Lecture Notes in Mathematics

Edited by A. Dold, B. Eckmann and F. Takens

1402

C. Carasso P. Charrier
B. Hanouzet J.-L. Joly (Eds.)

Nonlinear Hyperbolic Problems

Proceedings of an Advanced Research Workshop
held in Bordeaux, France, June 13–17, 1988

Springer-Verlag
Berlin Heidelberg New York London Paris Tokyo Hong Kong

Editors

Claude Carasso
Faculté des Sciences et Techniques
23 rue du Docteur P. Michelon
42023 Saint-Etienne Cedex 2, France

Pierre Charrier
Bernard Hanouzet
Jean-Luc Joly
U.E.R. de Mathématiques, Université de Bordeaux 1
351 Cours de la libération, 33405 Talence Cédex, France

Mathematics Subject Classification (1980): 35 L, 65 M, 70 N

ISBN 3-540-51746-4 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-51746-4 Springer-Verlag New York Berlin Heidelberg

Printing and binding: Druckhaus Beltz, Hemsbach/Bergstr.
2146/3140-543210 – Printed on acid-free paper

Organisé dans le cadre de l'année spéciale sur les phénomènes non linéaires, sous le patronage du CNRS et du MEN, le Colloque sur les problèmes hyperboliques non linéaires s'est tenu à Bordeaux, sur le campus de l'Université, dans les locaux de l'Amphi. Kastler du Lundi 13 Juin au Vendredi 17 Juin 1988.

Le Colloque a réuni 120 participants de 13 nationalités différentes (Algérie, Allemagne fédérale, Angleterre, Belgique, Cameroun, Chine, Etats-Unis d'Amérique, France, Italie, Israël, Japon, Pologne, Suède) dans la proportion de 80 participants français et 40 étrangers. On doit aussi remarquer qu'il y avait 28 participants non universitaires en provenance des secteurs industriels aussi bien public que privé.

Le financement du Colloque a été assuré grâce au concours de divers organismes CNRS-MEN, le Ministère des Affaires Etrangères, le Conseil Régional d'Aquitaine, l'Université de Bordeaux I, la SMAI, le GAMNI, la SMF, le CEA/CELV, le CEA/CESTA, la DRET, le CIMPA, le SERAM.

Le but scientifique du Colloque était de favoriser les interactions entre les multiples aspects de l'étude des ondes hyperboliques non linéaires.

Les ondes hyperboliques non linéaires modélisent en effet de très nombreuses situations physiques (dynamique des fluides, élastodynamique, écoulements réactifs, théorie des champs...) et sont à la base d'applications scientifiques et industrielles très importantes (aéronautique, industrie pétrolière, détonique, calcul d'impact, combustion...).

Les aspects tant théoriques que numériques de ces problèmes sont très variés et imbriqués : équations aux dérivées partielles, analyse, analyse numérique, géométrie.

Nous avons donc tenté de choisir des conférenciers représentatifs de toutes les tendances de façon à attirer des spécialistes du plus grand nombre de domaines de l'hyperbolique.

Les 21 conférences ont été ainsi données uniquement sur invitation du Comité scientifique du Colloque constitué de :MM. C. Bardos, S. Klainerman, A.Y. Le Roux, A. Majda, S. Osher, J. Rauch et du comité d'organisation formé de : MM. C. Carasso, P. Charrier, B. Hanouzet et J.L. Joly.

Le présent document a été établi en exécution du contrat n°88 1309 par la Direction des Recherches Etudes et Techniques Direction Scientifique.

Section Soutien à la Recherche Délégation Générale pour l'Armement.

CONTENTS

1- *Numerical Analysis*

a- General theory.

S. Benachour, A.-Y. Le Roux et M.-N. Le Roux : Approximation to nonlinear convection diffusion equations. 1

A. Lerat : Difference schemes for nonlinear hyperbolic systems - A general framework. 12

b- Main applications.

M. Ben-Artzi : Numerical calculations of reacting flows. 30

A. Bourgeade, H. Jourdren et J. Ovadia : Problème de Riemann en hydrodynamique et applications. 37

D. d'Humières, P. Lallemand and Y.H. Qian : Review of flow simulations using lattice gases. 56

B. Larrouturou and L. Fezoui : On the equations of multi-component perfect of real gas inviscid flow. 69

2 -*Hyperbolic P.D.E. theory*

A. Bachelot : Global existence of large amplitude solutions for Dirac-Klein-Gordon systems in Minkowski space. 99

J.-M. Bony : Analyse microlocale et singularités non linéaires. 114

D. Christodoulou and S. Klainerman : The nonlinear stability of the Minkowski metric in General Relativity. 128

R. De Vore and B. Lucier : High order regularity for solution of the inviscid Burger's equation. 147

F. John : Solutions of quasi-linear wave equations with small initial data. 155

B. Lee Keyfitz and H.C. Kranzer : A viscosity approximation to a system of conservation laws with no classical Riemann solution. 185

Li Ta-tsien and Chen Yun-mei : Global classical solutions to the Cauchy problem for nonlinear wave equations. 198

G. Métivier : Ondes de choc, ondes de raréfaction et ondes soniques multidimensionnelles. 203

G. Métivier et J. Rauch : The interaction of two progressing waves. 216

R. Rosales : Diffraction effects in weakly nonlinear detonation waves. 227

3 - *List of participants* 240

4 - *List of talks* 248

APPROXIMATION TO NONLINEAR CONVECTION DIFFUSION EQUATIONS

Said BENACHOUR

Institut de Mathématiques,
Université des Sciences et Techniques Houari Boumedienne
BP 9, Dar El Beida, Alger, ALGERIA

Alain Yves LE ROUX

and Marie Noëlle LE ROUX

UER Mathématiques et Informatique,
Université de Bordeaux 1,
33405 Talence, FRANCE.

We try to build a strong numerical method for convection diffusion equations in the nonlinear case, which gives L^∞ and Bounded Variation (=BV) stability on the gradient of the solution. This leads to a compactness argument in L^∞ for the approximate solution and then to a proof of convergence on a nonlinear diffusion term. Several examples are reported in order to show that hyperbolic techniques are suitable for such nonlinear parabolic models.

This paper is divided into 3 parts. A first example is detailed in Section 1, where the numerical method is described in a very simple way. Then the same method is adapted to the porous media equation in Section 2. Next, Section 3 is devoted to the two dimension case, including some numerical techniques adapted to the diffusion term. Then a Riemann solver is proposed for the first order term, which comes from the derivation of the diffusion term. This leads to a two dimension version of the Lax Friedrichs scheme, and a construction of the Godunov scheme using the same Riemann solver.

Other numerical methods presents same properties of stability, such as the one proposed in [6],[8],[9],[10] or [12]. However, the mathematical model studied here deals with the equation of velocity and the schemes proposed here too.

1.- AN EXAMPLE - We consider the equation

$$u_t = \left(u\, u_x \right)_x \tag{1}$$

together with the initial condition

$$u(x,0) = u_0(x)$$

where

$$u_0 \in W^{1,\infty}(\mathbb{R}) \ , \ u_0 \geq 0 \ , \quad \text{with compact support.}$$

For $i \in \mathbb{Z}$ and $n \in \mathbb{N}$, we denote by u_i^n the approximate value of $u(i\Delta x, n\Delta t)$, for a space increment Δx and a time increment Δt.
By the Euler scheme, we get

$$u_i^{n+1} = u_i^n + \frac{\Delta t}{2\Delta x^2}\Big[(u_{i+1}^n + u_i^n)(u_{i+1}^{n+1} - u_i^{n+1}) - (u_i^n + u_{i-1}^n)(u_i^{n+1} - u_{i-1}^{n+1})\Big] \qquad (2)$$

which leads to the estimates, provided all u_i^n to be non negative,

$$\forall\ i \in \mathbb{Z} \quad , \quad u_i^{n+1} \geq 0 \quad ,$$

$$\operatorname*{Max}_{i\in\mathbb{Z}} |u_i^{n+1}| \leq \operatorname*{Max}_{j\in\mathbb{Z}} |u_j^n| ,$$

and

$$\sum_{i\in\mathbb{Z}} | u_{i+1}^{n+1} - u_i^{n+1} | \leq \sum_{j\in\mathbb{Z}} | u_{j+1}^n - u_j^n | \quad .$$

This means that the scheme repserves the positiveness of u , and is L^∞ and BV (=Bounded Variation) stable (or is TVD, that is Total Variation Diminishing).

Let ψ be a test function in $C^2(\mathbb{R} \times \mathbb{R}_+)$. We set

$$\psi_i^n = \psi(i\Delta x, n\Delta t)$$

and

$$u_{i+1/2}^n = \frac{1}{2}\,(\, u_i^n + u_{i+1}^n\,) \qquad .$$

Then we get, by multiplying the scheme (2) by ϕ_i^n and summing,

$$\sum_{i,n} u_i^n \, \frac{\psi_i^n - \psi_i^{n-1}}{\Delta t}\, \Delta t \Delta x = \sum_{i,n} u_{i+1/2}^n \, \frac{u_{i+1}^{n+1} - u_i^{n+1}}{\Delta x} \, \frac{\psi_{i+1}^n - \psi_i^n}{\Delta x} \quad \Delta t \Delta x$$

However the estimates given above are not sufficient to enable us to go to the limit (for any subsequence) on the product

$$u_{i+1/2}^n \, \frac{u_{i+1}^{n+1} - u_i^{n+1}}{\Delta x} \qquad .$$

We need another estimate, which can be the uniform convergence on u . This can be deduced from L^∞ and BV estimates on u_x instead of u as above.

In order to get it, we set

$$v = - u_x \;.$$

Then we get the equation

$$v_t + (v^2)_x = (u\ v_x)_x \;. \qquad (3)$$

This equation will be discretized into two steps. The first one is devoted to the second order term, which correspond to a diffusion term. We set m = n + 1/2 , which will be used as the upper index for an intermadiary value between the times nΔt and (n+1)Δt. We compute

$$v_i^m = v_i^n + \frac{\Delta t}{\Delta x}2 \left[u^n_{i+1/2}(v^m_{i+1}-v^m_i) - u^n_{i-1/2}(v^m_i-v^m_{i-1}) \right] \qquad (4)$$

As above, this scheme preserves L^∞ and BV estimates for v . It is now sufficient to use a L^∞ and BV stable scheme for the discretization of the first order term. This can be done by using the Godunov scheme.

This scheme uses a Riemann solver associated with the scalar equation

$$v_t + (v^2)_x = 0 \qquad (5)$$

which allows to compute the fluxes on both sides of the cells.
This is performed as follows. We compute, for any i ∈ ℤ ,

$$v^m_{i+1/2} = \begin{cases} v^m_i & \text{if } v^m_i \geq 0 \text{ and } v^m_i + v^m_{i+1} \geq 0 , \\ 0 & \text{if } v^m_i \leq 0 \text{ and } v^m_{i+1} \geq 0 , \\ v^m_{i+1} & \text{if } v^m_{i+1} \leq 0 \text{ and } v^m_i + v^m_{i+1} \leq 0 . \end{cases}$$

and then,

$$v_i^{n+1} = v_i^m - \frac{\Delta t}{\Delta x} \left[(v^m_{i+1/2})^2 - (v^m_{i-1/2})^2 \right] \qquad (6)$$

This scheme is $L^\infty$and BV stable under the stability condition

$$\max_j |v_j^m| \frac{\Delta t}{\Delta x} \leq \frac{1}{2} \qquad (7)$$

We notice that this condition is the well known Courant Friedrichs Lewy condition, and the coefficient $\frac{1}{2}$ comes from the flux in (5), which is 2 v . This condition gives also the conservation of the positiveness of u , since we have

$$u^{n+1}_{i+1/2} = - \sum_{j\leq i} v_j^{n+1} \Delta x \geq 0 .$$

As a matter of fact,

$$-\sum_{j\leq i} v_j^{n+1} = - \sum_{j\leq i} v_j^m + \frac{\Delta t}{\Delta x} (v^m_{i+1/2})^2 \geq - w^m_{i+1/2}$$

by writing

$$w^m_{i+1/2} = \sum_{j\leq i} v_j^m .$$

And since we have

$$w^m_{i+1/2} = w^n_{i+1/2} + \frac{\Delta t}{\Delta x} u^n_{i+1/2}\left(w^m_{i+3/2} - 2\, w^m_{i-1/2} + w^m_{i-1/2}\right)$$

which is a linear system involving a M-matrix, we get

$$\forall\ i \in \mathbb{Z} \quad w^n_{i+1/2} \le 0 \quad \Rightarrow \quad \forall\ i \in \mathbb{Z} \quad w^m_{i+1/2} \le 0 \ .$$

This proves the conservation of the positiveness.

2.- THE POROUS MEDIA EQUATION. We are now concerned with the equation

$$\frac{\partial u}{\partial t} = \Delta\ \Phi(u) \tag{8}$$

where $\Phi \in C^2(\mathbb{R})$ is a nondecreasing function such that $\Phi'(0)=0$. We set

$$\phi(u) = \Phi'(u)\ ,$$

and

$$v = -\ \frac{\phi(u)}{u}\ \frac{\partial u}{\partial x}\ .$$

Then we get the convection equation

$$u_t + (\ u\ v\ )_x = 0\ . \tag{9}$$

Next we introduce

$$p(u) = \int_0^u \frac{\phi(y)}{y}\ dy$$

which corresponds physically to a pressure if u is a concentration. Then we get

$$v + p_x = 0 \tag{10}$$

which is known as the Darcy law; here v is a velocity.

From (9) and by using (10), we can derive the equation of the velocity and obtain

$$v_t + (v^2)_x = (\ \phi(u)\ v_x\ )_x \quad . \tag{11}$$

We propose a discretization of this equation.

Since ϕ is nonnegative, the previous scheme will work and we get $L^\infty$and BV estimates for v . We notice that in the first step (i.e. the discretization of the second order term), we only have to change $u^n_{i+1/2}$ into $\phi(u^m_{i+1/2})$, which can be written as a function of the pressure p by using the Darcy law (10). This will be denoted $d(p) = \phi(u)$, and for example

$$d(p) = \frac{p}{k} \qquad \text{if} \qquad \phi(u) = u^k .$$

This is possible only when the positiveness of the pressure is preserved during the two steps of the scheme. By writing, from the Darcy law (10),

$$p^m_{i+1/2} = p^m_{i-1/2} - v^m_i \, \Delta x \tag{12}$$

and

$$\phi^n_{i+1/2} = d(p^n_{i+1/2}) \; ,$$

we get

$$p^m_{i+1/2} = p^n_{i+1/2} + \frac{\Delta t}{\Delta x} \phi^n_{i+1/2} \left[p^m_{i+3/2} - 2p^m_{i+1/2} + p^m_{i-1/2} \right]$$

which involves a M-matrix. Then we get

$$\forall \; j \in \mathbb{Z} \;\; p^n_{j+1/2} \geq 0 \quad \Rightarrow \quad \forall \; i \in \mathbb{Z} \;\; p^m_{i+1/2} \geq 0 \; .$$

Next we have, by computing $p^{n+1}_{i+1/2}$ from the v^{n+1}_j as for the intermediary values in (12),

$$p^{n+1}_{i+1/2} = p^m_{i+1/2} + \Delta t \; (v^m_{i+1/2})^2 \; \geq \; p^m_{i+1/2} \; \geq 0 \; .$$

From these estimates we can deduce the convergence of a subsequence, from a compactness argument, towards a weak solution which can be defined, for example, as follows.

For any test function ψ with a compact support in $\mathbb{R}\times\mathbb{R}_+$, v and p satisfy

$$\iint_{\mathbb{R}\times\mathbb{R}_+} (\, v \, \psi_t + v^2 \psi_x \,) \, dx \, dt = \iint_{\mathbb{R}\times\mathbb{R}_+} (d'(p) \, v^2 \, \psi_x - d(p) \, v \, \psi_{xx}) \, dx \, dt$$

and

$$\iint_{\mathbb{R}\times\mathbb{R}_+} (\, v \, \psi - p \, \psi_x) \, dx \, dt = 0 \; .$$

Here the convergence on each product is possible since we have a uniform convergence for p and a strong L^1 convergence for v.

3. - THE TWO DIMENSION CASE. For a given non negative function ϕ in $C^1(\mathbb{R})$, with $\phi(0) = 0$, we consider the two dimension equation

$$w_t = \operatorname{div}(\phi(w)\ \nabla w) \tag{13}$$

The two space variables will be denoted x and y. We set

$$V = \begin{pmatrix} u \\ v \end{pmatrix} = - \frac{\phi(w)}{w} \nabla w.$$

Then for

$$p = \int_0^w \frac{\phi(\xi)}{\xi}\, d\xi \tag{14}$$

we obtain the Darcy law

$$V + \nabla p = 0, \tag{15}$$

and the convection equation

$$w_t + \operatorname{div}(w V) = 0. \tag{16}$$

Here p corresponds to a pressure, V is a velocity field and w is a concentration. This is physically meaningful when p and w has nonnegative values.

As above in the one dimension case, we compute the time derivative of V . We obtain successively,

$$\begin{aligned} V_t &= - \nabla p_t && \text{from (15)} , \\ &= - \nabla \left\{ \frac{\phi(w)}{w}\ w_t \right\} && \text{by using (14)} , \\ &= + \nabla \left\{ \frac{\phi(w)}{w}\ \operatorname{div}(w V) \right\} && \text{by using (16)} . \end{aligned}$$

From

$$\operatorname{div} (w V) = V . \nabla w + w \operatorname{div} (V) ,$$

we get the equation of the velovity

$$V_t + \nabla (|V|^2) = \nabla (d(p) \operatorname{div}(V)) \tag{17}$$

since from (14) we can find a function d of the pressure such that

$$d(p) = \phi(w) .$$

We propose now a two step numerical scheme for the discretization of (17). The first step deals with the second order term. We can use here a classical technique for diffusion equations.

For example, we can write

$$q = \mathrm{div}(V)$$

which satisfies

$$q_t = \Delta (d(p)\, q) .$$

This equation can be solved by using the implicit Euler method for the time discretization, and a finite difference method with a frosen d(p) for the space discretization. Then it remains to solve

$$\mathrm{div}\ (V) = q \quad ; \quad \mathrm{rot}(V) = 0 \quad ,$$

to get the velocity. Such an elliptic problem is studied in [3] or [5].

Other classical techniques can work too. Now, from this first step, we get intermediary values $V^m_{i,j}$ of the velocity field on each cell

$$M_{i,j} = \left]\ (i-\tfrac{1}{2})\Delta x\ ,\ (i+\tfrac{1}{2})\Delta x\ \right[\times \left](j-\tfrac{1}{2})\ \Delta y\ ,\ (j+\tfrac{1}{2})\Delta y\right[$$

where Δx and Δy denote the space meshsizes.

We are now concerned with the second step of the numerical scheme. This corresponds to a discretization of the non linear hyperbolic system

$$\begin{aligned} u_t + (u^2 + v^2)_x &= 0 \quad , \\ v_t + (u^2 + v^2)_y &= 0 \quad . \end{aligned} \tag{18}$$

Either for a Godunov scheme using alternated directions or for a Lax Friedrichs scheme with modified fluxes, we need a one dimension Riemann solver. In the x-direction for example, we have to solve the Riemann problem

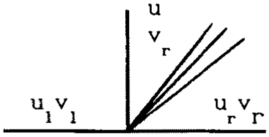
$$\begin{aligned} u_t + \ (u^2 + v^2) &= 0 \\ v_t \quad &= 0 \quad , \end{aligned} \tag{19}$$

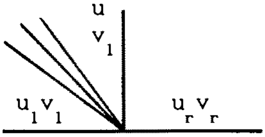
with the constant piecewise intial condition

$$(u(x,0),v(x,0)) = \begin{cases} (u_l,v_l) & \text{for} \quad x < 0 \ , \\ (u_r,v_r) & \text{for} \quad x > 0 \ , \end{cases}$$

where u_l, v_l, u_r, v_r are given real data. We can have either a wave travelling towards the right hand side (with a positive velocity), or a wave travelling towards the left hand side (with a negative velocity). This wave can be either a rarefaction wave or a shock wave.

In both cases we have a constant state (u, v_r) (in the first case) or (u, v_l) (in the second case) between the line $x=0$ and the wave.





This value u satisfies the following condition, which is a Rankine Hugoniot condition along the line $x=0$,

$$\begin{aligned} u^2 + v_r^2 &= u_l^2 + v_l^2 \qquad &&\text{(first case)} \\ u^2 + v_l^2 &= u_r^2 + v_r^2 \qquad &&\text{(second case)} \end{aligned} \tag{20}$$

For a shock wave with a positive velocity, we have necessary, from an entropy argument,

$$u + u_r > 0 \qquad \text{and} \qquad u > 0 \ .$$

Then we need in this case,

$$u_l^2 + v_l^2 \geq u_r^2 + v_r^2 \qquad \text{with} \qquad u \geq 0 \ .$$

For a rarefaction wave with a positive velocity, we have

$$0 \leq u \leq u_r$$

then we need

$$u_l^2 + v_l^2 \leq u_r^2 + v_r^2 \qquad \text{with} \qquad u \geq 0 \ .$$

For a shock wave with a negative velocity we have

$$u_l^2 + v_l^2 \leq u_r^2 + v_r^2 \qquad \text{with} \qquad u \leq 0 \ .$$

For a rarefaction wave with a negative velocity we have

$$u_l^2 + v_l^2 \geq u_r^2 + v_r^2 \qquad \text{with} \qquad u \geq 0 \ .$$

Another case can arise, when a rarefaction wave is spread on both sides of the line $x=0$. This is very seldom in practice, and we have

$$u_l < 0 < u_r \qquad \text{and} \qquad |v_l| = |v_r| \ .$$

From these remarks, we can solve the Riemann problem as follows.

If $u_l^2 + v_l^2 \geq u_r^2 + v_r^2$ then we have

if $v_l^2 \leq u_r^2 + v_r^2$ and $u_l < 0$, a rarefaction wave with a negative velocity,

else a shock wave with a positive velocity.

If $u_l^2 + v_l^2 \leq u_r^2 + v_r^2$ then we have

if $v_r^2 \leq u_l^2 + v_l^2$ and $u_r > 0$, a rarefaction wave with a positive velocity,

else a shock wave with a negative velocity.

Using this Riemann solver we are now able to compute

$$F_x(u_l, v_l, u_r, v_r) = \begin{cases} u_l^2 + v_l^2 & \text{(in the first case)} \\ u_r^2 + v_r^2 & \text{(in the second case)} \end{cases}$$

as for the first or the second case in (20), and

$$G(u, v, u, v) = \frac{2}{\Delta t \ \Delta x} \int_{-\Delta x/2}^{\Delta x/2} \int_0^{\Delta t/2} \left[u(x,t)^2 + v(x,t)^2 \right] dx\, dt \ .$$

By the same way, a Riemann solver associated with the Riemann problem

$$\begin{aligned} u_t &= 0 \ , \\ v_t + (u^2+v^2)_y &= 0 \ , \end{aligned} \tag{21}$$

$$(u(x,0), v(x,0)) = \begin{cases} (u_l, v_l) & \text{for } y < 0 \ , \\ (u_r, v_r) & \text{for } y > 0 \ . \end{cases}$$

can be built and we are also able to compute

$$F_y(u_l, v_l, u_r, v_r) \quad \text{and} \quad G_y(u_l, v_l, u_r, v_r) \ .$$

We can notice that the same Riemann solver can be used since we have only to change u into v and v into u .

Now we can write the Godunov scheme as follows,

$$F^m_{i+1/2,j} = F_x(u^m_{i+1,j}, v^m_{i+1,j}, u^m_{i,j}, v^m_{i,j}) \ ,$$

$$F^m_{i,j+1/2} = F_y(u^m_{i,j+1}, v^m_{i,j+1}, u^m_{i,j}, v^m_{i,j}) \ ,$$

$$u^{n+1}_{i,j} = u^m_{i,j} - \frac{\Delta t}{\Delta x}(F^m_{i+1/2,j} - F^m_{i-1/2,j}) \ ,$$

$$v^{n+1}_{i,j} = v^m_{i,j} - \frac{\Delta t}{\Delta y}(F^m_{i,j+1/2} - F^m_{i,j-1/2}) \ ,$$

or the modified Lax Friedrichs scheme as follows

$$G^m_{i+1/2,j} = G_x(u^m_{i+1,j}, v^m_{i+1,j}, u^m_{i,j}, v^m_{i,j})$$

$$G^m_{i,j+1/2} = G_y(u^m_{i,j+1}, v^m_{i,j+1}, u^m_{i,j}, v^m_{i,j})$$

$$\begin{aligned} u^{\mu}_{i+1/2,j+1/2} = \frac{1}{4}(u^m_{i,j} + u^m_{i,j+1} + u^m_{i+1,j} + u^m_{i+1,j+1}) \\ - \frac{\Delta t}{\Delta x}(G^m_{i+1,j+1/2} - G^m_{i,j+1/2}) \end{aligned}$$

$$\begin{aligned} v^{\mu}_{i+1/2,j+1/2} = \frac{1}{4}(v^m_{i,j} + v^m_{i+1,j} + v^m_{i,j+1} + v^m_{i+1,j+1}) \\ - \frac{\Delta t}{\Delta y}(G^m_{i+1/2,j+1} - G^m_{i+1/2,j}) \end{aligned}$$

which corresponds to new intermediary values denoted by μ), and

$$G^{\mu}_{i,j+1/2} = G_x(u^{\mu}_{i+1/2,j+1/2}, v^{\mu}_{i+1/2,j+1/2}, u^{\mu}_{i-1/2,j+1/2}, v^{\mu}_{i-1/2,j+1/2})$$

$$G^{\mu}_{i+1/2,j} = G_y(u^{\mu}_{i+1/2,j+1/2}, v^{\mu}_{i+1/2,j+1/2}, u^{\mu}_{i+1/2,j-1/2}, v^{\mu}_{i+1/2,j-1/2})$$

$$\begin{aligned} u^{n+1}_{i,j} = \frac{1}{4}(u^{\mu}_{i+1/2,j+1/2} + u^{\mu}_{i+1/2,j-1/2} + u^{\mu}_{i-1/2,j+1/2} + u^{\mu}_{i-1/2,j-1/2}) \\ - \frac{\Delta t}{\Delta x}(G^{\mu}_{i+1/2,j} - G^{\mu}_{i-1/2,j}) \end{aligned}$$

$$\begin{aligned} v^{n+1}_{i,j} = \frac{1}{4}(v^{\mu}_{i+1/2,j+1/2} + v^{\mu}_{i+1/2,j-1/2} + v^{\mu}_{i-1/2,j+1/2} + v^{\mu}_{i-1/2,j-1/2}) \\ - \frac{\Delta t}{\Delta y}(G^{\mu}_{i,j+1/2} - G^{\mu}_{i,j-1/2}) \end{aligned}$$

This scheme has been studied for the scalar equation in two dimensions and a proof of convergence is given in [2]. Here the fluxes are different from those given in [4], by Conway and Smoller. In this case the authors use an average in the two directions, which spreads out the approximate solution near a singularity.

-CONCLUSION- The idea which consists in taking out a convection term from the term of nonlinear diffusion is not really a new one (see e.g. [7]). We think that to apply this method to the equation of the velocity seems to be a new one. This can be also adapted to the case of a second arder term which is not a degenerated one or when a convection term already lies in the equation . This is the case in interdiffusion problems(see [1]), in cellular division, population dynamics and many other topics. This technique has a good behaviour near a the free boundary corresponding to the degeneration point (here for u = 0), or when a boundary condition is to be taken in account. A Riemann solver of the same conception appears in some problems of combustion and allows to introduce pointwise boundary conditions (see [11]). An antidiffusion technique adapted to this version of the Lax Friedrichs method has been analysed in [2] ; sufficient conditions for convergence are given for the quasi linear equation.

REFERENCES

[1]- D.Ansel,C.Fresnel,M.N.LeRoux, Résolution numérique d'un problème d'interdiffusion, C.R.Acad. Sc. Paris, t 306, 1, pp 143-146, 1988.

[2]- T.Boukadida, Convergences de schémas numériques adaptés à la convection non linéaire en dimension deux; Application à des couplages de modes en plasma. Thèse Bordeaux, 1988.

[3]- I.Bouvier, Un schéma tridimensionnel adapté à la propagation d'ondes élastiques. Thèse Bordeaux, 1988.

[4]- E.Conway, J.Smoller, Global solutions of the Cauchy problem for quasi linear first order equations in several space variables. Comm. Pure Applied Math. V.19, pp 95-105, 1966.

[5]- M.Crouzeix, A.Y.LeRoux, Ecoulement d'un fluide irrotationnel, Actes des Journées Eléments Finis, Rennes, 1976.

[6]- E.DiBenedetto, D.Hoff, An interface tracking algorithm for the porous medium equation, Trans. of the A.M.S., v.284, n° 2, pp 463-500, 1984.

[7]- J.L.Graveleau, P.Jamet, A finite difference appraoch to some degenerate nonlinear parabolic equations, SIAM J. of Appl. Math. V.20, pp 199-223, 1971.

[8]- M.N.LeRoux, H.Wilhelmsson, (to appear).

[9]- R.C.MacCamy, E.Scolovsky, A numerical procedure for the porous media equation, Comp. & Math with Appl. v.11, n° 1-3, pp 315-319, 1985.

[10]- M.Mimura, T.Nakaki, K.Tomoeda, A numerical approach of interface curves for some nonlinear diffusion equations, Japan J. of Appl. Math., V.1, pp 93-139, 1984.

[11]- D.Ribereau, Génération d'un logiciel de simulation de la combustion d'un bloc de propergol solide, Thèse, Bordeaux 1988.

[12]- M.Watanabe, An approach by difference to a quasi linear parabolic equation, Proc. of Japan Acad. V.59, n° 8, pp 375-378, 1983.

DIFFERENCE SCHEMES FOR NONLINEAR HYPERBOLIC SYSTEMS - A GENERAL FRAMEWORK

A. Lerat
ENSAM, 151 bd de l'Hôpital, 75013 Paris, France
and ONERA, 29 av. Div. Leclerc, 92320 Châtillon, France

ABSTRACT

For a hyperbolic system of conservation laws, the general form of conservative difference schemes involving two time-levels in an explicit or implicit way is obtained under natural assumptions. General results are shown on the schemes and this framework is used to study implicit schemes of second-order accuracy.

1. INTRODUCTION

After the pioneering works of Lax and Godunov in the late fifties, a lot of conservative difference schemes have been proposed for the solution of hyperbolic systems of conservation laws, using either a centred approximation in space or some upwinding. Since the late seventies, the schemes devised have been mostly implicit and free of severe stability constraints on the time step. Nowadays, a great number of explicit and implicit schemes are available and it would be useful to gather them in a general framework in order to unify their presentation, simplify their analysis, and make the search easier for new efficient methods. The aim of the present paper is to propose such a framework in the case of conservative schemes involving only two time - levels.

The construction of the general form of the schemes is developed in Section 2. Then some examples are given in Section 3 showing that the usual schemes can be simply identified. In Section 4, necessary and sufficient conditions are presented to obtain second-order accuracy, solvability, stability, dissipation and diagonal dominance. Section 5 describes the application of the general framework to the study of implicit schemes of second - order accuracy. Possible developments are indicated in the conclusions.

2. CONSTRUCTION OF THE TWO-LEVEL SCHEMES

Let us consider an initial-value problem for the system of m conservation laws :

$$w_t + f(w)_x = 0, \qquad x\in\mathbb{R}, \quad t > 0, \tag{2.1}$$

where the state-vector w(x,t) belongs to an open set Ω of $\mathbb{R}^m$ and the flux - function $f : \Omega \to \mathbb{R}^m$ is smooth. This system can also be written in the expanded form :

$$w_t + A(w) \ w_x = 0 \tag{2.2}$$

with the jacobian matrix A(w) = df(w) / dw.

System (2.1) is assumed to be hyperbolic, i.e. the matrix A(w) has m real eigenvalues and a complete set of eigenvectors.

We approximate System (2.1) by a finite-difference scheme with 2 time - levels :

$$\mathbf{S} \ (w_{j-J}, w_{j-J+1}, \ldots, w_{j+J} \ ; \Delta w_{j-J_1}, \ldots, \Delta w_{j+J_1} \ ; \sigma) = 0 \tag{2.3}$$

where $w_j \equiv w_j^n$ is the numerical solution at the old time-level $t = n\,\Delta t$ for $x = j\Delta x$, $\Delta w_j \equiv w_j^{n+1} - w_j^n$ is the increment of the numerical solution during a time step and σ denotes the step ratio :

$$\sigma = \Delta t \ / \ \Delta x.$$

Scheme (2.3) involves (at most) 2J+1 points at the old time-level and $2J_1+1$ points at the new one. It is explicit if $J_1 = 0$ or implicit otherwise.

Scheme (2.3) is assumed to be **conservative**, which means it can be written as :

$$\Delta w_j = - \sigma \ (h_{j+\frac{1}{2}} - h_{j-\frac{1}{2}}) \tag{2.4}$$

with a numerical flux :

$$h_{j+\frac{1}{2}} = h \ (w_{j-J+1}, \ldots, w_{j+J} \ ; \Delta w_{j-J_1+1}, \ldots, \Delta w_{j+J_1} \ ; \sigma)$$

where h is a Lipschitz-continuous function satisfying the consistency condition :

$$h \ (u, u, \ldots, u \ ; 0, 0, \ldots, 0 \ ; \sigma) = f \ (u), \qquad u\in\Omega. \tag{2.5}$$

Similarly as in the Lax and Wendroff paper for explicit schemes [1], one can easily show that if such a conservative scheme converges boundedly almost everywhere as Δx and Δt tend to zero, then it converges to a weak solution of System (2.1).

Furthermore, we assume that the scheme (2.4) involves **essentially 3 points,** i.e.

$$h\,(u_{-J+1},\ \dots,\ u_{-1},\ u,\ u,\ u_2,\ \dots,\ u_J\ ;$$

$$v_{-J_1+1},\ \dots,\ v_{-1},\ 0,\ 0,\ v_2,\ \dots,\ v_{J_1}\ ;\ \sigma) = f\ (u)\ , \qquad (2.6)$$

for any $u \in \Omega$, $u_p \in \Omega$ and $v_p \in \mathbb{R}^m$.

This property, first introduced for explicit schemes (see [2, Section 4] and [3]), is stronger than the consistency condition (2.5) , but it is satisfied by nearly all the schemes presently used in practice. Roughly speaking, it means that the consistency is ensured with the 3 central-points.

Finally, the scheme (2.4) is supposed to be either **explicit or linearly implicit,** i.e. its numerical flux is of the form :

$$h_{j+\frac{1}{2}} = h^{\mathrm{expl}}_{j+\frac{1}{2}} + \sum_{p=-J_1+1}^{J_1} (H_p)_{j+\frac{1}{2}}\ \Delta w_{j+p} \qquad (2.7)$$

with an explicit part :

$$h^{\mathrm{expl}}_{j+\frac{1}{2}} = h^{\mathrm{expl}}\ (w_{j-J+1},\ \dots,\ w_{j+J};\ \sigma)$$

and an implicit part with mxm matrix coefficients :

$$(H_p)_{j+\frac{1}{2}} = H_p\ (w_{j-J+1},\dots,w_{j+J};\ \sigma), \qquad p=-J_1+1,\ \dots,\ J_1.$$

With the numerical flux (2.7), the scheme (2.4) leads to the solution of an algebraic linear system at each time iteration. For simplicity reasons, all the usual schemes are explicit or linearly implicit.

Let us now give the general form of the schemes satisfying the above assumptions. To write this form down, we need two classical operators for the space-differencing :

$$(\mu\psi)_{j+\frac{1}{2}} \equiv \frac{1}{2}(\psi_j + \psi_{j+1})$$

$$(\delta\psi)_{j+\frac{1}{2}} \equiv \psi_{j+1} - \psi_j$$

where ψ_j is a mesh function defined at $x = j\Delta x$ for integer values of $2j$.

Theorem 1 - Any conservative scheme (2.4) involving essentially 3 points and being explicit or linearly implicit can be written in the simple form :

$$\Delta w_j + \frac{\sigma}{2}\,\delta\,[M\,\mu\,(\Delta w)]_j - \frac{1}{4}\,\delta\,[P\,\delta(\Delta w)]_j + \sigma\,(\delta h')_j = -\,\sigma\,\delta(\mu f)_j + \frac{1}{2}\,\delta\,(Q\,\delta w)_j \qquad (2.8)$$

with three mxm - matrices depending on the old time-level :

$$M_{j+\frac{1}{2}} = M\,(w_{j-J+1}, \ldots, w_{j+J}\,;\,\sigma)$$

$$P_{j+\frac{1}{2}} = P\,(w_{j-J+1}, \ldots, w_{j+J}\,;\,\sigma)$$

$$Q_{j+\frac{1}{2}} = Q\,(w_{j-J+1}, \ldots, w_{j+J}\,;\,\sigma)$$

and a m-vector :

$$h'_{j+\frac{1}{2}} = \sum_{p=-J_1+1}^{J_1} (\mathcal{H}_p)_{j+\frac{1}{2}}\,(\delta w_{j+\frac{1}{2}}\,,\,\Delta w_{j+p})$$

where the $(\mathcal{H}_p)_{j+\frac{1}{2}}$ are bilinear applications depending on the old-time level :

$$(\mathcal{H}_p)_{j+\frac{1}{2}} = \mathcal{H}_p\,(w_{j-J+1}, \ldots, w_{j+J}\,;\,\sigma), \quad \text{for } p\neq 0 \text{ and } 1$$

$$(\mathcal{H}_o)_{j+\frac{1}{2}} = (\mathcal{H}_1)_{j+\frac{1}{2}} = 0.$$

Remarks :

a) If $J_1 = 1$ (only 3 points at the new time-level), then the most complicated term disappears :

$$h'_{j+\frac{1}{2}} = 0.$$

Such a scheme is entirely characterized by the data of M, P and Q.

b) If $J_1 = 0$ (explicit scheme), then

$$M_{j+\frac{1}{2}} = P_{j+\frac{1}{2}} = 0 \qquad \text{and} \qquad h'_{j+\frac{1}{2}} = 0$$

and the scheme is defined by the only datum of Q, as in the work by Tadmor [3].

proof of theorem 1: For an explicit or linearly-implicit conservative scheme, the numerical flux function can be written as :

$$h\,(U\,;\,V;\,\sigma) = h^{expl}(U\,;\,\sigma) + h^{impl}\,(U\,;\,V;\,\sigma) \tag{2.9}$$

with

$$h^{impl}\,(U\,;\,V;\,\sigma) = \sum_{p=-J_1+1}^{J_1} H_p\,(U\,;\,\sigma)\,v_p \tag{2.10}$$

where the argument list has been shortened by setting :

$$U = (u_{-J+1},\, u_{-J+2},\, \ldots,\, u_J)$$

$$V = (v_{-J_1+1},\, v_{-J_1+2},\, \ldots,\, v_{J_1}).$$

The scheme involving essentially 3 points, the function h must satisfy :

$$h\,(U_o\,;\,V_o^*;\,\sigma) = f\,(u_o) \tag{2.11}$$

where

$$U_o = (u_{-J+1},\, \ldots,\, u_{-1},\, u_o,\, u_o,\, u_2,\, \ldots,\, u_J)$$

$$V_o^* = (v_{-J_1+1},\, \ldots,\, v_{-1},\, 0,\, 0,\, v_2,\, \ldots,\, v_{J_1}).$$

By choosing $v_p = 0$ for any p, from (2.9), (2.10) and (2.11) we can deduce :

$$h^{expl}\,(U_o;\,\sigma) = f\,(u_o). \tag{2.12}$$

Hence :

$$\sum_{p\,\epsilon\,\Pi_o} H_p\,(U_o\,;\,\sigma)\;v_p = 0$$

where $\Pi_o = \{-J_1+1,\, -J_1+2,\, \ldots,\, -1\} \cup \{2,\, 3,\, \ldots,\, J_1\}.$

This relation being valid for any value of v_p, we have ;

$$H_p\,(U_o\,;\,\sigma) = 0, \qquad p\,\epsilon\,\Pi_o \tag{2.13}$$

Since the numerical flux is Lipschitz continuous, so is the function H_p and there exists some bilinear application $\mathcal{H}_p\ (U\ ;\ \sigma)$ such that :

$$H_p\ (U\ ;\ \sigma) = H_p\ (U_o\ ;\ \sigma) + \mathcal{H}_p\ (U\ ;\ \sigma)\ (u_1 - u_o,\ .\ )$$

$$= \mathcal{H}_p\ (U\ ;\ \sigma)\ \ (u_1 - u_o,\ .\ ),\qquad p \in \Pi_o$$

Therefore, the implicit part of the numerical flux can be expressed as :

$$h^{impl}\ (U\ ;\ V\ ;\ \sigma) = H_o\ (U\ ;\ \sigma)\ v_o + H_1\ (U\ ;\ \sigma)\ v_1\ + h'\ (U\ ;\ V\ ;\ \sigma)$$

with

$$h'\ (U\ ;\ V\ ;\ \sigma) = \sum_{p\,=\,-J_1+1}^{J_1} \mathcal{H}_p\ (U\ ;\ \sigma)\ (u_1 - u_o,\ v_p)$$

and

$$\mathcal{H}_o = \mathcal{H}_1 = 0.$$

By introducing new matrices :

$$M\ (U\ ;\ \sigma) = 2\ (H_o + H_1)\ \ (U\ ;\ \sigma)$$

$$P\ (U\ ;\ \sigma) = 2\sigma\ (\ H_o - H_1)\ (U\ ;\ \sigma)$$

we can rewrite h^{impl} as :

$$h^{impl}\ (U\ ;\ V\ ;\ \sigma) = \frac{1}{4}\ M\ (U\ ;\ \sigma)\ \ (v_o + v_1) - \frac{1}{4\sigma}\ P\ (U\ ;\ \sigma)\ (v_1 - v_o) + h'\ (U\ ;\ V\ ;\ \sigma).$$

Let us now consider the explicit part of the numerical flux. There exists some matrices $\bar{A}\ (u_o,\ u_1)$ and $B(U\ ;\ \sigma)$ such that :

$$f\ (u_1) = f\ (u_o) + \bar{A}\ (u_o,\ u_1)\ (u_1 - u_o)$$

$$h^{expl}\ (U\ ;\ \sigma) = h^{expl}\ (U_o\ ;\ \sigma) + \ B\ (U\ ;\ \sigma)\ \ (\ u_1 - u_o).$$

Thus, by using the condition (2.12), we obtain :

$$h^{expl}\ (U\ ;\ \sigma) = \frac{1}{2}\ \left[f(u_o) + f(u_1)\right] + \left[B(U\ ;\ \sigma) - \frac{1}{2}\ \bar{A}\ (u_o,\ u_1)\right]\ (\ u_1 - u_o)$$

$$= \frac{1}{2}\ \left[f(u_o) + f(u_1)\right] - \frac{1}{2\sigma}\ \ Q\ (U\ ;\ \sigma)\ \ (u_1 - u_o)$$

where

$$Q(U\ ;\ \sigma)\ \ =\ \ \sigma\ \bar{A}(u_o,\ u_1)\ -\ \ 2\sigma\ B(U\ ;\ \sigma).$$

Finally by taking $u_p = w_{j+p}$ and $v_p = \Delta w_{j+p}$, we find the numerical flux at $j+\frac{1}{2}$:

$$h_{j+\frac{1}{2}} = \left[\mu\ f - \frac{1}{2\sigma}\ Q\ \delta w\ + \frac{1}{2}\ M\ \mu(\Delta w) - \frac{1}{4\sigma}\ P\ \delta(\Delta w) + h'\right]_{j+\frac{1}{2}}$$

corresponding to the scheme (2.8).

3. Some examples

3.1 Explicit schemes :

For an explicit scheme, the general form reduces to :

$$\Delta w_j = -\ \sigma\ \ \delta(\mu f)_j + \frac{1}{2}\ \ \delta(Q\ \delta\ w)_j . \tag{3.1}$$

Here, we want to emphasize that we are able to exhibit the Q - matrix for the usual schemes, even though they are defined in several steps.

a) Lax scheme

The Lax scheme [4] can be defined simply by :

$$Q_{j+\frac{1}{2}} = I \tag{3.2}$$

where I is the mxm identity matrix.

b) First-order Roe scheme

It is based on the splitting of some mxm-matrix calculated from the "Roe average" [5] :

$$(A_R)_{j+\frac{1}{2}} = A_R\ (w_j,\ w_{J+1})$$

which is defined by the following properties :

- $f(v) - f(u) = A_R(u,\ v)\ \ (v-u), \qquad u,\ v \in \Omega$ (3.3)
- $A_R(u,\ u) = A(u), \quad u\in\Omega$ (3.4)
- $A_R(u,\ v)$ has real eigenvalues and a complete set of eigenvectors for any u and v in Ω.

Roe has constructed this average for gas dynamics. Later, Harten and Lax have shown [6] that such an average exists for any hyperbolic system having a strictly convex entropy.

Splitting up the Roe matrix according to the sign of its eigenvalues :

$$A_R = A_R^+ + A_R^-,$$

and using an upwind differencing in space, one obtains the first-order Roe scheme [5]:

$$\Delta w_j = -\sigma\left[(A_R^- \ \delta w)_{j+\frac{1}{2}} + (A_R^+ \ \delta w)_{j-\frac{1}{2}}\right].$$

This scheme is really conservative and corresponds to

$$Q_{j+\frac{1}{2}} = \sigma \ |A_R|_{j+\frac{1}{2}} \tag{3.5}$$

where $|A_R| = A_R^+ - A_R^-$.

c) Lax - Wendroff scheme

It is a centred scheme [1] defined as :

$$\Delta w_j = -\sigma \ \delta(\mu f)_j + \frac{\sigma^2}{2}\left[\delta(\mu A) \ \delta f\right]_j$$

To find the Q - matrix of the Lax - Wendroff scheme, one can use the Roe average. From the relation (3.3), we deduce :

$$Q_{j+\frac{1}{2}} = \sigma^2 \ (\mu A)_{j+\frac{1}{2}} \ (A_R)_{j+\frac{1}{2}} \tag{3.6}$$

d) S_β^α schemes

These predictor - corrector schemes [7] can be expressed as :

$$\tilde{w}_{j+\beta}^{n+\alpha} = (1-\beta)w_j + \beta \ w_{j+1} - \alpha\sigma \ (f_{j+1} - f_j)$$

$$\Delta w_j = -\frac{\sigma}{2\alpha}\left[(\alpha-\beta) \ f_{j+1} + (2\beta-1) \ f_j + (1-\alpha-\beta) \ f_{j-1} + \tilde{f}_{j+\beta}^{n+\alpha} - \tilde{f}_{j+\beta-1}^{n+\alpha}\right]$$

where $\tilde{f}_{j+\beta}^{n+\alpha} = f(\tilde{w}_{j+\beta}^{n+\alpha})$, α and β being two parameters (real numbers, $\alpha \neq 0$).

This class of schemes contains some popular explicit methods such as the Richtmyer scheme [8] ($\alpha = \beta = \frac{1}{2}$) or the MacCormack scheme [9] ($\alpha = 1$, $\beta = 0$)

It is possible to write the S_β^α schemes in a single step by using again the Roe average. After some computation, one can express these schemes in the general form with

$$Q_{j+\frac{1}{2}} = \sigma \ \frac{\beta}{\alpha} \ (A_R - \tilde{A}_R)_{j+\frac{1}{2}} + \sigma^2 \ (\tilde{A}_R)_{j+\frac{1}{2}} \ (A_R)_{j+\frac{1}{2}} \tag{3.7}$$

where

$$(\tilde{A}_R)_{j+\frac{1}{2}} = A_R \ (w_j, \ \tilde{w}_{j+\beta}^{n+\alpha}).$$

3.2 Implicit schemes :

Let us consider the Beam and Warming schemes [10] :

$$\Delta w_j + \Theta\ \sigma\ \ \delta\left[\mu\ (A\ \Delta w)\right]_j = -\ \sigma\ \ \delta(\mu f)_j$$

where Θ is a parameter ($\Theta \in \mathbb{R}$).

In the general form (2.8), these implicit schemes correspond to $h'_{j+\frac{1}{2}} = 0$ and :

$$M_{j+\frac{1}{2}} = 2\ \Theta\ (\mu A)_{j+\frac{1}{2}}, \qquad P_{j+\frac{1}{2}} = -\ \Theta\sigma\ (\delta A)_{j+\frac{1}{2}} \tag{3.8}$$

$$Q_{j+\frac{1}{2}} = 0. \tag{3.9}$$

4. General results

4.1 Order of accuracy

Theorem 2 - Suppose that the exact flux and the numerical flux are sufficiently smooth (more precisely f is C^3, M and Q are C^2, P and $\mathcal{H}_p$ are C^1) and the step ratio σ is constant as Δt and Δx tend to zero. Then, the general two-level scheme (2.8) is at least **second-order accurate** if and only if :

$$Q\ (u,\ u,\ \ldots,\ u\ ;\ \sigma) = \sigma^2\ \left[A^2(u) - M(u,u,\ \ldots,\ u\ ;\ \sigma)\ A\ (u)\right], \qquad u\epsilon\Omega. \tag{4.1}$$

Otherwise, the scheme is first-order accurate.

Remarks :

a) The condition for second-order accuracy is independent of the functions P and $\mathcal{H}_p$, $p = -J_1+1,\ \ldots,\ J_1$. This means that in the scheme, the terms associated with P and h' are of order 2, at least.

b) For an explicit scheme, this condition reduces to :

$$Q(u,u,\ \ldots u\ ;\ \sigma) = \sigma^2\ A^2(u), \qquad u\epsilon\Omega. \tag{4.2}$$

Proof of theorem 2 : The local truncation error of the scheme (2.8) can be written as :

$$\epsilon_j(\Delta t\ ;\ \sigma) = \sum_{q=1}^{6}\ (e_q)_j$$

where

$$e_1 = \frac{\Delta w}{\Delta t}, \quad e_2 = \frac{\Delta t}{2\,\Delta x}\ \delta\left[M\,\mu\left(\frac{\Delta w}{\Delta t}\right)\right], \quad e_3 = -\frac{1}{4}\ \delta\left[P\,\delta\left(\frac{\Delta w}{\Delta t}\right)\right]$$

$$e_4 = \frac{\delta h'}{\Delta x}, \quad e_5 = \frac{\delta \mu f}{\Delta x}, \quad e_6 = -\frac{1}{2\,\Delta t}\ \delta(Q\ \delta w)$$

and w_j is here a smooth solution of the exact system (2.1) at $x=j\Delta x$ and $t = n\,\Delta t$.

By making Taylor expansions of the truncation-error terms, we obtain :

$$e_1 = \frac{\partial w}{\partial t} + \frac{\Delta t}{2}\ \frac{\partial^2 w}{\partial t^2} + 0\ (\Delta t^2)$$

$$e_2 = \frac{\Delta t}{2}\ \frac{\partial}{\partial x}\ \left[M\ (w, \ldots, w\ ; \sigma)\ \frac{\partial w}{\partial t}\right] + 0\ (\Delta t^2) + 0\ (\Delta t\ \Delta x)$$

$$e_3 = 0\ (\Delta x^2), \quad e_4 = 0\ (\Delta t\ \Delta x), \quad e_5 = \frac{\partial f(w)}{\partial x} + 0\ (\Delta x^2)$$

$$e_6 = \frac{\Delta t}{2\ \sigma^2}\ \frac{\partial}{\partial x}\ \left[Q(w, \ldots, w\ ; \sigma)\ \frac{\partial w}{\partial x}\right] + 0\ (\Delta t\ \Delta x).$$

Since w satisfies System (2.1), one can express the time derivatives in terms of space derivatives :

$$\frac{\partial w}{\partial t} = -\frac{\partial f(w)}{\partial x} = -\ A(w)\ \frac{\partial w}{\partial x}$$

$$\frac{\partial^2 w}{\partial t^2} = -\frac{\partial}{\partial x}\left[\frac{\partial f(w)}{\partial t}\right] = -\frac{\partial}{\partial x}\left[A(w)\ \frac{\partial w}{\partial t}\right] = \frac{\partial}{\partial x}\left[A(w)\ \frac{f(w)}{\partial x}\right] = \frac{\partial}{\partial x}\left[A^2(w)\ \frac{\partial w}{\partial x}\right].$$

Upon eliminating the time derivatives, we get :

$$\epsilon_j\ (\Delta t\ ; \sigma) = \frac{\Delta t}{2}\ \frac{\partial}{\partial x}\left\{\left[A^2(w) - M\ (w, \ldots, w\ ; \sigma)\ A(w) - \frac{1}{\sigma^2}\ Q\ (w, \ldots, w\ ; \sigma)\right]\ \frac{\partial w}{\partial x}\right\}_j + 0\ (\Delta t^2)$$

as $\Delta t \to 0$ with $\sigma=$ const. , which leads to the condition (4.1) for second - order accuracy.

Examples - Application of theorem 2 to the 3-point schemes introduced in the previous section is summarized in Table 1. For Roe, Lax-Wendroff and S^{α}_{β} schemes, we have computed Q (u, u ; σ) by using the property (3.4) of the Roe average.

SCHEME	Q(u,u ; σ)	M(u,u ; σ)	ORDER
Lax	I	0	1
Roe	$\sigma\ \lvert A(u)\rvert$	0	1
LW	$\sigma^2\ A^2(u)$	0	2
S^{α}_{β}	$\sigma^2\ A^2(u)$	0	2
BW	0	$2\ \Theta\ A(u)$	2, if $\Theta = \frac{1}{2}$ 1, otherwise

Table 1 : Order of accuracy of various schemes.

4.2 Solvability, stability, dissipation and diagonal dominance

To analyse the solvability, stability and dissipation of the general two-level scheme (2.8), we linearise the problem by assuming that the jacobian matrix A is constant. Similarly, we suppose that the matrices $M_{j+\frac{1}{2}}$, $P_{j+\frac{1}{2}}$ and $Q_{j+\frac{1}{2}}$ does not really depend on j (we drop the subscripts) and that the bilinear applications $(\mathcal{H}_p)_{j+\frac{1}{2}}$ are null, so that the scheme (2.8) is linear too. Often, the linearity of the scheme results from the linearity of the exact system.

We assume also that the matrices M, P and Q have the same eigenvectors as A and have real eigenvalues, which is true for all the usual schemes. We shall denote by $\lambda^{(k)}_D$ the k^{th} eigenvalue of some matrix D (D = A, M, P and Q, k = 1, 2, ... m).

By making a spatial Fourier analysis, we have obtained the following results for the discrete initial-value problem in L_2 :

Theorem 3 - The scheme (2.8) is linearly solvable (in L_2) if and only if

$$(1 + \lambda^{(k)}_P > 0) \quad \underline{\text{or}} \quad (1 + \lambda^{(k)}_P < 0 \quad \text{and} \quad \lambda^{(k)}_M \neq 0) \qquad (4.3)$$

for k = 1, 2, ..., m.

Remarks

a) Of course, the condition (4.3) involves only the matrices M and P relative to the implicit part of the numerical flux. It is trivially satisfied for an explicit scheme ($\lambda_p^{(k)} = 0$).

b) Theorem 3 yields that the scheme is not admissible if one of the eigenvalue of P is equal to -1. However in pratice, we cannot admit either a scheme satisfying the second possibility in (4.3) for some k, since it can proved that such a scheme is not stable, provided its matrices M and P be defined as continuous functions of A. Thus, we are usually interested in schemes for which all the eigenvalues of P are strictly greater than - 1.

Theorem 4 - Suppose that the scheme (2.8) is linearly solvable. Then it is linearly stable (in L_2) if and only if :

$$\sigma^2 (\lambda_A^{(k)} - \lambda_M^{(k)}) \quad \lambda_A^{(k)} \leq \lambda_Q^{(k)} \tag{4.4}$$

and

$$\lambda_Q^{(k)} \; (\lambda_Q^{(k)} - \lambda_P^{(k)} - 1) \leq 0 \tag{4.5}$$

for k = 1, 2, ..., m.

Remark - For an explicit scheme, conditions (4.4) - (4.5) reduce to :

$$(\sigma \, \lambda_A^{(k)})^2 \leq \lambda_Q^{(k)} \leq 1. \tag{4.6}$$

Theorem 5 - Suppose that the scheme (2.8) is linearly solvable. Then it is linearly dissipative (in the sense of Kreiss) if it and only if is linearly stable and satisfies

$$\lambda_Q^{(k)} \neq 0 \quad \text{and} \quad \lambda_Q^{(k)} \neq 1 + \lambda_P^{(k)} \tag{4.7}$$

for k = 1, 2, ..., m.

Remark - When the scheme is linearly dissipative, the order of dissipation is 4 if condition (4.1) holds, or 2 otherwise.

We can also express in terms of the eigenvalues of M and P a simple condition ensuring a good numerical resolution of the algebraic problem to solve at each time iteration. Consider a linear scheme in the form :

$$\sum_{p=-L}^{L} \quad B_p \quad (\Delta w)_{j+p} = \Delta w_j^{\text{expl}}$$

where the coefficients B_p are constant mxm matrices having the same eigenvectors. We shall say that this scheme is SDD (strictly diagonally dominant) if :

$$|\lambda_{B_o}^{(k)}| > \sum_{\substack{p=-L \\ p\neq 0}}^{L} |\lambda_{B_p}^{(k)}|, \qquad k = 1, 2, \ldots, m.$$

Using this definition, one can prove the next theorem.

Theorem 6 The scheme (2.8) is linearly SDD if and only if :

$$1 + \lambda_P^{(k)} > 0 \quad \underline{\text{and}} \quad \sigma\,|\lambda_M^{(k)}| < 2 + \lambda_P^{(k)} \qquad (4.8)$$

for k = 1, 2, ..., m.

Remark Note that condition (4.8) is stronger than the solvability condition (4.3)

Examples - Application of theorems 3-6 to some schemes introduced in Section 3 is summarized in Table 2. The CFL number is :

$$\text{CFL} = \sigma \max_k |\lambda_A^{(k)}|.$$

Scheme	$\lambda_M^{(k)}, \lambda_P^{(k)}, \lambda_Q^{(k)}$	Solvability	Stability	Dissipation	SDD
Roe	$0, 0, \sigma\|\lambda_A^{(k)}\|$	always	CFL $\leq$ 1	CFL < 1 $\lambda_A^{(k)} \neq 0$	always
LW	$0, 0, (\sigma\lambda_A^{(k)})^2$	always	CFL $\leq$ 1	CFL < 1 $\lambda_A^{(k)} \neq 0$	always
BW $\theta \geq \frac{1}{2}$	$2\theta\lambda_A^{(k)}, 0, 0$	always	always	never	CFL$<\frac{1}{\theta}$

Table 2: Necessary and sufficient conditions for solvability, stability, dissipation and SDD.

5 . A CLASS OF CENTRED IMPLICIT SCHEMES

We now purpose using our general framework for systematically constructing an interesting class of implicit schemes. We look for two - level conservative schemes which are linearly implicit and involve only 3 points at each time - level. These schemes are of the form (2.8) with $h'_{j+\frac{1}{2}} = 0$ and the matrices $M_{j+\frac{1}{2}}$, $P_{j+\frac{1}{2}}$, $Q_{j+\frac{1}{2}}$ depend only on w_j, w_{j+1} and on the step ratio σ.

An economical choice of these matrices consists in the polynomial expressions :

$$\sigma M_{j+\frac{1}{2}} = c_o^M I + c_1^M (\sigma \bar{A}_{j+\frac{1}{2}}) + c_2^M (\sigma \bar{A}_{j+\frac{1}{2}})^2$$

$$P_{j+\frac{1}{2}} = c_o^P I + c_1^P (\sigma \bar{A}_{j+\frac{1}{2}}) + c_2^P (\sigma \bar{A}_{j+\frac{1}{2}})^2$$

$$Q_{j+\frac{1}{2}} = c_o^Q I + c_1^Q (\sigma \bar{A}_{j+\frac{1}{2}}) + c_2^Q (\sigma \bar{A}_{j+\frac{1}{2}})^2$$

where the (nondimensionalized) polynomial coefficients are scalar and independent of w_j, w_{j+1} and σ ; the matrix $\bar{A}_{j+\frac{1}{2}}$ denotes some average of A_j and A_{j+1} defined by

$$\bar{A}_{j+\frac{1}{2}} = \bar{A} (w_j, w_{j+1})$$

with :

$$\bar{A} (v, u) = \bar{A} (u, v), \qquad u, v \in \Omega \tag{5.1}$$

$$\bar{A} (u, u) = A (u), \qquad u \in \Omega. \tag{5.2}$$

For example, we can choose one of the following average formulas :

$$\bar{A}_{j+\frac{1}{2}} = (\mu A)_{j+\frac{1}{2}}, \qquad A \left[(\mu w)_{j+\frac{1}{2}}\right] \quad \text{or} \quad (A_R)_{j+\frac{1}{2}}$$

where A_R is the Roe average defined in Section 3.1.

For simplicity, we want the schemes to be centered in space, i.e. their numerical flux to satisfy :

$$h (w_j, w_{j+1} ; \Delta w_j, \Delta w_{j+1} ; \sigma) = h (w_{j+1}, w_j ; \Delta w_{j+1}, \Delta w_j, - \sigma).$$

This condition is equivalent to :

$$M (w_j, w_{j+1} ; \sigma) = M (w_{j+1}, w_j ; -\sigma) \tag{5.3}$$

$$P (w_j, w_{j+1} ; \sigma) = P (w_{j+1}, w_j ; -\sigma) \tag{5.4}$$

$$Q (w_j, w_{j+1} ; \sigma) = Q (w_{j+1}, w_j ; -\sigma), \tag{5.5}$$

which gives :

$$c_o^M = c_2^M = 0, \qquad c_1^P = 0 \qquad \text{and} \qquad c_1^Q = 0. \tag{5.6}$$

We also prescribe second-order accuracy, that is the condition (4.1). Using (5.2), this condition becomes :

$$c_o^Q I + (c_1^Q + c_o^M)(\sigma A) + (c_2^Q + c_1^M - 1)(\sigma A)^2 + c_2^M (\sigma A)^3 = 0.$$

Taking into account (5.6), we obtain the additional requirements :

$$c_2^Q = 1 - c_1^M \qquad \text{and} \qquad c_o^Q = 0. \tag{5.7}$$

Finally, the 9 polynominal coefficients must satisfy the 6 independent linear equations (5.6) and (5.7). Thus, the schemes depend on 3 real parameters that we denote by α, β and γ and that we define as :

$$c_1^M = 2\alpha, \qquad c_2^P = -2\beta, \qquad c_o^P = -2\gamma. \tag{5.8}$$

These schemes are characterized by the matrices :

$$M_{j+\frac{1}{2}} = 2\alpha \; \bar{A}_{j+\frac{1}{2}}, \qquad P_{j+\frac{1}{2}} = -2\beta \; (\sigma \bar{A}_{j+\frac{1}{2}})^2 - 2\gamma I$$

$$\text{and} \qquad Q_{j+\frac{1}{2}} = (1-2\alpha) \; (\sigma \bar{A}_{j+\frac{1}{2}})^2$$

One can slightly modify these matrices by using not only one average $\bar{A}$ but several ones and also by adding some terms compatible with the centring in space and the second-order accuracy. For instance, one can consider :

$$M_{j+\frac{1}{2}} = 2\alpha \; (\mu A)_{j+\frac{1}{2}} \tag{5.9}$$

$$P_{j+\frac{1}{2}} = -2\beta \; \sigma^2 \; (\mu A)^2_{j+\frac{1}{2}} - 2\gamma \; I - \alpha\sigma \; (\delta A)_{j+\frac{1}{2}} \tag{5.10}$$

$$Q_{j+\frac{1}{2}} = (1-2\alpha) \; \sigma^2 \; (\mu A)_{j+\frac{1}{2}} \; (A_R)_{j+\frac{1}{2}} \tag{5.11}$$

This choice satisfies(5.3) - (5.5) and also (4.1). Its advantage is to contain as a particular case the scheme of Lax and Wendroff ($\alpha = \beta = \gamma = 0$) and also the scheme of Beam and Warming with $\Theta = 1/2$ ($\alpha = 1/2$, $\beta = \gamma = 0$). Inserting the matrices (5.9) - (5.11) in the general form of the scheme, we obtain :

$$\Delta w_j + \alpha \, \sigma \, \delta \; [\mu \; (A \; \Delta w)]_j + \beta \; \frac{\sigma^2}{2} \; \delta \; [(\mu A)^2 \; \delta(\Delta w)]_j + \frac{\gamma}{2} \; \delta^3 \; (\Delta w)_j$$

$$= -\sigma \; \delta \; (\mu f)_j + (1-2\alpha) \; \frac{\sigma^2}{2} \; \delta \; [(\mu A) \; \delta f]_j \tag{5.12}$$

This class of schemes was first proposed in [11].

Using the general results of Section 4, we can easily select a scheme in the class (5.12). Using theorems 3, 4 and 5 we find that the implicit schemes (5.12) are linearly solvable, stable and dissipative without condition (on the CFL and thus on Δt) if and only if :

$$\alpha < \frac{1}{2}, \quad \beta \leq \alpha - \frac{1}{2} \quad \text{and} \quad \gamma < \frac{1}{2} . \tag{5.13}$$

Furthermore, theorem 6 shows that the schemes are linearly SDD without condition if and only if :

$$\beta < - \frac{\alpha^2}{4(1-\gamma)} . \tag{5.14}$$

To end the selection of the parameters, one can use other criterions. For an unsteady problem, one can minimize the evolutionary truncation error as in [11]. For a steady problem solved as the limit of a time - dependent problem, it is more efficient to optimize the convergence rate to the steady state as in [12]. This gives the condition :

$$\beta = 2\alpha - 1 \tag{5.15}$$

The simplest choice satisfying (5.13) - (5.15) is :

$$\alpha = 0, \quad \beta = -1 \quad \text{and } \gamma = 0 \tag{5.16}$$

which corresponds to :

$$\Delta w_j - \frac{\sigma^2}{2} \delta \left[(\mu A)^2 \quad \delta(\Delta w)\right]_j$$

$$= - \sigma \delta (\mu f)_j + \frac{\sigma^2}{2} \delta \left[(\mu A) \quad \delta f\right]_j \tag{5.17}$$

Various applications have been developed with the extension of this scheme in two and three space-dimensions. Let us mention here the transonic calculations without artificial viscosity presented in [13].

6 . CONCLUSIONS

A general framework has been presented for two-level approximations of hyperbolic systems of conservation laws in one-space dimension. This framework applies to explicit and implicit schemes, with centred or upwind spatial-differencing. Using this framework, we have shown that it is very easy to deduce the basic properties of a particular approximation. Moreover, the framework is useful for systematically constructing a scheme submitted to some requirements.

Some properties of difference schemes have not been considered in the present paper, for instance the property of the numerical solution to be total-variation diminishing (for a single conservation law) or the property of satisfying a discrete entropy inequality. The former property can be investigated with the present framework, at least for explicit approximations (see [2] and [3]). The latter is more difficult to study ; however general results can be obtained when the scheme is applied only to reach a steady-state, by using some results of Osher [14] and Tadmor [15] on semi-discrete schemes. Following this idea, entropy corrections of implicit schemes have been proposed in [16]. Finally, the generalization of the framework to the case of several space-dimensions, can be possible subject to increasing technical difficulties.

REFERENCES

[1] LAX P.D. and WENDROFF B. - Systems of conservation laws, Comm. Pure Appl. Math., 13, pp. 217-237, 1960.

[2] HARTEN A. - On a class of high resolution total-variation-stable finite-difference schemes, SIAM J. Numer. Anal. , 21, pp. 1-23, 1984.

[3] TADMOR E. - Numerical viscosity and the entropy condition for conservative difference schemes, Math. Comput. , 43, pp. 369-381, 1984.

[4] LAX P.D. - Weak solutions of nonlinear hyperbolic equations and their numerical computation, Comm. Pure Appl. Math. , 7, pp. 159-193, 1954.

[5] ROE P.L. - Approximate Riemann solvers, parameter vectors and difference schemes, J. Comput. Phys., 43, pp. 357-372, 1981.

[6] HARTEN A - On the symmetric form of systems of conservation laws with entropy, J. Comput. Phys., 49, pp. 151-164, 1983.

[7] LERAT A. and PEYRET R. - Sur le choix de schémas aux différences du second ordre fournissant des profils de choc sans oscillation, C.R. Acad. Sci. Paris, 277 A, pp. 363-366, 1973.

[8] RICHTMYER R.D. and MORTON K.W. - Difference methods for initial-value problems, Interscience Publ., New York, 1967.

[9] MacCORMACK R.W. - The effect of viscosity in hypervelocity impact cratering, AIAA Paper n° 69-354, 1969.

[10] BEAM R. and WARMING R.F. - An implicit finite-difference algorithm for hyperbolic systems in conservation-law form, J. Comput. Phys., 22, pp. 87-110, 1976.

[11] LERAT A. - Une classe de schémas aux différences implicites pour les systèmes hyperboliques de lois de conservation, C.R. Acad. Sci. Paris, 288 A, pp. 1033-1036, 1979.

[12] DARU V. and LERAT A. - Analysis of an implicit Euler solver, in Numerical Methods for the Euler Equations of Fluid Dynamics, F. Angrand et al. Eds, SIAM Publ., pp. 246-280, 1985.

[13] LERAT A. and SIDES J. - Efficient solution of the steady Euler equations with a centered implicit method, Intern. Conf. Num. Meth. Fluid Dyn., Oxford, march 1988, To appear. Also TP ONERA 1988-128.

[14] OSHER S. - Riemann solvers, the entropy condition and difference approximations, SIAM J. Numer. Anal., 21, pp. 217-235, 1984.

[15] TADMOR E. - The numerical viscosity of entropy stable schemes for systems of conservation laws. I, Math. Comput., 49, pp. 91-103, 1987.

[16] KHALFALLAH K. and LERAT A. - Correction d'entropie pour des schémas numériques approchant un système hyperbolique, to appear.

NUMERICAL CALCULATIONS OF REACTING FLOWS

Matania Ben-Artzi
Institute of Mathematics
Hebrew University
Jerusalem 91904, Israel

1. Introduction

Consider the Euler equations that model the time dependent flows of an inviscid, compressible, reactive fluid through a duct of smoothly varying cross-section. In addition to the hydrodynamical pressure force, we allow an external conservative field, which does not vary with time. We are using here the quasi one-dimensional approximation, namely, the hypothesis that all flow variables are uniform across a fixed cross section. Note that our treatment applies in particular to all problems with planar, cylindrical or spherical symmetry. The latter arise, e.g., in astrophysics [8]. Denoting by r the spatial coordinate and by $A(r)$ the area of the cross section at r, our equations are

$$A \frac{\partial}{\partial t} U + \frac{\partial}{\partial r}(AF(U)) + A \frac{\partial}{\partial r} G(U) + AH(U) = 0 ,$$

$$U = \begin{pmatrix} \rho \\ \rho u \\ \rho E \\ \rho z \end{pmatrix}, \quad F(U) = \begin{pmatrix} \rho u \\ \rho u^2 \\ (\rho E + p)u \\ \rho z u \end{pmatrix}, \quad G(U) = \begin{pmatrix} 0 \\ p \\ 0 \\ 0 \end{pmatrix}, \quad H(U) = \begin{pmatrix} 0 \\ \rho\phi'(r) \\ 0 \\ k\rho \end{pmatrix}, \tag{1.1}$$

where ρ, p, u are, respectively, density, pressure and velocity. z is the mass fraction of the unburnt fluid, that is, $z = 1$ (resp. $z = 0$) represents the completely unburnt (resp. burnt) fluid. The total specific energy E is given by $E = e + \frac{1}{2} u^2 + \phi$, where $\phi = \phi(r)$ is the external potential whose derivative $\phi'(r)$ is the external force field in (1.1) and e is the specific internal energy (including chemical energy). We are assuming an equation-of-state of the form $p = p(e, \rho, z)$. The reaction rate $k = k(e, \rho, z)$ is assumed to be a positive function. Along a particle path we have $\frac{dz}{dt} = -k$.

Our purpose in this paper is to present a robust, high-resolution numerical scheme for the time integration of the equations (1.1). We work within the general technique of the GRP (Generalized Riemann Problem) method [1,2], which is an analytic extension of Godunov's first-order scheme, and has its origins in the work of van-Leer [7]. The basic ingredient in our approach is a careful analysis of solutions of (1.1) in the neighborhood of a jump discontinuity, where, unlike the standard Riemann problem, the initial values of the flow variables are not piecewise constant and their slopes are allowed to have a jump at the discontinuity. It is important to emphasize that a sharp resolution of discontinuities in the case at hand is much more crucial than in the non-reactive fluid dynamical case. Indeed, if "viscous" shocks can be tolerated in the latter case, smearing discontinuities over a few computational zones, they must be totally avoided in the first case. If not avoided, they interfere with the fine structure of the reaction zone, leading to non-physical bifurcating solutions. The reader is referred to [1,4] for more background material on the structure of reacting waves and the numerical difficulties encountered in their calculations.

2. The GRP (Generalized Riemann Problem) Method

The GRP can be described as follows: Let $U(r,t)$ be the (correct entropy) solution of (1.1), where the initial values ($t = 0$) are linearly distributed with a jump at $r = 0$,

$$U_0(r) = U_\pm + U_\pm' \cdot r \quad \text{for} \quad \pm r > 0 . \tag{2.1}$$

Here $U_\pm, U_\pm'$ are constant vectors.

The problem is: Find the values of $U(r,t)$ and $\frac{\partial}{\partial t} U(r,t)$ at the initial discontinuity, namely,

$$U_0 = \lim_{t \to 0} U(0,t), \quad \left(\frac{\partial U}{\partial t} \right)_0 = \lim_{t \to 0} \frac{\partial}{\partial t} U(0,t) . \tag{2.2}$$

We define the "Associated Riemann Problem" as the initial value problem for (1.1) where, however, we set $U_\pm' = 0$ in (2.1) and assume that in (1.1),

$$A(r) \equiv 1, \quad H(U) \equiv 0. \tag{2.3}$$

This, of course, leads to a (self-similar) solution depending only on $\frac{r}{t}$. It is known [1] that the wave patterns for the GRP and its associated RP are the same, and, furthermore, they have a common limiting value U_0 (see (2.2)). Thus, the determination of U_0 is reduced to the set of algebraic equations encountered in the solution of the Riemann problem. However, while $\left(\frac{\partial U}{\partial t} \right)_0 = 0$ necessarily for the Riemann problem (self-similarity), its determination in the GRP case is far from trivial. In what follows we shall outline our method for solving the GRP, trying to highlight the main ideas in the proofs. The reader may find the detailed treatment in [1,2,3].

Recall that the wave pattern for the solution of the GRP involves the non-linear waves, associated with the characteristic values $u \pm c$, separated by the linearly degenerate contact discontinuity. Across the latter, the pressure p and velocity u are continuous, hence the same is true for their directional derivatives along $dr = udt$, i.e., the time derivatives in the "Lagrangian frame". Denoting these derivatives by $\left(\frac{dp}{dt}\right)^*, \left(\frac{du}{dt}\right)^*$, one has:

Theorem 2.1. The derivatives $\left(\frac{du}{dt}\right)^*, \left(\frac{dp}{dt}\right)^*$, *satisfy a pair of linear (algebraic) equations,*

$$a_\pm \left(\frac{du}{dt}\right)^* + b_\pm \left(\frac{dp}{dt}\right)^* = d_\pm , \qquad (2.4)$$

where a_+, b_+, d_+ *(resp.* $a_-, b_- d_-$*) can be determined explicitly from the values of* U_0 *(see (2.2)) and* U_+, U_+' *(resp.* U_-, U_-' *) (see (2.1)).*

The idea of the proof is as follows. Suppose that the wave travelling to the right is a shock wave. Then, parameterizing its strength by u, say, we have a Rankine-Hugoniot relation $p = p(u)$, where p is the pressure behind the shock (note that this relation involves the state ahead of the shock as a parameter). Differentiating this relation and using the system (1.1) to replace spatial (resp. time) derivatives behind (resp. ahead of) the shock one gets a linear relation of the type (2.4). Of course, the situation is more complicated if instead of the shock we have a centered rarefaction wave. In this case, the argument above can be implemented in order to derive an equation of the type (2.4), *provided* that the derivatives along the *tail characteristic* of the rarefaction wave are known (they should replace the initial spatial derivatives in the first part of the argument). But clearly the initial conditions (2.1) yield only the derivatives along the *head characteristic*. Thus, we should look for a suitable technique which would enable us to propagate directional derivatives across a centered rarefaction wave. Indeed, this is the analogue of the Rankine-Hugoniot relation mentioned above in the case of a shock. Such a technique is readily provided by the well-known "propagation of singularities" method. To be specific, assume that we have a centered rarefaction wave travelling to the left. Thus, we have Γ^--characteristics (slope $u - c$) fanning out of the singularity at $r = t = 0$. The family of Γ^+-characteristics intersects the wave transversally.

Let us use characteristic coordinates (α,β) throughout the rarefaction, so that $\beta =$ const. (resp. $\alpha =$ const.) corresponds to a Γ^- (resp. Γ^+) characteristic. It is convenient to take β as the (normalized) slope of the Γ^- curves at the origin and to take α in such a way that $\alpha \to 0$ for Γ^+ curves approaching the origin. Observe that the origin itself can be viewed as a "degenerate Γ^+ characteristic", to which we assign the value $\alpha = 0$. Also, we let $\beta = 1$ (resp. $\beta = \beta^*$) be the head (resp. tail) characteristic of the rarefaction wave. Thus, throughout the wave, every flow variable Q (including the coordinates r, t themselves!) becomes a function $Q(\alpha,\beta)$, defined in a rectangle $0 \leq \alpha \leq \alpha^*$, $\beta^* \leq \beta \leq 1$, for some $\alpha^* > 0$.

According to the remarks following (2.3) the value β^* is determined by the associated RP (actually, it is the normalized slope of the tail characteristic of the corresponding centered wave in the RP). Using this framework, the derivatives along Γ^- curves, at the origin, are given by $\frac{\partial Q}{\partial \alpha}(\alpha = 0,\beta)$, $\beta^* \leq \beta \leq 1$. The information which is readily available to us, via the initial conditions, consists of the values of $\frac{\partial Q}{\partial \alpha}(\alpha = 0,\ \beta = 1)$, for all variables Q, and we need to find the corresponding derivatives at $\beta = \beta^*$.

Recall that the standard "propagation of singularities" principle, applied to first order hyperbolic systems [5], states that along characteristic curves, where the Cauchy problem is not normally solvable, the transversal derivatives satisfy a (usually coupled) system of first order ordinary differential equations. In our case, we want to apply the principle along the degenerate Γ^+ curve $\alpha = 0$ (parameterized by β). Remarkably, it turns out that the resulting system for the transversal derivatives $\frac{\partial Q}{\partial \alpha}(\alpha = 0,\beta)$, is particularly simple. More precisely, we have the following theorem.

Theorem 2.2. *Let $S(e,\rho,z)$ be the entropy. The functions $\frac{\partial S}{\partial \alpha}(0,\beta)$, $\frac{\partial z}{\partial \alpha}(0,\beta)$ can be obtained by simple integrations involving the Riemann invariants for the associated RP. The function $a(\beta) := \frac{\partial u}{\partial \alpha}(0,\beta)$ satisfies a differential relation of the form,*

$$a'(\beta) = W(\beta)\,, \tag{2.5}$$

where $W(\beta)$ is a known function in terms of $\frac{\partial S}{\partial \alpha}(0,\beta)$, $\frac{\partial z}{\partial \alpha}(0,\beta)$ and the solution of the associated RP.

The proof is carried out by first casting the system (1.1) in characteristic coordinates and then using a suitable (straightforward) elimination procedure, singling out $a(\beta)$. Full details can be found in [3]. However, let us point out the following. Throughout the centered rarefaction wave there cannot be any jumps in S, z or any other flow variable (in fact, z can jump only across a contact discontinuity). Nevertheless, S and z vary along streamlines due to the presence of a thermodynamical "source term" (or "heat release") $k\rho$ in (1.1) (the external field is conservative). Thus, in contrast with the non-reactive fluid dynamical case ($z \equiv$ const.), the entropy is not invariant along streamlines (the flow is not adiabatic). The rate of change of S reflects the strength of the "coupling" between the fluid dynamical and the chemical phases of the flow. For a strong coupling, the attempt to split the two phases in a numerical calculation, could lead to non-physical solutions. This point is discussed in [1].

3. Numerical Examples

The numerical discretization of (1.1) is straightforward. Suppose that we use equally spaced grid-points $r_i = i \cdot \Delta r$ along the r-axis and equal time intervals of size Δt. By "cell i" we shall refer to the interval extending between the "cell-boundaries" $r_{i \pm \frac{1}{2}} = \left(i \pm \frac{1}{2} \right) \cdot \Delta r$. We let Q_i^n denote the average value of a quantity Q over cell i at time $t_n = n \cdot \Delta t$. Similarly, we denote by $Q_{i+\frac{1}{2}}^{n+\frac{1}{2}}$ the value of Q at the cell-boundary $r_{i+\frac{1}{2}}$, averaged over the time interval between t_n and t_{n+1}. Our "Godunov-type" difference scheme for (1.1) is now given by,

$$\begin{aligned} U_i^{n+1} - U_i^n = & - \frac{\Delta t}{\Delta V_i} \left[A(r_{i+\frac{1}{2}}) F(U)_{i+\frac{1}{2}}^{n+\frac{1}{2}} - A(r_{i-\frac{1}{2}}) F(U)_{i-\frac{1}{2}}^{n+\frac{1}{2}} \right] \\ & - \frac{\Delta t}{\Delta r} \left[G(U)_{i+\frac{1}{2}}^{n+\frac{1}{2}} - G(U)_{i-\frac{1}{2}}^{n+\frac{1}{2}} \right] - \Delta t \cdot H(U_i^{n+\frac{1}{2}}), \end{aligned} \tag{3.1}$$

where $\Delta V_i = \int_{r_{i-\frac{1}{2}}}^{r_{i+\frac{1}{2}}} A(r) dr$.

