

# INRIA

UNITE DE RECHERCHE  
INRIA-SOPHIA ANTIPOLIS

Institut National  
de Recherche  
en Informatique  
et en Automatique

Domaine de Voluceau  
Rocquencourt  
BP 105  
78153 Le Chesnay Cedex  
France  
Tel (1) 39 63 55 11

## Rapports de Recherche

N° 1080

*Programme 1*  
*Programmation, Calcul Symbolique*  
*et Intelligence Artificielle*

### HOW TO PRESERVE THE MASS FRACTIONS POSITIVITY WHEN COMPUTING COMPRESSIBLE MULTI-COMPONENT FLOWS

**Bernard LARROUTUROU**

**Août 1989**



# HOW TO PRESERVE THE MASS FRACTIONS POSITIVITY WHEN COMPUTING COMPRESSIBLE MULTI-COMPONENT FLOWS

B. LARROUTUROU

INRIA, SOPHIA-ANTIPOLIS, 06560 VALBONNE, FRANCE

**ABSTRACT:** We are interested in the numerical investigation of the compressible flow of a gaseous mixture. We consider an hyperbolic system including the Euler equations for the mixture and a mass conservation equation for each species. We propose a new approximation scheme for the convective term of the species equations. This approximation, whose construction relies on some properties of the exact solution of the Riemann problem for the multi-component system, uses an upwind Godunov-type scheme for the Euler equations. Its main interest lies in the fact that it preserves the positivity and monotonicity of the mass fractions of all species.

## METHODES NUMERIQUES PRESERVANT LA POSITIVITE DES FRACTIONS MASSIQUES POUR LE CALCUL D'ECOULEMENTS COMPRESSIBLES MULTI-ESPECES

**RESUME:** Nous nous intéressons à l'étude numérique de l'écoulement d'un mélange compressible de plusieurs espèces gazeuses. Plus précisément, nous considérons un système hyperbolique d'équations non linéaires comprenant les équations d'Euler pour le mélange et une équation de conservation de la masse pour chaque espèce gazeuse. Nous proposons pour ce système un nouveau schéma d'approximation des équations des espèces. Cette approximation, qui repose sur l'utilisation d'un schéma décentré de type Godunov pour les équations d'Euler, est construite à partir de certaines propriétés de la solution exacte de la solution du problème de Riemann pour le système multi-espèces. Elle offre l'intérêt de préserver la positivité et la monotonie des fractions massiques de toutes les espèces.

## 1. INTRODUCTION

The numerical simulation of flows of gaseous mixtures, and in particular of chemically reactive flows, has received an increasing attention in the last years and is still the subject of numerous investigations. For some of these flows, such as transonic reactive flows (see e.g. [6]), hypersonic flows (see e.g. [4], [8]) or detonations (see e.g. [17]), the hyperbolic aspects of the phenomenon play a major role. This is the origin of our interest in the following system of the “multi-component Euler equations”, which contains the classical Euler equations written for the mixture and an additional “species equation”:

$$\begin{cases} \rho_t + (\rho u)_x = 0 , \\ (\rho u)_t + (\rho u^2 + p)_x = 0 , \\ E_t + [u(E + p)]_x = 0 , \\ (\rho Y)_t + (\rho u Y)_x = 0 , \end{cases} \quad (1.1)$$

(see the definition of the notations below).

We are interested in the numerical solution of (1.1) using upwind difference schemes. Essentially two different strategies have been used in previous studies for this problem. Either one uses for the first three equations in (1.1) one of the numerous available schemes aimed at solving the Euler equations and one solves separately the species equation; 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 “Euler schemes”. Some details about these two existing approaches are presented in Section 2 below.

Our aim is to present a third alternative approach for the solution of (1.1). This new approach is based on some properties of the exact solution of the Riemann problem for (1.1), which are derived in Section 3. Besides its major computational simplicity, the proposed method has over the two existing approaches the advantage of preserving the positivity (and if the case may be the monotonicity) of the mass fractions of all species.

Let us now describe more precisely the setting of our work. We are interested in the description of the multi-dimensional flow of a compressible gaseous mixture of  $N$  species. But, for the sake of simplicity, we will consider in a first step the one-dimensional flow of a mixture of two gaseous species  $\mathcal{S}_1$  and  $\mathcal{S}_2$ . In the absence of diffusive and reactive phenomena, this flow is described by system (1.1), where  $\rho$  is the density of the mixture,  $u$  is its velocity,  $p$  its pressure,  $E$  the total energy per unit volume of the mixture, and

$Y$  is the mass fraction of species  $\mathcal{S}_1$ . If one assumes that both species behave as perfect gases, then the pressure in (1.1) is given by the relation (see [19]):

$$p = (\gamma - 1) \left( E - \frac{1}{2} \rho u^2 \right), \quad (1.2)$$

where  $\gamma$  is the ratio of the specific heats *of the mixture*, given by:

$$\gamma = \frac{Y C_{v1} \gamma_1 + (1 - Y) C_{v2} \gamma_2}{Y C_{v1} + (1 - Y) C_{v2}}, \quad (1.3)$$

(the subscripts 1 and 2 in this expression refer to species  $\mathcal{S}_1$  and  $\mathcal{S}_2$  respectively). As already said the first three equations in (1.1) form together with (1.2) the classical Euler equations, written here for the mixture of both species. Note however that, if  $\gamma_1 \neq \gamma_2$ , the coefficient  $\gamma$  in (1.2) actually depends on  $Y$ , whereas it is constant in the classical (i.e. single-component) Euler equations. The fourth equation in (1.1) will be referred to as the species equation.

