gamma\nu$ on $\Gamma x\mathbb{R}_+$, of the Young measure ν associated with $\{u^h\}$, which implies

$$\lim_{h\to 0} T_\Gamma \geq \int_{\Gamma x\mathbb{R}_+} \langle \gamma\nu_{(x,t)}, f(\lambda) - f(k)\rangle \, n(x) \, \varphi(x,t) \, \operatorname{sgn}(a-k) \, d\Gamma(x)dt \qquad (3.43.b)$$

Now combining (3.35), (3.38) and (3.43.a) gives

$$\int_{\Omega x\mathbb{R}_+} \left\{\langle \nu, \eta\rangle \partial_t\varphi + \langle \nu, F\rangle \nabla_x\varphi\right\} dx\, dt + \int_\Omega \eta(u_o(x)) \, \varphi(x,0)\, dx$$

$$-\int_{\Gamma x\mathbb{R}_+} \langle \gamma\nu_{(x,t)}, f(\lambda) - f(k)\rangle \, n(x) \, \varphi(x,t) \, \operatorname{sgn}(a-k) \, d\Gamma(x)dt \;\geq\; \lim_{h\to 0} R_1^h + R_2^h$$

It thus remains to prove that the right hand side of (3.35) satisfy

$$\lim_{h\to 0} R_1^h + R_2^h \geq 0$$

The term R_1^h defined by (3.36) converges to 0 with h and τ since φ is of class C^2 and by using lemma.3.2. We now treat of R_2^h defined by (3.37), we have :

$$R_2^h = \sum_{n,K} \sum_{e\in\partial K} \int_{\tau n}^{\tau(n+1)} \int_e \Delta^n_{K,e}(F) \, (\varphi^n_K - \varphi(t,x)) \, d\Gamma dt$$

$$= \sum_{n,K} \sum_{e\in\partial K} \tau\,|e|\, \Delta^n_{K,e}(F) \left\{\varphi^n_K - \frac{1}{\tau|e|}\int_{\tau n}^{\tau(n+1)} \int_K \varphi(t,x)\, d\Gamma dt\right\}$$

$$= \sum_{n,K} \sum_{e\in\partial K} \tau\,|e|\, \Delta^n_{K,e}(F)\, \Phi^n_{K,e}$$

where $\Phi^n_{K,e}$ is defined by (3.23.a). Let us now recall result (3.24) of lemma 3.3, we have

$$(\forall K \in T_h)\,(\forall e \in K)\,(\forall n \in \mathbb{N}) \qquad \Delta^n_{K,e}(F) \;\leq\; \operatorname{sgn}(u^n_K - k)\, \Delta^n_{K,e}(f)$$

and, since $\Phi^n_{K,e} \leq 0$ (see lemma 3.2.), we have the following inequality :

$$R_2^h \geq \sum_{n,K} \sum_{e\in\partial K} \tau\,|e|\, \operatorname{sgn}(u^n_K - k)\, \Delta^n_{K,e}(f)\, \Phi^n_{K,e}$$

It suffices now to prove that the term R_h, defined by :

$$R_h = \sum_{n,K} \sum_{e\in\partial K} \tau\,|e|\, \operatorname{sgn}(u^n_K - k)\, \Delta^n_{K,e}(f)\, \Phi^n_{K,e}$$

tendes to 0 with h.

In view of (3.16.a), R_h may be equivalently written as :

$$R_h = \sum_{n,K} \sum_{e \in \partial K} \tau\, |e|\, \mathrm{sgn}(u_K^n - k)\, \Delta_e^n(u)\, C_{K,e}(u_K^n, u_{K_e}^n)\, \Phi_{K,e}^n$$

Since $(\mathrm{sgn}|u_K^n - k| \leq 1)$, and using lemma 3.2, we get the following inequality :

$$|R_h| \leq \sum_{n,K} \sum_{e \in \partial K} \tau\, |e|\, |\Delta_e^n(u)|\, C_{K,e}(u_K^n, u_{K_e}^n)\, (\tau+h)\, \|\varphi\|_{C^1(\Omega_{K,n})}$$

which, by using a summation on the edges e of the polyhedrization T_h, each edge having two contributions excepted the edges belonging to the boundary, we get :

$$|R_h| \leq \|\varphi\| \sum_n \sum_{I(\Omega) \supset \{e\}} \tau\, (\tau+h)\, |e|\, |\Delta_e^n(u)| \left\{ C_{K,e}(u_K^n, u_{K_e}^n) + C_{K_e,e}(u_{K_e}^n, u_K^n) \right\}$$

$$+ \|\varphi\| \sum_n \sum_{e \in \partial\Omega} \tau\, (\tau+h)\, |e|\, |u_{K(e)}^n - u_e^n|\, C_{K(e),e}(u_{K(e)}^n, u_e^n) \qquad (3.44)$$

where we have set : $I(\Omega) = \mathring{\Omega}$. From the conservativity of the flux $g_{K,e}$, and the definition of the coefficients $C_{K,e}$ and $D_{K,e}$, one can easilly verify that we have the following relation :

$$C_{K_e,e}(u_{K_e}^n, u_K^n) = D_{K,e}(u_K^n, u_{K_e}^n) \qquad (3.45)$$

since $g_{K,e}$ is an E-flux, the incremental coefficients satisfy the following relations

$$Q_e \geq C_{K,e} \geq 0, \qquad Q_e \geq D_{K,e} \geq 0, \qquad Q_e = C_{K,e} + D_{K,e}$$

inserting (3.45) in (3.44), gives

$$|R_h| \leq \|\varphi\| \sum_n \sum_{I(\Omega) \supset \{e\}} \tau\, (\tau+h)\, |e|\, |\Delta_e^n(u)\, |Q_e(u_K^n, u_{K_e}^n)|$$

$$+ \|\varphi\| \sum_n \sum_{e \in \partial\Omega} \tau\, (\tau+h)\, |e|\, |u_{K(e)}^n - u_e^n|\, C_{K(e),e}(u_{K(e)}^n, u_e^n) \qquad (3.46)$$

We now prove that the second term in the right hand side of (3.46) tends to zero with h. We write (since $C_{K(e),e}(u_{K(e)}^n, u_e^n) \leq 1$)

$$\|\varphi\|\, (\tau+h) \sum_n \sum_{e \in \partial\Omega} \tau\, |e|\, |u_{K(e)}^n - u_e^n|\, C_{K(e),e}(u_{K(e)}^n, u_e^n) \leq \|\varphi\|\, (\tau+h) \sum_n \sum_{e \in \partial\Omega} \tau\, |e|\, |u_{K(e)}^n - u_e^n|$$

$$\leq \|\varphi\|\, (\tau+h)\, T\, L(\Gamma)\, 2\, \|u_0\|_\infty$$

where $L(\Gamma)$ is the length of Γ.

Now, using (3.1.a), we get

$$|R_h| \leq \|\varphi\|\, (1+b) \sum_n \sum_{e \in \partial\Omega} \tau\, h\, |e|\, |\Delta_e^n(u)\, |Q_e(u_K^n, u_{K_e}^n)| + \|\varphi\|\, (\tau+h)\, T\, L(\Gamma)\, 2\, \|u_0\|_\infty$$

by the Holder inequality, we find :

$$|R_h| \leq \|\varphi\|\, (1+b) \left(\sum_{n,e} \tau\, |e|\, |\Delta_e^n(u)|^2\, Q_e^2(u_K^n, u_{K_e}^n) \right)^{1/2} \left(h^2 \sum_{n,e} \tau\, |e| \right)^{1/2}$$

$$+ \|\varphi\|\, (\tau+h)\, T\, L(\Gamma)\, 2\, \|u_0\|_\infty$$

and using estimate (3.14) of theorem 3.1. we get :

$$|R_h| \leq C(T,\Omega)\, h \left(\sum_{n,e} \tau\, |e|\right)^{1/2} + \|\varphi\|_{(\tau+h)\,T\,L(\Gamma)\,2}\, \|u_o\|_\infty$$

Using the assumptions of regularity made on the mesh, in paragraph 3.1, and (3.1.a)-(3.1.b), we get

$$\left(\sum_{n,e} \tau\, |e|\right)^{1/2} \leq \left(C \sum_{n,K} \tau \frac{|K|}{h}\right)^{1/2}$$

and then, we get

$$R \leq C' |\Omega|\, h^{1/2} + \|\varphi\|_{(\tau+h)\,T\,L(\Gamma)\,2}\, \|u_o\|_\infty$$

where C and C' are real constants and $|\Omega|$ is the surface of Ω. This ends the proof of theorem 3.1.

proof of lemma 3.4.