In order to evaluate the boundary terms $U_{i+\frac{1}{2}}^{n+\frac{1}{2}}$ in (3.1), we are using the GRP solution. Thus, assume that at time $t = t_n$ the variation of U across cell i is given by $(\Delta U)_i^n$. Take in (2.1),

$$U_+ = U_{i+1}^n - \frac{1}{2} (\Delta U)_{i+1}^n, \quad U_+' = \frac{(\Delta U)_{i+1}^n}{\Delta r}, \tag{3.2)(a}$$

$$U_- = U_i^n + \frac{1}{2} (\Delta U)_i^n, \quad U_-' = \frac{(\Delta U)_i^n}{\Delta r}, \tag{3.2)(b}$$

then finding the solution (2.2) for the resulting GRP we set,

$$U^n_{i+\frac{1}{2}} = U_0 \ , \quad \left(\frac{\partial U}{\partial t}\right)^n_{i+\frac{1}{2}} = \left(\frac{\partial U}{\partial t}\right)_0 , \tag{3.3}$$

and finally, in (3.1), we have,

$$U^{n+\frac{1}{2}}_{i+\frac{1}{2}} = U^n_{i+\frac{1}{2}} + \frac{\Delta t}{2} \cdot \left(\frac{\partial U}{\partial t}\right)^n_{i+\frac{1}{2}} . \tag{3.4}$$

Thus, the difference scheme (3.1) is linked directly to the analytic solution of the GRP, with a minimal amount of additional processing (such as dissipative mechanisms). The reader is referred again to [3] for more details related to the scheme (3.1) as well as the numerical examples which will be described next (very briefly).

Example 1 (Infinite Reflected Shock). This is a test problem proposed by W.F. Noh [6]. It has no chemistry, but its significance lies in the fact that it has spherical symmetry (hence a singularity at $r = 0$) and yet possesses an exact analytic solution. So, take an infinite sphere of gas which is initially cold (hence $p = 0$ and $\rho = 1$, say) and collapsing toward the center at a uniform (radial) speed, say $u = -1$. It is not difficult to see that the resulting flow consists of an outgoing shock wave (at maximal density ratio, i.e., "infinitely strong") which brings the incoming gas to rest. The density profile ahead of the shock reflects the effect of geometrical convergence. Figure 1 shows the resulting velocity and pressure profiles after 900 cycles (at $\Delta t = 0.25$). The analytic solution is plotted by a solid line and has an excellent matching with the computer solution depicted by dots.

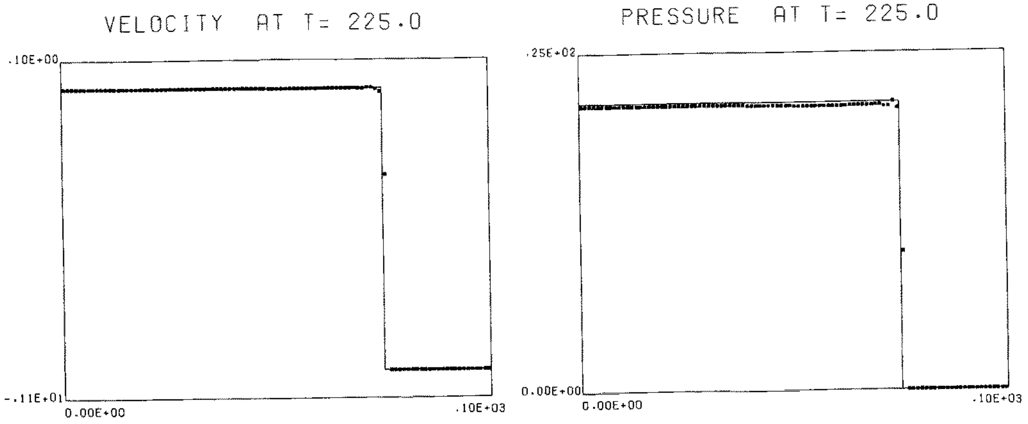


Fig. 1. Infinite reflected shock

Example 2. (Reacting gas in cluster). This example is of astrophysical character, where the geometry is again spherical. It is taken from [8]. We consider a uniform gas, initially at rest, and subject to an external potential simulating gravity. The gas starts to collapse under the external force (directed inward) and pressure, density and temperature begin to rise. When the temperature reaches a critical level, the gas is ignited (at the center) and the resulting outgoing reaction brings the gas to a halt (in a sphere around the center), balancing external force against pressure gradient. Profiles of pressure and velocity are shown in Figure 2, where the reaction front is moving out (ignition started at $r = 0$).

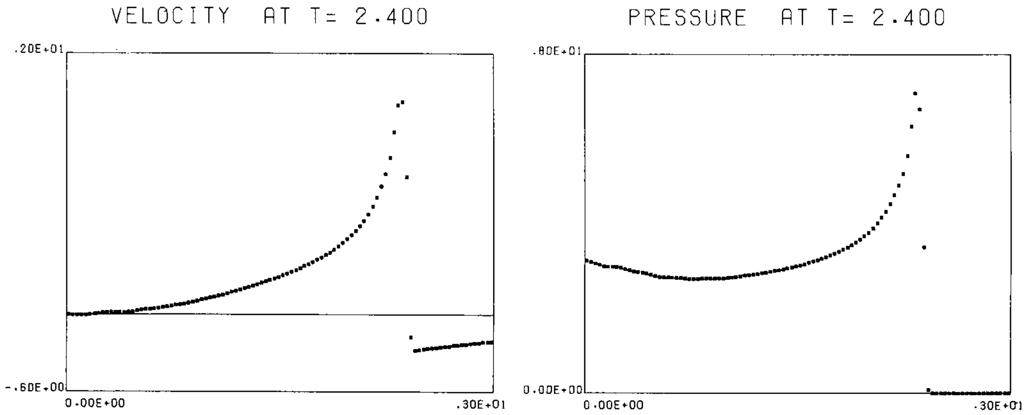


Fig. 2. Reacting gas in cluster

References

[1] M. Ben-Artzi, The generalized Riemann problem for reactive flows, J. Comput. Phys. (in press).

[2] M. Ben-Artzi and J. Falcovitz, A second order Godunov-type scheme for compressible fluid dynamics, J. Comput. Phys. *55* (1984), 1-32.

[3] M. Ben-Artzi and A. Birman, Computation of reactive duct flows in external fields, LBL-25283, Lawrence Berkeley Laboratory, May 1988.

[4] R. Courant and K.O. Friedrichs, "Supersonic Flow and Shock Waves", Interscience, New York, 1949.

[5] R. Courant and D. Hilbert, "Methods of Mathematical Physics II", Interscience, New York, 1962.

[6] W.F. Noh, Unpublished Memorandum, Lawrence Livermore Laboratory, 1982.

[7] B. van-Leer, Towards the ultimate conservative difference scheme, V, J. Comput. Phys. *32* (1979), 101-136.

[8] A. Yahil, M.D. Johnston, and A. Burrows, A conservative Lagrangian hydrodynamical scheme with parabolic spatial accuracy, submitted to J. Comput. Phys.

PROBLEME DE RIEMANN EN HYDRODYNAMIQUE ET APPLICATIONS

A. BOURGEADE, H. JOURDREN, J. OVADIA

Centre d'Etudes de Limeil-Valenton
B.P N° 27
94190 - VILLENEUVE-SAINT -GEORGES

Le problème de Riemann consiste à calculer à l'aide des équations de l'hydrodynamique l'écoulement résultant de la juxtaposition de deux fluides ayant des états à priori différents. Dans les codes numériques un tel problème se pose à chaque interface entre deux mailles et sa solution est fournie par un solveur de Riemann.

Les méthodes numériques faisant appel à un solveur de Riemann ont connu un grand développement ces dernières années. Par contre, leur introduction dans les codes d'hydrodynamique n'a pas eu le même succès. L'une des principales raisons en est que la plupart des solveurs de Riemann ont été écrits pour des gaz parfaits et ne s'appliquent pas à des équations d'état quelconques. En outre, si un solveur de Riemann s'introduit facilement dans un code unidimensionnel il n'en est plus de même pour les codes lagangiens multidimensionnels car il faut alors trouver la vitesse des noeuds qui n'est plus fournie par le solveur de Riemann.

Le but de cet article est de présenter plusieurs types de solveur de Riemann et de montrer comment ils ont été introduits dans différents codes de calcul : un code à grille variable, un code eulérien multifluide et un code eulérien pour les écoulements réactifs.

I - LES SOLVEURS DE RIEMANN

Suivant le type de schéma qui est considéré, plusieurs solveurs de Rìemann peuvent être envisagés.

- Le solveur de Riemann simple est utilisé dans les codes lagangiens et dans les codes eulériens ou à grille variable, comportant une phase lagrangienne. Il calcule la pression et la vitesse à l'interface pour le problème de Riemann considéré.

- Le solveur de Riemann direct est utilisé dans les codes eulériens, ou à grille variable, sans phase lagrangienne. Il calcule directement les flux de masse, de quantité de mouvement et d'énergie, à travers l'interface considérée en fonction de la vitesse de grille.

- Le solveur de Riemann généralisé /1/ est utilisé dans les codes faisant appel à un schéma d'ordre 2 de type Van Leer. Il peut être direct ou non et il calcule la solution du problème de Riemann et sa dérivée par rapport au temps à l'aide non seulement des états de part et d'autre de l'interface mais aussi des gradients des quantités considérées.

Il est possible de résoudre un problème de Riemann quelle que soit l'équation d'état. Néanmoins, afin d'éviter des calculs trop longs, il est préférable que l'équation d'état permette de définir analytiquement les isentropiques et les courbes de choc. La plus simple des équations d'état ayant cette propriété est l'équation d'état binomiale :

$$e = \frac{P + \gamma P_o}{(\gamma + 1)} (\tau + \tau_o) - e_o$$

Cette équation d'état est une généralisation de l'équation d'état des gaz parfaits qui permet de considérer non seulement des gaz ($P_o = 0$) mais aussi des solides. Le solveur de Riemann est alors, à l'introduction près des paramètres P_o et τ_o, identique à ceux utilisés pour l'équation d'état des gaz parfaits. Dans ce qui suit τ_o est toujours nul et, γ, P_o et e_o sont déterminés, pour chaque matériau, de façon à bien approcher la courbe d'Hugoniot définie à partir de l'état initial du matériau. Les calculs avec les différents codes montrent que le choix des paramètres n'influence que faiblement les résultats à condition de faire appel à la véritable équation d'état pour tout ce qui ne concerne pas le solveur de Riemann.

II - APPLICATIONS ET RESULTATS NUMERIQUES

II.1 - GAIA : UN CODE D'HYDRODYNAMIQUE BIDIMENSIONNEL ET À GRILLE VARIABLE.

GAIA est un code bidimensionnel qui utilise plusieurs méthodes dérivées des travaux de Godounov et ses collaborateurs. A chaque bras le solveur de Riemann direct est employé de sorte qu'il n'y a ni phase lagrangienne ni phase de projection. Le code ne fait pas appel à la méthode des directions alternées

SCHÉMAS DE GODOUNOV (x,y) ET (ξ,η)

Les équations de l'hydrodynamique en variables eulériennes :

$$\frac{\partial U}{\partial t} + \frac{\partial}{\partial x} F(U) + \frac{\partial}{\partial y} G(U) = 0$$

$$U = \begin{matrix} \rho \\ \rho u \\ \rho v \\ \rho e \end{matrix} \quad , \; F(U) = \begin{matrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ (\rho e + p)u \end{matrix} \quad , \; G(U) = \begin{matrix} \rho v \\ \rho uv \\ \rho v^2 + p \\ (\rho e + p)v \end{matrix}$$

s'écrivent sous forme intégrale : $\iint U \, dx \, dy + F \, dy \, dt + G \, dt \, dx = 0$.

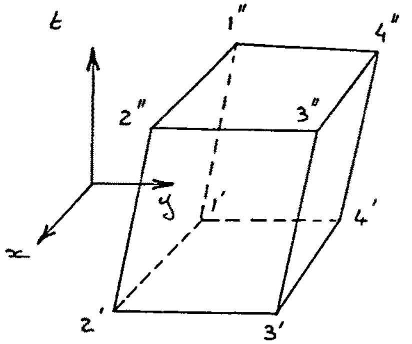


Fig. a

• Le schéma de Godounov grille variable cartésien est construit (cf. fig. a) en appliquant cette formulation à la surface fermée (1', 2', 3', 4', 1", 2", 3", 4") délimitée par les 4 sommets d'une maille au cours de son déplacement arbitraire sur le pas de temps Δt. En adoptant les notations :

$$\Omega^{n+1} = \iint_{(1",2",3",4")} dx \, dy \; , \; \Omega^{n} = \iint_{(1',2',3',4')} dx \, dy$$

$$\Omega_{ij} = \iint_{(i',j',j",i')} dx \, dy \; , \; \Phi_{ij} = \iint_{(i',j',j",i")} dy \, dt \; ,$$

$$\phi_{ij} = \iint_{(i',j',j",i")} dt \, dx \; , Q_{ij} = U_{ij} \, \Omega_{ij} + F_{ij} \, \Phi_{ij} + G_{ij} \, \psi_{ij}$$

où U_{ij}, F_{ij}, G_{ij} désignent les valeurs moyennes des quantités U, F et G sur le côté [ij] à l'instant t + Δt/2, le schéma numérique reliant U^n à U^{n+1}, quantités conservatives au centre de la maille aux instants t et t+Δt, s'écrit :

$$U^{n+1} \, \Omega^{n+1} = U^n \, \Omega^n - Q_{12} - Q_{23} - Q_{34} - Q_{41}$$

En calculant les positions moyennes des sommets :

$$x_i = \frac{1}{2} (x_i + x_{i"}) \; , \; y_i = \frac{1}{2} (y_{i'} + y_{i"})$$

en approchant les surfaces Ω_{ij}, Φ_{ij}, Ψ_{ij} par :

$$\Omega_{ij} = \frac{1}{2} \left[(x_{j''}-x_{i'})(y_{i''}-y_{j'})-(x_{i''}-x_{j'})(y_{j''}-y_{i'})\right] , \quad \Phi_{ij}= \Delta t(y_j-y_i), \quad \Phi_{ij}= \Delta t(x_i-x_j)$$

puis en définissant une vitesse de grille W^*_{ij} telle que :

$$\Omega_{ij} = \Delta t \; \ell_{ij} \; W^*_{ij} \text{ avec } \ell_{ij} = \sqrt{\;} (x_j - x_i)^2 + (y_i - y_i)^2$$

le flux numérique Q_{ij} prend la forme :

$$Q_{ij} = U_{ij} \; \Delta t \; \ell_{ij} \; W^*_{ij} + F_{ij} \; \Delta t \; (y_j - y_i) - G_{ij} \; \Delta t \; (x_j - w_i)$$

Cette expression devient, en introduisant les cosinus directeurs de la normale au côté $\alpha_{ij} = - (y_j - y_i)/\ell_{ij}$, $\beta_{ij} = (x_j - x_i)/\ell_{ij}$

$$Q_{ij} = \Delta t \; \ell_{ij} \; (U_{ij} \; W^*_{ij} - F_{ij} \; \alpha_{ij} - G_{ij} \; \beta_{ij})$$

ou encore, en faisant apparaitre la composante normale de vitesse matière $N_{ij} = \alpha_{ij} \; U_{ij} + \beta_{ij} \; V_{ij}$

$$Q_{ij} = \Delta t \; \ell_{ij} \begin{array}{l} R(W^* - N) \\ RU(W^* - N) - \alpha P \\ RV(W^* - N) - \beta P \\ RE(W^* - N) - NP \end{array}_{ij}$$

Les quantités $(R, P, N, E)_{ij}$ sont calculées par résolution du problème de Riemann selon la direction normale au côté. Si U^*_{ij} désigne la vitesse de la discontinuité de contact, la composante tangentielle de vitesse matière est choisie d'après :

$$T_{ij} = \begin{array}{ll} t^{\text{gauche}}_{ij} & , \text{ si } \; W^*_{ij} < U^*_{ij} \\ T^{\text{droite}}_{ij} & , \text{ si } \; W^*_{ij} > U^*_{ij} \end{array}$$

Les composantes cartésiennes de vitesse nécessaires au calcul du flux numérique Q_{ij} sont reconstituées par $U_{ij} = \alpha_{ij}N_{ij} + \beta_{ij}T_{ij}$, $V_{ij} = \beta_{ij}N_{ij}-\alpha_{ij}T_{ij}$.

- La version curviligne du schéma de Godounov 2-D est sensiblement plus complexe.

Un repère curviligne local (ξ, η) est défini pour chaque maille en début de chaque pas de temps. Les équations d'Euler sont considérées selon les composantes intrinsèques de vitesse (μ_k, ν_n) [resp. (ν_k, μ_n)] pour chaque ligne de coordonnée ξ [resp. η]. Les schémas numériques grille variable est alors bâti comme dans le cas cartésien, par intégration sur l'élément de volume. Des termes apparaissent en second membre des équations de l'impulsion, en raison de la courbure des lignes de coordonnée et de la non-orthogonalité du maillage.

La dérivation complète figure dans l'ouvrage [2]. La variante implantée dans le code GAIA est donnée en référence [3].

EXTENSION TVD D'ORDRE 2

a - Equation de l'advection

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) = 0 \quad \text{où} \quad f(u) = a\,u \quad , \quad a = \text{cste}$$

Le schéma décentré et le schéma de Lax-Wendroff, schémas respectivement d'ordre 1 et 2, s'écrivent en posant $\sigma = a\Delta x/\Delta t$:

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x}(f_{i+½} - f_{i-½})$$

$$f_{i+½}^{d} = \frac{1}{2}(f_i + f_{i+1}) - \frac{\text{sign}(\sigma)}{2}(f_{i+1} - f_i)$$

$$f_{i+½}^{LW} = \frac{1}{2}(f_i + f_{i+1}) - \frac{\sigma}{2}(f_{i+1} - f_i)$$

Une large classe de schémas TVD "quasi" d'ordre 2 s'obtient pour cette équation (voir par exemple /4/, /5/) en introduisant un limiteur ϕ permettant de retrouver le schéma d'ordre 1 sur les discontinuités, là où le schéma d'ordre 2 génèrerait des oscillations.

$$f_{i+½} = f_{i+½}^{d} + \phi\,[f_{i+½}^{LW} - f_{i+½}^{d}] \quad \text{avec } 0 \le \Phi \le 1.$$

b - Système hyperbolique non linéaire de lois de conservation

$$\frac{\partial U}{\partial t} + \frac{\partial}{\partial x} F(U) = 0 \quad \text{où} \quad A = \frac{\partial F}{\partial U}, \ \text{A à valeurs propres réelles.}$$

La méthode de linéarisation de Roe, établie pour les équations de la dynamique des gaz dans le cas gaz parfait, peut s'appliquer avec l'équation d'état binomiale. En calculant la matrice $\tilde{A}_{1+½} = A^{Roe}_{k} (F_i , F_{i+1})$ de valeurs propres $(\tilde{\lambda}_{i+½})$, de vecteurs propres $(\tilde{e}^{k}_{i+½})$, en déterminant $(\alpha^{k}_{i+½})$ d'après :

$$F_{i+1} - F_i = \sum_k (\tilde{\alpha}^{k}_{i+½})\ (\tilde{\lambda}_{i+½})\ (\tilde{e}_{i+½})$$

le schéma scalaire "quasi" d'ordre 2 précédent s'écrit pour le système linéarisé, après quelques manipulations algébriques.

$$F_{i+½} = \frac{1}{2} (F_i+F_{i+1}) - \sum_k (\tilde{\alpha}^{k}_{i+½})\ |\tilde{\lambda}^{k}_{i+½}|(\tilde{e}^{k}_{i+½}) + \frac{1}{2} \sum_k \phi_k\ [\mathrm{sign}(\sigma_k) - \sigma_k](\tilde{\lambda}^{k}_{i+½})\ (\tilde{e}^{k}_{i+½})^r$$

Les deux premiers termes constituent le flux décentré de Roe qui, pris isolément, aboutit à un schéma numérique du premier ordre. Le dernier terme s'interprète ainsi comme un flux d'antidiffusion.

La spécificité du schéma numérique d'ordre 2 développé dans le code GAIA est d'utiliser ce flux d'antidiffusion de Roe avec le flux de Godounov (i.e. solution du problème de Riemann).

c - Hydrodynamique 2-D grille variable (ξ, η)

L'implantation en deux dimensions d'espace est réalisée par la méthode traditionnelle de "splitting", mais en variables curvilignes c'est-à-dire en considérant dans la détermination du flux numérique selon la direction ξ du maillage logique

$$(\text{resp. } \eta)\ \frac{\partial}{\partial t} \Psi' + A' \frac{\partial}{\partial \xi} \Psi' = B',\ [\text{resp. } \frac{\partial}{\partial t} \Psi'' + A'' = B'']\text{avec } \Psi' = \begin{vmatrix} \mu_n \\ \nu_k \\ \rho \\ p \end{vmatrix},\ \Psi'' = \begin{vmatrix} \nu_n \\ \mu_k \\ \rho \\ p \end{vmatrix}.$$

Les flux d'antidiffusion sont calculés par la méthode précédemment exposée pour être ajoutés aux flux de Godounov.

A l'heure actuelle la discrétisation de certains termes sources, notamment ceux de non-orthogonalité du maillage, est encore réalisée à l'ordre 1.

RÉSULTATS NUMÉRIQUES

1/ Implosion sphérique

Les performances du code GAIA en calcul d'implosion sont illustrées fig. 1, 2, par le cas test sévère proposé par W. NOH /6/ associant une forte compression isentropique à un choc d'intensité infinie. Le schéma numérique de von Neumann-Richtmyer concède, en utilisant la formulation classique de pseudoviscosité, d'importantes erreurs au pôle et sur le niveau du choc retour. Ces erreurs sont particulièrement bien corrigées par le schéma GAIA d'ordre 2.

2/ Choc oblique

L'exemple suivant permet de vérifier la qualité du schéma quant aux phénomènes 2-D d'interaction entre matériaux. La figure 3 montre une géométrie correspondant à une transmission de choc (41,5 Gpa) d'un matériau d'impédance forte vers un autre matériau d'impédance faible avec une incidence de 17 degrés. Sur les figures 4 et 5, la très bonne linéarité des isobares et isodensités pour les chocs incidents transmis, ainsi que pour la détente réfléchie souligne l'excellente précision en présence d'un maillage grossier a priori imparfaitement adapté à la simulation du phénomène.

3/ Relèvement de cylindre

Le dernier exemple présente un relèvement de cylindre (explosif TATB-cuivre) obtenu en utilisant un maillage relativement grossier (cf. fig. 6). Nous donnons fig.7 et 8 le maillage utilisé avec le code GAIA traitant en grille variable le glissement et l'advection dans le sens longitudinal et le maillage utilisé avec un code lagrangien multibloc. Les courbes de la figure 9 soulignent à nouveau la parfaite "propreté" de la mise en vitesse.

II.2 - CEE-R : UN CODE EULÉRIEN MULTIFLUIDE

Le code CEE-R est un code bidimensionnel qui utilise la méthode des directions alternées. En outre, le schéma employé dans chaque direction est d'ordre 1 et se décompose en une phase lagrangienne utilisant un solveur de Riemann et une phase de projection basée sur la méthode SLIC.

Le code est multifluide, d'où quelques difficultés pour introduire un solveur de Riemann. Chaque maille peut contenir plusieurs matériaux et pour le solveur de Riemann il faut définir un état et une équation d'état de chaque côté du bras considéré.

L'équation d'état de mélange est obtenue à partir de la relation :

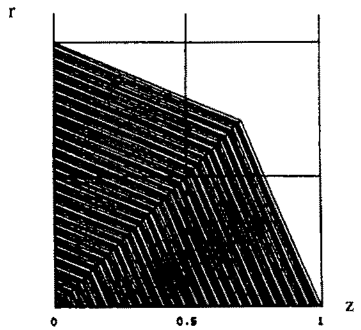


Fig. 1.a

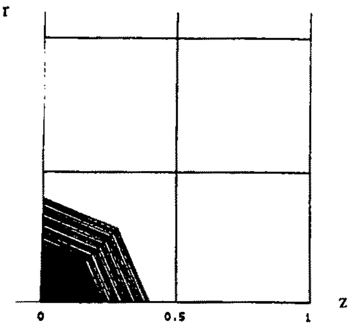
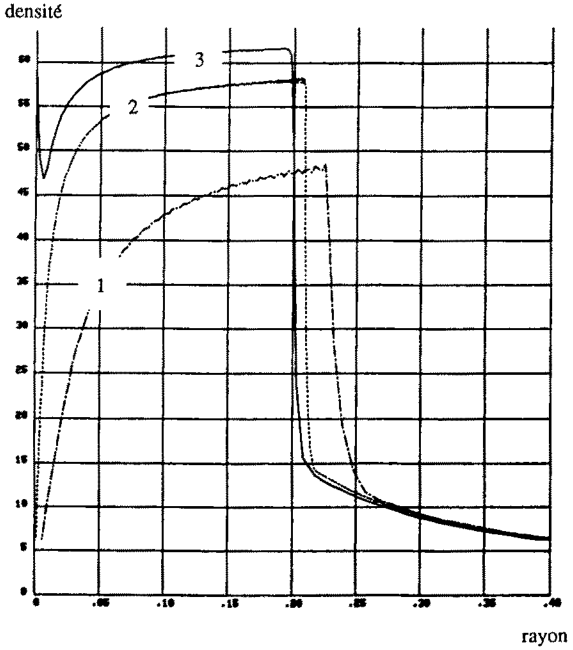


Fig. 1.b

Fig. 2.



IMPLOSION SPHERIQUE : cas test de Noh ($r_{max} = 1$, $\rho = 1$, $u_r = -1, p = 0$, $\gamma = 1.4$)

Fig 1.a Maillage initial GAIA

Fig 1.b Maillage final GAIA

Fig 2. Profils de densité ($\rho^{MAX}_{theorique} = 64$)

- courbe 1: schéma 1-D de **Rytchmyer** (pseudo quadratique) 100 couches
- courbe 2: schéma 1-D de **Rytchmyer** (pseudo quadratique) 400 couches
- courbe 3: schéma 2-D **GAIA** 100 couches

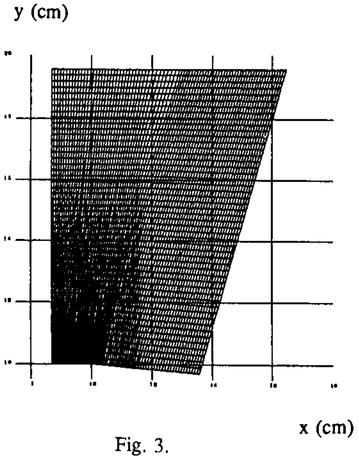


Fig. 3.

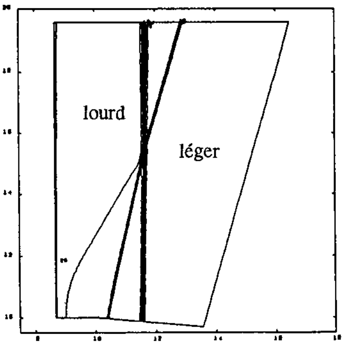


Fig 4.

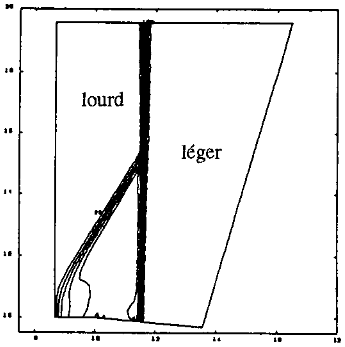


Fig. 5.

CHOC OBLIQUE 2-D : lourd - léger

Fig. 3. Géométrie et maillage

Fig. 4. Isodensités

Fig. 5. Isobares

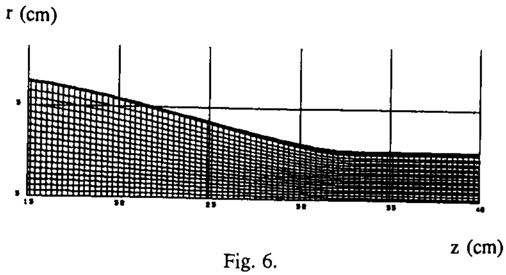


Fig. 6.

Fig. 7.

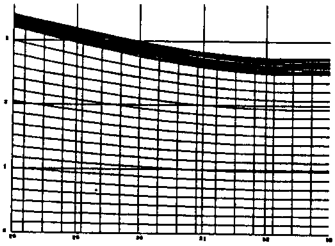
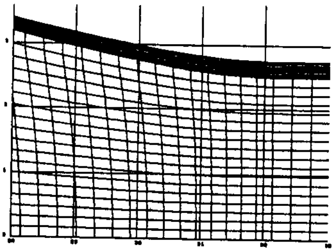


Fig. 8.



RELEVEMENT DE CYLINDRE : TATB - cuivre

Fig. 6. Géométrie et maillage

Fig. 7. Maillage GAIA grille variable monobloc

Fig. 8. Maillage code lagrangien multibloc avec glissement

Fig. 9. Courbes de mise en vitesse

- trait plein : schéma GAIA d'ordre 2
- trait pointillé : schéma de Wilkins multibloc

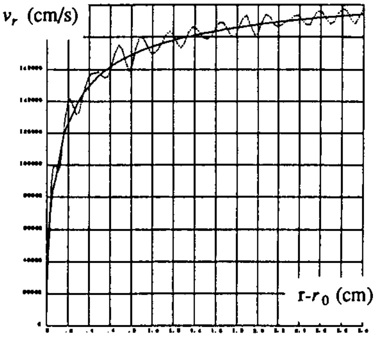


Fig. 9.

$$me = \sum_k m_k e_k ,$$

définissant l'énergie interne spécifique (e) d'une maille mixte (i.e. contenant plusieurs matériaux) à l'aide des énergies internes spécifiques de chacun des matériaux (e_k). En introduisant l'équation d'état binomiale pour le mélange et pour chaque matériau la relation devient :

$$\text{(III.1)} \qquad \rho\, V \frac{P + \gamma P^o}{(\gamma-1)\, \rho} = \sum_k \rho_k\, V_k \frac{P_k + \gamma_k\, P_k^o}{(\gamma_k-1)\, \rho_k}$$

Cette relation doit être toujours vérifiée. En particulier si le mélange est isobare elle s'écrit :

$$\frac{P + \gamma\, P^o}{\gamma - 1} V = \sum_k \frac{P + \gamma_k\, P_k^o}{(\gamma_k - 1)} V_k$$

Cette dernière égalité est valable quelle que soit la valeur de P, par conséquent nous obtenons :

$$\gamma = 1 + \frac{V}{\sum_k \frac{V_k}{\gamma_{k-1}}},$$

et

$$P^o = \frac{\gamma-1}{\gamma V} \sum_k \frac{\gamma_k\, P_k^o}{\gamma_{k-1}} V_k .$$

Reste maintenant à définir des valeurs moyennes dans les mailles mixtes. S'il est facile d'obtenir une densité moyenne à partir de la masse totale des matériaux et une vitesse moyenne à partir de la quantité de mouvement totale, le cas de la pression est plus délicat. Pour cela, il faut revenir à la relation (III.1) qui, en faisant intervenir la définition de γ et P^o, se simplifie pour donner :

$$\frac{PV}{\gamma-1} = \sum_k \frac{P_k\, V_k}{\gamma_{k-1}} ,$$

soit

$$P = \frac{(\gamma-1)}{V} \sum_k \frac{P_k\, V_k}{(\gamma_{k-1})} .$$

Le solveur de Riemann peut donc être utilisé à chaque bras, même dans le cas de mailles mixtes, et définit une pression et une vitesse sur le bras considéré. Si la phase lagrangienne s'effectue classiquement (en appliquant les lois de conservation) pour les mailles pures (avec un seul matériau), pour les autres il faut non seulement calculer les accroissements (algébriques) de volume, de quantité de mouvement et d'énergie mais aussi répartir ces accroissements entre les divers matériaux constituant la maille considérée.

Tout d'abord, les accroissements de volumes δV_k sont répartis en fonction des accroissements calculés à l'aide des équations d'état simplifiées. Les accroissements de quantité de mouvement sont répartis au prorata des masses. Cette répartition faite il est possible de calculer un accroissement d'énergie cinétique pour chaque matériau et donc aussi pour la maille. Finalement, l'accroissement d'énergie interne, obtenu par différence entre l'accroissement d'énergie totale et celui d'énergie cinétique, est réparti en fonction des produits $P_k \bullet \delta V_k$.

RÉSULTATS NUMÉRIQUES

1/ Instabilités de Rayleigh-Taylor

L'utilisation d'un solveur de Riemann a permis de passer des calculs d'instabilités en milieu quasi incompressible et d'obtenir des résultats dont la qualité peut être appréciée sur la planche 10 qui représente l'état de l'interface entre le gaz lourd et le gaz léger à différents instants.

2/ Implosion sphérique

Les figures de la planche 11 montrent l'évolution d'une sphère composée de trois matériaux concentriques (tantale, molybdène et air) et soumise à une pression extérieure de 500 kilobars. Bien que la méthode soit eulérienne et d'ordre un, la sphéricité est bien conservée.

II.3 - ARES : UN CODE POUR LES ÉCOULEMENTS RÉACTIFS

Pour résoudre les équations de l'hydrodynamique,

$$\frac{\partial \rho}{\partial t} + v \frac{\partial \rho}{\partial r} + \rho \frac{\partial v}{\partial r} + \chi \rho v = 0,$$

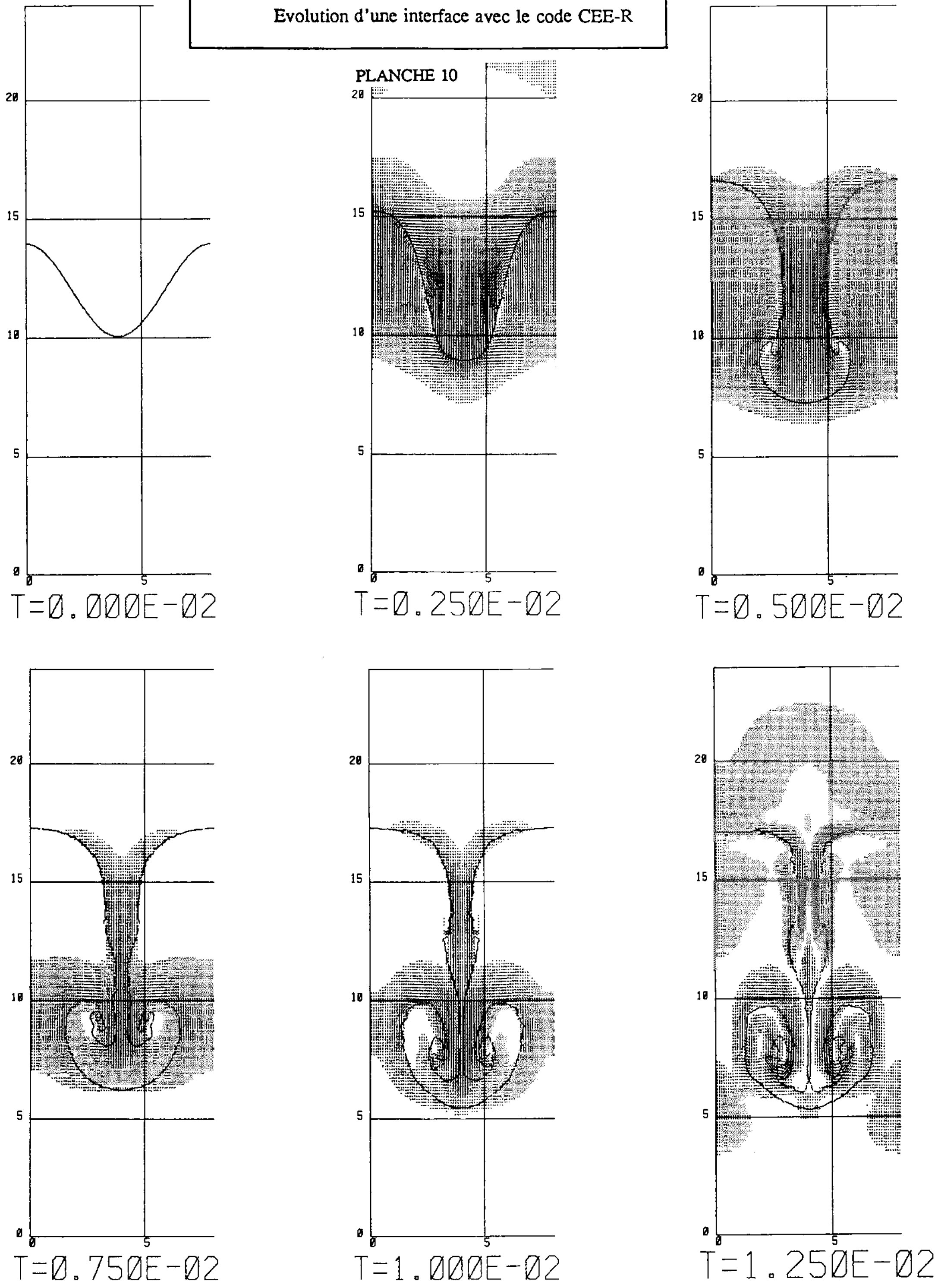
$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial r} + \frac{1}{\rho} \frac{\partial P}{\partial r} = 0,$$

$$\frac{\partial E}{\partial t} + v \frac{\partial E}{\partial r} + \frac{1}{\rho} \frac{\partial (Pv)}{\partial r} + \chi \frac{Pv}{\rho} = 0 \;;$$

INSTABILITES DE RAYLEIGH-TAYLOR :

Evolution d'une interface avec le code CEE-R

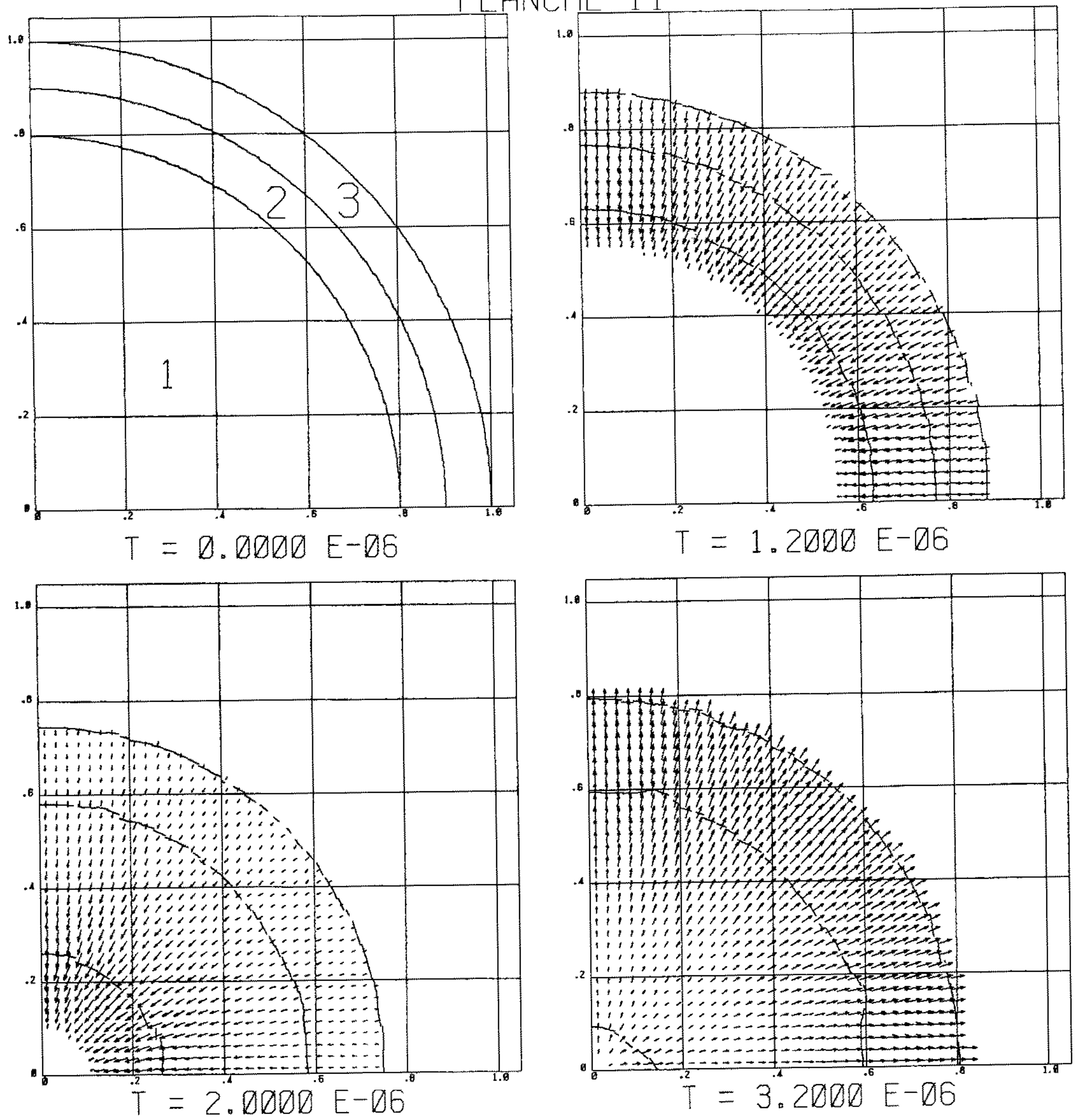
PLANCHE 10



IMPLOSOIR SPHERIQUE : $p_{initiale} = 1$ bar, $p_{exterieure} = 500$ kbar

- milieu 1 : air , milieu 2 : molybdène , milieu 3 : tantale
- géométries avant et après rebond avec le code CEE-R

PLANCHE 11



T = Ø.ØØØØ E-Ø6

T = 1.2ØØØ E-Ø6

T = 2.ØØØØ E-Ø6

T = 3.2ØØØ E-Ø6

couplées à la cinétique chimique

$$\frac{\partial \omega}{\partial t} + v\,\frac{\partial \omega}{\partial r} = f(P,\rho,\omega),$$

et à l'équation d'état binomiale.

$e = \frac{P + \gamma P^0}{(\gamma-1)\rho} - e^0$ où $e = E - \frac{1}{2}\,u^2$; γ, P^0 et e^0 sont des fonctions de ω ; χ est une fonction de r, la méthode utilisée dans ARES et basée à quelques détails près sur le schéma de B.Van Leer /7/ se décompose en deux phases : une phase lagrangienne qui permet de calculer l'évolution du fluide considéré et une phase de remaillage qui traduit cette évolution sur un maillage donné (en général eulérien). Après chacune de ces phases, une procédure de correction est nécessaire pour assurer une bonne stabilité au schéma. Chaque phase définit des valeurs moyennes et des gradients pour chaque maille ; les gradients sont modifiés lors des procédures de correction afin que le schéma soit monotone. L'ordre 2 est obtenu en utilisant une projection d'ordre 2 et un solveur pour le problème de Riemann généralisé tenant compte de la cinétique chimique. Cette méthode a été choisie car c'est une amélioration du schéma de GODOUNOV qui est considéré comme le meilleur schéma monotone d'ordre 1. Par ailleurs, la décomposition en deux phases permet de greffer plus facilement une méthode de suivi de front.

Pour les écoulements réactifs la méthode des directions alternées est légèrement modifiée en raison de la cinétique chimique. En effet, celle-ci ne peut être traitée à chaque balayage car la méthode ferait alors dégager l'énergie deux fois trop vite. Pour palier cet inconvénient, la cinétique n'intervient à chaque pas de temps que lors du second balayage, le premier balayage se contentant de convecter les fractions brûlées. Cette méthode a l'avantage d'être moins diffusive que celle qui consisterait à appliquer la cinétique à chaque balayage mais pour un demi pas de temps. Elle permet en particulier d'utiliser des maillages plus grossiers sans trop détériorer la solution.

Suivi de front unidimensionnel

Lors de la phase de projection d'un calcul unidimensionnel, le maillage sur lequel on projette la solution obtenue après la phase lagrangienne peut ne pas être fixe mais varier d'un instant à l'autre (i.e. la méthode qui a été mise en oeuvre peut s'appliquer aussi bien en Euler qu'en grille variable). En effet, pour améliorer la précision du calcul il est parfois utile de pouvoir suivre un front de discontinuité (front de détonation, choc, discontinuité de contact).

Pour introduire la position du front à un instant donné dans le maillage la première méthode envisagée consiste simplement à remplacer le noeud qui suit immédiatement le front par le front lui-même (voir figure P1). Mais cette méthode présente l'inconvénient d'introduire parfois des mailles de dimensions trop faibles. C'est pourquoi il est alors préférable de remplacer le noeud le plus proche du front et non plus systématiquement celui qui suit (voir figure P2).

Quelle que soit la méthode choisie il n'y a pas de modification à apporter à la phase de projection, une fois que le nouveau maillage a été défini. Par contre, pour la phase lagrangienne, il convient de distinguer entre les discontinuités de contact qui sont des surfaces fluides et les autres fronts.

Dans le premier cas la phase lagrangienne ne subit aucune modification et permet de définir la nouvelle position du front (voir figures Q1 et Q2).

Dans le second cas il convient d'ajouter à nouveau une maille correspondant à la masse de fluide traversée par le front entre l'instant t et l'instant t+Δt (voir figure R). Les valeurs moyennes dans les mailles de part et d'autre de la position du front à l'instant t+Δt sont déterminées par la conservation de la masse, de la quantité de mouvement et de l'énergie totale. Les pentes sont calculées à l'aide des valeurs connues ou obtenues de part et d'autre du front. Dans le cas d'une détonation C-J ces quantités sont fournies par la théorie, tandis que dans le cadre d'une discontinuité de contact ou d'un choc, elles sont données par la solution du problème de Riemann généralisé correspondant à la discontinuité considérée.

Les deux versions (bidimensionnelle et unidimensionnelle avec suivi du choc initiant la détonation) du code ARES ont été testées.

Dans le cas unidimensionnel l'exemple traité est une transition choc- détonation de type Forest Fire. Le choc initiant la détonation est réactif, c'est-à-dire qu'une partie de l'explosif est décomposée lors du passage du choc. La planche 12 représente les "Pop Plot" théorique et numérique reliant la pression (P) à la distance parcourue par le choc (x), les profils de pression et de fraction brûlée à différents instants et enfin les courbes de vitesse du choc et de pression sur le choc en fonction du temps.

Dans le cas bidimensionnel l'exemple traité est le contournement d'un coin par une onde de détonation. La planche 13 représente les isocontours de pression, densité, vitesses et fraction brûlée à l'issue du calcul.

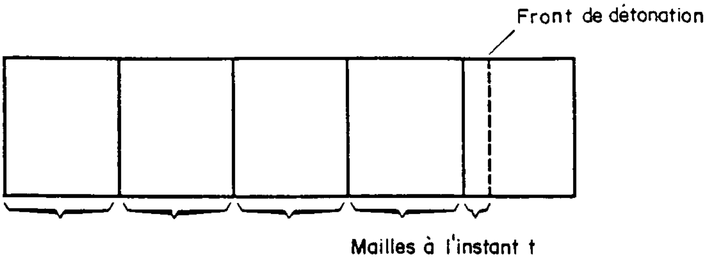


Figure P1

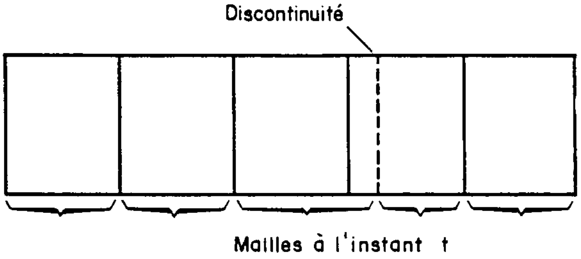
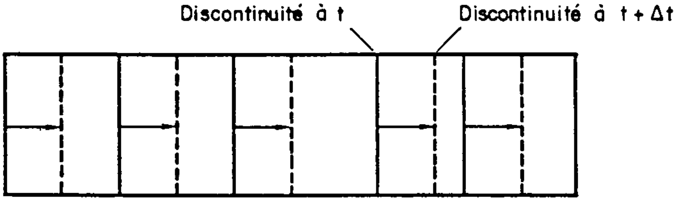
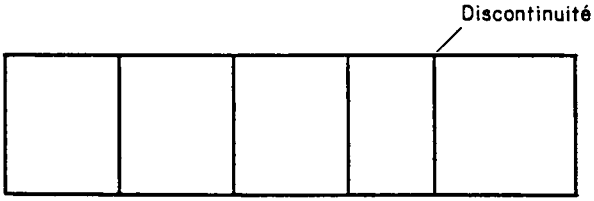


Figure P2



Maillage à l'instant t et maillage lagrangien à l'instant t + Δt

Figure Q1



Maillage à l'instant t + Δt

Figure Q2

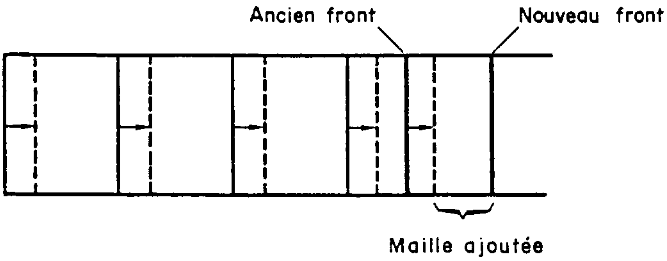
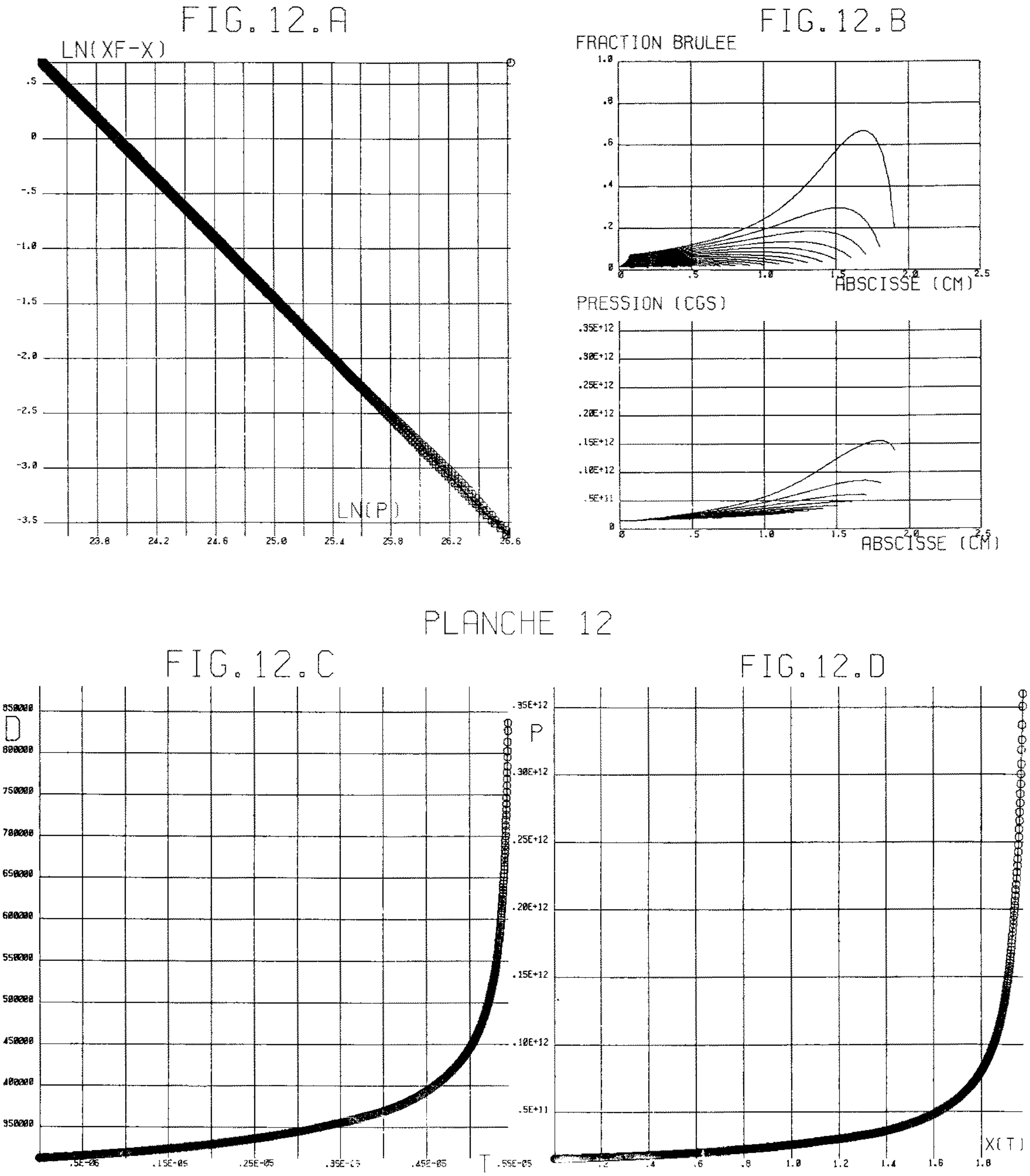


Figure R

PLANCHE 12



TRANSITION CHOC-DETONATION : explosif PBX-9404

Fig. 12.a Superposition des "Pop plots" théorique et numérique

Fig. 12.b Etablissement de la détonation

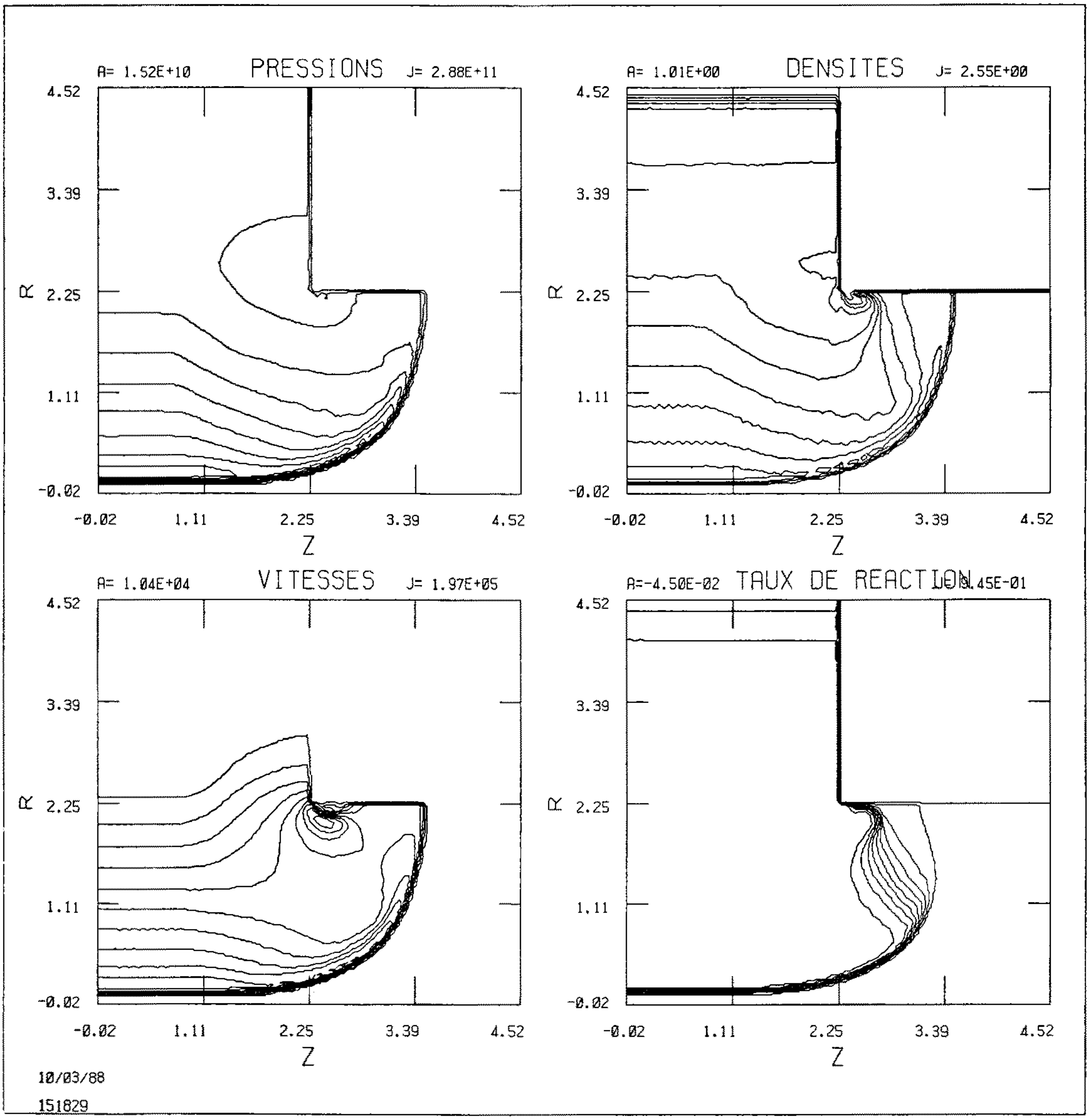
Fig. 12.c Vitesse du front réactif en fonction du temps

Fig. 12.d Pression sur le front réactif en fonction de la distance parcourue

CONTOURNEMENT D'UN COIN : explosif PBX-9502

- Amorçage plan, cinétique "Forest Fire"
- Isovaleurs après le contournement

PLANCHE 13



10/03/88

151829

IV - CONCLUSION

Les exemples précédents montrent qu'il est possible d'introduire avec succès les méthodes numériques utilisant un solveur de Riemann dans des codes d'hydrodynamique.

Il faut noter cependant qu'aucun des codes présentés n'est lagrangien. En fait, si le solveur de Riemann fournit une vitesse au milieu des bras, les codes lagrangiens ont besoin d'une vitesse aux noeuds. Malgré de nombreux travaux /8/, il semble qu'aucune solution vraiment satisfaisante n'ait été trouvée.

REFERENCES

/1/ A. BOURGEADE
"Quelques méthodes numériques pour le traitement des écoulements réactifs" Note CEA N-2570 (juin 1988).

/2/ S.K. GODOUNOV et al.
"Résolution des problèmes multidimensionnels de la dynamique des gaz" - Edition Mir, Moscou (1979).

/3/ H. JOURDREN
"Rencontre 87 CEA/DAM-LANL" - preprint (1987).

/4/ A. HARTEN
J. Comput. Phys. 49 (1983), 357.

/5/ P.K. SWEBY
SIAM J. Numer. Anal. 21 (1984), 995.

/6/ W.F. NOH
J. Comput. Phys. 72 (1988), 78.

/7/ B. VAN LEER
J. Comput. Phys. 32 (1979), 101.

/8/ C. CHERFILS
Thèse PARIS VI (1988).

REVIEW OF FLOW SIMULATIONS USING LATTICE GASES

D. d'Humières, P. Lallemand, Y.H. Qian
Laboratoire de Physique Statistique
associé au CNRS et à l'Université Pierre et Marie Curie
Ecole Normale Supérieure
24 Rue Lhomond
75231 Paris Cedex 05

Abstract

Lattice gases are first defined as a new way to perform molecular dynamics calculations in a simplified manner but at large speeds thus allowing to consider enough particles to simulate real flows. Some results of the statistical analysis of lattice gases are summarized showing that their macroscopic behaviour follows closely that of real fluid flows at low Mach numbers. An example of a two-dimensional flow simulation using a simple lattice gas model is given. Finally new results are presented concerning one-dimensional shocks studied by numerical simulations for two-dimensional lattice gases, showing good agreement with theoretical analysis.

I Introduction

Among the various techniques used to analyze fluid flows a general yet rarely used method for numerical simulations, consists in describing the fluid as a collection of interacting particles. This is called *Molecular Dynamics.* Usually N particles are considered. Their locations $\mathbf{r}_i(t)$ can be calculated by solving numerically Newton's equations with interparticle forces that are derived from a two-body potential $V(\mathbf{r}_i - \mathbf{r}_j)$. This has been used for the last 30 years or so, mostly to obtain detailed information about dense gases and liquids. In those cases where the aim was to determine thermodynamical and transport properties[1], usually a fairly small number of particles were required, typically $N \simeq 1000$. These numbers are obviously too small to be represent any macroscopic flow where macroscopic scales are equal to many times the particle mean free path, itself usually larger than the range of the interparticle potential. As a result the numerical calculations have to be performed with a discretized time scale that is much smaller than any macroscopic time scale of the system. With the availability of powerful computers some studies have been performed with significantly larger numbers of particles, 10^5 or more, and a few results of true molecular dynamics calculations of real flows are available in the literature[2].

In the context of the theory of fluids, the microscopic nature of real systems is usually considered in the framework of the Boltzmann equation. A large body of literature has been devoted to the question of the relationship between the microscopic and the macroscopic descriptions of fluids. Due to the difficulty of the problem, several approximate ways have been used. One of them consists in limiting drastically the phase space by taking a discrete set of possible velocities for the particles : the Broadwell gas[3] is a typical example of this approach that has allowed to obtain important results on shock waves[4]. Lattice gases, as first introduced by Hardy, de Pazzis and Pomeau (HPP)[5], involve even more drastic approximations. Particles move along the links of a regular lattice and time is discretized in such a way that particles occupy successively the nodes ot the lattice. "Collision" events occur between particles that occupy the same node of the lattice at a given time. As we have not specified the detailed dynamics of collisions of these point particles on nodes of the lattice there is no way to determine the outcome of a given precollision state. However we can choose at will the result of any collision event but it is preferable to try and satisfy conservation laws of physics if we wish to design models that have some physical relevance. For point particles we shall impose conservation of mass, linear momentum and energy.

A lattice gas will therefore be defined by the geometry of the lattice, the rules of their motions from node to node on the lattice and the choice of "collision" events. A description of some lattice gases will be given in part II of this review together with a summary of the theoretical analysis of such systems. Part III will give some information concerning their use for flow simulations. Part IV will present new results concerning shock waves in these systems that appear to be of interest in a situation which was considered by previous studies beyond their range of application.

II Lattice gas models

As indicated in the introduction to define a lattice gas we first need to choose a regular lattice. The early work of HPP used a square lattice with particles hoping from one node to one of its nearest neighbour. Particles are undistinguishable so that instead of using a lagrangian description, as it is commonly done in molecular dynamics, an eulerian description is preferred. The system is fully determined by the set of numbers $n_i(\mathbf{r}_j, t_k)$ where $i \in \{1, \cdots, 4\}$ represents the four directions of space from one node to its four nearest neighbours and $\mathbf{r}_j$ labels the nodes of the lattice. An additional assumption, which simplifies the subsequent numerical simulations, is the introduction of a Boolean character for the particles, which means that any of the numbers n_i is equal to 0 or 1. In a physicist's language one speaks of an exclusion principle like for electrons in a semiconductor where distribution functions are of the Fermi-Dirac type. Time t_k runs in a discrete manner, $t_k = k\tau$. If l is the distance between adjacent nodes of the lattice, one may define a unit velocity $c = l/\tau$. Most of the subsequent expressions will be given in terms of reduced values of time, velocity and mass if all particles have the same mass m.

The early HPP model has been extended in recent years to different lattice geometries[6,7] and in some cases to larger velocity spaces than just considering elementary motions to the nearest neighbours[8]. In addition in some cases particles carry a label as will be indicated later when discussing simulations of reaction-diffusion phenomena[9].

The microdynamics of the lattice gas is defined as the knowledge of the set of numbers $n_i(\mathbf{r}_j, t_k)$ for $i = 1, \cdots, b$ if b is the number of possible velocities $\mathbf{c}_i$, for all $j \in \mathcal{L}$, if $\mathcal{L}$ includes all the nodes of the lattice and for all values of the discretized time t_k.

Following the indications for the dynamics of the system given in the introduction, this is obtained by considering a succession of steps : a propagation step and a collision step for each time increment.

Both these steps can be represented by operators :

$$\mathcal{S} : n_i(\mathbf{r}, t+1) \rightarrow n_i(\mathbf{r} - \mathbf{c}_i, t)$$

for the propagation step and

$$\mathcal{C} : n_i(\mathbf{r}, t) \rightarrow n_i(\mathbf{r}) + \Delta_i(n(\mathbf{r})).$$

for the collision step. One complete time step of the evolution of the system is thus described by the microdynamical equation of motion

$$n_i(\mathbf{r}, t+1) = n_i(\mathbf{r} - \mathbf{c}_i, t).$$

The propagation step is deterministic, whereas the collision step can either be deterministic or not depending whether a given input precollision state $\{n_i\}$ leads to one or a number of postcollision state $\{n'_i\}$ with a probability $A(\{n_i\} \rightarrow \{n'_i\})$. Most lattice gas models involve only local collision which means that precollision states depend only upon the particles present on one node.

The collision probabilities A satisfy general properties. If we call s the initial precollision state $\{n_i\}$ and s' the postcollision state $\{n'_i\}$ we have

$$\begin{aligned} A(s \rightarrow s') &\geq 0 \\ \sum_{s'} A(s \rightarrow s') &= 1 \qquad \forall s \\ \sum_i (n'_i - n_i) A(s \rightarrow s') a^l_i &= 0, \qquad \forall s, s' \end{aligned}$$

where a^l_i is one of the appropriate functions of the cartesian components of the velocities $\mathbf{c}_i$, chosen in order to express conservation of mass, linear momentum or energy in collisions.

The microdynamics is well suited for numerical simulations on a digital computer, as $\mathcal{S}$ involves changes of address in the computer memory and $\mathcal{C}$ involves either consultation of a table of predetermined values of the outcome of a collision event or a small number of boolean operations to compute the value of the quantity $\Delta_i(n(\mathbf{r}))$.

However, as stated in the introduction we are interested in the macroscopic behaviour of the system. Thus we don't need to have a full microscopic description of the lattice gas. We could view the problem as a statistical one where we have to perform ensemble averages over microscopic realisations of the same macroscopic situation. However the lattice gas may present a non linear behaviour and thus ensemble averages are not appropriate, and thus we consider the behaviour of one system as a function of time (assuming ergodicity) and perform spatial averages of the required quantities over a small volume of space (for instance over 16×16 sites in the case of a two dimensional problem).

We shall summarize the results of statistical analysis first performed for the simpler lattice gases by HPP for the square model, then by several authors for other lattices[10-12]. Let Γ be the phase space of the lattice gas, a particular state of the gas is defined by the set of numbers

$s(\cdot) = \{n_i(\mathbf{r}^*), i = 1, \cdots, b,\ \mathbf{r}^* \in \mathcal{L}\}$. First we have to define ensemble averages of the numbers n_i over a set of representations $s(\cdot)$ that occur with probability $P(0, s(\cdot)) \geq 0$ and such that

$$\sum_{s(\cdot)\in\Gamma} P(0, s(\cdot)) = 1.$$

$$N_i(\mathbf{r}, t) = \sum_{s(\cdot)\in\Gamma} P(0, s(\cdot)) n_i(\mathbf{r}, t).$$

From the knowledge of the average quantities N_i, one may derive values for the local fluid density, mass flux and velocity

$$\rho(\mathbf{r}, t) = \sum_i N_i(\mathbf{r}, t).$$

$$\mathbf{j}(\mathbf{r}, t) = \sum_i \mathbf{c}_i N_i(\mathbf{r}, t).$$

$$\mathbf{j}(\mathbf{r}, t) = \rho(\mathbf{r}, t)\mathbf{u}(\mathbf{r}, t).$$

Although it has not been proven that the system is ergodic, it is possible to analyze the problem using tools similar to those developped for standard continuous models of real gases. In particular one may define an entropy

$$S = \sum_i \sum_{\mathbf{r}^*\in\mathcal{L}} N_i(\mathbf{r}^*) \log(N_i(\mathbf{r}^*)) + (1 - N_i(\mathbf{r}^*)) \log(1 - N_i(\mathbf{r}^*)).$$

Similarly one may write a discrete Liouville equation for the time evolution of the probability distribution P

$$P(t + 1, \mathcal{S}s'(\cdot)) = \sum_{s(\cdot)\in\Gamma} \prod_{\mathbf{r}^*\in\mathcal{L}} A(s(\mathbf{r}^*) \rightarrow s'(\mathbf{r}^*)) P(t, s(\cdot)).$$

This equation just expresses that the probability at $t + 1$ of a given configuration $s'(\cdot)$ is the sum of the probabilities at t of all possible original configurations $s(\cdot)$ times the transition probability. At equilibrium and assuming that there is decoupling of the probability distributions on different sites, the local distribution can be calculated from the collision transitions $A(s \rightarrow s')$. When these collision transitions satisfy the so-called semi-detailed balance condition

$$\sum_s A(s \rightarrow s') = 1, \quad \forall s',$$

then the N_i are given by Fermi-Dirac distributions

$$N_i = \frac{1}{1 + \exp(h + \mathbf{q} \cdot \mathbf{c}_i)}.$$

where h is a real number and $\mathbf{q}$ is a D-dimensional vector that are functions of the macroscopic properties (density and velocity).

From these values of the mean populations N_i, we may calculate the local density ρ and particle flux and thus the local velocity $\mathbf{u}$. Assigning values to the density and velocity, it is possible to obtain expressions for h and $\mathbf{q}$ when the speed is small compared to the particle velocities c. To second order in $\mathbf{u}$, one gets

$$N_i^{eq}(\rho, \mathbf{u}) = \frac{\rho}{b} + \frac{\rho D}{c^2 b} c_{i\alpha} u_\alpha + \rho G(\rho) Q_{i\alpha\beta} u_\alpha u_\beta + \mathcal{O}(u^3),$$

where

$$G(\rho) = \frac{D^2}{2c^4 b} \frac{b - 2\rho}{b - \rho} \quad \text{and} \quad Q_{i\alpha\beta} = c_{i\alpha} c_{i\beta} - \frac{c^2}{D} \delta_{\alpha\beta},$$

D is the dimensionality of space and greek indices are related to the spatial cartesian coordinates.

Note the presence of the factor $G(\rho)$ which vanishes for $\rho = b/2$, that is when there is an equal density of particles and "holes". The presence of this factor $G(\rho)$ is directly linked to the boolean character of the particles. It is possible to change its value by taking slightly more complicated models. Its effect will be seen on the macroscopic equations of motion of the system, but even for equilibrium properties it leads to the lack of galilean invariance of the model.

We now turn to the macrodynamical equations of the lattice gas . This can be done in several ways. We may perform an asymptotic analysis of situations in which macroscopic quantities (density and velocity) vary over a large spatial scale $\mathcal{O}(\epsilon^{-1})$ in terms of the number of lattice sites. We then "glue" together local thermodynamical equilibria with slowly varying parameters. We expect relaxation to local equilibrium to occur with a time scale ϵ^0, density perturbations propagating as sound waves to evolve on a time scale ϵ^{-1} and amplitude of waves to relax on a time scale ϵ^{-2}. This leads us to use several temporal and spatial variables : t_* (discrete), $t_1 = \epsilon t_*$, $t_2 = \epsilon^2 t_*$, $\mathbf{r}_*$ (discrete) and $\mathbf{r}_1 = \epsilon \mathbf{r}_*$ for a multi-scale analysis. In the first order we obtain the "macrodynamical Euler equations"

$$\partial_{t_1} \rho + \partial_{1\beta}(\rho u_\beta) = 0,$$

and

$$\partial_{t_1}(\rho u_\alpha) + \partial_{1\beta} P_{\alpha\beta} = 0.$$

$P_{\alpha\beta}$ is the momentum-flux tensor which is given (to leading order in u) by,

$$\begin{aligned} P_{\alpha\beta} &\equiv \sum_i c_{i\alpha} c_{i\beta} N_i^{eq} \\ &= \frac{c^2}{D} \rho \delta_{\alpha\beta} + \rho G(\rho) T_{\alpha\beta\gamma\delta} + \mathcal{O}(u^4), \end{aligned}$$

with

$$T_{\alpha\beta\gamma\delta} = \sum_i c_{i\alpha} c_{i\beta} Q_{i\gamma\delta}$$

where $G(\rho)$ and $Q_{i\alpha\beta}$ are given above.

We shall discuss later the implication of the tensor $T_{\alpha\beta\gamma\delta}$ upon the isotropy of the system. These equations allow to define a speed of sound that depends upon the geometry of the lattice, the set of velocities and the collision probabilities A. For the simple HPP gas the speed of sound c_s is $c/\sqrt{2}$ in terms of the particle speed c.

In the next order in the ϵ development we obtain the "macrodynamical Navier-Stokes equations" in the form

$$
\begin{aligned}
&\partial_t\rho + \partial_\beta(\rho u_\beta) = 0, \\
&\partial_t(\rho u_\alpha) + \partial_\beta(\rho G(\rho) T_{\alpha\beta\gamma\delta} u_\gamma u_\delta + \frac{c^2}{D}\rho\delta_{\alpha\beta}) \\
&\qquad + \partial_\beta\left[(\psi(\rho) + \frac{D}{2c^2 b}) T_{\alpha\beta\gamma\delta}\partial_\gamma(\rho u_\delta)\right] \\
&\qquad = \mathcal{O}(\epsilon u^3) + \mathcal{O}(\epsilon^2 u^2) + \mathcal{O}(\epsilon^3 u).
\end{aligned}
$$

The first of these macrodynamical equations is identical to the usual continuity equation for standard fluids. The second one bears a strong resemblance to the Navier-Stokes equation. As particles all have the same speed there is no additional equation similar to the heat equation. This can be obtained when considering more complicated models with larger velocity sets. A major feature of real fluids is their spatial isotropy. In order for the lattice gas to be isotropic we have to look for situations in which the tensor $T_{\alpha\beta\gamma\delta}$ is isotropic under arbitrary rotations. This question has been addressed by Frisch, Hasslacher and Pomeau[6] who replaced the early square model of HPP by a model constructed on a triangular lattice: the so-called FHP model. For three dimensional situations there is no lattice allowing to get an isotropic model where all particles have the same speed, however it is possible to find a four-dimensional model, the so-called Face-Centered-Hypercubic model (FCHC) which meets the requirements of isotropy[7]. In this model the set of velocity includes 24 elements given by

$$
\begin{array}{ll}
(\pm1, \pm1, 0, 0) & (\pm1, 0, \pm1, 0) \\
(\pm1, 0, 0, \pm1) & (0, \pm1, \pm1, 0) \\
(0, \pm1, 0, \pm1) & (0, 0, \pm1, \pm1)
\end{array}
$$

This four-dimensional model can then be projected onto three-dimensional space taking the fourth component of the velocity as a passive scalar, thus giving rise to pseudo-four dimensional three-dimensional model.

When using models with particles of different speeds it is possible to take a lattice that was not satisfactory in the simpler case by choosing properly the ratio of particles of different speeds. For instance in the two-dimensional case one may take[13] the HPP square lattice and a velocity set including 9 components: 1 corresponding to particles at rest, 4 corresponding to particles of speed 1 (that move to the nearest-neighbour) and 4 corresponding to particles of speed $\sqrt{2}$ (that move to the next nearest-neighbour). If d_0, d_1 and d_2 are respectively the densities per velocity component of these three types of particles, then isotropy is obtained when

$$d_1(1-d_1)(1-2d_1) = 4d_2(1-d_2)(1-2d_2)$$

The rest particles are created in collision events of the type: two speed 1 particles colliding at right angle produce a rest particle and a particle of speed $\sqrt{2}$ in such a way that momemtum is conserved. They are destroyed in the reverse process. This shows the crucial importance of the rest particles to couple the two HPP model: one with speed 1 and one with speed $\sqrt{2}$ lying on a lattice rotated by $\pi/4$ from the first one. In that case the densities d_0, d_1 and d_2 are related by

$$d_0 d_2(1-d_1)^2 = d_1^2(1-d_0)(1-d_2)$$

For models that satisfy the required isotropy we replace the coefficient $G(\rho)T_{\alpha\beta\gamma\delta}$ in the advection term of the "Navier-Stokes" equation by $g(\rho)$. This coefficient depends upon the density. If we consider incompressible or slightly compressible flows, then $g(\rho)$ is a constant that may be eliminated by making a change of scale of the velocities by a factor $g(\rho)$. We then recover a satisfactory expression for the advection term, however the question of galilean invariance is not solved by this renormalization of the velocities. The term involving $\psi(\rho)$ on the right hand side of the "Navier-Stokes" equation is related to the viscosity of the fluid.

The actual determination of the viscosity is performed in the Boltzmann approximation either starting from a microscopic analysis of a Couette flow situation, as performed by Hénon[14], or by solving the linearized Boltzmann equation as performed by Rivet and Frisch[10].

These analysis are performed assuming that molecular chaos is satisfied and that higher order distribution functions can be factorized in terms of one velocity distribution functions. Under these assumptions closed form expressions for the kinematic viscosity can be derived. In the case of models with rest particles there is a bulk viscosity, the determination of which involves the calculation of an appropriate eigenvalue of the linearized collision operator. As we shall be mostly interested in using lattice gases to study low speed flows then the significant fluid property is the kinematic shear viscosity that enters the definition of the Reynolds number. Here we used an effective Reynolds number following the renormalization of the velocities in the form

$$R_{\text{eff}} = \frac{g(\rho)VL}{\nu}$$

where V and L are typical velocities and dimensions of the flow and ν is the kinematic shear viscosity.

As mentionned above the viscosity coefficient can be obtained from a linearized analysis of the Boltzmann equation, thus it is possible to choose its value by a proper choice of the collision matrix elements $A(s \rightarrow s')$.

For models with b velocities and local collisions, the set of possible precollision states s includes 2^b states. For small values of b it is possible to adjust the collision matrix A by inspection, however for models like the FCHC model with $b = 24$ it is necessary to find some algorithm to perform that task. This has been done by Hénon[15] in order to minimize the value of the shear viscosity of the FCHC model.

In conclusion of this review of the properties of lattice gas models, it is possible to state that provided a lattice of the right symmetry is chosen, lattice gases can be used as a new fluid to simulate hydrodynamical flows at low Mach numbers. At the macroscopic level, they follow equations of motion similar to those of real fluids

$$\partial_t \rho + \text{div}(\rho \mathbf{u}) = 0$$

$$\partial_t(\rho u_\alpha) + \partial_\beta(g(\rho)\rho u_\alpha u_\beta) = -\partial_\alpha P(\rho, u^2) + \partial_\beta(\nu \partial_\beta(\rho u_\alpha)) + \partial_\alpha(\zeta \text{div}(\rho \mathbf{u}))$$

with

$$g(\rho) = \frac{\rho}{2\rho_m}\frac{1-2d}{1-d}, \quad P(\rho, u^2) = \frac{\rho_m}{2} - \frac{\rho}{2}g(\rho)(4c_s^2 - 1)u^2$$

where ρ is the total density, ρ_m the density of moving particles, d the mean number of particles per link of the lattice, c_s the velocity of sound and ν and ζ respectively the shear and bulk kinematic viscosities.

III Results of lattice gas simulations

Lattice gas simulations are fairly easy to perform on a digital computer. As indicated above they involve essentially two steps: displacement and collisions. In addition collisions with boundaries can be provided for instance by allowing particles that hit a boundary to reverse their velocity ("bounce-back condition" that corresponds to the stick condition) or to be specularly reflected by the boundary (that corresponds to the free-slip condition). Finally initial conditions can be set by starting some macroscopic distribution for the flow and taking microscopic conditions at random according to the Fermi-Dirac equilibrium distribution corresponding to the local values of the density and velocity of the fluid. More or less efficient algorithms have been written. The main choice is how to store the data in the computer memory. A natural way is to assign one memory location for one lattice site, provided the computer words include more bits than b the number of velocities of the model. A more efficient technique consists in storing in one word of the computer memory the value for a given velocity component for a number of sites equal to the number B of bits in the computer word (usually $B=$ 32 or 64). The most time consuming operation is the collision step. In the first case the value of the particle distribution at each site is used as address for a collision table (the so-called look-up table technique). In the second case the postcollision situation is computed for B sites simultaneously through a series of logical operations (the so-called logical technique). Examples of the logical operations that can be used for various two-dimensional models are given in Ref. 16. For models with a large number of velocities (for instance $b = 24$ for the FCHC model) it has not been possible to find a logical expression so that a look-up table technique is used. This obviously requires a computer with a very large random-access memory (2^{24} memory locations are required for the collision table, plus obviously the storage of the state of the lattice). The three-dimensional simulations have been performed on a Cray-2 machine[17]. Work is in progress to try and adapt the logical technique to the FCHC model.

We now give a summary of the work on lattice gas simulations.