**Remark 1:** Using the first and fourth equation in (1.1), one can rewrite the species equation in non conservative form as  $Y_t + uY_x = 0$ , which shows that the variable  $Y$  is purely convected by the flow. The latter non conservative equation is the basis of the donor-cell approximation (2.1) below. •

**Remark 2:** In fact the results presented in this paper apply to more general situations. We could equally well consider a mixture of  $N$  perfect gases: the last equation in (1.1) would then be replaced by  $N - 1$  equations  $(\rho Y_k)_t + (\rho u Y_k)_x = 0$  for  $1 \leq k \leq N - 1$  (see Remarks 10, 13 and 17 below). We could also consider the case where the last equation in (1.1) describes the convection of different quantities which may or may not affect the value of  $\gamma$  (the variable  $Y$  could for instance represent a passive scalar, or the variables  $k$  and  $\epsilon$  in the classical  $k - \epsilon$  turbulence model [23]...). Also, the numerical methods presented in this paper can be extended to compute the flow of a *real gas* mixture (see [8] for more details in this direction). •

Let us now introduce the notations:

$$W = \begin{pmatrix} \rho \\ \rho u \\ E \\ \rho Y \end{pmatrix} = \begin{pmatrix} W^1 \\ W^2 \\ W^3 \\ W^4 \end{pmatrix}, \quad F = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ u(E + p) \\ \rho u Y \end{pmatrix} = \begin{pmatrix} F^1 \\ F^2 \\ F^3 \\ F^4 \end{pmatrix}. \quad (1.4)$$

For the numerical solution of (1.1) (which can be written as  $W_t + F_x = 0$  in vector form), we will first consider explicit conservative schemes of the form:

$$\frac{W_i^{n+1} - W_i^n}{\Delta t} + \frac{\phi_{i+1/2} - \phi_{i-1/2}}{\Delta x} = 0, \quad (1.5)$$

where the numerical flux  $\phi_{i+1/2}$  is evaluated using a “numerical flux function”  $\Phi$  which defines the scheme under consideration:

$$\phi_{i+1/2} = \Phi(W_i^n, W_{i+1}^n). \quad (1.6)$$

The notations in (1.5) are classical:  $\Delta t$  and  $\Delta x$  represent the discretization steps in time and space respectively, and  $i$  and  $n$  are the spatial and temporal subscripts; we write  $\phi_{i+1/2}$  instead of  $\phi_{i+1/2}^n$  for the sake of simplicity.

The remaining of the paper is organized as follows. We recall some facts about the two existing numerical approaches in Section 2. The new positivity-preserving approximation of the species equation is then presented in Section 3, and some extensions are discussed in Section 4. Lastly, Section 5 is devoted to the presentation and comparison of some numerical results.

## 2. TWO EXISTING APPROACHES

We begin by presenting in this section two existing approaches for the numerical solution of (1.1), and by discussing their advantages and drawbacks. We will then propose a third new approach in Section 3.

### 2.1. The uncoupled approach

A first and simple approach to the numerical solution of (1.1) has been used in e.g. [15]. It consists in treating separately at each time step the Euler equations and the species equation in (1.1). For Euler equations, one uses one of the classical (single-component) “Euler schemes” (such as the approximate Riemann solver of Osher [22], or of Roe [24], or the flux vector splitting of Steger and Warming [26] or of Van Leer [28]), with a “frozen  $\gamma$ ”. This means that, for the evaluation of the first three components of  $\phi_{i+1/2}$  at time  $t^n$ , one uses the flux of one of the above Euler schemes computed using a frozen value  $\gamma_{i+1/2} = \gamma \left( \frac{Y_i^n + Y_{i+1}^n}{2} \right)$ . Beside this, one uses an upwind approximation of the donor-cell type for the species equation; defining  $u_{i+1/2}^n = \frac{1}{2}(u_i^n + u_{i+1}^n)$ , one evaluates the fourth component  $\phi_{i+1/2}^4$  of  $\phi_{i+1/2}$  by:

$$\phi_{i+1/2}^4 = u_{i+1/2}^n \times \begin{cases} (\rho Y)_i^n & \text{if } u_{i+1/2}^n > 0, \\ (\rho Y)_{i+1}^n & \text{if } u_{i+1/2}^n < 0. \end{cases} \quad (2.1)$$

In the sequel, this uncoupled approximation will be referred to as “approach (A)”.

## 2.2. The fully coupled approach

A second type of approximation for system (1.1) has been used in [1], [2], [4], [10], [14], and is discussed in detail in [19]. It consists in seeing (1.1) as a whole, and in extending to this bigger system of conservation laws the upwind schemes (of Osher, Roe, Van Leer, Steger and Warming...) which have been developed for the single-component case. This approach, where the Euler equations and the species equation are not separated, will be referred to in the sequel as “approach (B)”.

Referring to [19] for the details, we just recall here the basic facts about system (1.1): the components of the flux vector  $F$  can be expressed as functions of the conservative variables  $W^l$  ( $1 \leq l \leq 4$ ), and the system  $W_t + F(W)_x = 0$  is hyperbolic: the Jacobian matrix  $A(W) = \left( \frac{\partial F^l}{\partial W^m} \right)$  has four real eigenvalues  $\lambda_1 = u - c$ ,  $\lambda_2 = u$ ,  $\lambda_3 = u$  and  $\lambda_4 = u + c$ , where  $c = \sqrt{\frac{\gamma p}{\rho}}$  is the sound speed, and four real eigenvectors.

It is then possible to extend to (1.1) the classical upwind schemes developed for the Euler equations. We now briefly describe these extended schemes, referring to [19] and the references therein for more details.

### 2.2.1. The multi-component Steger and Warming scheme

As in the single-component case (see [26]), the definition of the multi-component Steger and Warming scheme relies on the homogeneity property:

$$F(rW) = rF(W) \text{ for any } r > 0, \quad (2.2)$$

which can be readily checked using (1.4), and implies the Euler identity:

$$F(W) = A(W)W. \quad (2.3)$$

We need here to introduce a classical notation: if  $B$  is a diagonalisable matrix, and if  $f$  maps  $\mathbb{R}$  into itself, we define the matrix  $f(B)$  as follows: we write the diagonalisation of  $B$  as  $B = T\Lambda T^{-1}$ ,  $\Lambda = \text{Diag}[\mu_1, \mu_2 \cdots \mu_n]$ , and set  $f(B) = Tf(\Lambda)T^{-1}$ , where  $f(\Lambda) = \text{Diag}[f(\mu_1), f(\mu_2) \cdots f(\mu_n)]$ . We will use this definition below for  $f(x) = |x|$ ,  $f(x) = x_+ = \max(x, 0)$  or  $f(x) = x_- = \min(x, 0)$ .