Let ϕ (resp. ψ) be a positive function in $C_o^1(\bar{\Omega}x\mathbb{R}_+)$ (resp. $C_o^1(\Gamma x\mathbb{R}_+)$), satisfying

$$\phi(x,t) = \begin{cases} 0 & \text{if} \quad \text{dist}(x,\Gamma) \geq \delta \\ \leq \psi(x,t) & \text{if} \quad \text{dist}(x,\Gamma) < \delta \\ \psi(x,t) & \text{if} \quad (x,t) \in \Gamma x\mathbb{R}_+ \end{cases}$$

we set $\phi_K^n = \phi(x_K,t_n)$

Using the same notations and techniques that in the proof of theorem 3.1. we get

$$\sum_{n,K} [\eta(u_K^{n+1}) - \eta(u_K^n)] |K| \phi_K^n + \tau \sum_{n,K} \sum_{e\in\partial K} \phi_K^n |e| \Delta_{K,e}^n(F) \leq 0 \tag{3.47}$$

We get for the first term of (3.47)

$$\sum_{n,K} [\eta(u_K^{n+1}) - \eta(u_K^n)] |K| \phi_K^n = -\int_{\Omega x\mathbb{R}_+} \eta(u^h)\, \partial_t\phi(t,x)\, dt\, dx$$

$$+ \sum_{n,K} \eta(u_K^{n+1}) \int_{t_n}^{t(n+1)} \int_K \{\partial_t\phi - \frac{1}{\tau}[\phi_K^{n+1} - \phi_K^n]\}\, dt\, dx \tag{3.48}$$

Let us now transform the second term of (3.47). Using the conservativity property of $\Pi_{K,e}$, we get :

$$\sum_{\substack{K\in T_h \\ e\in\partial K}} \int_e \Pi_{K,e}(u_K^n,u_{K_e}^n)\, \phi(x,t)\, d\Gamma(x)\, dt = \sum_{e\in\partial\Omega} \int_e \Pi_{K,e}(u_{K(e)}^n,u_e^n)\, \phi(x,t)\, d\Gamma(x)\, dt$$

$$\tau \sum_{n,K} \sum_{e\in\partial K} \phi_K^n |e| \Pi_{K,e}(u_K^n,u_{K_e}^n) = \sum_{n,K} \sum_{e\in\partial K} \int_{t_n}^{t(n+1)} \int_e \Pi_{K,e}(u_K^n,u_{K_e}^n)\, (\phi_K^n - \phi(x,t))\, d\Gamma(x)\, dt$$

$$+ \sum_{n} \sum_{e\in\partial\Omega} \int_{t_n}^{t(n+1)} \int_e \Pi_{K,e}(u_{K(e)}^n,u_e^n)\, \phi(x,t)\, d\Gamma(x)\, dt \tag{3.49}$$

similarly

$$\tau \sum_{n,K} \sum_{e\in\partial K} \phi_K^n |e| F_{e,K}(u_K^n) = \sum_{n,K} \sum_{e\in\partial K} \int_{t_n}^{t(n+1)} \int_e F_{e,K}(u_K^n)\, \phi(x,t)\, d\Gamma(x)\, dt$$

$$+ \sum_{n,K} \sum_{e\in\partial K} \int_{t_n}^{t(n+1)} \int_e F_{e,K}(u_K^n)\, [\phi_K^n - \phi(x,t)]\, d\Gamma(x)\, dt \tag{3.50}$$

where $\partial\Omega$ is the boundary of Ω, and K(e) denotes the polyhedron containing e (when e is on $\partial\Omega$). Combining (3.49) and (3.50) yields the following formula for the second term of inequality (3.47) :

$$\tau \sum_{n,K} \sum_{e\in\partial K} \phi_K^n |e| \Delta_{K,e}^n(F) = -T_\Gamma - \sum_{n,K} \sum_{e\in\partial K} \int_{t_n}^{t(n+1)} \int_e F_{e,K}(u_K^n)\, \phi(x,t)\, d\Gamma(x)\, dt$$

$$+ \sum_{n,K} \sum_{e\in\partial K} \int_{t_n}^{t(n+1)} \int_e \Delta_{K,e}^n(F)\, (\phi_K^n - \phi)\, d\Gamma dt \tag{3.51}$$

where T_Γ is given by by (3.41) where we replace φ by ϕ.

In view of (3.48), (3.51), and inequality (3.43.b), inequality (3.47) may be written as follows

$$\int_{\Omega\times R_+} \eta(u^h)\,\Delta_t\phi^h\ dtdx + \int_{\Gamma\times R_+} \mathrm{sgn}(u^h_\Gamma(x,t)-k)\ (W^h(x,t) - f(k))\ \phi(x,t)\ d\Gamma(x)dt$$

$$+ \sum_{n,K}\sum_{e\in\partial K} \int_{t^n}^{t^{(n+1)}} \int_e F_{e,K}(u^n_K)\ \phi(x,t)\ d\Gamma(x)\ dt \ \geq\ R_1^h + R_2^h \qquad (3.52)$$

where R_1^h and R_2^h are given by (3.36) and (3.37) if we replace φ by ϕ.

If we pass to the limit on δ in (3.52), the first term of the left hand side tends to zero (by definition of ϕ). On the other hand, in the last term of the left hand side remains only the edges belonging to the boundary $\partial\Omega$. Thus we get the following limit for this term

$$\lim_{\delta\to 0} \sum_{n,K}\sum_{e\in\partial K} \int_{t^n}^{t^{(n+1)}} \int_e F_{e,K}(u^n_K)\ \phi(x,t)\ d\Gamma(x)\ dt = \sum_{n}\sum_{e\in\partial\Omega} \int_{t^n}^{t^{(n+1)}} \int_e F_{e,K}(u^n_{K(e)})\ \psi(x,t)\ d\Gamma(x)\ dt$$

This yields to the following inequality

$$\int_{\Gamma\times R_+} \mathrm{sgn}(u^h_\Gamma(x,t)-k)\ (W^h(x,t) - f(k).n^h(x,t))\ \psi(x,t)\ d\Gamma(x)dt$$

$$+ \sum_{n}\sum_{e\in\partial\Omega} \int_{t^n}^{t^{(n+1)}} \int_e F_{e,K}(u^n_{K(e)})\ \psi(x,t)\ d\Gamma(x)\ dt \ \geq\ R_1^h + R_2^h \qquad (3.53)$$

Let us now pass to the limit on h in (3.53). Thank's to the weak estimate (3.14) we get (see proof of theorem 3.1.):

$$\lim_{h\to 0}\{R_1^h + R_2^h\} \geq 0$$

Using the definition of the trace γv associated to v, and the definition of the L^∞-weak star limit of W^h, we get

$$\int_{\Gamma\times R_+} \mathrm{sgn}(a(x,t)-k)\ (W(x,t) - f(k).n)\ \phi(x,t)\ d\Gamma(x)dt \geq$$

$$\int_{\Gamma\times R_+} \langle \gamma v_{(x,t)},\ \mathrm{sgn}(\lambda-k)(f(\lambda) - f(k))\rangle\ n\ \psi(x,t)\ d\Gamma(x)dt \qquad \forall\, k\in R$$

in particular for : $k \geq \mathrm{Max}(\|u_o\|_\infty, \sup_{(x,t)\in \mathrm{supp}(\psi)} a(x,t))$, and $k \leq \mathrm{Min}(\|u_o\|_\infty, \min_{(x,t)\in \mathrm{supp}(\psi)} a(x,t))$, which implies

$$\int_{\Gamma\times R_+} W(x,t)\ \psi(x,t)\ d\Gamma(x)dt = \int_{\Gamma\times R_+} \langle \gamma v_{(x,t)},\ f(\lambda)\rangle\ n\ \psi(x,t)\ d\Gamma(x)\ dt$$

where supp(ψ) is the support of ψ. This ends the proof of lemma 3.4.