Linear situations like sound wave or shear wave relaxation lead to accurate values for the speed of sound and the viscosity coefficients. The measured values of c_s agree with the theoretical one to better than a few 10^3 (limited by experimental errors). There is a remarkably good agreement between the measured values of the kinematic shear viscosity (to about 1 %) with the theoretical values obtained in the framework of the Boltzmann approximation[16]. This means that higher order distribution functions are well represented by products of one velocity distribution functions and that long time tail effects, as conjectured in real fluids, are not very important. There are however size dependent effects on the value of the viscosity both for two dimensional models[18] and for one dimensional models [19].

Non linear situations, that is study of flows around obstacles or inhomogeneous flows, have given quantitative agreement with real experiments or results of numerical results obtained by applying standard techniques to the continuous macroscopic Navier-Stokes equations. This was obtained provided velocity renormalization is properly performed[16].

As an example we consider a two-dimensional situation with a flow between linear boundaries with a sudden expansion. This problem of the backward facing step has been studied both experimentally and by numerical methods. The lattice gas technique has been applied in the following conditions: the FHP model with rest particles has been used with the set of collisions chosen to minimize the shear viscosity (FHP III model). The lattice includes 4096 × 512 nodes and at the input side the width of the channel is 256 over a length of 512. The FHP lattice being hexagonal, we have to specify the orientation of the channel with respect to the lattice. Here the axis of the channel is parallel to one of the 6 possible velocities. The sides of the lattice are set with the "bounce-back" condition to ensure $v_{\parallel} = 0$. The Reynolds number is computed using the maximum speed in the input region (close to a parabolic profile) and the height of the step. Starting from a not very physical flow: Poiseuille profile in the region before the step with maximum velocity v_x and an other Poiseuille profile beyond the step with maximum velocity $v_x/2$, we first observe a transient during which a recirculation zone develops beyond the step, with formation of a region of low velocity near the other side of the channel located beyond the step. The Reynolds numbers used in these simulations were not large enough to obtain recirculation in that region. Lattice gases involve small numbers of particles at each node of the lattice and thus velocity fields that are determined by this technique are very noisy. This can be improved by taking spatial averages over a number of neighbouring sites, typically 8 × 8 or 16 × 16 and, when the situation is either steady or slowly varying in time, this can be further improved by taking time averages. This was done in order to determine the length L_r of the recirculation zone defined by the distance between the step and the point near the boundary where the component v_x changes sign. The results for the ratio L_r/step height are the following:

Reynolds	measured L_r
53	2.1
92	2.8
129	3.5

These values are plotted in Fig. 1 together with results of standard calculations[20].

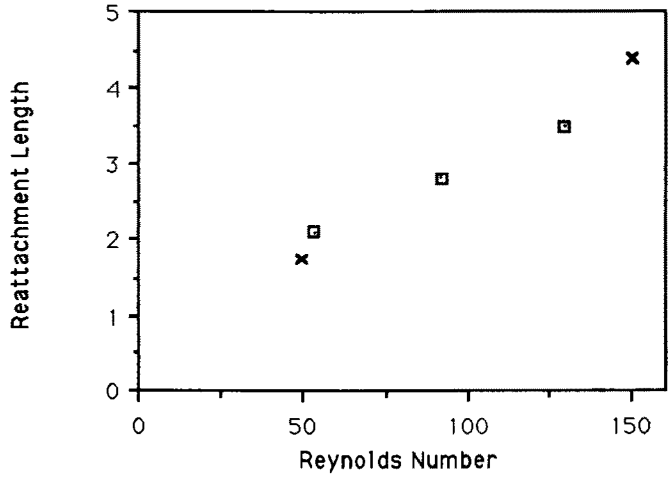


Fig. 1 Location of the reattachment point beyond a backward facing step. Crosses are obtained by standard resolution of the Navier-Stokes equation[20], boxes are computed by the lattice gas technique

IV Study of one dimensional shock waves

We are going now to describe the study of one-dimensional shocks using the FHP III model, as above. Let us give the particular values of the coefficients involved in the macrodynamical equations of motion ot that system. For the FHP model, the pressure P is given by

$$P = \frac{3\rho}{7}(1 - \frac{5}{6}g(\rho)u^2)$$

with

$$g(\rho) = \frac{7}{12}\frac{7-2\rho}{7-\rho}.$$

As we shall study one-dimensional shocks propagating either in direction x or y, we can simplify the equations of motion for our particular case, and we rewrite them in terms of the density ρ and momentum $j = \rho u$.

$$\partial_t \rho + \partial_x j = 0,$$
$$\partial_t j + \frac{9}{14}\partial_x(\frac{g}{\rho}j^2) = -c_s^2 \partial_x \rho.$$

Applying the transformation $x' = x - \xi t$, $t' = t$, and searching for steady state solutions with a jump of j from j_1 to j_2 and of ρ from ρ_1 to ρ_2 on either side of a front, we find expressions similar to the Rankine-Hugoniot relationships

$$\xi(\rho_2 - \rho_1) - (j_2 - j_1) = 0.$$
$$\xi(j_1 - j_2) - \frac{9}{14}(\frac{g(\rho_1)}{\rho_1}j_1^2 - \frac{g(\rho_2)}{\rho_2}j_2^2) - c_s^2(\rho_2 - \rho_1) = 0.$$

To obtain ξ we have to solve

$$\xi^2 - \frac{9}{14(j_2 - j_1)}(\frac{g_2}{\rho_2}j_2^2 - \frac{g_1}{\rho_1}j_1^2)\xi - c_s^2 = 0.$$

For the cases to be studied latter, j is chosen equal to 0 on one side of the front, so that we get

$$\xi = \frac{9}{28}\frac{g}{\rho}j \pm c_s\sqrt{1 + \frac{81}{392}\frac{g^2 j^2}{\rho^2 c_s^2}}.$$

Numerical study

A FHP lattice including 4096 × 512 nodes has been considered to obtain data on the propagation of shock waves. In a first study, the long side of the lattice is set perpendicular to one the velocities of the particles and we measure the behaviour of waves propagating parallel to the small side (Ox case). The lattice is given periodic boundary conditions along the long side. In the other direction solid boundaries are set with "bounce-back" conditions. Initial conditions are set so that the density is uniform and equal to ρ_0, and the velocity is uniform and equal to u_0 perpendicular to the solid boundaries. By reflection against these boundaries, two fronts appear. Velocities are 0 towards the wall, u_0 in the center of the lattice, whereas the density is ρ_0 in the center of the lattice. We then determine at various times the mean values of the density and momentum as a function of the space coordinate along Ox. Fronts are found. They are analyzed by a least square fit to the following shape

$$\rho(x) = \frac{1}{2}(\rho_0 + \rho_1) + \frac{1}{2}(\rho_0 - \rho_1)\tanh(\frac{x - x_0}{\delta})$$

where x_0 is the location the front and δ is related to its width.

The velocity of the compression shock is found to be smaller than that of the rarefaction wave. We first checked that the relationship between the change in ρ and that in j is verified. We then determined ξ both for the compression wave and for the depletion wave. The value of $\xi - c_s$ varies approximately linearly with j, so that we can determine the value of z in $\xi = c_s + zj$. Fig. 2 presents our measurements of the shock velocities as a function of the initial density, together with its theoretical value for a given value of the initial velocity of the gas $u_0 (= 3/14)$. Note that for $\rho > 0.5$, the lattice can be considered as being filled with "holes" at a density $1 - \rho$ and therefore we should obtain the same values for z. This is well verified. We then studied the case where the long side of the lattice is parallel to one of the particle velocities, (Oy case). By comparing the data obtained for the two geometries, propagation along Ox or Oy, we test the isotropy of the model. We have obtained the same values of z for the two cases, which allows us to say that the model is isotropic.

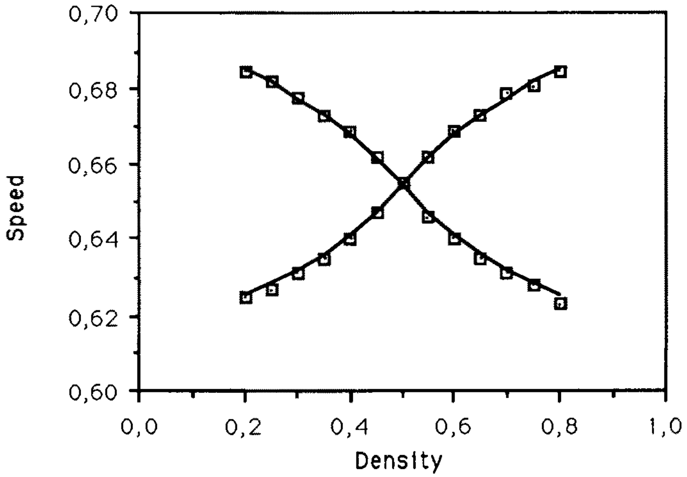


Fig. 2 Velocity of shock waves *vs* value of the particle density d per lattice link. Boxes are results of simulations, solid line is theoretical

The depletion wave broadens as a function of time, whereas the width of the compression wave reaches a steady state value when starting either from a very sharp front or from a wide front. The steady state width depends upon the density and the strength of the shock. It is of the order of a few mean free path. Its value does not agree with that derived from the Navier-Stokes equations. A microscopic analysis is required.

The very good agreement between the "experimental" data and the theoretical predictions for the propagation velocities of both the compression and depletion shocks show that higher order terms in the macroscopic equations of motion of the FHP lattice gas play no significant role. This shows that lattice gases may be useful in a situation not considered in previous studies that were limited to weakly compressible flows.

In conlusion of this review it is possible to state that lattice gas models can be used as a new "fluid" to simulate fluid flows in complex geometries. Their implementation on digital computers is much simpler than that of standard techniques used nowadays to compute the solutions of the continuous Navier-Stokes equations. The theoretical analysis show that lattice gases can be used only in incompressible situations. Experience acquired since the revival of the field by the paper of FHP shows that they are not limited to small values of the Mach number. However the Reynolds numbers of situations that can be studied are limited to moderate values unless extremely large lattices are considered. Efforts are being made to reduce the value of the viscosity and thus increase the Reynolds number. Finally we can just mention here that extensions of the simple FHP lattice gas model with either several speeds or with particle labelling allow thermal or reaction-diffusion situations to be simulated in an efficient way. This is especially true for problems with free fronts as lattice gas models exist in which phase separation phenomena occur naturally[9,21].

References

[1] Boon J.P. and Yip S., *Molecular Hydrodynamics*, Mac Graw Hill, (1980).

[2] Rapaport D.C. and Clementi F., *Phys. Rev. Letters*, **57**, 695 (1986). Meiburg E., *Phys. Fluids*, **29**, 3107 (1986). Koplik J., Banavar, J.R. and Willemsen J.F., *Phys. Rev. Letters*, **60**, 1282, (1988). Rapaport D.C., *Phys. Rev. Letters*, **60**, 2480 (1988).

[3] Broadwell J.C., *Phys. Fluids*, **7**, 1243 (1964).

[4] Gatignol R., "Théorie Cinétique des Gaz à Répartition Discrète de Vitesses", *Lecture Notes in Physics*, **36** (Springer, Berlin, 1975).

[5] Hardy J., Pomeau Y. and de Pazzis O. , *J. Math. Phys.*, **14**,1746 (1973). Hardy J., de Pazzis O. and Pomeau Y., *Phys. Rev.*, **A 13**,1949 (1976).

[6] Frisch U., Hasslacher B. and Pomeau Y., *Phys. Rev. Lett.*, **56**, 1505 (1986).

[7] d'Humières D., Lallemand P. and Frisch U., *Europhys. Lett.*, **2**, 291 (1986).

[8] Burges C. and Zaleski S., *Complex Systems*, **1**, 31, (1987).

[9] Clavin P., Lallemand P., Pomeau Y. and Searby G., *J. Fluid Mech.*, **188**, 437, (1988).

[10] Rivet J.P. and Frisch U., *Comp. Rend. Acad. Sci. Paris* **II 302**, 267, (1987).

[11] Wolfram S., *J. Stat. Phys.*, **45**, 471, (1986).

[12] Frisch U., d'Humières D., Hasslacher B., Lallemand P., Pomeau Y. and Rivet J.P.,*Complex Systems*, **1**,649, (1987).

[13] d'Humières D. and Lallemand P., *Helvetica Physica Acta*, **59**, 1231, (1986).

[14] Hénon M., *Complex Systems*, **1**, 763, (1987).

[15] Hénon M. to be published in *Proceedings of the Conference "Discrete kinetic theory, lattice gas dynamics and foundations of hydrodynamics"*, Torino, Sept. 1988.

[16] d'Humières D. and Lallemand P., *Complex Systems*, **1**, 1, 599, (1987).

[17] Rivet J.P., *Comp. Rend. Acad. Sci. Paris*, **II 305**, 751, (1987). Rivet J.P., Hénon M. , Frisch U. and d'Humières D., *Europhys. Lett.*, **7**, 231, (1988).

[18] Kadanoff L., McNamara G. and Zanetti G., "Size-dependence of the shear-viscosity for a two-dimensional lattice gas", preprint Univ. Chicago, (1987).

[19] d'Humières D., Lallemand P. and Qian Y.H., *Comp. Rend. Acad. Sci. Paris* to be published (1988).

[20] Glowinski R., Mantel B., Périaux J. and Tissier O. in *Note on Numerical Fluid Mechanics*, **9**, Morgan K., Périaux J. and Thomasset F. editors (Vieweg and Sons) (1984).

[21] Rothmann D.H. and Keller J.M., *J. Stat. Phys.*, **52**, 1119, (1988).

ON THE EQUATIONS OF MULTI-COMPONENT PERFECT OF REAL GAS INVISCID FLOW

B. LARROUTUROU and L. FEZOUI
INRIA, Sophia-Antipolis, 06560 VALBONNE, FRANCE

1. INTRODUCTION

In the last years, the development of efficient algorithms for the numerical solution of fluid dynamics problems and the increase of the available computing power has made possible to consider the numerical simulation of more and more complex fluid flows, whose investigation was out of reach in the previous decade. Among these newcomers in the set of problems addressed by the "computational fluid dynamicists" are the *flows of mixtures of several gaseous species*, and in particular the chemically reacting flows.

In fact, there exist many different kinds of multi-component flows, whose numerical simulations involve different difficulties. For example, most flame propagation phenomena are highly subsonic; then, except in some particular situations (for instance if one is interested in the flame-acoustics interaction), the hyperbolic effects play a minor role in these phenomena, which are dominated by the purely hydrodynamic (quasi-incompressible), the diffusive and the reactive effects (see e.g. [27] and the references therein). On the other hand, the aspects of wave propagation in the gaseous mixture, in other words the hyperbolic aspects, are of first importance in several other situations, including detonations (see e.g. [7], [12], [32]), transonic combustion (see e.g. [10]) or hypersonic reacting flows (see e.g. [9], [14], [33]).

The present paper is devoted to this last kind of multi-component flows; more precisely, the emphasis will be put on the *hyperbolic aspects of multi-species flows.*

To be more specific, we will neglect in the whole paper the effects of diffusion and chemical reactions in the gaseous mixture; we will also always assume that all species in the mixture are at thermal equilibrium (in other words, that one can use a single temperature, which may vary in space and time, but which is the same for all species), and that the total pressure in the mixture is the sum of the partial pressures of the individual components (Dalton's law). With these assumptions, we will consider the equations describing the *inviscid one-dimensional flow* of a gaseous mixture. These equations will take the form:

$$\begin{cases} \rho_t + (\rho u)_x = 0 \ , \\ (\rho u)_t + (\rho u^2 + p)_x = 0 \ , \\ E_t + [u(E+p)]_x = 0 \ , \\ (\rho Y_k)_t + (\rho u Y_k)_x = 0 \quad \text{for} \quad 1 \le k \le N-1 \ . \end{cases} \tag{1.1}$$

(see Section 2 below for the definition of the notations).

Essentially two different strategies can be applied for the numerical solution of this system. Either one uses for the first three equations in (1.1) one of the numerous available schemes aimed at solving the Euler equations (based on approximate Riemann solvers, or TVD, flux-splitting, flux-corrected transport methods or other approaches...), and one solves separately the species equations (last line in (1.1)), with an ad hoc upwind or viscosity term (for instance with a donorcell type approximation); or one considers the whole system (1.1) as a system of conservation laws and solves all equations in a coupled way by extending to (1.1) one of the above mentioned schemes designed for the solution of the single-component Euler equations. The first approach has been used for instance in [6], [14], [25]; the second approach has been employed in e.g. [1], [4], [9], [16]. In particular, some comparisons between both approaches are presented in [16], where the advantages of treating system (1.1) as a whole (that is, the advantages of the second approach) are clearly shown, from both points of view of accuracy and of preserving the positivity of those variables which need to remain positive (such as the densities of all species).

We review and discuss in this paper a set of problems related to this second approach. More precisely, considering (1.1) as a system of $N+2$ conservations laws (which will appear to be hyperbolic), we essentially address two questions: the *exact solution of the Riemann problem* for this system, and *the extension to this multi-component flow of the most classical numerical schemes used in the single-component case* (i.e. for the Euler equations). We will first examine in Section 2 the case where all species in the mixture behave as perfect gases, and then investigate in Section 3 the case of real gas mixtures.

The questions addressed in this paper have been the subject of a few recent papers [1], [4], [16], and also have relations with the problem of a single real gas investigated in [21], [30], [33]; for the sake of completeness, we will recall below some of the results of these works.

2. MULTI-COMPONENT PERFECT GAS FLOW

We consider in this section the one-dimensional inviscid flow of a mixture of N species Σ_1, Σ_2 ... Σ_N, each component being assumed to behave as a perfect gas.

2.1. Governing equations

The governing equations for this flow express the conservation of mass for each component, the conservation of momentum and of the total energy. They take the form (see e.g. [43]):

$$\begin{cases} (\rho Y_k)_t + (\rho u Y_k)_x = 0 \quad \text{for} \quad 1 \leq k \leq N \ , \\ (\rho u)_t + (\rho u^2 + p)_x = 0 \ , \\ \mathcal{E}_t + [u(\mathcal{E} + p)]_x = 0 \ , \end{cases} \tag{2.1}$$

where ρ is the mixture density, Y_k is the mass fraction of species Σ_k (that is, ρY_k is the separate density of species Σ_k, and $\sum_{k=1}^{N} Y_k = 1$), u is the mixture velocity (which is also the velocity of each species, since we neglect molecular diffusion), p is the total pressure in the mixture, and $\mathcal{E}$ is the total energy per unit volume.

We assume here that each species Σ_k obeys the perfect gas laws, and in particular has constant specific heats at constant volume and pressure C_{vk} and C_{pk}. We will also denote γ_k the ratio $\gamma_k = \dfrac{C_{pk}}{C_{vk}}$, and M_k the molecular weight of species Σ_k, which satisfies Mayer's relation:

$$M_k(C_{pk} - C_{vk}) = R \ , \tag{2.2}$$

the universal gas constant. The total pressure p is then given by Dalton's law:

$$p = \sum_{k=1}^{N} p_k \ , \tag{2.3}$$

the partial pressure p_k of species Σ_k being given by:

$$p_k = \rho Y_k \frac{R}{M_k} T \ ; \tag{2.4}$$

T is the temperature of the mixture (the same for all species). Considering that the N species may have different specific heats of formation h_k^0 , we write the total energy $\mathcal{E}$ as (see e.g. [8], [43]):

$$\mathcal{E} = \sum_{k=1}^{N} \left(\frac{1}{2} \rho Y_k u^2 + \rho Y_k C_{vk} T + \rho Y_k h_k^0 \right) \ . \tag{2.5}$$

Since the temperature and the partial pressures do not appear in the conservation relations (2.1), we can eliminate them and consider that, in equations (2.1), the pressure p is given by the following relation, which is deduced from (2.2)-(2.5):

$$p = (\gamma - 1) \left(\mathcal{E} - \frac{1}{2} \rho u^2 - \sum_{k=1}^{N} \rho Y_k h_k^0 \right) \ , \tag{2.6}$$

γ being the local ratio of the specific heats of the mixture:

$$\gamma = \frac{(C_p)_{mixture}}{(C_v)_{mixture}} = \frac{\sum\limits_k Y_k C_{pk}}{\sum\limits_k Y_k C_{vk}} = \frac{\sum\limits_k Y_k C_{vk} \gamma_k}{\sum\limits_k Y_k C_{vk}} \ . \tag{2.7}$$

The last equality in (2.7) shows that the local value of γ (which depends on the mixture composition) is a linear convex combination of the γ_k's.

System (2.1)-(2.6) (with γ given by (2.7)) can be rewritten in a different way. Defining E as the sum of the kinetic and thermal energies per unit volume:

$$E = \frac{1}{2} \rho u^2 + \sum_{k=1}^{N} \rho Y_k \ C_{vk} T = \mathcal{E} - \sum_{k=1}^{N} \rho Y_k h_k^0 \ , \tag{2.8}$$

we can rewrite (2.1)-(2.6) as:

$$\begin{cases} (\rho Y_k)_t + (\rho u Y_k)_x = 0 \quad \text{for} \quad 1 \le k \le N \ , \\ (\rho u)_t + (\rho u^2 + p)_x = 0 \ , \\ E_t + [u(E + p)]_x = 0 \ , \end{cases} \tag{2.9}$$

$$p = (\gamma - 1) \left(E - \frac{1}{2} \rho u^2 \right) \ , \tag{2.10}$$

(the energy equation in (2.9) is a linear combination of the energy and species conservation equations in (2.1)). Summing all species equations in (2.9), one can also get an equation for the total density ρ and rewrite (2.9) as:

$$\begin{cases} \rho_t + (\rho u)_x = 0 \ , \\ (\rho u)_t + (\rho u^2 + p)_x = 0 \ , \\ E_t + [u(E+p)]_x = 0 \ , \\ (\rho Y_k)_t + (\rho u Y_k)_x = 0 \quad \text{for} \ \ 1 \le k \le N-1 \ . \end{cases} \tag{2.11}$$

Remark 1: At this point, one may wonder if the three systems of conservation laws (2.1), (2.9) and (2.11) are really equivalent, from both points of view of the exact solutions of the corresponding initial value problem and of their numerical approximation using conservative schemes. The answer is of course positive, since only *linear* combinations of the *conservative* equations are used to transform one of these systems into another. Indeed, if we denote $W_t + F_x = 0$, with $W \in \mathbb{R}^{N+2}$, $F \in \mathbb{R}^{N+2}$ the vector form of (2.1), system (2.9) can be written in the form $\mathcal{W}_t + \mathcal{F}_x = 0$, with $\mathcal{W} = MW$ and $\mathcal{F} = MF$, M being an $N+2 \times N+2$ matrix *which is independent of* W. It is then straightforward to check that solving an initial value problem for $W_t + F_x = 0$, either exactly or using any of the conservative schemes considered below, is equivalent to solving it for the transformed system $\mathcal{W}_t + \mathcal{F}_x = 0$.

This shows that *using instead of (2.9) the system (2.11), where the N^{th} species does not play the same role as the other $N-1$ components, has no importance*, from both mathematical and numerical points of view. We will in fact use the form (2.11), where the three first equations are the familiar Euler equations. The preceding arguments also show that *treating the system (2.1)-(2.6) where the N species have different heats of formation exactly amounts to treating the system (2.9)-(2.10) where all heats of formation are equal to* 0. •

2.2. Homogeneity and hyperbolicity

From now on, we will restrict our attention to the case of a mixture made of only two species Σ_1 and Σ_2; *but all results presented below can be straightforwardly extended to mixtures consisting of any number of components N*. Simply denoting Y the mass fraction Y_1 of the first species, we consider the system:

$$\begin{pmatrix} \rho \\ \rho u \\ E \\ \rho Y \end{pmatrix}_t + \begin{pmatrix} \rho u \\ \rho u^2 + p \\ u(E+p) \\ \rho u Y \end{pmatrix}_x = 0 \ , \tag{2.12}$$

with:

$$p = (\gamma - 1)\left(E - \frac{1}{2}\rho u^2\right) \ , \tag{2.13}$$

and:

$$\gamma = \frac{Y C_{v1}\gamma_1 + (1-Y) C_{v2}\gamma_2}{Y C_{v1} + (1-Y) C_{v2}} \ . \tag{2.14}$$

We will use the classical notations W and F for the vectors of the conservative variables and of the fluxes:

$$W = \begin{pmatrix} \rho \\ \rho u \\ E \\ \rho Y \end{pmatrix} = \begin{pmatrix} \rho \\ m \\ E \\ \rho' \end{pmatrix} = \begin{pmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \end{pmatrix} \ , \quad F = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ u(E+p) \\ \rho Y \end{pmatrix} = \begin{pmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \end{pmatrix} \ . \tag{2.15}$$

Then, we have the two following simple results, which are also shown in e.g. [1], [4], [9], [16]:

Proposition 1:

The flux vector F is an homogeneous function of degree 1 of W. •

Proposition 2:

If the specific heat ratio γ_k of each species in the mixture satisfies the inequality:

$$\gamma_k > 1 \ , \tag{2.16}$$

then the system (2.12) is hyperbolic. •

PROOF of Proposition 1: Since we can write:

$$\gamma = \frac{\rho' C_{v1}\gamma_1 + (\rho - \rho')C_{v2}\gamma_2}{\rho' C_{v1} + (\rho - \rho')C_{v2}} = \gamma(\rho, \rho') = \gamma(W) \ , \tag{2.17}$$

we have:

$$F = \begin{pmatrix} m \\ \dfrac{m^2}{\rho} + (\gamma(W) - 1)\left(E - \dfrac{m^2}{2\rho}\right) \\ \dfrac{m}{\rho}\left(\gamma(W)E - (\gamma(W) - 1)\dfrac{m^2}{2\rho}\right) \\ \dfrac{m\rho'}{\rho} \end{pmatrix} , \tag{2.18}$$

which shows that the fluxes only depend on the conservative variables: $F = F(W)$. Moreover, one can notice on (2.17) that the function $\gamma(W)$ is homogeneous of degree 0:

$$\forall r > 0 \ , \ \ \gamma(rW) = \gamma(W) \ , \tag{2.19}$$

which implies that F is homogeneous of degree 1, as in the single component case:

$$\forall r > 0 \ , \ \ F(rW) = rF(W) \ . \bullet \tag{2.20}$$

PROOF of Proposition 2: Let us first say that (2.7) and (2.16) imply $\gamma > 1$. Now, the Jacobian matrix $A(W) = \dfrac{DF}{DW}$ has the following expression:

$$A(W) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \dfrac{(\gamma - 3)}{2}u^2 + X & (3 - \gamma)u & \gamma - 1 & X' \\ \dfrac{(\gamma - 1)}{2}u^3 - uH + uX & H - (\gamma - 1)u^2 & \gamma u & uX' \\ -uY & Y & 0 & u \end{pmatrix} ; \tag{2.21}$$

in (2.21), we have set $H = \dfrac{E + p}{\rho}$ (H is the specific enthalpy of the mixture), $X = \dfrac{p}{\gamma - 1}\gamma_\rho$, $X' = \dfrac{p}{\gamma - 1}\gamma_{\rho'}$, where γ_ρ and $\gamma_{\rho'}$ are the partial derivatives of γ given by (2.17). A straightforward calculation then shows that the eigenvalues of $A(W)$ are the roots of the polynomial:

$$\det[A(W) - \lambda Id] = (u-\lambda)^2 \left[(u-\lambda)^2 - \frac{\gamma p}{\rho} - (X + YX')\right] . \tag{2.22}$$

But $X + YX' = \frac{p}{\gamma - 1}(\gamma_\rho + Y\gamma_{\rho'}) = \frac{p}{\rho(\gamma-1)}(\rho\gamma_\rho + \rho'\gamma_{\rho'})$, which is identically 0 since $\gamma(\rho, \rho')$ is homogeneous of degree 0 (we use here Euler's property of homogeneous functions). Therefore, the eigenvalues of $A(W)$ are:

$$\lambda_1 = u - c\ , \quad \lambda_2 = u\ , \quad \lambda_3 = u\ , \quad \lambda_4 = u + c\ , \tag{2.23}$$

where the sound speed c has the usual expression:

$$c = \sqrt{\frac{\gamma p}{\rho}}\ , \tag{2.24}$$

but with $\gamma = \gamma(\rho, \rho')$. We should say at this point that this expression of the sound speed in the two-component mixture, which we have derived by evaluating the eigenvalues of the flux Jacobian matrix A, is equivalent to the usual expression $c = \frac{\partial p}{\partial \rho}\ |_S$ (derivative at constant entropy).

A set of right eigenvectors is easily found; one can take:

$$r_1 = \begin{pmatrix} 1 \\ u - c \\ H - uc \\ Y \end{pmatrix}\ , \quad r_2 = \begin{pmatrix} 1 \\ u \\ \dfrac{u^2}{2} - \dfrac{X}{\gamma - 1} \\ 0 \end{pmatrix}\ , \quad r_3 = \begin{pmatrix} 0 \\ 0 \\ -\dfrac{X'}{\gamma - 1} \\ 1 \end{pmatrix}\ , \quad r_4 = \begin{pmatrix} 1 \\ u + c \\ H + uc \\ Y \end{pmatrix}\ , \tag{2.25}$$

(of course, any combination of r_2 and r_3 is also a right eigenvector associated to the eigenvalue u). We have therefore found four independent eigenvectors, which shows that (2.12) is an hyperbolic system of conservation laws (although *non strictly hyperbolic* since $\lambda_2 = \lambda_3$; in this respect, a one-dimensional two-component flow has some similarity with a single-component two-dimensional flow: in both cases, the velocity becomes a double eigenvalue). •

Remark 2: The following relations are easy to check and will be useful in the sequel:

$$H = \frac{u^2}{2} + \frac{c^2}{\gamma - 1}\ , \tag{2.26}$$

$$X' = \frac{p}{\gamma - 1}\ \frac{C_{v1}C_{v2}(\gamma_1 - \gamma_2)}{\rho[YC_{v1} + (1-Y)C_{v2}]^2} = \frac{C_{v1}C_{v2}(\gamma_1 - \gamma_2)T}{YC_{v1} + (1-Y)C_{v2}}\ . \bullet \tag{2.27}$$

2.3. The Riemann problem

Let us now examine the solution of a Riemann problem for system (2.12). Using again the notations (2.14)-(2.15) and introducing two states W_L and W_R, we consider the problem:

$$\begin{cases} W_t + F(W)_x = 0 \quad \text{for } x \in I\!R\ ,\ \ t \geq 0\ , \\ W(x, 0) = \begin{cases} W_L & \text{if } x < 0\ , \\ W_R & \text{if } x > 0\ . \end{cases} \end{cases} \tag{2.28}$$

When trying to solve this problem, a first important question concerns the genuine nonlinearity or the degeneracy of the characteristic fields (see [28], [36]). As in the single-component case, the answer is here that *the first and last characteristic fields are genuinely non linear*, since:

$$\nabla_W \lambda_1 . r_1 = \sum_{l=1}^{4} \frac{\partial \lambda_1}{\partial W_l} . (r_1)_l = -\frac{(\gamma + 1)c}{2\rho} < 0\ , \tag{2.29}$$

$$\nabla_W \lambda_4 . r_4 = \sum_{l=1}^{4} \frac{\partial \lambda_4}{\partial W_l} . (r_4)_l = \frac{(\gamma+1)c}{2\rho} > 0 \; , \qquad (2.30)$$

whereas the characteristic fields associated with the eigenvalue u are linearly degenerate, since:

$$\nabla_W u . (\alpha_2 r_2 + \alpha_3 r_3) \equiv 0 \; , \qquad (2.31)$$

for any pair of real numbers (α_2, α_3).

Therefore, each characteristic field is either genuinely non linear or linearly degenerate. In the case of a *strictly* hyperbolic problem, this information is sufficient to state an existence result and describe the structure of the exact solution of the Riemann problem provided that the two states W_L and W_R are close enough to another. In the present case, further informations are needed since (2.28) is not strictly hyperbolic and since the states W_L and W_R are not necessarily close to each other. In fact, *it is possible to construct an entropic solution to problem (2.28)*, exactly as in the case of the single-component Euler equations, by analysing the shock or rarefaction waves associated with each of the non linear characteristic fields. We refer the reader to the Appendix and the references mentioned therein for the technical details, and simply describe here the structure of this solution.

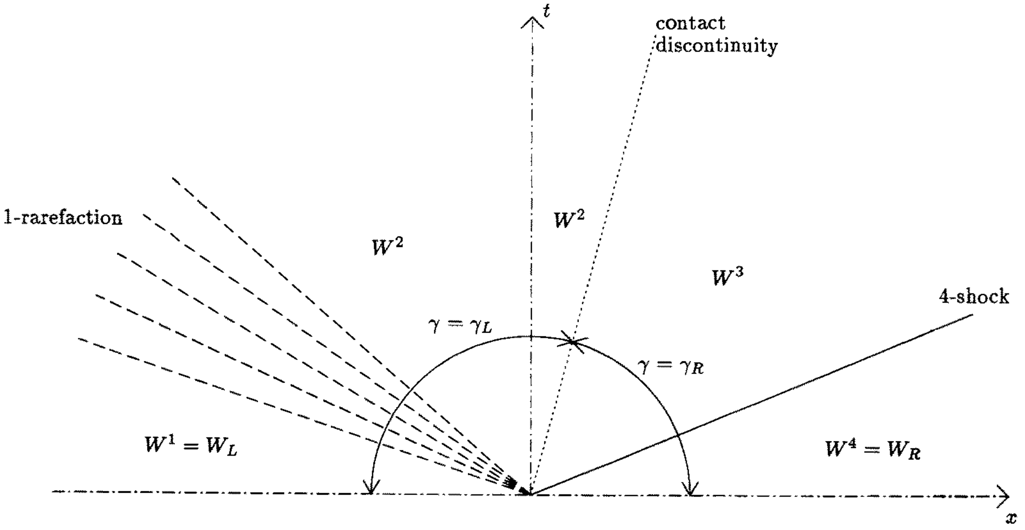


Figure 1: The solution of the multi-component Riemann problem

This solution is of course self-similar (i.e. $W(x,t)$ only depends on the ratio $\frac{x}{t}$), and consist, as in the single-component case, of four constant states W^1, W^2, W^3, W^4 separated by shocks, rarefaction waves or a contact discontinuity. More precisely, as shown on Figure 1, $W^1 = W_L$ and W^2 are separated by a 1-wave (i.e. a wave associated with the first characteristic field, either a 1-shock or a 1-rarefaction wave); W^2 and W^3 are separated by a 2-discontinuity or contact discontinuity; and W^3 and $W^4 = W_R$ are separated by a 4-wave. Also, the pressure p and the velocity u are continuous across the contact discontinuity. Last but not least, the mass fraction Y remains constant across the 1-wave and the 4-wave (whatever these waves are, shocks or rarefactions). This fact has important consequences. Indeed, γ *is constant on each side of the contact* *discontinuity*. On the left side of the discontinuity, the mixture has the composition of the state W_L, and *behaves as a single perfect gas* whose specific heat ratio is $\gamma_L = \gamma(W_L)$. Analogous conclusions hold for W^3 and W_R on the right side of the contact discontinuity. It is then straightforward to extend a single-component perfect gas Riemann solver into a multi-component perfect gas Riemann solver (see Remark A4 in the Appendix).

Remark 3: These results are not surprising if one keeps in mind that the first and fourth equations in (2.12) yield the following non conservative mass fraction equation:

$$Y_t + uY_x = 0 \ , \tag{2.32}$$

which shows that Y is constant along each "particle path", and if one realizes that the contact discontinuity is the trajectory of the "particle" which is initially located at $x = 0$. Then the left (resp. right) side of the contact discontinuity is filled with "particles" coming from the domain $x < 0$ (resp. $x > 0$), and therefore the mixture has there the composition of the state W_L (resp. W_R).

In fact, this simple and convincing empirical argument is not rigorous since the non conservative equation (2.32) does not a priori hold across a shock wave. •

2.4. Numerical schemes

We now turn to the numerical solution of an initial value problem associated with system (2.12):

$$\begin{cases} W_t + F(W)_x = 0 & \text{for } x \in I\!R \ , \ t \geq 0 \ , \\ W(x,0) = W^0(x) & \text{for } x \in I\!R \ . \end{cases} \tag{2.33}$$

We will restrict our attention to explicit, three-point, first-order accurate schemes written in conservative form. In other words, using very classical notations, we consider numerical schemes of the form:

$$\frac{W_j^{n+1} - W_j^n}{\Delta t} + \frac{F_{j+1/2}^n - F_{j-1/2}^n}{\Delta x} = 0 \ , \tag{2.34}$$

where the numerical flux $F_{j+1/2}^n$ is evaluated using a "numerical flux function" Φ:

$$F_{j+1/2}^n = \Phi(W_j^n, W_{j+1}^n) \ . \tag{2.35}$$

There exists many schemes of this type for the single-component Euler equations (see for instance [26] and the references mentioned below). We are going to show how four of these schemes, namely the Steger and Warming, Van Leer, Roe and Osher schemes can be extended to the solution of the two-component problem (2.33). We will very briefly recall the definition of each of these schemes for the Euler equations, using the notations:

$$W_E = \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix} \ , \quad F_E = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ u(E+p) \end{pmatrix} \ . \tag{2.36}$$

We refer to e.g. [15], [41] for comparisons of the qualities and defects of these four schemes, but recall that all of them are very widely used for gas dynamics calculations. Even the splitting of Steger and Warming, which may be considered as the most diffusive and therefore the least accurate of these schemes for a large class of problems is still very useful because of its very simplicity (for instance for the treatment of boundary conditions or for the design of implicit schemes ; see e.g. [13], [19], [25]).

2.4.1. The Steger and Warming scheme

The scheme proposed by Steger and Warming [38] for the Euler equations is based on the homogeneity property of the flux vector ($F_E(rW_E) = rF_E(W_E)$ for any $r > 0$), which implies:

$$F_E(W_E) = A_E(W_E)W_E \ , \tag{2.37}$$

where $A_E = \dfrac{DF_E}{DW_E}$.

We need here to introduce a (classical) notation: if B is a diagonalisable matrix, and if f maps $I\!R$ into itself, we define the matrix $f(B)$ as follows: we write the diagonalisation of B as:

$$B = T\Lambda T^{-1} \ , \ \Lambda = Diag[\mu_1, \mu_2 \cdots \mu_n] \ , \tag{2.38}$$

and set:

$$B = Tf(\Lambda)T^{-1} \ , \ \text{where } f(\Lambda) = Diag[f(\mu_1), f(\mu_2) \cdots f(\mu_n)] \ , \tag{2.39}$$

(in practice, we will use this definition for $f(x) = \mid x \mid$, $f(x) = x^+ = \max(x, 0)$ or $f(x) = x^- = \min(x, 0)$).

Using now (2.38)-(2.39), we can define A_E^+ and A_E^-. Since $A_E^+ + A_E^- = A_E$, it follows from (2.37) that the Steger and Warming numerical flux function:

$$\Phi_E(W_E^1, W_E^2) = A_E^+(W_E^1)W_E^1 + A_E^-(W_E^2)W_E^2 \tag{2.40}$$

is consistent with the Euler equations (i.e. satisfies $\Phi_E(W_E, W_E) = F_E(W_E)$).

* * *

Extending this scheme to the multi-component case is straightforward since the basic homogeneity property still holds for mixtures. We just set:

$$\Phi(W_L, W_R) = A^+(W_L)W_L + A^-(W_R)W_R \ . \tag{2.41}$$

As in the single-component case, the extended Steger and Warming scheme is a "flux-vector-splitting" scheme (see [26]) since, setting $F^+(W) = A^+(W)W$, $F^-(W) = A^-(W)W$, we have:

$$F(W) = F^+(W) + F^-(W) \ , \tag{2.42}$$

and:

$$\Phi(W_L, W_R) = F^+(W_L) + F^-(W_R) \ . \tag{2.43}$$

We also have here the property that, if all characteristic wave speeds associated with the state W are positive (resp. negative), i.e. if $u \geq c$ (resp. $u \leq -c$), then $F^+(W) = F(W)$ (resp. $F^-(W) = F(W)$). At this point arises the question of the validity of this flux decomposition, or of the stability of the resulting scheme: the scheme uses an upward (resp. downward) differencing for F^+ (resp. F^-), clearly because all wave speeds associated with F^+ (resp. F^-) are expected to be positive (resp. negative). But is it actually the case ? The answer is given by:

Proposition 3:

If the specific heat ratio γ_k of each species in the mixture satisfies the inequality:

$$1 < \gamma_k < \frac{5}{3} , \tag{2.44}$$

then all eigenvalues of the Jacobian matrix $\frac{DF^+}{DW}$ (resp. $\frac{DF^-}{DW}$) are real and positive (resp. negative). •

The proof is roughly analogous to the one of Lerat [29] who proved the same result in the single-component case. Since it is rather lengthy and technical (of course, the matrix $\frac{DF^+}{DW}$ is not equal to A^+ !), we will omit it, referring to [18] for the details.

2.4.2. The Van Leer scheme

The above Steger and Warming scheme has the drawback that the derivatives of the split fluxes F^+ and F^- are discontinuous when one of the eigenvalues of A changes sign (i.e. at sonic and stagnation points). To remedy this, Van Leer [40] introduced a continuously differentiable splitting. In the single-component case, this splitting is defined by the following expressions:

* if $u \geq \mathrm{c}$, $F_E^+(W_E) = F_E(W_E)$, $F_E^-(W_E) = 0$;

* if $-\mathrm{c} \leq u \leq \mathrm{c}$,

$$F_E^+(W_E) = \begin{pmatrix} F_1^+ \\ F_2^+ \\ F_3^+ \end{pmatrix} = \begin{pmatrix} \frac{\rho}{4\mathrm{c}}(u+\mathrm{c})^2 \\ F_1^+\left(u - \frac{u-2\mathrm{c}}{\gamma}\right) \\ \frac{\gamma^2}{2(\gamma^2-1)} \frac{(F_2^+)^2}{F_1^+} \end{pmatrix} , \tag{2.45}$$

$$F_E^-(W_E) = \begin{pmatrix} F_1^- \\ F_2^- \\ F_3^- \end{pmatrix} = \begin{pmatrix} -\frac{\rho}{4\mathrm{c}}(u-\mathrm{c})^2 \\ F_1^-\left(u - \frac{u+2\mathrm{c}}{\gamma}\right) \\ \frac{\gamma^2}{2(\gamma^2-1)} \frac{(F_2^-)^2}{F_1^-} \end{pmatrix} ; \tag{2.46}$$

* if $u \leq -c$, $F_E^+(W_E) = 0$, $F_E^-(W_E) = F_E(W_E)$.

We emphasize here that γ is a constant in (2.45)-(2.46) since these expressions concern the single-component case.

* * *

There is a natural way of extending this flux decomposition to the two-component case. We simply set $F^+(W) = F(W)$ when $u \geq c$, $F^+(W) = 0$ when $u \leq -\mathrm{c}$, and:

$$F^+(W) = \begin{pmatrix} F_1^+ \\ F_2^+ \\ F_3^+ \\ F_4^+ \end{pmatrix} = \begin{pmatrix} \dfrac{\rho}{4c}(u+c)^2 \\[2mm] F_1^+\left(u - \dfrac{u-2c}{\gamma}\right) \\[2mm] \dfrac{\gamma^2}{2(\gamma^2-1)}\,\dfrac{(F_2^+)^2}{F_1^+} \\[2mm] YF_1^+ \end{pmatrix} , \tag{2.47}$$

when $-c \leq u \leq c$. In (2.47), γ *is the local (non constant) specific heat ratio* (2.17). We define F^- by analogous formulae (or by the identity (2.42)). This flux splitting can be used to define a stable conservative scheme (using (2.43)) since we have the following result:

Proposition 4:

If the specific heat ratio γ_k of each species in the mixture satisfies the inequality:

$$1 < \gamma_k < 3 , \tag{2.48}$$

then all eigenvalues of the Jacobian matrix $\dfrac{DF^+}{DW}$ (resp. $\dfrac{DF^-}{DW}$) are real and positive (resp. negative). •

PROOF: We only need to examine the case where $-c \leq u \leq c$; we present the proof for $\dfrac{DF^+}{DW}$ (the result for $\dfrac{DF^-}{DW}$ follows by symmetry). To evaluate the eigenvalues of the Jacobian matrix $\dfrac{DF^+}{DW}$, we need to consider the determinant $\mathcal{D} = \det\left(\dfrac{DF^+}{DW} - \lambda Id\right)$, for $\lambda \in I\!R$ (but we do not want to evaluate all terms in this matrix !). The main idea is then to consider each component of F^+ as a function of W_1, W_2, W_3 and Y (which itself depends on W_1 and W_4), instead of seeing these components as functions of the four conservative variables W_l $(1 \leq l \leq 4)$. It can indeed be noticed that all components (2.47) of F^+ actually depend on Y through (2.24) and (2.14). Thus, we write, for $1 \leq l \leq 4$:

$$F_l^+ = F_l^+[W_1, W_2, W_3; Y(W_1, W_4)] . \tag{2.49}$$

Keeping in mind that $F_4^+ = YF_1^+$, we get:

$$\mathcal{D} = \begin{vmatrix} \dfrac{\partial F_1^+}{\partial W_1} + \dfrac{\partial F_1^+}{\partial Y}\dfrac{\partial Y}{\partial W_1} - \lambda & \dfrac{\partial F_1^+}{\partial W_2} & \dfrac{\partial F_1^+}{\partial W_3} & \dfrac{\partial F_1^+}{\partial Y}\dfrac{\partial Y}{\partial W_4} \\[3mm] \dfrac{\partial F_2^+}{\partial W_1} + \dfrac{\partial F_2^+}{\partial Y}\dfrac{\partial Y}{\partial W_1} & \dfrac{\partial F_2^+}{\partial W_2} - \lambda & \dfrac{\partial F_2^+}{\partial W_3} & \dfrac{\partial F_2^+}{\partial Y}\dfrac{\partial Y}{\partial W_4} \\[3mm] \dfrac{\partial F_3^+}{\partial W_1} + \dfrac{\partial F_3^+}{\partial Y}\dfrac{\partial Y}{\partial W_1} & \dfrac{\partial F_3^+}{\partial W_2} & \dfrac{\partial F_3^+}{\partial W_3} - \lambda & \dfrac{\partial F_3^+}{\partial Y}\dfrac{\partial Y}{\partial W_4} \\[3mm] F_1^+\dfrac{\partial Y}{\partial W_1} + Y\dfrac{\partial F_1^+}{\partial W_1} + Y\dfrac{\partial F_1^+}{\partial Y}\dfrac{\partial Y}{\partial W_1} & Y\dfrac{\partial F_1^+}{\partial W_2} & Y\dfrac{\partial F_1^+}{\partial W_3} & F_1^+\dfrac{\partial Y}{\partial W_4} + Y\dfrac{\partial F_1^+}{\partial Y}\dfrac{\partial Y}{\partial W_4} - \lambda \end{vmatrix} . \tag{2.50}$$

Replacing now the first column of this determinant by the sum of itself and of the fourth column multiplied by Y, and using the relation:

$$\frac{\partial Y}{\partial W_1} + Y\frac{\partial Y}{\partial W_4} = \frac{1}{W_1}\left(W_1\frac{\partial Y}{\partial W_1} + W_4\frac{\partial Y}{\partial W_4}\right) = 0\ , \tag{2.51}$$

which follows from the fact that $Y = Y(W_1, W_4) = \dfrac{W_1}{W_4}$ is a homogeneous function of degree 0, we obtain:

$$\mathcal{D} = \begin{vmatrix} \dfrac{\partial F_1^+}{\partial W_1} - \lambda & \dfrac{\partial F_1^+}{\partial W_2} & \dfrac{\partial F_1^+}{\partial W_3} & \dfrac{\partial F_1^+}{\partial Y}\dfrac{\partial Y}{\partial W_4} \\ \dfrac{\partial F_2^+}{\partial W_1} & \dfrac{\partial F_2^+}{\partial W_2} - \lambda & \dfrac{\partial F_2^+}{\partial W_3} & \dfrac{\partial F_2^+}{\partial Y}\dfrac{\partial Y}{\partial W_4} \\ \dfrac{\partial F_3^+}{\partial W_1} & \dfrac{\partial F_3^+}{\partial W_2} & \dfrac{\partial F_3^+}{\partial W_3} - \lambda & \dfrac{\partial F_3^+}{\partial Y}\dfrac{\partial Y}{\partial W_4} \\ Y\dfrac{\partial F_1^+}{\partial W_1} - \lambda Y & Y\dfrac{\partial F_1^+}{\partial W_2} & Y\dfrac{\partial F_1^+}{\partial W_3} & F_1^+\dfrac{\partial Y}{\partial W_4} + Y\dfrac{\partial F_1^+}{\partial Y}\dfrac{\partial Y}{\partial W_4} - \lambda \end{vmatrix} . \tag{2.52}$$

Substracting now of the fourth row Y times the first row, we further get:

$$\mathcal{D} = \begin{vmatrix} \dfrac{\partial F_1^+}{\partial W_1} - \lambda & \dfrac{\partial F_1^+}{\partial W_2} & \dfrac{\partial F_1^+}{\partial W_3} & \dfrac{\partial F_1^+}{\partial Y}\dfrac{\partial Y}{\partial W_4} \\ \dfrac{\partial F_2^+}{\partial W_1} & \dfrac{\partial F_2^+}{\partial W_2} - \lambda & \dfrac{\partial F_2^+}{\partial W_3} & \dfrac{\partial F_2^+}{\partial Y}\dfrac{\partial Y}{\partial W_4} \\ \dfrac{\partial F_3^+}{\partial W_1} & \dfrac{\partial F_3^+}{\partial W_2} & \dfrac{\partial F_3^+}{\partial W_3} - \lambda & \dfrac{\partial F_3^+}{\partial Y}\dfrac{\partial Y}{\partial W_4} \\ 0 & 0 & 0 & F_1^+\dfrac{\partial Y}{\partial W_4} - \lambda \end{vmatrix} . \tag{2.53}$$

Since $\dfrac{\partial Y}{\partial W_4} = \dfrac{1}{\rho}$, we finally get:

$$\mathcal{D} = \left(\frac{F_1^+}{\rho} - \lambda\right) . \begin{vmatrix} \dfrac{\partial F_1^+}{\partial W_1} - \lambda & \dfrac{\partial F_1^+}{\partial W_2} & \dfrac{\partial F_1^+}{\partial W_3} \\ \dfrac{\partial F_2^+}{\partial W_1} & \dfrac{\partial F_2^+}{\partial W_2} - \lambda & \dfrac{\partial F_2^+}{\partial W_3} \\ \dfrac{\partial F_3^+}{\partial W_1} & \dfrac{\partial F_3^+}{\partial W_2} & \dfrac{\partial F_3^+}{\partial W_3} - \lambda \end{vmatrix} . \tag{2.54}$$

The determinant in the right-hand side of (2.54) is clearly the determinant of the 3 × 3 matrix $\dfrac{DF_E^+}{DW_E} - \lambda Id$, which corresponds to the single component case; in other words, the Y-dependence has been removed in this 3 × 3 determinant which can be evaluated with $\gamma = \gamma(Y)$ considered as fixed. This single-component case has been analysed by Van Leer [40]: under the hypothesis $1 < \gamma < 3$, the matrix $\dfrac{DF_E^+}{DW_E}$ has one eigenvalue equal to zero and two positive eigenvalues. Thus, (2.54) says that considering a two-component mixture just introduces an additional eigenvalue $\lambda = \dfrac{F_1^+}{\rho} = \dfrac{(u+c)^2}{4c}$ which is positive. This completes the proof of Proposition 4. •

Remark 4: We have seen in Section 2.2 that adding a second component introduces (with respect to the single-component case) an additional eigenvalue equal to the velocity u for the flux Jacobian matrix $A = \dfrac{DF}{DW}$. In a completely similar way, considering a second component introduces for the split flux Jacobian matrix $\dfrac{DF^+}{DW}$ the additional eigenvalue $\dfrac{(u+c)^2}{4c}$, which, according to the splitting:

$$u = \frac{(u+c)^2}{4c} - \frac{(u-c)^2}{4c} = u^+ + u^- \tag{2.55}$$

used to define $F_1^+ = \rho u^+$ and $F_1^- = \rho u^-$, is nothing but the "positive part u^+ of the velocity". •

2.4.3. The Roe scheme

Roe [35] has proposed a conservative upwind scheme which uses an approximate Riemann solver based on a linearization of the fluxes. We refer the reader to [35] and also to [26], where a very clear presentation of this scheme inside the framework of the Godunov-type schemes is given. We simply recall here the basic results concerning this scheme in the single-component case, before considering its extension to the two-component case.

The numerical flux function of this scheme has the form:

$$\Phi_E(W_E^1, W_E^2) = \frac{F_E(W_E^1) + F_E(W_E^2)}{2} + \frac{1}{2} \mid \tilde{A}_E \mid (W_E^1 - W_E^2) \,, \tag{2.56}$$

where $\tilde{A}_E = \tilde{A}_E(W_E^1, W_E^2)$ is a diagonalisable matrix which satisfies the property:

$$F_E(W_E^1) - F_E(W_E^2) = \tilde{A}_E(W_E^1 - W_E^2) \;; \tag{2.57}$$

This property has many interesting consequences (mainly because of its similarity with the Rankine-Hugoniot relations; see [26], [35]). In particular, it yields simpler expressions of the numerical flux function:

$$\begin{aligned} \Phi_E(W_E^1, W_E^2) &= F_E(W_E^1) - \tilde{A}_E^-(W_E^1 - W_E^2) \\ &= F_E(W_E^2) + \tilde{A}_E^+(W_E^1 - W_E^2) \,. \end{aligned} \tag{2.58}$$

There are several different ways of choosing a matrix $\tilde{A}_E$ satisfying (2.57). Roe [35] proposed to define $\tilde{A}_E$ as follows: $\tilde{A}_E$ is equal to the flux Jacobian matrix A_E evaluated for some state $\tilde{W}_E = \tilde{W}_E(W_E^1; W_E^2)$ known as "Roe's average of W_E^1 and W_E^2". More precisely:

$$\tilde{A}_E = A_E[\tilde{W}_E(W_E^1, W_E^2)] \,, \tag{2.59}$$

where $\tilde{W}_E = (\tilde{\rho}, \tilde{\rho}\tilde{u}, \tilde{E})^T$ is defined by the relations:

$$\left(\tilde{\rho} = \frac{\rho^1\sqrt{\rho^1} + \rho^2\sqrt{\rho^2}}{\sqrt{\rho^1} + \sqrt{\rho^2}} \,, \right) \quad \tilde{u} = \frac{u^1\sqrt{\rho^1} + u^2\sqrt{\rho^2}}{\sqrt{\rho^1} + \sqrt{\rho^2}} \,, \tag{2.60}$$

$$\tilde{H} = \frac{H^1\sqrt{\rho^1} + H^2\sqrt{\rho^2}}{\sqrt{\rho^1} + \sqrt{\rho^2}} \,, \tag{2.61}$$

(in fact, defining $\tilde{\rho}$ is not useful here since only $\tilde{u}$ and $\tilde{H}$ are needed to evaluate the Jacobian matrix $A_E(\tilde{W}_E)$).

* * *

The extension of Roe's scheme to the two-component system (2.33) has been derived independently by Abgrall [1] and Fernandez and Larrouturou [16]. The two-component scheme relies on the relations:

$$\begin{aligned}\Phi(W_L,W_R) &= \frac{F(W_L)+F(W_R)}{2}+\frac{1}{2}\mid\tilde{A}\mid(W_L-W_R)\\ &= F(W_L)-\tilde{A}^{-}(W_L-W_R)\\ &= F_E(W_R)+\tilde{A}^{+}(W_L-W_R)\ ,\end{aligned} \tag{2.62}$$

where $\tilde{A}=\tilde{A}(W_L,W_R)$ satisfies the property:

$$F(W_L)-F(W_R)=\tilde{A}(W_L-W_R)\ . \tag{2.63}$$

Again we can define the averaged state $\tilde{W}=(\tilde{\rho},\tilde{\rho}\tilde{u},\tilde{E},\tilde{\rho}\tilde{Y})^T$ by:

$$\left(\tilde{\rho}=\frac{\rho_L\sqrt{\rho_L}+\rho_R\sqrt{\rho_R}}{\sqrt{\rho_L}+\sqrt{\rho_R}}\ ,\right)\quad \tilde{u}=\frac{u_L\sqrt{\rho_L}+u_R\sqrt{\rho_R}}{\sqrt{\rho_L}+\sqrt{\rho_R}}\ , \tag{2.64}$$

$$\tilde{H}=\frac{H_L\sqrt{\rho_L}+H_R\sqrt{\rho_R}}{\sqrt{\rho_L}+\sqrt{\rho_R}}\ , \tag{2.65}$$

$$\tilde{Y}=\frac{Y_L\sqrt{\rho_L}+Y_R\sqrt{\rho_R}}{\sqrt{\rho_L}+\sqrt{\rho_R}}\ , \tag{2.66}$$

(again determining $\tilde{\rho}$ is not necessary). But in this two-component context (unless both species in the mixture have the same specific heat ratio $\gamma_1=\gamma_2$, that is unless $\gamma=\gamma(W)$ is a constant), the flux Jacobian matrix $A(\tilde{W})$ does not satisfy property (2.63). Therefore, the matrix $\tilde{A}$ is to be chosen different from $A(\tilde{W})$ (but close to the latter since we want our extension to reduce to the usual Roe scheme when both species are the same). The result given in [1], [16] is the following:

$$\tilde{A}=\begin{pmatrix} 0 & 1 & 0 & 0\\ \frac{(\tilde{\gamma}-3)}{2}\tilde{u}^2-\tilde{Y}\tilde{X}' & (3-\tilde{\gamma})\tilde{u} & \tilde{\gamma}-1 & \tilde{X}'\\ -\tilde{u}\tilde{H}+\frac{(\tilde{\gamma}-1)}{2}\tilde{u}^3-\tilde{u}\tilde{Y}\tilde{X}' & \tilde{H}-(\tilde{\gamma}-1)\tilde{u}^2 & \tilde{\gamma}\tilde{u} & \tilde{u}\tilde{X}'\\ -\tilde{u}\tilde{Y} & \tilde{Y} & 0 & \tilde{u}\end{pmatrix}\ , \tag{2.67}$$

where $\tilde{\gamma}=\gamma(\tilde{W})$, but where $\tilde{X}'$ is not equal to $X'(\tilde{W})$ given by (2.27). In order to insure that property (2.63) holds, one has to choose:

$$\tilde{X}'=\frac{C_{v1}C_{v2}(\gamma_1-\gamma_2)\tilde{T}}{\tilde{Y}C_{v1}+(1-\tilde{Y})C_{v2}}\ , \tag{2.68}$$

with:

$$\tilde{T}=\frac{T_L\sqrt{\rho_L}+T_R\sqrt{\rho_R}}{\sqrt{\rho_L}+\sqrt{\rho_R}}\neq T(\tilde{W})\ . \tag{2.69}$$

The matrix $\tilde{A}$ defined by (2.67)-(2.69) is then diagonalisable: its eigenvalues are $\tilde{u}-\tilde{c}$, $\tilde{u}$, $\tilde{u}$, $\tilde{u}+\tilde{c}$, where $\tilde{c}^2=(\tilde{\gamma}-1)\left(\tilde{H}-\frac{\tilde{u}^2}{2}\right)$ from (2.26) (or $\tilde{c}^2=\frac{\tilde{\gamma}\tilde{p}}{\tilde{\rho}}$ if one actually defines $\tilde{\rho}$), and its eigenvectors are given by expressions which are analogous to (2.25).

We leave it to the reader to check that $\tilde{A}$ defined by (2.67)-(2.69) actually satisfies (2.63) (the details of the proof can also be found in [1]). This verification is easily done using the following arithmetic rules, which hold for any variable U:

$$\Delta UV = \hat{U}\Delta V + \tilde{V}\Delta U \ , \quad (\rho\hat{U}) = \hat{\rho}\tilde{U} \ , \tag{2.70}$$

where:

$$\Delta U = U_L - U_R \ , \quad \tilde{U} = \frac{U_L\sqrt{\rho_L} + U_R\sqrt{\rho_R}}{\sqrt{\rho_L} + \sqrt{\rho_R}} \ , \quad \hat{U} = \frac{U_L\sqrt{\rho_R} + U_R\sqrt{\rho_L}}{\sqrt{\rho_L} + \sqrt{\rho_R}} \ . \tag{2.71}$$

Remark 5: We have said in Remark 1 that both systems (2.9) (with one equation for each species) and (2.11) (with one equation for the mixture density ρ and one equation for all but one species) are equivalent. Of course we found it helpful for all preceding calculations to use system (2.11) whose first three equations have the form of the usual Euler equations. But this formulation (2.11) (or (2.12) in the case of two species) also has a small drawback: the expression (2.68) for the extended Roe scheme is less easy to extend to N-component mixtures that the expression given in [1], which is equivalent to (2.68), but which is derived by using a system written under the form (2.9) where all species play the same role. •

2.4.4. The Osher scheme

The scheme proposed by Osher and Solomon [34], which is referred to as Osher's scheme, is based on the following expression of the numerical flux function:

$$\Phi_E(W_E^1, W_E^2) = \frac{F_E(W_E^1) + F_E(W_E^2)}{2} + \int_{W_E^1}^{W_E^2} \mid A_E(W_E) \mid dW_E \ , \tag{2.72}$$

where the integration is carried out on a path connecting W_E^1 and W_E^2 in the state-space. The integration path proposed in [34] is piecewise parallel to the right eigenvectors of the flux Jacobian matrix A_E, and the evaluation of the integral in (2.72) relies on the knowledge of the Riemann invariants associated with each eigenvectors (see [34] for the details).

* * *

The extension of Osher's scheme to multi-component flows is straightforward (exactly in the same way as the extension of exact Riemann solvers is straightforward, as we have seen in Section 2.3), and has been done by Abgrall and Montagné [4]. The extended scheme is of course defined by the analogue of (2.72):

$$\Phi(W_L, W_R) = \frac{F(W_L) + F(W_R)}{2} + \int_{W_L}^{W_R} \mid A(W) \mid dW \ , \tag{2.73}$$

and the evaluation of the integral again uses the Riemann invariants. Let us recall here that $\phi^{(m)} = \phi^{(m)}(W)$ is an m-Riemann invariant if:

$$\nabla_W \phi^{(m)}(W).r_m(W) = \sum_{l=1}^{4} \frac{\partial \phi^{(m)}}{\partial W_l}.(r_m)_l = 0 \ ; \tag{2.74}$$

as a consequence, a m-Riemann invariant is constant along a curve $W(s)$ in the state-space which is parallel to the m^{th} right eigenvector of the Jacobian matrix A, i.e. along which $\frac{dW}{ds}$ is always colinear with $r_m(W(s))$.

Using the expression (2.25) of the right eigenvectors, it is easy to check that the following quantities are Riemann invariants:

m	1	2	3	4
λ_m	$u-c$	u	u	$u+c$
$\phi^{(m)}$	$u+\frac{2c}{\gamma(W)-1}$ $\frac{p}{\rho^{\gamma(W)}}$ Y	u p ρY	u p ρ	$u-\frac{2c}{\gamma(W)-1}$ $\frac{p}{\rho^{\gamma(W)}}$ Y

Table 1: Riemann invariants in the two-component case.

Several comments are needed here. In each column of Table 1, the two first Riemann invariants are the Riemann invariants of the single-component case (but γ is no longer a constant). In particular, the invariant $\frac{p}{\rho^{\gamma(W)}}$ can still be interpreted as a function of the entropy (of the two-component mixture). Beside this, the fact that the velocity u is an m-Riemann invariant for $m = 2$ and $m = 3$ was already known from (2.31). Lastly, we should keep in mind here that, the system being not strictly hyperbolic, there is some arbitrariness in our choice of the eigenvectors associated with the double eigenvalue u. Thus, if $\alpha_2(W)$ and $\alpha_3(W)$ are real numbers, then:

$$r_u(W) = \alpha_2(W)r_2(W) + \alpha_3(W)r_3(W) \tag{2.75}$$

is also an eigenvector of the Jacobian matrix, and one may consider the equation:

$$\nabla_W\phi(W).r_u(W) = \sum_{l=1}^{4} \frac{\partial\phi}{\partial W_l}.(r_u)_l = 0\ . \tag{2.76}$$

It is clear that $\phi = u$ and $\phi = p$ satisfy (2.76) for any choice of α_2 and α_3 in (2.75). But such is not the case for the other invariants ρY and ρ appearing on the last line of Table 1. Therefore, u and p are the only "intrinsic Riemann invariants associated with the eigenvalue u" (there are only two such intrinsic Riemann invariants because u is a double eigenvalue).

Once the Riemann invariants are known, the problem of evaluating the integral in (2.73) is essentially analogue to the similar problem in the single-component case, since Y (and therefore γ) is constant along those pieces of the integration path which are parallel to either r_1 or r_4. We refer the reader to [4] for the details.

2.5. Remarks

We gather in this section several remarks concerning the numerical approximation of the two-component problem (2.33).

Remark 6: We have restricted in Section 2.4 our attention to explicit first-order accurate schemes for the one-dimensional problem (2.33). But the results of that section can be used for many further extensions: the generalization to N-component flows is straightforward, and the extensions to implicit schemes, second-order accurate schemes or even schemes operating on unstructured meshes for the simulation of multi-dimensional multi-component flows can be done exactly as in the case of the Euler equations (see e.g. [10], [13], [17], [19]). •

Remark 7: For the sake of simplicity, we have just presented the extension of the considered schemes to the two-component case, and we have not discussed how the different properties of these schemes (such as: exact resolution of discontinuities, need for entropy corrections ...) are or are not modified in the multi-component case. Concerning this matter, the general answer is that most of the conclusions which hold for a single perfect gas will also hold for a mixture of perfect gases. But there are exceptions to this "rule"; one of these exceptions is presented below.

When using the extended Roe scheme presented in Section 2.4.3 for a two-component shock tube problem, Abgrall [1] noticed slight pressure oscillations at the contact discontinuity. We now show that this difficulty may well appear with *any* upwind schemes and is intrinsically related to the presence of *several* species.

Following Abgrall [1], we consider a Riemann problem (2.28) where the two states W_L and W_R are supersonic and can be separated by a contact discontinuity (more precisely, we assume that $u_L = u_R = \bar{u}$, $u_L > c_L$, $u_R > c_R$, $p_L = p_R = \bar{p}$), and we further assume that we use an upwind scheme which satisfies:

$$\Phi(W_L, W_R) = F(W_L) \ , \tag{2.77}$$

a natural condition since the states W_L and W_R are supersonic. After spatial discretization, we have at time $t = 0$: $W_j^0 = W_L$ for $j \leq i-1$, $W_j^0 = W_R$ for $j \geq i$. Using the consistency relation $\Phi(W, W) = F(W)$, it is easy to see that the values W_j^1 (updated values after one time step) are equal to W_j^0 for all $j \neq i$, and that W_i^1 is given by:

$$W_i^1 = W_R - \frac{\Delta t}{\Delta x}[F(W_R) - F(W_L)] \ . \tag{2.78}$$

Calling $\nu = \dfrac{\bar{u}\Delta t}{\Delta x}$ the Courant number ($0 \leq \nu \leq 1$ from the CFL stability condition), we easily deduce from (2.78) that:

$$\rho_i^1 = (1-\nu)\rho_R + \nu\rho_L \ , \tag{2.79}$$

$$Y_i^1 = \frac{(1-\nu)\rho_R Y_R + \nu\rho_L Y_L}{(1-\nu)\rho_R + \nu\rho_L} \ , \tag{2.80}$$

and:

$$u_i^1 = \bar{u} \ , \tag{2.81}$$

$$\frac{p_i^1}{\gamma(Y_i^1) - 1} = \bar{p}\left(\frac{1-\nu}{\gamma(Y_R) - 1} + \frac{\nu}{\gamma(Y_L) - 1}\right) \ . \tag{2.82}$$

Thus, after one time step, the velocity has the correct value at each node, but the pressure is modified at node i (since in general $p_i^1 \neq \bar{p}$). This inability of the scheme to reproduce from one time level to another the constant pressure $\bar{p}$ may well cause after several time steps the oscillations observed in [1], and is directly related to the fact that γ is not constant, i.e. that the fluid is a mixture of several species.

To conclude this remark, we want to point out that the preceding observation holds as soon as the assumption (2.77) is satisfied; in particular it holds even if an exact Riemann solver is used at each mesh interface $x_{j+1/2}$, as in the Godunov method [23]. Only Glimm's scheme [22], which does not use the average process involved in Godunov-type schemes, would provide the correct pressure $p_i^1 = \bar{p}$. •

Remark 8: Incidentally, the preceding remark rises the following question: with the different conservative schemes considered in Section 2.4, is the relation (2.77) automatically true as soon as W_L and W_R are supersonic and satisfy $u_L \geq c_L$, $u_R \geq c_R$? The answer to this question is obviously positive.for the Steger and Warming scheme and Van Leer's scheme; it is also positive for Osher's scheme. But the answer is negative for Roe's scheme, even in the single-component case ! Even if we further assume, as in Remark 7, that the supersonic states W_L and W_R are separated by a contact discontinuity (i.e. satisfy $u_L = u_R, p_L = p_R$), then the relation (2.77) is not automatically true in the two-component case (we leave it to the reader to find counter examples; using the second equality in (2.62), it suffices to examine if $\tilde{A}^- = 0$). This particularity of Roe's scheme, which does not seem to have a major importance in practice, does not contradict the fact that this scheme is an upstream scheme in the sense of the definition given in [26]. Furthermore, what is observed here for Roe's scheme is also true for other schemes, like the Q-scheme [39] or the scheme of Vijasundaram [42]. •

3. MULTI-COMPONENT REAL GAS FLOW

We now consider the one-dimensional inviscid flow of an N-component real gas mixture. As we have said in the introduction, we will still neglect diffusion and chemical reactions and still use a thermal equilibrium assumption and Dalton's law.

3.1. Governing equations

We give below three examples of equations which describe the class of real gas mixtures which we are going to consider:

Example 1:

We consider here a gaseous mixture in which Dalton's law (2.3) still holds, in which the partial pressure of each component is still given by the perfect gas equation (2.4), but in which the expression of the total energy $\mathcal{E}$ is more complex than (2.5) and has the form:

$$\mathcal{E} = \frac{1}{2}\rho u^2 + \sum_{k=1}^{N} \rho Y_k e_k(T) - p \; ; \tag{3.1}$$

such a case may happen for instance if the specific heats C_{pk} of the species are assumed to depend on the temperature, leading to:

$$\mathcal{E} = \frac{1}{2}\rho u^2 + \sum_{k=1}^{N} \rho Y_k \left(h_k^0 + \int_{T_0}^{T} C_{pk}(T')dT' \right) - p \; ; \tag{3.2}$$

a relation like (3.1) also arises when vibrational energies of multi-atome species are taken into account, as in [14]:

$$\mathcal{E} = \frac{1}{2}\rho u^2 + \sum_{k=1}^{N} \rho Y_k \left(C_{pk} T + h_k^0 + \frac{R}{M_k} \frac{\theta_k}{\exp\left(\frac{\theta_k}{T}\right) - 1} \right) - p \ . \tag{3.3}$$

To summarize, in this first example, the governing equations consist of the system of conservation laws (2.1), with the additional relations (3.1) and:

$$p = \sum_{k=1}^{N} \rho Y_k \frac{R}{M_k} T \ .\bullet \tag{3.4}$$

Example 2:

In this second more general example, we consider a case where the energy is again given by (3.1), where Dalton's law still holds, but where the partial pressure p_k of the component Σ_k is given by:

$$p_k = \rho Y_k f_k(T) \ . \tag{3.5}$$

In other words, we assume that each separate component obeys Boyle's law (also known as Mariotte's law: the ratio $\dfrac{p_k}{\rho Y_k}$ only depends on the temperature; see e.g. [5], [24]).

Therefore, we consider in this second example that the system of conservation laws (2.1) is completed with two relations of the form:

$$p = \sum_{k=1}^{N} \rho Y_k f_k(T) \ , \tag{3.6}$$

$$\mathcal{E} = \frac{1}{2}\rho u^2 + \sum_{k=1}^{N} \rho Y_k g_k(T) \ , \tag{3.7}$$

(we have used (3.6) to rewrite (3.1) under the form (3.7)). •

Example 3:

In the framework of Example 2, one may often assume that equation (3.7) can always be solved for the temperature, in other words that (3.7) uniquely determines T as a function of $\mathcal{E}$, ρ, u and the mass fractions Y_k's (this is the case if all functions g_k are monotone increasing):

$$T = T\left(\mathcal{E} - \frac{1}{2}\rho u^2, \rho Y_k\right) \ . \tag{3.8}$$

Thus, (3.6) now gives the pressure as a function of the same arguments:

$$p = p\left(\mathcal{E} - \frac{1}{2}\rho u^2, \rho Y_k\right) \ . \tag{3.9}$$

In this third example, we simply consider that the system of conservation laws (2.1) is completed by a pressure equation of the form (3.9). •

Remark 9: Of course, we are not claiming that all equations of state used in practice to describe real gases belong to the framework of Examples 2 or 3. In particular, using the Van der Waals or the Virial equations for the partial pressure p_k would not lead to an expression like (3.6) (see e.g. [5], [24]). But equations (3.6)-(3.7) are nevertheless very general, and may be considered as sufficient to adequately describe mixtures at low or moderate pressures, in a wide range of temperature (since Boyle's law essentially fails at high pressures and very high temperatures; see [5], [24]). •

3.2. Homogeneity and hyperbolicity

For the sake of simplicity, we will restrict our attention as in Section 2.2 to a two-component mixture; but again, all results presented below can be straightforwardly extended to N-component mixtures. In the framework of Example 3, we consider again the equations (we now write E instead of $\mathcal{E}$ in order to use the same notations as in Section 2.2):

$$\begin{pmatrix} \rho \\ \rho u \\ E \\ \rho Y \end{pmatrix}_t + \begin{pmatrix} \rho u \\ \rho u^2 + p \\ u(E+p) \\ \rho u Y \end{pmatrix}_x = 0 , \tag{3.10}$$

with now:

$$p = p\left(E - \frac{1}{2}\rho u^2, \rho, \rho'\right) . \tag{3.11}$$

We will denote $\epsilon = E - \frac{1}{2}\rho u^2$ the internal energy per unit volume of the mixture.

Again it is easy to see that the flux vector F only depends on the conservative variables: $F = F(W)$. But an important remark is the following:

Proposition 5:
Under the hypotheses of Example 2, the flux vector F is an homogeneous function of degree 1 of W. •

PROOF: Let $r > 0$. Clearly, if we replace $\mathcal{E}$ by $r\mathcal{E}$, ρ by $r\rho$ and do not modify u nor the Y_k's in (3.7), we do not modify the value of the temperature. Hence $T(W)$ is an homogeneous function of degree 0, and (3.6) then implies that $p(W)$ is an homogeneous function of degree 1, which ends the proof. •

Because of this result, we will restrict from now on our attention to cases where the function p appearing in (3.11) is homogeneous of degree 1:

$$\forall r > 0 , \quad p(r\epsilon, r\rho, r\rho') = p(\epsilon, \rho, \rho') . \tag{3.12}$$

Remark 10: This homogeneity property should not be seen as a restrictive hypothesis; this assumption is by no means essential since most of the results presented below still hold without (3.12) (see [18]; but (3.12) will appear to bring in the following developments several nice simplifications which may result in a non negligible saving in computer time). Anyway, it appears in Proposition 5 that *the flux homogeneity is an intrinsic property of gas flow under the very general conditions described in Remark 9.* •

Remark 11: Several authors have recently studied the numerical simulation of an inviscid flow of a *single* real gas, with no homogeneity assumption (see [14], [21], [30], [33]). In several of these works, the single real gas is in fact a *mixture* for which one assumes *chemical equilibrium*. The mixture is then described by equations like (3.6)-(3.7), but one considers that the mass fractions, instead of being independent variables, can be evaluated as functions of ρ and T using the laws of chemical equilibrium (see [14]). It appears clearly now that this is the chemical equilibrium assumption which makes the flux vector non homogeneous in these studies; this contributes to make the simulation of equilibrium flows (which moreover involves the solution of the non linear chemical equilibrium equations) not really simpler than the simulation of non equilibrium **flows.** •

We **can** now turn to the question of the hyperbolicity of problem (3.10)-(3.11). We first evaluate the Jacobian matrix $A(W) = \dfrac{DF}{DW}$:

$$A(W) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ (p_\epsilon - 2)\dfrac{u^2}{2} + p_\rho & (2 - p_\epsilon)u & p_\epsilon & p_{\rho'} \\ p_\epsilon \dfrac{u^3}{2} - uH + up_\rho & H - p_\epsilon u^2 & u(1 + p_\epsilon) & up_{\rho'} \\ -uY & Y & 0 & u \end{pmatrix} ; \qquad (3.13)$$

the enthalpy H is still defined here by $H = \dfrac{E+p}{\rho}$, and $p_\epsilon, p_\rho, p_{\rho'}$ are the derivatives of the function p in (3.11). A straightforward calculation then shows that the determinant $\det[A(W) - \lambda Id]$ can be put in the form:

$$\det[A(W) - \lambda Id] = (u - \lambda)^2 \left((u - \lambda)^2 - \left[p_\rho + Y p_{\rho'} + p_\epsilon \frac{\epsilon + p}{\rho} \right] \right) . \qquad (3.14)$$

Thus, the system (3.10)-(3.11) can be hyperbolic only if the quantity between brackets is positive (this quantity will be the squared sound speed c^2). But the homogeneity property (3.12) implies (using again Euler's property of homogeneous functions):

$$p = \epsilon p_\epsilon + \rho p_\rho + \rho' p_{\rho'} . \qquad (3.15)$$

Therefore, the homogeneity property substantially simplifies the expression of the sound speed:

$$\begin{aligned} c^2 &= p_\rho + Y p_{\rho'} + p_\epsilon \frac{\epsilon}{\rho} + p_\epsilon \frac{p}{\rho} \\ &= (p_\epsilon + 1)\frac{p}{\rho} . \end{aligned} \qquad (3.16)$$

We are then led to introduce the notation:

$$\gamma(W) = p_\epsilon(W) + 1 ; \qquad (3.17)$$

with this notation, relations (3.13)-(3.16) can be summarized as follows. The Jacobian matrix:

$$A(W) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{(\gamma-3)}{2}u^2 + p_\rho & (3-\gamma)u & \gamma - 1 & p_{\rho'} \\ \frac{(\gamma-1)}{2}u^3 - uH + up_\rho & H - (\gamma-1)u^2 & \gamma u & up_{\rho'} \\ -uY & Y & 0 & u \end{pmatrix} \tag{3.18}$$

has almost the same form as in (2.21). The eigenvalues of A are:

$$\lambda_1 = u - c \ , \quad \lambda_2 = u \ , \quad \lambda_3 = u \ , \quad \lambda_4 = u + c \ , \tag{3.19}$$

where the sound speed c still has its usual expression:

$$c = \sqrt{\frac{\gamma p}{\rho}} \ ! \tag{3.20}$$

Lastly, we can exhibit a complete set of eigenvectors:

$$r_1 = \begin{pmatrix} 1 \\ u-c \\ H-uc \\ Y \end{pmatrix} , \quad r_2 = \begin{pmatrix} 1 \\ u \\ \frac{u^2}{2} - \frac{p_\rho}{\gamma-1} \\ 0 \end{pmatrix} , \quad r_3 = \begin{pmatrix} 0 \\ 0 \\ -\frac{p_{\rho'}}{\gamma-1} \\ 1 \end{pmatrix} , \quad r_4 = \begin{pmatrix} 1 \\ u+c \\ H+uc \\ Y \end{pmatrix} , \tag{3.21}$$

which are almost the same as in (2.25). We have therefore proved the analogue of Proposition 2:

Proposition 6: If the pressure equation (3.11) satisfies (3.12) and is such that, for any state W, $p_\epsilon(W) > 0$ (or equivalently $\gamma(W) > 1$), then system (3.10) is hyperbolic. •

Remark 12: In the studies [14], [21], [30], [33] of a single real gas with no homogeneity assumption, it is not clear to decide which (if any) of the quantities $\frac{\rho c^2}{p}$, $\frac{p}{\epsilon} + 1$ or $p_\epsilon + 1$ should be called γ; most of the referenced authors introduce in fact several γ's corresponding to several of these quantities (which all coincide in the case of a single- or multi-component perfect gas). In the present framework where the homogeneity assumption holds, there is a natural and obvious choice for γ: γ is defined by (3.17), the usual expression of the sound speed still holds, and the perfect gas relation $p = (\gamma - 1)\epsilon$ is replaced by:

$$p = (\gamma - 1)\epsilon + \rho p_\rho + \rho' p_{\rho'} \ , \tag{3.22}$$

which follows from (3.15) and (3.17). Of course, our definition (3.17) of γ, which involves a pressure derivative, may not be convenient in cases where the pressure is not given by an analytical expression but evaluated using tabulated data (see for instance [33] where a Mollier diagram is used). •

3.3. The Riemann problem

Although the results of the preceding section are very close to those obtained in Section 2.2

for a perfect gas mixture, we will now meet much more difficulties than in Section 2.3 to describe the solution of a Riemann problem for system (3.10)-(3.11).