Then, exactly as in the single-component case, we define the extended Steger and Warming flux function by:

$$\Phi(W_L, W_R) = A_+(W_L)W_L + A_-(W_R)W_R . \quad (2.4)$$

The extended Steger and Warming scheme is a “flux-vector-splitting” scheme (see [16]), since, setting  $F_+(W) = A_+(W)W$ ,  $F_-(W) = A_-(W)W$ , we have:

$$F(W) = F_+(W) + F_-(W) , \quad (2.5)$$

and:

$$\Phi(W_L, W_R) = F_+(W_L) + F_-(W_R) . \quad (2.6)$$

We also have here the property that, if all characteristic wave speeds associated with the state  $W$  are positive (resp. negative), that is if  $u \geq c$  (resp.  $u \leq -c$ ), then  $F_+(W) = F(W)$  (resp.  $F_-(W) = F(W)$ ). Moreover, the following result, which says that this flux splitting can be used to define a stable conservative scheme, is proved in [19]:

**Proposition 1:**

If the specific heat ratio  $\gamma_k$  of each species in the mixture satisfies the inequality  $1 < \gamma_k < \frac{5}{3}$ , then all eigenvalues of the Jacobian matrix  $\frac{DF_+}{DW}$  (resp.  $\frac{DF_-}{DW}$ ) are real and positive (resp. negative). •

**2.2.2. The multi-component Van Leer scheme**

The extension of the continuously differentiable flux splitting of Van Leer [28] to multi-component flows was first derived in [19]. The numerical flux function has again the form (2.6), where  $F_+$  and  $F_-$  are defined by (2.5) and the following expressions:

\* if  $u \geq c = \sqrt{\frac{\gamma P}{\rho}}$ ,  $F_+(W) = F(W)$ ;

\* if  $-c \leq u \leq c$ ,

$$F_+(W) = \begin{pmatrix} F_+^1 \\ F_+^2 \\ F_+^3 \\ F_+^4 \end{pmatrix} = \begin{pmatrix} \frac{\rho}{4c}(u+c)^2 \\ F_+^1 \left( u - \frac{u-2c}{\gamma} \right) \\ \frac{\gamma^2}{2(\gamma^2-1)} \frac{(F_+^2)^2}{F_+^1} \\ Y F_+^1 \end{pmatrix} , \quad (2.7)$$

(the first three components of  $F_+$  have the same expression as in the single-component case, but with the non constant coefficient  $\gamma$  given by (1.3));

\* if  $u \leq -c$ ,  $F_+(W) = 0$ .

The analogue of Proposition 1 also holds here (see [19] for the proof):

**Proposition 2:**

If the specific heat ratio  $\gamma_k$  of each species in the mixture satisfies the inequality  $1 < \gamma_k < 3$ , then all eigenvalues of the Jacobian matrix  $\frac{DF_+}{DW}$  (resp.  $\frac{DF_-}{DW}$ ) are real and positive (resp. negative). •

**2.2.3. The multi-component Roe scheme**

The extension of the conservative upwind scheme proposed by Roe [24] to mixtures has been derived in [1] and [10]. The numerical flux function of the extended Roe scheme has the same expression as in the single-component case:

$$\Phi(W_L, W_R) = \frac{F(W_L) + F(W_R)}{2} + \frac{1}{2} |\tilde{A}| (W_L - W_R), \quad (2.8)$$

where  $\tilde{A} = \tilde{A}(W_L, W_R)$  is a diagonalisable matrix which satisfies the property:

$$F(W_L) - F(W_R) = \tilde{A}(W_L - W_R). \quad (2.9)$$

When  $\gamma$  is constant, that is when  $\gamma_1 = \gamma_2$  in (1.3), then the matrix  $\tilde{A}(W_L, W_R)$  is equal to the Jacobian matrix  $A(\tilde{W})$ , where the state  $\tilde{W}$  is ‘‘Roe’s average’’ of  $W_L$  and  $W_R$ . When  $\gamma_1 \neq \gamma_2$ , the matrix  $A(\tilde{W})$  does no longer satisfy (2.9), and the matrix  $\tilde{A}$  has to be chosen close to but different from  $A(\tilde{W})$  (see [1], [10], [19]).

**2.2.4. The multi-component Osher scheme**

The multi-component Osher scheme has been derived in [2]. As in the single-component case, the numerical flux function of the extended Osher scheme has the expression:

$$\Phi(W_L, W_R) = \frac{F(W_L) + F(W_R)}{2} + \int_{W_L}^{W_R} |A(W)| dW, \quad (2.10)$$

where the integration is carried out along a path connecting  $W_L$  to  $W_R$  in the state-space. The integration path is piecewise parallel to the right eigenvectors of the flux Jacobian matrix  $A$ , and the evaluation of the integral in (2.10) requires the knowledge



of the Riemann invariants associated with each eigenvector (see [2] and [19] for the details).

### 2.3. Discussion

The main advantage of approach (A) is its simplicity: even in two or three space dimensions, following this approach to transform a single-component Euler code into a multi-component code requires a very small programming and computational effort. This approach also has the advantage of preserving the positivity of the mass fraction  $Y$  (the proof of this fact is similar to the one of Lemma 3 below). But this approach also has important drawbacks, which will be discussed in more detail below.

Compared to approach (A), all schemes of type (B), except the extended Van Leer scheme, are substantially more expensive from the computational point of view. This is especially true for the extended Roe or Steger and Warming schemes:  $4 \times 4$  Jacobian matrices have to be handled (that is,  $N + 2 \times N + 2$  matrices if one considers a mixture of  $N$  species), while only  $3 \times 3$  matrices are involved in the corresponding schemes of type (A). Nevertheless, approach (B), which is more satisfactory from a theoretical point of view, actually gives better results than approach (A): the accuracy of the results is greatly improved (see Section 5 below).