3.4. Derivation of the weak estimate

Now, the remark to make is that it is sufficient to derive inequality (3.14) of theorem 3.1. to prove that the Young measure ν associated with $\{u^h\}$ is an entropy measure valued solution of equation (1.1).

This "weak" estimate will be deduced, like in [BCV.1], from a "sharp" estimation of the entropy dissipation generated by the finite volume method. This estimation will be derived for finite volume schemes based on numerical fluxes associated with E-schemes.

Using a local estimate of the entropy dissipation in monodimensional finite difference scheme, derived by Benharbit, Chalabi and Vila in [BCV.1], we will derive a global estimate depending of the numerical viscosity defined by (3.15).

Lemma 3.5.

Under the CFL condition (3.2), the sequence $(u_{K,e}^{n+1})$ given by (3.18) satisfies the following inequality

$$\eta(u_{K,e}^{n+1}) - \eta(u_K^n) + \frac{\tau P_K}{|K|}\left[\Pi_{K,e}(u_K^n,u_{K_e}^n) - \Pi_{K,e}(u_K^n,u_K^n)\right] \le -2\varepsilon\, |\Delta_e^n(u)|^2\, C_{K,e}^2(u_K^n,u_{K_e}^n) \qquad (3.54)$$

where ε is defined by the CFL condition (3.2), and is independent of τ and h. ∎

Proof of Lemma 3.5.

Since the flux used in relation (3.18) is E-flux, we may use the local estimate of the entropy dissipation derived in [BCV.1], which in our case, may be written as :

$$\eta(u_{K,e}^{n+1}) - \eta(u_K^n) + \frac{\tau P_K}{|K|}\left[\Pi_{K,e}(u_K^n,u_{K_e}^n) - \Pi_{K,e}(u_K^n,u_K^n)\right] \le -2\varepsilon\, |\Delta_e^n(u)|^2\, C_{K,e}^2(u_K^n,u_{K_e}^n)$$

which ends the proof of lemma 3.6.

In view of all this results, we may announce the following theorem :

Theorem 3.2.

Let $\{u^h\}$ be a sequence given by the finite volume method (3.5)-(3.6), where $g_{K,e}$ is an E-flux and u_o is a function in $L^\infty(\Omega)$. Assume that $\{u^h\}$ satisfies the L^∞-estimate (3.13), and let T be a positive constante, then under the CFL condition (3.2) we have the following estimate :

$$\sum_{n\le T}\sum_{I(\Omega)\supset\{e\}} \tau\,|e|\,|\Delta_e^n(u)|^2\, Q_e^2(u_K^n,u_{K_e}^n) \le C(T,\Omega) \qquad (3.55)$$

where : $I(\Omega) = \overset{\circ}{\Omega}$ and $C(T,\Omega)$ is a positive constant dependent of ε and Ω and independent of h.

Proof of theorem 3.2.

After multiplication of inequality (3.54) of lemma 3.5. by $\frac{|e|}{P_K}$ and summation on the edges e of K, the convexity of η yields the following inequality :

$$\eta(u_K^{n+1}) - \eta(u_K^n) + \frac{\tau}{|K|} \sum_{e \in \partial K} |e|\, \Delta_{K,e}^n(F) \le - 2\varepsilon \sum_{e \in \partial K} \frac{|e|}{P_K} |\Delta_e^n(u)|^2 C_{K,e}^2(u_K^n, u_{K_e}^n)$$

Multiplying this inequality by $|K|$ and summing the left hand side on K and $n \le T$, and the right hand side on e and $n \le T$, we get :

$$\sum_{n,K} |K| \eta(u_K^{n+1}) - \sum_{n,K} |K| \eta(u_K^n) + \tau \sum_{n,K} \sum_{e \in \partial K} |e|\, \Delta_{K,e}^n(F) \le$$

$$- 2\varepsilon \sum_{n} \sum_{I(\Omega) \supset \{e\}} \tau\, |e|\, |\Delta_e^n(u)|^2 \left\{ C_{K,e}^2(u_K^n, u_{K_e}^n) + C_{K_e,e}^2(u_{K_e}^n, u_K^n) \right\}$$

$$- 2\varepsilon \sum_{n} \sum_{e \in \partial \Omega} \tau\, |e|\, |u_{K(e)}^n - u_e^n|^2 C_{K(e),e}^2(u_{K(e)}^n, u_e^n) \qquad (3.56)$$

Using relation (3.52) we may write

$$\sum_{n,K} |K| \eta(u_K^{n+1}) - \sum_{n,K} |K| \eta(u_K^n) + \tau \sum_{n,K} \sum_{e \in \partial K} |e|\, \Delta_{K,e}^n(F) \le \qquad (3.57)$$

$$-2\varepsilon \sum_{n} \sum_{I(\Omega) \supset \{e\}} \tau\, |e|\, |\Delta_e^n(u)|^2 \left\{ C_{K,e}^2(u_K^n, u_{K_e}^n) + D_{K,e}^2(u_K^n, u_{K_e}^n) \right\} - 2\varepsilon \sum_{n,\, e \in \partial \Omega} \tau |e|\, |u_{K(e)}^n - u_e^n|^2 C_{K(e),e}^2(u_{K(e)}^n, u_e^n)$$

now since : $-2(a^2+b^2) \le -(a+b)^2$ we get

$$- 2\varepsilon \sum_{n} \sum_{I(\Omega) \supset \{e\}} \tau\, |e|\, |\Delta_e^n(u)|^2 \left\{ C_{K,e}^2(u_K^n, u_{K_e}^n) + D_{K,e}^2(u_K^n, u_{K_e}^n) \right\} \le$$

$$- \varepsilon \sum_{n} \sum_{I(\Omega) \supset \{e\}} \tau\, |e|\, |\Delta_e^n(u)|^2 Q_e^2(u_K^n, u_{K_e}^n)$$

which implies for (3.57), since the second sum in the right hand side of (3.57) is non positive :

$$\varepsilon Q_T \le - \sum_{n,K} |K| \eta(u_K^{n+1}) + \sum_{n,K} |K| \eta(u_K^n) - \tau \sum_{n,K} \sum_{e \in \partial K} |e|\, \Delta_{K,e}^n(F) \qquad (3.58)$$

where $$Q_T = \sum_{n \le T} \sum_{I(\Omega) \supset \{e\}} \tau\, |e|\, |\Delta_e^n(u)|^2 Q_e^2(u_K^n, u_{K_e}^n)$$