In particular, the first and fourth characteristic fields are no longer necessarily genuinely non linear, since the quantities:

$$\nabla_W \lambda_1 . r_1 = -\frac{c}{2\rho}\left[(\gamma+1)+\frac{p}{\gamma}p_{\epsilon\epsilon}\right] , \tag{3.23}$$

$$\nabla_W \lambda_4 . r_4 = \frac{c}{2\rho}\left[(\gamma+1)+\frac{p}{\gamma}p_{\epsilon\epsilon}\right] \tag{3.24}$$

may vanish in the absence of additional assumptions. The evaluation of the quantities (3.23)-(3.24) is straightforward and uses the relation $\epsilon p_{\epsilon\epsilon} + \rho p_{\rho\epsilon} + \rho' p_{\rho'\epsilon} = 0$ which follows from the homogeneity of p_ϵ (much more complex expressions would be obtained in the absence of the homogeneity assumption). Beside this, we still have the degeneracy relation:

$$\nabla_W u . (\alpha_2 r_2 + \alpha_3 r_3) \equiv 0 , \tag{3.25}$$

for any pair of real numbers (α_2, α_3).

It is shown in the Appendix that, as in the case of a perfect gas mixture, the mass fraction Y is locally constant everywhere in the (x,t)-plane except possibly on a contact discontinuity associated with the degenerate eigenvalue (it is also shown that u and p are continuous across such a discontinuity). But fully describing the solution of the Riemann problem in the present context remains a much more complex task than in the case of a perfect gas mixture (see Remark A3 in the Appendix).

3.4. Numerical schemes

We now examine how the Steger and Warming, Van Leer, Roe and Osher schemes can be extended to the simulation of multi-component real gas flows governed by system (3.10)-(3.11).

3.4.1. The Steger and Warming scheme

Since we still have $F(W) = A(W)W$ from Proposition 5, there is no difficulty in extending the Steger and Warming scheme. We again set:

$$F^+(W) = A^+(W)W \ , \quad F^-(W) = A^-(W)W \ . \tag{3.26}$$

This generalization is much simpler than the extensions proposed in [30] and [33] for a single real non homogeneous gas, which involve more or less arbitrary choices. Apart from this, it seems that there is no chance to prove without additional assumptions that all wave speeds associated with F^+ (resp. F^-) are positive (resp. negative) (see the discussion of this point in [33], for a different extension of the Steger and Warming scheme).

3.4.2. The Van Leer scheme

A generalizarion of the Van Leer scheme for a single real gas has been proposed in [30] and [33], leading to a one-parameter family of flux-splitting schemes. Referring to [30], [33] for the details, we mention the result: when $-c \le u \le c$, F^+ is defined by the relations (we adopt the formulation of [30] for the energy split flux; the result of [33] is equivalent, although expressed under a different form):

$$F_1^+ = \frac{\rho}{4c}(u+c)^2 \ , \tag{3.27}$$

$$F_2^+ = F_1^+ \left(u - \frac{u-2c}{\gamma} \right) \ , \tag{3.28}$$

as for a perfect gas, and:

$$F_3^+ = F_1^+ [H - x(u-c)^2] \ , \tag{3.29}$$

where x is a real number to be chosen.

To extend these results to the present two-component case, we simply add to relations (3.27)-(3.29) the following relation:

$$F_4^+ = Y F_1^+ \ . \tag{3.30}$$

We can still apply to this real gas splitting scheme the arguments of the proof of Proposition 4: the matrix $\frac{DF^+}{DW}$ has the eigenvalue $u^+ = \frac{(u+c)^2}{4c}$, and its three other eigenvalues are the eigenvalues of the 3×3 Jacobian matrix $\frac{D(F_1^+, F_2^+, F_3^+)}{D(W_1, W_2, W_3)}$ which can be evaluated with Y fixed. Unfortunately, it has been observed in [30], [33] that the latter eigenvalues do not always have the expected sign: therefore the wave speeds associated with F^+ (resp. F^-) are not necessarily positive (resp. negative).

3.4.3. The Roe scheme

We now consider the extension of Roe scheme to a two-component real gas (we refer to [21], [30], [33] for generalizations of this scheme to a single real gas). Two states W_L and W_R being given, the major point is to find a diagonalisable matrix $\tilde{A}$ satisfying Roe's property (2.63). We try to find $\tilde{A}$ under the form:

$$\tilde{A} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ (\tilde{p}_1 - 2)\frac{\tilde{u}^2}{2} + \tilde{p}_2 & (2 - \tilde{p}_1)\tilde{u} & \tilde{p}_1 & \tilde{p}_3 \\ \tilde{p}_1 \frac{\tilde{u}^3}{2} - \tilde{u}\tilde{H} + \tilde{u}\tilde{p}_2 & \tilde{H} - \tilde{p}_1 \tilde{u}^2 & \tilde{u}(1 + \tilde{p}_1) & \tilde{u}\tilde{p}_3 \\ -\tilde{u}\tilde{Y} & \tilde{Y} & 0 & \tilde{u} \end{pmatrix} , \tag{3.31}$$

which is close to the form (3.13) of $A(\tilde{W})$, $\tilde{W}$ being again Roe's average defined by (2.64)-(2.66); but $\tilde{A}$ is not equal to $A(\tilde{W})$ since we will not choose $[\tilde{p}_1, \tilde{p}_2, \tilde{p}_3]$ equal to $[p_\epsilon(\tilde{W}), p_\rho(\tilde{W}), p_{\rho'}(\tilde{W})]$. One can easily check that the matrix $\tilde{A}$ given by (3.31) satisfies the property (2.63) as soon as the following relation holds:

$$\Delta p = \tilde{p}_1 \Delta\epsilon + \tilde{p}_2 \Delta\rho + \tilde{p}_3 \Delta\rho' \ , \tag{3.32}$$

(where the operator Δ is defined as in (2.71)). At this point, several choices are possible. A simple strategy inspired from [30] consists in choosing $[\tilde{p}_1, \tilde{p}_2, \tilde{p}_3]$ which minimizes the quantity:

$$[\tilde{p}_1 - p_\epsilon(\tilde{W})]^2 + [\tilde{p}_2 - p_\rho(\tilde{W})]^2 + [\tilde{p}_3 - p_{\rho'}(\tilde{W})]^2 \tag{3.33}$$

among all possible candidates satisfying (3.32). Other choices are proposed in [21] and [33] for a single real gas. But it does not seem clear that any of these choices actually *guarantees* that the matrix $\tilde{A}$ is diagonalisable (i.e. that the quantity $\tilde{c}^2 = \tilde{p}_2 + \tilde{Y}\tilde{p}_3 + \tilde{p}_1 \dfrac{\tilde{\epsilon} + \tilde{p}}{\tilde{\rho}}$ is positive). This difficulty does not seem to have appeared in the numerical experiments presented in [21], [30] and [33].

Finally, let us mention that in the simpler framework of Example 2, and provided that the functions f_k and g_k in (3.6)-(3.7) are monotone increasing, there exists a "natural" extension of Roe's scheme, as in the case of a perfect gas mixture (that is, without any arbitrary choice); the matrix $\tilde{A}$ constructed in this natural extension is always diagonalisable (see [2]).

3.4.4. The Osher scheme

It is easy to check that u and p remain Riemann invariants associated with the degenerate eigenvalue, and that Y is still a Riemann invariant associated with the first and fourth characteristic fields. But we are not presently able to write the exact expressions of all Riemann invariants for system (3.10)-(3.11); thus, we cannot generalize the Osher scheme in the present context (see however an approximate generalization in [4]).

Only in one particular case, namely when it is assumed that the second derivative $p_{\epsilon\epsilon}$ identically vanishes can we show that all invariants given in Table 1 are still Riemann invariants for the system (3.10)-(3.11). Therefore, in the case where $p_{\epsilon\epsilon} \equiv 0$, the first and fourth characteristic fields are genuinely non linear from (3.23)-(3.24), and analytical expressions are available for all Riemann invariants, a situation which is much favorable for constructing an exact Riemann solver and extending Osher's scheme. But this remark is probably of little interest in practice: in the framework of Example 2, the assumption $p_{\epsilon\epsilon} \equiv 0$ implies that either $\dfrac{f_k'(T)}{g_k'(T)}$ is a constant independent of k and T or that there exists a function $f_0(T)$ such that all ratios $\dfrac{f_k'(T)}{f_0(T)}$ and $\dfrac{g_k'(T)}{f_0(T)}$ are independent of T; one may wonder if, in practice, this does not happen only when all species behave as perfect gases !

4. CONCLUSIONS

We have examined the system of equations governing the one-dimensional inviscid flow of a multi-component perfect or real gas. This system appears to be hyperbolic, which makes possible to solve it as a whole system of conservation laws by extending the available upwind schemes designed for the Euler equations instead of treating separately the mixture conservation equations and the species conservation equations.

With the assumption that all species in the gaseous mixture obey the perfect gas laws, there is no difficulty in solving the multi-component Riemann problem; the extension to these mixtures of the most classical upwind schemes aimed at solving the Euler equations is also straightforward.

The situation is more complex in the case of real gas mixtures. The homogeneity property of the fluxes has been shown to hold for a large class of real gas mixtures and to lead to a natural definition of γ. But several difficulties arise for the description of the exact solution of the Riemann problem and the definition of robust, accurate and efficient numerical schemes.

We have not presented any numerical result; we refer the reader to [1], [4], [16] for results concerning one-dimensional flows of perfect gas mixtures, and to [21], [30], [33] for the simulation of the one-dimensional flow of a *single* real gas. Some of the "multi-component schemes" presented in this paper have also been applied to two-dimensional reactive or inert flows in [9], [10], [16]. We also refer to [18] for numerical experiments and comparisons which use the schemes presented in this paper (and also other schemes) to simulate mixture flows.

APPENDIX

We consider in this Appendix a general Riemann problem for *either a perfect gas or a real gas multi-component mixture*. We use very general hypotheses: in particular, no assumption of flux homogeneity or of genuine non linearity is needed in the sequel.

Let us first briefly recall some facts about the general Riemann problem (i.e. the Riemann problem for a general system of conservation laws), referring to Liu [31] and the references therein for more details. A solution of this problem is sought in the class of self-similar solutions consisting of shocks, rarefactions and contact discontinuities; this solution may involve a quite complex set of waves; in particular contact discontinuities associated with any of the eigenvalues $u-c$, u or $u+c$ may exist if the characteristic fields are not genuinely non linear (a contact discontinuity associated with the eigenvalue λ_k is defined as a discontinuity whose speed coincides with the eigenvalue λ_k on each side of the discontinuity).

We are going to prove the following result (which is stated here for a two-component mixture, but can be easily generalized to N-component flows):

Proposition A1:

If the general Riemann problem for an two-component mixture has a solution in the class of solutions consisting of shocks, rarefactions and contact discontinuities, then the mass fraction Y may change only across a contact discontinuity associated with the degenerate eigenvalue u. •

At the end of the Appendix, we will explain why this result makes it straightforward to extend a single-component perfect gas Riemann solver into a multi-component perfect gas Riemann solver.

PROOF: To begin with, we consider a 1-rarefaction wave. This wave corresponds to a region of the (x,t)-plane where $W(x,t) = V(\frac{x}{t})$, V being a C^1 function such that (we set σ for $\frac{x}{t}$):

$$\lambda_1[V(\sigma)] = \sigma \ , \tag{A.1}$$

$\frac{dV}{d\sigma}$ being parallel to $r_1[V(\sigma)]$ (see e.g. Lax [28] or Smoller [36]). Then the quantity:

$$\frac{d}{d\sigma}\left(\phi^{(1)}[V(\sigma)]\right) = \nabla_W \phi^{(1)} . \frac{dV}{d\sigma} \tag{A.2}$$

identically vanishes for any 1-Riemann invariant. Since the mass fraction Y is always a 1-Riemann invariant (this is true under the most general hypotheses, in particular for a real gas mixture with no homogeneity assumption), this shows that Y is constant in a 1-rarefaction wave.

The same argument applies for a 4-rarefaction wave. Beside this, a rarefaction wave associated with the degenerate eigenvalue u cannot exist; indeed, by differencing (A.1), it appears that one cannot construct a mapping $V(\sigma)$ satisfying (A.1) and such that $\frac{dV}{d\sigma}$ is parallel to an eigenvector associated with the degenerate eigenvalue.

Let us now consider a discontinuity, located along a line $x = st$ in the (x,t)-plane (s is the speed of the discontinuity); this discontinuity separates two states W_L and W_R which satisfy the Rankine-Hugoniot relations (see [28], [36]):

$$F(W_L) - F(W_R) = s(W_L - W_R) \; ; \tag{A.3}$$

in particular, (A.3) implies (we use the notation ΔU for $U_L - U_R$):

$$\Delta(\rho u) = s\Delta\rho \; . \tag{A.4}$$

$$\Delta(\rho Y u) = s\Delta(\rho Y) \; , \tag{A.5}$$

Using now the rules (2.70)-(2.71) for the operator Δ, we can rewrite (A.5) as:

$$\tilde{Y}\Delta(\rho u) + \hat{\rho}\tilde{u}\Delta Y = s(\tilde{Y}\Delta\rho + \hat{\rho}\Delta Y) \; , \tag{A.6}$$

which together with (A.4) yields:

$$\hat{\rho}(\tilde{u} - s)\Delta Y = 0 \; . \tag{A.7}$$

Therefore, the only discontinuities across which Y may change are such that $s = \tilde{u}$ (we assume that $\hat{\rho} > 0$ since Y is not defined in a vacuum region). But (A.4) then writes $\Delta(\rho u) = \tilde{u}\Delta\rho$. Since (2.70) yields $\Delta(\rho u) = \tilde{u}\Delta\rho + \hat{\rho}\Delta u$, we obtain $\Delta u = 0$, whence:

$$u_L = u_R = s \; , \tag{A.8}$$

which shows that the considered discontinuity is precisely a contact discontinuity associated with the degenerate eigenvalue u (furthermore, it is easy to deduce from (A.3) and (A.8) that the pressure is also continuous across this discontinuity: $p_L = p_R$). •

Remark A1: We want to point out that the above proof does not require the existence of a diagonalisable matrix $\tilde{A}$ satisfying Roe's property (2.63). We have just used (following an idea of Abgrall [3]) the arithmetic rules (2.70)-(2.71) of the operator Δ to analyse the Rankine-Hugoniot relations. •

Remark A2: In the case where the first and fourth characteristic fields are genuinely non linear, one can also use Lax's conditions to deduce from (A.7) that $\Delta Y = 0$ across entropic 1-shocks or 4-shocks. Indeed, an entropic 1-shock satisfies (see Lax [28]):

$$u_R - c_R < s < u_L - c_L \; , \quad s < u_R \; , \tag{A.9}$$

which shows that $s < \min(u_L, u_R)$. But $\tilde{u} \geq \min(u_L, u_R)$ from (2.64); thus $\tilde{u} - s > 0$, and (A.7) implies $\Delta Y = 0$. •

Remark A3: Proposition A1 implies that, except at a contact discontinuity associated with the eigenvalue u where its composition changes, the mixture behaves *as a single real gas*. Nevertheless, constructing an exact Riemann solver for real gas mixtures remains a difficult task and is still the subject of current research: under additional hypotheses on the pressure equation (3.11) which insure the genuine non linearity of the first and last characteristic fields (such as a convexity assumption $p_{\epsilon\epsilon} > 0$), it suffices to extend the results obtained by Colella and Glaz [11] for a single real gas; without such assumptions, one needs to extend the more general results of Liu [31]. •

Remark A4: In the case of a *perfect gas* mixture, the result of Proposition A1 makes possible to completely solve the Riemann problem. As in the single component case (see e.g. [20], [36], [37]), it is convenient to construct the solution of the Riemann problem by intersecting in the (u,p)-plane the curve $\mathcal{C}_L$ of those states W^2 which can be linked to W_L by either a 1-rarefaction or an entropic 1-shock (W_L being on the left side of the rarefaction or shock wave), and the curve $\mathcal{C}_R$ of those states W^3 which can be linked to W_R by either a 4-rarefaction or an entropic 4-shock (W_R being on the right side of the rarefaction or shock wave). Proposition A1 implies that these curves can be described by the same analytical expressions as in the single-component case, with γ constant and equal to $\gamma(W_L)$ on the curve $\mathcal{C}_L$ and γ constant and equal to $\gamma(W_R)$ on $\mathcal{C}_R$. Then, it just remains to check that these curves actually intersect (see [1] and [4] for the details), and that two states which have the same velocity and the same pressure can be separated by a contact discontinuity. The argument for this last point is the same as in the single component case, since the equalities $u_L = u_R$, $p_L = p_R$ yield the Rankine-Hugoniot relations:

$$F(W_L) - F(W_R) = u_L(W_L - W_R) \ , \qquad (A.10)$$

which exactly says that the states W_L and W_R can be separated by a contact discontinuity moving at speed $s = u_L$ (no considerations on the entropy variation across this discontinuity is needed here since the fluid does not cross the discontinuity but moves parallel to it). •

ACKNOWLEDGEMENTS

We thank our colleagues R. Abgrall and A. Dervieux for several helpful conversations we had with them about this work and for their help in proofreading the typescript.

REFERENCES

[1] R. ABGRALL, "Généralisation du schéma de Roe pour le calcul d'écoulements de mélanges de gaz à concentrations variables", to appear in La Recherche Aérospatiale.

[2] R. ABGRALL, "Preliminary results on the extension of Roe's scheme to a class of real gas mixtures", to appear.

[3] R. ABGRALL, private communication.

[4] R. ABGRALL & J. L. MONTAGNE, "Généralisation du schéma d'Osher pour le calcul d'écoulements de mélanges de gaz à concentrations variables et de gaz réels", submitted to La Recherche Aérospatiale.

[5] J. A. BEATTIE & I. OPPENHEIM, "Principles of thermodynamics", Studies in modern thermodynamics, **2**, Elsevier, Amsterdam, (1979).

[6] F. BENKHALDOUN, A. DERVIEUX, G. FERNANDEZ, H. GUILLARD & B. LARROUTUROU, "Some finite-element investigations of stiff combustion problems: mesh adaption and implicit time-stepping", Mathematical modelling in combustion and related topics, Brauner & Schmidt-Lainé eds., pp. 393-409, NATO ASI Series E, Nijhoff, Doordrecht, (1988).

[7] A. BOURGEADE, "Quelques méthodes numériques pour le traitement des écoulements réactifs", CEA Report CEA-N-2570, (1988).

[8] J. D. BUCKMASTER & G. S. S. LUDFORD, "Theory of laminar flames", Cambridge Univ. Press, Cambridge, (1982).

[9] G. V. CANDLER & R. W. McCORMACK, "The computation of hypersonic flows in chemical and thermal nonequilibrium", Paper N^o 107, Third National Aero-Space Plane Technology Symposium, (1987).

[10] D. CHARGY, A. DERVIEUX & B. LARROUTUROU, "Upwind adaptive finite-element investigations of two-dimensional transonic reactive flows", to appear.

[11] P. COLELLA & H. M. GLAZ, "Efficient solution algorithms for the Riemann problem for real gases", J. Comp. Phys., **59**, pp. 264-289, (1985).

[12] R. COURANT & K. O. FRIEDRICHS, "Supersonic flow and shock waves", Appl. Math. Sciences, **21**, Springer Verlag, New-York, (1948).

[13] A. DERVIEUX, "Steady Euler simulations using unstructured meshes", Partial differential equations of hyperbolic type and applications, Geymonat ed., pp. 33-111, World Scientific, Singapore, (1987).

[14] J. A. DESIDERI, N. GLINSKY & E. HETTENA, "Hypersonic reactive flow computations", to appear in Computers and Fluids.

[15] G. FERNANDEZ, "Implicit conservative upwind schemes for strongly transient flows", INRIA Report 873, (1988).

[16] G. FERNANDEZ & B. LARROUTUROU, "On the use of hyperbolic schemes for multicomponent Euler flows", pròceedings of the Second Int. Conf. on hyperbolic problems, to appear.

[17] L. FEZOUI, "Résolution des équations d'Euler par un schéma de Van Leer en éléments finis", INRIA Report 358, (1985).

[18] L. FEZOUI & B. LARROUTUROU, "Upwind conservative schemes for multi-component perfect or real gas flows", INRIA Report, to appear.

[19] L. FEZOUI & B. STOUFFLET, "A class of implicit upwind schemes for Euler simulation with unstructured meshes", to appear in J. Comp. Phys..

[20] H. GILQUIN, "Analyse numérique d'un problème hyperbolique multidimensionnel en dynamique des gaz avec frontières mobiles", Thesis, Université de Saint-Etienne, (1984).

[21] P. GLAISTER, "An approximate linearised Riemann solver for the Euler equations for real gases", J. Comp. Phys., **74**, pp. 382-408, (1988).

[22] J. GLIMM, "Solutions in the large for non linear hyperbolic systems of equations", Comm. Pure Appl. Math., **18**, pp. 697-715, (1965).

[23] S. K. GODUNOV, "A difference scheme for numerical computation of discontinuous solutions of equations of fluid dynamics", Math. Sbornik, **47**, pp. 271-306, (1959) (in russian).

[24] N. A. GOKCEN, "Thermodynamics", Techscience Inc., Hawthorne, (1975).

[25] A. HABBAL, A. DERVIEUX, H. GUILLARD & B. LARROUTUROU, "Explicit calculations of reactive flows with an upwind finite-element hydrodynamical code", INRIA Report 690, (1987).

[26] A. HARTEN, P. D. LAX & B. VAN LEER, "On upstream differencing and Godunov type schemes for hyperbolic conservation laws", SIAM Review, **25**, pp. 35-61, (1983).

[27] B. LARROUTUROU, "Introduction to mathematical and numerical modelling in gaseous combustion", Applied Mathematics, Gordon and Breach, to appear.

[28] P. D. LAX, "Hyperbolic systems of conservation laws and the mathematical theory of shock waves", CBMS regional conference series in applied mathematics, **11**, SIAM, Philadelphia, (1972).

[29] A. LERAT, "Propriété d'homogénéité et décomposition des flux en dynamique des gaz", J. Méca. Théor. Appl., **2**, (2), pp. 185-213, (1983).

[30] M. S. LIOU, B. VAN LEER & J. S. SHUEN, "Splitting of inviscid fluxes for real gases", NASA Technical memorandum 100856, (1988).

[31] T. P. LIU, "The Riemann problem for general systems of conservation laws", J. Diff. Equ., **18**, pp. 218-234, (1975).

[32] A. MAJDA, "High Mach number combustion", Combustion and chemical reactors, Ludford ed., pp. 109-184, Lecture in Appl. Math., **24**, (1), AMS, Providence, (1986).

[33] J. L. MONTAGNE, H. C. YEE & M. VINOKUR, "Comparative study of high-resolution shock capturing schemes for a real gas", NASA Technical memorandum 100004, (1987).

[34] S. OSHER & F. SOLOMON, "Upwind schemes for hyperbolic systems of conservation laws", Math. Comp., **38**, (158), pp. 339-374, (1982).

[35] P. L. ROE, "Approximate Riemann solvers, parameters vectors and difference schemes", J. Comp. Phys., **43**, pp. 357-372, (1981).

[36] J. SMOLLER, "Shock waves and reaction-diffusion equations", Springer Verlag, New-York, (1983).

[37] G. A. SOD, "A survey of several finite-difference methods for systems of nonlinear hyperbolic conservation laws", J. Comp. Phys., **27**, pp. 1-31, (1977).

[38] J. L. STEGER & R. F. WARMING, "Flux vector splitting for the inviscid gas dynamic equations with applications to finite-difference methods", J. Comp. Phys., **40**, (2), pp. 263-293, (1981).

[39] B. VAN LEER, "Towards the ultimate conservative difference scheme III - Upstream centered finite-difference schemes for ideal compressible flow", J. Comp. Phys., **23**, pp. 263-275, (1977).

[40] B. VAN LEER, "Flux-vector splitting for the Euler equations", Eighth international conference on numerical methods in fluid dynamics, Krause ed., pp. 507-512, Lecture notes in physics, **170**, Springer-Verlag, (1982).

[41] B. VAN LEER, J. L. THOMAS, P. L. ROE & R. W. NEWSOME, "A comparison of numerical flux formulas for the Euler and Navier-Stokes equations", AIAA paper 87-1104, (1987).

[42] G. VIJAYASUNDARAM, "Transonic flow simulations using an upstream-centered scheme of Godunov in finite elements", J. Comp. Phys., **63**, (1986).

[43] F. A. WILLIAMS, "Combustion theory", second edition, Benjamin Cummings, Menlo Park, (1985).

Global existence of large amplitude solutions for Dirac-Klein-Gordon systems in Minkowski space

Alain BACHELOT

Département de Mathématiques Appliquées
Université de Bordeaux I
351, Cours de la Libération
33405 TALENCE

INTRODUCTION

The purpose of this paper is to prove the existence of some global solutions, with *large* energy, of Dirac-Klein-Gordon systems with quadratic coupling and cubic autointeractions in Minkowski space. We know that some algebraic conditions on the nonlinearities, allow to solve the global Cauchy problem for classical fields with *small* initial data [2] [9] : the notion of compatibility of a product with a differential system, introduced by B. Hanouzet and J.L. Joly [5,6,7], and the null condition of S. Klainerman [9]. These both conditions are related to the Lorentz invariance.

In this work we show that the global Cauchy problem is wellposed again for arbitrarly large initial data if the non linearities and the data satisfy some algebraic properties ; more precisely we assume the system is Lorentz-invariant and the polarization of the Cauchy data is such that the *chiral invariant* is small.

Let's consider the mass Dirac-Klein-Gordon system in Minkowski space $\mathbb{R}^{3+1}$ with Lorentz metric $g_{\mu,\nu}$=diag(1,-1,-1,-1)

$$-i\gamma^{\mu}\partial_{\mu}\psi + M\psi = f(\varphi,\psi), \tag{1}$$

$$\Box\varphi + m^2\varphi = g(\varphi,\psi) . \tag{2}$$

We suppose the masses are non null

$$M \neq 0 , m \neq 0 . \tag{3}$$

Now we introduce the Lorentz invariants

$$\overline{\psi}\psi = \tilde{\psi}\gamma^{o}\psi, \ \tilde{\psi}=\text{transposate conjugate of } \psi,$$

$$\overline{\psi}\gamma^{5}\psi, \ \gamma^{5}=-i\gamma^{o}\gamma^{1}\gamma^{2}\gamma^{3} \ ,$$

and we define the vector space $\mathcal{M}$ of 4×4 matrices

$$\mathcal{M} = \{\alpha I+i\beta\gamma^{5}, \ (\alpha,\beta)\in\mathbb{R}^{2}\} \ .$$

The hypotheses on the nonlinearities are following :

$$f(\varphi,\psi)=\varphi V\psi+F(\overline{\psi}\psi,i\overline{\psi}\gamma^{5}\psi)\psi \tag{4}$$

where V is a 4×4 matrix with constant coefficients and

$$V\in\mathcal{M}, \ F\in C^{\infty}(\mathbb{R}^{2},\mathcal{M}), \ |F(u,v)|=O(|u|+|v|), |u|+|v|\to 0 \ , \tag{5}$$

$$g(\varphi,\psi)=G(\overline{\psi}\psi,i\overline{\psi}\gamma^{5}\psi)-k\varphi^{3} \tag{6}$$

where k is a real constant and

$$G\in C^{\infty}(\mathbb{R}^{2},\mathbb{R}), |G(u,v)|=O(|u|+|v|), |u|+|v|\to 0 \ . \tag{7}$$

Obviously, to obtain large solutions, we must assume

$$k \geqslant 0 \ . \tag{8}$$

Many models of the relativistic fields theory satisfy these hypotheses : the scalar and pseudoscalar Yukawa models of the nuclear forces, the interactions of Heisenberg, Federbusch, the magnetic monopole of G. Lochak.

Now, we recall that J. Chadam and R.T. Glassey established in [4] the existence of global solutions to the scalar Yukawa model, for which the Dirac system and the Klein-Gordon equation are decoupled and $\overline{\psi}\psi\equiv 0$.

Here, we solve the global Cauchy problem for (1)-(8) in a neighborhood of such a decoupling solution. More precisely, we choose

$$\psi\Big|_{t=0} = \Psi_{o} + \varepsilon\chi_{o} \ , \ 0<\varepsilon \tag{9}$$

$$\Psi_{o}, \ \chi_{o}\in\mathcal{D}(\mathbb{R}^{3}_{x},\mathbb{C}^{4}) \tag{10}$$

$$\varphi\Big|_{t=0} = \varphi_0, \quad \partial_t\varphi\Big|_{t=0} = \varphi_1 \tag{11}$$

$$\varphi_j \in \mathcal{D}(\mathbb{R}^3_x, \mathbb{R}) \quad . \tag{12}$$

The algebraic hypothese on the polarization of Ψ_0 is

$$\Psi_0 = z\gamma^2\Psi_0^+, z\in\mathbb{C}, |z|=1, \Psi_0^+ = \text{conjugate of } \Psi_0 \tag{13}$$

where

$$\gamma^2 = \begin{bmatrix} 0 & \sigma^2 \\ -\sigma^2 & 0 \end{bmatrix}, \quad \sigma^2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} .$$

(13) is Majorana's condition generalized by G. Lochak [11].
In the first part we prove condition (13) is time independent for the solution of a Dirac system with scalar or pseudoscalar time dependent potential.
In the second part we make energy estimates and uniform decay estimates in Sobolev spaces associated with Lorentz metric for the nonlinear Klein-Gordon equation

$$\Box\varphi + \varphi = -\varphi^3 \quad .$$

We solve global Cauchy problem (1)-(13) in part III ; we obtain asymptotically free solutions.

I - CHIRAL INVARIANT OF DIRAC FIELD.

We consider a solution ψ of the Dirac system

$$-i\gamma^\mu\partial_\mu\psi + M\psi = A\psi, \quad M\in\mathbb{R} \tag{14}$$

where the time dependent potential A satisfies

$$A, \ \partial_\mu A \in L^\infty(\mathbb{R}_t\times\mathbb{R}^3_x; \mathcal{M}) . \tag{15}$$

Following G. Lochak [11] we introduce the *chiral invariant* of ψ, $\rho(\psi)$

$$\rho^2 = |\overline{\psi}\psi|^2 + |\overline{\psi}\gamma^5\psi|^2 \quad . \tag{16}$$

We are concerned by the solution ψ for which the Chiral invariant is null.

PROPOSITION I.1 :

Let ψ be a solution of (14) and $\psi \in C^0(\mathbb{R}_t(L^2(\mathbb{R}^3_x))^4)$, $\psi\Big|_{t=0} = \psi_0 \in (L^2(\mathbb{R}^3_x))^4$.

Then the following assertions are equivalent :

(i) $\psi_0 = z\gamma^2\psi_0^+$, $z \in \mathbb{C}$, $|z|=1$, ψ_0^+ conjugate of ψ_0,

(ii) $\forall x \in \mathbb{R}^3$, $\rho(\psi_0(x)) = 0$,

(iii) $\forall (t,x) \in \mathbb{R}^{1+3}$, $\rho(\psi(t,x)) = 0$.

Proof : We use the bispinorial representation of Weyl by putting

$$\psi = 2^{-½}(\gamma^0+\gamma^5)\begin{pmatrix}\xi \\ \eta\end{pmatrix}, \quad \xi,\eta \in \mathbb{C}^2 \quad . \tag{17}$$

We verify easily that

$$\bar{\psi}\psi = \xi^+\eta+\eta^+\xi \quad , \quad \bar{\psi}\gamma^5\psi = \xi^+\eta-\eta^+\xi \quad .$$

Therefore $\rho=0$ if and only if

$$\xi^+\eta = 0 \quad .$$

This condition is equivalent to

$$\xi = z\sigma^2\eta^+ \quad , \quad z \in \mathbb{C} \quad , \quad |z|=1 \quad .$$

By using (17) we see that this equality means

$$\psi = z\gamma^2\psi^+$$

and we conclude that

$$\psi = z\gamma^2\psi^+ \iff \rho(\psi) = 0 \quad . \tag{18}$$

Now it is obvious that it is sufficient to prove (ii) $\iff$ (iii) for $\psi_0 \in (H^1(\mathbb{R}^3_x))^4$ with compact support. Equation (14) can be written

$$\partial_0\psi + \sum_{j=1}^{3}\gamma^0\gamma^j\partial_j\psi + iM\gamma^0\psi = i\gamma^0 A\psi \quad . \tag{19}$$

By multiplying (19) by $\tilde{\psi}$ we find

$$\partial^{o}|\psi|^{2} + \sum_{j=1}^{3} \partial_{j}(\tilde{\psi}\gamma^{o}\gamma^{j}\psi) = 0 \ .$$

We integrate (20) over $\mathbb{R}^3_x$ and we obtain the charge conservation

$$\int_{\mathbb{R}^3} |\psi(t,x)|^2 dx = \text{cst} \ . \tag{21}$$

Now we multiply (19) by ${}^t\psi\gamma^2$, ${}^t\psi$ is the transposate of ψ and it follows

$$\partial_{o}({}^t\psi\gamma^{2}\psi) + \sum_{j=1}^{3} \partial_{j}({}^t\psi\gamma^{2}\gamma^{o}\gamma^{j}\psi) = 0 \ .$$

We integrate over $\mathbb{R}^3_x$ again and we obtain the conservation law

$$\int_{\mathbb{R}^3} {}^t\psi(t,x)\gamma^2\psi(t,x)dx = \text{cst} \ . \tag{22}$$

Let z be a complex number of modulus one. We have

$$|\psi - z\gamma^2\psi^+|^2 = 2|\psi|^2 + 2\mathcal{R}e(\overline{z}\ {}^t\psi\gamma^2\psi) \ .$$

Then we have thanks to (21) and (22)

$$\int_{\mathbb{R}^3} |\psi(t,x) - z\gamma^2\psi^+(t,x)|^2 dx = \text{cst} \ . \tag{23}$$

By (18) and (23) we conclude that $\rho(\psi_o)$ is equivalent to $\rho(\psi) \equiv 0$. Q.E.D.

II - ESTIMATES FOR THE NONLINEAR KLEIN-GORDON EQUATION.

We define the Sobolev norms associated with the Lorentz metric ; for any test function $u \in \mathcal{D}(\mathbb{R}_t \times \mathbb{R}^3_x)$ and any integer N, we put

$$\|u(t)\|^2_N = \sum_{|\lambda| \leq N} \|\Gamma^\lambda u(t)\|^2_{L^2(\mathbb{R}^3_x)} \tag{24}$$

$$|u(t)|_N = \operatorname*{Sup}_{|\lambda| \leq N} |\Gamma^\lambda u(t)|_{L^\infty(\mathbb{R}^3_x)} \tag{25}$$

where

$$\Gamma^{\lambda}=\Gamma_1^{\lambda_1}\ldots\Gamma_{10}^{\lambda_{10}} \quad , \quad \lambda\in\mathbb{N}^{10}$$

$$|\lambda|=\lambda_1+\ldots\lambda_{10} \quad ,$$

and $(\Gamma_\sigma)_{1\leqslant\sigma\leqslant 10}$ are the generators of Poincaré group

$$(\Gamma_\sigma)_{1\leqslant\sigma\leqslant 10} = (\partial_\mu=\frac{\partial}{\partial x^\mu} \quad , \quad x_\mu\partial_\nu-x_\nu\partial_\mu)_{0\leqslant\mu,\nu\leqslant 3} \quad . \tag{26}$$

We will writte so

$$\partial_0=\partial_t \quad , \quad x^0=t \quad , \quad x=(x^1,x^2,x^3) \, .$$

In this part, our purpose is to estimate with these norms the solution u of the nonlinear Klein-Gordon equation

$$\Box u + u = -gu^3, \quad 0\leqslant g \quad , \tag{27}$$

$$u\Big|_{t=0}\in\mathcal{D}(\mathbb{R}_x^3,\mathbb{R}) \, , \partial_t u\Big|_{t=0}\in\mathcal{D}(\mathbb{R}_x^3,\mathbb{R}) \, . \tag{28}$$

PROPOSITION II.1 -

The solution u of (27) (28) satisfies for any integer N :

$$\sup_{t\in\mathbb{R}} \|u(t)\|_N < +\infty \tag{29}$$

$$\sup_{t\in\mathbb{R}} (1+|t|)^{3/2}|u(t)|_N < +\infty \quad . \tag{30}$$

<u>Proof</u> : First, we prove by iteration on N the following assertion P_N :

$$(P_N) \quad \begin{cases} \text{there exists } d_N>1 \text{ such that} \\ \sup_{t\in\mathbb{R}} (\|u(t)\|_N+\|u'(t)\|_N+(1+|t|)^{d_N}|u(t)|_{N-1})<\infty \end{cases}$$

where $u'=(\partial_\mu u)_{0\leqslant\mu\leqslant 3}$.

Recall the result of C. Morawetz and W. Strauss [12] :

$$\sup_{t\in\mathbb{R}} (1+|t|)^{3/2}|u(t)|_0 < +\infty . \tag{31}$$

Now we note Ω_N an element of order $\leqslant N$, of the Lie algebra spanned by the generators of Poincaré group

$$\Omega_N = \sum_{\text{finite}} C_\lambda \Gamma^\lambda, \ C_\lambda\in\mathbb{C}, \ \lambda\in\mathbb{N}^{10}, \ |\lambda|\leqslant N . \tag{32}$$

The Lorentz invariance of the Klein-Gordon equation implies

$$\Box\Omega_1 u+\Omega_1 u = -g\Omega_1(u^3) .$$

It follows

$$\|u(t)\|_1+\|u'(t)\|_1\leqslant C(1+\int_0^t \|u(s)\|_1 |u(s)|_0^2 ds)$$

and by using Gronwall's lemma

$$\sup_{t\in\mathbb{R}} (\|u(t)\|_1+\|u'(t)\|_1)\leqslant C \exp(C\int |u(s)|_0^2 ds) . \tag{33}$$

We conclude by (31) and (33) that (P_1) is verified. Now, assume (P_N) is proved ; we have again

$$\Box\Omega_{N+1}u+\Omega_{N+1}u = -g\Omega_{N+1}(u^3)$$

and thus

$$\|\Omega_{N+1}u(t)\|_0 +\|\Omega_{N+1}u'(t)\|_0 \leqslant C(1+\int_0^t \|\Omega_{N+1}(u^3)(s)\|_0 \ ds) .$$

We note

$$\Omega_{N+1}(u^3) = \sum_{\text{finite}} \{(\Omega_{N+1}u)u^2+(\Omega_N u)(\Omega_1 u)u+\prod_{j=1}^{3}(\Omega_{p_j} u)\}$$

where

$$p_j \leqslant N-1 \ , \ \sum_1^3 p_j \leqslant N+1$$

and then

$$\|u^3(s)\|_{N+1}\leqslant C(\|u(s)\|_{N+1}|u(s)|_0^2+\|u(s)\|_{N-1}|u(s)|_{N-1}^2$$
$$+\sum_{\text{finite}}\|\Omega_N u(s)\|_{L^4(\mathbb{R}^3_x)}\cdot\|\Omega_1 u(s)\|_{L^4(\mathbb{R}^3_x)}\cdot|u(s)|_0)\,.$$

Thanks to the Sobolev injection $H^1(\mathbb{R}^3_x)\subset L^4(\mathbb{R}^3_x)$, (P_1) and (P_N) we find.

$$\|u(t)\|_{N+1}+\|u'(t)\|_{N+1}\leqslant C(1+|\int_0^t\|u(s)\|_{N+1}(1+|s|)^{-d_N}ds|)$$

and Gronwall's lemma gives

$$\sup_{t\in\mathbb{R}}(\|u(t)\|_{N+1}+\|u'(t)\|_{N+1})<+\infty\ . \tag{34}$$

Now, we recall that the solution v of

$$\Box v+v=0,\ v\Big|_{t=0}=0\ ,\ \partial_0 v\Big|_{t=0}(x)=g(x)\,,$$

verifies for $0\leqslant\alpha\leqslant 1$

$$|v(t)|_0\leqslant C|t|^{-\alpha-(1-\alpha)(3/2)}\|g\|^{\alpha}_{W^{1,1}(\mathbb{R}^3_x)}\|g\|^{1-\alpha}_{W^{2,1}(\mathbb{R}^3_x)}\,,$$

$$|v(t)|_0\leqslant C\|g\|_{W^{1,2}(\mathbb{R}^3_x)}\,,$$

where

$$\|g\|_{W^{n,p}(\mathbb{R}^3_x)}=\sum_{|\alpha|\leqslant n}\|\partial^\alpha_x g\|_{L^p(\mathbb{R}^3_x)}\,.$$

It follows that

$$|\Omega_N u(t)|_0\leqslant C\Big[(1+|t|)^{-3/2}+|\int_0^t(1+|t-s|)^{-\alpha-(1-\alpha)(3/2)} \tag{35}$$
$$(\|\Omega_N u^3(s)\|_{W^{1,2}(\mathbb{R}^3_x)}+\|\Omega_N u^3(s)\|^{\alpha}_{W^{1,1}(\mathbb{R}^3_x)}\cdot\|\Omega_N u^3(s)\|^{1-\alpha}_{W^{2,1}(\mathbb{R}^3_x)}ds|\Big]$$

We write again

$$\Omega_N(u^3)=\sum_{\text{finite}}{}'\ \{(\Omega_N u)u^2+(\Omega_{N-1}u)(\Omega_1 u)u+\prod_{j=1}^{3}(\Omega_{p_j}u)\}$$

where

$p_j\leqslant N-2$.

Then hypothese (P_N) and (34) imply

$$\|\Omega_N(u^3)(s)\|_{W^{1,2}(\mathbb{R}^3_x)\cap W^{1,1}(\mathbb{R}^3_x)} \leqslant C(1+|s|)^{-d_N} \tag{36}$$

and thanks to (34)

$$\sup_{t\in\mathbb{R}} \|\Omega_N(u^3)(s)\|_{W^{2,1}(\mathbb{R}^3_x)} < +\infty \ . \tag{37}$$

The inequalities (35) (36) (37) yield

$$|u(t)|_N \leqslant C((1+|t|)^{-3/2} + \left|\int_0^t (1+|t-s|)^{-\alpha-(1-\alpha)3/2}(1+|s|)^{-d_N} ds\right|) .$$

We choose

$$\alpha = 3(2d_N+1)^{-1} \in]d_N^{-1},1[\ ,$$

$$d_{N+1} = \alpha+(1-\alpha)(3/2) = \alpha d_N > 1 \ .$$

Thus

$$|u(t)|_N \leqslant C(1+|t|)^{-d_{N+1}}$$

this ends the proof of (P_{N+1}). To obtain the uniform decay of $t^{-3/2}$ we apply the L^2-L^∞ estimate for the Klein-Gordon equation [2] :

$$|u(t)|_N \leqslant C(1+|t|)^{-3/2}(1+\int_{\mathbb{R}} \|u^3(s)\|_{N+4} ds)$$

(P_{N+5}) implies

$$\|u^3(s)\|_{N+4} \leqslant C(1+|s|)^{-2d_{N+5}} \in L^1(\mathbb{R}_s)$$

and we conclude that

$$\sup_{t\in\mathbb{R}} (1+|t|)^{3/2}|u(t)|_N < +\infty \qquad \text{Q.E.D}$$

III - GLOBAL EXISTENCE OF LARGE AMPLITUDE SOLUTIONS.

MAIN THEOREM

There exists $\varepsilon_0>0$ depending only on the derivatives of initial data Ψ_0, χ_0, φ_0,φ_1 of order $\leqslant 10$, such that for any $0\leqslant\varepsilon\leqslant\varepsilon_0$, the Cauchy problem (1) to (13) has a unique solution (ψ,φ) in $C^\infty(\mathbb{R}^4)$. Moreover, this solution is asymptotically free : there exists $\psi^\pm$, $\varphi^\pm$ satisfying :

$$\psi^\pm\in\bigcap_k C^k(\mathbb{R}_t,H^{10-k}(\mathbb{R}^3_x)),-i\gamma^\mu\partial_\mu\psi^\pm+M\psi^\pm=0\ ,$$

$$\varphi^\pm\in\bigcap_k C^k(\mathbb{R}_t,H^{11-k}(\mathbb{R}^3_x)),\Box\varphi^\pm+m^2\varphi^\pm=0\ ,$$

$$\forall k\in\mathbb{N},\ \lim_{t\to\pm\infty}\|\partial_t^k\psi(t)-\partial_t^k\psi^\pm(t)\|_{H^{10-k}}+\|\partial_t^k\varphi(t)-\partial_t^k\varphi^\pm(t)\|_{H^{11-k}}=0$$

where H^s is the Sobolev space $W^{s,2}(\mathbb{R}^3_x)$.

Proof : Let (Ψ,Φ) be the solution of

$$-i\gamma^\mu\partial_\mu\Psi + M\Psi = \Phi V\Psi\ , \tag{38}$$

$$\Box\Phi + m^2\Phi = -k\Phi^3\ , \tag{39}$$

$$\Psi\Big|_{t=0} = \Psi_0\ , \tag{40}$$

$$\Phi\Big|_{t=0} = \varphi_0,\ \partial_t\Phi\Big|_{t=0}=\varphi_1\ . \tag{41}$$

Lochak-Majorana's condition (13) and Proposition I.1 imply

$$\overline{\Psi}\Psi = \overline{\Psi}\gamma^5\Psi \equiv 0\ . \tag{42}$$

Now we put

$$\psi = \Psi+\chi\ ,\ \varphi = \Phi+u \tag{43}$$

and to solve the Cauchy problem for (ψ,φ) we study the problem

$$-i\gamma^\mu\partial_\mu\chi+M\chi = \tilde{f}(\chi,u;\Psi,\Phi) \tag{44}$$

$$\Box u+m^2u = \tilde{g}(\chi,u;\Psi,\Phi) \tag{45}$$

$$\chi\Big|_{t=0}=\varepsilon\chi_0\ ,\ u\Big|_{t=0}=0\ ,\ \partial_t u\Big|_{t=0}=0 \tag{46}$$

where $\tilde{f}$ and $\tilde{g}$ are C^∞ functions of its variables and verify

$$|h(\chi,u;\Psi,\Phi)|=0((|\chi|+|u|)(|\chi|+|u|+|\Psi|+|\Phi|)(1+|\chi|+|u|+|\Psi|+|\Phi|))$$

as

$$|\chi|+|u| \to 0, \ h=\tilde{f},\tilde{g} \ . \tag{47}$$

As usual we define the sequence $(\chi^\nu,u^\nu)_{\nu\geq 0}$ by putting

$$\chi^0 \equiv 0, \ u^0 \equiv 0 \ , \tag{48}$$

and for $\nu\geq 1$

$$-i\gamma^\mu\partial_\mu\chi^\nu+M\chi^\nu=\tilde{f}(\chi^{\nu-1},u^{\nu-1};\Psi,\Phi) , \tag{49}$$

$$\Box u^\nu+m^2u^\nu=\tilde{g}(\chi^{\nu-1},u^{\nu-1};\Psi,\Phi) , \tag{50}$$

$$\chi^\nu\Big|_{t=0}=\varepsilon\chi_0 \ , \ u^\nu\Big|_{t=0}=\partial_t u^\nu\Big|_{t=0}=0 \ . \tag{51}$$

To estimate the norms $\|\chi^\nu(t)\|_N$ we replace in (24) (25) the operators $(\Gamma_\sigma)_{1\leq\sigma\leq 10}$ by the Fermi operators

$$(\tilde{\Gamma}_\sigma)_{1\leq\sigma\leq 10} = (\partial_\mu, x_\mu\partial_\nu-x_\nu\partial_\mu+½\gamma_\mu\gamma_\nu)_{0\leq\mu,\nu\leq 3}$$

which define obviously equivalent norms, and commute with the Dirac system. The commutation relations for $\tilde{\Gamma}_\sigma$ the charge conservation for the Dirac system and the usual energy equality for the Klein-Gordon equation imply

$$\begin{aligned}\|\chi^\nu(t)\|_N+\|u^\nu(t)\|_N+\|(u^\nu)'(t)\|_N \leq \\ \leq C\Big[\varepsilon+\Big|\int_0^t (\|\chi^{\nu-1}(s)\|_N+\|u^{\nu-1}(s)\|_N) \\ \times(|\chi^{\nu-1}(s)|_{[N/2]}+|u^{\nu-1}(s)|_{[N/2]}+|\Psi(s)|_N+|\Phi(s)|_N) \\ \times(1+|\chi^{\nu-1}(s)|_{[N/2]}+|u^{\nu-1}(s)|_{[N/2]}+|\Psi(s)|_N+|\Phi(s)|_N)ds\Big|\Big]\end{aligned}$$

where

$$(u^\nu)' = (\partial_\mu u^\nu)_{0\leq\mu\leq 3} \ .$$

Following Proposition II.1, we have

$$|\Psi(s)|_N+|\Phi(s)|_N \leq C_N(1+|s|)^{-3/2} \ .$$

We deduce that

$$\begin{aligned}\|\chi^{\nu}(t)\|_{N}+\|u^{\nu}(t)\|_{N}+\|(u^{\nu}(t))'\|_{N} &\leqslant \\ \leqslant\Big[\varepsilon+\Big|\int_{0}^{t}&(\|\chi^{\nu-1}(s)\|_{N}+\|u^{\nu-1}(s)\|_{N}) \\ &\times(|\chi^{\nu-1}(s)|_{[N/2]}+|u^{\nu-1}(s)|_{[N/2]}+(1+|s|)^{-3/2}) \\ &\times(1+|\chi^{\nu-1}(s)|_{[N/2]}+|u^{\nu-1}(s)|_{[N/2]})ds\Big|\Big] .\end{aligned} \tag{52}$$

We define for $n\in\mathbb{N}$, $t\in\mathbb{R}$

$$a_{n}(t)=\sup_{\substack{|s|\leqslant|t| \\ 0\leqslant\nu\leqslant n}}(\|\chi^{\nu}(s)\|_{N}+\|u^{\nu}(s)\|_{N}+\|(u^{\nu}(s))'\|_{N})$$

$$b_{n}(t)=\sup_{\substack{|s|\leqslant|t| \\ 0\leqslant\nu\leqslant n}}((1+|s|)^{3/2}(|\chi^{\nu}(s)|_{[N/2]}+|u^{\nu}(s)|_{[N/2]})) .$$

Inequality (52) can be written

$$a_{n}(t)\leqslant C(\varepsilon+\Big|\int_{0}^{t}a_{n-1}(s)(1+b_{n-1}(s))^{2}(1+|s|)^{-3/2}ds\Big|) . \tag{53}$$

At present our L^{2}-L^{∞} estimate for Klein-Gordon equation [2] gives

$$\begin{aligned}|\chi^{\nu}(t)|_{[N/2]}+|u^{\nu}(t)|_{[N/2]}&\leqslant C(1+|t|)^{-3/2} \\ (\varepsilon+\Big|\int_{0}^{2t}&(\|\chi^{\nu-1}(s)\|_{[N/2]+4}+\|u^{\nu-1}(s)\|_{[N/2]+4}) \\ \times(|\chi^{\nu-1}(s)|_{[([N/2]+4)/2]}&+|u^{\nu-1}(s)|_{[([N/2]+4)/2]} \\ &+|\Psi(s)|_{[N/2]+4} + |\Phi(s)|_{[N/2]+4}) \\ \times(1+|\chi^{\nu-1}(s)|_{[([N/2]+4)/2]}&+|u^{\nu-1}(s)|_{[([N/2]+4)/2]} + \\ &+|\Psi(s)|_{[N/2]+4}+|\Phi(s)|_{[N/2]+4})ds\Big|) .\end{aligned}$$

We choose N such that

$$\left[\frac{N}{2}\right]+4 \leqslant N, \quad \left[\frac{[N/2]+4}{2}\right] \leqslant \left[\frac{N}{2}\right]$$

i.e.

$$10 \leqslant N . \tag{54}$$

Then we have

$$b_n(t)\leqslant C(\varepsilon+|\int_0^{2t} a_{n-1}(s)(1+b_{n-1}(s))^2(1+|s|)^{-3/2}ds|). \qquad (55)$$

Relations (53) and (55) show that if a_{n-1} and b_{n-1} are in $L^\infty_{loc}(\mathbb{R})$, then a_n and b_n are in $L^\infty_{loc}(\mathbb{R})$. Now $a_0=b_0\equiv 0$, then

$$\forall n\in\mathbb{N},\ a_n,b_n\in L^\infty_{loc}(\mathbb{R}). \qquad (56)$$

We can apply the Gronwall lemma to (53) by noting $a_n(t)$ and $b_n(t)$ are creasing functions of n and t

$$a_n(t)\leqslant C\varepsilon\ \exp\{C(1+b_{n-1}(t))^2\} \qquad (57)$$

where C is independing on n .
Let A_n, B_n be

$$A_n = \operatorname*{Sup}_{t\in\mathbb{R}} a_n(t)\ ,\ B_n = \operatorname*{Sup}_{t\in\mathbb{R}} b_n(t)\ .$$

(55) and (57) imply

$$A_n\leqslant C\varepsilon\ \exp\{C(1+B_{n-1})^2\} \qquad (58)$$

$$B_n\leqslant C(\varepsilon+A_{n-1}(1+B_{n-1})^2) \qquad (59)$$

and we have

$$A_0 = B_0 = 0\ . \qquad (60)$$

We choose $0<\varepsilon_0$ such that

$$C\varepsilon(1+4C\ \exp\ 4C)\ \leqslant\ 1\ . \qquad (61)$$

Suppose

$$0\leqslant\varepsilon\leqslant\varepsilon_0\ ,\ A_{n-1}\leqslant C\varepsilon\exp 4C\ ,\ B_{n-1}\leqslant 1\ , \qquad (62)$$

(58) and (62) imply

$$A_n\ \leqslant\ C\varepsilon\ \exp\ 4C\ , \qquad (63)$$

and (59) and (62) imply

$$B_n\ \leqslant\ C(\varepsilon+4C\varepsilon\ \exp\ 4C),$$

and thanks to (61)

$$B_n \leqslant 1 \ . \tag{64}$$

We conclude by (60) (62) (63) (64) that

$$\sup_n (A_n + B_n) < +\infty \ . \tag{65}$$

Now, the existence of global solution follows from classical method (see e.g.[2]). At present we prove the asymptotic freedom. We note respectively D(t) and U(t) the propagators associated to the free equations of Dirac and Klein-Gordon

$$D(t) = \exp it\,\mathcal{A} \ , \ \mathcal{A} = \sum_{j=1}^{3} i\gamma^0\gamma^j\partial_j - M\gamma^0$$

$$U(t) = \exp it\,A \ , \ A = -i\begin{pmatrix} 0 \ , & 1 \\ \Delta_x - m^2 \ , & 0 \end{pmatrix}.$$

To obtain $\psi^{\pm}$, $\varphi^{\pm}$ it is sufficient to prove the convergence of $D(-t)\psi(t)$ and $U(-t)(\varphi(t),\partial_t\varphi(t))$ respectively in $(H^{10}(\mathbb{R}^3_x))^4$ and $H^{11}(\mathbb{R}^3_x)\times H^{10}(\mathbb{R}^3_x)$ as t → ±∞ .
We have

$$D(-t)\psi(t) = \psi\Big|_{t=0} + \int_0^t D(-s)f(\varphi(s),\psi(s))\,ds$$

$$U(-t)(\varphi(t),\partial_t\varphi(t)) = (\varphi,\partial_t\varphi)\Big|_{t=0} + \int_0^t U(-s)(0,g(\varphi(s),\psi(s))\,ds \ .$$

The propagators D(t) and U(t) being uniformly bounded on the Sobolev spaces we have to prove only

$$\|f(\varphi(s),\psi(s))\|_{(H^{10}(\mathbb{R}^3_x))^4} \in L^1(\mathbb{R}_s) \tag{66}$$

$$\|g(\varphi(s),\psi(s))\|_{H^{10}(\mathbb{R}^3_x)} \in L^1(\mathbb{R}_s) \ . \tag{67}$$

We deduce from (65) that

$$\sup_t \{\|\psi(t)\|_{10} + \|\varphi(t)\|_{10} + (1+|t|)^{3/2}(|\psi(t)|_5 + |\varphi(t)|_5)\} < +\infty \ .$$

We conclude that the norms in (66) and (67) are $O((1+|t|)^{-3/2})$ and this ends the proof.

BIBLIOGRAPHY

[1] A. Bachelot, *Equipartition de l'énergie pour les systèmes hyperboliques et formes compatibles*. Ann. Inst. Henri Poincaré, Physique Théorique, vol.46, n°1, 1987, P.45-76.

[2] A. Bachelot, *Problème de Cauchy global pour des systèmes de Dirac-Klein-Gordon*. Ann. Inst. Henri Poincaré, Physique Théorique, vol.48, n°4, 1988, p.387-422.

[3] A. Bachelot, *Global existence of large amplitude solutions for nonlinear massless Dirac equation*. To appear in Portugaliae Math.

[4] J. Chadam, R. Glassey, *On certain global solutions of the Cauchy problem for the (classical) coupled Klein-Gordon-Dirac equations in one and three space dimensions*. Arch. Rat. Mech. Anal. 54, 1974, p.223-237.

[5] B. Hanouzet, J.L. Joly, *Applications bilinéaires sur certains sous-espaces de type Sobolev*. C.R. Acad. Sc. Paris, série I, t.294, 1982, p.745-747.

[6] B. Hanouzet, J.L. Joly, *Bilinear maps compatible with a system*. Research Notes in Mathematics, 89, Pitman, 1983, p.208-217.

[7] B. Hanouzet, J.L. Joly, *Applications bilinéaires compatibles avec un système hyperbolique*. Ann. Inst. Henri Poincaré, Analyse non linéaire, vol.4, n°4, 1987, p.357-376.

[8] S. Klainerman, *Uniform decay estimates and the Lorentz invariance of the classical wave equations*. Comm. Pure and Appl. Math. 38, 1985, p.321-332.

[9] S. Klainerman, *The null condition and global existence to nonlinear wave equations*. Lectures in Appl. Math., vol. 23, 1986, p.293-326.

[10] S. Klainerman, *Global existence of small amplitude solutions to nonlinear Klein-Gordon equations in four space time dimensions*. Comm. Pure and Appl. Math. 38, 1985, p.631-641.

[11] G. Lochak, *Wave equation for a magnetic monopole*. Int. J. Theor. Phys. 24, n°10, 1985, p.1019-1050.

[12] C.S. Morawetz, W.A. Strauss, *Decay and scattering of solutions of a nonlinear relativistic wave equation*. Comm. Pure and Appl. Math. 25, 1972, p.1-31.

ANALYSE MICROLOCALE ET SINGULARITÉS NON LINÉAIRES

Jean-Michel Bony
Centre de Mathématiques, École Polytechnique
91128 Palaiseau

Cet exposé a pour but de donner une idée des résultats concernant la propagation et l'interaction des singularités pour les équations aux dérivées partielles non linéaires, obtenus depuis une dizaine d'années par des méthodes d'analyse microlocale. Cette nouvelle branche de l'analyse, qui a permis des développements spectaculaires de la théorie des équations aux dérivées partielles linéaires, s'avère aussi indispensable dans le cas non linéaire pour obtenir des renseignements précis, non seulement sur l'existence ou l'absence de singularités, mais sur leur localisation.

Nous ne traiterons ici que de singularités faibles (*grosso modo* plus régulières que les ondes de choc), renvoyant à l'exposé de G. Métivier pour les interventions de l'analyse microlocale pour des singularités plus fortes. Après avoir rappelé au § 1 les concepts fondamentaux et les résultats relatifs aux équations linéaires, nous décrivons au § 2 la problématique générale. Les §§ 3, 4 et 5 sont consacrés aux résultats, de plus en plus raffinés mais aux hypothèses de plus en plus strictes, sur la propagation non linéaire.

Si nous avons essayé de donner une idée des méthodes utilisées, il est clair qu'il était impossible dans le cadre de cet exposé de donner des preuves. Nous nous sommes contentés de donner des énoncés précis, en renvoyant aux mémoires originaux pour les démonstrations.

1 Équations linéaires et analyse microlocale

1.1 Régularité locale

Nous nous placerons dans l'espace $\mathbf{R}^n$, dont la dernière coordonnée x_n sera parfois notée t, en posant $x' = (x_1, \ldots, x_{n-1})$. Nous ne considèrerons pour simplifier que des opérateurs différentiels strictement hyperboliques

$$P(x,\partial) = \sum_{|\alpha|\leq m} a_\alpha(x)\partial^\alpha \ .$$

où α désigne un multiindice $(\alpha_1, \ldots, \alpha_n) \in \mathbf{N}^n$, et où $|\alpha| = \alpha_1 + \cdots + \alpha_n$ est l'ordre de la dérivation ∂^α. La condition d'hyperbolicité stricte signifie que l'équation en τ

$$p_m(x', t, \xi', \tau) = 0$$

admet m racines réelles distinctes, pour tout $(x', t) \in \mathbf{R}^n$ et $\xi' \in (\mathbf{R}^{n-1} \setminus 0)$. On a noté p_m le symbole principal de l'opérateur

$$p_m(x,\xi) = \sum_{|\alpha|=m} a_\alpha(x)\xi^\alpha \ , \tag{1}$$

où ξ^α est le monôme $\xi_1^{\alpha_1} \cdots \xi_n^{\alpha_n}$.

Pour de tels opérateurs, on dispose de théorèmes d'existence, d'unicité et de régularité du problème de Cauchy, une solutions u appartenant à l'espace de Sobolev H^s lorsque les données de Cauchy $\partial_t^j u(x', 0)$ appartiennent respectivement à H^{s-j} , $j = 0, \ldots, m-1$. Toutefois, de tels résultats d'appartenance à un espace fonctionnel ne fournissent pas de réponse satisfaisante lorsque les données de Cauchy appartiennent à H^s partout mais sont plus régulières en dehors d'un ensemble donné. On souhaiterait alors pouvoir en déduire que la solution u est elle-même plus régulière en dehors d'un certain ensemble et déterminer celui-ci.

Les concepts adaptés à ce type de questions sont les suivants.

Définition 1.1 *On dit que u appartient à H^s* localement au point x_0 , *ce que l'on notera $u \in H^s_{x_0}$, s'il existe un voisinage ω de x_0 tel que, pour toute fonction φ de classe C^∞ à support dans ω, on ait $\varphi u \in H^s$, ce qui s'exprime encore par*

$$\widehat{\varphi u}(\xi)\left(1+|\xi|^2\right)^{s/2} \in L^2(\mathbf{R}^n)\,.$$

Le support singulier d'ordre s *de u est par définition l'ensemble des points x tels que u n'appartienne pas à H^s_x. C'est un sous-ensemble fermé noté* H^s-Supp Sing(u). *Le cas $s = +\infty$ correspond bien sûr à la régularité locale C^∞.*

La connaissance des supports singuliers d'une fonction u donne une description géométrique très satisfaisante de la régularité de u. L'inconvénient est qu'il n'existe pas de bons théorèmes relatifs à ces concepts. La connaissance du support singulier d'une solution dans le passé (ou la connaissance du support singulier des données de Cauchy) ne détermine pas celui-ci dans l'avenir.

Par exemple, si une solution de l'équation des ondes est singulière en un point (x_0', t_0), on peut affirmer qu'elle est singulière le long de l'une au moins des génératrices du cône d'onde issu de ce point, mais pour déterminer laquelle ou lesquelles de ces génératrices portent effectivement des singularités, des informations supplémentaires sont nécessaires.

1.2 Régularité microlocale

L'une des idées principales de l'analyse microlocale est d'associer à u des sous-ensembles de l'*espace des phases* $\mathbf{R}^n\times(\mathbf{R}^n\backslash 0)$, et non plus de $\mathbf{R}^n$, qui donnent une représentation plus fine des singularités de u.

Définition 1.2 *Soient $x_0 \in \mathbf{R}^n$ et $\xi_0 \in \mathbf{R}^n \setminus 0$. On dit que u* appartient à H^s microlocalement en (x_0, ξ_0) , *ce que l'on notera $u \in H^s_{x_0,\xi_0}$ s'il existe un voisinage ω de x_0 et un voisinage* conique Γ *de ξ_0 tels que l'on ait*

$$\widehat{\varphi u}(\xi)\left(1+|\xi|^2\right)^{s/2} \in L^2(\Gamma)\,.$$

On appelle front d'onde d'ordre s *de u (on dit simplement front d'onde lorsque $s = +\infty$) l'ensemble des points (x, ξ) tels que u n'appartienne pas à $H^s_{x,\xi}$. C'est un sous-ensemble fermé et conique en ξ noté $WF_s(u)$.*

L'exemple suivant donne une bonne idée de ce que signifie cette localisation des singularités à la fois dans l'espace ambiant $\mathbf{R}^n_x$ et dans 'l'espace des fréquences' $\mathbf{R}^n_\xi$. Considérons une fonction u qui est de classe C^∞ jusqu'au bord de part et d'autre d'une hypersurface lisse Σ mais qui peut avoir un saut le long de Σ. Le front d'onde de u est alors contenu dans l'ensemble des (x, ξ) tels que x appartienne à Σ et que ξ soit normal à Σ au point x.

La connaissance des régularités microlocales de u entraîne la connaissance des régularités locales. Pour que u appartienne à $H^s_{x_0}$, il faut et il suffit que pour tout $\xi \neq 0$, on ait $u \in H^s_{x_0,\xi}$. En d'autres termes, le support singulier d'ordre s de u est la projection sur $\mathbf{R}^n$ de $WF_s(u)$.

Les théorèmes qui vont suivre assurent que la propagation des singularités se décrit de manière très simple au niveau des régularités microlocales. Au niveau des régularités locales, il s'agit donc de la projection sur $\mathbf{R}^n$ d'un comportement très simple dans $\mathbf{R}^{2n}$, ce qui peut apparaître comme fort complexe et difficilement compréhensible si on reste à ce niveau.

1.3 Propagation des singularités

Le symbole principal p_m de l'opérateur différentiel P défini par (1) est une fonction définie sur l'espace des phases à laquelle sont attachés deux concepts géométriques importants.

On appelle *variété caractéristique* de P l'ensemble $\mathrm{Car}(P)$ des points (x,ξ) vérifiant $p_m(x,\xi)=0$. Un point de l'espace des phases appartenant à $\mathrm{Car}(P)$ est dit *caractéristique.*

On appelle *bicaractéristique* une courbe intégrale du champ hamiltonien de p_m, c'est-à-dire une courbe $s\mapsto(x(s),\xi(s))$ solution du système différentiel

$$\frac{dx_i}{ds}=\frac{\partial p_m}{\partial\xi_i}(x(s),\xi(s)) \quad , \quad \frac{d\xi_i}{ds}=-\frac{\partial p_m}{\partial x_i}(x(s),\xi(s)) \ .$$

Il est facile de voir que p_m reste constant le long d'une bicaractéristique. Si une bicaractéristique contient un point de $\mathrm{Car}(P)$ elle est donc tout entière tracée sur $\mathrm{Car}(P)$. Nous n'aurons à considérer que des bicaractéristiques de ce type, dites *bicaractéristiques nulles.*

Bien entendu, étant donné un point (x_0,ξ_0) de $\mathrm{Car}(P)$, il existe d'après le théorème de Cauchy-Lipschitz une et une seule bicaractéristique nulle passant par (x_0,ξ_0).

Les résultats fondamentaux suivants sont dûs à Hörmander (voir [20]), et n'exigent pas toutes les hypothèses que nous avons faites sur l'opérateur P (il suffit que la partie principale de l'opérateur soit à coefficients réels).

Théorème 1.3 *Soit $s\in\mathbf{R}$ et soit u une solution de l'équation $Pu=f$, où $f\in H^{s-m+1}$.*
(a) *En tout point (x_0,ξ_0) non caractéristique, on a $u\in H^{s+1}_{x_0,\xi_0}$.*
(b) *Si (x_1,ξ_1) est un point caractéristique, et si (x_2,ξ_2) est situé sur la même bicaractéristique, on a l'équivalence suivante*

$$u\in H^s_{x_1,\xi_1}\iff u\in H^s_{x_2,\xi_2} \ .$$

Ce théorème permet de déterminer complètement et simplement la régularité de u en chaque point $x_0=(x'_0,t_0)$ avec $t_0>0$, connaissant la régularité de u dans le passé. Pour prouver que $u\in H^s_{x_0}$, nous avons vu qu'il est équivalent de montrer que, pour tout $\xi\in\mathbf{R}^n\setminus 0$, on a $u\in H^s_{x_0,\xi}$. Si le point (x_0,ξ) est non caractéristique, la partie (a) du théorème nous assure qu'il en est bien ainsi. Pour chaque ξ tel que (x_0,ξ) soit caractéristique, on considère la bicaractéristique issue de ce point, et on choisit un point (x_2,ξ_2) de cette bicaractéristique situé dans le passé, l'hyperbolicité stricte assurant que de tels points existent. De deux choses l'une :

- ou bien, pour chaque ξ, la distribution u appartient microlocalement à H^s au point (x_2,ξ_2) associé. D'après la partie (b) du théorème, on a alors $u\in H^s_{x_0,\xi}$ pour tout ξ et donc $u\in H^s_{x_0}$.
- ou bien il existe un ξ tel que u n'appartienne pas microlocalement à H^s au point (x_2,ξ_2) associé, et il en résulte que u n'appartient pas localement à H^s au point x_0.

2 Équations non linéaires : position du problème

Nous considérerons maintenant une équation non linéaire d'ordre m, dont la forme générale est la suivante

$$F(x,u,\nabla u,\dots,\nabla^m u)=0 \ . \qquad (2)$$

Si u est une solution suffisamment régulière de 2, on peut définir l'opérateur linéarisé le long de u (en notant u_α la variable de F correspondant à $\partial^\alpha u$)

$$\mathcal{L}_u\varphi=\sum_{|\alpha|\le m}\frac{\partial F}{\partial u_\alpha}(x,u,\dots,\nabla^m u)\partial^\alpha\varphi \ .$$

Si u appartient à H^s avec $s>n/2+m$ (cette hypothèse peut être affaiblie lorsque (2) n'est pas totalement non linéaire), $\mathcal{L}_u$ est un opérateur linéaire à coefficients de classe H^{s-m}, et on peut définir son symbole principal

$$l(x,\xi) = \sum_{|\alpha|=m} \frac{\partial f}{\partial u_\alpha}(x,u,\ldots,\nabla^m u)\xi^\alpha .$$

Ce symbole dépend en général non seulement de l'équation, mais de la solution u elle-même. Cela dit, il en est indépendant dans le cas important des équations *semi-linéaires* où la fonction F de (2) est linéaire (à coefficients ne dépendant que de x) par rapport aux dérivées d'ordre maximum de u. Dans tous les cas, on définit $\mathrm{Car}(\mathcal{L}_u)$ et les bicaractéristiques à partir de $l(x,\xi)$ comme au n°1.3.

Pour pouvoir obtenir des résultats semi-globaux, nous supposerons la solution u définie dans un ouvert Ω de $\mathbf{R}^n$ contenant l'origine. Nous poserons $\Omega^\pm = \Omega \bigcap \{\pm t > 0\}$ et nous ferons les hypothèses suivantes :

- L'opérateur linéarisé $\mathcal{L}_u$ est strictement hyperbolique dans Ω. Cette condition dépend, sauf dans le cas semi-linéaire, de la solution u elle-même, mais est une condition ouverte.
- L'ouvert Ω^+ est inclus dans le domaine d'influence de Ω^-. Cela signifie que les bicaractéristiques nulles issues d'un point de $\Omega^+ \times \mathbf{R}^n$ et dirigées vers le passé rencontrent $\Omega^- \times \mathbf{R}^n$ avant de sortir de $\Omega \times \mathbf{R}^n$.
- On connaît une régularité minimale de u dans tout Ω, c'est-à-dire que $u \in H^s(\Omega)$ pour un s donné.
- La régularité de u est connue dans le passé, c'est-à-dire que l'on connait, pour chaque σ, l'ensemble $WF_\sigma(u) \bigcap \{t < 0\}$.

Le problème posé est alors le suivant :
Étant donnés $x_0 \in \Omega^+$ et $\sigma > s$, peut on déterminer si u appartient ou non à H^σ au point x_0?

Nous ne pourrons donner de résultats significatifs que pour des solutions suffisamment régulières, c'est-à-dire pour s supérieur à un indice critique s_0 qui dépend de l'équation. La valeur $s_0 = n/2 + m + 1$ convient toujours, et cette valeur s'abaisse pour des équations quasi- ou semi-linéaires, mais la régularité imposée exige toujours la continuité des termes non linéaires apparaissant dans (2) et exclut donc l'apparition de chocs.

La solution du problème posé ci-dessus dépend fortement des valeurs relatives de s et σ. Le §3 étudie le cas $\sigma \leq 2s - s_0$, pour lequel on obtient une réponse complète : les singularités se propagent comme dans le cas linéaire, et on ne perçoit pas d'interaction.

Le §4 étudie le cas $\sigma \leq 3s - s_1$. On obtient une très bonne réponse au problème posé. Les singularités ne peuvent parvenir au point x_0 que par propagation directe, ou par une seule interaction de singularités propagées.

Le §5 étudie le cas général, σ étant éventuellement égal à $+\infty$. Il faut alors prendre en compte la possibilité d'un grand nombre d'interactions. A l'exception du cas de la dimension 1 d'espace, on n'obtient de résultats intéressants qu'en supposant les singularités de type 'conormal' dans le passé. Sous cette hypothèse, on dispose de résultats positifs dans un certain nombre de cas relativement simples, ainsi que de résultats très généraux lorsque la géométrie des singularités incidentes est analytique.

3 Calcul paradifférentiel et singularités 'jusqu'à $2s$'

Le calcul paradifférentiel est un calcul symbolique, analogue au calcul pseudo-différentiel, mais où les symboles ont une régularité limitée. Il est particulièrement adapté à l'étude des équations linéaires à coefficients peu réguliers et à celle des équations non linéaires. Plutôt que d'entrer dans les détails techniques, pour lesquels nous renvoyons à [6] (et à [9] pour une présentation plus élémentaire), nous nous bornerons à en décrire quelques conséquences.

Le 'principe' suivant n'est bien sûr pas un théorème, mais il décrit bien à la fois l'efficacité et les limitations de ce calcul.

Principe 3.1 *Si u est une solution de (5.5) appartenant à H^s dans Ω, avec $s > s_0$, on peut*
- *faire comme si u était solution de son équation linéarisée*

$$\mathcal{L}_u u = f \tag{3}$$

avec un second membre f mal connu mais appartenant à H^{2s-s_0-m+1},
- *faire comme si l'opérateur $\mathcal{L}_u$ était à coefficients C^∞.*

'En général', les conclusions que l'on déduira par des 'procédés raisonnables' de ces premisses fausses seront exactes.

Si le lecteur veut bien nous croire, il constatera que si on perturbe u par une fonction de classe H^{2s-s_0+1}, on obtient une solution d'une équation (3) dont le second membre a changé mais a la même régularité. On ne peut donc espérer, en ce qui concerne la régularité de u que des résultats modulo H^{2s-s_0+1}. D'autre part, si on voulait appliquer ce principe pour $s \leq s_0$, le fait que le second membre appartienne à H^{2s-s_0-m+1} ne permettrait de controler la propagation des singularités que jusqu'à H^σ avec $\sigma < s$ (théorème 1.3), ce qui est sans intérêt puisque l'on a supposé $u \in H^s(\Omega)$.

La raison de la validité de ce principe est la suivante. On peut démontrer que u est solution d'une équation

$$\widetilde{\mathcal{L}_u} u = f \tag{4}$$

où f appartient à H^{2s-s_0-m+1}, et où $\widetilde{\mathcal{L}_u}$ est un opérateur ***paradifférentiel*** dont le symbole principal est le même que celui de $\mathcal{L}_u$. Et le calcul paradifférentiel est suffisamment proche du calcul pseudo-différentiel usuel (que l'on pourrait appliquer à $\mathcal{L}_u$ si ses coefficients étaient C^∞) pour que des démonstrations fondées sur du calcul symbolique (inversion, conjugaison, ...) ou des estimations d'énergie, ..., se laissent transposer (plus ou moins facilement) à la vraie équation (4).

Par exemple, le théorème suivant serait une 'conséquence' immédiate du principe précédent et du théorème 1.3. Nous renvoyons à [6] pour une véritable démonstration.

Théorème 3.2 *Soit u une solution de (2) appartenant à $H^s(\Omega)$, avec $s \geq s_0$. Pour tout $\sigma \leq (2s - s_0)$, on a alors*
(a) *En tout point (x_0, ξ_0) non caractéristique, on a $u \in H^{\sigma+1}_{x_0,\xi_0}$.*
(b) *Si (x_1, ξ_1) est un point caractéristique, et si (x_2, ξ_2) est situé sur la même bicaractéristique, on a l'équivalence suivante*

$$u \in H^\sigma_{x_1,\xi_1} \iff u \in H^\sigma_{x_2,\xi_2} \ .$$

On dispose maintenant de théorèmes analogues concernant la réflexion ou la réfraction des singularités [29] ainsi que sur la diffraction. Les singularités d'ordre σ se comportent comme pour les équations linéaires, à condition de se limiter à $\sigma \leq 2s - s_0$.

Remarque **3.3** Pour les équations non linéaires elliptiques, le calcul paradifférentiel ne redonne que des résultats connus. Par contre, il permet d'obtenir des théorèmes de régularité dans des situations où le linéarisé est hypoelliptique. Nous renvoyons à [31] [21] et ne décrivons que le résultat ci-dessous, dû à Xu Chao Jiang [30].

Soit J une fonctionnelle du premier ordre dans $\Omega \subset \mathbf{R}^n$

$$J(u) = \int_\Omega G\left(x, u(x), \nabla u(x)\right) dx \ ,$$

où F est une fonction C^∞ de ses arguments. On suppose que u_0 est une fonction réalisant un minimum local 'suffisamment strict' de J au sens suivant :

$$\forall \varphi \in C_0^\infty(\Omega) \qquad J(u+\varphi) \geq J(u) + C^{\text{te}} \left\|\varphi\right\|^2_{H^\sigma} \ ,$$

avec $\sigma > 0$.

La conclusion est : si u_0 est de classe C^3 alors u est de classe C^∞.

L'idée est que le caractère strict du minimum va entraîner une inégalité

$$\|\varphi\|^2_{H^\sigma} \leq C^{te}(|(\mathcal{L}_{u_0}\varphi \mid \varphi)| + \|\varphi\|^2_{L^2})$$

pour le linéarisé $\mathcal{L}_{u_0}$ de l'équation d'Euler-Lagrange et que cette inégalité va se transférer à l'opérateur paradifférentiel $\widetilde{\mathcal{L}_{u_0}}$ associé à cette même équation. Il est ensuite possible de transposer à l'équation (4) des arguments classiques sur l'hypoellipticité des opérateurs linéaires.

4 Première interaction et singularités 'jusqu'à $3s$'

Il s'agit de résultats dûs à M. Beals [4] [5] pour des équations semi-linéaires du second ordre et à J.-Y. Chemin [16] dans le cas général. Nous décrivons ici une forme simplifiée du théorème principal.

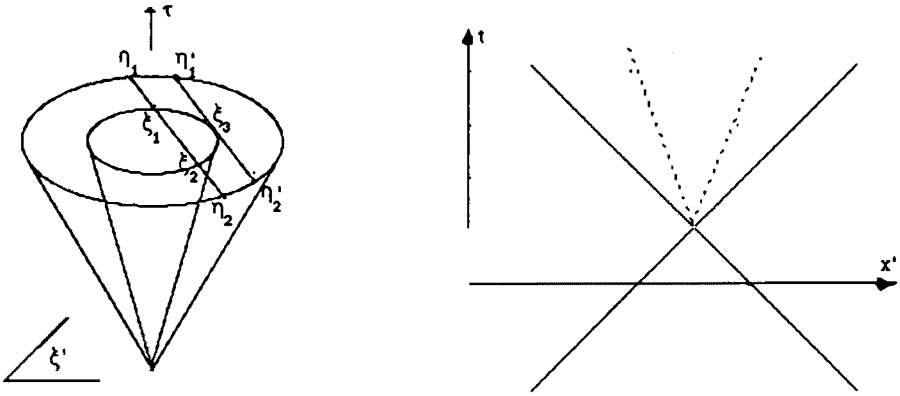


Figure 1: Pour une équation d'ondes non linéaires à deux vitesses $(\partial_t^2 - c_1^2\Delta)(\partial_t^2 - c_2^2\Delta) = f(u)$ deux bicaractéristiques de F_1, porteuses de singularités 'd'ordre s ', peuvent se croiser en arrivant en (x, ξ_1) et (x, ξ_2). Le point (x, η_1) appartient alors à G_1 et la bicaractéristique issue de ce point, incluse dans F_2 est en général porteuse d'une singularité 'd'ordre $2s$'

On considère une solution de (2) qui appartient à $H^s(\Omega)$ pour un $s > s_0$. A partir de l'ensemble $\mathrm{WF}^- = (\Omega^- \times \mathbf{R}^n) \cap \mathrm{WF}\, u$ qui décrit les singularités de u dans le passé, on définit les ensembles suivants.

$$F_1 = \left\{(x,\xi) \in \mathrm{Car}(\mathcal{L}_u) \;\middle|\; \begin{array}{l}\text{il existe une bicaractéristique orientée vers}\\ \text{l'avenir joignant un point de } \mathrm{WF}^- \text{ à } (x,\xi)\end{array}\right\}.$$

L'ensemble G_1 est la 'somme fibre à fibre' de F_1 et de lui-même, en y incluant le cas limite décrit par $\widetilde{G_1}$

$$G_1 = \left\{(x,\xi) \;\middle|\; \begin{array}{l}\text{il existe } \xi_1 \text{ et } \xi_2 \text{ avec } \xi = \xi_1 + \xi_2 \text{ tels que}\\ (x,\xi_1) \text{ et } (x,\xi_2) \text{ appartiennent à } F_1\end{array}\right\} \bigcup \widetilde{G_1}$$

où $\widetilde{G_1}$ est défini par

$$\widetilde{G_1} = \left\{(x,\xi) \;\middle|\; \begin{array}{l}\text{il existe } \xi_3 \text{ tel que } (x,\xi_3) \text{ appartienne à } F_1 \text{ et}\\ \text{que } (\xi - \xi_3) \text{ soit tangent à } \mathrm{Car}(\mathcal{L}_u)_x \text{ en } \xi_3\end{array}\right\}.$$

On a noté $\mathrm{Car}(\mathcal{L}_u)_x$ l'ensemble des ξ tels que $(x,\xi) \in \mathrm{Car}(\mathcal{L}_u)$.