However, we need to make precise here that the “discrete maximum principle” for the mass fraction does not hold with any of the two approaches (that is, the inequalities  $0 \leq Y \leq 1$  are not preserved by these schemes). But in practice, the mass fraction values obtained with approach (B) remain much closer to the interval  $[0, 1]$  than those obtained with approach (A) (see Section 5 and [5]).

Intuitively, one can feel the origin of the inaccuracy of the computed mass fraction values in approach (A) as follows: the mass fraction  $Y$  is evaluated as the ratio of two conservative variables:

$$Y = \frac{W^4}{W^1} = \frac{\rho Y}{\rho} ; \quad (2.11)$$

in approach (A), these two variables  $W^1$  and  $W^4$  are advanced in time using two different approximations (more precisely, two different upwinding techniques). On the opposite, a single global and coupled approach is used for all conservative variables  $W^l$  in the schemes of type (B).

This remark leads us to propose a third approach, which we present in the next section.

### 3. A THIRD APPROACH

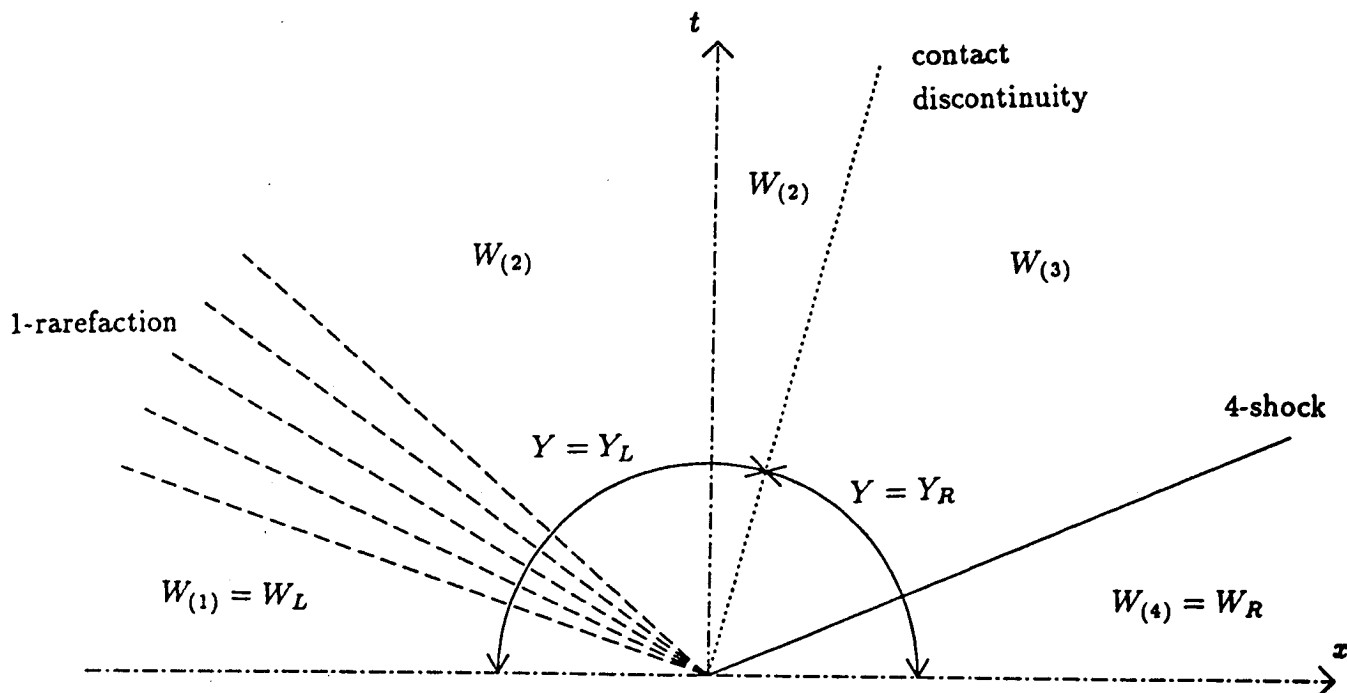
Before presenting this new approach, we need to state some simple properties of the exact solution of the Riemann problem for system (1.1).

#### 3.1. The multi-component Riemann problem

Two states  $W_L$  and  $W_R$  being given, we consider the Riemann problem:

$$\begin{cases} W_t + F(W)_x = 0 & \text{for } x \in \mathbb{R}, t \geq 0, \\ W(x, 0) = \begin{cases} W_L & \text{if } x < 0, \\ W_R & \text{if } x > 0. \end{cases} \end{cases} \quad (3.1)$$

This problem is solved in [1], [2], [19]. Its exact solution  $W^{\mathcal{R}}(x, t)$  is presented on Figure 1. It is of course self-similar (i.e.  $W^{\mathcal{R}}(x, t)$  only depends on the ratio  $\frac{x}{t}$ ), and consists, 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 wave associated with the first characteristic field, that is with the eigenvalue  $\lambda_1 = u - c$ , either a 1-shock or a 1-rarefaction wave;  $W_{(2)}$  and  $W_{(3)}$  are separated by a contact discontinuity (associated with the eigenvalue  $u$ );  $W_{(3)}$  and  $W_{(4)} = W_R$  are separated by a 4-wave, associated with  $\lambda_4 = u + c$ . 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) and only varies across the contact discontinuity.