Since $$\sum_{e \in \partial K} |e|\, F_{e,K}(u_K^n) = F(u_K^n) \sum_{e \in \partial K} |e|\, n_{e,K} = 0$$

and thank's to the conservativity of the numerical entropy flux $\Pi_{K,e}$, the second term in the left hand side of (3.57) gives :

$$\sum_{n,K} \sum_{e \in \partial K} \tau\, |e|\, \Pi_{K,e}(u_K^n, u_{K_e}^n) = \sum_{n} \sum_{e \in \partial \Omega} \tau\, |e|\, \Pi_{K(e),e}(u_{K(e)}^n, u_e^n)$$

replacing in (3.57), and since the last sum in the right hand side of (3.57) is negative, we get :

$$\varepsilon Q_T \le \sum_K |K|\eta(u_K^0) - \sum_K |K|\eta(u_K^{n_T}) - \sum_n \sum_{e\in\partial\Omega} \tau\, |e|\Pi_{K(e),e}(u_{K(e)}^n, u_e^n) \qquad (3.59)$$

with $n_T=[T/\tau]$. Let us now estimate the second sum in the right hand side of (3.59). For this purpose, by using the local convex combination used in (3.41) of the numerical entropy flux to estimate $\Pi_{K(e),e}$, and the definitions (3.39.a) and (3.39.b) of the Godounov's and the Modified-Lax-Friedrichs numerical entropy fluxes we get, since $\{u^h\}$ satisfies the L^∞-estimate (3.13)

$$\left|\Pi_{K(e),e}(u_{K(e)}^n, u_e^n)\right| \le 2\,\|F(u^h)\|_\infty + \frac{1}{2\lambda}\|\eta(u^h)\|_\infty \le C(\eta,F,u^0)$$

where $C(\eta,F,u^0)$ is a real positive constant independent of h. Using the L^∞-estimate (3.13) we get

$$\left|\sum_K |K|\eta(u_K^0) - \sum_K |K|\eta(u_K^{n_T})\right| \le 2\,\|\eta(u^0)\|_{L^1(\Omega)}$$

This implies for (3.59)

$$\varepsilon Q_T \le 2\,\|\eta(u^0)\|_{L^1(\Omega)} + C(\eta,F,u^0)\, L(\Gamma)\, T$$

where $L(\Gamma)$ is the length of Γ. This ends the proof of theorem 3.2.

4. Convergence of second order finite volume scheme

4.1. Introduction

In this section, we consider the class of high order accurate explicit finite volume schemes. We seek the approximate solution $U^h(t,x)$ in the space V_h^1 defined by

$$V_h^1 = \{f(x); \forall K \in T_h, f \text{ over } K \text{ is a linear function}\}$$

We assume $U^h(t_n,x)$ in V_h^1, and we look for $U^h(t_{n+1},x)$ in V_h^1. We need in this section, in addition to the averages values of $U^h(t,x)$ over K, one more degree of freedom to represent $U^h(t,x)$ in V_h^1. The "slopes" of $U^h(t,x)$ over each K in T_h will be choosen to be this second degree of freedom (see paragraph 4.2. for the details of calculations).

4.2. Description of the scheme

4.2.1. Notations

In addition to the notations of section 3, we introduce the following ones

* $\vec{S_K}$: denotes the slope of $U^h(t,x)$ over K. $\vec{S_K} = (S_K^x; S_K^y)^T$

* $u_K^n(e)$: is the value of $U^h(t,x)$ at the edge e of K, given by

$$u_K^n(e) = u_K^n + \vec{S_K} \cdot \overrightarrow{X_K X_e} \tag{4.1}$$

where X_K (resp. X_e) are the cartesian coordinates of the barycenter of K (resp. the center of the edge e). (see figure 4.1). Henceforward, for any element K of T_h, and any edge e, K_e will denote the element of T_h such that : $K \cap K_e = \{e\}$. We will also denote by r_K the ratio : $r_K = \frac{\tau}{|K|}$.

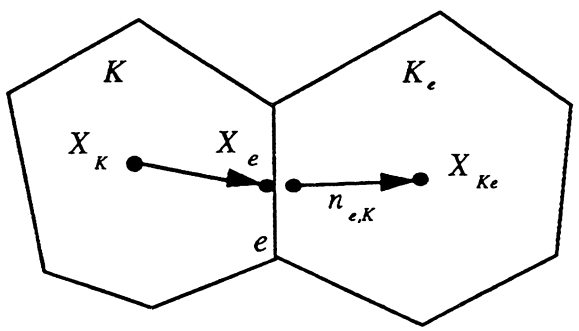


Figure 4.1.

4.2.2. Second order finite volume scheme

With the notations above, the second order accurate finite volume scheme may be written as

$$u_K^{n+1} = u_K^n - r_K \sum_{e \in \partial K} |e|\, g_{K,e}(u_K^n(e), u_{K_e}^n(e)) \tag{4.2.a}$$

$$u_K^0 = \frac{1}{|K|} \int_K u_0(x,y)\, dxdy \tag{4.2.b}$$

with

$$u_K^n(e) = u_K^n + \overrightarrow{S_K} \cdot \overrightarrow{X_K X_e} \tag{4.3.a}$$

$$u_{K_e}^n(e) = u_{K_e}^n + \overrightarrow{S_{K_e}} \cdot \overrightarrow{X_{K_e} X_e} \tag{4.3.b}$$

where $g_{K,e}$ is an E-flux, satisfying the properties of consistency and conservativity defined in section 3. We recall also that this numerical flux is TVD under a stability condition of the form (3.2).

4.2.2.a. Prediction

The calculation of the slopes is made in two steps. First, by giving a first guess, using a technique available in the case of a non regular mesh, which consists of resolving the following problem

$$\min_{(S^x, S^y)^t \in R^2} \sum_{e \in \partial K} \left(u_K^n - u_{K_e}^n - \overrightarrow{S} \cdot \overrightarrow{X_K X_e} \right)^2 \tag{4.4}$$

If we use the slope given by (4.4) in the scheme (4.2)-(4.3), we obtain an approximate solution with great oscillations. It is so necessary to "limit" the slopes in order to avoid oscillations .

4.2.2.b. Correction

The technique used in the limitation consists of putting the slopes equal to zero when a local extrema is created. The idea of limitation may be summarized as follow

$\forall K \in T_h, \forall e \in \partial K, \exists \lambda_e(K) \in [0,1]$, such that

$$\overrightarrow{S_K} \cdot \overrightarrow{X_K X_e} = \lambda_e(K).(u_{K_e}^n - u_K^n) \tag{4.5}$$

Relation (4.5) means that the interfaces values $u_K^n(e)$ must lie between the values u_K^n and $u_{K_e}^n$, where K and K_e are such that : $K \cap K_e = \{e\}$.

It is well known that we may assume, without loss of accuracy, that the correction term vanish with the mesh size, in the following sens : there exist α in]0,1[and a constant C>0 such that :

$$\forall K \in T_h, \forall e \in \partial K, \qquad |\overrightarrow{S_K} \cdot \overrightarrow{X_K X_e}| \le C\, h^{\alpha} \tag{4.6}$$

We recall that an uniform L^∞ estimate is needed in order to define the Young measure associated with the sequence $\{u^h\}$ defined by the scheme (4.2)-(4.4).

4.3. L^∞-stability

To prove the L^∞-stability, we have to put the scheme (4.2.a) under an incremental form. For this purpose, we define the following incremental coefficients, related to a given edge e of the mesh T_h. Let K and K_e two elements of T_h such that : $K \cap K_e = \{e\}$ and the normal to e : $n_{e,K}$ is outward to K. The incremental coefficients related to the numerical flux $g_{K,e}$ are defined by :

$$C(u,v) = \frac{g_{K,e}(u,u) - g_{K,e}(u,v)}{v-u} \tag{4.7}$$

$$D(u,v) = \frac{g_{K,e}(v,v) - g_{K,e}(u,v)}{v-u} \tag{4.8}$$

It is well known that if $g_{K,e}$ is an E-flux, then under a CFL stability condition of the form (3.2), the three points scheme associated with $g_{K,e}$ is TVD, and then the coefficients C and D satisfy

$$\forall\, (u,v) \in \mathbb{R}^2, \quad C(u,v) \geq 0, \quad \text{and} \quad D(u,v) \geq 0.$$

We set

$$\tilde{C}_e = C(u_K^n(e), u_{K_e}^n(e)) \qquad C_e = C(u_K^n(e), u_K^n) \tag{4.9}$$

$$\tilde{D}_e = D(u_K^n(e), u_{K_e}^n(e)) \qquad D_e = D(u_K^n(e), u_K^n) \tag{4.10}$$

We have the following result

Proposition 4.1.