La raison de l'introduction de G_1 provient du fait suivant. Si (x,ξ_1) et (x,ξ_2) appartiennent respectivement aux fronts d'onde de v et w, on voit facilement (en regardant la décroissance du produit de convolution des transformées de Fourier) que $(x,\xi_1+\xi_2)$ appartiendra en général au front d'onde du produit vw.

Si on pense que, par propagation comme dans le cas linéaire, les points de F_1 ont toutes les chances d'appartenir à WF u, les points de G_1 ont alors toutes les chances d'appartenir au front d'onde des termes non-linéaires de l'équation (2), et donc à WF u lui-même. Ces nouvelles singularités peuvent se propager le long des bicaractéristiques nulles, ce qui justifie l'introduction de l'ensemble suivant.

$$F_2 = \left\{(x,\xi) \in \mathrm{Car}(\mathcal{L}_u) \;\middle|\; \begin{array}{l}\text{il existe une bicaractéristique orientée vers l'avenir}\\ \text{joignant un point de } G_1 \cap \mathrm{Car}(\mathcal{L}_u) \text{ à } (x,\xi)\end{array}\right\} .$$

Le théorème général est le suivant, où on devra prendre $s_1 = n+2m+2$ dans le cas général, mais où cette valeur s'abaisse pour des équations moins violemment non-linéaires.

Théorème 4.1 *Avec les notations précédentes, on a*
(a) *La fonction u appartient microlocalement à H^{2s-s_0} en dehors de F_1.*
(b) *La fonction u appartient microlocalement à H^{3s-s_1} en dehors de $F_1 \cup G_1 \cup F_2$.*

La partie (a) n'est qu'une reformulation du théorème 3.2, et la partie (b) affirme que 'jusqu'à $3s$' on ne voit géométriquement que la propagation caractéristique, assortie d'une interaction au plus.

Remarque **4.2** Le théorème de Chemin [16] est plus précis : pour $\sigma \in [s, 3s-s_1]$, à partir de la collection des ensembles $\mathrm{WF}_\sigma(u) \cap \{t<0\}$, on construit des ensembles $F_{1,\sigma}$, $G_{1,\sigma}$ et $F_{2,\sigma}$ en.dehors desquels la solution appartiendra microlocalement à H^σ.

5 Singularités conormales et contrôle de l'interaction 'jusqu'à C^∞'

Le théorème 4.1 pourrait laisser espérer la construction d'une suite infinie $F_1, G_1, F_2, G_2, \ldots$ de sous-ensembles de l'espace des phases, exprimables à partir des singularités de u dans le passé, et prenant en compte la possibilité de $1, 2, \ldots$ interaction successives, tels que u soit *grosso modo* microlocalement de classe H^{ks} en dehors de $F_1 \cup G_1 \cup \ldots \cup F_k$. Malheureusement, il n'est pas possible d'obtenir de tels résultats en général, comme le montrent des contre-exemples de M. Beals [4] et de Chemin [16].

Par exemple (voir [4]), pour une équation $\Box u + \beta u^3 = 0$, où β est une fonction de classe C^∞, il existe des solutions $u \in H^s$ dont les données de Cauchy sont C^∞ hors de l'origine, alors que u est singulière non seulement à la surface du cône d'onde, mais aussi à l'intérieur de ce cône (où elle n'est *grosso modo* que de classe H^{3s}). Un tel contre-exemple semble complètement ruiner l'idée de décrire la propagation des singularités 'au-delà de $3s$' à partir de la propagation sur les bicaractéristique, et de l'interaction.

Toutefois, en faisant sur les singularités de u dans le passé des hypothèses un peu plus fortes que la seule localisation du front d'onde (singularités conormales), nous allons voir qu'il est possible d'obtenir des théorèmes de propagation des singularités jusqu'à C^∞ conformes aux espérances ci-dessus, au moins dans les cas où la géométrie est soit relativement simple, soit (sous-)analytique.

Il faut mettre à part le cas de la dimension 1 d'espace, où les résultats de Rauch et Reed [26] dans le cas semi-linéaire, et de Chemin [15] fournissent une réponse très complète. A partir de la connaissance d'une fonction $\rho(x)$ telle que les données de Cauchy de u appartiennent à $H^{\rho(x)}$ au voisinage de x, on peut déterminer (par propagation caractéristique et interactions successives) une fonction $\sigma(x,t)$ telle que la solution u appartienne à $H^{\sigma(x,t)}$ au voisinage de (x,t).

5.1 Distributions conormales

Nous ne donnerons une définition précise que dans le cas géométriquement le plus simple.

Définition 5.1 *Désignons par Σ soit une sous-variété lisse, soit la réunion de deux hypersurfaces lisses Σ_1 et Σ_2 se coupant transversalement. On dit que u appartient à l'espace $H^{s,k}_{\Sigma}$ si $u \in H^s$ et si on a pour $l \leq k$*

$$Z_1 \circ Z_2 \circ \cdots \circ Z_l \, u \in H^s \tag{5}$$

quels que soient les champs de vecteurs Z_j (identifiés à des opérateurs différentiels du premier ordre) tangents à Σ.

Les exemples types d'éléments de $H^{s,\infty}_{\Sigma}$ sont les fonctions homogènes de la distance à Σ dans le premier cas, et les produits de telles fonctions relatives à Σ_1 et Σ_2 dans le second.

Si $u \in H^{s,k}_{\Sigma}$, il est facile de voir que u appartient localement à H^{s+k} en dehors de Σ. De plus, dans le cas où Σ est lisse, $\mathrm{WF}_{s+k}\, u$ est contenu dans le conormal de Σ (ensemble des (x, ξ) avec $x \in \Sigma$ et ξ orthogonal à Σ en x). Dans le cas où $\Sigma = \Sigma_1 \cup \Sigma_2$, l'ensemble $\mathrm{WF}\, u$ est contenu dans la réunion des conormaux à Σ_1, à Σ_2 et à $\Sigma_1 \cap \Sigma_2$.

L'appartenance à $H^{s,k}_{\Sigma}$ est toutefois une propriété strictement plus forte que les inclusions du front d'onde ci-dessus. Par exemple, les distributions vérifiant ces inclusions ne forment pas une algèbre dès que k dépasse $s - n/2$, alors que l'on a le résultat suivant.

Théorème 5.2 *Pour $s > n/2$ et $k = 0, 1, \ldots, \infty$ les espaces $H^{s,k}_{\Sigma}$ forment une algèbre pour la multiplication, et sont stable par $u \mapsto f \circ u$ lorsque $f \in C^\infty$.*

Il s'agit d'une conséquence simple des formules de dérivation d'un produit et d'une fonction composée, et du fait que H^s est une algèbre pour $s > n/2$.

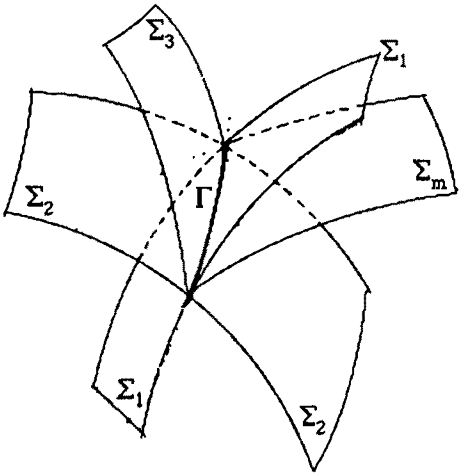
5.2 Quelques cas géométriquement simples

Interaction de deux ondes Le résultat suivant est dû à J.-M. Bony [8] pour les équations semi-linéaires, et à S. Alinhac [3] dans le cas général. On considère une solution u de l'équation (2) appartenant à $H^s(\Omega)$, $s > n/2 + m + 4$. On considère une sous variété Γ de codimension 2 et de classe C^1, ne rencontrant pas Ω^- et de type espace au sens suivant : par Γ passent exactement m (l'ordre de l'opérateur) hypersurfaces caractéristiques $\Sigma_1, \ldots, \Sigma_m$ de classe C^1 se coupant transversalement sur Γ.

Pour $t < 0$, on suppose que les hypersurfaces Σ_1 et Σ_2 sont de classe C^∞, que la fonction u est C^∞ hors de $\Sigma_1 \cup \Sigma_2$ et que près de Σ_i on a $u \in H^{s,\infty}_{\Sigma_i}$ pour $i = 1, 2$.

Théorème 5.3 *Sous les hypothèses précédentes, on a*
(a) *La fonction u est de classe C^∞ en dehors de Σ_1, de Σ_2 et des demi-hypersurfaces Σ_j^+ , $j = 3, \ldots, m$ limitées par Γ et tournées vers l'avenir.*
(b) *La sous-variété Γ est de classe C^∞, et il en est de même de Σ_1, de Σ_2 et des Σ_j^+ en dehors de Γ.*
(c) *Localement près de $\Sigma_i \setminus \Gamma$, $i = 1, 2$, on a $u \in H^{s,\infty}_{\Sigma_i}$*
(d) *Localement près de $\Sigma_j^+ \setminus \Gamma$, $j = 3, \ldots, m$, on a $u \in H^{2s-s_0,\infty}_{\Sigma_j}$*

On a représenté ci-contre ce phénomène d'interaction pour une équation d'ordre plus grand que 2. Pour une équation du second ordre, aucun phénomène d'interaction n'apparaît géométriquement. En effet, les seules hypersurfaces caractéristiques issues de Γ sont Σ_1 et Σ_2 et, sous les hypothèses du théorème la solution est régulière dans l'avenir comme dans le passé en dehors de ces hypersurfaces.
Pour une équation d'ordre quelconque, si u n'est singulière dans le passé que sur une hypersurface caractéristique Σ_1 (où elle a une singularité conormale), alors il en est de même dans l'avenir où elle appartient exactement au même espace $H^{\sigma,k}_{\Sigma_1}$ (voir [7] [1] [2] [28]).



Interaction de trois ondes On se limite ici à une équation des ondes non linéaires en dimension 2 d'espace

$$\Box u = \partial_t^2 u - \partial_x^2 u - \partial_y^2 u = f(t, x, y, u) \tag{6}$$

où on suppose que la solution u appartient à H^s, $s > 3/2$. On se donne trois surfaces caractéristiques Σ_1, Σ_2 et Σ_3 se coupant transversalement en un point N situé dans Ω^+. On note Γ^+ le demi-cône d'onde d'avenir issu de N. On a alors le résultat suivant [10] [11] (un résultat très voisin a été démontré indépendamment par Melrose-Ritter [25], d'autre part J.-Y. Chemin [17] a étendu ce résultat au cas où le membre de droite de (6) dépend aussi du gradient de u).

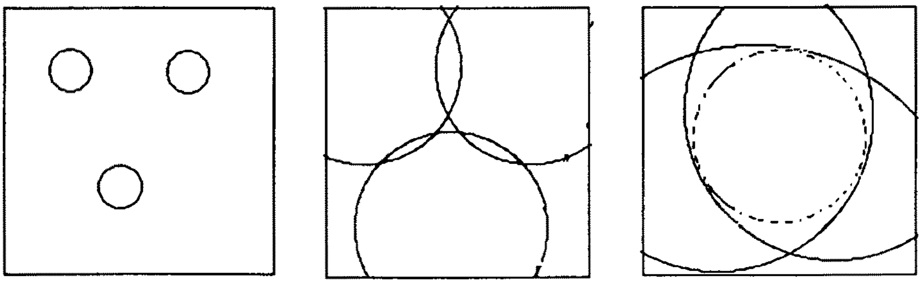


Figure 2: Interaction de trois ondes circulaires

Théorème 5.4 *Supposons que dans Ω^-, la solution u appartienne à $H^{\sigma+k}$, avec $\sigma > 3/2$, hors des Σ_i, et appartienne à $H^{\sigma,k}_{\Sigma_i \cup \Sigma_j}$ près de $\Sigma_i \cup \Sigma_j$. On a alors pour tout $\sigma' < \sigma$*
(a) *$u \in H^{\sigma'+k}$ hors de $\cup_j \Sigma_j \cup \Gamma^+$*
(b) *$u \in H^{\sigma',k}_{\Sigma_j}$ près de $\Sigma_j \setminus (\cup_{i \neq j} \Sigma_i \cup \Gamma^+)$*
(c) *$u \in H^{\tau,l}_{\Gamma^+}$ près de $\Gamma^+ \setminus (\cup \Sigma_j)$ si τ et l vérifient $\tau < \sigma + k$, $\tau < 3\sigma - 3$ et $\tau + l < \sigma + k$.*

On a représenté le film montrant ce phénomène d'interaction. Des exemples de Rauch et Reed [27] montrent que la création d'une nouvelle singularité sur Γ se produit effectivement.

Problème de Cauchy La figure 3 illustre trois cas où des résultats de propagation des singularités non linéaires à partir des données de Cauchy sont connus. Les trois énoncés sont du même type : on suppose que les données de Cauchy appartiennent à des espaces $H^{s,\infty}$ (les énoncés existent pour $H^{s,k}$) par rapport à un sous-ensemble donné dans l'hyperplan $t = 0$, et on en déduit l'existence d'un sous-ensemble de l'espace-temps en dehors duquel la solution est de classe C^∞ et au voisinage duquel elle est de classe $H^{s,k}$. Nous renvoyons aux articles cités ci-dessous pour les énoncés précis.

Données de Cauchy singulières sur une hypersurface Δ [12] [3] Les solutions sont régulières en dehors des m hypersurfaces caractéristiques s'appuyant sur Δ.

Données de Cauchy singulières sur p courbes de $\mathbf{R}^2$ *se coupant 2 à 2 transversalement en un point* [12] Les solutions de l'équation des ondes non linéaires en dimension 2 d'espace sont alors régulières en dehors des $2p$ surfaces caractéristiques issues des courbes et du cône d'onde issu de leur point d'intersection.

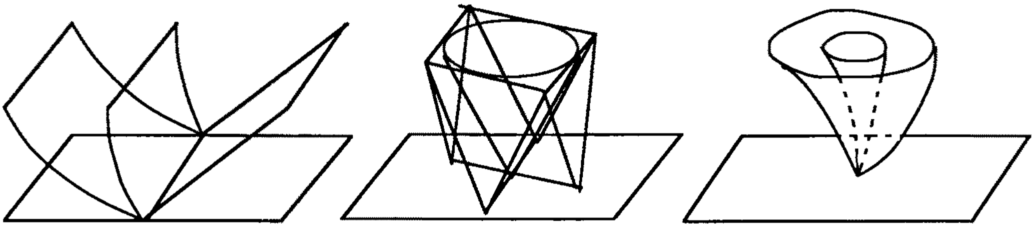


Figure 3: Problème de Cauchy

Données de Cauchy singulières en un point [18] [19] On considère une équation totalement non linéaire (2), et une solution u dont les données de Cauchy sont conormales par rapport à l'origine, telles que l'équation à coefficients constants obtenue en gelant à l'origine les coefficients de l'équation linéarisée ait son cône d'onde lisse. Les bicaractéristiques nulles issues de l'origine engendrent une hypersurface Γ, tangente à ce cône à l'origine et C^∞ hors de l'origine. La solution u est régulière en dehors de Γ.

5.3 Démonstrations et microlocalisations d'ordre supérieur

L'idée de départ est très simple, mais ne fonctionne telle quelle que dans des cas eux aussi très simples. On considère l'espace des champs de vecteurs tangents aux hypersurfaces qui doivent porter les singularités, et un système fini de générateurs Z_i de cet espace. L'appartenance de u à $H^{s,k}$ dans le passé nous assure que les

$$Z^I u \overset{\text{def}}{=} Z_{i_1} \circ Z_{i_2} \circ \cdots \circ Z_{i_l} u$$

appartiennent à H^s pour $t < 0$ lorsque l'entier l (la longueur $|I|$ du multiindice I) est inférieur à k, et il suffit de démontrer qu'il en est de même dans l'avenir.

On s'efforce donc, en dérivant l'équation selon les Z_i, d'obtenir un système d'équations portant sur la famille $\left\{ Z^I u \;\middle|\; |I| \leq k \right\}$, à laquelle on pourra appliquer un théorème de propagation de la régularité hyperbolique. On s'aperçoit très vite que l'on pourra obtenir ainsi un système clos d'équations si de bonnes relations de commutation existent entre l'équation et les Z_i (essentiellement, les commutateurs doivent pouvoir s'exprimer à partir des Z_i et de l'équation elle-même).

La méthode fonctionne comme décrit ci-dessus dans le cas des équations semi-linéaires, et pour des solutions singulières dans le passé le long d'une hypersurface, ou bien de deux hypersurfaces si l'ordre est égal à 2. Pour le théorème 5.3, dès que l'ordre est > 2, et même dans le cas semi-linéaire, on doit remplacer les Z_i par des opérateurs pseudo-différentiels P_i, ce qui introduit une difficulté nouvelle : s'il existe une formule bien connue pour exprimer $Z_i(f \circ u)$, le mélange d'opérateurs pseudo-différentiels et d'opérations non linéaires est plus délicat.

Pour les équations non semi-linéaires, la situation est encore plus complexe : dans le théorème 5.3, ni les Σ_i ni les Z_i (ou les opérateurs qui les remplacent) ne sont C^∞ près de Γ, et il faut dans une même récurrence prouver une régularité conormale limitée des Z_i, définir les distributions conormales associées et prouver la régularité de la solution.

Lorsque la situation géométrique se complique, les opérateurs pseudo-différentiels ne suffisent plus (c'est-à-dire que l'on ne peut plus trouver de système d'opérateurs pseudo-différentiels tels que les relations de commutation escomptées soient vérifiées). Il faut faire appel à des opérateurs d'un type nouveau, où on dispose d'un calcul symbolique, mais où on autorise les symboles à être singuliers le long de certaines sous-variétés.

Nous avons été ainsi amenés à introduire dans ce contexte des opérateurs *2-microdifférentiels*, à symboles singuliers près du point N pour démontrer le théorème 5.4. Assez curieusement, alors que l'analyse microlocale usuelle suffit pour étudier la propagation des singularités dans le cas linéaire (au moins dans le cas strictement hyperbolique), les équations non-linéaires semblent exiger une analyse beaucoup plus raffinée (voir [11] [12] [25] [24]).

Pour pouvoir traiter des situations plus complexes, nous avons introduit, avec N. Lerner, les microlocalisations d'ordre supérieur [13] [14]. Il devient alors possible de définir les espaces de distributions conormales par rapport à des hypersurfaces ayant des singularités d'ordre arbitrairement élevé.

5.4 Cas des singularités localisées sur des variétés analytiques

Il s'agit de résultats de G. Lebeau relatifs au problème de Cauchy pour l'équation des ondes non-linéaire

$$\begin{aligned} \Box u &= f(u) && (7)\\ u_{|t=0} &= u_0 && (8)\\ \partial_t u_{|t=0} &= u_1 && (9) \end{aligned}$$

On suppose que la solution u appartient à H^s pour $s > 2$, et que les u_i sont conormales relativement à une hypersurface *analytique réelle* ou que, plus généralement, les u_i sont des distributions intégrales de Fourier relatives à une sous-variété lagrangienne analytique lisse de l'espace cotangent à l'hyperplan $t = 0$. On a alors le résultat suivant [22] [23].

Théorème 5.5 (a) *Pour tout $\sigma \in \mathbf{R}$, il existe un ensemble sous-analytique homogène lagrangien L_σ de l'espace cotangent tel que $\mathrm{WF}_\sigma\, u \subset L_\sigma$. En conséquence, pour tout entier k, la solution u est de classe C^k dans un ouvert dense de Ω.*
(b) *L'ensemble L_σ peut être déterminé de manière explicite à partir de la variété lagrangienne associée aux données de Cauchy de u.*

Les ensembles L_σ se calculent par une récurrence, un peu analogue à une poursuite de la construction des ensembles F_1, G_1, F_2, $\cdots$ au § 4. Il faut en fait considérer des ensembles de suites de z_n, ζ_n complexes tendant vers des x, ξ réels, faire intervenir les stabilités par propagation caractéristique, par somme fibre à fibre, et par des procédures limites.

Un cas particulier particulièrement frappant est celui du pincement d'une onde simple, en dimension 2, dans le cas où les données de Cauchy sont conormales par rapport à (par exemple)

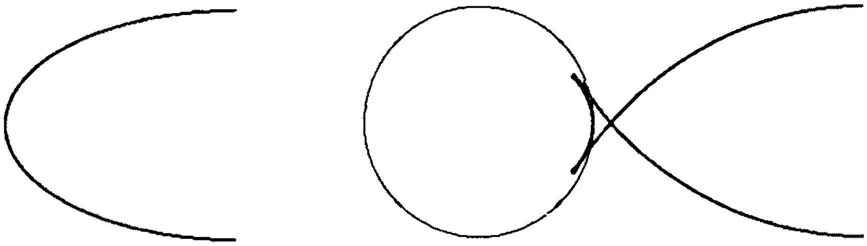


Figure 4: Caustique non linéaire

une parabole. Dans le cas linéaire, les singularités se propageraient sur une queue d'aronde de l'espace-temps (apparition de caustiques), dont on a dessiné une section à t fixé en trait gras. Il s'y ajoute dans le cas non-linéaire un cône d'onde issu du point de naisance de la singularités, dont la section est le cercle en traits fins.

La solution u est de classe C^∞ en dehors des surfaces précédentes. Ce résultat est fourni par le calcul explicite des L_σ, qui restent fixes à partir d'une certaine valeur, et se projettent sur ces surfaces.

Le résultat analogue lorsque les données de Cauchy sont conormales relativement à une courbe de classe C^∞ est toujours un problème ouvert.

Bibliographie

[1] S. Alinhac. Paracomposition et opérateurs paradifférentiels *Comm. in P.D.E.* **11-4** (1986) 87–121

[2] S. Alinhac. Évolution d'une onde simple pour des équations non-linéaires générales *Current Topics in P.D.E.* Kinokuniya Tokyo (1986) 63–90

[3] S. Alinhac. Interaction d'ondes simples pour des équations complètement non-linéaires *Ann. Sci. Éc. Norm. Sup. (4^e série)* **21** (1988) 91–132

[4] M. Beals. Self-spreading and strength of singularities for solutions to semi-linear wave equations *Ann. of Math.* **118** (1983) 187–214

[5] M. Beals. Propagation of smoothness for nonlinear second order strictly hyperbolic equations *Proc. of Symposia in Pure Math.* **43** (1985) 21–45

[6] J.-M. Bony. Calcul symbolique et propagation des singularités pour les équations aux dérivées partielles non linéaires. *Ann. Sci. Ec. Norm. Sup. (4ème série) 14* (1981) 209–246

[7] J.-M. Bony. Propagation des singularités ... *Sém. Goulaouic-Schwartz Ec. Polytechnique* (1979–80) n° 22

[8] J.-M. Bony. Interaction des singularités ... *Sém. Goulaouic-Schwartz Ec. Polytechnique* (1981–82) n° 2

[9] J.-M. Bony. Propagation et interaction des singularités *Proc. Int. Cong. Math. , Varsovie* (1983) 1133–1146

[10] J.-M. Bony. Interaction des singularités pour les équations de Klein-Gordon non linéaires. *Sém. Goulaouic-Meyer-Schwartz Ec. Polytechnique* (1983–84) n° 10

[11] J.-M. Bony. Second microlocalization and interaction of singularities for non linear P.D.E. *Hyperbolic Eq. and related topics, Mizohata ed., Kinokuniya* (1986) 11–49

[12] J.-M. Bony. Singularités des solutions de problèmes hyperboliques non linéaires. *Advances in Microlocal Analysis, Garnir ed., NATO ASI Series 168, Reidel* (1985) 15–39

[13] J.-M. Bony et N. Lerner. Quantification asymptotique ... *Sém. E.D.P. Ec. Polytechnique* (1986–87) n° 2 et 3

[14] J.-M. Bony et N. Lerner. Quantification asymptotique et microlocalisations d'ordre supérieur *Ann. Sci. Ec. Norm. Sup.* (à paraître)

[15] J.-Y. Chemin. Calcul paradifférentiel précisé et applications aux équations aux dérivées partielles non semilinéaires *Duke Math. Journ.* **56-3** (1988) 431–469

[16] J.-Y. Chemin. Interaction contrôlée dans les E.D.P. non linéaires strictement hyperboliques *Bull. Soc. Math. France* (1988)

[17] J.-Y. Chemin. Interaction de trois ondes dans les équations semilinéaires strictement hyperboliques d'ordre 2 (*à paraître*)

[18] J.-Y. Chemin. Régularité de la solution d'un problème de Cauchy fortement non linéaire à données singulières en un point *Ann. Inst. Fourier* **39** (1989)

[19] J.-Y. Chemin. Évolution d'une singularité ponctuelle dans des équations strictement hyperboliques non linéaires (*à paraître*)

[20] L. Hörmander. The Analysis of Linear Partial Differential Operators IV *Springer-Verlag* 1985

[21] J. Hong et C. Zuily. Existence of C^∞ local solutions for the Monge-Ampère equation *Invent. Math.* **89** (1987) 645–661

[22] G. Lebeau. Problème de Cauchy semi-linéaire en 3 dimensions d'espace. Un résultat de finitude *Journ. Funct. Anal.* **77** (1988)

[23] G. Lebeau. Équations des ondes semi-linéaires II. Contrôle des singularités et caustiques non linéaires (prépublication Université Paris-Sud)

[24] R. Melrose. Semi-linear waves with cusp singularities *Actes Journées E.D.P. S^t Jean de Monts* (1987) n° 10

[25] R. Melrose et N. Ritter. Interaction of progressing waves for semi-linear wave equation *Ann. of Math.* **121** (1985)

[26] J. Rauch et M. Reed. Nonlinear microlocal analysis of semi-linear hyperbolic systems in one space dimension *Duke Math. Journ.* **49** (1982) 397–475

[27] J. Rauch et M. Reed. Singularities produced by the nonlinear interaction of three progressing waves; examples *Comm in P.D.E.* **7** (1982) 1117–1133

[28] J. Rauch et M. Reed. Classical, conormal, semilinear waves *Séminaire E.D.P. Ec. Polytechnique* (1985) n°5

[29] M. Sablé-Tougeron. Régularité microlocale pour des problèmes aux limites non linéaires *Ann. Inst. Fourier* **36-1** (1986) 39–82

[30] C. J. Xu. *Thèse Université Paris 11, Orsay* 1984 et *C. R. Acad.Sc. Paris* (1985) 267–270

[31] C. J. Xu. Régularité des solutions d'équations aux dérivées partielles associées un système de champs de vecteurs *Ann. Inst. Fourier* **37** (1987) 105–113

The nonlinear stability of the Minkowski Metric in General Relativity

D. Christodoulou and S. Klainerman

The aim of this report is to provide a very short summary of our work on the nonlinear gravitational stability of the Minkowski space-time. This work, which is still in progress, accomplishes the following goals:

1. It provides a constructive proof of global, smooth, nontrivial, solutions to the Einstein Vacuum Equations which look , in the large, like the Minkowski space. In particular, these solutions are free of black holes and singularities.
2. It provides a detailed description of the sense in which these solutions are closed to the Minkowski space, in all directions, and gives a rigorous derivation of the laws of gravitational radiation proposed by Bondi.
3. It obtains these solutions as dynamic developments of all initial data sets, which are close, in a precise manner, to the initial data set of the Minkowski space , and thus establishes the global dynamic stability of the latter.
4. Though our results are established only for developments of initial data sets which are uniformly close to the trivial one, they should in fact be valid in the complement of the domain of influence of a sufficiently large compact subset of the initial manifold of any "strongly asymptotically flat" initial data set.

According to Einstein the underlying geometry of space-time is that given by a pair $(\mathbf{M},\mathbf{g})$ where $\mathbf{M}$ is a 3+1 dimensional manifold and $\mathbf{g}$ is an Einstein metric on $\mathbf{M}$, that is, a smooth, nondegenerate, 2-covariant tensor field with the property that at each point one can choose 3+1 vectors e_0, e_1, e_2, e_3 such that $\mathbf{g}(e_\alpha, e_\beta) = \eta_{\alpha\beta}$; $\alpha, \beta = 0, 1, 2, 3$ where η is the diagonal matrix with entries -1,1,1,1. The Einstein metric divides the nonzero vectors X in the tangent space at each point into time-like, null or space-like vectors according to whether the quadratic form $< X, X >= \mathbf{g}_{\alpha\beta} X^\alpha X^\beta$ is, respectively, negative, zero or positive.

The set of null vectors form a double cone, called the null cone of the corresponding point. The set of time-like vectors form the interior of this cone. It has two connected components whose boundaries are the corresponding components of the null cone. The set of space-like vectors is the exterior of the null cone, a connected open set. Any physically meaningful space-time should be time orientable, that is, one can choose in a continuous fashion a future directed component of the set of time-like vectors. This allows us to specify the causal future and past of any point in space-time. More generally, the causal future of a set $S \subset \mathbf{M}$, denoted by $J^{+}(S)$, is defined as the set of points q which can be reached by a future directed causal curve [1] which initiates at S. Similarly $J^{-}(S)$ consists of the set of all points q which can be reached, from S, by a past directed causal curve.

The boundaries of past and future sets of points in $\mathbf{M}$ are null geodesic cones, often called light cones. Their specification defines the *causal structure* of the space-time which, up to a conformal factor, uniquely determines the metric.

A hypersurface M in $\mathbf{M}$ is said to be space-like if its normal direction is time-like at every point on M. We denote by g the Riemannian metric induced by $\mathbf{g}$ on M. The covariant differentiaton on the space-time $\mathbf{M}$ will be denoted by $\mathbf{D}$, while that on M will be written with the symbols D or ∇. Similarly we denote by $\mathbf{R}$, resp.R, the Riemann curvature tensors of $\mathbf{M}$, resp. M. Recall that for any given vector fields X, Y, Z on $(\mathbf{M}, \mathbf{g})$,

$$\mathbf{D}_X \mathbf{D}_Y - \mathbf{D}_Y \mathbf{D}_X Z = \mathbf{R}(X, Y)Z + \mathbf{D}_{[X,Y]} Z$$

or, in components, relative to an arbitrary frame e_α, $\alpha = 0, 1, 2, 3$,

$$\mathbf{D}_\beta \mathbf{D}_\alpha Z^\gamma = \mathbf{D}_\alpha \mathbf{D}_\beta Z^\gamma + \mathbf{R}^\gamma_{\ \sigma\beta\alpha} Z^\sigma$$

The extrinsic curvature, or second fundamental form, of M in $\mathbf{M}$ will be denoted by k. Recall that, if T denotes the future directed unit normal to M we have,

$$k_{ij} = - < \mathbf{D}_{e_i} T, e_j > = < T, \mathbf{D}_{e_i} e_j >$$

with e_i, $i = 1, 2, 3$, an arbitrary frame on M.

We will use the notation $\in_{\alpha\beta\gamma\delta}$ to express the components of the volume element $d\mu_{\mathbf{M}}$ relative to an arbitrary frame. Similarly, if e_i, $i = 1, 2, 3$ is an arbitrary frame on M, then $\in_{ijk} = \in_{oijk}$ are the components of $d\mu_M$, the volume element of M, with respect to the frame $e_0 = T, e_1, e_2, e_3$.

The Riemann curvature tensor $\mathbf{R}$ of the space-time satisfies the following,

Bianchi Identities

$$\mathbf{D}_{[\epsilon} \mathbf{R}_{\alpha\beta]\gamma\delta} = \frac{1}{3}(\mathbf{D}_\epsilon \mathbf{R}_{\alpha\beta\gamma\delta} + \mathbf{D}_\alpha \mathbf{R}_{\beta\epsilon\gamma\delta} + \mathbf{D}_\beta \mathbf{R}_{\epsilon\alpha\gamma\delta}) = 0.$$

[1] A differentiable curve $\lambda(t)$ whose tangent at every point is a future directed time-like or null vector

The traceless part of the curvature tensor is the Weyl tensor,

$$\begin{aligned}\mathbf{C}_{\alpha\beta\gamma\delta} =&\mathbf{R}_{\alpha\beta\gamma\delta} - \frac{1}{2}(\mathbf{g}_{\alpha\gamma}\mathbf{R}_{\beta\delta} + \mathbf{g}_{\gamma\delta}\mathbf{R}_{\alpha\beta} - \mathbf{g}_{\beta\gamma}\mathbf{R}_{\alpha\delta} - \mathbf{g}_{\alpha\delta}\mathbf{R}_{\beta\gamma})\\ &+ \frac{1}{6}(\mathbf{g}_{\alpha\beta}\mathbf{g}_{\gamma\delta} - \mathbf{g}_{\alpha\delta}\mathbf{g}_{\beta\gamma})\mathbf{R}\end{aligned}$$

where the 2-tensor $\mathbf{R}_{\alpha\beta}$ and scalar $\mathbf{R}$ are respectively the Ricci tensor and the scalar curvature of the space-time. We notice that the Riemann curvature tensor has twenty independent components while the Weyl and Ricci tensors have ten components each.

The Weyl tensor is a particular example of a spin-2 tensor. This refers to arbitrary four tensors W which satisfy all the symmetry properties of the curvature tensor and, in addition are traceless. We say that such a W satisfies the spin-2 equation if, with respect to the covariant differentiation on $\mathbf{M}$,

Spin-2 Equation

$$\mathbf{D}_{[\epsilon}W_{\alpha\beta]\gamma\delta} = 0.$$

For a spin-2 tensor W the following definitions of left and right Hodge duals are equivalent:

$${}^\star W_{\alpha\beta\gamma\delta} = \frac{1}{2} \in_{\alpha\beta\mu\nu} W^{\mu\nu}{}_{\gamma\delta}$$

$$W^\star_{\alpha\beta\gamma\delta} = W_{\alpha\beta}{}^{\mu\nu}\frac{1}{2} \in_{\mu\nu\gamma\delta}$$

where $\in^{\alpha\beta\gamma\delta}$ are the components of the volume element in $\mathbf{M}$. One easily checks that, ${}^\star W = W^\star$ is also a spin-2 field and ${}^\star({}^\star W) = -W$. Given an arbitrary vector field X, we can define the *electro-magnetic decomposition* of W to be the pair of 2-tensors formed by contracting W with X accordingly to the formulas,

$$ii_X(W)_{\alpha\beta} = W_{\mu\alpha\nu\beta}X^\alpha X^\beta$$

$$ii_X({}^\star W)_{\alpha\beta} =^\star W_{\mu\alpha\nu\beta}X^\alpha X^\beta$$

These new tensors are both symmetric, traceless, and orthogonal to X. Moreover they completely determine W, provided that X is not null (see [Ch-Kl]).

One can associate to the Weyl tensor or, more general, to any spin-2 field W, a 4-tensor which is quadratic in W and plays precisely the same role, for solutions of the spin-2 equations, as the energy-momentum tensor of an electromagnetic field plays for solutions of the Maxwell equations.

Bell-Robinson Tensor

$$Q_{\alpha\beta\gamma\delta} = \frac{1}{2}(W_{\alpha\mu\beta\nu}W_\gamma{}^\mu{}_\delta{}^\nu + {}^\star W_{\alpha\mu\beta\nu}\,{}^\star W_\gamma{}^\mu{}_\delta{}^\nu)$$

Q is fully symmetric and traceless, moreover it satisfies the positive energy condition, $Q(X, Y, X, Y)$ is positive whenever X, Y are future directed time-like vectors (see [Ch-Kl], for a proof of the above properties of Q). Moreover,

$$\mathbf{D}^{\delta} Q_{\alpha\beta\gamma\delta} = 0$$

whenever W satisfies the spin-2 equations.

It is well known that the causal structure of an arbitrary Einstein space time can have undesirable pathologies. All these can be avoided by postulating the existence of a Cauchy hypersurface in M, i.e. a hypersurface Σ with the property that any causal curve intersects it at precisely one point. [2] Einstein space-times with this property are called ***globally hyperbolic.*** Such space-times are in particular stable causal, i.e. they allow the existence of a globally defined differentiable function t whose gradient, $\mathbf{D}t$, is everywhere time-like. We call t a *time function,* and the foliation given by its level surfaces a *t-foliation.* We denote by T the future directed unit normal to the foliation.

Topologically, a space-time foliated by the level surfaces of a time function is diffeomorphic to a product manifold $\Re \times \Sigma$ where Σ is a three dimensional manifold. Indeed the space-time can be parametrized by points on the slice $t = 0$ by following the integral curves of $\mathbf{D}t$. Moreover relative to this parametrization, the space-time metric takes the form,

$$ds^2 = -\phi^2(t, x)dt^2 + \sum_{i,j=1}^{3} g_{ij}(t, x)dx^i dx^j \tag{0.1}$$

where $x = (x^1, x^2, x^3)$ are arbitrary coordinates on the slice $t = 0$. The function

$$\phi(t, x) = \frac{1}{< \mathbf{D}t, \mathbf{D}t >^{1/2}}$$

is called the *lapse function* of the foliation, g_{ij} its first fundamental form. We refer to (0.1) as the canonical form of the space-time metric with respect to the foliation.

The foliation is said to be normal if

Normal Foliation Condition

$$\phi \to 1 \text{ as } x \to \infty \text{ on each leaf } \Sigma_t.$$

The second fundamental form of the foliation, i.e. the extrinsic curvature of the leaves Σ_t, is given by,

$$k_{ij} = -(2\phi)^{-1}\partial_t g_{ij} \tag{0.2}$$

We denote by ∇ the induced covariant derivative on the leaves Σ_t and by R_{ij} the corresponding Ricci curvature tensor. Since Σ_t is three dimensional, we recall that the

[2] In particular Σ is a space-like hypersurface.

Ricci curvature R_{ij} completely determines the induced Riemann curvature tensor R_{ijkl} accordingly to the formula,

$$R_{ijkl} = g_{ik}R_{jl} + g_{jl}R_{ik} - g_{jk}R_{il} - g_{il}R_{jk} - \frac{1}{2}(g_{ik}g_{jl} - g_{jk}g_{il})R$$

where R is the scalar curvature $g^{ij}R_{ij}$. The second fundamental form k, the lapse function ϕ and the Ricci curvature tensor R_{ij} of the foliation are connected to the space-time curvature tensor $\mathbf{R}_{\alpha\beta\gamma\delta}$ accordingly to the following,

The Structure Equations of the Foliation

$$\begin{aligned} \partial_t k_{ij} &= -\nabla_i\nabla_j\phi + \phi(\mathbf{R}_{iTjT} - k_{ia}k^a{}_j) && (0.3a) \\ \nabla_i k_{jm} - \nabla_j k_{im} &= \mathbf{R}_{mTij} && (0.3b) \\ R_{ij} - k_{ia}k^a{}_j + k_{ij}\mathrm{tr}k &= \mathbf{R}_{iTjT} + \mathbf{R}_{ij} && (0.3c) \end{aligned}$$

where ∂_t denotes the partial derivative with respect to t and $\mathbf{R}_{iTjT}$, $\mathbf{R}_{mTij}$ are the components $\mathbf{R}(\partial_i, T, \partial_j, T)$ and, respectively, $\mathbf{R}(\partial_m, T, \partial_i, \partial_j)$ of the space-time curvature relative to arbitrary coordinates on Σ. The equations $(0.3a)$ are the second variation formulas, while $(0.3b)$ and $(0.3c)$ are the classical Gauss-Codazzi and, respectively, Gauss equations of the foliation.

In view of $(0.3c)$, the equation $(0.3a)$ becomes,

$$\partial_t k_{ij} = -\nabla_i\nabla_j\phi + \phi(-\mathbf{R}_{ij} + R_{ij} + \mathrm{tr}k k_{ij} - 2k_{ia}k^a{}_j) \qquad (0.3d)$$

Taking the trace of the equations $(0.3c)$, $(0.3b)$ and $(0.3a)$ respectively, we derive,

$$\begin{aligned} R - |k|^2 + (\mathrm{tr}k)^2 &= 2\mathbf{R}_{TT} + \mathbf{R} && (0.4a) \\ \nabla^j k_{ji} - \nabla_i \mathrm{tr}k &= \mathbf{R}_{Ti} && (0.4b) \\ \partial_t \mathrm{tr}k &= -\Delta\phi + \phi(\mathbf{R}_{TT} + |k|^2) && (0.4c) \end{aligned}$$

where $|k|^2 = k_{ij}k^{ij}$.

By contrast to Riemannian geometry where the basic covariant equations one encounters are of elliptic type, in Einstein geometry the basic equations are hyperbolic. The causal structure of the space-time is tied to the evolutions feature of the corresponding equations. This is particularly true for the Einstein field equations where the space itself is the dynamic variable.

The Einstein field equations were proposed by Einstein as a unified theory of space-time and gravitation. As mentioned above the space-time $(\mathbf{M}, \mathbf{g})$ is the unknown; one has to find an Einstein metric $\mathbf{g}$ such that,

Einstein Field Equations

$$\mathbf{G}_{\mu\nu} = 8\pi\mathbf{T}_{\mu\nu}$$

where $\mathbf{G}_{\mu\nu}$ is the tensor $\mathbf{R}_{\mu\nu} - \frac{1}{2}\mathbf{g}_{\mu\nu}\mathbf{R}$, with $\mathbf{R}_{\mu\nu}$ the Ricci curvature of the metric, $\mathbf{R}$ its scalar curvature and $\mathbf{T}_{\mu\nu}$, the energy momentum tensor of a matter field (e.g. the Maxwell equations). Contracting twice the Bianchi identities $\mathbf{D}_{[\epsilon}\mathbf{R}_{\alpha\beta]\gamma\beta} = 0$ we derive

Contracted Bianchi Identities

$$\mathbf{D}^{\nu}\mathbf{G}_{\mu\nu} = 0$$

which are equivalent to the divergence equations of the matter-field,

$$\mathbf{D}^{\nu}\mathbf{T}_{\mu\nu} = 0.$$

In the simplest situation of the physical vacuum, i.e. $\mathbf{T} = 0$, the Einstein equations take the form,

Einstein-Vacuum Equations

$$\mathbf{R}_{\mu\nu} = 0.$$

In view of the four contracted Bianchi identities mentioned above, the Einstein-Vacuum equations, for short E-V, can be viewed as a system of 10-4=6 equations for the 10 components of the metric tensor $\mathbf{g}$. The remaining 4 degrees of freedom correspond to the general covariance of the equations. Indeed if $\Phi : M \longrightarrow M$ is a diffeomorphism then the pairs $(M, \mathbf{g})$ and $(M, \Phi^{*}\mathbf{g})$ represent the same solution of the field equations.

Written explicitly in an arbitrary system of coordinates the E-V equations lead to a degenerate system of equations. The well posedness of the Cauchy problem, which we discuss below, was proved however by Y.Choquet-Bruhat in harmonic coordinates (see [Br]), yet as she has pointed out later these are unstable in the large. This problem of finding a globally stable, well posed coordinate conditions is the first major difficulty one has to overcome in the construction of global solutions to the Einstein equations.

To emphasize the dynamic character of the E-V equations it is helpful to express them in terms of the parameters ϕ, g, k of an arbitrary t-foliation. Thus assuming that the space-time $(\mathbf{M}, \mathbf{g})$ can be foliated by the level surfaces of a time function t, and writing $\mathbf{g}$ in its canonical form (0.1), the E-V equations are equivalent to the following,

Constraint Equations for E-V

$$\nabla^{j}k_{ji} - \nabla_{i}\mathrm{tr}k = 0 \tag{0.5a}$$

$$R - |k|^{2} + (\mathrm{tr}k)^{2} = 0 \tag{0.5b}$$

Evolution Equations for E-V

$$\partial_{t}g_{ij} = -2\phi k_{ij} \tag{0.6a}$$

$$\partial_{t}k_{ij} = -\nabla_{i}\nabla_{j}\phi + \phi(R_{ij} + \mathrm{tr}k k_{ij} - 2k_{ia}k^{a}{}_{j}) \tag{0.6b}$$

Indeed the equivalence of the equations (0.5*a*), (0.5*b*), (0.6*a*) (0.6*b*) with the E-V is an immediate consequence of (0.4*a*), (0.4*b*), and (0.5*a*).

Also, $(0.4c)$ becomes

$$\partial_t \mathrm{tr} k = -\Delta\phi + \phi(R + (\mathrm{tr} k)^2) \tag{0.7}$$

Given a t-foliation we denote by E, H the electro-magnetic decomposition of the curvature tensor $\mathbf{R}$, of an E-V manifold with respect to T, the future oriented unit normal to the time foliation. Clearly E, H are symmetric traceless 2 tensors tangent to the foliation. In view of these definition the equations $(0.3b)$ and $(0.3c)$ become

$$\nabla_i k_{jm} - \nabla_j k_{im} = \in_{ij}\ {}^l H_{lm} \tag{0.8a}$$

$$R_{ij} - k_{ia} k^{aj} + k_{ij} \mathrm{tr} k = E_{ij} \tag{0.8b}$$

Note that the total number of unknowns in the Evolution Equations $(0.6a)$ and $(0.6b)$ is 13 while the total number of equations is only 12. This discrepancy corresponds to the remaining freedom of choosing the time function t which defines the foliation.

We also note that, in view of the twice contracted Bianchi identities, if g, k satisfy the Evolution Equations, then the Constraint Equations $(0.5a)$ and $(0.5b)$ are automatically satisfied on any Σ_t provided they are satisfied on a given initial slice Σ_{t_0}. Therefore they can be regarded as constraints on given initial conditions for g and k. Accordingly to this an *initial data set* for E-V is defined to be a triplet (Σ, g, k) consisting of a three dimensional manifold Σ together with a Riemannian metric g and covariant symmetric 2-tensor k which satisfy the constraint equations $(0.5a)$ and $(0.5b)$ on Σ.

A development of an initial data set consists of an Einstein-Vacuum space-time $(\mathbf{M}, \mathbf{g})$ together with an embedding $i : \Sigma \longrightarrow \mathbf{M}$ such that g and k are the induced first and second fundamental forms of Σ in $\mathbf{M}$. The central question in the mathematical theory of E-V equations is the study of the evolution of general initial data sets.

The simplest solution of E-V equations is the Minkowski space-time $\mathbf{R}^{3+1}$, i.e. the space $\mathbf{R}^4$ together with a given Einstein metric $<,>$ and a canonical coordinate system (x^0, x^1, x^2, x^3) such that

$$< \partial_\alpha, \partial_\beta >= \eta_{\alpha\beta};\ \alpha, \beta = 0, 1, 2, 3$$

The issue we want to address in our work is that of the global nonlinear stability of the Minkowski space. More precisely we want to investigate whether Cauchy developments of initial data sets which are close, in an appropriate sense, to the trivial data set lead to global, smooth, geodesically complete solutions of the Einstein-Vacuum equations which remain close, in an appropriate, global sense, to the Minkowski space. We like to stress the fact that at the present time is not even known whether is *any smooth, geodesically complete solution which becomes flat at infinity on any given space-like direction.* Any attempt to significantly simplify the problem by looking for solutions with additional symmetries, fails as a consequence of the well known results of Lichnerowics for static solutions, [3] and Birkhoff for spherically symmetric solutions. Accordingly to the

[3] A space-time is said to be stationary if there exists a one parameter group of isometries whose orbits are time-like curves. It is said to be static if, in addition, the orbits of the group are orthogonal to a space-like hypersurface

first, a static solution which is geodesically complete and flat at infinity on any space-like hypersurface must be flat. The Birkhoff theorem asserts that all spherically symmetric solutions of the E-V equations are static. Thus, disregarding the Schwartzchild solution which is not geodesically complete, the only such solution which becomes flat at space-like infinity, is the Minkowski space.

The problem of stability of the Minkowski space is closely related with that of characterizing the space-time solutions of the Einstein-Vacuum equations which are *globally asymptotically flat* i.e., as defined in the physics literature, space-times which become flat as we approach infinity in any direction. Despite the central importance that such space-times have in General Relativity, as corresponding to isolated physical systems, it is not at all settled how to define them correctly, consistent with the field equations. Attempts to develop such a notion have been made however in the last 25 years (see [Ne-To] for a survey) beginning with the work of Bondi ([Bo-Bu-Me],[Bo]) (see also [Sa]) who introduced the idea to analyze solutions of the field equations along null hypersurfaces. The present state of understanding was set by Penrose ([Pe2],[Pe1]) who formalized the idea of asymptotic flatness by adding a boundary at infinity attached through a smooth *conformal compactification*. However it remains questionable whether there exists any nontrivial [4] solution of the field equations which satisfy the Penrose requirements. Indeed his regularity assumptions translate into fall-off conditions of the curvature which, we believe, are too stringent and thus fail to be satisfied by any solution which could allow gravitational waves. Moreover the picture given by the conformal compactification fails to address the crucial issue of the relationship between conditions in the past and behavior in the future.

We believe that a real understanding of asymptotically-flat spaces can only be accomplished by constructing them from initial data, and studying their asymptotic behaviour. This is precisely the objective we set out to achieve.

To make our discussion more precise we have to introduce the notion of an asymptotically-flat initial data set. By this we understand an initial data set (Σ, g, k) with the property that the complement of a finite set in Σ is diffeomorphic to the complement of a ball in R^3 (i.e. Σ is diffeomorphic to R^3 at *infinity*) and the notion of energy, linear and angular momentum are well defined and finite. These can be unambiguously defined for the following class of initial data sets we will refer to as ***strongly asymptotically flat.***

We say that an initial data set (Σ, g, k) satisfies the S.A.F. condition if g, k are sufficiently smooth, and there exists a coordinate system (x^1, x^2, x^3) defined in a neighborhood of infinity such that, as, $r = [\Sigma_{i=1}^3 (x^i)^2]^{\frac{1}{2}} \to \infty$

S.A.F. Initial Data Sets [5]

$$g_{ij} = (1 + 2M/r)\delta_{ij} + o_4(r^{-\frac{3}{2}}) \tag{0.9a}$$

[4] Namely a nonstationary solution.

[5] A function f is said to be $o_m(r^{-k})$, resp. $O_m(r^{-k})$, as $r \to \infty$ if $\partial^l f(x) = o(r^{-k-l})$, resp. $O(r^{-l-k})$, for any $l = 0, 1, \ldots, m$, where ∂^l denote all the partial derivatives of order l relative to the coordinates (x^1, x^2, x^3)

$$k_{ij} = o_3(r^{-\frac{5}{2}}) \qquad (0.9b)$$

We shal call the leading term, $(1+2M/r)\delta_{ij}$ in the expansion $(0.9a)$ *the Schwarzchild Part* of the metric g. [6]

Given such a data set the ADM (Arnowitt, Deser and Misner) definitions of energy E, linear momentum P and angular momentum J are given by,

$$E = \frac{1}{16} \lim_{r\to\infty} \int_{S_r} \sum_{i,j} (\partial_i g_{ij} - \partial_j g_{ii}) N^j dA$$

$$P_i = \frac{1}{8} \lim_{r\to\infty} \int_{S_r} (k_{ij} - \mathrm{tr}k g_{ij}) N^j dA, \; i = 1,2,3$$

$$J_i = \frac{1}{8} \lim_{r\to\infty} \int_{S_r} \epsilon_{iab}\, x^a (k^{bj} - g^{bj}\mathrm{tr}k) N_j dA, \; i = 1,2,3$$

where S_r is the coordinate sphere of radius r, N is the exterior unit normal to it and dA its area element. Clearly the limits on the right hand side of the formulas defining E and P exist and are finite. To check that J is well defined one has to remark that the difference between the integrals on two different spheres r_1, r_2, can be written as a volume integral of an expression which involves, as higher order term, $\partial_j k_b{}^j - \partial_b(\mathrm{tr}k) = \nabla_j k_b{}^j - \nabla_b(\mathrm{tr}k) + o(r^{-\frac{9}{2}})$. The assertion follows then with the help of the constraint equations $(0.5a)$.

Moreover, due to our conditions $(0.9a)$ and $(0.9b)$ we have

$$E = M, \quad P = 0.$$

Thus the S.A.F. conditions implies that the initial data set is in a center of mass frame. In view of the positive mass theorem M must be a positive number vanishing only if the initial data set is flat.

The definition of the energy-momentum (E, P_1, P_2, P_3) and the angular momentum (J_1, J_2, J_3) are independent of the particular choice of the coordinates x^1, x^2, x^3 in the definition of S.A.F. initial data sets. [7] Moreover, they are presevеd by the evolution equations $(0.6a)$ and $(0.6b)$ of a normally foliated (see definition on page 5) E-V spacetime. This can be easily checked by taking the time derivatives of the expressions defining E, P, J.

We believe that the question we are investigating here, namely the stability of the Minkowski space, requires initial data sets with finite energy, linear and angular momentum.

[6] It is the same as that of a space-like hypersurface, orthogonal to the Killing vector field of a Schwarzchild space-time.

[7] Indeed, first remark that the definitions are invariant under rigid transformations of the coordinates x^1, x^2, x^3. It thus suffices to show that the variations of the integrals defining E, P, J, with respect to one parameter groups of diffeomorphisms generated by vector fields $\psi = O_3(1)$ as $r \to \infty$, vanish in the limit.

In its least precise version our main result asserts the following, **First Version of the Main Theorem**

Any Strongly Asymptotically Flat initial data set which satisfies, in addition, a Global Smallness Assumption, leads to a unique, globally hyperbolic, smooth and geometrically complete development, solution of the Einstein-Vacuum Equations. Moreover this development is globally asymptotically flat, by which we mean that its Riemann curvature tensor approaches zero on any causal or space-like geodesic, as the corresponding affine parameter tends to infinity.

The main difficulties one encounters in the proof of our reslut are the following:

The main difficulties one encounters in the proof of our result are the following:

1. The problem of coordinates.
2. The strongly nonlinear hyperbolic features of the Einstein equations.
3. The logarithmic divergence of the light cones.

1. The problem of coordinates is, as we have mentioned above the first major difficulty one has to overcome when trying to solve the Cauchy problem of the Einstein equations. Our strategy is based on two ideas. First, we describe our space-time by specifying, instead of full coordinate conditions, only a time function whose level hypersurfaces are *maximal.* [8] More precisely we impose, in addition to the equations (0.5*a*), (0.5*b*), the constraint

$$\mathrm{tr}k = 0 \tag{0.10}$$

With this choice we remove the indeterminancy of the evolution equations (0.6*a*), (0.6*b*) and obtain the following *determined* system of equations for the maximal foliation of an E-V space-time:

Constraint Equations of a Maximal Foliation (0.10)

$$\mathrm{tr}k = 0 \tag{0.11a}$$
$$\nabla^j k_{ji} = 0 \tag{0.11b}$$
$$R = |k|^2 \tag{0.11c}$$

Evolution Equations of a Maximal Foliation (0.11)

$$\partial_t g_{ij} = -2\phi k_{ij} \tag{0.12a}$$
$$\partial_t k_{ij} = -\nabla_i \nabla_j \phi + \phi(R_{ij} - 2k_{ia}k^a{}_j) \tag{0.12b}$$

[8] In Einstein geometry a maximal hypersurface is one which is space-like and maximizes the volume among all possible compact perturbations of it.

Lapse Equation of a Maximal Foliation

$$\Delta\phi = |k|^2\phi \tag{0.13}$$

Note that the time function is defined by especifying the level sets only up to a transformation of the form $t \to f \circ t$ with f any, orientation preserving, diffeomorphism of the real line. However we can specify a unique t by requiring that, regarded as a parameter on an integral curve Γ_x of t which passes through a point x of Σ_0, it converges to the arclength on Γ_x as x tends to infinity on Σ_0. This is equivalent to the condition that ϕ tends to 1 at infinity on each Σ_t, which is precisely the normal foliation condition introduced above. Indeed, with the exception of the Minkowski space-time, the above definition specifies a unique time function. This is due to the fact that, when the A.D.M. energy E is non-zero, there is a unique maximal foliation with respect to which the linear momentum P vanishes. In physical terms, this foliation constitutes the center of mass frame of the corresponding isolated system.

The second idea is to make use in a fundamental way, of the Bianchi identities of the space-time and the Bell-Robinson tensor introduced below. The basic observation is that, once we have good estimates for the curvature tensor $\mathbf{R}$, all the parameters of the foliation, i.e. g, k, ϕ, are determined purely by solving the elliptic system,

$$R_{ij} - k_{ia}k^a{}_j = E_{ij} \tag{0.14a}$$

$$\text{curl}\, k_{ij} = H_{ij} \tag{0.14b}$$

$$\nabla^j k_{ji} = 0 \tag{0.14c}$$

$$\text{tr}k = 0 \tag{0.14d}$$

together with the lapse equation (0.13).

The equations (0.14a), (0.14b) are immediate consequences of, respectively, (0.8b), (0.8a) with curl $k_{ij} = \in_i^{ab}\, k_{jab}$. Thus all the evolution features of the Einstein equations are contained in the Bianchi identities, which have the great advantage of being covariant.

2. The other major obstacle in the study of the Einstein equations consists in their hyperbolic and strongly nonlinear character. The only powerful analytic tool we have in the study of nonlinear hyperbolic equations, in the physical space-time dimension, are the energy estimates. Yet the classical energy estimates are limited to proving estimates which are local in time. The difficulty has to do with the fact that, in order to control the higher energy norms of the solutions, one has to control the integral in time of their bounds in uniform norm. In recent years however, new techniques were developed, based on modified energy estimates and the invariance property of the corresponding linear equations, which were applied to prove global or long time existence results for nonlinear wave equations (see [Kl3], [Kl1]). More precisely, one uses the Killing and conformal Killing vector fields generated by the conformal group of the Minkowski space to define a global energy norm which is invariant relative to the linear evolution. The

precise asymptotic behaviour, including the uniform bounds mentioned above, are then an immediate consequence of a global version of the Sobolev inequalities (see [Kl3], [Kl2], [Ho]).

The relevant linearized equations for the E-V field equations are the spin-2 equations (see page 3) in Minkowski space. As a first preliminary step in our program, we have analyzed the complete asymptotic properties of the spin-2 equations in Minkowski space by using only energy estimates and the conformal invariance properties of the equations in the spirit of the ideas outlined above (see [Ch-Kl]).

However to derive a global existence result one also needs to investigate the structure of the nonlinear terms [9]. It is well known that arbitrary quadratic nonlinear perturbations of the scalar wave equation, even when derivable from a Lagrangian could lead to formation of singularities unless a certain structural condition, which we have called the *Null Condition,* is satisfied (see [Ch], [Kl1]). It turns out that the appropriate, tensorial version of this structural condition is satisfied by the Einstein equations. One could say that the troublesome nonlinear terms, which could have led to formation of singularities are in fact excluded due to the covariance and algebraic properties of the Einstein equations.

3. In implementing the strategy outlined in (1) and (2) one encounters a very serious technical difficulty. The *mass term* which appears in the Schwartzchild part of an (S.A.F.) initial data set, (0.9a), has the long range effect of distorting the asymptotic position of the null geodesic cones. They are expected to diverge logarithmically from their corresponding position in flat space. In addition to this their asymptotic shear differs drastically from that in the Minkowski space-time. This difference reflects the presence of gravitational radiation in any nontrivial perturbation the Minkowski space-time [10]. To take this effect into account one has to appriopriately modify the Killing and conformal Killing vector fields used in the definition of the basic energy norm. We achieve this by an elaborate construction of an *optical function* whose level surfaces are outgoing null hypersurfaces, related by a translation at infinity. The construction of the optical function and the approximate Killing and conformal Killing vector fields related to it requires more than a half of our work. The most demanding part in the construction is taken by the angular momentum vector fields [11]. These are particularly important to our construction as they are crucial in circumventing the problem of slow decay at infinity of the initial data set. Thus we do not estimate directly **R** from the Bianchi identities but only its Lie derivatives with respect to these vector fields. This allows us to consider higher weighted norms than will be possible for **R**. Yet, as it turns out, the latter can be easily estimated in terms of the former [12]. Similarly we use the approximate Killing

[9] Generated each time we commute the Bianchi Identities with one of the vector fields used in the definition of the global energy norm.

[10] For more details of this fact we refer to the next section

[11] I.e. the vector fields which can be viewed as deformation of $\omega_{ij}=x_i\partial_j-x_j\partial_i$, for $i,j=1,2,3$, of Minkowski space.

[12] This fact seems entirely plausible in view of the Birkhoff Theorem.

vector field T, the unit normal to the foliation, to allow higher weighted norms for the Lie derivatives of the curvature tensor with respect to T [13].

As outlined above our construction requires initial data sets which satisfy in addition to the constraint equations, the maximal condition tr$k = 0$. We will refer to them as maximal, in what follows.

To make the statement of our main theorem precise we need also to define what we mean by the global smallness assumption. Before stating this condition, we assume the metric g to be complete and introduce the following quantity:

$$
\begin{aligned}
Q(x_{(0)}, a) = &\sup_{\Sigma}\{a^{-2}(d_0^2 + a^2)^3 |Ric|^2\} + \\
&a^{-3}\{\int_{\Sigma} \sum_{\ell=0}^{3} (d_0^2 + a^2)^{\ell+1} |\nabla^{\ell} k|^2 + \int_{\Sigma} \sum_{\ell=0}^{1} (d_0^2 + a^2)^{\ell} + 3|\nabla^{\ell} B|^2\}
\end{aligned}
$$

where $d_0(x) = d(x_{(0)}, x)$ is the Riemannian geodesic distance between the point x and a given point $x_{(0)}$ on $\Sigma, |Ric|^2 = \mathbf{R}^{ij}\mathbf{R}_{ij}, \nabla^l$ denote the l covariant derivatives and B is the symmetric, traceless 2-tensor tensor [14],

$$
B_{ij} = \in_j {}^{ab} \nabla_b (R_{ia} - \frac{1}{4} g_{ia} R)
$$

We say that an S.A.F. initial data set, (Σ, g, k), satisfies the global smallness assumption:

The Global Smallness Assumption

The metric g is complete and there exists a sufficiently small positive ϵ s.t.

$$
\inf_{x_{(0)} \in \Sigma, a \geq 0} Q(x_{(0)}, a) \leq \epsilon \qquad (0.15)
$$

Second Version of the Main Theorem [15]

Any Strongly Asymptotically Flat, Maximal, initial data set which satisfies the Global Smallness Assumption (0.15), *leads to a unique, globally hyperbolic, smooth and geodesically complete solution of the Einstein-Vacuum Equations foliated by a normal, maximal time foliation. Moreover this development is globally asymptotically flat.* [16]

[13] In view of the Lichnerowitz theorem, this procedure allows us to obtain information about $\mathbf{R}$ itself.

[14] Remark that B is dual to the tensor $\mathbf{R}_{ijk} = \nabla_k \mathbf{R}_{ij} + \nabla_j \mathbf{R}_{ik} + \frac{1}{4}(g_{ik}\nabla_j \mathbf{R} - g_{ij}\nabla_k \mathbf{R})$ whose vanishing characterizes locally conformally flat three-dimensional manifolds (see [Eisen]). Thus, up to lower order terms, the Schwarzchild part of g does not affect it.

[15] The first version of the Theorem is not an immediate consequence of the second. It can be proved however by, first, developing the initial data set locally in time and, then, imbedding in it a maximal hypersurface. Imbedding results of the type one needs were obtained by Bartnik (see [Ba]).

[16] A precise statement of the asymptotic behaviour for the curvature tensor $\mathbf{R}$ and also for the lapse function ϕ and second fundamental form k of the foliation is too technical for the purpose of this report.

We next indicate how to construct maximal initial data sets which are asymptotically flat and satisfy (0.15). This is based on the observation that the constraint equations (0.11a) and (0.11b) are conformal invariant. More precisely they are invariant with respect to the transformation, $g_{ij} \to \Phi^4 g_{ij}$ and $k_{ij} \to \Phi{-}2k_{ij}$. Thus, given arbitrary solutions $\tilde{g}, \tilde{k}$ to the equations,

$$\mathrm{tr}_{\tilde{g}}\tilde{k} = 0 \qquad (0.16a)$$

$$\tilde{\nabla}^j \tilde{k}_{ji} = 0 \qquad (0.16b)$$

where $\tilde{\nabla}$ denotes the covariant differentiation with respect to the metric $\tilde{g}$, we infer that $g_{ij} = \Phi^4 \tilde{g}_{ij}$ and $k_{ij} = \Phi{-}2\tilde{k}ij$, are solutions to the same equations for arbitrary function Φ. To satisfy also the equation (0.11c) we have to subject Φ to the Lichnerowitz equation

$$\tilde{\Delta}\Phi - \frac{1}{8}\tilde{R}\Phi + |\tilde{k}|^2_{\tilde{g}}\Phi^{-7} = 0 \qquad (0.17a)$$

In practice one does not solve directly the Lichnerowitz equation. The standard approach is to look for Φ of the form $\Phi = \Omega\Psi$ where Ω and Ψ are the conformal factors corresponding to transformations which take, first, an arbitrary solution of the equations (0.16a), (0.16b) to a solution $\bar{g}, \bar{k}$ of the same equations and, then, take $\bar{g}, \bar{k}$ to the desired solution g, k. The first conformal factor Ω is chosen so that the Ricci curvature $\bar{R}$ of $\bar{g}$ vanishes identically. Thus Ω has to be a solution of the linear equation

$$\tilde{\Delta} - \frac{1}{8}\tilde{R}\Omega = 0 \qquad (0.17b)$$

The second conformal factor Ψ is chosen such that the transformed variables g, k satisfy $R = |k|^2$. For this to happen Ψ has to be a solution of the linear equation

$$\bar{\Delta}\Psi + \frac{1}{8}|\bar{k}|^2_{\bar{g}}\Psi^{-7} = 0 \qquad (0.17c)$$

Note that, by virtue of the maximal principle, the equation (0.17c) has always a smooth solution, $\Psi \geq 1$, with $\Psi \to \infty$ on Σ. On the other hand a sufficient condition so that the equation (0.17b) has a positive solution with the same property is that the $L^{\frac{3}{2}}$ norm of the negative part of $\bar{R}$ is sufficiently small. Therefore, (Σ, g, k) is an initial data set satisfying the S.A.F. conditions (0.9a), (0.9b) provided that the corresponding solutions $\tilde{g}_{ij}, \tilde{k}$ of (0.16a), (0.16b) verify

$$\tilde{g}_{ij} = \delta_{ij} + o_4(r^{-3/2})$$

$$\tilde{k}_{ij} = o_3(r^{-5/2})$$

and the relative part of $\bar{R}$ satisfies the smallness condition mentioned above. Moreover g, k satisfy the Global Smallness Assumption of the Theorem provided that the metric $\tilde{g}$ is complete and, there exists a small positive ϵ such that,

$$\inf_{x_{(0)}\in\Sigma, a\geq 0}\{\sup_{\Sigma}(a^2+\tilde{d}_0^2)^3|\widetilde{Ricc}|^2+\int_\Sigma\sum_{\ell=0}^{2}(a^2+\tilde{d}_0^2)^{\ell+2}|\tilde{\nabla}\widetilde{Ricc}|^2$$
$$+\int_\Sigma\sum_{\ell=0}^{3}(a^2+\tilde{d}_0^2)^{\ell+1}|\tilde{\nabla}^\ell\tilde{k}|^2\}<\epsilon$$

where $\tilde{d}_0(x)$ denotes the Riemannian geodesic distance relative to $\tilde{g}$ between the point x and a given point $x_{(0)}$ on Σ.

It only remains to discuss whether the equations (0.16a), (0.16b) have solutions verifying the above properties. This can be done using the orthogonal York decomposition of any symmetric, traceless 2-covariant tensor h, on a three-dimensional Riemannian manifold $(\Sigma, \tilde{g})$, into a divergence-free part $\tilde{k}$ and the traceless part of the deformation tensor of a vector field X,

$$h_{ij}=\tilde{k}_{ij}+\widehat{\mathcal{L}_{\setminus}}\tilde{g}_{ij}.$$

The vector field X has to be a solution of the York equation,

$$\tilde{\nabla}^i(\tilde{\nabla}_iX_j+\tilde{\nabla}_jX_i-\frac{2}{3}\tilde{g}_{ij}\tilde{\nabla}^\ell X_\ell)=\tilde{\nabla}^i\tilde{h}_{ij}.$$

Thus, for given $\tilde{g}=\delta_{ij}+o_4(r^{\frac{-3}{2}})$, we select an appropriate $\tilde{k}$ by decomposing any symmetric traceless tensor $h=o_3((r^{\frac{-5}{2}})$ according to the definition above, where X is a solution to the York equation. For details of how to achieve this we refer to [Ch-Mu].

The proof of the Main Theorem, hinges on an elaborate comparison argument with the Minkowski space-time at the level of the three geometric structures with which this is equipped.

- *The canonical space-like foliation* of Minkowski space-time is given by any choice of a one parameter family of parallel space-like hyperplanes, the level sets of the time function $t=x^0=$ const.

- *The null structure* of the Minkowski space-time is specified by one family of future null cones and another of past null cones with vertices on a time-like geodesic orthogonal to the canonical space-like foliation. These families are the level sets of the *optical* functions $u=r-t$ and, respectively, $v=r+t$, where $r=(\Sigma_{i=1}^3|x^i|^2)^{\frac{1}{2}}$. The null vectors $e_+=\partial_t+\partial_r$ and $e_-=\partial_t-\partial_r$ are parallel to their respective gradients and span all the asymptotic null directions.

- *The conformal group structure* is given by the 15 parameter group of translations, Lorentz rotations, scaling and inverted translations. The corresponding infinitesimal generators of the group are,

1. The 4 generators of translations,

$$T_\mu=\partial_\mu,\qquad \mu=0,1,2,3.$$

2. The 6 generators of the Lorentz group,

$$\Omega_{\mu\nu} = x_\mu \partial_\nu - x_\nu \partial_\mu, \qquad \mu, \nu = 0, 1, 2, 3$$

where, $x_\mu = \eta_{\mu\nu} x^\nu$.

3. The scalling vector field,

$$S = x^\mu \partial_\mu$$

4. The 4 acceleration vector fields,

$$K_\mu = -2x_\mu S + < x, x > \partial_\mu, \qquad \mu = 0, 1, 2, 3$$

We recall that the vector fields in the first two groups are Killing while all the others are conformal Killing [17]. In particular the deformation tensors of S and K_0 are given by,

$${}^{(S)}\pi = 2\eta, \qquad {}^{(K_0)}\pi = 4t\eta.$$

As small perturbations of the Minkowski space-time, the solutions of the E-V which we want to construct will mirror the structures outlined above. In other words we construct them together with the following:

- A maximal space-like foliation of the type described above.
- An appropriate defined optical function u whose level surfaces describe the structure of future null infinity.
- A family of almost Killing and conformal Killing vector fields.

[17] A vector field S in a space-time (M,g) is called Killing, resp. conformal Killing, if its deformation tensor ${}^{(X)}\pi = \mathcal{L}_X g$ is zero, resp. proportional to g.

References

[Ba] R. Bartnik.
Existence of maximal surfaces in asymptotically flat Space-Times.
Math. Phys. 94, 1984, 155-175.

[Bo] H. Bondi
Nature 186, 1960, 535

[Bo-Bu-Me] H. Bondi - M.G.J. van der Burg - A.W.K. Metzner.
Gravitational Waves in General Relativity vii.
Proc. Roy. Soc. Lond. A269, 1962, 21-52.

[Br] Y. Bruhat.
Théorème d'existence pour certaines systèmes d'équations
aux dérivées partielles non linéaires.
Acta Mathematica 88 (1952), 141-225.

[Ch] D. Christodoulou.
Global solutions for nonlinear hyperbolic equations for small data.
Comm. Pure Appl. Math. 39, 1986, 267-282.

[Ch-Kl] D. Christodoulou - S. Klainerman.
Asymptotic properties of linear field equations in Minkowski space.
preprint.

[Ch-Mu] D. Christodoulou - N.O'Murchadha.
The Boost Problem in General Relativity.
Comm. Math. Phys. 80, 1981, 271-300.

[Eisen] L.P.Eisenhart.
Riemannian Geometry.
Princeton University Press 1926.

[Ho] L.Hörmander.
On Sobolev spaces associated with some Lie Algebras.
Report NO:4, Inst. Mittag-Leffler 1985.

[Kl1] S. Klainerman.
The null condition and global existence to nonlinear wave equations
Lect. Appl. Math. 23, 1986, 293-326.

[Kl2] S. Klainerman.
Remarks on the global Sobolev Inequalities in Minkowski Space.
Comm. Pure Appl. Math. 40, 1987, 111-117.

[Kl3] S. Klainerman.
Uniform decay estimates and the Lorentz invariance of the classical wave equation.
Comm. Pure Appl. Math. 38, 1985, 321-332.

[Ne-To] E.T.Newman - K.P.Todd.
Asymptotically flat space-times.
General Relativity and Gravitation, vol 2, A. Held, Plenum. 1980

[Pe1] R. Penrose.
Structure of Space-Time.
Battelle Rencontre. C.M.DeWitt and J.A.Wheeler 1967

[Pe2] R. Penrose.
Zero rest mass fields including gravitation: Asymptotic Behaviour.
Proc. Roy. Soc. Lond. A284, 1962, 159-203.

[Sa] R.K.Sacks
Gravitational waves in General Relativity viii.
Proc. Roy. Soc. Lond. A270, 1962, 103-126.

HIGH ORDER REGULARITY FOR SOLUTIONS OF THE INVISCID BURGERS EQUATION

RONALD A. DEVORE
Department of Mathematics
University of South Carolina
Columbia, South Carolina 29208

BRADLEY J. LUCIER
Department of Mathematics
Purdue University
West Lafayette, Indiana 47907

Abstract. We discuss a recent Besov space regularity theory for discontinuous, entropy solutions of quasilinear, scalar hyperbolic conservation laws in one space dimension. This theory is very closely related to rates of approximation in L^1 by moving grid, finite element methods. In addition, we establish the Besov space regularity of solutions of the inviscid Burgers equation; the new aspect of this study is that no assumption is made about the local variation of the initial data.

Key words. regularity, moving grid finite elements, Besov spaces, conservation laws

AMS(MOS) subject classifications. 35L65, 35D10, 41A25, 46E35

1. INTRODUCTION

A regularity theory is developed in [2] and [8] for discontinuous, entropy solutions $u(x,t)$ of the scalar hyperbolic conservation law

$$\begin{aligned} &u_t + f(u)_x = 0, \qquad && x \in \mathsf{R}, \quad t > 0, \\ &u(x,0) = u_0(x), \qquad && x \in \mathsf{R}, \end{aligned} \tag{C}$$

under the assumption that f is uniformly convex and $u_0 \in \mathrm{BV}(\mathsf{R})$ has bounded support. In this theory one measures the regularity of $u(\,\cdot\,,t)$ in Besov spaces $B_q^\alpha(L^p(I))$; functions in these spaces have, roughly speaking, $\alpha > 0$ "derivatives" in $L^p(I)$, where I is a bounded interval, and q is a secondary index of regularity. (See §2 for precise definitions.)

Whereas the solutions of many evolution equations (such as the heat equation) have enough regularity in Sobolev spaces to be approximated to high order in $L^p(I)$ by piecewise polynomial splines defined on *uniform* grids with grid spacing $1/n$, discontinuous solutions of (C) can be approximated by splines on uniform grids to at most $O(n^{-1/p})$ in $L^p(I)$. Thus, if one would like high-order approximation by splines, one is led to consider approximations drawn from the class of piecewise polynomials defined on *arbitrary* grids with n intervals, i.e., *free knot* splines. Such approximations occur in moving grid finite element approximations to time-dependent partial differential equations, such as those used by Miller [9], Glimm et al. [5], and Lucier [7]. The following questions then arise: What regularity is needed to ensure high-order approximation in $L^p(I)$ by free knot splines, and do solutions of (C) maintain this regularity as time progresses? The answers are that regularity in certain Besov spaces is necessary and sufficient for certain orders of approximation by free knot splines, and that solutions of (C) retain this regularity if one considers approximation in L^1.

At this point, it is useful to contrast the approximation properties of functions in Sobolev spaces $W^{\alpha,p}(I)$, $\alpha > 0$, $p > 1$, with those of functions in the Besov spaces

$B_q^\alpha(L^q(I))$, $\alpha > 0$, $q = 1/(\alpha + 1/p)$, $p > 0$. For $u \in L^p(I)$, I a finite interval, $p > 1$, define

$$s_n(u)_p := s_{n,r}(u)_p := \inf_{P \in S_n} \|u - P\|_{L^p(I)},$$

where $S_n := S_{n,r}$ is the set of all piecewise polynomials of degree less than r on a *uniform* grid of size $|I|/n$. (We use the notation $:=$ to mean "is defined as".) Then it can be shown that

$$u \in W^{\alpha,p}(I) \iff \begin{cases} \sup\limits_{n>0} n^\alpha s_n(u)_p < \infty, & \alpha = r \in \mathbf{Z}, \\ \left(\sum\limits_{n=1}^{\infty} [n^\alpha s_n(u)_p]^p n^{-1}\right)^{1/p} < \infty, & r - 1 < \alpha < r \in \mathbf{Z}, \end{cases} \tag{1.1}$$

and that the quantities on the right of (1.1) are equivalent to the seminorm $|u|_{W^{\alpha,p}(I)}$. (For α not an integer, the Sobolev space $W^{\alpha,p}(I)$ is the same as the Besov space $B_p^\alpha(L^p(I))$; see [1, p. 223].)

Recent results of Petrushev [10], [11] and DeVore and Popov [3], [4] provide a characterization of functions that can be approximated to high order by piecewise polynomials in $\Sigma_n := \Sigma_{n,r}$, the set of all piecewise polynomials of degree less than r on *arbitrary* grids with n intervals. Define for $p > 0$

$$\sigma_n(u)_p := \sigma_{n,r}(u)_p := \inf_{P \in \Sigma_n} \|u - P\|_{L^p(I)},$$

and let $q = 1/(\alpha + 1/p)$. Then for $\alpha < r$,

$$u \in B_q^\alpha(L^q(I)) \iff \left(\sum_{n=1}^{\infty} [n^\alpha \sigma_n(u)_p]^q n^{-1}\right)^{1/q} < \infty, \tag{1.2}$$

where the right hand side of (1.2) is equivalent to the "seminorm" $|u|_{B_q^\alpha(L^q(I))}$. (This "seminorm" does not satisfy the triangle inequality if $q < 1$, in which case $B_q^\alpha(L^q(I))$ is not locally convex, but only locally quasiconvex.)

This suggests that perhaps there are certain spaces $X := B_q^\alpha(L^q(I))$ that are regularity spaces for (C); i.e., spaces X for which $u_0 \in X$ implies $u(\cdot, t) \in X$. In this direction, the following general theorem is proved in [2]:

THEOREM 1.1. *Assume that r is a positive integer and that $u_0 \in \mathrm{BV}(\mathbf{R})$ has support in $I := [0, 1]$. Then there exists a constant $C_1 := C_1(r)$ such that the following statements are valid. Let $\Omega = \{y \mid |y| < C_1 \|u_0\|_{L^\infty(\mathbf{R})}\}$. Assume that there is a constant C_2 such that for all $\xi \in \Omega$, $|f^{(r+1)}(\xi)| < C_2$ and $f''(\xi) \geq 1/C_2$. Then for any positive $\alpha < r$ and time $t > 0$ there exists a constant C such that if $u_0 \in B^\alpha(I) := B_q^\alpha(L^q(I))$, where $q = 1/(\alpha+1)$, then $u(\cdot, t)$, the solution of (C), has support in $I_t = [\inf_{\xi\in\Omega} f'(\xi)t, 1 + \sup_{\xi\in\Omega} f'(\xi)t]$ and $\|u(\cdot, t)\|_{B^\alpha(I_t)} \leq C(\|u_0\|_{B^\alpha(I)} + 1)$.*

In this paper we examine the special case of the inviscid Burgers equation,

$$\begin{aligned} &u_t + (u^2)_x = 0, && x \in \mathbf{R}, \quad t > 0, \\ &u(x, 0) = u_0(x), && x \in \mathbf{R}. \end{aligned} \tag{B}$$

In this case, we are able to avoid the requirement that the total variation of u_0 be bounded, and we can show that the Besov space norm of u is bounded independently of time. Furthermore, the proof is simpler. Thus, in §3 we prove the following theorem:

THEOREM 1.2. *Assume that $u_0 \in L^1(\mathbf{R})$ has support in $I := [0,1]$ and that $f(u) = u^2$. Then for any positive α there exists a constant C such that if $u_0 \in B^\alpha(I) := B^\alpha_q(L^q(I))$, where $q = 1/(\alpha+1)$, then $u(\,\cdot\,,t)$ has support in $I_t = [-(8\|u_0\|_{L^1(\mathbf{R})}t)^{1/2}, 1+(8\|u_0\|_{L^1(\mathbf{R})}t)^{1/2}]$ and $\|u(\,\cdot\,,t)\|_{B^\alpha(I_t)} \le C\|u_0\|_{B^\alpha(I)}$.*

2. PRELIMINARIES

In this section, we recall the definition of Besov spaces, present relevant properties of the solutions of (C), and, finally, restate lemmas found in [2] that will be useful here.