**Figure 1:** The solution of the multi-component Riemann problem (3.1).

For  $\sigma \in \mathbb{R}$  and  $t \geq 0$ , we will denote  $\mathcal{W}(\sigma; W_L, W_R)$  (or sometimes simply  $\mathcal{W}(\sigma)$ ) the value of  $W^{\mathcal{R}}(\sigma t, t)$ , which is independent of  $t$ . We can then state the main property of this section:

**Lemma 1:**

For any states  $W_L$  and  $W_R$ , the following equality holds:

$$F^4[\mathcal{W}(0)] = F^1[\mathcal{W}(0)] \times \begin{cases} Y_L & \text{if } F^1[\mathcal{W}(0)] > 0, \\ Y_R & \text{if } F^1[\mathcal{W}(0)] < 0. \bullet \end{cases} \quad (3.2)$$

PROOF: Let us call  $u^*(W_L, W_R)$  the speed of the contact discontinuity. For any  $\sigma \in \mathbb{R}$ , we have from Figure 1:

$$Y(\mathcal{W}(\sigma)) = \begin{cases} Y_L & \text{if } \sigma < u^*, \\ Y_R & \text{if } \sigma > u^*. \end{cases} \quad (3.3)$$

Using the definitions of  $F^1 = \rho u$  and  $F^4 = \rho u Y$ , we can write:

$$F^4[\mathcal{W}(0)] = F^1[\mathcal{W}(0)] \times \begin{cases} Y_L & \text{if } u^* > 0, \\ Y_R & \text{if } u^* < 0, \end{cases} \quad (3.4)$$

and the end of the proof then relies on Lemma 2 below.  $\bullet$

**Remark 3:** It can be noticed that (3.2) and (3.4) remain valid if  $F^1(\mathcal{W}(0)) = 0$  or if  $u^* = 0$ , since we then have  $F^4(\mathcal{W}(0)) = F^1(\mathcal{W}(0)) = 0$  (we recall that, by definition of the contact discontinuity,  $u^*$  is also the velocity of the two states separated by this discontinuity:  $u(W_{(2)}) = u(W_{(3)}) = u^*$ ). •

**Remark 4:** We have assumed in the proof of Lemma 1 that a contact discontinuity actually exists in the structure of the exact solution  $W^{\mathcal{R}}$ . If this is not case, then  $Y(x, t)$  is necessarily constant,  $Y(x, t) \equiv Y_L = Y_R$ , and (3.2) holds again. •

**Lemma 2:**

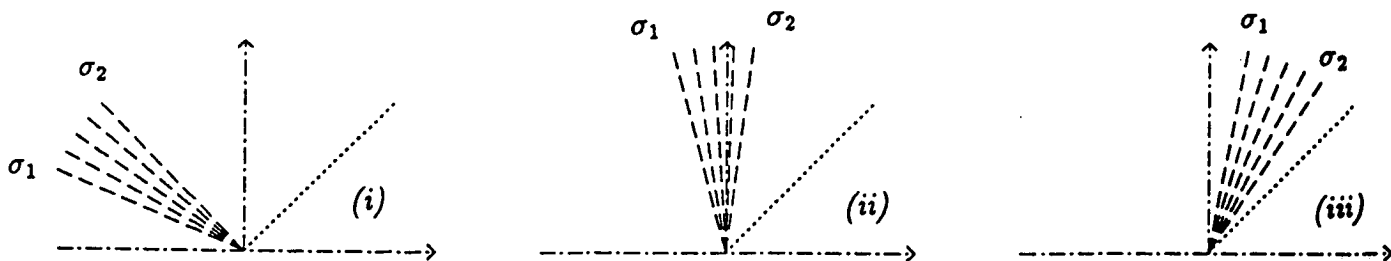
For any states  $W_L$  and  $W_R$ , the quantities  $F^1(\mathcal{W}(0; W_L, W_R))$  and  $u^*(W_L, W_R)$  have the same sign. •

**PROOF:** It consists in considering the structure of the exact solution  $W^{\mathcal{R}}$  of the Riemann problem and in using the classical properties of the rarefaction and shock waves. Let us assume, without loss of generality, that  $u^* > 0$ ; we are going to show that  $F^1(\mathcal{W}(0)) > 0$ .

a) Consider first the situation where the 1-wave is a rarefaction wave, located in the region  $\sigma_1 \leq \frac{x}{t} \leq \sigma_2$ , with  $\sigma_2 < u^*$ . Since this wave is associated with the first eigenvalue  $\lambda_1 = u - c$ , we have (see [20]):

$$(u - c)(\mathcal{W}(\sigma)) = \sigma, \quad \forall \sigma \in [\sigma_1, \sigma_2]. \quad (3.5)$$

We then have to consider three different cases (see Figure 2):



**Figure 2:** The three different configurations for the 1-rarefaction.

(i) If  $\sigma_2 < 0$ :

Then  $\mathcal{W}(0) = W_{(2)}$ ,  $u(\mathcal{W}(0)) = u^* > 0$ , whence  $F^1(\mathcal{W}(0)) > 0$ .

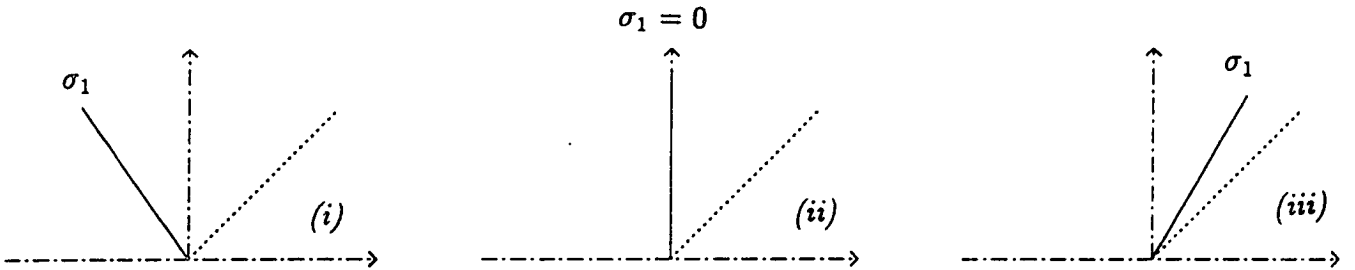
(ii) If  $\sigma_1 \leq 0 \leq \sigma_2$ :

Applying (3.5) for  $\sigma = 0$ , we see that  $u(\mathcal{W}(0)) > 0$ , whence  $F^1(\mathcal{W}(0)) > 0$ .