If the numerical flux $g_{K,e}$ used in the finite volume scheme (4.2)-(4.4) is associated with an E-scheme, then under the following CFL stability condition

$$r_K \sum_{e \in \partial K} |e| \left\{ \tilde{C}_e.(1-\lambda_e(K_e)) + C_e.\lambda_e(K) + (\tilde{C}_e + D_e) \sum_{\substack{e' \in \partial K \\ e' \neq e}} \lambda_{e'}(K) \right\} \leq 1 \tag{4.11}$$

the familly $\{u^h\}$ defined by (4.2)-(4.4) satisfies the following maximum principle :

$$(\forall n \in N)\ (\forall K \in T_h) \qquad |u_K^n| \leq \|u_o\|_\infty \tag{4.12}$$

Proof of proposition 4.1.

The scheme (4.2) may be written as :

$$\begin{aligned} u_K^{n+1} = u_K^n &- r_K \sum_{e \in \partial K} |e| \left\{ g_{K,e}(u_K^n(e), u_{K_e}^n(e)) - g_{K,e}(u_K^n(e), u_K^n(e)) \right\} \\ &- r_K \sum_{e \in \partial K} |e| \left\{ g_{K,e}(u_K^n(e), u_K^n(e)) - g_{K,e}(u_K^n(e), u_K^n) \right\} \\ &- r_K \sum_{e \in \partial K} |e| \left\{ g_{K,e}(u_K^n(e), u_K^n) - g_{K,e}(u_K^n, u_K^n) \right\} - r_K \sum_{e \in \partial K} |e|\, g_{K,e}(u_K^n, u_K^n) \end{aligned} \tag{4.13}$$

The last term is equal to zero, thanks to the consistency of $g_{K,e}$ with $f_{e,K}$,

$$\sum_{e \in \partial K} |e|\, g_{K,e}(u_K^n, u_K^n) = \sum_{e \in \partial K} |e|\, f_{e,K}(u_K^n) = f(u_K^n) \sum_{e \in \partial K} |e|\, n_{e,K} = 0$$

Using the incremental coefficients, relation (4.13) may be equivalently written as

$$u_K^{n+1} = u_K^n - r_K\ (A+B+C) \tag{4.14}$$

with $A = \sum_{e\in\partial K} |e|\ \tilde{C}_e\ (u_K^n(e)-u_{K_e}^n(e)), \qquad B = \sum_{e\in\partial K} |e|\ C_e.(u_K^n-u_K^n(e)),$

and $C = \sum_{e\in\partial K} |e|\ D_e.(u_K^n(e)-u_K^n)$

On the other hand, we have

$$u_K^n(e)-u_{K_e}^n(e) = (u_K^n(e)-u_K^n) - (u_{K_e}^n(e)-u_{K_e}^n) + u_K^n - u_{K_e}^n$$

$$= (u_K^n-u_{K_e}^n) + \vec{S}_K.\overrightarrow{X_KX_e} - \vec{S}_{K_e}.\overrightarrow{X_{K_e}X_e}$$

and since $\sum_{e\in\partial K} \overrightarrow{X_KX_e} = \vec{0}$ and $\vec{S}_K.\overrightarrow{X_KX_e} = \lambda_e(K).(u_{K_e}^n - u_K^n)$

then we have $u_K^n(e)-u_{K_e}^n(e) = (u_K^n-u_{K_e}^n) - \vec{S}_K.\sum_{\substack{e'\in\partial K\\ e'\neq e}} \overrightarrow{X_KX_{e'}} - \lambda_e(K_e).(u_K^n - u_{K_e}^n)$

$$= (1 - \lambda_e(K_e))\ (u_K^n-u_{K_e}^n) + \sum_{\substack{e'\in\partial K\\ e'\neq e}} \lambda_{e'}(K)(u_K^n-u_{K_{e'}}^n)$$

wich gives for A : $$A = \sum_{e\in\partial K} |e|\ \tilde{C}_e.\left\{(1 - \lambda_e(K_e))\ (u_K^n-u_{K_e}^n) + \sum_{\substack{e'\in\partial K\\ e'\neq e}} \lambda_{e'}(K)(u_K^n-u_{K_{e'}}^n)\right\} \tag{4.15}$$

Using (4.3.a) we find for B :

$$B = \sum_{e\in\partial K} |e|\ C_e.\lambda_e(K)(u_K^n-u_{K_e}^n) \tag{4.16}$$

Let us now transform the term C, we write :

$$u_K^n(e)-u_K^n = -\vec{S}_K.\overrightarrow{X_KX_e} = \vec{S}_K.\sum_{\substack{e'\in\partial K\\ e'\neq e}} \overrightarrow{X_KX_{e'}} = \sum_{\substack{e'\in\partial K\\ e'\neq e}} \lambda_{e'}(K)(u_K^n-u_{K_{e'}}^n)$$

which gives for C

$$C = \sum_{e\in\partial K} |e|\ D_e.\sum_{\substack{e'\in\partial K\\ e'\neq e}} \lambda_{e'}(K)(u_K^n-u_{K_{e'}}^n) \tag{4.17}$$

Replacing (4.15)-(4.17) in (4.14), we finally find

$$u_K^{n+1} = u_K^n\left[1- r_K \sum_{e\in\partial K} |e|\left\{\tilde{C}_e\left[(1 - \lambda_e(K_e)) + \sum_{\substack{e'\in\partial K\\ e'\neq e}} \lambda_{e'}(K)\right] + \lambda_e(K)\ C_e + D_e \sum_{\substack{e'\in\partial K\\ e'\neq e}} \lambda_{e'}(K)\right\}\right]$$

$$+ u_{K_e}^n\, r_K \sum_{e\in\partial K} |e|\left\{(1 - \lambda_e(K_e)).\tilde{C}_e + \lambda_e(K)C_e\right\} + r_K \sum_{e\in\partial K} |e|\left\{\sum_{\substack{e'\in\partial K\\ e'\neq e}} u_{K_{e'}}^n.\ \lambda_{e'}(K)\left(\tilde{C}_e+D_e\right)\right\}$$

Under the CFL stability condition (4.11), all the coefficients of u_K^n and $u_{K_e}^n$ are positive, which allows us to derive the following inequality

$$|u_K^n| \le |u| \left\{ 1 - r_K \sum_{e\in\partial K} |e| \left\{ \tilde{C}_e.(1 - \lambda_e(K_e)) + \lambda_e(K).C_e + (\tilde{C}_e + D_e) \sum_{\substack{e'\in\partial K \\ e'\neq e}} \lambda_{e'}(K) \right\} \right.$$

$$\left. + r_K \sum_{e\in\partial K} |e| \left\{ (1 - \lambda_e(K_e)).\tilde{C}_e + \lambda_e(K).C_e \right\} + r_K \sum_{e\in\partial K} |e| \left(\tilde{C}_e + D_e \right) \left(\sum_{\substack{e'\in\partial K \\ e'\neq e}} \lambda_{e'}(K) \right) \right\}$$

$$\le |u| \le \|u_o\|_\infty$$

with : $|u| = \sup_K |u_K^n|$. This ends the proof of proposition 4.1.

4.4. Convergence

Since the L^∞ estimate is obtained, one can define the Young measure ν associated with u^h. In the following, we will prove that ν is a mv-solution to problem (1.1-1.2).

Remark 4.1.

For the sake of simplicity, we will consider in this section, only the pure initial value problem (1.1-1.2). The convergence for the boundary value problem (1.1-1.3) is similarely obtained.

Theorem 4.1.

Let $\{u^h\}$ be the sequence given by the second order accurate finite volume scheme (4.2)-(4.4), under the CFL stability condition (4.11), where $g_{K,e}$ is an E-flux and u_o is a function in $L^\infty(\mathbb{R}^2)$. We suppose also that condition (4.6) is satisfied. Then, the Young measure ν associated with $\{u^h\}$ is mv-solution of to problem (1.1-1.2).