Let I be a finite interval. Fix $0 < \alpha < \infty$, $0 < q \le \infty$ and $0 < p < \infty$, and pick an integer $r > \alpha$. Define the $L^p(I)$ modulus of continuity $\omega_r(f,t)_p$ to be the supremum over all $0 < h < t$ of $\|\Delta^r_h f\|_{L^p(I_h)}$, where $I_h = \{x \in I \mid x + rh \in I\}$, and $\Delta^0_h f(x) := f(x)$ and $\Delta^r_h f(x) := \Delta^{r-1}_h f(x+h) - \Delta^{r-1}_h f(x)$. The Besov space $B^\alpha_q(L^p(I))$ is defined to be the set of all functions $f \in L^p(I)$ for which

$$|f|_{B^\alpha_q(L^p(I))} := \left(\int_0^\infty [t^{-\alpha}\omega_r(f,t)_p]^q \, dt/t\right)^{1/q}$$

is finite. Set $\|f\|_{B^\alpha_q(L^p(I))} := \|f\|_{L^p(I)} + |f|_{B^\alpha_q(L^p(I))}$.

We are particularly interested in the spaces $B^\alpha(I) := B^\alpha_q(L^q(I))$, $\alpha > 0$, where $q := 1/(\alpha+1)$. These spaces have the property that if $\alpha' > \alpha$ then $B^{\alpha'}(I)$ is continuously embedded in $B^\alpha(I)$, which in turn is continuously embedded in $L^1(I)$. We define $B^0(I) := L^1(I)$.

The spaces $B^\alpha(I)$, $\alpha > 0$, form a real interpolation family. The real method of interpolation using K-functionals can be described as follows: For any two linear, complete, quasi-normed spaces X_0 and X_1 continuously embedded in a linear Hausdorff topological space X, define the following functional for all f in $X_0 + X_1$:

$$K(f,t,X_0,X_1) := \inf_{f=f_0+f_1} \{\|f_0\|_{X_0} + t\|f_1\|_{X_1}\},$$

where $f_0 \in X_0$ and $f_1 \in X_1$. The new space $X_{\theta,q} := (X_0,X_1)_{\theta,q}$ $(0 < \theta < 1,\ 0 < q \le \infty)$ consists of functions f for which

$$\|f\|_{X_{\theta,q}} := \|f\|_{X_0+X_1} + \left(\int_0^\infty [t^{-\theta}K(f,t,X_0,X_1)]^q \, dt/t\right)^{1/q} < \infty,$$

where $\|f\|_{X_0+X_1} := K(f,1,X_0,X_1)$. DeVore and Popov [3] showed that if $\beta > \gamma > \alpha \ge 0$, $q = 1/(\gamma+1)$, and θ is defined by $\gamma = (1-\theta)\alpha + \theta\beta$, then $(B^\alpha(I), B^\beta(I))_{\theta,q} = B^\gamma(I)$. In particular, $(L^1(I), B^\beta(I))_{\alpha/\beta,1/(\alpha+1)} = B^\alpha(I)$.

As we have noted earlier, the Besov spaces $B^\alpha(I)$ are intimately related to approximation by piecewise polynomials with free knots. For each pair of positive integers n and r, let $\Sigma_n := \Sigma_{n,r}$ denote the collection of all piecewise polynomials on I of degree less than r with at most 2^n pieces. (This is slightly different than in §1.) If f is in $L^1(I)$ and $n \ge 0$, we let

$$\sigma_n(f)_1 := \sigma_{n,r}(f)_1 := \inf_{v \in \Sigma_n} \|f - v\|_{L^1(I)}$$

be the error in approximating f in the $L^1(I)$ norm by the elements of Σ_n; $s_{-1}(f)_1 := \|f\|_{L^1(I)}$. As a special case of (1.2) we have that a function f is in $B^\alpha(I)$ with $\alpha > 0$ if and only if

$$\|f\|_{\mathcal{A}_q^\alpha(L^1(I))} := \left(\sum_{n=-1}^{\infty} (2^{n\alpha}\sigma_n(f)_1)^q \right)^{1/q} < \infty, \tag{2.1}$$

and $\|f\|_{\mathcal{A}_q^\alpha(L^1(I))}$ is equivalent to $\|f\|_{B^\alpha(I)}$. More generally (see [4]), if $\beta > \alpha$ and $0 < q \le \infty$ then $\mathcal{A}_q^\alpha(L^1(I)) = (L^1(I), B^\beta(I))_{\alpha/\beta,q}$. The characterization given here of the equivalence between approximation and regularity is more suited to our present purposes than the one given in §1.

We will now relate certain properties of conservation laws, all of which can be found in the monograph by Lax [6]. When $f(u) = u^2$, (C) is the inviscid Burgers equation, (B). Given x and t, Lax shows that $u(x,t) = u_0(y)$, where $y := y(x,t)$ is a solution (there may be many) of the implicit equation $y = x - 2tu_0(y)$, and furthermore $y(x,t)$ is an increasing function of x for each fixed $t > 0$. It follows that when u_0 has support in $[0,1]$ and is piecewise polynomial of degree less than r with 2^n pieces then $u(x,t)$ is piecewise an algebraic curve in x for each t. On each piece, u satisfies the equation

$$u = P_i(x - 2ut), \tag{2.2}$$

where P_i is one of the polynomial pieces of u_0. Furthermore, u can only have jump discontinuities that decrease. It follows that there are no more than $r2^n$ pieces in the definition of $u(\,\cdot\,,t)$ for all $t > 0$.

Lax also shows that for any $u_0 \in L^1([0,1])$ the support of $u(\,\cdot\,,t)$ is contained in $I_t := [-(8\|u_0\|_{L^1(\mathbf{R})}t)^{1/2}, 1 + (8\|u_0\|_{L^1(\mathbf{R})}t)^{1/2}]$. Also, if $u(x,t)$ and $v(x,t)$ are solutions of (B) with initial data u_0 and v_0 respectively, then

$$\|u(\,\cdot\,,t) - v(\,\cdot\,,t)\|_{L^1(\mathbf{R})} \le \|u_0 - v_0\|_{L^1(\mathbf{R})}. \tag{2.3}$$

Thus, if v_0 is a best piecewise polynomial approximation in $L^1([0,1])$ from Σ_n to u_0 (i.e., $\|u_0 - v_0\|_{L^1(I)} = \sigma_n(u_0)_1$) and $U_n(x,t) := v(x,t)$ is the solution of (B) with initial data v_0, then

$$\|u(\,\cdot\,,t) - U_n(\,\cdot\,,t)\|_{L^1(I_t)} \le \|u_0 - U_n(\,\cdot\,,0)\|_{L^1(I)} = \sigma_n(u_0)_1. \tag{2.4}$$

It will be useful to redefine the values of $U_n(x,t)$ for $x \notin I_t$ to be zero. Then (2.4) remains valid because $u(x,t) = 0$ for $x \notin I_t$.

We will need the following lemmas, which are proved in [2]. Let

$$\|g\|_p^*(I) := \left(\frac{1}{|I|} \int_I |g|^p \right)^{1/p}.$$

LEMMA 2.1 (Equivalence of Norms). *Let ϕ and ψ be defined on an interval I as the functional inverses of polynomials P and Q of degree $\le d$; assume that ϕ and ψ are monotone on I. Then for all* $1 \le p < d/(d-1)$

$$\|\phi - \psi\|_p^*(I) \le C(p,d)\|\phi - \psi\|_1^*(I). \tag{2.5}$$

LEMMA 2.2 (Bounded Oscillation). *Assume that P and Q are polynomials with real coefficients in two variables of total degree less than r. Let ϕ and ψ be functions that are real analytic in the interior of an interval I and satisfy $P(x,\phi)=0$ and $Q(x,\psi)=0$ for $x \in I$. Let $A=\phi-\psi$. Then for $k=0,1,\ldots,r+1$ either $A^{(k)}$ is identically zero on I or $A^{(k)}(x)=0$ has finitely many solutions x in I. The number of solutions depends only on r.*

LEMMA 2.3 (Inverse Inequality). *Let v be twice continuously differentiable on an open interval I and assume that v, v', and v'' each have one sign on I. If numbers p and q are given such that $0<p\leq 1$ and $qp<q-p$, then there exists a constant C such that whenever $v \in L^q(I)$ then $v' \in L^p(I)$ and*

$$\|v'\|_p^*(I) \leq C|I|^{-1}\|v\|_q^*(I). \tag{2.6}$$

3. PROOF OF THEOREM 1.2

If $u_0 \in B^\alpha(I) := B_q^\alpha(L^q(I))$, $q=1/(\alpha+1)$, then by (2.1) u_0 can be approximated well in $L^1(I)$ by piecewise polynomial functions of degree less than r; inequality (2.4) then shows that $u(\cdot,t)$ can be approximated well by piecewise algebraic curves of a certain degree. The proof of Theorem 1.2 consists of showing that good approximation by algebraic curves of the form (2.2) implies $u(\cdot,t)\in B^\alpha(I_t)$.

Assume first that α is less than, but close to, an integer r, and $u_0 \in B^\alpha(I)$. Then by the characterization (2.1), $\sum[2^{n\alpha}\sigma_n(u_0)_1]^q<\infty$. From (2.4) we obtain that $U_n(\cdot,t)$ converges to $u(\cdot,t)$ in $L^1(I_t)$ and therefore

$$u = U_0+\sum_{n=0}^{\infty}(U_{n+1}-U_n)=\sum_{n=-1}^{\infty}T_n,$$

where $T_{-1}:=U_0$ and for later use we define $U_{-1}:=0$.

From the form of the function $U_n(x,t)$ discussed in §2, we can write for $n=-1,0,\ldots$

$$T_n=\sum_{j=1}^{N}A_j, \qquad N\leq C2^n,$$

where C depends on r. Here $A_j=(\phi_j-\psi_j)\chi_j$ with ϕ_j and ψ_j algebraic functions and χ_j the characteristic function of an interval I_j. The intervals I_j, $j=1,\ldots,N$ are piecewise disjoint. We can further assume by Lemma 2.2 that $A_j^{(k)}$ has one sign on I_j for each $k=0,\ldots,r+1$ and $1\leq j\leq N$.

We fix j and measure the smoothness of $A:=A_j$. For this, fix h and consider the sets Γ of all x such that $\{x,x+h,\ldots,x+rh\}\subset I:=I_j$, Γ' of all $x\notin\Gamma$ for which $\{x,x+h,\ldots,x+rh\}\cap I\neq\phi$, and Γ'' of all remaining $x\in\mathbf{R}$.

For $x\in\Gamma''$, $\Delta_h^r(A,x)=0$, so

$$\int_{\Gamma''}|\Delta_h^r(A,x)|^q\,dx=0. \tag{3.1}$$

For $x \in \Gamma'$, $\Delta_h^r(A,x) \le 2^r(|A(x)| + \cdots + |A(x+rh)|)$. Since Γ' has measure no greater than $2r\min(h,|I|)$, we have for a fixed $p > 1$ with $p < 1 + 1/r$, by Hölder's inequality

$$\int_{\Gamma'} |\Delta_h^r(A,x)|^q\,dx \le C[\min(h,|I|)]^{1-q/p}\left(\int_I |A(x)|^p\right)^{q/p}. \tag{3.2}$$

We can write $A = \phi - \psi$ where ϕ is a piece of U_{n+1} and ψ is a piece of U_n. From (2.2), we can write ϕ as

$$\phi = \frac{x - (I + 2tP_1)^{-1}(x)}{2t}, \tag{3.3}$$

where P_1 is one of the polynomial pieces in the definition of $U_{n+1}(0)$; similarly for ψ. Therefore,

$$\begin{aligned} \|\phi - \psi\|_p^*(I) &= \frac{1}{2t}\|(I+2tP_1)^{-1} - (I+2tP_2)^{-1}\|_p^*(I) \\ &\le \frac{C}{2t}\|(I+2tP_1)^{-1} - (I+2tP_2)^{-1}\|_1^*(I) \\ &= C\|\phi - \psi\|_1^*(I). \end{aligned} \tag{3.4}$$

Here the first equality is (3.3) and the inequality that follows is by Lemma 2.1. Therefore, from (3.2) and (3.4) we can conclude that

$$\int_{\Gamma'} |\Delta_h^r(A,x)|^q\,dx \le C[\min(h,|I|)]^{1-q/p}|I|^{-q+q/p}\left(\int_I |A(x)|\,dx\right)^q. \tag{3.5}$$

We next consider $x \in \Gamma$. Because $A^{(r)}$ is monotone on I, we know that for each x there is a ξ such that

$$|\Delta_h^r(A,x)| = C(r)h^r|A^{(r)}(\xi)| \le Ch^r\max(|A^{(r)}(x)|, |A^{(r)}(x+rh)|).$$

Without loss of generality assume that the maximum is attained by the first term. For a number $\epsilon > 0$ to be specified in a moment, let $\alpha_r := \alpha$ and $\alpha_k := \alpha_{k+1} - 1 - \epsilon$, $k = r-1, \ldots, 0$, and let $q_k := 1/(\alpha_k + 1)$. Then by choosing ϵ appropriately, we will have $q_0 = p$, where p is as in (3.4). (Here we must assume that α is close enough to r.) We also have that $0 < q_k \le 1$ for $k = r, \ldots, 1$, and that $q_k q_{k-1} < q_{k-1} - q_k$; therefore, Lemma 2.3 implies that

$$\|A^{(r)}\|_{q_r}^*(I) \le C|I|^{-1}\|A^{(r-1)}\|_{q_{r-1}}^*(I) \le \cdots \le C|I|^{-r}\|A\|_{q_0}^*(I).$$

We then apply (3.4) to find that

$$\begin{aligned} \int_\Gamma |\Delta_h^r(A,x)|^q\,dx &\le Ch^{rq}\int_I |A^{(r)}(x)|^q\,dx \\ &\le Ch^{rq}|I|^{-rq+1}\left(\frac{1}{|I|}\int_I |A(x)|^p\,dx\right)^{q/p} \\ &\le Ch^{rq}|I|^{-rq-q+1}\left(\int_I |A(x)|\,dx\right)^q. \end{aligned} \tag{3.6}$$

Because $\Gamma = \phi$ if $h > |I|/r$, (3.6), (3.5), and (3.1) imply that

$$\int_{\mathbf{R}} |\Delta_h^r(A,x)|^q \, dx \\ \leq C\left([\min(h,|I|)]^{1-q/p}|I|^{-q+q/p} + |I|^{-rq-q+1}h^{rq}\chi(h)\right)\left(\int_I |A(x)|\,dx\right)^q,$$

where χ is the characteristic function of $[0,|I|/r]$. It follows that $\omega_r(A,h)_q^q$ is also less than the right hand side of our latest inequality. Therefore,

$$\begin{aligned}\int_0^\infty & h^{-\alpha q}\omega_r(A,h)_q^q \, dh/h \\ &\leq C\left(|I|^{-q+q/p}\int_0^{|I|} h^{-\alpha q - q/p}\,dh + |I|^{1-q}\int_{|I|}^\infty h^{-\alpha q-1}\,dh\right. \\ &\qquad \left. + |I|^{-rq-q+1}\int_0^{|I|} h^{(r-\alpha)q-1}\,dh\right)\left(\int_I |A(x)|\,dx\right)^q \\ &\leq C|I|^{-\alpha q-q+1}\left(\int_I |A(x)|\,dx\right)^q \\ &= C\left(\int_I |A(x)|\,dx\right)^q,\end{aligned} \tag{3.7}$$

because $-\alpha q - q + 1 = 0$.

We can now estimate the smoothness of $T_n = T_n(\,\cdot\,,t)$. Because $q < 1$, we know that

$$\omega_r(T_n,h)_q^q \leq \sum_{j=1}^N \omega_r(A_j,h)_q^q. \tag{3.8}$$

Hence, (3.7) and Hölder's inequality imply that

$$\begin{aligned}\int_0^\infty h^{-\alpha q}\omega_r(T_n,h)_q^q \, dh/h &\leq C\sum_{j=1}^N \left(\int_{I_j} |A(x)|\,dx\right)^q \\ &\leq CN^{1-q}\|T_n\|_{L^1(I_t)}^q \\ &\leq CN^{\alpha q}\|T_n\|_{L^1(I_t)}^q.\end{aligned} \tag{3.9}$$

Consider now the expression for u, $u(\,\cdot\,,t) = \sum_{n=-1}^\infty T_n$. Using (3.8) and the continuous embedding of $B^\alpha([0,1])$ into $L^1([0,1])$, we obtain

$$\begin{aligned}\int_0^\infty \omega_r(u,h)_q^q h^{-\alpha q-1}\,dh &\leq \sum_{n=-1}^\infty \int_0^\infty \omega_r(T_n,h)_q^q h^{-\alpha q-1}\,dh \\ &\leq C\sum_{n=-1}^\infty 2^{n\alpha q}\|T_n\|_{L^1(I_t)}^q \\ &\leq C\sum_{n=-1}^\infty 2^{n\alpha}\sigma_n(u_0)_1^q \\ &\leq C\|u_0\|_{B^\alpha([0,1])}^q + C\|u_0\|_{L^1([0,1])}^q \\ &\leq C\|u_0\|_{B^\alpha([0,1])}^q,\end{aligned} \tag{3.10}$$

because from (2.3), for $n = -1, 0 \ldots,$

$$\begin{aligned}\|T_n(t)\|_{L^1(I_t)} &= \|U_{n+1}(t) - U_n(t)\|_{L^1(I_t)} \\ &\leq \|U_{n+1}(t) - u(t)\|_{L^1(I_t)} + \|u(t) - U_n(t)\|_{L^1(I_t)} \\ &\leq \|U_{n+1}(0) - u_0\|_{L^1(I_t)} + \|u_0 - U_n(0)\|_{L^1(I_t)} \\ &\leq 2\sigma_n(u_0)_1.\end{aligned}$$

By (3.10), $\|u(\,\cdot\,,t)\|_{B^\alpha(I_t)} \leq C\|u_0\|_{B^\alpha([0,1])}$.

This proves the theorem for α close to r. The proof for other values of $\alpha < r$ can be completed using interpolation; see [2]. □

ACKNOWLEDGMENTS

The first author was supported in part by the National Science Foundation under grant DMS–8620108. The second author was supported in part by the National Science Foundation under grants DMS–8403219 and DMS–8802734.

REFERENCES

[1] R. A. ADAMS, *Sobolev Spaces*, Academic Press, New York, 1975.

[2] R. A. DEVORE AND B. J. LUCIER, *High order regularity for conservation laws*, Purdue University Center for Applied Mathematics, Tech. Rep. 85, Aug. 1988.

[3] R. A. DEVORE AND V. A. POPOV, *Interpolation of Besov spaces*, Trans. Amer. Math. Soc., 305 (1988), pp. 397–414.

[4] ———, *Interpolation spaces and non-linear approximation*, in Function Spaces and Applications, M. Cwikel, J. Peetre, Y. Sagher, and H. Wallin, ed., Springer Lecture Notes in Mathematics, Vol. 1302, Springer-Verlag, New York, 1988, pp. 191–205.

[5] J. GLIMM, B. LINDQUIST, O. MCBRYAN, AND L. PADMANABHAN, *A front tracking reservoir simulator, five-spot validation studies and the water coning problem*, in Mathematics of Reservoir Simulation, R. E. Ewing, ed., SIAM, Philadelphia, 1983.

[6] P. D. LAX, *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*, Regional Conference Series in Applied Mathematics, Vol. 11, SIAM, Philadelphia, 1973.

[7] B. J. LUCIER, *A moving mesh numerical method for hyperbolic conservation laws*, Math. Comp., 46 (1986), pp. 59–69.

[8] ———, *Regularity through approximation for scalar conservation laws*, SIAM J. Math. Anal., 19 (1988), pp. 763–773.

[9] K. MILLER, *Alternate modes to control the nodes in the moving finite element method*, in Adaptive Computational Methods for Partial Differential Equations, I. Babuska, J. Chandra, J. Flaherty, ed., SIAM, Philadelphia, 1983.

[10] P. PETRUSHEV, *Direct and converse theorems for best spline approximation with free knots and Besov spaces*, C. R. Acad. Bulgare Sci., 39 (1986), pp. 25–28.

[11] ———, *Direct and converse theorems for spline and rational approximation and Besov spaces*, in Function Spaces and Applications, M. Cwikel, J. Peetre, Y. Sagher, and H. Wallin, ed., Springer Lecture Notes in Mathematics, Vol. 1302, Springer-Verlag, New York, 1988, pp. 363–377.

Solutions of Quasi-Linear Wave Equations with Small Initial Data.

The Third Phase.

Fritz John

This paper deals with the behavior of solutions of the initial value problem of quasi-linear wave equations in three space dimensions. The solution $u(t, x_1, x_2, x_3) = u(t, x)$ has derivatives denoted by Du with $D_0 = \partial/\partial t$, $D_i = \partial/\partial x_i$ for $i = 1, 2, 3$, combined into a vector $u' = (D_0 u, D_1 u, D_2 u, D_3 u)$. The differential equations in question have the form

$$\Box u = a_{\alpha\beta}(u') D_\alpha D_\beta u = 0 \tag{1a}$$

with

$$\Box = D_0^2 - D_1^2 - D_2^2 - D_3^2; \qquad a_{\alpha\beta} \in C^\infty(R_4); \qquad a_{\alpha\beta}(0) = 0. \tag{1b}$$

We use the summation convention with subscripts α, β, γ, ... always ranging over 0, 1, 2, 3, and i, j, k, ... over 1, 2, 3. For infinitesimal u, (1a) reduces to the classical linear wave equation $\Box u = 0$. We only consider solutions of (1a) which belong to C^∞ in a strip

$$S_\tau : 0 \leq t < \tau, \quad x \in R_3. \tag{1c}$$

The s corresponds to the largest[1] strip S_s to which u can be extended, defines the "life-span" T of u.

We consider solutions u that correspond to prescribed initial values of the form

$$u = \varepsilon f(x), \quad D_0 u = \varepsilon g(x) \qquad \text{for } t = 0, \; x \in R_3 \tag{2a}$$

where f, $g \in C_0(R_3)$, and ε is a positive constant. We are interested in the behavior of solutions with "small" initial values for large times. In

[1] Solutions defined in one strip can only be extended in a unique way as solutions in a larger strip. (See p. 8). Thus T is the supremum, (possibly infinite), of all s with a C^∞-solution defined in S_s which agrees with u in S_τ.

this paper the "smallness requirement" means that ε for a fixed choice of functions f, g, $a_{\alpha\beta}$ is sufficiently small.[2]

For small ε and bounded t the solution u of (1a), (2a) clearly is approximated by $\varepsilon U(t,x)$, where U is the solution of the *linear* problem

$$\Box U = 0 \tag{4a}$$

$$U = f(x), \quad D_0 u = g(x) \qquad \text{for } t = 0. \tag{4b}$$

Here U and its derivatives decay in time like $1/t$. More precisely, following G. Friedlander, we have for fixed f, g, and $x = r\xi$, $r = |x|$, (the euclidean distance from the origin in R_3) that

$$U(t,x) = t^{-1}k(\xi, r-t) + 0(t^{-2}). \tag{5a}$$

Here k can be expressed in terms of plane integrals of f and g:

$$k(\xi,p) = \frac{1}{4\pi}\iint\limits_{y\cdot\xi=p} g(y)\,dS_y - \frac{1}{4\pi}\frac{\partial}{\partial p}\iint\limits_{y\cdot\xi=p} f(y)\,dS_y \tag{5b}$$

for $\xi \in S^2$, $p \in R$. Analogous formulae hold for the derivatives of U. (see [4], pp. 103–4).

It turns out that εU for small ε is a good approximation to the solution of the nonlinear problem (1a), (2a) for a surprisingly long time, namely as long as $\varepsilon \log t$ is negligible compared to 1. During this *first phase* of the evolution of u we have for $n \geq 1$ and $\varepsilon t > 1$

$$D_0^n u(t,x) = \frac{\varepsilon}{t} D_0^n k(\xi, r-t) + 0\left(\frac{\varepsilon^2 \log(1/\varepsilon)}{t}\right) \tag{6a}$$

with similar formulae holding for other derivatives of u. This implies in particular for the life-span T of u in its dependence on ε that

$$\liminf_{\varepsilon\to 0} \varepsilon \log T(\varepsilon) > 0. \tag{6b}$$

(See [1], [2]).

[2] It is not essential that the initial data are *strictly linear* in ε. All results stay valid for more general initial data of the form

$$u = F(\varepsilon,x), \quad D_0 u = G(\varepsilon,x), \qquad \text{for } t = 0,\ x \in R_3 \tag{3a}$$

where F and G are of compact support in x uniformly in ε, and

$$F(0,x) = G(0,x) = 0; \qquad F_\varepsilon(0,x) = f(x), \quad F_\varepsilon(0,x) = g(x). \tag{3b}$$

The interpretation of "smallness" used here is, of course, somewhat artificial. It does not permit to decide when initial data not depending explicitly on a parameter ε are sufficiently small.

For t so large that $\varepsilon \log t$ is of the order 1 the effect of the nonlinear terms in (1a) becomes noticeable, and εU ceases to be a good approximation for u. During this *second phase* the derivatives of u still decay like $1/t$. The duration of this phase is determined by a constant H, that depends on the plane integrals of f, g and on the leading nonlinear terms in (1a). Let

$$z_{\alpha\beta\gamma} = \left(\frac{\partial a_{\alpha\beta}(u')}{\partial D_\gamma u} \right)_{u'=0} . \tag{7a}$$

With a 3-vector $\xi = (\xi_1, \xi_2, \xi_3)$ we associate the 4-vector

$$X = (-1, \xi_1, \xi_2, \xi_3) \tag{7b}$$

and the cubic function

$$Z(\xi) = -\frac{1}{2} z_{\alpha\beta\gamma} X_\alpha X_\beta X_\gamma . \tag{7c}$$

Denote by k', k'', ... successive p-derivatives of the function $k(\xi, p)$ defined by (5b). We then arrive at the fundamental constant[3]

$$H = \sup Z(\xi) k''(\xi, p) \qquad \text{for } \xi \in S^2, \; p \in R. \tag{7d}$$

The second phase lasts as long as

$$\varepsilon H \log t < q \tag{7e}$$

with a fixed $q < 1$. This implies (more precisely than (6b)) that

$$\liminf_{\varepsilon \to 0} \varepsilon \log T(\varepsilon) \geq \frac{1}{H} . \tag{7f}$$

During this second phase the derivatives of u decay approximately to values of order $\exp(-1/H\varepsilon)$. (See [3], [4]).

We shall exclude the case where $H = 0$ from consideration. In that case either u is the trivial solution or the differential equation (1a) satisfies Klainerman's *null-condition*, which implies that u for sufficiently small ε is a "global" solution for all $t > 0$. (See [5], [6]). The subject of the present paper is the *third phase*. During this phase there is a radical change in the behavior of the derivatives of u (except of those of first order). They stop decaying altogether, but instead increase by an amount equal to a

[3] The constant H can be expressed directly in terms of U and the $a_{\alpha\beta}$ by the formula

$$H = \lim_{\delta \to +0} \lim_{t \to \infty} \sup_{|D_0 u| > \delta/t} \frac{t a_{\alpha\beta}(U') D_\alpha D_\beta U}{2 D_0 U} \tag{7g}$$

as follows easily from (5a).

(fractional) power of the previous decay. (The first derivatives continue to decrease like $1/t$). The duration of this phase is comparable to that of the second phase. It can be dated from a time t_1, at which $\sup_x |D_0^2 u(t,x)|$ is of the order of $\exp(-1/\varepsilon H)$ to a time t_2 when $\sup_x |D_0^2 u(t,x)|$ is of order $\exp(-1/\varepsilon H + \theta/\varepsilon H)$ with a positive θ not depending on ε. The third phase is characterized by the fact that the derivatives of u satisfy approximate ordinary differential equations along each ray with direction ξ, essentially without coupling between different rays, as if we were in one space dimension. Moreover the growth of higher derivatives can be described in terms of polynomials in that of the second derivative.

Beyond $t = t_2$, with $\varepsilon \log t$ still close to $1/H$, one would expect $D_0^2 u$ rapidly to become infinite, so that actually

$$\lim_{\varepsilon \to 0} \varepsilon \log T(\varepsilon) = \frac{1}{H}. \tag{8}$$

This has not been proved, except in the special case of *radial* solutions. (See [7]). Generally we can expect a *fourth phase* in which cross effects between different rays (represented by "angular" derivatives) can no more be neglected; it is difficult to deduce their asymptotic behavior in the same way as during the third phase. It is conceivable that for some types of equations and data, blow-up is delayed, or even prevented altogether.[4]

Outline of proof

The methods used are based on those in [4], with the difference that now tu_{tt} is allowed to be large. We have to establish suitable a priori estimates for a finite number of derivatives of u. These are extracted from relations between their L_∞- and L_2-norms, and in the case of second derivatives, also their L_1-norms.

Energy estimates give bounds for the L_2-norms of *higher* derivatives (here those of order ≤ 9) in terms of time integrals of L_∞-norms of *lower* derivatives (those of orders ≤ 5). These estimates can be extended without change to all *generalized* derivatives associated with the d'Alembertian $\Box$, including the *angular* derivatives with respect to the Minkowski metric.

Using Klainerman's extension of Sobolev's inequality [10], we can in turn estimate the L_∞-norms of the lower (generalized) derivatives in terms of the L_2-norms of the higher ones, with additional information on their decay with increasing t or diminishing $|x|$.

A third source of information consists of the approximate ordinary differential equations that describe the evolution of the lower order derivatives

[4] For certain types of equations, blow-up of non-radial solutions with initial data restricted by inequalities had been established, (see [8], [9]), without showing, however, that the life-span $T(\varepsilon)$ satisfies (8).

on each ray in R_3 along "pseudo-characteristic curves", similar to the equations satisfied along characteristics in the one-dimensional case. The error terms in these equations that can be neglected during the third phase (and are partly of higher order), either involve products of decaying quantities (arising from nonlinearities) or angular derivatives multiplied with higher negative powers of t. In this critical period, when $\varepsilon \log t$ is close to $1/H$, it is convenient to measure growth of L_∞- and L_2-norms not in terms of powers of t, but rather in terms of powers of the quantity

$$w(t) = \sup_x Z(x/|x|)tD_0^2u(t,x) \tag{9}$$

which essentially describes the growth of the second derivatives. This is analogous to the one-dimensional situation, where higher derivatives grow like polynomials of second derivatives along characteristics.

An extra consideration is needed to show that the first derivatives of u continue to decay like $1/t$ during the third phase. For that purpose we have to derive bounds for the L_1-norm of $|x|D_0^2u(t,x)$ along a ray.

The estimates derived (which depend on each other) are finally put together in a scheme yielding a priori bounds for all quantities for sufficiently small ε, provided the growth of $w(t)$ does not exceed that of a certain fractional power of t.

Generalized derivatives

Following Klainerman, [11], we introduce the linear differential operators ("generalized derivatives")

$$\begin{array}{lll}
\Gamma_0 = tD_0 + x_iD_i; & \Gamma_1 = x_1D_0 + tD_1; & \Gamma_2 = x_2D_0 + tD_2; \\
\Gamma_3 = x_3D_0 + tD_3; & \Gamma_4 = x_2D_3 - x_3D_2; & \Gamma_5 = x_3D_1 - x_1D_3; \\
\Gamma_6 = x_1D_2 - x_2D_1; & \Gamma_7 = D_0; & \Gamma_8 = D_1; \\
\Gamma_9 = D_2; & \Gamma_{10} = D_3. &
\end{array}$$

Here $\Gamma_1, \ldots, \Gamma_6$ are the generators of the Lorentz group ("angular derivatives" in the Minkowski metric), Γ_0 the generator of the homethetic transformations in R_4, and $\Gamma_7, \ldots, \Gamma_{10}$ those of the translation group. For commutation with the d'Alembertian, we have the rule

$$\Box\Gamma_m - \Gamma_m\Box = 2\delta_{0m}\Box \qquad \text{for } m = 0,\ 1, \ldots, 10. \tag{10a}$$

For multi-indices $A = (A_0, A_1, \ldots, A_{10})$ with non-negative integral elements we define addition in the obvious way, and also

$$|A| = \sum_{m=0}^{10} A_m; \qquad |A|* = \sum_{m=0}^{6} A_m; \qquad A! = \sum_{m=0}^{10} A_m! \tag{10b}$$

Here $A = 0$, if all $A_m = 0$. As usual

$$\Gamma^A = (\Gamma_0)^{A_0}(\Gamma_1)^{A_1} \dots (\Gamma_{10})^{A_{10}}. \tag{10c}$$

In what follows we use the symbol $\overline{\sum}$ to stand for a "finite linear combination with constant numerical coefficients". One easily verifies that

$$\Gamma^A\Gamma^B = \Gamma^{A+B} + \mathop{\overline{\sum}}\limits_{C} \Gamma^C \qquad \text{with } |C| < |A| + |B| \tag{11a}$$

$$\Gamma^A D_\alpha - D_\alpha \Gamma^A = \mathop{\overline{\sum}}\limits_{C,\beta} D_\beta \Gamma^C \qquad \text{with } |C| < |A| \tag{11b}$$

$$\Gamma^A \Box - \Box \Gamma^A = \mathop{\overline{\sum}}\limits_{C} \Gamma^C \Box \qquad \text{with } |C| < |A|, \;\; |C|^* < |A|^* \tag{11c}$$

For scalars $u(t,x)$, $v(t,x)$

$$\Gamma^A(uv) = \sum_{B,C} \frac{A!}{B!\,C!}(\Gamma^B u)(\Gamma^C v) \qquad \text{with } B + C = A. \tag{11d}$$

Setting $d_\alpha = \partial/\partial D_\alpha u$, we have by the chain rule

$$\Gamma^A a_{\alpha\beta}(u') = (d_\gamma a_{\alpha\beta})(\Gamma^A D_\gamma u) + \mathop{\overline{\sum}}\limits_{\gamma, A^{(n)}, \gamma_n} (d^\nu a_{\alpha\beta}) \prod_{n=1}^{\nu} (\Gamma^{A^{(n)}} D_{\gamma_n} u) \tag{11e}$$

where d^ν stands for a monomial of degree ν in the d_γ, and

$$2 \leq \nu \leq |A|; \qquad \sum_{n=1}^{\nu} A^{(n)} = A; \qquad A^{(n)} \neq 0. \tag{11f}$$

Norms and 0-notation

For a scalar $v = v(t,x)$ of compact support in x, uniformly in t, and a non-negative integer N we define the norms

$$|v(t,x)|_N = \sup_A |\Gamma^A(v(t,x)| \qquad \text{with } |A| \leq N \tag{12a}$$

$$|v'(t,x)|_N = \sup_{A,\alpha} |\Gamma^A D_\alpha v(t,x)| \qquad \text{with } |A| \leq N \tag{12b}$$

$$\|v'(t,x)\|_N = \left(\iiint_{R_3} (|v'(t,x)|_N)^2 \, dx_1 \, dx_2 \, dx_3 \right)^{1/2}. \tag{12c}$$

In particular we write

$$|v'(t,x)| = |v'(t,x)|_0 = \sup_\alpha |D_\alpha v(t,x)|. \tag{12d}$$

Given a set S in the two-dimensional pq-plane, (whose definition may involve the functions f, g, $a_{\alpha\beta}$ and ε), we write

$$p = 0(q) \quad \text{in } S \tag{13a}$$

if there exists functionals C and E of f, g, $a_{\alpha\beta}$ such that

$$|p| < Cq \quad \text{for all } (p,q) \in S \quad \text{provided } 0 < \varepsilon < E. \tag{13b}$$

If the set S depends on a parameter M, and E (but not C) also depends on M, we write

$$p = 0_M(q). \tag{13c}$$

Energy inequalities

The assumption f, $g \in C_0^\infty(R_3)$ implies that there exists a smallest S such that

$$f(x) = g(x) = 0 \qquad \text{for } |x| > S. \tag{14a}$$

Then also (see [8], pp. 48–51)

$$u(t,x) = 0 \qquad \text{for } |x| > S + t,\ \ 0 \leq t < T. \tag{14b}$$

For the solution u of (1a), (2a) we define the *linear* differential operator L_u acting on functions $v(t,x)$ by

$$L_u v = \Box v - a_{\alpha\beta}(u')D_\alpha D_\beta v. \tag{15a}$$

For any $v(t,x) \in C^2(t,x)$ and vanishing for $|x| > S + t$ we have the integral identity

$$\frac{d}{dt} \iiint_{R_3} I(t,x)\, dx_1\, dx_2\, dx_3 = \iiint_{R_3} J(t,x)\, dx_1\, dx_2\, dx_3 \tag{15b}$$

where

$$\begin{aligned} I(t,x) &= (D_\alpha v)(D_\alpha v) - a_{00}(D_0 v)^2 + a_{ik}(D_i v)(D_k v) && (15c) \\ J(t,x) &= 2(D_0 v)L_u v) - (D_0 a_{00} + 2D_i a_{i0})(D_0 v)^2 \\ &\qquad - 2(D_k a_{ik})(D_0 v)(D_i v) + (D_0 a_{ik})(D_i v)(D_k v) && (15d) \end{aligned}$$

valid for $0 \leq t < T$, $x \in R_3$. (The $a_{\alpha\beta}$ have argument u'). For $0 \leq t_0 \leq t < T$ then

$$\begin{aligned}\iiint I(t,x)\,dx_1\,dx_2\,dx_3 \\ = \quad \iiint I(t_0,x)\,dx_1\,dx_2\,dx_3 + \int_{t_0}^{t} ds \iiint J(s,x)\,dx_1\,dx_2\,dx_3 .\end{aligned} \tag{15e}$$

Since $a_{\alpha\beta}(0) = 0$, there exists a $\rho > 0$ such that

$$|a_{\alpha\beta}(u')| < \frac{1}{8} \qquad \text{for } |u'| < \rho. \tag{16a}$$

We can choose here

$$\rho = \min\bigl(1, (\sup_{\alpha,\beta,\gamma,\eta} 24\, d_\gamma a_{\alpha\beta}(\eta))^{-1}\bigr) \quad \text{with } \eta = (\eta_0, \eta_1, \eta_2, \eta_3);\ |\eta_\alpha| < 1. \tag{16b}$$

If then

$$|u'(t,x)| < \rho \qquad \text{for } t_0 \leq t < \tau, \quad x \in R_3 \tag{16c}$$

we have

$$\frac{1}{2}|v'|^2 \leq \frac{1}{2}(D_\alpha v)(D_\alpha v) \leq I \leq \frac{3}{2}(D_\alpha v)(D_\alpha v) \leq 6|v'|^2 \tag{16d}$$

and hence from (15e)

$$||v'(t)||^2 \leq 12||v'(t_0)||^2 + 2\int_{t_0}^{t} ds \iiint J(s,x)\,dx_1\,dx_2\,dx_3 . \tag{16e}$$

Uniqueness

If there were two solutions u and u^* of (1a), (2a) with $|u'| < \rho$, $|u^{*\prime}| < \rho$ for $0 \leq t \leq \tau$, $x \in R_3$, we apply (16e) to $v = u^* - u$. Here

$$L_u v = \bigl(a_{\alpha\beta}(u^{*\prime}) - a_{\alpha\beta}(u')\bigr) D_\alpha D_\beta u^* = 0(|v'||u^*|_2)$$

$$J = 0\bigl((|u^*|_2 + |u|_2)|v'|^2\bigr).$$

Since $||v'(0)|| = 0$ we find

$$||v'(t)||^2 = 0\left(\sup\bigl(|u^*(s,x)|_2 + |u(s,x)|_2\bigr) \int_0^t ||v'(s)||^2\,ds\right)$$

for $0 \leq s \leq \tau$, $x \in R_3$, which implies $||v'(t)|| = 0$, and hence $u^* = u$.

Energy inequalities for higher derivatives

We apply the energy inequality (16e) to the function $v = \Gamma^A u$ with $1 \leq |A| \leq N$ for $t_0 \leq t < \tau < T$, assuming that (16c) is satisfied. Since $L_u u = 0$, we have by (15a)

$$\begin{aligned} L_u \Gamma^A u &= (\Box\Gamma^A - \Gamma^A\Box)u - a_{\alpha\beta}(u')(D_\alpha D_\beta \Gamma^A - \Gamma^A D_\alpha D_\beta)u \\ &\qquad - a_{\alpha\beta}(u')\Gamma^A D_\alpha D_\beta u + \Gamma^A a_{\alpha\beta}(u') D_\alpha D_\beta u. \qquad (17a) \end{aligned}$$

Using (11a,b,c,d,e) crudely, we arrive at a representation

$$L_u \Gamma^A u = \overline{\sum_{B,C}} (\Gamma^B a_{\alpha\beta})(\Gamma^C D_\alpha D_\beta u) \qquad \text{with } |B| + |C| \leq |A|, \quad |C| < |A| \qquad (17b)$$

leading to the estimate (using $a_{\alpha\beta}(u') = 0(|u'|)$)

$$L_u \Gamma^A u = 0\left(\emptyset(N) \sup_{\nu, n_m} \prod_{m=1}^{\nu} |u'|_{n_m}\right) \qquad (17c)$$

$$\text{with } \nu \geq 2; \quad n_1 + \cdots + n_\nu \leq N + 1; \quad n_m \leq N \qquad (17d)$$

with a suitable numerical factor $\emptyset(N)$. For $n' = \lfloor (N+1)/2 \rfloor$, (where $\lfloor z \rfloor =$ the largest integer not exceeding z) we find

$$\begin{aligned} L_u \Gamma^A u &= 0(\emptyset(N)(|u'|_{N'} + |u'|^N_{N'})|u'|_N) \qquad (17e) \\ J &= 0(\emptyset(N)(|u'|_{N'} + |u'|^N_{N'})|u'|^2_N) \qquad (17f) \end{aligned}$$

and finally

$$\iiint J(s,x)\, dx_1\, dx_2\, dx_3 = 0(\emptyset(N) a_N(s) ||u'||^2_N) \qquad (17g)$$

where

$$a_N(s) = \sup_x (\emptyset(N)(|u'(s,x)|_{N'}) + |u'(s,x)|^N_{N'}). \qquad (17h)$$

By (16e), (11b)

$$\begin{aligned} ||(\Gamma^A u)'(t)||^2 &= ||v'(t)||^2 \\ &= 0\left(||u'(t_0)||^2_N + \int_{t_0}^{t} a_N(s) ||u'(s)||^2_N\, ds\right) \qquad (17i) \\ ||u'(t)||^2_N &= \iiint |u'(t,x)|^2_N\, dx_1\, dx_2\, dx_3 \\ &= 0(\emptyset(N) \sup_{|A| \leq N} ||(\Gamma^A u)'(t)||^2) \\ &= 0\left(\emptyset(N) ||u'(t_0)||^2 + \int_{t_0}^{t} a_N(s) ||u'(s)||^2_N\, ds\right). \qquad (17j) \end{aligned}$$

Klainerman's inequality

The energy estimates for a solution of (1a), (2a) are supplemented by a general inequality due to S. Klainerman[5] connecting the L_∞-norms and L_2-norms of generalized derivatives of any scalar v:

$$v(t,x) = 0\big((1+t)^{-1/2}(1+t+|x|)^{-1}\|v(t)\|_2\big). \tag{18a}$$

This implies (with a suitable numerical $\emptyset(N)$) that

$$\begin{aligned} |v(t,x)|_N &= 0\big(\emptyset(N)(1+t)^{-1/2}(1+t+|x|)^{-1}\|v(t)\|_{N+2}\big) \\ &= 0\left(\emptyset(N)\frac{1}{1+t}\|v(t)\|_{N+2}\right). \end{aligned} \tag{18b}$$

As an application take inequality (17h) for the case where $N > 5$, and hence $N' + 2 < N$. Then $a_{N'}(s)$ can be estimated in terms of $\|u'(s)\|_{N-1}$. By Gronwall's Lemma[6]we can obtain bounds for $\|u'(s)\|_N$ in terms of those for $\|u'(t_0)\|_N$ and $\|u'(s)\|_{N-1}$. It follows by induction that a priori bounds for $\|u'(t)\|_5$ imply bounds for all $\|u'(t)\|_N$.

The first phase

We apply (17h) for $t_0 = 0$, $N = 8$. Assume also that

$$|u'(t,x)|_4 < 1, \quad |u'(t,x)| < \rho \qquad \text{for } 0 \le t < \tau,\ x \in R_3. \tag{20a}$$

[5] See [10]. For the purposes of the present paper one could get along with Klainerman's earlier inequality connecting $|v|_N$ with $\|v\|_{N+4}$. (See [11]).

[6] Here and elsewhere in this paper we use Gronwall's Lemma in the form: If $z(t) \ge 0$ for $t_0 < t < \tau$ and satisfies

$$z^2(t) \le C\left(z_0^2 + \int_{t_0}^{t} \big(a(s)z^2(s) + b(s)z(s)\big)\,ds\right) \tag{19a}$$

where

$$z_0 > 0; \quad a(s) < a^*(s); \quad b(s) < b^*(s); \quad a^*(s) > 0; \quad b^*(s) > 0$$

then

$$z(t) \le \sqrt{C}\emptyset(t)z_0 + \frac{1}{2}C\int_{t_0}^{t} \frac{\emptyset(t)}{\emptyset(s)} b^*(s)\,ds \tag{19b}$$

with

$$\emptyset(t) = \exp\left(\frac{C}{2}\int_{t_0}^{t} a^*(s)\,ds\right). \tag{19c}$$

Then by (17f), (18b)

$$a_N(s) = 0\left(\frac{1}{1+s}||u'(s)||_8\right). \tag{20b}$$

Since also $||u'(0)||_8 = 0(\varepsilon)$, there exists a C depending on f, g, $a_{\alpha\beta}$ but not on ε or τ, such that

$$||u'(t)||_8^2 \leq C^2\left(\varepsilon^2 + \int_0^t \frac{1}{1+s}||u'(s)||_8^3\, ds\right) \tag{20c}$$

for $0 \leq t \leq \tau$. This implies that

$$||u'(t)||_8 \leq \frac{C\varepsilon}{1 - C^3\varepsilon\log(1+t)/2} = 0(\varepsilon) \tag{20d}$$

provided

$$\varepsilon\log(1+t) < C^{-3}. \tag{20e}$$

It follows from (18b) that

$$|u'(t,x)|_6 = 0\left(\frac{\varepsilon}{1+t}\right)$$

which implies that

$$|u'(t,x)|_4 < \frac{1}{2}; \qquad |u'(t,x)| < \frac{1}{2}\rho \tag{20f}$$

for $0 \leq t < \tau$, $x \in R_3$, provided $\varepsilon < E$ with a suitable E. This proves that

$$||u'(t)||_8 = 0(\varepsilon); \qquad |u'(t,x)|_6 = 0\left(\frac{\varepsilon}{1+t}\right) \tag{20g}$$

for all ε, t satisfying (20e).

Assuming then (20e) to hold, we apply (16e) to the function $v = \Gamma^A(u - \varepsilon U)$, where U is the solution of the linear problem (4a,b). Here by (17c), (20f) for $|A| \leq 8$

$$\begin{aligned} L_u v &= L_u\Gamma^A u - \varepsilon a_{\alpha\beta}D_\alpha D_\beta \Gamma^A U \\ &= 0(|u'|_4|u'|_8 + \varepsilon|u'||U'|_9 && \text{(21a)} \\ J &= 0(|u' - \varepsilon U'|_8|u'|_4|u'|_8 + \varepsilon|u' - \varepsilon U'|_8|u'||U'|_9 + |u'|_1|u' - \varepsilon U'|_8^2) \\ &= 0(|u'|_4|u' - \varepsilon U'|_8^2 + \varepsilon|u'|_4|U'|_9|u' - \varepsilon U'|_8). && \text{(21b)} \end{aligned}$$

Since $||u' - \varepsilon U'||_8 = 0(\varepsilon^2)$ for $t = 0$, we find from (16e) that $z(t) = ||(u - \varepsilon U)'(t)||_8$ satisfies an inequality of the form

$$z(t) = 0\left(\varepsilon^2 + \int_0^t \left(a(s)z^2(s) + b(s)z(s)\right) ds\right)$$

with

$$\begin{aligned} a(s) &= \sup_x |u'(s,x)|_4 = 0\left(\frac{\varepsilon}{1+s}\right) \\ b(s) &= \varepsilon \sup_x (|u'(s,x)|_4 \|U'(s)\|_9) = 0\left(\frac{\varepsilon^2}{1+s}\right). \end{aligned}$$

By (20e)

$$\int_0^t a(s)\,ds = 0\big(\varepsilon \log(1+t)\big) = 0(1).$$

Applying (19b) we find

$$\|(u-\varepsilon U)'(t)\|_8 = 0\big(\varepsilon^2(1+\log(1+t))\big) \tag{21c}$$

for ε, t satisfying (20e). By (18b) we conclude that

$$|u'(t,x) - \varepsilon U'(t,x)|_6 = 0\left(\frac{\varepsilon^2(1+\log(1+t))}{(1+t)(1+\big|t-|x|\big|)^{1/2}}\right) \tag{21d}$$

expression the degree to which u' is approximated by $\varepsilon U'$ during the first phase.

By Huygen's principle $U(t,x) = 0$ for $\big|t-|x|\big| > S$. It follows from (21d) that

$$\int_0^{t+S} r|u'(t,r\xi)|_6\,dr = 0\big(\varepsilon^2(1+\log(1+t))(1+t)^{1/2}\big). \tag{21e}$$

The starting time for our later estimates will be

$$t_0 = \left(\varepsilon \log\frac{1}{\varepsilon}\right)^{-2}. \tag{22a}$$

We notice that at that time by (20g), (21c), (21d), (21e)

$$\|u'(t_0)\|_8 = 0(\varepsilon); \qquad |u'(t_0,x)|_6 = 0\left(\frac{\varepsilon}{1+t_0}\right) \tag{22b}$$

$$|u'(t_0,x) - \varepsilon U'(t_0,x)|_6 = 0\left(\frac{\varepsilon^2\log(1/\varepsilon)}{(1+t_0)(1+\big|t_0-|x|\big|)^{1/2}}\right) \tag{22c}$$

$$\int_0^\infty r|u'(t_0,r\xi)|_6\,dr = 0(\varepsilon). \tag{22d}$$

In particular we find from (22c), (6a) that for $x = r\xi$

$$t_0 D_0^2 u(t_0,x) = k''(\xi, r-t_0) + 0\left(\varepsilon^2 \log\frac{1}{\varepsilon}\right) \tag{23a}$$

and hence by (9), (7d)

$$w(t_0) = \varepsilon H + 0\left(\varepsilon^2 \log\frac{1}{\varepsilon}\right). \tag{23b}$$

This implies that

$$\frac{\varepsilon}{w(t_0)} = \frac{1}{H} + 0\left(\varepsilon\log\frac{1}{\varepsilon}\right) = 0(1). \tag{23c}$$

L_2-estimates in the second and third phase

Let the function $w(t)$ be defined by (9). We write w_0 for $w(t_0)$, where t_0 is the number given by (22a). We assume that in the interval $t_0 \leq t < \tau$ the function $w(t)$ is Lipschitz continuous with a derivative that satisfies

$$\frac{dw}{dt} > \frac{w^2}{2t} \tag{24a}$$

wherever it exists. Assume that in the same interval

$$|u'(t,x)| < \rho; \quad |u'(t,x)|_m < 1 \qquad \text{for } m = 1,\, 2,\, 3,\, 4 \tag{24b}$$

$$|u'(t,x)|_m < \frac{M\varepsilon}{t}\left(\frac{w(t)}{w_0}\right)^{2m-1} \qquad \text{for } m = 1,\, 2,\, 3,\, 4 \tag{24c}$$

with a certain $M > 0$. We shall prove then by induction over N that for $t_0 \leq t < \tau$

$$||u'(t)||_N = 0\left((1+M)^N \varepsilon\left(\frac{w(t)}{w_0}\right)^{\mu M + 2N}\right) \tag{24d}$$

for $N = 0,\, 1,\, 2,\, \ldots,\, 8$, with a constant μ satisfying

$$\mu = 0(1). \tag{24e}$$

For the proof we apply (16a) to $v = \Gamma^A u$ with $|A| \leq N \leq 8$. By (17b) and assumption (24b)

$$L_u \Gamma^A u = 0(\sup_{n_1,n_2} |u'|_{n_1} |u'|_{n_2}) \qquad \text{with } n_1,\, n_2 \leq N,\;\; n_1 + n_2 \leq N+1.$$

It follows from (15d) that

$$J = 0(|u'|_1 u'|_N^2 + \sup_n |u'|_n |u'|_{N+1-n} |u'|_N)$$

with $2 \leq n \leq (N+1)/2$. (The last term only occurs for $N \geq 2$). Thus

$$\begin{aligned}\iiint & J\, dx_1\, dx_2\, dx_3 \\ &= 0\big(\sup_x |u'(t,x)|_1 ||u'(t)||_N^2 + \sup_{x,n} |u'(t,x)|_n ||u'(t)||_N ||u'(t)||_{N+1-n}\big)\end{aligned}$$

with $2 \leq n \leq (N+1)/2$. It follows from (16e) that

$$||u'(t)||_N^2 \leq C\left(||u'(t_0)||_N^2 + \int_{t_0}^{t} \left(a(s)||u'(s)||_N^2 + b_N(s)||u'(s)||_N\right) ds\right) \quad (25a)$$

with

$$a(s) = \sup_x |u'(s,x)|_1 \quad (25b)$$

$$b_N(s) = \sup_{x,n} |u'(t,x)|_n ||u'(t)||_{N+1-n} \qquad \text{with } 2 \leq n \leq \frac{N+1}{2} \quad (25c)$$

so that $b_N(s) = 0$ for $N = 0,\ 1$. By assumptions (24a,c)

$$a(s) \leq \frac{M\varepsilon}{s}\frac{w(s)}{w_0} \leq \frac{2M\varepsilon}{w_0}\frac{d\log w(s)}{ds} = a^*(s).$$

Hence here

$$\Phi(t) = \exp\left(\frac{C}{2}\int_{t_0}^{t} a^*(s)\,ds\right) = \left(\frac{w(t)}{w_0}\right)^{\mu M} \quad (25d)$$

with

$$\mu = \frac{C\varepsilon}{w_0} = \frac{C}{H} + 0\left(\varepsilon\log\frac{1}{\varepsilon}\right) = 0(1) \quad (25e)$$

by (23c). Since also $||u'(t_0)||_N = 0(\varepsilon)$, it follows from (19b) that (24d) holds for $N = 0,\ 1$.

Assume that (24d) has been established already with N replaced by $N-1$, where $N \geq 2$. Since $n \leq \left((N+1)/2\right) \leq 4$, and $N+1-n \leq N-1$ we find from (24d), (24c), (25c) that

$$b_N(s) = 0\left(\varepsilon^2 s^{-1}(1+M)^N \left(\frac{w(s)}{w_0}\right)^{\mu M + 2N + 1}\right).$$

Using (25d), (24a), (23c) we conclude from Gronwall's lemma (19b) that (24d) also holds for N, thus completing the proof of (24d) by induction.

Restrictions on the growth of w

We showed that (24d) holds, whenever assumptions (24a,b,c) are satisfied for $t_0 \leq t < \tau,\ x \in R_3$. We now introduce a further assumption (really a restriction on τ), namely that

$$\frac{w(t)}{w_0} < 2\left(\frac{t}{t_0}\right)^{\theta} \qquad \text{for } t_0 \leq t < \tau,\ x \in R_3 \quad (26a)$$

where

$$\theta = \theta(M) = \frac{1}{10}\left(\frac{CM}{H} + 21\right)^{-1}. \quad (26b)$$

Here C is the constant (depending only on f, g, $a_{\alpha\beta}$) entering (20c), H is defined by (7d), and M taken from assumptions (24c). Our present aim is to show that then

$$|u'(t,x)| = 0_M\left(\frac{\varepsilon}{t}\right) \tag{27a}$$

$$|u'(t,x)|_m = 0_M\left(\frac{\varepsilon}{t}\left(\frac{w(t)}{w_0}\right)^{2m-1}\right) \qquad \text{for } m = 1,\ 2,\ 3,\ 4. \tag{27b}$$

Using (23b) we observe that (27b) implies that

$$(\mu M + 21)\theta < \frac{1}{10} + 0\left(\frac{H}{C}\varepsilon\log\frac{1}{\varepsilon}\right) < \frac{1}{8} \tag{28a}$$

for sufficiently small ε. By (24d), (18b) for $x = r\xi$

$$\begin{aligned} \|u'(t)\|_8 &= 0\left((1+M)^8\varepsilon\left(\frac{w(t)}{w_0}\right)^{\mu M+16}\right) \\ &= 0\left((1+M)^8 2^{\mu M+16}\left(\frac{t}{t_0}\right)^{(\mu M+16)\theta}\right) \end{aligned} \tag{28b}$$

$$\begin{aligned} |u'(t,x)|_6 &= 0\left(\frac{(1+M)^8\varepsilon}{t(1+|t-r|)^{1/2}}\left(\frac{w(t)}{w_0}\right)^{\mu M+16}\right) \\ &= 0\left(\frac{(1+M)^8 2^{\mu M+16}\varepsilon}{t(1+|t-r|)^{1/2}}\left(\frac{t}{t_0}\right)^{(\mu M+16)\theta}\right). \end{aligned} \tag{28c}$$

Estimates in the noncritical zone

We divide the region $t_0 \leq t < \tau$, $0 \leq r < t + S$ in the rt-plane into a *noncritical* zone

$$0 \leq r \leq t - t^{1/4}, \qquad t_0 \leq t < \tau \tag{29a}$$

and a *critical* zone

$$t - t^{1/4} < r < t + S, \qquad t_0 \leq t < \tau. \tag{29b}$$

In the noncritical zone (29a) we can immediately verify (27a,b). Indeed by (28c), (29a), (22a), (28a) we have

$$\begin{aligned} |u'(t,x)|_6 &= 0\left(\frac{\varepsilon}{t}(1+M)^8 2^{\mu M+16} t^{-1/8}\left(\frac{t}{t_0}\right)^{(\mu M+16)\theta}\right) \\ &= 0\left(\frac{\varepsilon}{t}(1+M)^8 2^{\mu M+16}\left(\varepsilon\log\frac{1}{\varepsilon}\right)^{1/4}\right) = 0_M\left(\frac{\varepsilon}{t}\right). \end{aligned} \tag{30}$$

The approximate differential equations along pseudo-characteristics in the critical region

To establish (27a,b) also in the critical region we introduce for each direction ξ the curves given by the ordinary differential equation

$$\frac{dr}{dt} = c = 1 + \frac{1}{2} a_{\alpha\beta}(u') X_\alpha X_\beta \tag{31}$$

in the rt-plane, with $u' = u'(t, r\xi)$, and the X_α defined by (7b). We call these curves *pseudo-characteristics*.[7]

By (30) we have on the separating line $r = t - t^{1/4}$

$$1 - c = 0(u') = 0_M \left(\frac{\varepsilon}{t}\right) < \frac{1}{4} t^{-3/4} \tag{32}$$

for $\varepsilon < E(M)$ with a suitable $E(M)$. As a consequence, a pseudo-characteristic continued backwards from a point in the critical region either intersects the line $r = t - t^{1/4}$ in a point with $t > t_0$, or intersects the line $t = t_0$ in a point with $r > t_0 - t_0^{1/4}$. A forward pseudo-characteristic from a point in the critical region does not leave that region for $t < \tau$.

The definitions of the generalized derivatives $\Gamma_0, \ldots, \Gamma_{10}$ give rise to the identities

$$\begin{aligned} D_i &= -\xi_i D_0 + \frac{1}{t}\Gamma_i + \frac{\xi_i}{t+r}\Gamma_0 - \frac{r\xi_i\xi_k}{t(t+r)}\Gamma_k && \text{(33a)} \\ D_0 &= \frac{1}{t^2 - r^2}(t\Gamma_0 - x_i\Gamma_i). && \text{(33b)} \end{aligned}$$

It follows that for a function $v(t, x) = v(t, r\xi)$ in the critical zone (29b)

$$\begin{aligned} D_\alpha v &= -X_\alpha D_0 v + 0\left(\frac{1}{t}|v|_1\right) && \text{(33c)} \\ D_\alpha v &= 0\left(\frac{1}{|t-r|}|v|_1\right). && \text{(33d)} \end{aligned}$$

We introduce the *radial* derivative

$$R = \xi_i D_i = \frac{1}{r} x_i D_i = \frac{d}{dr}. \tag{33e}$$

Then

$$(D_0 + R)v = 0\left(\frac{1}{t}|v|_1\right). \tag{33f}$$

[7] Up to terms of higher order in u' the relation $dr/dt = c$ represents the characteristics for the operator L_u when applied to functions v depending only on r and t.

One verifies easily the further relations

$$D_\alpha D_\beta = X_\alpha X_\beta D_0^2 v + 0\left(\frac{1}{t}|v'|_1\right) = X_\alpha X_\beta R^2 v + 0\left(\frac{1}{t}|v'|_1\right) \tag{33g}$$

$$(D_0 + R)^2 v = 0\left(\frac{1}{t^2}|v|_2\right). \tag{33h}$$

The operator $L_u v$ can be written in the form

$$\begin{aligned} L_u v &= \Box v - a_{\alpha\beta}(u') D_\alpha D_\beta v \\ &= \frac{2}{t}(D_0 + cR) t D_0 v - (D_0 + R)^2 v - \frac{2}{r}(D_0 + R) v - 2\frac{t-r}{rt} D_0 v \\ &\quad - 2(c-1)(R + D_0) D_0 v - \frac{\delta_{ik} - \xi_i \xi_k}{t^2} \Gamma_i \Gamma_k v + \frac{2\xi_i}{rt} \Gamma_i v \\ &\quad - a_{\alpha\beta}(D_\alpha D_\beta v - X_\alpha X_\beta D_0^2 v). \end{aligned} \tag{34}$$

Thus along a pseudo-characteristic

$$\frac{d}{dt} t D_0 v = (D_0 + cR) t D_0 v = \frac{1}{2} t L_u v + 0\left(\left(\frac{1}{t} + |u'|\right)|v|_2\right). \tag{35a}$$

We shall apply this equation to $v = \Gamma^A u$ with $1 \le |A| \le 4$. In that case for $x = r\xi$ we have by (11a), (12a), (18b), (28b)

$$\begin{aligned} |v(t,x)|_2 &= 0(|u(t,x)|_6) = 0(\sup_{|B|\le 6} |\Gamma^B u(t,x)|) \\ &= 0\left(\sup_{|B|\le 6} \int_r^{t+S} |R\Gamma^B u(t, s\xi)|\, ds\right) \\ &= 0\left(\int_r^{t+S} |u(t,s\xi)|_6\, ds\right) = 0\left(\frac{1}{t}\|u'(t)\|_8 \int_{t-t^{1/4}}^{t+S} \frac{ds}{(1+|t-s|)^{1/2}}\right) \\ &= 0(t^{-7/8}\|u'(t)\|_8) = 0\left(\varepsilon(1+M)^8 t^{-7/8}\left(\frac{w(t)}{w_0}\right)^{\mu M + 16}\right) \end{aligned} \tag{35b}$$

$$\begin{aligned} \frac{1}{t} + |u'| &\le \frac{1}{t} + |u'|_1 = 0\left(\frac{1}{t}\left(1 + \varepsilon M \frac{w(t)}{w_0}\right)\right) \\ &= 0\left(\frac{1}{t}(1 + M\varepsilon)\frac{w(t)}{w_0}\right) = 0_M\left(\frac{1}{t}\frac{w(t)}{w_0}\right). \end{aligned} \tag{35c}$$

Estimates for $tD_0^2 u$ and w

We apply (35a) to $v = D_0 u$. Here by (7a,c),(33g)

$$L_u D_0 = (D_0 a_{\alpha\beta})(D_\alpha D_\beta u) = 2Z(\xi)(tD_0^2 u)^2 + 0\left(\left(\frac{1}{t} + |u'|\right)|u'|_1^2\right).$$

Hence by (35a)

$$\frac{d}{dt} tD_0^2 u = \frac{1}{t} Z(\xi)(tD_0^2 u)^2 + b \tag{36a}$$

where by (24c)

$$\begin{aligned} b &= 0\left(t \left(\frac{1}{t} + |u'|_1 \right)^2 |u'|_2 \right) = 0\left(\frac{\varepsilon M}{t^2} \left(1 + M\varepsilon \frac{w}{w_0} \right)^2 \left(\frac{w}{w_0} \right)^3 \right) \\ &= 0\left(\frac{\varepsilon M}{t^2} (1 + M\varepsilon)^2 \left(\frac{w}{w_0} \right)^5 \right) = 0_M \left(\frac{\varepsilon M}{t^2} \left(\frac{w}{w_0} \right)^5 \right). \end{aligned} \tag{36b}$$

Then also

$$\frac{d}{dt} Z(\xi) tD_0^2 u = \frac{1}{t} \left(Z(\xi) tD_0^2 u \right)^2 + 0_M \left(\frac{\varepsilon M}{t^2} \left(\frac{w}{w_0} \right)^5 \right). \tag{36c}$$

By definition there exists a unit vector ξ^0 and a number r_0 such that

$$w_0 = Z(\xi_0) t_0 D_0^2 u(t_0, r_0 \xi^0). \tag{37a}$$

Here $r_0 > t_0 - S > t_0 - t_0^{1/4}$ by (23a,b), since $k''(\xi, p) = 0$ for $|p| > S$. We can apply (36c) along the pseudo-characteristic issuing from the point $(t_0, r_0 \xi^0)$, which lies in the critical zone. By (23b), (26a,b)

$$\begin{aligned} 0_M \left(\frac{\varepsilon M}{t^2} \left(\frac{w}{w_0} \right)^5 \right) &= 0_M \left(\frac{\varepsilon M 2^{5\theta}}{t^{2-5\theta} t_0^{5\theta}} \right) = 0_M \left(\frac{\varepsilon M}{t t_0} \right) \\ &= 0_M \left(\frac{1}{t} \varepsilon^3 M \log^2 \frac{1}{\varepsilon} \right) < \frac{1}{t} w_0^2 \end{aligned} \tag{37b}$$

for ε less than a suitable $E(M)$. Hence the inequality

$$Z(\xi) tD_0^2 u(t, r\xi) > w_0$$

persists all along the pseudo-characteristic.

Then also

$$w(t) = \sup_{\xi, r} Z(\xi) tD_0^2 u(t, r\xi) > w_0 \tag{37c}$$

for $t_0 < t < \tau$. In the noncritical zone we have by (30), (28a), (22a)

$$\begin{aligned} Z(\xi) tD_0^2 u &= 0(t|u'|_6) = 0\left(\varepsilon(1+M)^8 2^{\mu M + 16} t^{-1/8} \left(\frac{t}{t_0} \right)^{(\mu M + 16)\theta} \right) \\ &= 0_M \left(\varepsilon(1+M)^8 2^{\mu M + 16} t^{-1/8} \right) < w_0. \end{aligned} \tag{37d}$$

It follows that $Z(\xi) tD_0^2 u$ for fixed t assumes its supremum $w(t)$ only at points $x = r\xi$, for which (r, t) lies in the interior of the critical zone.

By definition (37c) $w(t)$ is locally Lipschitz continuous for $t_0 < t < \tau$, since $Z(\xi)tD_0^2u(t,r\xi)$ is Lipschitz continuous in t. Let for a certain t

$$w(t) = Z(\eta)tD_0^2(t,\sigma\eta) \tag{38a}$$

with a certain unit vector η and $\sigma > t - t^{1/4}$. For $h > 0$ the increase in w from t to $t+h$ can not be less than the increase in $Z(\xi)tD_0^2u(t,r\xi)$ along the pseudo-characteristics through the point $(t,\sigma\eta)$. For $h \to 0$ we find from (36c) that

$$\liminf_{h\to+0} \frac{w(t+h)-w(t)}{h} \geq \frac{1}{t}w^2(t) + 0_M\left(\frac{\varepsilon M}{t^2}\left(\frac{w(t)}{w_0}\right)^5\right) > 0. \tag{38b}$$

This shows that $w(t)$ is a monotone increasing absolutely continuous function of t. Similarly we find for $h > 0$ that the increase in w from $t-h$ to t can not exceed the increase in $Z(\xi)tD_0^2u(t,r\xi)$ along the pseudo-characteristic above. Hence

$$\liminf_{h\to+0} \frac{w(t)-w(t-h)}{h} \leq \frac{1}{t}w^2(t) + 0_M\left(\frac{\varepsilon M}{t^2}\left(\frac{w(t)}{w_0}\right)^5\right). \tag{38c}$$

It follows that

$$\frac{dw}{dt} = \frac{1}{t}w^2 + 0_M\left(\frac{\varepsilon M}{t^2}\left(\frac{w}{w_0}\right)^5\right) \tag{38d}$$

whenever dw/dt exists.

Returning to the differential equation (36a) we see that

$$\frac{d}{dt}tD_0^2u = atD_0^2u + b \tag{39a}$$

where

$$a = \frac{1}{t}Z(\xi)tD_0^2u \leq \frac{1}{t}w = \frac{1}{w}\frac{dw}{dt} + 0_M\left(\frac{\varepsilon M}{t^2 w_0}\left(\frac{w}{w_0}\right)^4\right). \tag{39b}$$

The pseudo-characteristics in the critical zone originate at points $(t_1, r_1\xi^1)$ where either $r_1 = t_1 - t_1^{1/4}$, $t_1 > t_0$, or $t_1 = t_0$, $r_1 \geq t_0 - t_0^{1/4}$. In either case by (30) or (20g)

$$t_1 D_0^2 u(t_1, r_1\xi^1) = 0_M(\varepsilon). \tag{39c}$$

Since also by (39b), (36b) for $t_1 < t_2 < t$

$$\begin{aligned} \int_{t_2}^t a\,ds &\leq \log\frac{w(t)}{w(t_2)} + 0_M\left(\int_{t_2}^t \frac{M2^{4\theta}}{w_0 s^{2-4\theta}t_0^{4\theta}}\,ds\right) \\ &\leq \log\frac{w(t)}{w_0} + 0_M\left(\frac{\varepsilon M}{w_0 t_0}\right) = \log\frac{w(t)}{w_0} + 0_M(1) \qquad (39d) \\ \int_{t_2}^t |b|\,ds &= 0_M\left(\int_{t_2}^t \frac{\varepsilon M 2^{5\theta}}{s^{2-5\theta}t_0^{5\theta}}\,ds\right) = 0_M(\varepsilon) \end{aligned}$$

we find that

$$tD_0^2u(t,x) = 0_M\left(\varepsilon\frac{w(t)}{w_0}\right). \tag{39e}$$

More generally we have then from (24c), (33g) that

$$\begin{aligned} tD_\alpha D_\beta u &= X_\alpha X_\beta tD_0^2u + 0(|u'|_1) \\ &= 0_M\left(\varepsilon\frac{w(t)}{w_0}\right) + 0\left(\frac{M\varepsilon}{t}\frac{w(t)}{w_0}\right) = 0_M\left(\varepsilon\frac{w(t)}{w_0}\right). \end{aligned} \tag{39f}$$

Estimates for $|u'|$

In order to prove (27a) in the critical zone, it is sufficient to show that

$$\int_0^{t+S} p|D_\alpha D_\beta u(t,p\xi)|\,dp = 0_M(\varepsilon). \tag{40a}$$

Indeed, (40a) implies for $r > t - t^{1/4}$ that

$$\begin{aligned} tD_\beta u(t,r\xi) &= -\frac{t}{r}\int_r^{t+S} r\xi_i D_i D_\beta u(t,p\xi)\,dp \\ &= 0\left(\sup_{\alpha,\beta}\int_0^{t+S} p|D_\alpha D_\beta u(t,p\xi)|\,dp\right) = 0_M(\varepsilon). \end{aligned} \tag{40b}$$

In order to establish (40a) we rewrite the identity (see (35a))

$$\begin{aligned} \frac{d}{dt}tD_0^2u &= \frac{t}{2}L_uD_0u + 0\left(\left(\frac{1}{t} + |u'|\right)|u'|_2\right) \\ &= \frac{t}{2}(D_0a_{\alpha\beta})(D_\alpha D_\beta u) + 0\left(\left(\frac{1}{t} + |u'|\right)|u'|_2\right) \end{aligned} \tag{40c}$$

in the form

$$\begin{aligned} &D_0\left(\left(1 - \frac{1}{2}a_{\alpha\beta}X_\alpha X_\beta\right)(r|D_0^2u|)\right) + R(r|D_0^2u|) \\ &= \frac{r}{2}\left((D_0a_{\alpha\beta})(D_\alpha D_\beta - X_\alpha X_\beta D_0^2)u - a_{\alpha\beta}X_\alpha X_\beta(R + D_0)D_0^2u\right)\operatorname{sgn} D_0^2u \\ &\qquad + 0\left(\left(\frac{1}{t} + |u'|\right)|u'|_2\right) \\ &= 0\left(\left(\frac{1}{t} + |u'|_1\right)|u'|_2\right) = 0_M\left(\frac{1}{t}\frac{w(t)}{w_0}\frac{\|u'(t)\|_4}{t(1 + |t - r|)^{1/2}}\right) \\ &= 0_M\left(\frac{\varepsilon(1 + M)^5 2^{\mu M + 10}}{t^2(1 + |t - r|)^{1/2}}\left(\frac{t}{t_0}\right)^{(\mu M + 10)\theta}\right). \end{aligned} \tag{40d}$$

(See (35c), (24d)). By (16a), (24b)

$$\frac{1}{2} < 1 - \frac{1}{2}a_{\alpha\beta}(u')X_\alpha X_\beta < \frac{3}{2}. \tag{40e}$$

Integrating (40d) over t and r, it follows from (40e), (21e) that

$$\begin{aligned}
&\int_0^{t+S} r|D_0^2 u(t,r\xi)|\,dr \\
&= 0(\varepsilon) + 0_M\left(\int_{t_0}^{t} ds \int_0^{s+S} \frac{\varepsilon(1+M)^5 2^{\mu M+10}}{s^2(1+|s-p|)^{1/2}}\left(\frac{s}{t_0}\right)^{(\mu M+10)\theta} dp\right) \\
&= 0(\varepsilon) + 0_M\left(\int_{t_0}^{t} \varepsilon(1+M)^5 2^{\mu M+10} s^{-3/2}\left(\frac{s}{t_0}\right)^{(\mu M+10)\theta} ds\right) \\
&= 0(\varepsilon) + 0_M(\varepsilon(1+M)^5 2^{\mu M+10} t_0^{-1/2}) = 0_M(\varepsilon).
\end{aligned} \tag{40f}$$

Consequently by (33g), (28c)

$$\begin{aligned}
&\int_0^{t+S} r|D_\alpha D_\beta u(t,r\xi)|\,dr \\
&\qquad = \int_0^{t+S} r|X_\alpha X_\beta D_0^2 u(t,r\xi)|\,dr + 0\left(\int_0^{t+S} \frac{r}{t}|u'(t,r\xi)|_1\,dr\right) \\
&\qquad = 0_M\left(\varepsilon + \varepsilon t^{-1/2}(1+M)^8 2^{\mu M+16}\left(\frac{t}{t_0}\right)^{(\mu M+16)\theta}\right) = 0_M(\varepsilon).
\end{aligned}$$

This completes the proof of (27a).

Estimates for ordinary derivatives

We shall prove that

$$D_0^{n+1}u = 0_M\left(\frac{\varepsilon}{t}\left(\frac{w}{w_0}\right)\right)^{2n-1} \tag{41a}$$

for $2 \le n \le 4$ by induction over n. (It holds for $n = 1$ by (39e)). (41a) implies by (33g), (24c), (26a,b) that also

$$\begin{aligned}
D_0^{n-1}D_\alpha D_\beta u &= 0\left(|D_0^{n+1}u| + \frac{1}{t}|u'|_n\right) \\
&= 0_M\left(\frac{\varepsilon}{t}\left(\frac{w}{w_0}\right)^{2n-1} + \frac{M\varepsilon}{t^2}\left(\frac{w}{w_0}\right)^{2n-1}\right) \\
&= 0_M\left(\frac{\varepsilon}{t}\left(\frac{w}{w_0}\right)^{2n-1}\right).
\end{aligned} \tag{41b}$$

Let then $n \geq 2$ and

$$D_0^{N-1} D_\alpha D_\beta u = 0_M \left(\frac{\varepsilon}{t} \left(\frac{w}{w_0} \right)^{2N-1} \right) \qquad \text{for } 1 \leq N < n. \qquad (41c)$$

By (17a)

$$L_u D_0^n u = \sum_{p=1}^{n} \binom{n}{p} (D_0^p a_{\alpha\beta})(D_0^{n-p} D_\alpha D_\beta u). \qquad (41d)$$

Generally by (11e) for $|A| \leq 4$

$$\Gamma^A a_{\alpha\beta}(u') = (d_\gamma a_{\alpha\beta})(\Gamma^A D_\gamma u) + 0(|u'|_4^2). \qquad (41e)$$

Thus for $2 \leq p \leq n-1$ by (41c)

$$(D_0^p a_{\alpha\beta})(D_0^{n-p} D_\alpha D_\beta u) = 0_M \left(\frac{\varepsilon^2}{t^2} \left(\frac{w}{w_0} \right)^{2n} + |u'|_4^3 \right).$$

Hence by (33g), (7c)

$$\begin{aligned} L_u D_0^n u &= (d_\gamma a_{\alpha\beta}) \big((D_0^n D_\gamma u) + n (D_0 D_\gamma u)(D_0^{n-1} D_\alpha D_\beta u) \big) \\ &\quad + 0_M \left((1 - \delta_{n2}) \frac{\varepsilon^2}{t^2} \left(\frac{w}{w_0} \right)^{2n-1} + |u'|_4^3 \right) \\ &= 2(n+1) Z(\xi) (D_0^2 u)(D_0^{n+1} u) \\ &\quad + 0_M \left(|u'|_4^3 + \frac{1}{t} |u'|_1 |u'|_4 + (1 - \delta_{n2}) \frac{\varepsilon^2}{t^2} \left(\frac{w}{w_0} \right)^{2n} \right) \\ &= 2(n+1) Z(\xi) (D_0^2 u)(D_0^{n+1} u) \\ &\quad + 0_M \left((1+M)^3 \frac{\varepsilon^2}{t^3} \left(\frac{w}{w_0} \right)^{21} + (1 - \delta_{n2}) \frac{\varepsilon^2}{t^2} \left(\frac{w}{w_0} \right)^{2n} \right) \end{aligned} \qquad (42a)$$

It follows from (35a), (28c) that

$$\frac{d}{dt} t D_0^{n+1} u = (n+1) a t D_0^{n+1} u + b + 0_M \left((1 - \delta_{n2}) \frac{\varepsilon^2}{t} \left(\frac{w}{w_0} \right)^{2n} \right) \qquad (42b)$$

with a as in (39b) and

$$\begin{aligned} b &= 0_M \left(\frac{1}{t} |u'|_5 + t^{-2} (1+M)^3 \varepsilon^2 \left(\frac{w}{w_0} \right)^{21} \right) \\ &= 0_M \left(\frac{(1+M)^8 2^{(\mu M + 21)}}{t^{2 - (\mu M + 21)\theta} t_0^{(\mu M + 21)\theta}} \right) \end{aligned}$$

by (28c), (42a). Thus, using (28a),

$$\int_{t_0}^{t} |b(x)|\, ds = 0(\varepsilon).$$

Moreover, for $n > 2$

$$\int_{t_0}^{t} \frac{\varepsilon^2}{s} \left(\frac{w}{w_0}\right)^{2n} \left(\frac{w(t)}{w(s)}\right)^{n+1} ds = 0_M \left(\varepsilon \left(\frac{w}{w_0}\right)^{2n-1}\right). \tag{42c}$$

It follows then from (42b), (39d), (30) that (41a) holds.

Estimates for generalized derivatives

The inequality (27b) to be proved is equivalent to

$$\Gamma^A u' = 0_M \left(\frac{\varepsilon}{t} \left(\frac{w}{w_0}\right)^{2|A|-1}\right) \tag{43a}$$

for $1 \leq |A| \leq 4$. It is sufficient to show (43a) for Γ^A of the form

$$\Gamma^A = \Gamma^B D_0^n \tag{43b}$$

with

$$|B|^* = |B| = |A|^* = m; \qquad N = |A| = m + n \leq 4. \tag{43c}$$

For if more generally

$$A = B + C; \qquad |B|^* = |B| = |A|^* = m; \qquad |C| = n;$$
$$|C|^* = 0; \qquad N = |A| = m + n \leq 4$$

then by (33a), (43b), (28c), (28a)

$$\begin{aligned}
\Gamma^A D_\gamma u &= \Gamma^B \Gamma^C D_\gamma u = 0(|\Gamma^B D_0^{n+1} u| + \frac{1}{t}|u|_{n+m+1}) \\
&= 0_M \left(\varepsilon \left(\frac{w}{w_0}\right)^{2N-1} + \varepsilon(1+M)^8 2^{\mu M+16} t^{-15/8} \left(\frac{t}{t_0}\right)^{(\mu M+16)\theta}\right) \\
&= 0_M \left(\varepsilon \left(\frac{w}{w_0}\right)^{2N-1}\right).
\end{aligned}$$

In particular (41a) implies that (43a) holds for $|A|^* = 0$.