(iii) If  $\sigma_1 > 0$ :

Then  $\mathcal{W}(0) = \mathcal{W}(\sigma_1) = W_L$ , and (3.5) yields  $(u - c)(W_L) = \sigma_1 > 0$ , whence  $u(\mathcal{W}(0)) = u(W_L) > 0$ , and again  $F^1(\mathcal{W}(0)) > 0$ .

b) Consider now the situation where the 1-wave is a shock wave propagating with the speed  $\sigma_1 < u^*$ . We again consider three different cases (see Figure 3):



**Figure 3:** The three different configurations for the 1-shock.

(i) If  $\sigma_1 < 0$ :

Then  $u(\mathcal{W}(0)) = u(W_{(2)}) = u^* > 0$ , and  $F^1(\mathcal{W}(0)) > 0$ .

(ii) If  $\sigma_1 = 0$ :

The Rankine-Hugoniot relations for the steady 1-shock imply (see [20]):

$$\rho u(W_L) = \rho u(W_{(2)}) = \rho(W_{(2)})u^* > 0, \quad (3.6)$$

and  $F^1(\mathcal{W}(0))$  is well defined and positive since  $F^1(W_L) = F^1(W_{(2)}) > 0$ .

(iii) If  $\sigma_1 > 0$ :

Then  $\mathcal{W}(0) = W_L$ , and the Lax conditions for the 1-shock imply (see [20]):

$$(u - c)(W_L) > \sigma_1 > 0. \quad (3.7)$$

This yields  $u(W_L) > 0$ , whence  $F^1(\mathcal{W}(0)) > 0$ . •

### 3.2. The approximate multi-component Riemann flux

We can now present the third approach (C), which is based on property (3.2).

In order to introduce it, let us first recall that in all upwind schemes, and in particular in the so-called “Godunov-type” schemes based on approximate Riemann solvers (see [16]), the numerical flux  $\phi_{i+1/2}$  can be seen as an approximation of the flux  $F[\mathcal{W}(0; W_i^n, W_{i+1}^n)]$  of the exact solution of the Riemann problem constructed with the neighbouring states  $W_i^n$  and  $W_{i+1}^n$ . One even has the equality  $\phi_{i+1/2} = F[\mathcal{W}(0; W_i^n, W_{i+1}^n)]$  in the original Godunov method [13].

With this idea in mind, we define approach (C) as follows. The first three components  $\phi_{i+1/2}^1, \phi_{i+1/2}^2, \phi_{i+1/2}^3$  of the flux are evaluated as in approach (A), using a classical “Euler scheme” for the Euler equations, with  $\gamma$  locally frozen, but the fourth component of the flux is defined by the next relation, which mimics (3.2):

$$\phi_{i+1/2}^4 = \phi_{i+1/2}^1 \times \begin{cases} Y_i^n & \text{if } \phi_{i+1/2}^1 > 0, \\ Y_{i+1}^n & \text{if } \phi_{i+1/2}^1 < 0. \end{cases} \quad (3.8)$$

Therefore, this third approach is as simple as (A) from the computational point of view. But, like (B), approach (C) is constructed by considering as a whole the multi-component system (1.1), and retains some coupling between the Euler equations and the species equation: approach (C) defines an approximate Riemann flux for the whole multi-component system, and uses the same discrete mass fluxes for the species equation and the continuity equation.

Since the approximations used for the variables  $W^1$  and  $W^4$  are now well coupled, we may expect better result for the mass fraction  $Y = \frac{W^4}{W^1}$ . This is the object of the next lemma and of the following remarks:

**Lemma 3:**

Under the following CFL-like condition:

$$\frac{\Delta t}{\Delta x} \left[ \frac{\max(\phi_{i+1/2}^1, 0)}{\rho_i^n} - \frac{\min(\phi_{i+1/2}^1, 0)}{\rho_{i+1}^n} \right] \leq 1 \quad \forall i, \quad (3.9)$$

the schemes of type (C) preserve the maximum principle for the mass fraction  $Y$ : for all  $i$  and  $n \geq 0$ :

$$\min_j Y_j^0 \leq Y_i^n \leq \max_j Y_j^0. \bullet \quad (3.10)$$

PROOF: Setting  $\lambda = \frac{\Delta t}{\Delta x}$ , we have:

$$\rho_i^{n+1} = \rho_i^n - \lambda(\phi_{i+1/2}^1 - \phi_{i-1/2}^1), \quad (3.11)$$

$$\rho_i^{n+1} Y_i^{n+1} = \rho_i^n Y_i^n - \lambda(\phi_{i+1/2}^4 - \phi_{i-1/2}^4). \quad (3.12)$$

Let us temporarily assume that  $\phi_{i+1/2}^1 > 0$ ,  $\phi_{i-1/2}^1 > 0$ . Then, (3.8), (3.11) and (3.12) imply:

$$Y_i^{n+1} = \frac{Y_i^n(\rho_i^n - \lambda\phi_{i+1/2}^1) + Y_{i-1}^n(\lambda\phi_{i-1/2}^1)}{(\rho_i^n - \lambda\phi_{i+1/2}^1) + (\lambda\phi_{i-1/2}^1)}; \quad (3.13)$$

if (3.9) holds, (3.13) shows that  $Y_i^{n+1}$  is a convex linear combination of  $Y_i^n$  and  $Y_{i-1}^n$ .

It is easy to check that the same conclusion also holds with different hypotheses on the signs of  $\phi_{i+1/2}^1$  and  $\phi_{i-1/2}^1$ , which concludes the proof. •

**Remark 5 :** It is implicitly assumed in the above proof that  $\rho_i^{n+1}$  does not vanish. Therefore, rigorously speaking, the property (3.10) only holds if the ‘‘Euler scheme’’ used to evaluate  $\phi^1$ ,  $\phi^2$ ,  $\phi^3$  preserves the positivity of the density (this is of course the least one may ask to this ‘‘Euler scheme’’ !). •

**Remark 6:** One may think at first glance that the preceding proof does not use any property of the approximate mass flux  $\phi^1$ . This is not really the case. Indeed, it is easy to see that the CFL-like condition (3.9) which insures the discrete maximum principle for the mass fraction may well be impossible to fulfill in practice if one uses a centered approximation  $\phi_{i+1/2}^1 = \frac{1}{2}[(\rho u)_i + (\rho u)_{i+1}]$  (take for instance  $u_j \equiv \hat{u} > 0$  for all  $j$ ,  $\rho_i^n = \epsilon > 0$ ,  $\rho_{i+1}^n = 1$ : (3.9) then imposes that  $\frac{\hat{u}\Delta t}{\Delta x} \leq \frac{2\epsilon}{1+\epsilon}$  !). On the other hand, several numerical experiments have shown that the condition (3.9) is not more restrictive than the usual CFL condition when an upwind scheme is used to evaluate the first three components of  $\phi_{i+1/2}$  (see [5]). •