Let us first notice that we may write scheme (4.2.a) as follows :

$$u_K^{n+1} = \bar{u}_K^{n+1} - r_K \sum_{e\in\partial K} |e| \left[g_{K,e}(u_K^n(e), u_{K_e}^n(e)) - g_{K,e}(u_K^n, u_{K_e}^n) \right] \tag{4.18}$$

with $$\bar{u}_K^{n+1} = u_K^n - r_K \sum_{e\in\partial K} |e| \, g_{K,e}(u_K^n, u_{K_e}^n)$$

which may be written, thanks to the consistence of $g_{K,e}$ with $f_{e,K}$,

$$\bar{u}_K^{n+1} = u_K^n + r_K \sum_{e\in\partial K} |e| \, C_{K,e}^n \, \Delta_e u_K^n \tag{4.19}$$

where we have set :

$$\Delta_e u_K^n = u_{K_e}^n - u_K^n$$

and $$C_{K,e}^n = \left[g_{K,e}(u_K^n, u_K^n) - g_{K,e}(u_K^n, u_{K_e}^n) \right] / \Delta_e u_K^n \tag{4.20}$$

We know that under the CFL stability condition (3.2), $C_{K,e}^n$ satisfies the following inequality

$$r_K P_K C_{K,e}^n \le 1/2 - \varepsilon \tag{4.21}$$

$\bar{u}_K^{n+1}$ is given by a first order accurate finite volume scheme with flux $g_{K,e}$. For the following, we set

$$\varepsilon_{K,e}^n = g_{K,e}(u_K^n(e),u_{K_e}^n(e)) - g_{K,e}(u_K^n,u_{K_e}^n) \tag{4.22}$$

and will assume that

$$|\varepsilon_{K,e}^n| \le C\, h^\alpha\, Q_e^n \tag{4.23}$$

Relation (4.23) is a "limitation" of the slopes which is not very restrictive and which acts only in zones of discontinuities of the solution to avoid oscillations. It is simillar to the "classical" limitation given by : $|\varepsilon_{K,e}^n| \le H\, h^\alpha$, where H is a constant independent of h.

To prove theorem 4.1. we need the following lemma.

Lemma 4.1.

Let $\{u^h\}$ be a sequence given by the scheme (4.2)-(4.3), where $g_{K,e}$ is an E-flux and u_o is in $L^\infty(\Omega)$. Assume that $\{u^h\}$ satisfies the L^∞-estimate (4.12). Let η be the entropy of equation (1.1) given by $\eta(u)=u^2/2$, then under the CFL condition (3.2), $\{u^h\}$ satisfies the following inequality for every n in $\mathbb{N}$:

$$\sum_{K\in T_h} [\eta(\bar{u}_K^{n+1}) - \eta(u_K^n)]\,|K| \le -\varepsilon \sum_{e\in T_h} \tau\,|e|\,|\Delta_e^n(u)|^2\, Q_e^2(u_K^n,u_{K_e}^n) \tag{4.24}$$

where $Q_e(u,v)$ is the numerical viscosity given by (3.15).

Proof of lemma 4.1.

The sequence $(\bar{u}_K^{n+1})$ is given by a first order accurate finite volume scheme with flux $g_{K,e}$, then under the CFL stability condition (3.2) it satisfies relation (3.58), which we write for every n in $\mathbb{N}$

$$\varepsilon \sum_e \tau\,|e|\,|\Delta_e^n(u)|^2\, Q_e^2(u_K^n,u_{K_e}^n) \le -\sum_{K\in T_h} [\eta(\bar{u}_K^{n+1}) - \eta(u_K^n)]\,|K| - \tau \sum_{K,e} |e|\,\Delta_{K,e}^n(F) \tag{4.25}$$

where F is an entropy flux associated with η. The conservativity of the numerical entropy flux gives

$$\sum_{K,e} |e|\,\Delta_{K,e}^n(F) = \sum_{K,e} |e|\,\Pi_{K,e}^n = 0$$

This gives the required inequality, and ends the proof of lemma 4.1.

Thanks to inequality (4.24) of lemma 4.1. we will derive in the following lemma a "weak" estimate which is equivalent to estimate (3.55) of theorem 3.2. in section 3.

Lemma 4.2.

Let $\{u^h\}$ be a sequence given by the finite volume scheme (4.2)-(4.3), where $g_{K,e}$ is an E-flux and u_o is in $L^\infty(\Omega)$. Assume that $\{u^h\}$ satisfies the L^∞-estimate (4.12). Under the CFL condition (3.2), $\{u^h\}$ satisfies the following inequality :

$$\sum_n \sum_{e\in T_h} \tau\,|e|\,|\Delta_e^n(u)|^2\, Q_e^2(u_K^n,u_{K_e}^n) \le C_o\, h^{2\alpha-1} \tag{4.26}$$

where C_o is a constant indepent of h.

Proof of lemma 4.2.

We start from inequality (4.24) of lemma 4.1. which may be written as follows

$$\sum_{K\in T_h} \left[\eta(u_K^{n+1}) - \eta(u_K^n)\right] |K| \le -\varepsilon \sum_{e\in T_h} \tau\,|e|\,|\Delta_e(u)|^2\, Q_e^2(u_K^n, u_{K_e}^n) + \sum_{K\in T_h} \left[\eta(u_K^{n+1}) - \eta(\bar{u}_K^{n+1})\right] |K| \qquad (4.27).$$

Let us transform the last term in the right hand side of (4.27). The convexity of η gives ($\eta(u)=u^2/2$)

$$\eta(u_K^{n+1}) - \eta(\bar{u}_K^{n+1}) \le \eta'(u_K^{n+1})\,[u_K^{n+1} - \bar{u}_K^{n+1}] = u_K^{n+1}\,[u_K^{n+1} - \bar{u}_K^{n+1}]$$

If we set

$$B = \sum_{K\in T_h} \left[\eta(u_K^{n+1}) - \eta(\bar{u}_K^{n+1})\right] |K|$$

then thanks to (4.18) and (4.22) we get

$$B \le C$$

where

$$C = -\sum_{K\in T_h} u_K^{n+1}\,|K|\,r_K \sum_{e\in\partial K} |e|\,\varepsilon_{K,e}^n$$

$$= -\sum_{K\in T_h} u_K^{n+1}\,\tau \sum_{e\in\partial K} |e|\,\varepsilon_{K,e}^n$$

This term C may be written by summing on the edges of the mesh, it gives

$$C = \sum_{e\in T_h} \varepsilon_{K,e}^n\,\tau\,|e|\,\Delta_e u_K^{n+1} \qquad (4.28)$$

where we have set : $\Delta_e u_K^{n+1} = [u_{K_e}^{n+1} - u_K^{n+1}]$. By (4.18) and (4.22) we have

$$\Delta_e u_K^{n+1} = u_{K_e}^n + r_{K_e} \sum_{e'\in\partial K_e} |e'|\,C_{K_e,e'}^n\;\Delta_{e'} u_{K_e}^n - r_{K_e} \sum_{e'\in\partial K_e} |e'|\,\varepsilon_{K_e,e'}^n$$

$$- u_K^n - r_K \sum_{e''\in\partial K} |e''|\,C_{K,e''}^n\,\Delta_e u_K^n + r_K \sum_{e''\in\partial K} |e''|\,\varepsilon_{K,e''}^n$$

$$= \Delta_e u_K^n \left\{1 + r_{K_e}\,|e|\,C_{K_e,e}^n - r_K\,|e|\,C_{K,e}^n\right\} - |e|\left\{r_{K_e}\,\varepsilon_{K_e,e}^n - r_K\,\varepsilon_{K,e}^n\right\}$$

$$+ r_{K_e} \sum_{\substack{e'\in\partial K_e \\ e'\neq e}} |e'|\,C_{K_e,e'}^n\;\Delta_{e'} u_{K_e}^n - r_{K_e} \sum_{\substack{e'\in\partial K_e \\ e'\neq e}} |e'|\,\varepsilon_{K_e,e'}^n$$

$$- r_K \sum_{\substack{e''\in\partial K \\ e''\neq e}} |e''|\,C_{K,e''}^n\,\Delta_{e'} u_K^n + r_K \sum_{\substack{e''\in\partial K \\ e''\neq e}} |e''|\,\varepsilon_{K,e''}^n$$