We prove first that

$$t\Gamma_p D_0 u = 0\left(\varepsilon \frac{w}{w_0}\right) \tag{44a}$$

corresponding to the case $m = 1$, $n = 0$ in (43b). By (11b,c), (17a), (39f), (27a), (33c,g) we have

$$\begin{aligned} L_u \Gamma_p u &= \overline{\sum} a_{\alpha\beta} D_\gamma D_\delta u + (\Gamma_p a_{\alpha\beta})(D_\alpha D_\beta u) \\ &= 2Z(\xi)(D_0^2 u)(D_0 \Gamma_p u) + 0_M \left(\frac{\varepsilon^2}{t^2} \frac{w}{w_0} + |u'||u'|_1^2 + \frac{1}{t} |u|_2 |u'|_1 \right) . \end{aligned}$$

Using (35a,b), (24c), it follows that $tD_0\Gamma_p u$ satisfies an ordinary differential equation along pseudo-characteristics

$$\frac{d}{dt} t D_0 \Gamma_p u = a t D_0 \Gamma_p u + b + 0_M \left(\frac{\varepsilon^2}{t} \frac{w}{w_0} \right) \tag{44b}$$

with a as in (39b) and

$$\begin{aligned} b &= 0_M \left(\frac{1}{t} |u|_3 + t|u'||u'|_1^2 + |u|_2 |u'|_1 \right) \\ &= 0_M \left(\frac{\varepsilon}{t} (1+M)^8 t^{-7/8} \left(\frac{w}{w_0} \right)^{\mu M + 17} \right) . \end{aligned} \tag{44c}$$

We solve the linear differential equation (44b) explicitly, starting at a point (r_1, t_1) on the boundary of the critical zone at which (44a) is sure to hold. Using that by (26a), (28a)

$$\int_{t_1}^{t} |b(s)|\, ds = 0_M(\varepsilon) \tag{44d}$$

we find from (39d) that

$$t D_0 \Gamma_p u = 0_M \left(\varepsilon \frac{w}{w_0} + \varepsilon^2 \frac{w}{w_0} \log \frac{t}{t_0} \right) . \tag{44e}$$

From (24a) we conclude that

$$\frac{w}{w_0} \geq \left(1 - \frac{1}{2} w_0 \log \frac{t}{t_0} \right)^{-1} \tag{44f}$$

and thus

$$\varepsilon \log \frac{t}{t_0} = 0 \left(w_0 \log \frac{t}{t_0} \right) = 0(1). \tag{44g}$$

Hence (44a) follows from (44d).

We next show (43a) by induction over m for Γ^A of the form (43b,c) with $m \geq 1$, $N > 1$, assuming as known that

$$t \Gamma^C u' = 0_M \left(\varepsilon \left(\frac{w}{w_0} \right)^{2|C|-1} \right) \tag{45a}$$

for $|C| < N$ or for $|C| = N$, $|C|^* < m$.

By (17a), (11b,c,d)

$$\begin{aligned} L_u \Gamma^B D_0^n u &= \overline{\sum_{\substack{|E|<|B| \\ |E|^*<|B|^*}}} \Gamma^E D_0^n a_{\alpha\beta} D_\gamma D_\delta u \\ &\qquad + \Gamma^B \sum_{p=1}^{n} \binom{n}{k} (D_0^p a_{\alpha\beta})(D_0^{n-p} D_\alpha D_\beta u) \\ &\qquad + \Gamma^B a_{\alpha\beta} D_0^n D_\alpha D_\beta u - a_{\alpha\beta} \Gamma^B D_0^n D_\alpha D_\beta u \\ &= \overline{\sum} (\Gamma^C D_0^p a_{\alpha\beta})(\Gamma^D D_0^{n-p} D_\gamma D_\delta u) \end{aligned} \tag{45b}$$

with

$$|C| + |D| \le |B| = m; \qquad |D| + n - p + 1 \le m + n = N. \tag{45c}$$

Here by (41e) and the induction assumption

$$(\Gamma^C D_0^p a_{\alpha\beta})(\Gamma^D D_0^{n-p} D_\gamma D_\delta u) = 0_M \left(\frac{\varepsilon^2}{t^2} \left(\frac{w}{w_0} \right)^{2N} + |u'|_4^3 \right) \tag{45d}$$

unless $|C| = |B|$, $p = n$ or $|D| = |B|$, $p = 1$. This leaves

$$\begin{aligned} L_u \Gamma^B D_0^n u &= (\Gamma^B D_0^n a_{\alpha\beta})(D_\alpha D_\beta u) + n(D_0 a_{\alpha\beta})(\Gamma^B D_0^{n-1} D_\alpha D_\beta u) \\ &\qquad + 0_M \left(\frac{\varepsilon^2}{t^2} \left(\frac{w}{w_0} \right)^{2N} + |u'|_4^3 \right) \\ &= 2(n+1) Z(\xi)(D_0^2 u)(D_0 \Gamma^B D_0^n u) \\ &\qquad + 0_M \left(\frac{\varepsilon^2}{t^2} \left(\frac{w}{w_0} \right)^{2N} + |u'|_4^3 + \frac{1}{t} |u'|_1 |u|_4 \right). \end{aligned} \tag{45e}$$

Consequently by (35a,b)

$$\frac{d}{dt} t D_0 \Gamma^B D_0^n u = (n+1) a t D_0 \Gamma^B D_0^n u + b + 0_M \left(\frac{\varepsilon^2}{t} \left(\frac{w}{w_0} \right)^{2N} \right) \tag{46a}$$

with a as in (39b), and

$$b = 0_M \left(t |u'|_4^3 + |u'|_1 |u|_4 + \frac{1}{t} |u|_6 \right). \tag{46b}$$

Here again (44d) holds. Moreover by (24a)

$$\int_{t_1}^{t} \frac{\varepsilon^2}{s} \left(\frac{w(s)}{w_0} \right)^{2N} \left(\frac{w(t)}{w(s)} \right)^{n+1} ds = \frac{\varepsilon^2 w^{n+1}(t)}{w_0^{2N}} \int_{t_1}^{t} \frac{1}{s} w^{2N-n-1}(s)\, ds$$

$$\begin{aligned}
&\leq \; 2\frac{\varepsilon^2 w^{n+1}(t)}{w_0^{2N}} \int_{t_1}^{t} w^{2N-n-3}(s)\frac{dw}{ds}\, ds \\
&= \; 0\left(\varepsilon \left(\frac{w(t)}{w_0}\right)^{2N-1}\right)
\end{aligned}$$

since by assumption $N > 1$, $m \geq 1$, and thus

$$2N - n - 3 = N + m - 3 \geq 0.$$

Solving the differential equation (46a) then yields (43a) for Γ^A of the form (43b,c).

This completes the proof of (27b) in the critical zone. We notice that the estimate (27a) for $|u'|$ has been used in the proof of (27b) only for the case $m = 1$ with $\Gamma^A = \Gamma_p$. We also observe that assumptions (24c), (26a) imply that actually

$$|u'(t,x)| \leq |u'(t,x)_1 < \frac{\rho}{2}; \qquad |u'(t,x)|_m < \frac{1}{2} \quad \text{for } m = 1,\, 2,\, 3,\, 4 \tag{47a}$$

when $\varepsilon < E(M)$.

Assumption (24a) can be replaced by the assumption that

$$Z(\xi)tD_0^2u(t,x) < \frac{1}{2}w(t) \qquad \text{for } |x| < t - t^{1/4} \tag{47b}$$

once (24b,c), (26a,b) are postulated. Indeed, (47b) implies that $Z(\xi)tD_0^2u(t,x)$ reaches its supremum $w(t)$ only in the interior of the critical zone. We make use of (36c), which follows directly from (24c) without involving any L_2-estimates. This leads to (38b,c,d) and the monotonicity of $w(t)$. Finally, (38d), (26a) imply that

$$\begin{aligned}
\frac{dw}{dt} &= \frac{1}{t}w^2\left(1 + 0\left(\frac{\varepsilon M w^3}{t w_0^5}\right)\right) \\
&= \frac{1}{t}w^2\left(1 + 0\left(\frac{\varepsilon M}{w_0^2 t^{1-3\theta} t_0^{3\theta}}\right)\right) = \frac{1}{t}w^2\left(1 + 0_M\left(\frac{\varepsilon M}{w_0^2 t_0}\right)\right) \\
&= \frac{1}{t}w^2\left(1 + 0_M\left(M\varepsilon \log^2\frac{1}{\varepsilon}\right)\right) > \frac{1}{2t}w^2.
\end{aligned} \tag{47c}$$

The L_2-estimates can then be carried out and the estimate (28c) derived. This shows that actually in the non-critical zone

$$\begin{aligned}
Z(\xi)tD_0^2u(t,x) &= 0\left(\frac{(1+M)^8\varepsilon w(t)}{w_0(1+|t-r|)^{1/2}}\left(\frac{w(t)}{w_0}\right)^{\mu M+15}\right) \\
&= 0\left(\frac{(1+M)^8 2^{\mu M+15}\varepsilon w(t)}{w_0 t^{1/8}}\left(\frac{t}{t_0}\right)^{\mu M+15}\right)
\end{aligned}$$

$$= \quad 0\left(\frac{(1+M)^8 2^{\mu M+15} w(t)}{w_0 t_0^{1/8}}\right)$$
$$= \quad 0\left((1+M)^8 2^{\mu M+15}\left(\varepsilon \log \frac{1}{\varepsilon}\right)^{1/4}\right) < \frac{1}{3} w(t) \qquad (47\mathrm{d})$$

for ε less than a suitable $E(M)$.

Final estimates

The preceding estimates can be summarized as follows: For given f, g, $a_{\alpha\beta}$ there exists a constant C_1 and a function $E(M)$ such that whenever $\varepsilon < E(M)$ and (47b), (24b,c), (26a,b) are satisfied in a strip $t_0 \leq t < \tau$, $x \in R_3$ for the solution u of (1a), (2a), then in the same strip

$$Z(\xi) t D_0^2 u(t,x) < \frac{1}{3} w(t) \qquad \text{for } |x| < t - t^{1/4} \tag{48a}$$

$$|u'(t,x)| < \frac{\rho}{2}; \qquad |u'(t,x)|_m < \frac{1}{2} \quad \text{for } m = 1,\ 2,\ 3,\ 4 \tag{48b}$$

$$|u'(t,x)|_m < C_1 \frac{\varepsilon}{t}\left(\frac{w(t)}{w_0}\right)^{2m-1} \qquad \text{for } m = 1,\ 2,\ 3,\ 4 \tag{48c}$$

Obviously we can replace here C_1 by any larger number and $E(M)$ by any smaller one. We turn now to the question of finding M and τ, such that (47b), (24b,c), (26a,b) are satisfied. According to (20g), (23a) there exist positive constants C_0 and ε_0 such that for $0 < \varepsilon < \varepsilon_0$

$$Z(\xi) t_0 D_0^2 u(t_0,x) < \frac{1}{3} w(t_0) \tag{49a}$$

$$|u'(t_0,x)| < \frac{\rho}{2}; \qquad |u'(t_0,x)|_m < \frac{1}{2} \quad \text{for } m = 1,\ 2,\ 3,\ 4 \tag{49b}$$

$$|u'(t_0,x)|_m < C_1 \frac{\varepsilon}{t_0}. \tag{49c}$$

We can assume that

$$C_1 > C_0; \qquad E(M) < \varepsilon_0. \tag{49d}$$

We now choose

$$M = 2C_1; \qquad \theta = \theta(2C_1) = \frac{1}{10}\left(\frac{2}{H} C C_1 + 21\right)^{-1} \tag{50a}$$

and a fixed $\varepsilon < E(2C_1)$. The inequalities (24b,c), (26a), (47b) will then be satisfied for $t_0 \leq t < \tau$, $x \in R_3$, if only $\tau - t_0$ is sufficiently small, as a consequence of (49a,b,c) and of $2C_1 > C_0$, $w(t_0) = w_0$. There exists a largest $\tau = t_2$ such that (24b,c), (26a), (47) hold for $t_0 \leq t < \tau$, $x \in R_3$ with the values (50a) for M and θ. At $t = t_2$ one of the inequalities (47b) (24b,c), (26a) has to become an equality. It can not be (24b,c), (47b) since actually the stronger inequalities (48a,b,c) hold for $t_0 \leq t < t_2$, $x \in R_3$. Hence we must have

$$\frac{w(t_2)}{w_0} = 2\left(\frac{t_2}{t_1}\right)^{\theta} \tag{50b}$$

while for $t_0 \leq t < t_2$

$$\frac{w(t)}{w_0} < 2\left(\frac{t}{t_0}\right)^{\theta}; \qquad |u'(t,x|_m < 2C_1 \frac{\varepsilon}{t}\left(\frac{w}{w_0}\right)^{2m-1} \qquad \text{for } m = 1,\, 2,\, 3,\, 4.$$

By (47c)

$$w^{-2}\frac{dw}{dt} = \frac{1}{t} + 0\left(\frac{\varepsilon}{w_0^2 t^{2-3\theta} t_0^{3\theta}}\right) \tag{51a}$$

and hence

$$\frac{w(t)}{w_0} = \left(1 - w_0 \log\frac{t}{t_0} + 0\left(\varepsilon^2 \log^2\frac{1}{\varepsilon}\right)\right)^{-1}. \tag{51b}$$

It follows that for $w_0 \log(t/t_0) < 1/3$

$$\frac{w}{w_0} < \frac{3}{2} + 0\left(\varepsilon^2 \log^2\frac{1}{\varepsilon}\right) < 2$$

and hence by (50b) that

$$w_0 \log\frac{t_2}{t_0} > \frac{1}{3}. \tag{51c}$$

On the other hand, we conclude from (51b) that for

$$\frac{1}{3} < w_0 \log\frac{t}{t_0} < 1 - \varepsilon^2 \log^3\frac{1}{\varepsilon}$$

we have

$$\frac{w(t)}{w_0} \leq \left(\varepsilon^2 \log^3\frac{1}{\varepsilon}\right)^{-1} + 0\left(\varepsilon^2 \log^4\frac{1}{\varepsilon}\right)^{-1} < 2\exp\frac{\theta}{3w_0}$$

for sufficiently small ε. Hence, by (50b), (51c)

$$w_0 \log\frac{t_2}{t_0} > 1 - \varepsilon^2 \log^3\frac{1}{\varepsilon}. \tag{51d}$$

It follows from (50b) that

$$\begin{aligned} \frac{w(t_2)}{w_0} &> 2\exp\left(\frac{\theta - \theta\varepsilon^2 \log^3(1/\varepsilon)}{w_0}\right) \\ &= 2\exp\frac{\theta}{w_0}\left(1 - 0\left(\varepsilon \log^3\frac{1}{\varepsilon}\right)\right). \end{aligned} \tag{51e}$$

By (51d)

$$\log\frac{t_2}{t_0} > \frac{1}{w_0} - 0\left(\varepsilon\log^3\frac{1}{\varepsilon}\right)$$

while by (51b)

$$\log\frac{t_2}{t_0} < \frac{1}{w_0} + 0\left(\varepsilon\log^2\frac{1}{\varepsilon}\right). \tag{51f}$$

From (39a) and the definition (9) of $w(t)$ we know that

$$w^{-1}(t)\max_x|tD_0^2u(t,x)|$$

for sufficiently small ε lies between two positive bounds, that only depend on f, g, $a_{\alpha\beta}$. Thus we can use w/t as a measure for the supremum over x of $|D_0^2u(t,x)|$. The function $t^{-1}w(t)$ assumes its minimum at some point t_1 of the interval $t_0 \le t \le t_2$. From (38d), (26a) we find that

$$\frac{d}{dt}\frac{1}{t}w(t) = t^{-2}w\left(w - 1 + 0\left(\frac{\varepsilon w^4}{tw_0^5}\right)\right) = t^{-2}w\left(w - 1 + 0\left(\frac{\varepsilon}{w_0t_0}\right)\right) \tag{52a}$$

whenever dw/dt exists. This shows that $t_0 < t_1 < t_2$, since $w(t) = 0(\varepsilon) < 1$ near t_0 and by (51e) $w > 2$ near $t = t_2$. We also deduce from (52a) that

$$w(t_1) = 1 + 0\left(\frac{\varepsilon}{w_0t_0}\right) = 1 + 0\left(\varepsilon^2\log^2\frac{1}{\varepsilon}\right). \tag{52b}$$

The differential equation (47c) implies that

$$\frac{w(t_2)}{w(t_1)} = \left(1 - w(t_1)\log\frac{t_2}{t_2} + 0\left(\frac{\varepsilon w(t_1)}{w_0^2t_1^{1-3\theta}t_0^{3\theta}}\right)\right)^{-1}.$$

Consequently

$$\frac{t_2}{t_1} \le \exp\left(1 + 0\left(\varepsilon^2\log^2\frac{1}{\varepsilon} + \frac{\varepsilon}{w_0^2t_0}\right)\right) = e + 0\left(\varepsilon\log^2\frac{1}{\varepsilon}\right). \tag{52c}$$

We see from (52e,f) that

$$\frac{w(t_2)}{t_2} > \frac{2w_0}{t_0}\left(\exp\frac{\theta}{w_0}\right)\left(\exp-\frac{1}{w_0}\right)\left(1 - 0\left(\varepsilon\log^3\frac{1}{\varepsilon}\right)\right) \tag{52d}$$

and from (52b,c) that

$$\frac{w(t_1)}{t_1} = \frac{w(t_1)}{t_2}\frac{t_2}{t_1} < \frac{e}{t_0}\left(\exp-\frac{1}{w_0}\right)\left(1 + 0\left(\varepsilon\log^2\frac{1}{\varepsilon}\right)\right) \tag{52e}$$

where by (23b), (22a)

$$w_0 = \varepsilon H + 0\left(\varepsilon^2\log\frac{1}{\varepsilon}\right); \qquad t_0 = \left(\varepsilon\log\frac{1}{\varepsilon}\right)^{-2}. \tag{52f}$$

This completes the analysis of the behavior of the derivatives of u in the third phase, as announced in the introduction.

Acknowledgment. The research for this paper was supported in part by the National Science Foundation under Grant #DMS-8806726.

REFERENCES

[1] Klainerman, S. 1984. Weighted L^{∞} and L^1 estimates for solutions to the classical wave equations in three space dimensions. *Comm. Pure Appl. Math.*, 37:269–88.

[2] John, F. and Klainerman, S. 1984. Almost global existence to nonlinear wave equations in three space dimensions. *Comm. Pure Appl. Math.*, 37:443–55.

[3] Hörmander, L. 1985. The lifespan of classical solutions of nonlinear hyperbolic equations. Revised version *Institut Mittag-Leffler*, Report No. 5:1–67.

[4] John, F. 1987. Existence for large times of strict solutions of nonlinear wave equations in three space dimensions for small initial data. *Comm. Pure Appl. Math.*, 40:79–109.

[5] Christodoulou, D. 1986. Global solutions of nonlinear hyperbolic equations for small intial data. *Comm. Pure Appl. Math.*, 39:267–82.

[6] Klainerman, S. 1986. The null condition and global existence to nonlinear wave equations. *Lectures in Applied Mathematics*, 23:293–326.

[7] John, F. 1985. Blow-up of radial solutions of $u_{tt} = c^2(u_t)\Box u$ in three space dimensions. *Mathematica Aplicada e Computacional*, 4:3–18.

[8] John, F. 1981. Blow-up for quasi-linear wave equations in three space dimensions. *Comm. Pure Appl. Math.*, 34:29–51.

[9] John, F. 1985. Non-existence of global solutions of $\Box u = \frac{\partial}{\partial t}F(u_t)$ in two and three space dimensions. *Supplemento al Rendiconti del Circolo Matematico di Palermo*, Serie II numero 8:229–49.

[10] Klainerman, S. 1987. Remarks on the global Sobolev inequalities in the Minkowski space R^{n+1}. *Comm. Pure Appl. Math.*, 40:111–17.

[11] Klainerman, S. 1985. Uniform decay estimates and the Lorentz invariance of the classical wave equation. *Comm. Pure Appl. Math.*, 38:321–32.

A VISCOSITY APPROXIMATION TO A SYSTEM OF CONSERVATION LAWS WITH NO CLASSICAL RIEMANN SOLUTION

Barbara Lee Keyfitz[1]
Department of Mathematics, University of Houston
Houston, Texas 77004

and

Herbert C. Kranzer
Department of Mathematics, Adelphi University
Garden City, New York 11530

ABSTRACT: There are examples of systems of conservation laws which are strictly hyperbolic and genuinely nonlinear but for which the Riemann problem can be solved only for states which are sufficiently close together. For one such example, we introduce a particular type of artificial viscosity and show how it suggests a possible definition of "generalized" solution to the Riemann problem.

I. INTRODUCTION

The model system

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

$$U(x,0) = \begin{bmatrix} u \\ v \end{bmatrix}(x,0) = \begin{cases} U_L, & x < 0 \\ U_R & x \geq 0 \end{cases} \tag{2}$$

presents an example of a system of conservation laws satisfying the classical assumptions (strict hyperbolicity and genuine nonlinearity) which has no solution for some pairs of states U_L and U_R. On carrying out the standard construction of a solution to the Riemann problem (a shock or rarefaction of the slower family followed by a

[1] Research supported in part by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant Number AFOSR 86-0088, and by a grant from the Energy Laboratory, University of Houston. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

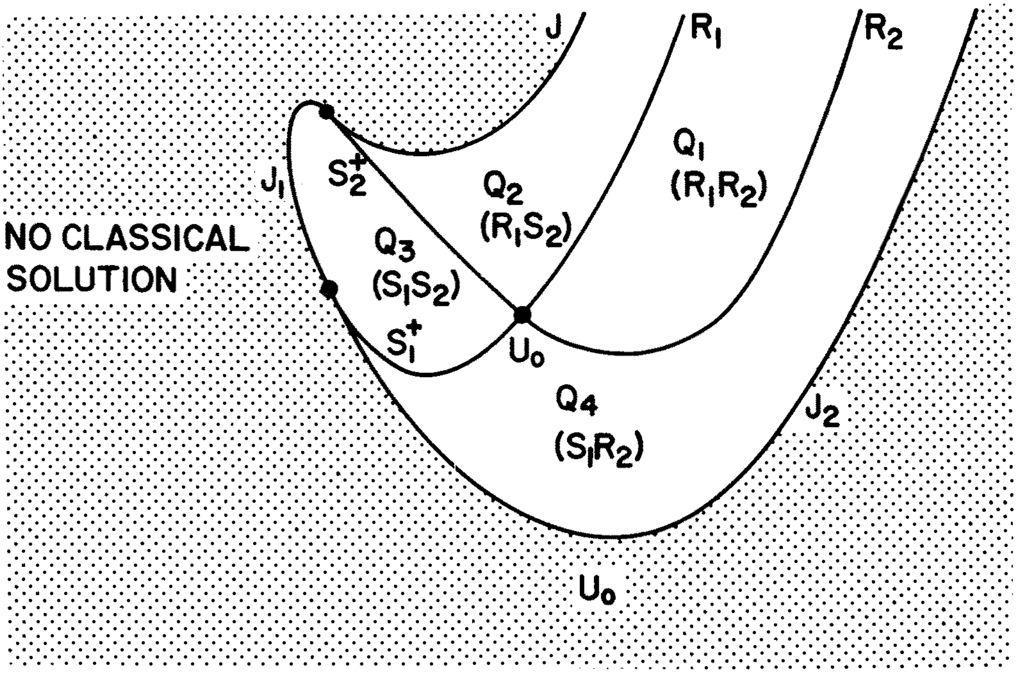


Figure 1

wave of the second), one finds that for a given left state, $U_L = U_0$, the classical type of solution exists in the four curvilinear quadrants Q_1, Q_2, Q_3, and Q_4 pictured in Figure 1. The defining equations are simple polynomial or algebraic curves. If U_R is outside the union, Q, of these sets, then no classical solution of (1), (2) exists.

If one is searching for conditions sufficient to guarantee existence for large data of solutions to systems of conservation laws, then it is interesting to try to understand how the solution to the Riemann problem breaks down for a system like (1). This motivates the study of systems that approximate (1) in some sense. One idea, first developed by Tupciev [9] and by Dafermos [2], is to approximate a system of conservation laws by the system

$$U_t + F_x = \varepsilon t U_{xx} . \tag{3}$$

The initial-value problem (3), (2) becomes a two-point boundary-value problem for a nonautonomous system of ordinary differential equations in the variable $\xi = x/t$ (we let $= d/d\xi$):

$$\varepsilon\ddot{U} = (A(U) - \xi)\dot{U} \ , \tag{4}$$

$$U(\xi) \to \begin{cases} U_L & \text{as} \quad \xi \to -\infty \\ U_R & \text{as} \quad \xi \to +\infty \end{cases} . \tag{5}$$

In this note, we study the solutions of (4), (5) for small ε using perturbation methods; we find that this system admits solutions with the structure of "singular shocks": boundary layers in which are embedded more singular solutions which are unbounded as $\varepsilon \to 0$. In a future paper [5], we will show, using the construction of Dafermos [2], that solutions of (4), (5) actually exist in a rigorous and not just an asymptotic sense, and that they have the qualitative properties exhibited here. By means of the construction presented here, we can identify "shock speeds" and "shock strengths", which we use to define a "generalized" Riemann solution.

Another source of interest in (1) is that it arises in some model equations for a problem in elastoplasticity considered by Colombeau and Le Roux in [1], when these equations are put in conservation form. Colombeau and Le Roux solve (numerically) the system

$$\begin{cases} u_t + uu_x = \sigma_x \\ \sigma_t + u\sigma_x = u_x \end{cases} \tag{1'}$$

which is related to (1) by the change of variables

$$\sigma = v - \frac{u^2}{2} .$$

The weak solutions of (1) and (1'), even if of classical type in both cases, would be inequivalent. Le Floch, in [6], establishes a theoretical framework for the study of "nonconservation laws" of this type, and formulates admissibility conditions for generalized shocks. Le Floch's method, applied to (1'), also gives a solution only in one quadrant of the plane, while the solution obtained by Colombeau and Le Roux is highly dependent on the specific form of the numerical scheme. However, the theory of generalized functions developed by Colombeau and used in [1] provides a tool whereby the relation of the approximation (4), (5) to the limiting system (1), (2) might be studied. This will be the subject of a future study.

II. THE SINGULAR SOLUTION

We consider the system (3). This approximation was designed for looking at solutions to the Riemann problem (initial data (2)), because it has solutions of the form $U = U^{\varepsilon}(x/t) = U^{\varepsilon}(\xi)$. In fact, Dafermos [2] proved a convergence result, as $\varepsilon \to 0$, for Riemann problems for 2×2 systems with enough restrictions that a global, bounded solution to the inviscid problem exists, and Dafermos and DiPerna [3] extended the convergence result to include the case of isentropic gas dynamics where the solution may be unbounded. The example we are considering in this paper does not fit into either of these categories. The existence of the approximate solutions and their convergence will be discussed in [5].

To simplify the notation, we drop the superscript ε when we look at solutions to the system (4), (5) of ordinary differential equations. Note that, unlike the systems generally used in studying viscous approximation to a single shock, this system depends explicitly on ε and on ξ.

The classical solutions to (4) are approximations either to <u>rarefactions</u>, in which U, $\dot{U}$ and $\ddot{U}$ are bounded as $\varepsilon \to 0$, or to <u>shocks</u>, in which U is bounded but $\dot{U}$ and $\ddot{U}$ are not. We do not expect (4), (5) to have a classical solution unless $U_R \in Q(U_L)$. We consider the possibility that <u>singular</u> <u>solutions</u> of (4) exist, in which U is unbounded for ξ near some value, s. Thus, let

$$\tilde{U}(\xi) = \begin{pmatrix} \frac{1}{\varepsilon^p}\tilde{u}\left(\frac{\xi - s}{\varepsilon^q}\right) \\ \frac{1}{\varepsilon^r}\tilde{u}\left(\frac{\xi - s}{\varepsilon^q}\right) \end{pmatrix} . \tag{6}$$

If we let $\eta = \frac{\xi - s}{\varepsilon^q}$, $' = \frac{d}{d\eta}$, then (4) becomes

$$\varepsilon^{1-q-p}\,\tilde{u}'' = (2\tilde{u}\,\varepsilon^{-p} - \varepsilon^q\eta - s)\tilde{u}'\varepsilon^{-p} - \varepsilon^{-r}\tilde{v}'$$

$$\varepsilon^{1-q-r}\,\tilde{v}'' = (\tilde{u}^2\,\varepsilon^{-2p} - 1)\tilde{u}'\varepsilon^{-p} - (\varepsilon^q\eta + s)\varepsilon^{-r}\tilde{v}' \ .$$

For nontrivial solutions to exist we must balance at least two terms in each equation. Thus we set $1 - q - r = -3p$ in the second and either $1 - q - p = -2p$ or $1 - q - r = -r$ in the first; either implies the other and yields $r = 2p$, $q = 1 + p$ and hence

$$\begin{cases} \tilde{u}'' = 2\tilde{u}\tilde{u}' - \tilde{v}' - \varepsilon^{p}(s\tilde{u}' + \varepsilon^{p+1}\eta\tilde{u}') \\ \tilde{v}'' = \tilde{u}^{2}\tilde{u}' \qquad\quad - \varepsilon^{p}(s\tilde{v}' + \varepsilon^{p}\tilde{u}' + \varepsilon^{p+1}\eta\tilde{v}') \end{cases} . \tag{7}$$

Now we expand $\tilde{u}$, $\tilde{v}$ as series in ε:

$$\tilde{u} = \tilde{u}_0(\eta) + o(1) \qquad\qquad \tilde{v} = \tilde{v}_0(\eta) + o(1) ,$$

to obtain

$$\begin{cases} \tilde{u}_0'' = 2\tilde{u}_0\tilde{u}_0' - \tilde{v}_0' \\ \tilde{v}_0'' = \tilde{u}_0^{\,2}\tilde{u}_0' \end{cases} . \tag{8}$$

We note that from (6) we must have $\tilde{u}_0, \tilde{v}_0 \to 0$ as $|\eta| \to \infty$, under the assumption that the singular behavior is confined to a neighborhood of $\xi = s$ (or $\eta = 0$); hence $\tilde{u}_0', \tilde{v}_0' \to 0$ as $|\eta| \to \infty$ and when we integrate (8) once we obtain

$$\begin{cases} \tilde{u}_0' = \tilde{u}_0^{\,2} - \tilde{v}_0 \\ \tilde{v}_0' = \frac{1}{3}\tilde{u}_0^{\,3} \end{cases} . \tag{9}$$

Simplifying the notation again, we see that we wish to study solutions of

$$\begin{cases} x' = x^{2} - y \\ y' = \frac{1}{3}x^{3} \end{cases} \tag{10}$$

which approach (0,0) as $|\eta| \to \infty$. The linearization of (10) at (0,0) gives the matrix

$$\begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix}$$

which is nilpotent (double zero eigenvalue). This apparently nongeneric behavior is stable: if we had begun with

$$\begin{cases} u_t + (f(u) - v)_x = 0 \\ v_t + \quad (g(u))_x = 0 \end{cases}$$

and had assumed $f \to u^2$, $g \to \frac{1}{3}u^3$ as $|u| \to \infty$, in following the derivation of equations (6), (7), and (8), we would have ended up with (9). The system is invariant under the group action $x \to -x$, $\eta \to -\eta$.

Now, it happens that we can integrate (10). Let

$$z = y - kx^2$$

where k is either root of

$$k^2 - k + \frac{1}{\sigma} = 0 .$$

Then

$$z' = 2kxz$$

and

$$\frac{d}{d\eta}\left[x^2 z^m + \frac{z^{m+1}}{2k-1} \right] = 0 ,$$

where $m = \frac{k-1}{k}$. Hence, along trajectories,

$$z^m\left[z + (2k - 1)x^2\right] = \text{constant}.$$

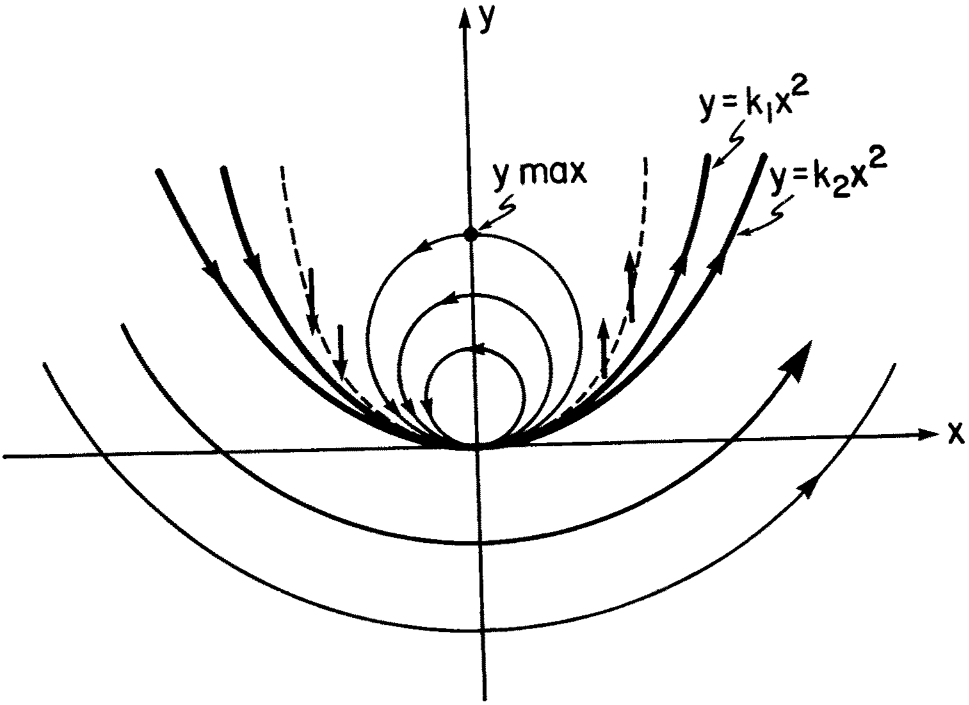


Figure 2

We see that the two parabolas $y = k_i x^2$,

$$k_1 = \frac{1}{2}\left[1 + \frac{1}{\sqrt{3}}\right] \qquad\qquad k_2 = \frac{1}{2}\left[1 - \frac{1}{\sqrt{3}}\right] ,$$

are trajectories of the flow $(z = 0)$. By choosing $k = k_1$, we see that if $z(0) = y(0) - k_1(x(0))^2 > 0$ then the trajectory is bounded: it is, in fact, a homoclinic orbit in the upper half plane, making second-order contact with $y = k_1 x^2$. For all these orbits, we may choose the symmetric solution: $x(0) = 0$, $x(-\eta) = - x(\eta)$; then each orbit is characterized by $y(0) = z(0) = y_{max} > 0$. We also have $y(-\eta) = y(\eta)$. The asymptotic behavior is

$$x = \frac{c}{\eta} + \mathcal{O}\left[\frac{1}{\eta^2}\right]$$

$$y = \frac{d}{\eta^2} + \mathcal{O}\left[\frac{1}{\eta^3}\right]$$

where $c = -6k_1 = -3 - \sqrt{3}$ and $d = c^2 + c = 9 + 5\sqrt{3}$ for all the orbits. In the sectors between the parabolas $y = k_i x^2$ the flow is radially inward (if $x < 0$) or outward (if $x > 0$); below $y = k_2 x^2$ the flow follows unbounded trajectories from left to right. This is summarized in Figure 2. Finally, we identify $(\tilde{u}_0, \tilde{v}_0)$ with (x, y) for one of the homoclinic orbits. We see that there are many singular solutions of (9) and hence of type (6).

III. COMPLETION OF THE BOUNDARY LAYER SOLUTION

The singular solution constructed in Section II has its essential support in a layer of width $|\xi - s| = \mathcal{O}(\varepsilon^q) = \mathcal{O}(\varepsilon^{p+1})$. Since $p > 0$, this solution is narrower than a classical shock (which has width $|\xi - s| = \mathcal{O}(\varepsilon)$); also, far away from $\xi = s$, it tends to zero. Thus, by itself it does not solve any Riemann problems. We are led to the idea of embedding a singular shock in a shock profile of the usual type: a solution $\bar{U}(\tau) = \bar{U}(\frac{\xi - s}{\varepsilon})$ of (4) which is bounded in a layer $|\xi - s| = \mathcal{O}(\varepsilon)$ outside the singular layer and whose derivatives are $\mathcal{O}(\frac{1}{\varepsilon})$ outside the singular layer. We shall call this two-component region, $\varepsilon^{p+1} < |\xi - s| < \varepsilon$, the boundary layer. Now, in terms of $\tau = \frac{\xi - s}{\varepsilon}$, equation (4) can be written

$$\frac{d^2\bar{U}}{d\tau^2} = (A(\bar{U}) - \varepsilon\tau - s)\,\frac{d\bar{U}}{d\tau} ,$$

or

$$\frac{d}{d\tau}\left[\frac{d\bar{U}}{d\tau} - F(\bar{U}) + s\bar{U}\right] = -\varepsilon\tau\frac{d\bar{U}}{d\tau}\,. \qquad (11)$$

We illustrate the scalings in Figure 3. If we expand $\bar{U} = \bar{U}_0 + o(1)$ in the boundary layer, then, since by assumption the right hand side of (11) is $\mathcal{O}(\varepsilon)$ there, we have

$$\frac{d}{d\tau}\left[\frac{d\bar{U}_0}{d\tau} - F(\bar{U}_0) + s\bar{U}_0\right] = 0$$

in each separate interval of the boundary layer, $\tau < 0$ and $\tau > 0$, and so

$$\frac{d\bar{U}_0}{d\tau} - F(\bar{U}_0) + s\bar{U}_0 = C_{\mp}\;; \qquad (12)$$

the two constants being constants of integration in the two intervals.

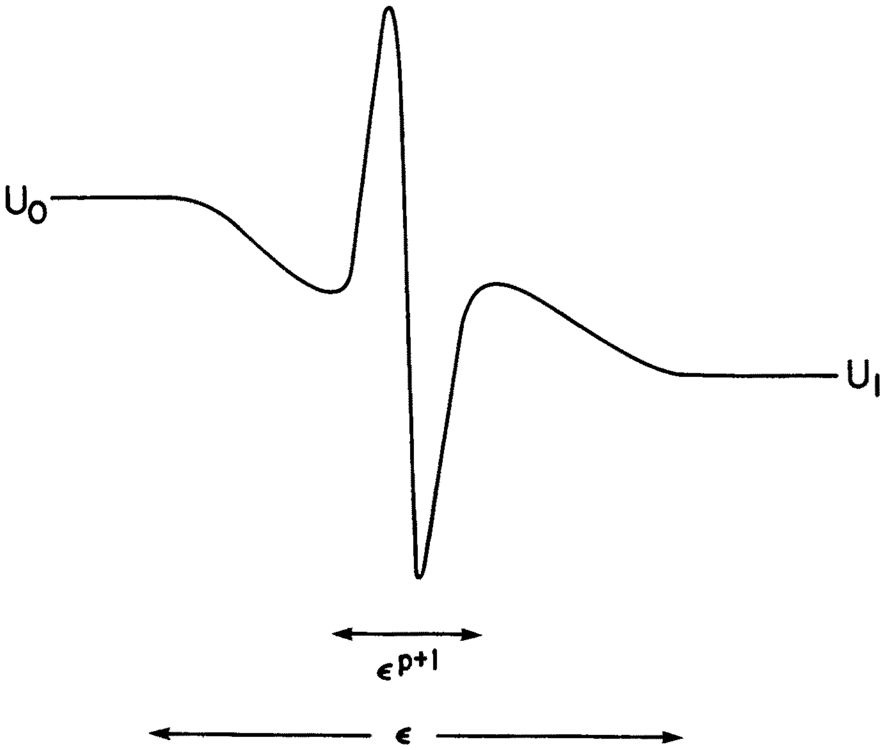


Figure 3

Furthermore, integrating (11), we have

$$\left\{ \frac{d\bar{U}}{d\tau} - F(\bar{U}) + s\bar{U} \right|_{\tau<0}^{\tau>0} = -\varepsilon \int_{\tau<0}^{\tau>0} \tau \frac{d\bar{U}}{d\tau}\, d\tau. \tag{13}$$

Now, let $\bar{U}(\tau) = \tilde{U}(\xi)$ in the integrand on the right side of (13); then, using (12), we see that

$$C_+ - C_- = \lim_{\varepsilon \to 0} \left(-\varepsilon \int_{-\infty}^{\infty} \varepsilon^p \eta \begin{pmatrix} \frac{1}{\varepsilon^p} \tilde{u}' \\ \frac{1}{\varepsilon^{2p}} \tilde{v}' \end{pmatrix} d\eta \right) = \lim_{\varepsilon \to 0} \begin{pmatrix} -\varepsilon \int \eta \tilde{u}_0' d\eta \\ -\varepsilon^{1-p} \int \eta \tilde{v}_0' d\eta \end{pmatrix}. \tag{14}$$

Now, $\tilde{u}_0' \approx \frac{c}{\eta^2}$ when $|\eta| \to \infty$, so $\eta \tilde{u}_0'$ is not absolutely integrable. However, it is an odd function, so its PV integral is zero. On the other hand, $\int \eta \tilde{v}_0' d\eta$ exists for all the homoclinic trajectories, and

$$\int_{-\infty}^{\infty} \eta \tilde{v}_0' d\eta = \eta \tilde{v}_0 \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \tilde{v}_0 d\eta = - \int_{-\infty}^{\infty} \tilde{v}_0 d\eta$$

has a different finite value for each trajectory. Thus we get a nontrivial result in (14) if $p = 1$. Then

$$C_+ - C_- = \begin{pmatrix} 0 \\ c \end{pmatrix}, \quad \text{where} \quad c = \int_{-\infty}^{\infty} \tilde{v}_0 d\eta . \tag{15}$$

Finally, a shock in the boundary layer which approaches constant values, $\bar{U}_0 \to U_\pm$ as $\tau \to \pm\infty$, and $d\bar{U}_0/d\tau \to 0$ as $|\tau| \to \infty$, must satisfy, from (12),

$$sU_- - F(U_-) = C_- , \qquad\qquad sU_+ - F(U_+) = C_+$$

and hence

$$s(U_+ - U_-) - (F(U_+) - F(U_-)) = C_+ - C_- = \begin{pmatrix} 0 \\ c \end{pmatrix}. \tag{16}$$

This is the <u>Generalized</u> <u>Rankine-Hugoniot</u> <u>Condition</u> for singular shocks.

Which states (U_-, U_+) can be joined by an <u>admissible</u> singular shock? That is, when does there exist a trajectory $\bar{U}(\tau)$ joining U_- and U_+? We note that for any pair of states (u_-, v_-) and (u_+, v_+), we have solutions to (16) given by

$$s = \frac{[-v]+[u^2]}{[u]} = u_+ + u_- - \frac{v_+-v_-}{u_+-u_-} \qquad (17)$$

and

$$c = s[v] - \frac{1}{3}[u^3] - [u] = [v]\left(\frac{[u^2]-[v]}{[u]}\right) + [u] - \frac{1}{3}[u^3] \text{ .}$$

But for trajectories to exist, we need at least one positive eigenvalue at U_- and one negative eigenvalue at U_+ in the linearized matrix $A(U) - sI$. We conjecture [5], that trajectories exist if and only if there are two such eigenvalues: that is,

$$\lambda_2(u_-) > \lambda_1(u_-) \geq s \geq \lambda_2(u_+) > \lambda_1(u_+). \qquad (18)$$

Since s depends on $[v]$, this defines a sector, $S_\delta(U_-)$, for each U_-, by

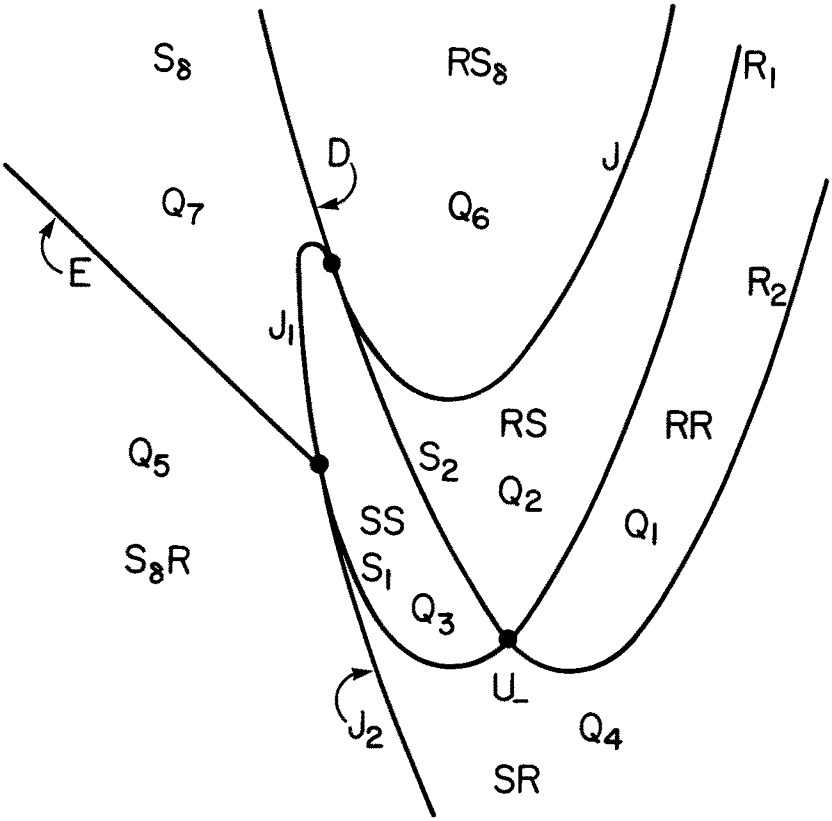


Figure 4

$$S_\delta(U_-) = (U \notin Q_3 |\ u < u_- ;\ u(u_- - 1) - u_-(u_- - 1)$$
$$\le v - v_- \le u^2 + (1 - u_-)u - u_- \} \ . \qquad (19)$$

On the lower boundary, E, of S_δ, $s = \lambda_2(u)$, and thus a singular shock to E can be continued via a 2-rarefaction to any point in a sector Q_5 below E. On the upper boundary, $D(U_-)$, $s = \lambda_1(u_-)$, and so in the region Q_6 above D there is a solution to the Riemann problem consisting of 1-rarefactions followed by singular shocks joining U to U_m, where $U_m \in R_1(U_-)$, and $U \in D(U_m)$. The regions Q_5 and Q_6 adjoin the sectors Q_2 and Q_4, in which there are classical solutions. Thus, letting Q_7 denote the sector S_δ in which the solution to the Riemann problem consists of a singular shock alone, we have described a "generalized Riemann solution" of our original problem in the entire plane. This is illustrated in Figure 4.

An analytical justification of the asymptotic derivation presented here, consisting of an existence theorem for solutions of (4) and (5) for all $\varepsilon > 0$ and a demonstration of the qualitative behavior of the solutions in sectors Q_5, Q_6 and Q_7 will be the subject of [5].

IV. CONCLUSIONS

We have presented, in (1), an example of a system of conservation laws for which the "large data" Riemann problem may not have a solution. The obstruction to solving the Riemann problem for (1) can be described as a consequence of the (easily verified) compactness of the Hugoniot locus in the u-v plane; this is related (although we have not established how general is the connection) to the fact that the two families of characteristic speeds are not globally distinct: the <u>global</u> character of this system is not that of a strictly hyperbolic problem. In fact, the system

$$\begin{cases} u_t + (u^2 - v)_x = 0 \\ v_t + (\frac{1}{3}u^3 - k^2 u)_x = 0 \end{cases} \qquad (20)$$

(which happens to correspond to the example considered in [1]), for which the characteristic speeds are $u \pm k$, can be rescaled ($u \mapsto ku$, $v \mapsto k^2 v$, $x \mapsto kx$) to the form (1), with a similar correspondence of solutions to the Riemann problem. In the limiting case, $k = 0$, (which is "unphysical" in [1], since k is a Hooke's constant), (20) becomes a nonlinear system with globally coincident characteristics. In this case the quadrant, Q, of classical Riemann solutions degenerates to a single, semi-infinite curve.

It is instructive to note that the linearization of this degenerate system about a constant state $u = a$, say, leads to the equation

$$w_{tt} + 2aw_{tx} + a^2 w_{xx} = 0 \tag{21}$$

for $w = u - a$; this equation is of a degenerate type (it might be called weakly hyperbolic at best). The solution of (21) with Cauchy data

$$w(x, 0) = w_0(x), \qquad w_t(x, 0) = v_0(x),$$

is

$$w(x, t) = w_0(x + at) + t[v_0(x + at) - aw_0'(x + at)]. \tag{22}$$

The solution in (22) of the linear problem (21) exhibits both growth in t and loss of a derivative; one can read both these features in the approximate solution constructed asymptotically in this paper: although the limit solution defined by (16) - (19) is described by finite-valued states, the limiting process itself involves a sequence which is unbounded. It does not appear possible to speak of a solution to (1) and (2), even in the weakest sense, without invoking functions which are more singular than the data. Thus, although it is possible that the behavior of the approximate solutions constructed in this paper is strongly affected by the approximation we have chosen, there is also the possibility that this behavior is characteristic of a type of global failure of strict hyperbolicity which should be further investigated.

ACKNOWLEDGEMENT

This paper was delivered at the "Seminaire International sur les Problemes Hyperboliques de Bordeaux". Discussion after the talk pointed out the relation between this example and the failure of strict hyperbolicity, especially some comments of Tai-Ping Liu. We also benefited from discussions with J. F. Colombeau about the relation between his theory of generalized distributions and the correct notion of convergence for the approximations we describe here.

V. REFERENCES

1. J. F. Colombeau and A. Y. LeRoux, "Numerical techniques in elastoplasticity," in Nonlinear Hyperbolic Problems, (ed Carasso, Raviart and Serre), Lecture Notes in Math. 1270 (1987), Springer-Verlag, Berlin, 103-114.

2. C. M. Dafermos, "Solution of the Riemann problem for a class of hyperbolic systems of conservation laws by the viscosity method," Arch. Rat. Mech. Anal. 52 (1973), 1-9.

3. C. M. Dafermos and R. J. DiPerna, "The Riemann problem for certain classes of hyperbolic systems of conservation laws," Jour. Diff. Eqns. 20 (1976), 90-114.

4. B. L. Keyfitz and H. C. Kranzer, "Existence and uniqueness of entropy solutions to the Riemann problem for hyperbolic systems of two nonlinear conservation laws," Jour. Diff. Eqns. 27 (1978), 444-476.

5. B. L. Keyfitz and H. C. Kranzer, "A system of conservation laws with no classical Riemann solution," preprint (1988).

6. Ph. Le Floch, "Nonlinear hyperbolic systems under nonconservative form," to appear in Comm in PDE (1988).

7. P. D. Lax, "Shock waves and entropy," in Contributions to Nonlinear Functional Analysis, (ed Zarantonello), Academic Press, New York (1971), 603-634.

8. M. Slemrod, "A limiting 'viscosity' approach to the Riemann problem for materials exhibiting change of phase," Univ. of Wisconsin preprint (1987).

9. V. A. Tupciev, "On the method of introducing viscosity in the study of problems involving decay of a discontinuity," Dokl. Akad. Nauk. SSR 211 (1973), 55-58; translated in Soviet Math. Dokl. 14.

GLOBAL CLASSICAL SOLUTIONS TO THE CAUCHY PROBLEM FOR NONLINEAR WAVE EQUATIONS

Li Ta-tsien (Li Da-qian)* and Chen Yun-mei**

§1. Introduction

Consider the cauchy problem for nonlinear wave equations

(1.1) $$\Box u=F(u,Du,D_xDu),\quad (t,x)\varepsilon R_+ x R^n,$$

(1.2) $$t=0: u=\varepsilon\phi(x),\quad u_t=\varepsilon\psi(x),\quad x\varepsilon R^n,$$

where

(1.3) $$\Box=\frac{\partial^2}{\partial t^2}-\sum_{i=1}^{n}\frac{\partial^2}{\partial x_i^2},$$

(1.4) $$D_x=\left(\frac{\partial}{\partial x_1},\dots,\frac{\partial}{\partial x_n}\right),\quad D=\left(\frac{\partial}{\partial t},\frac{\partial}{\partial x_1},\dots,\frac{\partial}{\partial x_n}\right),$$

$\varepsilon>0$ is a small parameter and

(1.5) $$\phi,\Psi\varepsilon C_0^\infty(R^n).$$

Let

(1.6) $$\hat{\lambda}=(\lambda;(\lambda_i),i=0,1,\dots,n;(\lambda_{ij}),i,j=0,1,\dots,n,i+j\geq 1).$$

Suppose that $F(\hat{\lambda})$ is a sufficiently smooth function satisfying

(1.7) $$F(\hat{\lambda})=0(|\hat{\lambda}|^{1+\alpha})$$

in a neighborhood of $\hat{\lambda}=0$, where α is an integer ≥ 1.

Only based on the decay estimates for solutions to the linear homogeneous wave equation and the energy estimates for solutions to linear inhomogeneous wave equations, we can use the contraction mapping principle in a suitable space to get directly the following global existence theorem: Under hypothesis

(1.8) $$\frac{n-1}{2}\left(1-\frac{2}{\alpha n}\right)\alpha>1,$$

if ε is suitably small, then Cauchy problem (1.1)-(1.2) admits a unique global classical solution on $t\geq 0$ and this solution has some decay pro-

*Dept. of Math. and Institute of Math., Fudan Univ., Shanghai, P.R.C.

**Dept. of Appl. Math., Tongji Univ., Shanghai, P.R.C.

perties as $t\to\infty$ just as solutions to the linear wave equation.

The relationship between n and α given by (1.8) can be expressed as follows

$\alpha=$	1	2,3	4,5,...
$n\geq$	5	3	2

.

This result generalizes the results obtained by S.Klainerman [1] (in the case that F does not explicitly depend on u), D.Christodoulou [2] (in the case that $\alpha=1$ and n is odd) and A.Matsumuta [3] (for a somewhat special kind of quasilinear wave equations and $\alpha>1$).

By differentiation, we only need to consider the Cauchy problem for the following general kind of quasilinear wave equations

$$(1.9)\qquad \Box u=\sum_{i,j=1}^{n} b_{ij}(u,Du)u_{x_ix_j}+2\sum_{j=1}^{n} a_{0j}(u,Du)u_{tx_j}+F(u,Du),(t,x)\in R_+\times R^n,$$

$$(1.2)\qquad t=0:u=\varepsilon\phi(x),\quad u_t=\varepsilon\psi(x),\quad x\in R^n.$$

Let

$$(1.10)\qquad \tilde{\lambda}=(\lambda;(\lambda_i),i=0,1,\dots,n).$$

Suppose that in a neighborhood of $\tilde{\lambda}=0$, $b_{ij}(\tilde{\lambda})$, $a_{0j}(\tilde{\lambda})$ and $F(\tilde{\lambda})$ are sufficiently smooth functions satisfying

$$(1.11)\qquad b_{ij}(\tilde{\lambda})=b_{ji}(\tilde{\lambda})\quad (i,j=1,\dots,n),$$

$$(1.12)\qquad b_{ij}(\tilde{\lambda}),\ a_{0j}(\tilde{\lambda})=O(|\tilde{\lambda}|^{\alpha})\quad (i,j=1,\dots,n),$$

$$(1.13)\qquad F(\tilde{\lambda})=O|\tilde{\lambda}|^{1+\alpha})$$

and

$$(1.14)\qquad \sum_{i,j=1}^{n} a_{ij}(\tilde{\lambda})\xi_i\xi_j\geq m_0|\xi|^2,\forall\ \xi\in R^n\quad (m_0>0,\ \text{constant}),$$

where α is an integer ≥ 1 and

$$(1.15)\qquad a_{ij}(\tilde{\lambda})=\delta_{ij}+b_{ij}(\tilde{\lambda}),$$

in which δ_{ij} is the Kronecker delta.

§2. Case $1\leq\alpha\leq 3$

In this section we give the precise statement of our result and a sketch of the proof for the case that α is an integer such that $1\leq\alpha\leq 3$ (correspondingly, $n>2$), cf.Li Ta-tsien and Chen Yun-mei [4].

Following S. Klainerman [1], introduce a set of partial differential operators

(2.1) $\Gamma=(L_\circ;(\partial_a),a=0,1,\ldots,n;(\Omega_{ab}),a,b=0,1,\ldots,n)$,

where

(2.2) $\partial_o=-\frac{\partial}{\partial t}\ ,\quad \partial_i=\frac{\partial}{\partial x_i}\quad (i=1,\ldots,n)$,

(2.3) $\Omega_{ab}=x_a\partial_b-x_b\partial_a \quad (a,b=0,1,\ldots,n;x_\circ=t)$,

(2.4) $L_o=t\partial_t+x_1\partial_1+\cdots+x_n\partial_n$,

and for any function $u=u(t,x)$ such that all norms appearing on the right hand side below are bounded, define

(2.5) $$\|u(t,\cdot)\|_{\Gamma,s,p}=\Big(\sum_{|k|\leq s}\|\Gamma^k u(t,\cdot)\|^2_{L^p(R^n)}\Big)^{\frac{1}{2}},\quad t\geq 0,$$

where $1\leq p\leq+\infty$, $k=(k_1,\ldots,k_\sigma)$ is a multi-index, $|k|=k_1+\cdots+k_\sigma$, σ is the number of partial differential operators in $\Gamma:\Gamma=(\Gamma_1,\ldots,\Gamma_\sigma)$ and

(2.6) $\Gamma^k=\Gamma_1^{k_1}\cdots\Gamma_\sigma^{k_\sigma}$.

For any given integers $s_\circ$ and s such that $s_\circ\geq n+10$, $s_\circ+n+1\leq s\leq 2s_\circ-9$ and any positive real number E, we introduce the following set of functions

(2.7) $$X_{s_\circ,s,E}=\{v=v(t,x)\,|\,D_{s_\circ,s}(v)\leq E\ ,$$
$$\partial_t^\ell v(0,x)=u_\ell^{(0)}(x)\,(\ell=o,1,\ldots,s+1)\},$$

where

(2.8) $$D_{s_\circ,\varepsilon}(v)=\sup_{t\geq 0}(1+t)^{\frac{n-1}{2}(1-\frac{2}{\alpha n})}\|v(t,\cdot)\|_{\Gamma,s_\circ,\alpha n}$$
$$+\sup_{t\geq 0}\|v(t,\cdot)\|_{\Gamma,s,2}+\sup_{t\geq 0}\|Dv(t,\cdot)\|_{\Gamma,s+1,2},$$

(2.9) $u_\circ^{(0)}=\varepsilon\phi(x)\ ,\quad u_1^{(0)}=\varepsilon\psi(x)$

and $u_\ell^{(0)}(x)\,(\ell=2,\ldots,s+1)$ are the values of $\partial_t^\ell\ u(t,x)$ at $t=0$ formally determined from equation (1.9) and initial condition (1.2).

Endowed with the metric

(2.10) $\rho(\bar{v},\bar{\bar{v}})=D_{s_\circ,s}(\bar{v}-\bar{\bar{v}})\ ,\ \forall\ \bar{v},\bar{\bar{v}}\in X_{s_\circ,s,E}$,

$X_{s_\circ,s,E}$ is a nonempty complete metric space, provided that $\varepsilon>0$ is suitably small.

Let $\tilde{X}_{s_\circ,s,E}$ be the subset of $X_{s_\circ,s,E}$ composed of all elements in $X_{s_\circ,s,E}$ with compact support in the variable x for any fixed $t\geq 0$.

we define a map

(2.11) $M:v\to u=Mv$

by solving the following initial value problem for linear wave equations for any $v\in\tilde{X}_{s_\circ,s,E}$

$$(2.12)\qquad \square u=\sum_{i,j=1}^{n} b_{ij}(v,Dv)u_{x_ix_j}+2\sum_{j=1}^{n} a_{\circ j}(v,Dv)u_{tx_j}+F(v,Dv),$$

$$(2.13)\qquad t=0:u=\varepsilon\phi(x),\quad u_t=\varepsilon\psi(x).$$

By means of some $L^p(p\geq 2)$ decay estimates for solutions to linear wave equations and some refined estimates on composite functions which can be used in the course of the proof to distinguish estimations for the solution itself and its derivatives, we can prove that if ε and E are suitably small, then M maps $\tilde{X}_{s_\circ,s,E}$ into itself and M is a contraction with respect to the metric of $X_{s_\circ-1,s-1,E}$. Therefore,, the contraction mapping principle can be used to get the following

THEOREM 1: Under assumptions (1.10)-(1.15), if $1\leq\alpha\leq 3$ and (1.8) holds, then for any integers $s_\circ$ and s such that $s_\circ\geq n+10$, $s_\circ+n+1\leq s\leq 2s_\circ-9$, there exist positive constants $\varepsilon_\circ$ and E so small that for any ε with $0<\varepsilon\leq\varepsilon_\circ$, Cauchy problem (1.9) (1.2) admits on $t\geq 0$ a unique global classical solution $u\varepsilon\tilde{X}_{s_\circ,s,E}$. Moreover, with eventual modification on a set with zero measure on $[0,\infty)$, for any T>0 we have

$$(2.14)\qquad u\varepsilon C([0,T];\ H^{s+1}(R^n)),$$

$$(2.15)\qquad u_t\varepsilon C([0,T];H^s(R^n)),$$

$$(2.16)\qquad u_{tt}\varepsilon C([0,T];H^{s-1}(R^n)).$$

§3. Case $\alpha\geq 4$

In this section we give the precise statement of our result and a sketch of the proof for the case that α is an integer ≥ 4 (correspondingly, n=2), cf. Li Ta-tsien and Chen Yun-mei [5].

For any given integers $s_\circ$ and s such that $s_\circ\geq 1$ and $s\geq s_\circ+n+1$ and any positive real number E, we introduce the following set of functions

$$(3.1)\qquad X_{s_\circ,s,E}=\{v=v(t,x)\ |\bar{D}_{s_\circ,s}(v)\leq E,\bar{\bar{D}}_{s_\circ,s}(v)\leq C_\circ E\}\ ,$$

where

$$(3.2)\qquad \bar{D}_{s_\circ,s}(v)=\sup_{t\geq 0}(\|v(t,\cdot)\|_{H^{s+1}(R^n)}+\|v_t(t,\cdot)\|_{H^s(R^n)})$$

$$+\sup_{t\geq 0}(1+t)^{\frac{n-1}{2}}(\|v(t,\cdot)\|_{W^{s_\circ+1,\infty}(R^n)}+\|v_t(t,\cdot)\|_{W^{s_\circ,\infty}(R^n)}),$$

$$(3.3)\qquad \bar{\bar{D}}_{s_\circ,s}(v)=\sup_{t\geq 0}\|v_{tt}(t,\cdot)\|_{H^{s-1}(R^n)}+\sup_{t\geq 0}(1+t)^{\frac{n-1}{2}}\|v_{tt}(t,\cdot)\|_{W^{s_\circ-1,\infty}(R^n)})$$

and C_0 is a positive constant to be determined.

Endowed with the metric

$$\rho(\bar{v},\bar{\bar{v}})=\bar{D}_{s_0,s}(\bar{v}-\bar{\bar{v}})+\bar{\bar{D}}_{s_0,s}(\bar{v}-\bar{\bar{v}}), \tag{3.4}$$

$X_{s_0,s,E}$ is a nonempty complete metric space for any fixed $C_0>0$.

We still define a map M by (2.11)-(2.13) for any $v\varepsilon X_{s_0,s,E}$. By means of some L^∞ decay estimates for solutions to linear wave equations, we can prove that if ε and E are suitably small, then there exists a positive constant C_0 such that M maps $X_{s_0,s,E}$ into itself and M is a contraction with respect to the metric of $X_{s_0-1,s-1,E}$. Therefore, the contraction mapping principle can be used to get the following

THEOREM 2: Under assumptions (1.10)-(1.15), if $\alpha\geq 4$ and (1.8) holds, then for any integers s_0 and s such that $s_0\geq 1$ and $s\geq s_0+n+1$, there exist positive constants ε_0 and E so small that for any ε with $0<\varepsilon\leq\varepsilon_0$, cauchy problem (1.9),(1.2) admits on $t\geq 0$ a unique global classical solution $u\varepsilon X_{s_0,s,E}$. Moreover, with eventual modification on a set with zero measure on $[0,\infty)$, for any T>0 we still have (2.14)-(2.16).

REFERENCES

[1] S.Klainerman, Uniform decay estimates and the Lorentz invariance of the classical wave equation, Comm. Pure Appl. Math., 38(1985), 321-332.

[2] D.Christodoulou, Global solutions of nonlinear hyperbolic equations for small initial data, Comm. Pure Appl. Math., 39(1986), 267-282.

[3] A.Matsumura, Initial value problems for some quasilinear partial differential equations in mathematical physics, Doctor Thesis, Kyoto Univ., 1980.

[4] Li Ta-tsien and Chen Yun-mei, Initial value problems for nonlinear wave equations, Commu. in Partial Differential Equations, 13 (1988), 383-422.

[5] Li Ta-tsien and Chen Yun-mei, A note on the global existence of classical solutions to nonlinear wave equations, to appear in Chin. Ann. of Math.

ONDES DE CHOC, ONDES DE RARÉFACTION ET ONDES SONIQUES MULTIDIMENSIONNELLES

G.MÉTIVIER
IRMAR
Université de Rennes I
Campus Beaulieu
35042 Rennes cedex

Introduction

L'étude des systèmes de lois de conservation unidimensionnels, notamment pour la résolution du problème de Riemann, fait intervenir un certain nombre "d'ondes simples" : ondes de choc, ondes de raréfaction, discontinuités de contact ; à cette liste, il est bon d'ajouter les ondes soniques (sound waves), qu'on appellera plutôt ici ondes de gradient, et qui apparaissent comme des discontinuités du gradient de la solution (ou de dérivées d'ordre supérieur).

Naturellement, ces constructions fournissent des solutions "mono-dimensionnelles" (solutions qui ne dépendent que d'une variable d'espace) de systèmes multidimensionnels. Une question immédiate est d'étudier la stabilité de telles soulutions 1-D , vis-à-vis de perturbations multi-D, d'autant plus que la justification des modèles 1-D consiste assez souvent à négliger des variables dans des modèles 3-D. Une question voisine est de constuire des solutions "proches" de ces solutions particulières 1-D.

Le but de cet exposé est de présenter un certain nombre de résultats récents allant dans ce sens, qui concernent les chocs (A.Majda [Ma1] [Ma2]), les ondes de raréfaction (S.Alinhac [A1] [A2] [A3]), les ondes soniques et les ondes "stratifiées" ([Mé1]), et aussi les chocs faibles. On voudrait aussi donner une méthode d'approche du problème et discuter quelques points significatifs.

1. Notations - Remarques préliminaires

1.1 Notations : on considère un système de lois de conservation :

$$\partial_t u + \sum_{1 \leq j \leq n} \partial_j f_j(u) = 0 \tag{1.1}$$

On note A_j la matrice jacobienne de f_j ; on convient que $A_o = Id$ et $\partial_o = \partial_t$. Ce système est supposé hyperbolique symétrique, c'est-à-dire qu'il existe une matrice symétrique définie positive, $S(u)$, telle que les SA_j sont toutes symétriques.

Pour $\theta \in \mathbb{R}^{n-1}$ (voisin de 0) on suppose que $\lambda(u,\ \theta)$ est une valeur propre simple de :

(1.2) $$A_n(u) - \sum_{1 \le j < n} \theta_j A_j(u)$$

on notera $r(u,\ \theta)$ un vecteur propre associé, et, lorsque λ est vraiment non linéaire, on fera la normalisation habituelle : $r.\nabla_u\lambda = 1$. Les ondes que nous allons étudier seront associées à cette valeur propre λ.

1.2 Changement de variables : la première difficulté que l'on rencontre, est que le front des ondes que l'on veut étudier, est inconnu ; pour rigidifier la géométrie, on utilise des changements de variables :

(1.3) $$(y, \tilde{x}_n) \to (y, x_n) \qquad \text{avec} \quad x_n = \phi(y, \tilde{x}_n) \qquad \text{et} \quad y = (t, x_1, \ldots, x_{n-1})$$

(1.4) $$\tilde{u}\,(y, \tilde{x}_n) = u(y, x_n) = u(y, \phi(y, \tilde{x}_n))$$

et, là où les fonctions sont de classe C^1, il est clair que (1.1) équivaut à :

(1.5) $$L(\tilde{u}, \phi)\,\tilde{u} = 0$$

où:

(1.6) $$L(v, \phi) = \partial_t + \sum_{1 \le j < n} A_j(v)\,\partial_j + \frac{1}{\partial_n\phi} A_n(v, \partial_y\phi)\,\partial_n$$

(1.7) $$A_n(v, \partial_y\phi) = A_n(v) - \sum_{0 \le j < n} \partial_j\phi\ A_j(v)$$

La forme du changement de variables (1.3) dépend évidemment du problème que l'on veut traiter. Il faut aussi bien comprendre que dans les équations (1.5) obtenues, ϕ est une des inconnues.

1.3 Linéarisation : un point important à mettre en évidence est la structure du linéarisé des équations (1.5).

LEMME 1 : *le linéarisé en* $(v,\ \psi)$ *de* $\mathcal{F}(v,\ \psi) = L(v,\ \psi)\,v$ *est un opérateur de la forme* :

(1.8) $$(u, \phi) \to L(v, \psi)\,u' + B(v, \psi)\,u' + \phi\ (\partial_n\psi)^{-1}\,\partial_n\mathcal{F}(v, \psi)$$

où $B(v,\ \psi)$ *est l' opérateur de multiplication par une matrice* B *dont les coefficients sont des fonctions de* $(v,\ \nabla v,\ \nabla\phi)$, *alors que* :

(1.9) $$u' = u - \phi\,\frac{\partial_n v}{\partial_n\psi}$$

L'apparition de la "bonne" inconnue u', s'explique très simplement : dans les variables initiales, le linéarisé en $\hat{v}$ de (1.1) est de la forme :

$$(1.10) \qquad \hat{u} \to \sum_{0 \leq j \leq n} A_j(\hat{v})\,\partial_j \hat{u} \; + \; B(\hat{v})\,\hat{u}$$

alors que la linéarisation de (1.4) donne : $u(y, \tilde{x}_n) = \hat{u}(y, \psi) + \phi\, \partial_n \hat{v}(y,\psi)$ et :

$$u'(y, \tilde{x}_n) = \hat{u}(y, \psi(y, \tilde{x}_n))$$

(1.8) s'obtient alors en reportant ce changement de variables et de fonctions dans (1.10).

REMARQUE : dans [Ma1], A.Majda ne calcule le linéarisé que sur un état v constant, auquel cas $u'=u$ et il n'y a pas de terme en ϕ ni $\nabla\phi$ dans (1.8). La remarque fondamentale faite par S.Alinhac, est que le changement $u \to u'$ efface de toutes façons les termes (gênants) en $\nabla\phi$.

2. Chocs

2.1 Les équations : on demande au changement de variables (1.3) de redresser le front Σ du choc en la surface $\{\tilde{x}_n = 0\}$, et alors Σ sera d'équation $x_n = \varphi(y)$ avec $\varphi(y) = \phi(y, 0)$. En oubliant les ~, on obtient les équations :

$$(2.1) \qquad \begin{cases} L(u^{\pm}, \phi^{\pm})\, u^{\pm} = 0 & \textit{dans} \quad \{\pm x_n > 0\} \\ [f_n(u)] = \sum\limits_{0 \leq j < n} \partial_j \varphi\, [f_j(u)] \quad \textit{et} \quad \phi = \varphi & \textit{sur} \quad \{x_n = 0\} \end{cases}$$

la condition aux limites sur $\{x_n = 0\}$ étant simpement l'expression de la condition de Rankine-Hugoniot. On remarque que ϕ n'est pas déterminé par ces équations (seulement φ), ce qui est naturel puisqu'on a seulement demandé au changement de variables de redresser Σ. A.Majda lève cette indétermination en cherchant ϕ sous la forme :

$$(2.2) \qquad \phi(y, x_n) = x_n + \varphi(y)$$

2.2 Stabilité uniforme : les équations linéarisées sont de la forme :

$$(2.3) \qquad \begin{cases} L(v^{\pm}, \psi)\, u^{\pm} = F^{\pm} & \textit{dans} \quad \{\pm x_n > 0\} \\ [A_n(v, \psi) u] - \sum\limits_{0 \leq j < n} \partial_j \varphi\, [f_j(u)] = G & \textit{sur} \quad \{x_n = 0\} \end{cases}$$

La condition de stabilité uniforme de A.Majda, consiste à dire que, pour ce problème la "condition de Lopatinski uniforme" est satisfaite. Cela se traduit par les estimations (maximales) suivantes, valables pour γ assez grand :

(2.4) $$\sqrt{\gamma}\,|u|_{o,\gamma} + |\Gamma u|_{o,\gamma} + |\varphi|_{1,\gamma} \leq C\{\frac{1}{\sqrt{\gamma}}\,|F|_{o,\gamma} + |G|_{o,\gamma}\}$$

où $\Gamma u = (\Gamma^+ u^+ , \Gamma^- u^-)$ désigne les traces de u^+ et u^- sur $\{x_n = 0\}$; pour a définie sur $\{\pm x_n > 0\}$ ou $\{x_n = 0\}$, $|a|_{o,\gamma}$ désigne la norme de $e^{-\gamma t} a$ dans l'espace L^2 sur le domaine correspondant ; enfin, $|\varphi|_{1,\gamma} = \gamma\,|\varphi|_{o,\gamma} + |\partial_y \varphi|_{o,\gamma}$.

REMARQUES 1 On renvoie à [Ma1] pour une traduction algébrique (sur les symboles) de cette condition de Lopatinski uniforme.

2 On renvoie aussi à [Ma1] pour une discussion de la pertinence de cette notion du point de vue des applications, notamment pour le système d'Euler de la dynamique des gaz.

3 Le contrôle de $|\Gamma u|_{o,\gamma}$ n'est envisageable que parce que le problème est non caractéristique.

4 Le gain d'une dérivée pour φ suppose que le système $\sum [f_j(v)]\,\partial_j$ soit elliptique ; cela ne peut avoir lieu que si $N \geq n$ et exclut donc les lois scalaires multidimensionnelles. Pour ces lois scalaires on a une estimation :

(2.5) $$\sqrt{\gamma}\,|u|_{o,\gamma} + |\Gamma u|_{o,\gamma} + \gamma|\phi|_{o,\gamma} \leq C\{\frac{1}{\sqrt{\gamma}}\,|F|_{o,\gamma} + |G|_{o,\gamma}\}$$

et une question intéressante serait de savoir si, de façon générale, une telle estimation suffit, par exemple pour construire des chocs.

<u>2.3 Construction de chocs</u> : dans [Ma2], A.Majda résout (2.1) avec la donnée de Cauchy :

(2.6) $$u^{\pm}\big|_{t=0} = u_o^{\pm} \qquad\qquad \varphi\big|_{t=0} = \varphi_o$$

Pour simplifier, on supposera $u_o^{\pm}$ C^∞ sur $\{\pm x_n \geq 0\}$ et φ_o C^∞ avec $\varphi_o(0) = \varphi_o'(0) = 0$.

Bien entendu, il faut que la donnée u_o vérifie un certain nombre de conditions de compatibilités. On suppose ici que $\lambda(v, \theta)$ est une valeur propre vraiment non linéaire (pour θ voisin de 0) et que $v^+ = U(a, v^-, \theta')$ est la courbe de Rankine-Hugoniot associée à cette valeur propre. La première condition de compatibilité consiste à dire qu'il existe une fonction $a(y')$ ($y' = (x_1, . . , x_{n-1})$) telle que :

(2.7) $$u_o^+(y') = U(a(y'), u_o^-(y'), \partial_y' \varphi_o)$$

Avec la normalisation habituelle, on suppose aussi que :

$$(2.8) \qquad a(y') \leq c < 0$$

de sorte que les conditions de Lax sont satisfaites. Les compatibilités d'ordre supérieur relient les traces sur $\{x_n = 0\}$ de $\partial_n^k u_o^+$ et de $\partial_n^k u_o^-$; pour des détails on renvoie à [Ma 2].

THÉORÈME 1(A.Majda [Ma 2]) : *supposons les données s -compatibles, avec* $s > \frac{n+3}{2}$. *Alors, le problème* (2.1) (2.6) *possède (au voisinage de* 0) *une solution telle que* $u^{\pm}$ *soit dans l'espace de Sobolev* H^s *sur* $\{\pm x_n \geq 0\}$ *et* φ *et* ϕ *soient dans* H^{s+1}.

En fait, A.Majda a démontré ce théorème pour $s > n + 7$, et A.Mokrane ([Mo]) a montré qu'on pouvait réduire s au niveau indiqué.

3. Ondes de gradient

3.1 Les équations stratifiées : le front Σ d'une onde de gradient est une surface caractéristique ; on peut choisir le changement de variable en sorte qu'il rectifie Σ seul, ou bien en sorte qu'il redresse toute une famille de surfaces caractéristiques parallèles. Considérons d'abord ce deuxième cas ; on obtient alors les équations :

$$(3.1) \qquad \begin{cases} L(u, \phi)\, u = 0 \\ \partial_t \phi = \lambda(u, \partial_y' \phi) \end{cases}$$

où $\partial_y' = (\partial_1, \ldots, \partial_{n-1})$. Notons que ce problème n'est pas un problème aux limites : les équations ont lieu dans tout l'espace.

Dans l'étude de ce problème, il est naturel (et intéressant) de travailler dans des espaces anisotropes. Notons $H^{o,s}$ l'espace des v tels que $\partial_y^\alpha v \in L^2$ pour $|\alpha| \leq s$, et E^s l'espace des $v \in H^{o,s}$ tels que $\partial_n v \in H^{o,s-2}$. Remplaçant L^2 par L^∞, on définit des espaces notés $L^{\infty,s}$ et Λ^s.

On peut énoncer un premier résultat concernant les **ondes "stratifiées"**:

THÉORÈME 2 (cf [Mé 1]) *si* (u, ϕ) *est une solution de* (3.1) *dans* $t<0$, *avec* $\partial_n \phi \neq 0$, *de régularité* $E^s \cap \Lambda^3$, *alors* (u, ϕ) *se prolonge en solution de régularité* $E^s \cap \Lambda^3$, *sur tout un voisinage de* 0.

Ce théorème est complété par des estimations a-priori qui expriment que, si

T_* est le temps d'existence de la solution stratifiée, et si T_* est fini, alors la norme de (u, ϕ) dans $\Lambda^3(\,t < T\,)$ n'est pas bornée lorsque $T \to T_*$.

On peut aussi résoudre le problème de Cauchy pour (3.1), mais les conditions de compatibilité ne sont pas explicites (cf néanmoins ci-dessous). Il faut alors écrire la condition initiale sous la forme :

(3.2) $$u = v \quad \text{et} \quad \phi = \psi \quad \text{sur } t=0$$

où v et ψ sont tels que $L(v, \psi)\, v$ et $\partial_t \psi - \lambda(v, \partial'_y \psi)$ sont nuls à l'ordre s sur $t=0$.

Pour les **ondes soniques** , on peut expliciter les conditions de compatibilité : si les données de Cauchy :

(3.3) $$u\big|_{t=0} = u_o \qquad \phi\big|_{t=0} = \phi_o$$

sont telles que les restrictions de u_o et ϕ_o à $\{\pm x_n \geq 0\}$ sont C^∞ et que :

(3.4) $$[\phi_o]=0 \qquad \text{et} \qquad \left|\partial_n \phi_o\right| \geq c > 0$$

la première compatibilité est simplement :

(3.5) $$[u_o] = 0$$

On peut ensuite expliciter les conditions d'ordre supérieur, qui à nouveau relient les traces sur $\{x_n = 0\}$ de $\partial_n^k u_o^+$ et de $\partial_n^k u_o^-$, et il est facile de construire des données compatibles (cf [Mé 1]).

Le théorème 2 s'applique, mais ne fournit pas la régularité H^s (en x_n) à laquelle on s'attend pour $x_n \neq 0$. En fait, dans ce problème, il est naturel de remplacer l'espace $H^{o,s}$ "stratifié" par l'espace "conormal" $H^{o,s}_\Sigma$ des v tels que $(x_n \partial_n)^k \partial_y^\alpha v \in L^2$ pour $k+|\alpha| \leq s$. L'espace E^s est alors remplacé par un espace que l'on note E^s_* ; procédant de même, Λ^μ est remplacé par un espace noté Λ^μ_* .