**Remark 7 :** The proof of Lemma 3 shows that the definition (3.8) of  $\phi_{i+1/2}^4$  is exactly what is needed in order to insure the maximum principle for the mass fraction. Schemes of type (A) or (B) do not satisfy (3.10) (see Section 5 below). •

**Remark 8:** Arguing as in the proof of Lemma 3, one can show that the schemes of type (C) preserve the monotonicity of the mass fractions: if  $Y_j^0 \leq Y_{j+1}^0$  for all  $j$ , then  $Y_j^n \leq Y_{j+1}^n$  for all  $j$  and  $n \geq 0$ . More generally, these schemes are *TVD* (Total Variation Diminishing) for the mass fraction  $Y$ . •

**Remark 9:** Method (C) can be used in conjunction with different Euler schemes: one may use any of the classical schemes (and in particular any of the schemes mentioned in Section 2) for the Euler equations, and then discretize the species equation using (3.8). However, there appears some particularity when the scheme of Van Leer [28] is used for the Euler equations. Indeed, we have seen that the multi-component Van Leer scheme

(approach (B)) gives the following expressions for the first and fourth components of the numerical flux:

$$\phi_{i+1/2}^1 = F_+^1(W_i^n) + F_-^1(W_{i+1}^n) , \quad (3.14)$$

$$\phi_{i+1/2}^4 = Y_i F_+^1(W_i^n) + Y_{i+1} F_-^1(W_{i+1}^n) . \quad (3.15)$$

Thus, the Van Leer scheme of type (B) uses the same discrete mass fluxes  $F_{\pm}^1$  for the continuity equation and the species equation, as the schemes of type (C). Moreover, since  $F_+^1(W) > 0$  and  $F_-^1(W) < 0$  for any  $W$ , it is easy to see that this Van Leer scheme of type (B) also preserves the maximum principle for the mass fraction. Therefore, there is no interest in using approach (C) with the Van Leer scheme. On the opposite, it can be checked that the Steger and Warming scheme of type (B) does not preserve the maximum principle for the mass fraction, although it is also a flux vector splitting method (one has not  $F_{\pm}^4(W) = Y F_{\pm}^1(W)$  for the Steger and Warming scheme). Thus, among the classical Euler schemes mentioned above, the approach (C) is relevant for the scheme of Steger and Warming and the Godunov-type schemes of Roe and Osher.

•

**Remark 10:** Consider the flow of a mixture of  $N$  species, with  $N \geq 2$ . Then, taking  $W = (\rho, \rho u, E, \rho Y_1, \dots, \rho Y_{N-1})$ , one will preserve the inequalities  $0 \leq Y_k \leq 1$  for  $1 \leq k \leq N - 1$  if approach (C) is used for the  $N - 1$  species equations. Moreover, it is easy to see that the last mass fraction  $Y_N$ , evaluated as  $(Y_N)_i^n = 1 - \sum_{k=1}^{N-1} (Y_k)_i^n$  will also remain in the interval  $[0, 1]$ . •

**Remark 11:** When a diffusive term and a consumption source term appear in the right-hand side of the species equations, which then has the form:

$$(\rho Y)_t + (\rho Y u)_x = Y_{xx} - KY , \quad (3.16)$$

it is easy to see that the maximum principle for the mass fraction is again preserved under an appropriate stability restriction on the time step if the convective term is approximated using approach (C) and if classical approximations are used for the diffusive and reactive terms (see [21]). This maximum principle now takes the following form:

$$\text{if } \min_j Y_j^0 \geq 0 , \text{ then } 0 \leq Y_i^n \leq \max_j Y_j^0 \quad \forall i, \forall n . \bullet \quad (3.17)$$

#### 4. EXTENSIONS

Before discussing in detail the advantages and drawbacks of approach (C) and comparing it to approaches (A) and (B) in Section 5, we now present how this type



of schemes can be extended to second-order accuracy, to implicit time-stepping or to multi-dimensional flows while keeping the basic property (3.10).

#### 4.1. Second-order accuracy

Let us therefore consider second-order extensions of the preceding schemes. More precisely, we will first consider schemes which are second-order accurate in space but remain first-order accurate in time. Starting from the previous first-order accurate schemes, the second-order spatial accuracy is obtained by using piecewise linear variables instead of piecewise constant variables, following the "MUSCL" approach of Van Leer [27]. This method involves three steps:

(a) At each time step, starting from the values  $W_i^n$ , one first evaluates slopes  $s_i^n$  for all variables which are chosen to be piecewise linear. Several choices are possible at this stage (for instance, one can choose either the conservative variables  $\rho, \rho u, E, \rho Y$  or the "physical variables"  $\rho, u, p, Y$  to vary linearly in each computational cell); here, we take  $Y$  (and not  $\rho Y$ ) as a piecewise linear variable.

(b) Slope limiters are then used in order to avoid the creation of new extrema; here again different strategies exist to evaluate the limited slopes. We assume that the slopes are constrained such that, taking the variable  $Y$  as an example, we have:

$$\min_{|j-i|\leq 1} Y_j^n \leq Y_i^n \pm \frac{\Delta x}{2} s_i^n \leq \max_{|j-i|\leq 1} Y_j^n. \quad (4.1)$$

(c) The limited slopes are then used to evaluate cell-interface values  $W_{i+1/2,\pm}^n$  (one sets  $Y_{i+1/2,-}^n = Y_i^n + \frac{\Delta x}{2} s_i^n$ ,  $Y_{i-1/2,+}^n = Y_i^n - \frac{\Delta x}{2} s_i^n$ ), and the solution is advanced in time according to relation (1.5), where we now take:

$$\phi_{i+1/2} = \Phi(W_{i+1/2,-}^n, W_{i+1/2,+}^n). \quad (4.2)$$

This construction (a)-(c) can therefore be applied to any of the schemes presented in the preceding sections; in particular, we can construct a limited second-order accurate scheme of type (C) if we use in (4.2) a numerical flux function of type (C), that is with the fourth component  $\phi^4$  of the flux evaluated using (3.8).