Replacing in C and using (4.23) we get

$$C \le \sum_{e\in T_h} |\varepsilon_{K,e}^n|\,\tau\,|e|\,\Delta_e u_K^n \left\{1 + 1/2 - \varepsilon + 1/2 - \varepsilon\right\} + \sum_{e\in T_h} \tau\,|e|^2\,|\varepsilon_{K,e}^n|\left\{r_{K_e}\,\varepsilon_{K_e,e}^n + r_K\,\varepsilon_{K,e}^n\right\}$$

$$+ \sum_{e\in T_h} \tau\,|e|\,r_{K_e} \sum_{\substack{e'\in\partial K_e \\ e'\neq e}} |e'|\,C_{K_e,e'}^n\;\Delta_{e'} u_{K_e}^n + \sum_{e\in T_h} \tau\,|e|\,r_{K_e} \sum_{\substack{e'\in\partial K_e \\ e'\neq e}} |e'|\,\varepsilon_{K_e,e'}^n$$

$$+\sum_{e\in T_h}\tau|e|\,r_K\sum_{\substack{e''\in\partial K\\ e''\neq e}}|e''|\,C^n_{K,e''}\,\Delta_e u^n_K+\sum_{e\in T_h}\tau|e|\,r_K\sum_{\substack{e''\in\partial K\\ e''\neq e}}|e''|\,\varepsilon^n_{K,e''}$$

Let : $M=\sup_{n,e}(Q^n_e)$, $\delta=\sup_{K,e}(r_K|e|)$ and n_e = the maximum of the number of edges of an element of the mesh. With theses notations we get

$$C\le 2\sum_{e\in T_h}\tau|e|\,\Delta_e u^n_K\,H\,Q^n_e\,h^\alpha+n_e\sum_{e\in T_h}\tau|e|\,\delta M^2H^2h^{2\alpha}$$

$$+2\delta n_e\sum_{e\in T_h}\tau|e|\,\Delta_e u^n_K\,H\,Q^n_e\,h^\alpha \qquad (4.29)$$

On the other hand, thanks to assumptions (3.1) made on the mesh, one can easily prove that

$$\sum_{e\in T_h}|e|\le L/h \qquad (4.30)$$

where L is a constant independent of h. Now summing (4.27) on $n\le n_T=[T/\tau]$, and using (4.29) and (4.30) we get

$$\sum_{K\in T_h}\eta(u^{nT}_K)+\varepsilon\sum_{n,e}\tau|e|\,|\Delta^n_e(u)|^2\,Q^2_e(u^n_K,u^n_{K_e})\le\sum_{K\in T_h}\eta(u^o_K)+D\sum_{n,e}\tau|e|\,\Delta_e u^n_K\,Q^n_e\,h^\alpha$$

$$+n_e\sum_{n,e}\tau|e|\,\delta M^2H^2h^{2\alpha}$$

$$\sum_{K\in T_h}\eta(u^{nT}_K)+\varepsilon\sum_{n,e}\tau|e|\,|\Delta^n_e(u)|^2\,Q^2_e(u^n_K,u^n_{K_e})\le C_1\sum_{n,e}\tau|e|\,\Delta_e u^n_K\,Q^n_e\,h^\alpha$$

$$+\sum_{K\in T_h}\eta(u^o_K)+C_2\,T\,h^{2\alpha-1} \qquad (4.31)$$

where C_1 and C_2 are positives constants independent of h.

Using the Holder inequality we get

$$\sum_{n,e}\tau|e|\,\Delta_e u^n_K\,Q^n_e\,h^\alpha\le h^{\alpha-1/2}\left(\sum_{n,e}\tau|e|\right)^{1/2}\left(\sum_{n,e}\tau|e|\,(\Delta_e u^n_K)^2\,(Q^n_e)^2\right)^{1/2}$$

$$\le C_3\,h^{\alpha-1/2}\left(\sum_{n,e}\tau|e|\,(\Delta_e u^n_K)^2\,(Q^n_e)^2\right)^{1/2}$$

which inserted in (4.31) gives

$$D\le C_3\,h^{\alpha-1/2}\,D^{1/2}+E \qquad (4.32)$$

where we have set

$$D=\sum_{n,e}\tau|e|\,(\Delta_e u^n_K)^2\,(Q^n_e)^2$$

and E depend on u_o and is bounded. From (4.32) one can easily deduce the following inequality

$$D\le C_4\,h^{2\alpha-1} \qquad (4.33)$$

where C_4 is independent of h. This ends the proof of lemma 4.2.

Proof of theorem 4.1.

Let (U,F_1,F_2) be an entropy for equation (1.1). Let us write scheme (4.2.a) as follows

$$u_K^{n+1} = \bar{u}_K^{n+1} - r_K \sum_{e\in\partial K} |e| \left[g_{K,e}(u_K^n(e),u_{K_e}^n(e)) - g_{K,e}(u_K^n,u_{K_e}^n) \right] \tag{4.34}$$

with

$$\bar{u}_K^{n+1} = u_K^n - r_K \sum_{e\in\partial K} |e|\, g_{K,e}(u_K^n,u_{K_e}^n) \tag{4.35}$$

$\bar{u}_K^{n+1}$ is given by a first order accurate finite volume scheme with flux $g_{K,e}$, then we have the following discrete entropy inequality

$$U(\bar{u}_K^{n+1}) - U(u_K^n) + r_K \sum_{e\in\partial K} |e|\, \Pi_{K,e}(u_K^n,u_{K_e}^n) \le 0 \tag{4.36}$$

where $\Pi_{K,e}$ is a numerical entropy flux consistent with $F_{e,K}$.

Let us now derive a similar result for the sequence (u_K^{n+1}) given by the scheme (4.2.a). We can write inequality (4.36) in the following form

$$U(u_K^{n+1}) - U(u_K^n) + r_K \sum_{e\in\partial K} |e|\, \Pi_{K,e}(u_K^n,u_{K_e}^n) \le U(u_K^{n+1}) - U(\bar{u}_K^{n+1}) \tag{4.37}$$

Since U is convex, we have :

$$U(u_K^{n+1}) - U(\bar{u}_K^{n+1}) \le U'(u_K^{n+1})(u_K^{n+1} - \bar{u}_K^{n+1})$$

which inserted in (4.37) and thanks to (4.34) gives

$$U(u_K^{n+1}) - U(u_K^n) + r_K \sum_{e\in\partial K} |e|\, \Pi_{K,e}(u_K^n,u_{K_e}^n) \le$$

$$U'(u_K^{n+1})\, r_K \sum_{e\in\partial K} |e| \left[g_{K,e}(u_K^n(e),u_{K_e}^n(e)) - g_{K,e}(u_K^n,u_{K_e}^n) \right] \tag{4.38}$$

Let now φ be a positive function in $C^2(R^+xR^2)$ with compact support in $]0,+\infty[xR^2$. Set :

$$\varphi_K^n = \inf_{(t,x)\in I_{K,n}} \varphi(t,x).$$

After multiplication by $|K|\varphi_K^n$ and summation on all n and K such that $I_{K,n}\cap Supp(\varphi)\neq\emptyset$ (where $Supp(\varphi)$ denotes the support of φ), inequality (4.38) gives

$$\sum_{n,K} \left[U(u_K^{n+1}) - U(u_K^n) \right] |K|\, \varphi_K^n + \tau \sum_{n,K} \sum_{e\in\partial K} \varphi_K^n\, |e|\, \Pi_{K,e}(u_K^n,u_{K_e}^n) \le G^h \tag{4.39}$$

with

$$G^h = \tau \sum_{n,K} \sum_{e\in\partial K} \varphi_K^n\, |e|\, U'(u_K^{n+1}) \left[g_{K,e}(u_K^n,u_{K_e}^n) - g_{K,e}(u_K^n(e),u_{K_e}^n(e)) \right] \tag{4.40}$$