THÉORÈME 3 : *si* ϕ_o *et* u_o *vérifient* (3.4) (3.5) *et sont s-compatibles, avec* $s > \frac{n}{2}+6$, *alors le problème* (3.1) (3.3) *possède une (unique) solution au voisinage de* 0 *dans* $E^s_* \cap \Lambda^4_*$ *qui vérifie* $[u]=0$ *et* $[\phi]=0$.

A nouveau, ce théorème est accompagné d'estimations a-priori.

3.2 Les équations non stratifiées : si on ne redresse qu'une seule surface caractéristique, on obtient les équations suivantes :

$$(3.6)\qquad \begin{cases} L(u^{\pm}, \phi^{\pm})\, u^{\pm} = 0 & \textit{dans} \quad \{\pm\, x_n > 0\} \\ [u] = 0 \qquad [\phi] = 0 & \textit{sur} \quad \{x_n = 0\} \\ \partial_t \varphi = \lambda(u, \partial'_y \varphi) & \textit{sur} \quad \{x_n = 0\} \end{cases}$$

où $\varphi = \Gamma^{\pm} \phi^{\pm}$. Les données de Cauchy sont :

$$(3.7)\qquad u\,\big|_{t=0} = u_o \qquad \varphi\,\big|_{t=0} = \varphi_o$$

où u_o est comme avant et φ_o est C^{∞}. La première compatibilité est toujours (3.5), et la construction de données compatibles est la même que précédemment.

THÉORÈME 4 : *si* φ_o *et* u_o *sont s-compatibles, avec* $s > \frac{n}{2}+7$ *, alors le problème* (3.6) (3.7) *possède une solution au voisinage de* 0 *dans* $\mathcal{E}^s \cap \mathcal{L}^4$.

Dans cet énoncé, $\mathcal{E}^s$ [resp $\mathcal{L}^{\mu}$] désigne l'espace des $u \in E^s_*$ [resp Λ^{μ}_*] tels que $\partial_n u \in E^{s-2}_*$ [resp $\Lambda^{\mu-2}_*$]. Dans la preuve de ce théorème, on détermine ϕ à partir de φ en cherchant

$$(3.8)\qquad \phi = x_n + R\varphi$$

où R est un opérateur de relèvement de traces convenable.

3.3 Le linéarisé de (3.6) : le lemme 1 donne la linéarisation de l'équation d'intérieur, en faisant intervenir la "bonne" inconnue u'. La linéarisation de la condition $[v] = 0$, conduit trivialement à :

$$(3.9)\qquad [u] = [u'] + \varphi\,[z] = 0$$

(où $z = (\partial_n \psi)^{-1}\, \partial_n v$). Si on note $F = \mathcal{F}(v, \psi)$, on a :

$$(3.10)\qquad A_n(v, \partial_y \psi)\, z = F - \sum_{j=0}^{n-1} A_j(v)\, \partial_j v$$

et, si $[v] = 0$ et $[\psi] = 0$, on en déduit que :

$$(3.11)\qquad A_n(v, \partial_y \psi)\,[z] = 0$$

En reportant dans (3.9), on obtient avec le lemme 1 le problème linéarisé suivant :

$$(3.12)\qquad \begin{cases} L(v^{\pm}, \psi^{\pm})\, u'^{\pm} = F^{\pm} & \textit{dans} \quad \{\pm x_n > 0\} \\ A_n(v, \psi)\,[u'] = 0 & \textit{sur} \quad \{x_n = 0\} \end{cases}$$

La remarque fondamentale que l'on fait alors, est que les conditions aux limites dans (3.12) sont maximales dissipatives, ce qui donne des estimations d'énergie de la forme :

(3.13) $$\gamma\ |u'|_{o,\gamma} \leq C\ |F|_{o,\gamma}$$

Si on note $\|.\|_{s,\gamma}$ la norme, pour les poids $e^{-\gamma t}$, de l'espace E^s_* (introduit avant le théorème 3), on a aussi :

(3.14) $$\gamma\ \|u'\|_{2,\gamma} \leq C\ \|F\|_{2,\gamma}$$

Comme on a, en notant $|.|_{s,\gamma}$ la norme à poids $e^{-\gamma t}$ de $H^s\,(x_n{=}0)$:

(3.15) $$\sqrt{\gamma}\ \ |\Gamma u'|_{s-1,\gamma} \leq C\ \|u'\|_{s,\gamma}$$

on voit que (3.14) fournit un contrôle de $|\Gamma u'|_{1,\gamma}$.

Lorsque $[z]$ est partout non nul, et si ℓ est tel que $\ell.[z]\neq 0$, on tire de (3.9) que:

(3.16) $$\varphi = (\ell.\,[z]\,)^{-1}\ (\ell.[u'])$$

ce qui nous fournit directement un contrôle de φ puis de ϕ par (3.8).

Mais on peut aussi, sans aucune hypothèse sur $[z]$, ce qui lui permet de s'annuler, remplacer (3.16) par le linéarisé de la dernière équation de (3.6) :

(3.17) $$\partial_t\,\varphi \;=\; \sum_{j=1}^{n-1} \frac{\partial\lambda}{\partial\theta_j}(v,\partial'_y\psi)\ \partial_j\varphi \;+\; \frac{\partial\lambda}{\partial v}(v,\partial'_y\psi)\,u$$

qui fournit (pour γ assez grand) une estimation de la forme :

(3.18) $$\gamma\ |\varphi|_{1,\gamma} \leq\ C\ |\Gamma u|_{1,\,\gamma}$$

et donne le contrôle de φ.

4. Chocs faibles

On peut d'abord remarquer que la notion de choc faible a bien un sens :

LEMME 2 : *soit* $u^{\pm}$ *et* ϕ *une solution de classe* C^1 *de* (2.1) (2.6), *telle que* $|[u_{|t=0}]| \leq \varepsilon$ [*resp* $|[u_{|t=0}]| \geq \varepsilon$] . *Alors sur un voisinage de* 0 *indépendant de* ε *on a* : $|[u]| \leq C\varepsilon$ [*resp* $|[u]| \geq C^{-1}\,\varepsilon$].

Dans toute la suite de ce paragraphe, les résultats, ne concernent (pour le moment) que le système d'Euler des gaz isentropiques.

4.1 Résultats : on revient au problème (2.1) (2.6), $u_o^{\pm}$ et φ_o restant dans des bornés de fonctions C^∞. On suppose toujours que la condition (2.7) est satisfaite, mais on remplace (2.8) par :

(4.1) $\quad a(y') = -\varepsilon\, \hat{a}(y') \qquad$ avec $\qquad |Log\,(\hat{a})| \leq C$

THÉORÈME 5 *: on suppose les données s-compatibles avec* $s>\frac{n}{2}+7$ *. Alors, il existe un voisinage de* 0 *indépendant de* $\varepsilon \in]0, \varepsilon_0]$ *, tel que le problème* (2.1) (2.6) *possède, sur ce voisinage, une solution dans* $\mathcal{E}^s \cap \mathcal{L}^4$ *, de norme majorée indépendamment de* ε*.*

De nouveau, dans la preuve de ce théorème, ϕ et φ sont reliés par un choix du type $\phi = x_n + R\varphi$.

Le théorème 4 apparait simpement comme le cas limite $\varepsilon=0$ dans le théorème 5. On peut construire des familles $(u_o^\varepsilon\ , \varphi_o^\varepsilon)$ de données s-compatibles qui verifient (2.7) avec (4.1), et qui convergent vers $(u_o\ , \varphi_o)$; on remarque alors que les données $(u_o\ , \varphi_o)$ sont s-compatible au sens du théorème 4. Avec les théorèmes 4 et 5 on a donc une famille de chocs faibles $(u^\varepsilon\ , \phi^\varepsilon)$ et une onde sonique $(u\ , \phi)$ définies sur un même voisinage de 0, et bornées dans $\mathcal{E}^s \cap \mathcal{L}^4$. Si on a fixé la correspondance $\varphi \rightarrow \phi$, on a **convergence des chocs faibles vers l'onde sonique** :

THÉORÈME 6 : *étant donnée une famille* $(u_o^\varepsilon\ , \varphi_o^\varepsilon)$ *de données s-compatibles qui converge vers* $(u_o\ , \varphi_o)$ *comme indiqué ci-dessus, on peut construire les chocs faibles* $(u^\varepsilon\ , \phi^\varepsilon)$ *et l'onde sonique* $(u\ , \phi)$ *des théorèmes* 5 *et* 6, *de sorte que l' on ait en plus convergence dans* L^2*, de* $(u^\varepsilon\ , \phi^\varepsilon)$ *vers* $(u\ , \phi)$.

4.2 Stabilité "uniforme en ε" : considérons une famille $(v_\varepsilon\ , \psi_\varepsilon)$ de chocs vérifiant (4.1) avec $Log\ |\hat{a}_\varepsilon|$ borné dans $W^{1,\infty}$. Notons $(2.3)_\varepsilon$ le système (2.3) correspondant à $(v_\varepsilon\ , \psi_\varepsilon)$. Parce que la matrice $A_n(v_\varepsilon\ , \partial_y\psi_\varepsilon)$ a une valeur propre de l'ordre de ε, on ne peut pas espérer d'estimation (2.4)) uniforme en ε pour les solutions de $(2.3)_\varepsilon$. Néanmoins,dans le cas du système d'Euler de la dynamique des gaz, on a les estimations suivantes pour $\gamma \geq \gamma_0$:

(4.2) $$\sqrt{\gamma}\,|u|_{o,\gamma} + \sqrt{\varepsilon}\,|\Gamma u|_{o,\gamma} + \sqrt{\varepsilon}\,|\varphi|_{1,\gamma} \le C\{\frac{1}{\sqrt{\gamma}}\,|F|_{o,\gamma} + \frac{1}{\sqrt{\varepsilon}}\,|G|_{o,\gamma}\}$$

avec C et γ_o indépendants de ε (assez petit).

On voit bien dans cette estimation comment on perd le contrôle de $|\Gamma u|_{o,\gamma}$. Le terme en $(\sqrt{\varepsilon})^{-1}\,|G|_{o,\gamma}$ pourrait inquiéter, mais la procédure de linéarisation conduit en fait à appliquer l'estimation (4.4) à des fonctions G qui sont déjà de la forme $\varepsilon\, G'$.

Il est cependant clair que l'on a besoin d'un contôle uniforme en ε de Γu et ϕ. En reprenant les notations du paragraphe 3.3, on peut montrer que :

(4.3) $$\sqrt{\gamma}\,\|u\|_{2,\gamma} + \gamma\,|\Gamma u|_{1,\gamma} + \gamma^2\,|\varphi|_{1,\gamma} \le C\{\frac{1}{\sqrt{\gamma}}\,\|F\|_{2,\gamma} + \frac{1}{\sqrt{\varepsilon}}\,|G|_{2,\gamma}\}$$

Quand on compare cette estimation avec (2.4), on constate une perte de régularité des traces Γu et de γ (compensée par un gain de poids sur γ). En outre, il y a perte de régularité entre G et Γu.

Mais d'autre part, pour $G=0$ (ou au moins pour $G=\varepsilon\, G'$), il est clair que le problème $(2.3)_\varepsilon$ "converge" vers le problème (3.12), et (3.14) (3.15) (3.18) signifient que l'estimation (4.3) est encore vraie pour $\varepsilon=0$.

5. Ondes de raréfaction

5.1 Les équations : dans les variables initiales, le motif géométrique généralise celui bien connu de la dimension 1 ; il est constitué de deux surfaces Σ et Σ', issues d'une surface Σ_0 (portant les discontinuités de la donnée initiale) et limitant un "dièdre" $\mathcal{W}$; en dehors de $\mathcal{W}$ la solution u est régulière jusqu'au bord alors que dans $\mathcal{W}$, u se comporte comme une fonction régulière des coordonnées "cylindriques" (t, y, θ) , θ étant quelque chose comme l'angle polaire dans $\mathcal{W}$.

Du côté des variables redressées, on a donc trois régions : $D^-=\{x_n < 0\}$, $D^+=\{x_n > 1\}$ et $D=\{0 < x_n < 1\}$. Le changement de variables Φ aura donc trois déterminations ϕ^+, ϕ^- et ϕ respectivement dans D^+, D^-, et D. Ces fonctions sont reliées par les relations :

(5.1) $$\phi^+(y, 1) = \psi(y, 1) \qquad \text{et} \qquad \phi^-(y, 0) = \psi(y, 0)$$

qui assurent la continuité de Φ. Σ et Σ' sont les images de $\{x_n = 0\}$ et $\{x_n = 1\}$. Si on note $x_n = \varphi_o(y')$ l'équation de Σ_o, on doit avoir :

(5.2) $$\phi^+(0, y', 1) = \phi(0, y', 1) = \phi(0, y', 0) = \phi^-(0, y', 0) = \varphi_o(y')$$

Enfin, on exprime que ϕ a exactement la singularité des coordonnées cylinriques :

(5.3) $$\partial_n \phi(t, y', \tilde{x}_n) = c\,t \quad \text{avec } c \text{ fonction} > 0$$

Suivant S.Alinhac, un représentant de l'onde de raréfaction est la donnée de $U = (u^+, u^-, u)$ et $\Phi = (\phi^+, \phi^-, \phi)$ qui vérifient (5.1-2-3) et :

(5.4) $$\begin{cases} L(u^\pm, \phi^\pm)\, u^\pm = 0 & \textit{dans} \quad D^\pm \\ L(u, \phi)\, u = 0 & \textit{dans} \quad D \\ u^+ = u & \textit{sur} \quad x_n = 1 \\ u^- = u & \textit{sur} \quad x_n = 0 \end{cases}$$

Les données initiales sont (5.2) et :

(5.5) $$\begin{cases} u^+_{|t=0} = u_o^+ & \textit{pour} \quad x_n > 1 \\ u^-_{|t=0} = u_o^- & \textit{pour} \quad x_n < 0 \end{cases}$$

$\lambda(v, \theta)$ étant à nouveau supposée vraiment non linéaire, on note $\mathcal{U}(a, v, \theta)$ la courbe intégrale de $r(v, \theta)$, issue de v pour $a = 0$. La première condition de compatibilité s'énonce : il existe une fonction $a(y')$ telle que :

(5.6) $$a(y') \geq c > 0$$

(5.7) $$u_o^+(y') = \mathcal{U}(a(y'), u_o^-(y'), \partial'_y \varphi_o)$$

Comme précédemment, les compatibilités d'ordre supérieur relient les traces sur $\{x_n = 0\}$ de $\partial_n^k u_o^+$ et de $\partial_n^k u_o^-$; pour des détails on renvoie à [A. 2].

THÉORÈME 7 (S. Alinhac [A. 2]) : *supposons les données k–compatibles, avec k assez grand. Alors le problème* (5.4) (5.5) *avec* (5.1-2-3), *possède, au voisinage de* 0, *une solution telle que :* $u^\pm$, u , $\phi^\pm$ et ϕ *sont de régularité* H^{k-d} *en dehors de* $x_n = 1$ et $x_n = 0$, *et de régularité* $H^{(k-d)/2}$ *pour* $t > 0$, *près de* $x_n = 1$ *ou* $x_n = 0$.

On renvoie à [A.2] pour une discussion de l'indice d . Signalons seulement qu'on peut le prendre égal à 0 dans le cas du système d'Euler isentropique.

Notons pour finir que S.Alinhac a montré l'unicité (dans les coordonnées initiales !) de la solution ayant un motif d'onde de raréfaction comme ci-dessus.

5.2 Le problème linéarisé : il est de la forme :

$$(5.8)\qquad \begin{cases} L(v^{\pm}, \psi^{\pm})\, u^{\pm} = F^{\pm} & dans \quad D^{\pm} \\ L(v, \psi)\, u = F & dans \quad D \\ A_n(v^{+}, \psi^{+})\, \{u^{+} - u\} = 0 & sur \quad \{\tilde{x}_n = 1\} \\ A_n(v^{-}, \psi^{-})\, \{u - u^{-}\} = 0 & sur \quad \{\tilde{x}_n = 0\} \end{cases}$$

(les v et ψ sont tels que : $v^{+}=v$, $\psi^{+}=\psi$ sur $\{\tilde{x}_n=1\}$ et $v^{-}=v$, $\psi^{-}=\psi$ sur $\{\tilde{x}_n=0\}$).

A nouveau les conditions aux limites sont maximales dissipatives. La difficulté nouvelle est que $L(v, \psi)$ est un opérateur singulier, puisque ψ est un changement de variables singulier sur $\{t=0\}$. En fait, $L(v, \psi)$ est de la forme :

$$(5.9)\qquad L = t\,\partial_t + \sum_{j=1}^{n-1} t A_j \partial_j + A_n \partial_n$$

Dans ces conditions, il est naturel de travailler avec des poids de la forme t^{γ}. En fait, S.Alinhac a montré que pour d assez grand :

$$(5.10)\quad \left|t^{-d-1} u^{\pm}\right|_o + \left|t^{-d-1/2} u\right|_o \leq C \{ \left|t^{-d-1} F^{\pm}\right|_o + \left|t^{-d-1/2} F\right|_o \}$$

où $|.|_o$ désigne ici la norme L^2 prise sur $]0,T[\times \mathbb{R}^n$, T étant assez petit.

On a évidemment intérêt à prendre d le plus petit possible, et il est remarquable, comme le souligne S.Alinhac, que pour le système d'Euler isentropique on puisse prendre $d = 0$.

Comme on l'a fait pour les ondes soniques, il faut ensuite travailler un peu plus pour obtenir des estimations sur les traces de u, et S.Alinhac utilise ensuite (3.16) pour contrôler Φ.

6. Remarques

Considérons le problème de Cauchy pour (1.1) avec donnée de Cauchy :

$$(6.1)\qquad u\,|_{t=0} = u_o$$

avec u_o discontinue sur une hypersurface Σ_o ; ce problème se présente comme une perturbation du classique problème de Riemann 1-D. On pourrait s'attendre à ce que la solution soit, comme dans le cas 1-D, la juxtaposition d'ondes simples. Néanmoins il reste à comprendre le phénomène des discontinuités de contact multi-D, qui, au moins dans le cas d'Euler isentropique, semblent être violemment instables. Ce problème reste donc largement ouvert.

Dans la cas où le système (1.1) n'a que des valeurs propres vraiment non linéaires, on peut espérer résoudre le problème (1.1) (6.1) en juxtaposant des chocs et des ondes de raréfaction (avec éventuellement des ondes de gradient). Un résultat dans ce sens a été donné en [Me 2], pour des systèmes 2 × 2 lorsque les deux ondes sortantes sont des chocs.

Pour finir, citons le travail de E.Harabetian [H] qui résout le problème de Cauchy (1.1) (6.1) lorsque les données sont analytiques. Ce travail qui construit des solutions multi-D "semblables" aux solutions 1-D, est évidemment très intéressant, mais il contourne les problèmes de stabilité C^∞ ou Sobolev (par exemple, les temps d'existence dépendent des domaines d'analyticité des données).

Bibliographie

[A 1] S.Alinhac : Existence d'ondes de raréfaction pour des écoulements isentropiques ; Séminaire Ecole Polytechnique 86-87.

[A 2] S.Alinhac : Existence d'ondes de raréfaction pour des systèmes quasi-linéaires hyperboliques multidimensionnels ; Comm. in Partial Diff. Equ.,1988, à paraître.

[A 3] S.Alinhac : Unicité d'ondes de raréfaction pour des systèmes quasilinéaires hyperboliques multidimensionnels ; preprint.

[H] E.Harabetian : a convergent series expansion for hyperbolic systems of conservation laws ; Trans. Amer. Math. Soc.,294 (1986) pp 383-424.

[Ma1] A.Majda : The stability of multidimensional shock fronts ; Mem. Amer. Math. Soc., n° 275 (1983).

[Ma2] A.Majda : The existence of multidimensional shock fronts ; Mem. Amer. Math. Soc., n° 281 (1983).

[Mé1] G.Métivier : Ondes soniques ; Séminaire Ecole Plytechnique 88-89.

[Mé2] G.Métivier : Interaction de deux chocs pour un système de deux lois de conservation en dimension deux d'espace ; Trans. Amer. Math. Soc., 296 (1986) pp.431-479.

[Mo] A.Mokrane : Problèmes mixtes hyperboliques non linéaires ; Thèse 3$^{\underline{\text{ème}}}$ cycle Rennes (1987).

THE INTERACTION OF TWO PROGRESSING WAVES

Guy Métivier
IRMAR
Universite de Rennes I, Campus Beaulieu
35042 Rennes Cedex
France

Jeffrey Rauch
Department of Mathematics
University of Michigan
Ann Arbor, MI. 48109
USA

Progressing waves of discontinuities are well modelled by conormal distributions with their singular support on the wavefront, Σ. For a regular embedded hypersurface Σ, a distribution, u, defined on a neighborhood of Σ is said to be conormal iff their is an $s \in \mathbb{R}$ such that for any finite set of smooth vector fields $V_1, \ldots, V_N$ tangent to Σ we have $V_1 \cdots V_N u \in H^s_{loc}$ [H§18.2]. Examples are single and double layers on Σ, the solutions of $\Box u = \delta$ at $t^2 = |x^2| > 0$, and, the piecewise smooth functions singular accross Σ. In the latter case one can take s equal to the order of the lowest derivative which is discontinuos accross Σ.

Suppose that Σ is a regular characteristic surface for the strictly hyperbolic semilinear differential system

$$P_m(D)u = F(x, D^{m-1}u).$$

It is known that there are many conormal solutions singular along Σ. If solutions are conormal or piecewise smooth in the past they remain so in the future provided Σ remains regular and $D^{m-1}u$ is locally bounded on a set, Ω, so large that Σ is in the domain of determinacy of $\Omega\cap\{t<0\}$ [Bo1,M,A]. Good surveys are presented in [Bo2] and [Be].

Subclasses, called classical conormal, have a more easily managed symbolic calculus and are therefore useful in constructing solutions with desired properties. Rauch and Reed [RR4] introduced a class which has too advantages: 1. The analysis is made almost exclusively in the t,x variables, and 2. Jump discontinuities in $D^{m-1}u$ are permitted.

The smallest class of classical conormal distibutions are the piecewise smooth functions which are singular only at Σ. Hadamard [H] analysed such solutions and Courant and Lax continued the study including the construction of solutions The analysis of the propagation of singularities and the construction of solutions in the linear case. This matierial is presented in [C §VI.4]. Stability under propagation by semilinear hyperbolic equations is proved in [RR2].

Bony [Bo1] proved that the class of conormal solutions is stable under pairwise interaction in the following sense. Suppose that Σ_1 and Σ_2 are two characteristic surfaces which cross transversally in $\Gamma\equiv\Sigma_1\cap\Sigma_2$. Let $\Sigma_3,\Sigma_4,\ldots.,\Sigma_\mu$ be the other characteristic surfaces passing through Δ. If a solution has the following properties:

i. The surfaces are regular

ii. $D^{m-1}u$ is locally bounded on the domanin of definition, Ω, of u, and $\Sigma_j \cap \{t>0\}$ is in the domain of determinacy of $\Omega\cap\{t<0\}$ for all j.

iii. In t<0, the solution is conormal with respect to Σ_1 and

Σ_2 and has sinqular support disjoint from Γ.

Then, the solution has singular support in the union of the Σ_j and is conormal at all points of $\Sigma_j \backslash \Gamma$.

If in addition, the solution is known to be piecewise smooth in $t<0$ one can ask whether it remains so in $t>0$. For linear problems the response is yes by simple superposition. For nonlinear problems the response is often yes. Note that the surfaces Σ locally cut space-time into 2μ wedges and piecewise smooth means smooth in the closure of each wedge.

Theorem. [MR] The interaction of two piecewise smooth waves as described above results in a piecewise smooth solution provided that the locus of interaction, Γ, is contained in a spacelike hypersurface.

Two important special cases are the case of systems which have only two sound speeds and the case of systems in one space dimension. In both cases, the hypothesis on Γ is automatically satisfied, and, the corresponding stability of piecewise smooth solution under interaction had been previously proved [RR-1,3,5]. Similarly, if one considers the Cauchy problem with piecewise smooth data singular accross $\Gamma \subset \{t=0\}$ there is local existence of a piecewise smooth solution singular along the characteristic hypersurfaces through Γ [MR].

The interaction of more than two progressing waves and the behavior when Σ degenerates are more complicated geometrically and analyticallv. We make no attemot at describinq the important work of Bony, Melrose, Ritter, Lehrner, Beals and Lebeau concerning these questions.

Without the hypothesis on Γ, the above Theorem would not be correct. The interaction of piecewise smooth waves can generate logarithmic singularities which propagate along the surfaces Σ_j. Even with the hypothesis on Γ, logarithmic singularities appear when classical conormal waves of the sort studied in [RR4] interact. The example, due to Piriou [P2], is a variant of a now classical example

$$(\partial_t \pm \partial_x)u_\pm = 0, \qquad u_\pm(0,x) = |(x \mp 1/2)^\pm|^{1/2} ,$$

$$\partial_t w = u_+ u_- \qquad w(0,x) = 0.$$

Two approaching "root x" singularities interact at (1,1) and logarithms are present in the outgoing singularity along the characteristic x=0 for t>0. In all the known examples producing unwanted logarithms, these terms do not appear in the principal symbol. This suggest the following problem.

Open problem. Show that the interaction of classsical conormal waves without logarithms produce waves which have no logarithms in their principal part. The natural starting point would be incoming waves which are piecewise smooth. Next, incoming waves as in [RR4]. Finally, incoming waves as in [Mel] with the proviso that logarithms are not present until terms one derivative smoother than the principal part.

In the remainder of this note, we will present an example exhibiting the production of logarithmic singularities from the interaction of piecewise smooth waves along a Γ which is not contained in a spacelike manifold.

In $\mathbb{R}^3=\mathbb{R}_t\times\mathbb{R}^2_x$ we denote by $\Box_c$ the d'Alembertian with speed $c>0$,

$$\Box_c \equiv \partial_t^2 - c^2\Delta.$$

We begin by considering the system of equations

$$\begin{aligned} &\Box_1 u = 0\\ (1)\qquad &\Box_1 v = 0\\ &\Box_2 w = uv. \end{aligned}$$

The solutions u and v are characteristic functions of halfspaces,

$$u \equiv \chi_{\{x_1-t<0\}},$$

$$v \equiv \chi_{\{x_2-t<0\}}.$$

Then, uv is the characteristic function of a wedge, W, whose edge is the line $x_1=x_2=t$ which moves with speed $2^{1/2} < 2$. For a $\Box_2$ observer this is slower than light, that is, inside the forward light cone. To find a w we take advantage of the Lorentz invariance of $\Box_2$. There is a Lorentz transformation $t,x\longmapsto t'x'$ preserving $\Box_2$ which maps W onto a wedge whose edge is the t'-axis, $\{x'=0\}$. The image wedge is therefore a cartesian product $\mathbb{R}_{t'}\times W'$ with W' a wedge in $\mathbb{R}^2_{x'}$. The equation for w is transformed to

$$\Box_2 w' = \chi_{W'}(x').$$

We can find solutions which do not depend on t' by solving

$$(2)\qquad \Delta_{x'}w'(x') = \chi_{W'}(x').$$

Note that the right hand side of (2) is piecewise smooth. As we will see, the solution is not piecewise smooth.

Let V_1 and V_2 be constant vector fields on $\mathbb{R}^2_{x'}$, parallel to the sides of W'. With a nonzero constant, a, we have

$$V_1V_2\chi_{W'} = a\delta(x').$$

Applying V_1V_2 to both sides of (2) yields

$$\Delta_{x'}V_1V_2w' = a\delta(x').$$

Therefore

$$V_1V_2w' = 2\pi a \ln(|x'|) + C^\infty.$$

In the original coordinates, we have constructed a solution which has a second derivative with a logarithmic singularity along the curve of intersection of the incoming waves. In particular, the solution is not piecewise smooth.

The construction above can be altered to remove three apparent flaws. First, the system (1) is not strictly hyperbolic. Second, the set of interaction is unbounded, extending to $t=-\infty$. Third, the failure of piecewise smoothness occurs uniquely along the line of interaction.

For the first, replacing one of the speed 1 d'Alembertians by speed $1+\sigma$, with $|\sigma|$ small and changing t to $(1+\sigma)t$ in the definition of v changes Γ to a line which still has speed less than two. In this way we get a strictly hyperbolic system.

For the second, we modify the construction in two stages. First, we replace the equation for w by

$$\Box_2\underline{w} = \eta(t,x)uv, \tag{3}$$

$$\underline{w}(0,x) = \underline{w}_t(0,x) = 0, \tag{4}$$

where η is a cutoff function supported near $(1,1,1)$, precisely

$$\text{supp}\ \eta \subset \{(t-1)^2+|x-(1,1)|^2 < \varepsilon^2/4\},$$

$$\eta \equiv 1 \text{ on } \{(t-1)^2+|x-(1,1)|^2 < \varepsilon^2/16\},$$

with $\epsilon>0$ small. Then ηuv vanishes for $t<1-\epsilon/2$ and the same is therefore true of $\underline{w}$.

Returning to the example u,v,w above, note that in the primed coordinates w' and the product $u'v'$ are independent of t'. Thus,

$$WFw' \cup WFu'v' \subset \{\tau' = 0\} \subset Ell(\Box_2),$$

where Ell, the set of elliptic points, is the complement of the characteristic variety. By invariance, we have

(5) $$WFw \cup WFuv \subset Ell(\Box_2)$$

in the original coordinates. In particular,

$$WF\eta uv \cap char(\Box_2) = \phi.$$

Propagating from $t<1-\epsilon$ using Hormander's Theorem yields

(6) $$WF\underline{w} \cap char(\Box_2) = \phi.$$

Combining (5) and (6) we see that

(7) $$WF(w-\underline{w}) \cap char(\Box_2) = \phi.$$

On the other hand we have

$$\Box_2(w-\underline{w}) = 0 \quad on \quad \{(t-1)^2+|x-(1,1)|^2 < \epsilon^2/16\}.$$

The microlocal elliptic regularity theorem then implies that

(8) $$WF(w-\underline{w}) \subset char(\Box_2) \text{ over } \{(t-1)^2+|x-(1,1)|^2 < \epsilon^2/16\}.$$

Combining (7) and (8) we see that w and $\underline{w}$ differ by a smooth function on the set $(t-1)^2+|x-(1,1)|^2 < \epsilon^2/16$. In particular, $\underline{w}$ is not piesewise smooth.

Having localized the interaction, we can now truncate the initial data of u,v. Let u,v be the solution of the initial value problem

(9) $$\Box_1\underline{u} = 0, \quad \Box_1\underline{v} = 0,$$

(10) $$u(0,x) = \varphi(x)u(0,x), \quad \underline{u}_t(0,x) = \varphi(x)u_t(0,x)$$

(11) $$\underline{v}(0,x) = \psi(x)v(0,x), \quad \underline{v}_t(0,x) = \psi(x)v_t(0,x).$$

Here $\varphi \in \mathcal{D}(\mathbb{R}^2)$ has support in a disc of radius ε about (1,0) and is identically equal to one on the disc of radius $\varepsilon/2$. The function ψ performs a similar cutoff at the point (0,1). Then

$$\operatorname{supp} \underline{u} \subset \{x_1-t < 0\}, \quad \operatorname{supp} \underline{v} \subset \{x_2-t < 0\},$$

$$\operatorname{supp} \underline{uv} \subset \text{the wedge } W.$$

The backward speed one light cone from $(1+\varepsilon/2,1,1)$ intersects the support of the Cauchy data inside the discs of radius $\varepsilon/2$ centered at (1,0) and (0,1). In these discs the data is the same as that for u,v so $\underline{uv}=uv$ inside this backward light cone. In particular, $\eta uv=\eta\underline{uv}$. Thus $\underline{w}$ satisfies

(12) $$\Box_2\underline{w} = \eta\underline{uv}.$$

Equations (4) and (9-12) give a globally solvable initial value problem and $\underline{w}$ has a second derivative with a logarithmic singularity along $x_1=x_2=t$.

For the third objection we add a fourth equation whose purpose is to transport the singularites of $\underline{w}$. Near (1,1,1), $\underline{w}$ is singular along $t=x_1=x_2$ and has wavefront set containing the conormal variety to this line,

$$WF\underline{w} \supset \{(\tau,\xi):\tau+\xi_1+\xi_2=0\}.$$

We want to use w as a source term in a hyperbolic equation whose characteristic variety meets this set and so that the entire system remains strictly hyperbolic. The simplest choice is

(13) $\partial_t z = \zeta(t,x)\underline{w}$

(14) $z(0,x) = 0.$

Here, ζ is a cutoff supported where η is identically one and itself equal to one on a neighborhood of (1,1,1).

To study z, make a linear change of variables $t,x \longmapsto T,X$ such that

i. $\partial_t = \partial_T$,

ii. The set $t=x_1=x_2$ is mapped to $\{T=0=X_1\}$, and

iii. $(1,1,1) \longmapsto (0,0,0)$.

Since $x'=0$ iff $T=0=X_1$ it follows that the quotient of $|x'|^2$ by $T^2+X_1^2$ is a smooth function which is strictly positive on a neighborhood of $x'=0$. Thus, on a small neighborhood of the origin in T,X space,

$$\partial_T V_1 V_2 z = 2\pi a \ \ln(|x'|) + C^\infty = 2\pi a \ \ln(T^2+X_1^2) + C^\infty.$$

An integration shows that

$$V_1 V_2 z = 2\pi a \int_0^T \ln(T^2+X_1^2) \ dT + f(X) + C^\infty,$$

with $f = V_1 V_2 z|_{T=0} \in \mathcal{D}'(\mathbb{R}^2_X)$. The integral is equal to

$$2X_1 \arctan(T/X_1) + T \ln(T^2+X_1^2) - 2T.$$

The second X_1 derivative of this expression tends to infinity as X_1 tends to zero. The rate of explosion despends on T so cannot be cancelled by the f(X) term. We conclude that z is not piecewise smooth at the characteristic hyperplane $X_1=0$. This completes our construction.

Readers who prefer first order systems to second order systems can easily replace the d'Alembertians by the standard first order alias,

$$\partial_t + c\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\partial_{x_1} + c\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\partial_{x_2}.$$

Readers who prefer purely second order systems are invited to use $\Box_s$, with $0<s<2^{1/2}$, in place of ∂_t in equation (13). The verification is then not as simple.

Those who prefer sinqle scalar equations of high order are encouraged to change their minds.

References

[A] S. Alinhac, Evolution d'une onde simple pour les équations non-linéaires génerales, preprint.

[Be] M. Beals, Presence and absence of weak singularities in nonlinear waves, in *Dynamical Problems in Continuum Physics, eds.* Bona,Dafermos,Erikson, and Kinderlehrer, Springer-Verlag, New York, 1987, p. 23-41.

[Bo1] J-M. Bony, Interaction des singularités pour les équations aux déerivées partielles non-linéaires, Séminaire Goulaouic-Meyer-Schwartz, 1981-82, exposé #2.

[Bo2] J-M. Bony, Propagation et interaction des singularités par les solutions des équations aux dérivées partielles non linéaires, Proc. Int. Cong. Math., Warsaw (1983), 1133-1147

[C] R. Courant, *Methods of Mathematical Physics,* vol.II, Interscience, New York, 1966.

[CL] R. Courant and P.D. Lax, The propagation of discontinuities in wave motion, Proc. Natl. Acad. Sci. USA 42(1956), p. 872-876.

[Ha] J. Hadamard, *Lecons sur la Propagation des Ondes et les*

Equations de l'Hydrodynamique, A. Hermann, Paris, 1903.

[Ho] L. Hormander, *The Analysis of Linear Partial Differential Operators*, vol. III, Springer-Verlag, New York, 1985.

[Mel] R. Melrose, Interaction of progressing waves through a nonlinear potential, Séminaire Goulaouic-Meyer-Schwartz, 1983/84, expose #12.

[M] G. Metivier, The Cauchy problem for semilinear hyperbolic systems with discontinuous data, Duke Math. J. 53(1986), p. 983-1011.

[MR] G. Metivier and J. Rauch, article in preparation.

[NP1] B. Nadir and A. Piriou, Symboles pour deux ondes conormales sans interaction, C.R.A.S. Paris, t.305, serie 1(1987).

[NP2] B. Nadir and A. Piriou, Ondes semi-linêare conormales par rapport á deux hypersurfaces transverses, to appear.

[P1] A. Piriou, Calcul symbolique nonlinêare pour une onde conormale simple, C.R.A.S. Paris t.304, serie 1, #4(1987).

[P2] A. Piriou, private communication.

[P3] A. Piriou, article to appear in Ann. Inst. Fourier.

[RR1] J. Rauch and M. Reed, Jump discontinuities of semilinear, strictly hyperbolic systems in two variables: creation and propagation, Comm. Math. Phys. 81(1981) p. 203-227.

[RR2] J. Rauch and M.Reed, Discontinuous progressing waves for semilinear systems, Comm. P.D.E. 10(1985), p. 1033-1075.

[RR3] J. Rauch and M. Reed, Striated solutions of semilinear two-speed wave equations, Indiana Math. J. 34(1985) p. 337-353.

[RR4] J. Rauch and M. Reed, Classical Conormal Solutions of Semilinear Systems, Comm. P.D.E. 13(1988) p. 1297-1335.

[RR5] J. Rauch and M.Reed, Bounded stratified and striated solutions of hyperbolic systems, in Nonlinear Partial Differential Equations and thier Applications, Séminaire Collège de France 1987, H. Brezis and J.L. Lions eds.

Diffraction Effects in Weakly Nonlinear Detonation Waves

Rodolfo R. Rosales*

Department of Mathematics, Room 2-337
Massachusetts Institute of Technology, Cambridge, MA 02139

In the limit of small heat release, large activation energy and weak nonlinearity, the propagation of detonation waves obeys a Geometrical Optics approximation. These equations develop caustic singularities, where the approximation fails. Here we present a derivation of a modified set of equations for weakly nonlinear detonation waves incorporating lateral diffraction effects. The modified set of equations does not fail at caustics.

1. Introduction

In [14] equations governing the propagation of a weakly nonlinear detonation wave in a reacting polytropic gas in the limit of large activation energy and small heat release are derived using Weakly Nonlinear Geometrical Optics asymptotics. The propagation of the detonation front $\psi(\boldsymbol{x}) = t$ is then governed by the Eikonal equation

$$(\nabla\psi)^2 = 1 \tag{1.1}$$

in appropriate nondimensional variables. When the wave moves into a uniform state and transport effects are neglected, the amplitude equations are

$$\frac{d\sigma}{dt} + \{\frac{1}{4}\frac{\gamma+1}{\gamma-1}\sigma^2 + q\lambda\}_\theta = -\frac{1}{2}(\Delta\psi)\sigma\,, \tag{1.2a}$$

$$\lambda_\theta = -\phi(\sigma, \lambda)\,, \tag{1.2b}$$

where $\sigma = \sigma(\theta, \boldsymbol{x}, t)$ controls the variation of the fluid dynamical variables across the wave (to leading order they are all proportional to $\epsilon\sigma, 0 < \epsilon << 1$, the proportionality constant being 1 for the temperature), $\lambda = \lambda(\theta, \boldsymbol{x}, t)$ is the reaction parameter to leading order (a reaction controlled

*This work was performed in part while the author was visiting the Department of Mathematics at Stanford University, Stanford, California. The author was partially supported by grants from the AFOSR, NSF and the Wade Foundation.

by a single parameter $0 \leq \lambda \leq 1$ is assumed with $\lambda = 0, 1$ in the fresh, burnt mixture respectively), γ is the usual γ of ideal gas laws, $q > 0$ is a constant (the heat release), $\theta = (1/\epsilon)(\psi - t)$ is the phase function, ϕ is the reaction rate (ignition temperature kinetics are assumed so that ϕ vanishes for, say, $\sigma \leq 0$) and d/dt denotes derivation along the characteristics of (1.1). Specifically

$$\frac{d}{dt}\boldsymbol{x} = \boldsymbol{\nabla}\psi\,,\ \frac{d}{dt}\sigma = \sigma_t + (\boldsymbol{\nabla}\sigma)\cdot(\boldsymbol{\nabla}\psi), \tag{1.3}$$

where $\psi(\boldsymbol{x}) = 0$ is the initial position of the reaction front.

We note that equation (1.1) says that the reaction front $\psi = t$ moves normal to itself at constant velocity 1. Thus a concave front will eventually focus and fold — forming arêtes and caustics. Once this happens the expansion in [14] breaks down, as it is built up on the implicit assumption of a single smooth reaction front. In fact

$$\Delta\psi = \sum_j \kappa_j, \tag{1.4}$$

where the κ_j are the principal curvatures of the front. Thus the right hand side of (1.2a), that takes care of amplification and damping of the wave by geometrical effects, becomes unbounded near caustics and arêtes.

The breakdown pointed out above is the same as that that occurs for the equations governing the propagation of weak shock fronts in compressible gases (set $q = 0$ in (1.2a) and ignore λ) and generally for Geometrical Optics approximations. In the case of weak shocks (by allowing "slow," parallel to the front, dependence of the variables in the expansion) asymptotic equations that may remain valid (see remark 1.1) near arêtes are proposed in [4]. In this paper we do the same for (1.1) and (1.2).

The equations, derived in section 3, are as follows:

Asymptotic Equations

$$\frac{d}{dt}\sigma + \{\frac{1}{4}\frac{\gamma+1}{\gamma-1}\sigma^2 + q\lambda\}_x + \frac{\gamma-1}{2}\eta_y \;=\; 0, \tag{1.5a}$$

$$(\gamma-1)\eta_x \;=\; \sigma_y, \tag{1.5b}$$

$$\lambda_x \;=\; -\phi(\sigma,\lambda), \tag{1.5c}$$

for propagation into a uniform state and assuming two space dimensions, with $\boldsymbol{x} = (X, Y)$. Here $\sigma = \sigma(x, y, \boldsymbol{x}, t)$, $\lambda = \lambda(x, y, \boldsymbol{x}, t), \gamma, q$ and ϕ have the same meaning as in (1.2), with

$$\frac{d}{dt} = \partial_t + \partial_X \tag{1.6}$$

and $\eta = \eta(x, y, \boldsymbol{x}, t)$ is related to the flow velocity component parallel to the front. The role of θ in (1.2) is taken over by $x = (1/\epsilon)(X - t)$ and $y = (1/\sqrt{\epsilon})Y$ is the new "slow" variable.

This formulation allows for curved fronts of $O(1)$ curvature in the $\boldsymbol{x}$ variables, e.g.

$$\begin{aligned} x &= f(y,\boldsymbol{x},t) \text{ or} \\ X &= t+\epsilon f(\frac{1}{\sqrt{\epsilon}}Y,\boldsymbol{x},t) \end{aligned} \tag{1.7}$$

for some function f, but no singular terms appear in the expansion as focusing occurs, nor do we have to deal with the troublesome folding and crossing of fronts of the Eikonal equation. Note that the trivial, plane wave solution $\psi = X$ of (1.1) is incorporated in the expansion.

The asymptotic equations (1.5) incorporate a certain amount of diffraction effects through the presence of y derivatives in them, but only at the linear level. Nonlinear effects are kept only in the direction normal to the front. We note that when $q = 0$, and λ is ignored, the equations (1.5) reduce to the equations in [4] — as they should.

The general idea and physical motivation behind the scalings that go into the derivation of (1.5) is that, near the points where (1.1) and (1.2) fail because the front develops a fold through (1.1), the various "branches" of the front are nearly parallel. Thus we can keep a variable playing the role of θ in the description, with a second - "slower" - variable to distinguish branches. Generally one must consider an expansion where the relevant independent variables are $\theta = (1/\epsilon)(\psi - t)$ — where ψ is a nonsingular solution of (1.1) — and a new, slower, variable $(1/\sqrt{\epsilon})\zeta(\boldsymbol{x},t)$ transversal to θ. We have chosen here the simplest case and leave consideration of the more general one for a later publication. The more general expansion may also be useful for propagation in the presence of obstacles, when "singular rays" and "shadow" boundaries appear. An in-depth exploration of these ideas in the non-reacting case can be found in [6].

Remark 1.1 The ideas sketched in the prior paragraph originate in the theory of caustics for linear equations (see [1], [10] and [12]) where they are known to work. The point is that in the linear case the resulting asymptotic equations can be solved by separation of variables and linear superposition. This is not so in the nonlinear case. While (1.5) are clearly simpler than the full set of equations governing the phenomenae of interest, they are complicated enough that no useful exact solutions are known. Furthermore, (1.5) are a canonical minimal set of equations and no further reduction by asymptotic techniques seem possible without losing the phenomenae of interest. A numerical study of them seems necessary and unavoidable and we are currently involved in that, together with A. Stuart at M.I.T. and Bath University.

Remark 1.2 The expansions leading to (1.1), (1.2) and to (1.5) remain valid for weak solutions where σ and η — but not λ — may have discontinuities. This follows because the expansions may be arranged so that no derivatives of σ and η appear in them. The proper conservation forms for weak solutions are those displayed.

The plan of this paper is as follows: In section 2 we review some of the issues concerning the failures of Geometrical Optics at caustics, arêtes, etc. and what is known from experiments. Section 3 has a derivation of the equations in (1.5).

2. Review

In this section we briefly review Geometrical Optics and some of the issues that concern the failures of the theory at arêtes, caustics, shadow boundaries, etc. No attempt at completeness is made. Our only purpose is to provide motivation and some background for the material in this paper. The section is organized in the following subsections: 2.1 General Considerations, 2.2 Geometrical Optics (linear theory), 2.3 Caustics (linear theory), 2.4 Weakly Nonlinear Geometrical Optics and 2.5 Experimental Results.

2.1 General Considerations

The propagation of sharp wave fronts — including shocks — and of high frequency progressive waves in linear and quasilinear hyperbolic p.d.e.'s is described by the asymptotic theories of:

i) Geometrical Optics in the linear limit of infinitesimal amplitude. See [10] and [16].

ii) Weakly Nonlinear Geometrical Optics for small but finite amplitudes. See [2], [7], [9] and [13].

In both cases the propagation of the wave fronts is described by a first order Hamilton-Jacobi equation (the *Eikonal* equation) for the phase function and by an associated *transport* equation for the amplitude of the waves.

Generally the Eikonal equation can lead to focusing of the wave fronts, with folds and other singularities appearing (*caustics, arêtes,* etc.). At those places the theories in (i) and (ii) above cease to be valid and, in particular, infinite amplitudes are predicted. Difficulties arise also at the *shadow boundaries* when applying (i) and (ii) to the study of wave propagation in the presence of obstacles (*Singular* rays).

In spite of the problems mentioned above, these two theories are very useful in the study of hyperbolic p.d.e. phenomenae, and a resolution of the difficulties mentioned above is very important. In the context of linear Geometrical Optics this has been done, and the behavior of waves at caustics, singular rays, foci, etc., is well understood (see [1], [10] and [12] for example). On the other hand, for Weakly Nonlinear Geometrical Optics, the problem is still wide open, although a substantial amount of work exists in the field (see [4], [6] and [8], for example).

The type of difficulties one must face in nonlinear theory are twofold:

a) First, in nonlinear situations, different wave modes interact and new modes are generated. Thus multiple wave situations are very difficult, and presently can be handled only in certain special circumstances (see [7], [9] and [13]). Clearly, at the places where wave fronts cross and fold, a multiple wave situation arises — even if the original wave was locally monocromatic. In the nonlinear case we can expect new waves to be produced; not so in the linear case — where the principle of linear superposition applies.

b) Second, the simplified, canonical, asymptotic model equations that can be derived to model the behavior of the waves near caustics, singular rays, etc. in the weakly nonlinear case are not well understood. They are simple, but not sufficiently so as to have known and useful exact solutions. (A numerical study seems indicated). It should be pointed out that their linear counterparts can be solved by separation of variables and linear superposition so that this second difficulty is not too different in nature from the first one above.

2.2 Geometrical Optics (linear theory)

Consider a linear system of totally hyperbolic p.d.e.'s (see [16]). For simplicity assume that we can write it in the form

$$\boldsymbol{u}_t + \sum A_j \boldsymbol{u}_{x_j} = 0\,, \tag{2.1}$$

where $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x}, t)$ is m-column vector valued, $\boldsymbol{x} = (x_1, \ldots, x_n)^t$ and the $A_j = A_j(\boldsymbol{x}, t)$ are $m \times m$ matrices.

Consider a locally (in space and time) monocromatic high frequency wave solution of (2.1), with wave number $\boldsymbol{k} = \boldsymbol{k}(\boldsymbol{x}, t)$ and wave frequency $\omega = \omega(\boldsymbol{x}, t)$ defined locally. By a locally monocromatic wave solution we mean a solution for which it is possible to define $\boldsymbol{k}$ and ω. Such a solution must necessarily be high frequency, for near any point $(\boldsymbol{x}, t)$ many wave fronts must be present to give a meaning to $\boldsymbol{k}$ and ω.

It is then clear that $\boldsymbol{k} = (k_1, \ldots, k_n)^t$ and ω must satisfy the *plane wave relationship*

$$\det[-\omega I + \sum k_j A_j] = 0 \tag{2.2}$$

to leading order in the frequency, with $\boldsymbol{u}$ proportional to the corresponding right eigenvector. Then if the wavefronts are given by $\varphi = \varphi(\boldsymbol{x}, t) =$ constant, we can choose φ so that

$$\omega = -\varphi_t \text{ and } \boldsymbol{k} = \boldsymbol{\nabla}\varphi\,. \tag{2.3}$$

Equations (2.2) and (2.3) are a system of equations for the phase φ of the wave: the *Eikonal* equation. The bicharacteristics of these equations are called *rays* in Geometrical Optics and correspond to a *particle view* of the phenomenae described by (2.1) with the wave propagating along these rays.

Example 2.1 In the case of the wave equation

$$u_{tt} - \text{div}(c^2 \text{grad} u) = 0, \tag{2.4}$$

with $c = c(\boldsymbol{x}) > 0$, we can take $\varphi(\boldsymbol{x}, t) = \psi(\boldsymbol{x}) - t$ and then we get the classical Eikonal equation

$$c^2(\nabla\psi)^2 = 1. \tag{2.5}$$

In this case the rays are normal to the wavefronts and these propagate along the rays at the local wave speed c.

The formal asymptotic expansion corresponding to these ideas is

$$\boldsymbol{u} = \{a\boldsymbol{r} + O(\epsilon)\}e^{i\theta}, \theta = \frac{1}{\epsilon}\varphi, \tag{2.6}$$

where $0 < \epsilon << 1$, $a = a(\boldsymbol{x}, t)$ is the amplitude and $\boldsymbol{r} = \boldsymbol{r}(\boldsymbol{x}, t)$ is the right eigenvector corresponding to (2.2), properly normalized. The Eikonal equation must then be complemented by a *transport* equation for the amplitude a. We will not give its general form here, but in the example 2.1 above it is

$$(a^2)_t + \mathrm{div}(a^2c^2\boldsymbol{k}) = 0\,, \tag{2.7}$$

so that a^2 is conserved on ray tubes.

2.3 Caustics (linear theory)

Generally the rays associated with the Eikonal equation will intersect, leading to singularities and multiple values in the solution (cusps and folds in the wave fronts). In addition, whenever obstacles are present, limiting ("singular") rays will appear separating regions of space accessible and not reached by the rays.

At all the places indicated in the prior paragraph the expansions in subsection 2.2 fail, as infinities and other singularities appear. For example at any point where a ray tube collapses, (2.7) predicts an infinite amplitude.

Of particular interest are:

Arêtes: places where a singularity appears for the first time, by focusing of an infinitesimal area of the initial wavefronts. After this first time the wave fronts cross and fold on themselves.

Caustics: these are the locations of the folding places in the wave fronts. Often, but not always, they begin at arêtes.

Foci: same as arêtes, but a finite region of the initial wave fronts is focused.

Singular rays: rays separating "illuminated" zones from "dark" zones due to the presence of obstacles and if diffraction of the waves from the points of contact of the singular rays and the object are ignored.

The resolution of all these difficulties are well understood by now (see [1], [10] and [12] for example). The multiple values given by the solution of the Eikonal equation simply mean that the wave is no longer monocromatic in those regions. Because of the linear superposition principle this does not represent a problem. Near the singular regions (caustics, etc.) the expansion must be supplemented by local — internal layer — expansions to resolve the infinities. Thus (linear) Geometrical Optics is valid everywhere — including multiple valued regions — provided we take care appropriately of the inner layers appearing near the regions of trouble. For example, near a caustic, the two branches of the wave front — incident and reflected — are nearly parallel. Thus an expansion very much like the one in subsection 2.2, but incorporating deviations from the plane mode form via a weak dependence (of $O(1/\sqrt{\epsilon})$ rather than $O(1/\epsilon)$ as in θ) on the transverse direction solves the problem.

The main point is that the asymptotic equations valid near the inner layers can be solved by separation of variables and superposition. (For example, near caustics the switch from waves to no waves occurs — in general — via Airy functions). This is an advantage that is lost in the nonlinear case, even though one can still derive equations that should be valid near (at least some of) the trouble spots.

2.4 Weakly Nonlinear Geometrical Optics

We can still deal with quasilinear systems of hyperbolic p.d.e.'s — such as (2.1) with the A_j functions of $\boldsymbol{u}$ also — provided we limit the amplitude to be small, of $O(\epsilon)$ with ϵ as in (2.6).

A small modification of the ideas in subsection 2.2 can be used. The main difference is that the form of the wave is not known in advance and nonlinear deformation of the wave forms must be allowed. Exponentials can no longer be used and (2.6) is replaced by

$$\boldsymbol{u} \sim \boldsymbol{u}_0(\boldsymbol{x},t) + \epsilon\sigma(x,t,\theta)\boldsymbol{r}(\boldsymbol{x},t) + \dots\,, \tag{2.8}$$

where $\boldsymbol{u}_0$ is a known solution of the equations (e.g. a constant). The Eikonal equation still applies for $\varphi = \epsilon\theta$, with the A_j's evaluated at $\boldsymbol{u} = \boldsymbol{u}_0$, and $\boldsymbol{r}$ has the same meaning as before. The transport equation is now nonlinear and has the form

$$\frac{d}{dt}\sigma + (\frac{1}{2}p\sigma^2)_\theta = -d\sigma\,, \tag{2.9}$$

where $\frac{d}{dt}$ indicates derivation along the rays, p is a nonlinearity coefficient and d is a focusing/defocusing coefficient, related to the curvature of the wave fronts.

Clearly this expansion will suffer from the same difficulties spelled out for the linear case in subsection 2.3. In this context this is an open problem and a satisfactory resolution of the situation is not known, as explained in subsection 2.1 and again at the end of subsection 2.3.

Remark 2.1 The expansion remains valid when waves break and shocks form (assume (2.1) has an associated conservation form) and (2.9) is the proper conservation form. Thus it can be used to study the propagation of weak shocks.

2.5 Experimental Results

Sturtevant and Kulkarny undertook a careful experimental investigation of the focusing of weak shocks in non-reacting gases. Their results, reported in [15], have enormous relevancy to the subject matter of subsection 2.4: failure of Weakly Nonlinear Geometrical Optics near arêtes, perfect foci, caustics, etc. No similar experiments have been carried out (to our knowledge) for weak detonation waves.

Sturtevant and Kulkarny produced converging weak shocks of controllable wave front shape by reflecting initially plane front shocks from concave end walls in a large shock tube (note that this approach would not work for detonation waves). We briefly summarize some of their results next. The interested reader should consult [15] for a full account.

It is clear that as a shock front focuses, its strength increases and, consequently, so does its speed. Thus the parts of the front where more focusing occurs will move at speeds farther (larger) and farther away from the (linear) acoustical speed predicted by the Eikonal equation. Clearly this (nonlinear) effect works to *prevent* focusing of the front itself — as it occurs in linear theory — not just merely to stop infinities.

It is found in [15] that beyond a certain critical shock strength (e.g. of about Mach number $M = 1.2$ for the incident plane shock in the case of a perfect focus in section 3.1 there), the effect in the paragraph above seems to dominate. As the front focuses, the focusing parts speed up and become nearly plane with true focusing entirely avoided. At the ends of this region the shock front develops corners, associated with the appearance of triple points there. The Mach stem and vortex line (the *new waves* being generated by nonlinearity, see (a) in subsection 2.1) associated with the triple points trail the main shock front as the triple points fly apart from each other (see Fig. 18d in [15]).

As the strength of the shock decreases and approaches criticality the paths of the triple points — which for large enough strength diverge — become more complicated. They first diverge, then stop, start approaching each other, stop again, and finally diverge again (see Fig. 18c in [15]). Below the critical strength, after first diverging and stopping, the triple points approach each other and collide (see Fig. 18b in [15]). After the collision the resulting wave fronts take on an appearance very similar to the "folded on itself" wave front predicted by the Eikonal equation. However, at the places where the front would merely fold on itself according to the Eikonal equation (caustics) *new waves* (generated by nonlinear interactions, no doubt) are observed and triple point-like structures seem to occur there. For weaker and weaker shocks the "new waves" become fainter and fainter, the interval between formation and collision of the triple points goes to zero and generally the whole observed pattern resembles more and more that predicted by linear theory.

The behavior for strong and moderately strong (well above critical in fact) converging shocks seems to agree reasonably well with that predicted by Whitham's Geometrical Shock Dynamics theory [16], provided one equates the "shock-shocks" of the theory with the triple points. Unfortunately the theoretical foundations of this theory are not well understood. Furthermore, this theory cannot predict the transition behaviors described above.

In an effort to explain the observed behavior, up to and possibly including the critical shock strength transition described above, Cramer and Seebass proposed a model in [4]. Their derivation is based on the idea that (at least for arêtes and provided the original shock front is not too convoluted) even after Weakly Nonlinear Geometrical Optics fails, the wave fronts present are all nearly parallel and thus a nearly one dimensional approximation may apply.

The model proposed in [4], which is a special case of our model derived in section 3 when combustion is ignored, has as its main component the equations

$$\sigma_t + \left\{\frac{1}{4}\frac{\gamma+1}{\gamma-1}\sigma^2\right\}_x + \frac{\gamma-1}{2}\eta_y = 0\,, \quad (\gamma-1)\eta_x = \sigma_y\,. \tag{2.10}$$

Equivalently, in the smooth part of the flow, we have

$$\sigma_{tx} + \left\{\frac{1}{4}\frac{\gamma+1}{\gamma-1}\sigma^2\right\}_{xx} + \frac{1}{2}\sigma_{yy} = 0, \tag{2.11}$$

where we have eliminated η by cross differentiation in (2.10). We note that this is in fact the same as the Time Dependent Small Disturbance Transonic Flow equation (see [3] and [11]).

The problem with the model in [4], as well as with our own model in this paper, is that very little is known concerning the behavior of equations such as (2.10) in what regards the problems considered here and that motivate their derivation. We hope that the numerical computations we are currently involved in (with A. Stuart) will shed some light on this.

While (2.10) has been proposed for arêtes, nonlinear Tricomi equations of the form

$$\sigma_t + (y\sigma + \sigma^2)_x + \eta_y = 0\,, \quad \eta_x = \sigma_y \tag{2.12}$$

have been proposed near caustics and singular rays [6]. The remarks in the prior paragraph are also valid here.

3. Derivation of the Asymptotic Equations

In this section we derive the equations in (1.5) for a weakly nonlinear detonation wave incorporating diffraction effects. The section is organized in the following subsections: 3.1 The Basic Nondimensional Equations of Reacting Gas Flow, 3.2 Derivation of the Asymptotic Equations, and 3.3 The Asymptotic Equations.

3.1 The Basic Nondimensional Equations of Reacting Gas Flow

If we neglect transport effects, assume a single, $0 \leq \lambda \leq 1$, progress variable for the reaction (with $\lambda = 0$ in the fresh mixture and $\lambda = 1$ in the burnt products), assume a polytropic gas with the same γ throughout the reaction and assume a constant small heat release (i.e. $\frac{\partial}{\partial\lambda}e = \text{const.}$, where e is the internal energy density) then the equations for a reacting gas can be written in the nondimensional form (see [5] and [14])

$$\dot{\tau} - \epsilon\tau\text{div}\boldsymbol{u} \;=\; 0\,, \tag{3.1a}$$

$$\gamma\epsilon\dot{\boldsymbol{u}} + \text{grad}T - \rho T\text{grad}\tau \;=\; 0\,, \tag{3.1b}$$

$$\dot{T} + \epsilon(\gamma-1)T\text{div}\boldsymbol{u} \;=\; 2\epsilon^2\frac{\gamma}{\gamma-1}q\dot{\lambda}\,, \tag{3.1c}$$

$$\epsilon\dot{\lambda} \;=\; w\,, \tag{3.1d}$$

where $0 < \epsilon << 1$, τ is the specific volume and $\rho = 1/\tau$ is the density, $\boldsymbol{u}$ is the flow speed, T is the temperature, $q > 0$ is a constant (the heat release), w is the reaction rate and a dot denotes the material derivative

$$\dot{p} = p_t + \epsilon\boldsymbol{u}\cdot\text{grad}p\,. \tag{3.2}$$

Introducing now a large activation energy hypothesis, we can write

$$w = W\left(\tau, \lambda, T, \frac{T-1}{\epsilon}\right) \tag{3.3}$$

for the reaction rate. An ignition temperature assumption will also be necessary, so that w vanishes for $T < 1$.

Actually, except for the form of the reaction rate (3.3), our hypothesis here are more restrictive than those in [14]. We do this for simplicity, as there is no difficulty in using the more general equations in [14].

Finally, let us define

$$\phi(\sigma,\lambda) = W(1,\lambda,1,\sigma). \tag{3.4}$$

3.2 Derivation of the Asymptotic Equations

Consider the case of two space dimensions (the generalization to three space dimensions is straightforward, see remark 3.1 at the end of this section 3), with $\boldsymbol{x} = (X,Y)$. Now let $x = (1/\epsilon)(X-t)$, $y = (1/\sqrt{\epsilon})Y$ and consider an expansion of the form

$$T = 1 + \epsilon\tau_1 + \epsilon^{3/2}\tau_2 + \epsilon^2\tau_3 + \dots, \tag{3.5a}$$

$$\boldsymbol{u} = \boldsymbol{u}_1 + \sqrt{\epsilon}\boldsymbol{u}_2 + \epsilon\boldsymbol{u}_3 + \dots, \tag{3.5b}$$

$$T = 1 + \epsilon T_1 + \epsilon^{3/2}T_2 + \epsilon^2 T_3 + \dots, \tag{3.5c}$$

$$\lambda = \lambda_0 + \sqrt{\epsilon}\lambda_1 + \dots, \tag{3.5d}$$

where $\boldsymbol{u}_j = (U_j, V_j)$ and the variables are functions of x, y, X, Y and t. We assume that, as $x \to \infty$, the variables reach a limit and that derivation commutes with the limit $x \to \infty$. The expansions for τ and T start with 1 as a result of the nondimensionalization.

Substituting this expansion first into the fluid dynamics part of (3.1) (i.e., the first three equations) we find, at leading order in ϵ, O(1), the equations

$$\tau_{1x} + U_{1x} = 0, \tag{3.6a}$$

$$\gamma U_{1x} - T_{1x} + \tau_{1x} = 0, \tag{3.6b}$$

$$\gamma V_{1x} = 0, \tag{3.6c}$$

$$T_{1x} - (\gamma-1)U_{1x} = 0. \tag{3.6d}$$

Clearly then

$$\tau_1 = \frac{-1}{(\gamma-1)}\sigma + \overline{\tau}_1\,, \quad U_1 = \frac{1}{(\gamma-1)}\sigma + \overline{U}_1\,, \quad V_1 = \overline{V}_1\,, \quad T_1 = \sigma, \tag{3.7}$$

where $\sigma = \sigma(x,y,X,Y,t)$ is arbitrary and the barred variables are functions of y, X, Y and t — but not of x. We assume that (as $x \to \infty$) $\sigma \to \overline{\sigma}$, with the temperature ahead of the wave ($T \sim 1 + \epsilon\overline{\sigma} + \dots$) below ignition.

At the next order, $O(\sqrt{\epsilon})$, in the fluid dynamics equations we have

$$\tau_{2x} + U_{2x} = -V_{1y}, \tag{3.8a}$$

$$\gamma U_{2x} - T_{2x} + \tau_{2x} = 0, \tag{3.8b}$$

$$\gamma V_{2x} = T_{1y} - \tau_{1y}, \tag{3.8c}$$

$$T_{2x} - (\gamma - 1)U_{2x} = (\gamma - 1)V_{1y}. \tag{3.8d}$$

Thus we must have $V_1 = \overline{V}_1 = \overline{V}_1(X, Y, t)$ not a function of y and

$$\tau_2 = \frac{-1}{(\gamma-1)}T_2 + \overline{\tau}_2, \quad U_2 = \frac{1}{(\gamma-1)}T_2 + \overline{U}_2, \quad V_2 = \eta, \tag{3.9}$$

where $\eta = \eta(x, y, X, Y, t)$ satisfies

$$(\gamma - 1)\eta_x = \sigma_y - \frac{\gamma-1}{\gamma}\overline{\tau}_{1y}, \tag{3.10}$$

and we assume both T_2 and η have limits $\overline{T}_2$ and $\overline{\eta}$ as $x \to \infty$.

Finally, at $O(\epsilon)$ in the fluids, we find

$$\tau_{3x} + U_{3x} = \tau_{1t} + U_1\tau_{1x} - \tau_1 U_{1x} - U_{1X} - V_{2y} - V_{1Y}, \tag{3.11a}$$

$$\gamma U_{3x} - T_{3x} + \tau_{3x} = \gamma U_{1t} + \gamma U_1 U_{1x} + T_{1X} - \tau_{1X} - (T_1 - \tau_1)\tau_{1x}, \tag{3.11b}$$

$$\gamma V_{3x} = \gamma V_{1t} + \gamma U_1 V_{1x} + T_{2y} + T_{1Y} - \tau_{2y} - \tau_{1Y}, \tag{3.11c}$$

$$\begin{aligned} T_{3x} - (\gamma-1)U_{3x} &= T_{1t} + U_1 T_{1x} + (\gamma-1)T_1 U_{1x} + (\gamma-1)U_{1X} \\ &+ (\gamma-1)V_{2y} + (\gamma-1)V_{1Y} + 2\frac{\gamma}{\gamma-1}q\lambda_{0x}. \end{aligned} \tag{3.11d}$$

The solvability condition for this system of equations is

$$\begin{aligned} \sigma_t + \sigma_X + \overline{U}_1\sigma_x &+ \frac{1}{2}\frac{\gamma+1}{\gamma-1}\sigma\sigma_x + \frac{1}{2}(\gamma-1)\eta_y + q\lambda_{0x} = \\ &- \gamma(\overline{U}_{1t} + \overline{U}_{1X}) + (\overline{\tau}_{1t} + \overline{\tau}_{1X}) - \gamma\overline{V}_{1Y}. \end{aligned} \tag{3.12}$$

Substituting now (3.5) into (3.1d) and using (3.3) and (3.4) we find

$$\lambda_{0x} = -\phi(\sigma, \lambda_0). \tag{3.13}$$

If the wave is moving into an equilibrium rest state, then we can take the barred variables constant and we have, from (3.7), (3.9), (3.10), (3.12) and (3.13):

3.3 The Asymptotic Equations (wave moving into a rest state)

$$\frac{d}{dt}\sigma + \left\{ \overline{U}_1\sigma + \frac{1}{4}\frac{\gamma+1}{\gamma-1}\sigma^2 + q\lambda_0 \right\}_x + \frac{\gamma-1}{2}\eta_y = 0\,, \tag{3.14a}$$

$$(\gamma-1)\eta_x = \sigma_y\,, \tag{3.14b}$$

$$\lambda_{0x} = -\phi(\sigma,\lambda_0)\,, \tag{3.14c}$$

where

$$\tau = 1 + \epsilon\left(\overline{\tau}_1 - \frac{1}{\gamma-1}\sigma\right) + O(\epsilon^{3/2})\,, \tag{3.15a}$$

$$U = \overline{U}_1 + \frac{1}{\gamma-1}\sigma + O(\sqrt{\epsilon})\,,\quad V = \overline{V}_1 + \sqrt{\epsilon}\eta + O(\epsilon)\,, \tag{3.15b}$$

$$T = 1 + \epsilon\sigma + O(\epsilon^{3/2})\,, \tag{3.15c}$$

$$\lambda = \lambda_0 + O(\sqrt{\epsilon}) \tag{3.15d}$$

and $\overline{U}_1$, $\overline{\tau_1}$ and $\overline{V}_1$ are constants. It is clear that, except for a trivial Galileian transformation, (1.5) and (3.14) are the same.

Remark 3.1 The extension to three space dimensions is straightforward. In this case η and y become 2-vectors, with η_y replaced by $\operatorname{div}\eta$ and σ_y replaced by $\operatorname{grad}\sigma$ (div and grad relative to the y variables).

References

[1] Buchal, R. N. and Keller, J. B., "Boundary Layer Problems in Diffraction Theory," *Comm. Pure Appl. Math.*, vol. **13**, pp. 85-114, 1960.

[2] Choquet-Bruhat, Y., "Ondes Asymptotiques et Approchées pour des systèmes d'Équations aux Dérivées Partielles non Linéaires," *J. Math. Pures et Appl.*, vol. **48**, pp. 117-158, 1969.

[3] Cole, J. D., "Modern Developments in Transonic Flow," *SIAM J. Appl. Math.*, vol. **29**, pp. 763-787, 1975.

[4] Cramer, M. S. and Seebass, A. R., "Focusing of Weak Shock Waves at an Arête," *J. Fluid Mech.*, vol. **88**, pp. 209-222, 1978.

[5] Fickett, W. and Davis, W. C., *Detonation*, Univ. of California Press, Berkeley, 1979.

[6] Hunter, J. K., "Transverse Diffraction of Nonlinear Waves and Singular Rays," *SIAM J. Appl. Math.,* vol. **48**, pp. 1-37, 1988.

[7] Hunter, J. K. and Keller, J. B., "Weakly Nonlinear, High-frequency Waves," *Comm. Pure Appl. Math.,* vol. **36**, pp. 547-569, 1983.

[8] Hunter, J. and Keller, J. B., "Caustics of Nonlinear Waves," *Wave Motion,* vol. **9**, pp. 429-443, 1987.

[9] Hunter, J. K., Majda, A. and Rosales, R., "Resonantly Interacting Weakly Nonlinear Hyperbolic Waves. II. Several Space Variables," *St. Appl. Math.,* vol. **75**, pp. 187-226, 1986.

[10] Keller, J. B., "Rays, Waves and Asymptotics," *Bull. Am. Math. Soc.,* vol. **84**, pp. 727-750, 1978.

[11] Kevorkian, J. and Cole, J. D., *Perturbation Methods in Applied Mathematics,* Springer-Verlag, New York, 1980.

[12] Ludwig, D., "Uniform Asymptotic Expansions at a Caustic," *Comm. Pure Appl. Math.,* vol. **19**, pp. 215-250, 1966.

[13] Majda, A. and Rosales, R., "Resonantly Interacting Weakly Nonlinear Hyperbolic Waves. I. A Single Space Variable," *St. Appl. Math.,* vol. **71**, pp. 149-179, 1984.

[14] Rosales, R. R. and Majda, A., "Weakly Nonlinear Detonation Waves," *SIAM J. Appl. Math.,* vol. **43**, pp. 1086-1118, 1983.

[15] Sturtevant, B. and Kulkarny, V. A., "The Focusing of Weak Shock Waves," *J. Fluid Mech.,* vol. **73**, pp. 651-671, 1976.

[16] Whitham, G. B., *Linear and Nonlinear Waves,* John Wiley and Sons, New York, 1974.

LIST OF PARTICIPANTS

Luc ARNAUD
Centre d'Etudes de Gramat (France)

Alberto AROSIO
Universita di Parma (Italie)

Davide ASCOLI
Universita di Torino (Italie)

Luc BARBET
Université de Poitiers (France)

Agnès BACHELOT
Université de Bordeaux I (France)

Alain BACHELOT
Université de Bordeaux I (France)

Saïd BENACHOUR
Université d'Alger (Algérie)

Matania BEN-ARTZI
TECHNION, Haïfa (Israël)

Claude BARDOS
ENS Ulm (France)

Michèle BERNARD
CEA-CEN Cadarache Aix-en-Provence (France)

Enrico BERNARDI
Université de Bologne (Italie)

Max BEZARD
Ecole Polytechnique, Palaiseau (France)

Pierre BONNEMASON
CEA Limeil (France)

Jean-Michel BONY
Ecole Polytechnique (France)

Antoine BOURGEADE
CEA Limeil (France)

Jean-Jacques BOUYER
Société d'Etudes AERO/Paris (France)

Antonio BOVE
Université de Bologne (Italie)

M. BUGNON
Aérospatiale St Médard en Jalles (France)

Claude CARASSO
Université de St Etienne (France)

Vincent CASELLES-COSTA
Université de Besançon (France)

Pierre CHARRIER
Université de Bordeaux I (France)

Jean-Yves CHEMIN
Ecole Polytechnique (France)

Jean-François COLOMBEAU
ENS Lyon (France)

Peter CONSTANTIN
University of Chicago (USA)

Olivier COULAUD
Université de Nancy I (France)

Frédéric DABBENE
CISI Ingénierie, Mérignac (France)

Jean-Marc DELORT
Université de Rennes I (France)

Patrick DE LUCA
Centre d'Etudes de Gramat (France)

Marcel DOSSA
Université de Yaoundé (Caméroun)

Jim DOUGLAS Jr.
Purdue University (USA)

Denise DRIOLLET
CISI Ingénierie, Mérignac (France)

Bruno DUBROCA
CEA/CESTA Le Barp (France)

Pierre FABRIE
Université de Bordeaux I (France)

Mr FANGET
Centre d'Etudes de Gramat (France)

Sonia FEZOUI
INRIA-Valbonne (France)

Alain FORESTIER
CEA-CEN , Saclay (France)

Heinrich FREISTUHLER
Institut für Mathematik
AACHEN (RFA)

Jean-Philippe GAILLARD
CISI Cadarache, Aix-en-Provence (France)

Gérard GALLICE
CEA/CESTA Le Barp (France)

Catherine GAUDY
CEA-CEN , Saclay (France)

Jean GAY
CEA/CESTA Le Barp (France)

Patrick GERARD
ENS, rue d'Ulm (France)

Hervé GILQUIN
ENS de Lyon (France)

Paul GODIN
Université de Bruxelles (Belgique)

Marc GRANDOTTO
CEA-CEN Cadarache Aix-en-Provence (France)

Olivier GUES
Université de Rennes I (France)

Laurence HALPERN
Ecole Polytechnique (France)

Kamel HAMDACHE
CNRS-ENSTA/GHN, Palaiseau (France)

Bernard HANOUZET
Université de Bordeaux I (France)

Amiram HARTEN
University of Tel-Aviv (Israël)

Yves HAUGAZEAU
Université de Bordeaux I (France)

Thierry HOCQUELET
Centre d'Etudes de Gramat (France)

Fritz JOHN
Courant Institute, New-York (USA)

Jean-Luc JOLY
Université de Bordeaux I (France)

Hervé JOURDREN
CEA Limeil (France)

Shuichi KAWASHIMA
Université de Paris VI (France)

Barbara KEYFITZ
University of Houston (USA)

Sergiu KLAINERMAN
Princeton University (USA)

Rupert KLEIN
Institüt für Allg. Mechanik
AACHEN (RFA)

Witold KOSINSKI
Polish Academy of Sciences, Varsovie (Pologne)

Patrick LABORDE
Université de Bordeaux I (France)

Chantal LACOMBLEZ
Université de Bordeaux II (France)

Mr LADONNE
Centre d'Etudes de Gramat (France)

André LAFON
ONERA-CERT, Toulouse (France)

Frédéric LAFON
Université de Bordeaux I (France)

Pierre LALLEMAND
ENS, rue d'Ulm (France)

Michel LANGLAIS
Université de Bordeaux II (France)

Bernard LARROUTUROU
INRIA/SOPHIA-ANTIPOLIS (France)

Gérard LASSALLE-BALIER
CNES/TOULOUSE (France)

Peter LAX
Courant Institute, New-York (USA)

Philippe LE FLOCH
Ecole Polytechnique (France)

Alain LERAT
ENSAM, Paris (France)

Alain LE ROUX
Université de Bordeaux I (France)

Marie-Noëlle LE ROUX
Université de Bordeaux I (France)

Bernard LEROY
CEA-CESTA/LE BARP (France)

Hans LINDBLAD
Lunds Universitet (Suède)

Taï-Ping LIU
University of Maryland (USA)

Bradley LUCIER
University of Maryland (USA)

Andrew MAJDA
Princeton University (USA)

Tetu MAKINO
University of Osaka Sangyo Daigaku (Japon)

Guy METIVIER
Université de Rennes I (France)

Paul MOREL
Université de Bordeaux I (France)

Jean-Pierre MORREEUW
CEA Limeil (France)

M.K.V. MURTHY
Université de Pise (Italie)

Gawtum NAMAH
Université de Bordeaux I (France)

Roberto NATALINI
Universita di Roma (Italie)

Anne NOURI
Université de Nice (France)

Ahmed NOUSSAIR
Université de Bordeaux I (France)

Frédéric OELHOFFEN
CISI Ingénierie, Mérignac (France)

Jean OVADIA
CEA Limeil (France)

Harmut PECHER
Bergische Universität-GH, Wuppertal (RFA)

Bernard PERROT
Université de Bordeaux I (France)

Alain PIRIOU
Université de Nice (France)

Mr POIREE
DRET (France)

Yue Hong QIAN
Ecole Normale Supérieure (France)

Reinhard RACKE
Universität Bonn (RFA)

Jeffrey RAUCH
University of Michigan (USA)

Dominique RIBEREAU
Université de Bordeaux I (France)

Rodolfo ROSALES
MIT, University Stanford (USA)

Philippe ROULPH
CISI Ingénierie, Mérignac (France)

Michel ROUZE
Centre Spatial/Toulouse (France)

Muriel SESQUES
Université de Bordeaux I (France)

Stefan SCHLECHTRIEM
Lehr und Forschungsgebiet Mechanik, Aachen (RFA)

Steve SCHOCHET
University of Tel-Aviv (Israël)

Jalal SHATAH
Courant Institute, New-York (USA)

Evelyne SIBE
CEA-CESTA, Le Barp (France)

Anders SZEPESSY
Chalmers University of Technology, Göteborg (Suède)

Mr TADIE
Sussex University (ENGLAND)

Jean-Marc TALBOT
MERLIN-GERIN, Grenoble (France)

Li TA-TSIEN
Fudan University, Shangaï (Chine)

Brigitte TESSIERAS
CISI, Saclay (France)

Monique TOUGERON
Université de Rennes I (France)

Klaus-Dieter WERNER
Institute of Geometry and Practical Mathematics
AACHEN (RFA)

Wang XIANG
Institut für Geometrie und Praktische Mathematik
AACHEN (RFA)

LIST OF TALKS

F. JOHN (Courant Institute, New-York, USA)
Solutions of quasi linear wave equations with small initial data. The third phase.

A. BACHELOT (Univ. de Bordeaux I, FRANCE)
Solutions globales des systèmes de Dirac-Klein-Gordon.

J. SHATAH (Courant Institute, New-York, USA)
Harmonic maps in Minkowski space.

S. KLAINERMAN (Princeton University, USA)
On the stability of the Minkowski metric in general relativity.

P. LALLEMAND (E.N.S. Paris, FRANCE)
Lattice gases for flow simulations.

J. DOUGLAS Jr. (Purdue University, West Lafayette, USA)
Waves in two-phase generalizations of Biot media.

B. LUCIER (Purdue University, West Lafayette, USA)
High order regularity for discontinuous solutions of hyperbolic conservation laws.

M. BEN-ARTZI (Technion, Haïfa, ISRAEL)
Numerical calculations of reactive flows.

A. BOURGEADE, (CEA/Limeil-Valenton,FRANCE)
Algorithms with Riemann solver for the computation of 2D multifluid flows.

T.P. LIU (University of Maryland, College Park, USA)
Shock waves for compressible Euler and Navier Stokes equations.

P.D. LAX (Courant Institute, New-York, USA)
Systems of hyperbolic conservation laws in more than one space variable.

A.J. MAJDA (Princeton University, USA)
The non linear development instabilities in supersonic vortex sheets.

B.L. KEYFITZ (University of Houston, USA)
A viscosity approximation to a system of conservation laws with no classical Riemann solution.

A. LERAT (ENSAM, Paris, FRANCE)
Implicit difference schemes for hyperbolic systems of conservation laws.

A.Y. LE ROUX (Univ. de Bordeaux I, FRANCE)
Approximation to non linear convection diffusion problems

J.M. BONY (Ecole Polytechnique, Paris, FRANCE)
Analyse microlocale et singularités non linéaires

J. RAUCH (University of Michigan, Ann Arbor, USA)
Progressing semilinear waves.

G. METIVIER (IRMAR, Rennes, FRANCE)
Singularités fortes pour les solutions de systèmes de lois de conservation multidimensionnels.

B. LARROUTUROU (INRIA, Sophia-Antipolis, FRANCE)
On the equations of multi-component perfect or real gas flow.

R. ROSALES (Stanford University, USA)
Diffraction effects in weakly nonlinear detonation waves.

A. HARTEN (Tel-Aviv University, ISRAEL)
E N O schemes with subcell resolution.