Using the same technique that in section 3, we can derive the following inequality satisfied by the sequence $\{u^h\}$

$$\iint_{R_+xR^2} \left[U(u^h)\, \partial_t\varphi(t,x) + F(u^h)\, \nabla_x\varphi \right] dx\, dt \ge R_1^h + R_2^h - G^h \tag{4.41}$$

where R_1^h and R_2^h are defined, like in section 3, by

$$R_1^h = \sum_{n,K} U(u_K^{n+1}) \int_{t^n}^{t^{(n+1)}} \int_K \left[\partial_t\varphi - \frac{1}{\tau}\{\varphi_K^{n+1} - \varphi_K^n\}\right] dt\, dx \tag{4.42}$$

and

$$R_2^h = \sum_{n,K} \sum_{e\in\partial K} \int_{t^n}^{t^{(n+1)}} \int_e \Pi_{K,e}(u_K^n, u_{K_e}^n)\,(\varphi_K^n - \varphi(t,x))\, d\Gamma dt \tag{4.43}$$

By definition of the Young measure ν associated with u^h, the left side of (4.41) converges to :

$$\iint_{R_+ xR^2} \left\{ <\nu,U>\, \partial_t\varphi + <\nu,F(u)>\, \nabla_x\varphi \right\} dx\, dt$$

It thus remains to prove that the right hand side of (4.41) satisfy

$$\lim_{h\to 0} (R_1^h + R_2^h - G^h) \geq 0 \tag{4.44}$$

The term R_1^h converges to 0 with h since φ is of class C^2 and by using lemma.3.2. of section 3.

We now treat of R_2^h defined by (4.43), we have :

$$R_2^h = \sum_{n,K} \sum_{e\in\partial K} \tau\, |e|\, \Pi_{K,e}(u_K^n, u_{K_e}^n)\, \Phi_{K,e}^n \tag{4.45}$$

with

$$\Phi_{K,e}^n = \varphi_K^n - \frac{1}{\tau|e|} \int_{t^n}^{t^{(n+1)}} \int_e \varphi(t,x)\, d\Gamma dt$$

Using the same technic that in section 3 and the estimate (4.26) of lemma 4.2. we get

$$\left| \sum_{n,K} \sum_{e\in\partial K} \tau\, |e|\, \Pi_{K,e}(u_K^n, u_{K_e}^n)\, \Phi_{K,e}^n \right| \leq C\, h^{1/2} \sum_{n,e} \tau\, |e|\, (\Delta_e u_K^n)^2\, (Q_e^n)^2$$

$$\leq C\, h^{2\alpha - 1/2} \tag{4.46}$$

where C is a constant independent of h.

Using the same technique we prove that : $\lim_{h\to 0} G^h \leq 0$. This ends the proof of theorem 4.1.

REFERENCES BIBLIOGRAPHIQUES

[BLN] C.BARDOS, A.Y. LEROUX & J.C. NEDELEC, First order quasilinear equations with boundary conditions, Comm. P.D.E. 4 (9), pp. 1017-1034, 1979.

[BCV.1] S. BENHARBIT, A. CHALABI, J.P. VILA, Convergence of finite difference methods for hyperbolic conservation laws with boundary conditions, to appear.

[BCV.2] S. BENHARBIT, A. CHALABI, J.P. VILA, Convergence of finite volume methods for hyperbolic conservation laws with boundary conditions, to appear.

[CGH] S. CHAMPIER, T. GALLOUET, R. HERBIN Convergence of an upstream finite volume scheme for a non linear hyperbolic equation on a triangular mesh Les prepublications du L.A.M.A., 91-08 (1991).

[CCL] B. COCKBURN, F. COQUEL, Ph. LeFLOCH, C.W. SHU, Convergence of finite volume methods.

[CL.1] F. COQUEL, Ph LeFLOCH, Convergence of finite differences schemes for conservation laws in several space dimensions : General theory, S.I.A.M. Numer. Anal. (1989).

[CL.2] F. COQUEL, Ph LeFLOCH, Convergence of finite differences schemes for conservation laws in several space variables : application to the corrected antidiffusive fluxes method. Math. of Comp. (1990).

[CM] M. CRANDALL, A. MAJDA, Monotone difference approximations for scalar conservation laws, Math. of Comp. vol.34, (1980), pp.1-21.

[DP.1] R.J. DI PERNA, Convergence of approximate solutions to conservation laws, Arch. Rat. Mech. Anal. 82, (1983), pp. 27-70.

[DP.2] R.J. DI PERNA, Measure-valued solutions for conservation laws, Arch. Rat. Mech. Anal. 88, (1985), pp. 223-270.

[Fez] F. FEZOUI, Résolution des équation d'Euler par un schéma de Van Leer en éléments finis, Rapport de recherche n° 358 INRIA Rocquencourt (1985).

[GH] T. GALLOUET, R. HERBIN, A uniqueness result for measure valued solutions of nonlinear hyperbolic equations, Les prepublications du L.A.M.A., 92-04b (1992).

[Gl] P. GLAISTER, An approximate linearised Riemann solver for the Euler equations for real gases, Jour. Comp. Phys., 74 (1988), pp. 382-408.

[God] S.K. GODOUNOV, A difference method for numerical calculation of discontinous equations of hydrodynamics, Mat. Sb., 47 (89), (1959), 271-300.

[Her] J.M. HERARD, Ecoulement transitoire compressible dans un puit de cuve, Note HE-41/89.43.

[Kru] S.N. KRUZKOV, First order quasi-linear equations in several variables, Math. USSR Sbornik, 10, (1970), pp. 217-243.

[Ler] LEROUX, Thèse d'état, Université de Rennes - 1979.

[Lit] LI TA-TSIEN, Problèmes aux limites et solutions discontinues pour les systèmes hyperboliques quasi-linéaires d'ordre 1, Séminaires collège de France.

[Med] L. MEDARO, Les explosifs occasionnels, Ed. Techniques et documentation, Lavoisier.

[Osh] S. OSHER, Riemann solvers, the entropy condition, and difference approximations, SIAM J. Numer. Anal., 21, (1984), pp. 217-235.

[Roe] P.L. ROE, Approximate Riemann Solvers, parametrer vectors, and difference schemes, J. Comp. Phys., 43 (1981), pp. 357-372.

[RP] P.L. ROE, J. PIKE, Comp. Meth. App. Sc. Eng. VI, Elsevier Sc. Pub., (1984), pp. 449-518.

[San] R. SANDERS, On convergence of monotone finite difference schemes with variable spatial differencing, Math. of Comp., 40, (1983), pp. 91-106.

[Sch] M. SCHATZMAN, Introduction à l'analyse des systèmes hyperboliques de lois de conservation non-linéaires, (Publication de l'université de Lyon (1985)).

[Smi] R.G. SMITH, The Riemann problem in gaz dynamics, Trans. Am. MAth. Soc., Vol. 249, 2, pp. 1-50.

[Smo] J.A. SMOLLER, Shocks waves. Reaction-Diffusion equations, Springer Verlag, N.Y. (1983)

[Sze] A. SZEPESSY, Measure valued solutions of scalar conservation laws with boundary conditions, Arch. Rational Mech. Anal. 107, n° 2, 1989, pp. 181-193.

[Tad] E. TADMOR, Numerical viscosity and the entropy condition for consevative difference schemes, Math. of Comp., 43, (1984), pp. 369-381.

[To] I. Toumi , Thèse de Doctorat de l'Université Paris VI, (1989).

[Vil.1] J.P. VILA, Thèse de Doctorat de l'Université Paris VI, (1986).

[Vil.2] J.P. VILA, An analysis of a class of second-order accurate Godounov-type schemes, SIAM J. NUMER. ANAL., vol 26, n° 4, pp. 830-850, (1989).

[Vil.3] J.P. VILA, Construction of Roe type matrices, Proceeding of Third International Conference on Hyperbolic Problems, Ed. by Engquist and Gustafsson, Uppsala, Sweden (1990), pp. 913-922.

[Van] VAN LEER B., "Towards the ultimate conservative difference scheme, V", J. Comp. Phys., 32 (1979), 101-136.