Then, we have the following:

**Lemma 4:**

Under the more restrictive CFL-like condition:

$$\frac{\Delta t}{\Delta x} \left[ \frac{\max(\phi_{i+1/2}^1, 0)}{\rho_i^n} - \frac{\min(\phi_{i+1/2}^1, 0)}{\rho_{i+1}^n} \right] \leq \frac{1}{2} \quad \forall i, \quad (4.3)$$

the spatially second-order accurate schemes of type (C) constructed above preserve the maximum principle for the mass fraction  $Y$ : for all  $i$  and  $n \geq 0$ , we have:

$$\min_j Y_j^0 \leq Y_i^n \leq \max_j Y_j^0. \bullet \quad (4.4)$$

PROOF: We will assume that  $\phi_{i+1/2}^1 > 0$ ,  $\phi_{i-1/2}^1 > 0$  (the proof is very similar in the other cases). Using relations (3.11)-(3.12), we have:

$$\begin{aligned} Y_i^{n+1} &= \frac{\rho_i^n Y_i^n - \lambda \phi_{i+1/2}^1 Y_{i+1/2,-}^n + \lambda \phi_{i-1/2}^1 Y_{i-1/2,-}^n}{\rho_i^n - \lambda \phi_{i+1/2}^1 + \lambda \phi_{i-1/2}^1} \\ &= \frac{(\rho_i^n - \lambda \phi_{i+1/2}^1) \left( \frac{\rho_i^n Y_i^n - \lambda \phi_{i+1/2}^1 Y_{i+1/2,-}^n}{\rho_i^n - \lambda \phi_{i+1/2}^1} \right) + \lambda \phi_{i-1/2}^1 Y_{i-1/2,-}^n}{(\rho_i^n - \lambda \phi_{i+1/2}^1) + \lambda \phi_{i-1/2}^1}; \end{aligned} \quad (4.5)$$

Let  $a$  and  $b$  be two real numbers such that  $a \leq Y_j^n \leq b$  for all  $j$ . Since  $Y_{i-1/2,-}^n \in [a, b]$  from (4.1), it suffices to check that:

$$\hat{Y} = \frac{\rho_i^n Y_i^n - \lambda \phi_{i+1/2}^1 Y_{i+1/2,-}^n}{\rho_i^n - \lambda \phi_{i+1/2}^1} \in [a, b]. \quad (4.6)$$

Using the definition of  $Y_{i+1/2,-}^n$  and setting  $X = \lambda \frac{\phi_{i+1/2}^1}{\rho_i^n}$ , we get:

$$\hat{Y} = Y_i^n - \frac{X}{1-X} \frac{\Delta x}{2} s_i^n, \quad (4.7)$$

and (4.6) holds from (4.1) since  $0 \leq X \leq \frac{1}{2}$  from (4.3).  $\bullet$

**Remark 12:** The preceding proof clearly shows that it is preferable to use limited slopes for the variable  $Y$  and not for  $\rho Y$  when the approach (C) is used to discretize the species equation. Nevertheless, this strategy is almost but not completely sufficient to guarantee the preservation of the maximum principle for the mass fractions when a

fully second-order accurate scheme (in both space and time) is used. This question is discussed in the Appendix. •

**Remark 13:** In the case of a mixture of two species  $\mathcal{S}_1$  and  $\mathcal{S}_2$ , the preceding proof shows that the mass fractions of both species will remain in the interval  $[0, 1]$  when the above spatially second-order accurate method is used: since  $Y_1 = Y \in [0, 1]$ , then  $Y_2 = 1 - Y \in [0, 1]$ . But additional difficulties arise when one considers mixtures of  $N$  species with  $N > 2$ . Applying the above method to species  $k$ , we get  $(Y_k)_i^n \in [0, 1]$  for all  $k$ ,  $1 \leq k \leq N - 1$ ; but it is no more clear whether  $(Y_N)_i^n = 1 - \sum_{k=1}^{N-1} (Y_k)_i^n$  also is in the interval  $[0, 1]$  (this comes from the fact that the limiters are by their very essence *nonlinear*). A way to remedy this difficulty is the following: instead of treating separately the  $N^{th}$  species, one could treat all species in the same way, setting  $(Y_k)_{i+1/2,-} = (Y_k)_i^n + \frac{\Delta x}{2} (s_k)_i^n$  and applying the scheme of type (C) for all  $k$ ,  $1 \leq k \leq N$  (and not  $N - 1$ ). The maximum principle will be preserved if the slopes are constrained such that:

$$\min_{|j-i| \leq 1} (Y_k)_j^n - (Y_k)_i^n \leq \frac{\Delta x}{2} |(s_k)_i^n| \leq \max_{|j-i| \leq 1} (Y_k)_j^n - (Y_k)_i^n, \quad (4.8)$$

(which is equivalent to (4.1)), and:

$$\sum_{k=1}^N (s_k)_i^n = 0, \quad (4.9)$$

for all  $i$  and  $n$  (the last relation imposes that the equality  $\sum_{k=1}^N (Y_k)_i^n = 1$  is preserved). •

## 4.2. Implicit time stepping

Let us now consider the extension to an implicit time stepping. As in the explicit case, the implicit (C) schemes will be based on the principle that the species equation is integrated after the Euler equations, using the discrete mass fluxes of the continuity equation.

Let us therefore consider as given a conservative implicit scheme for the Euler equations, which we write as:

$$\frac{(W^l)_i^{n+1} - (W^l)_i^n}{\Delta t} + \frac{\phi_{i+1/2}^{l,n+1} - \phi_{i-1/2}^{l,n+1}}{\Delta x} = 0, \quad (4.10)$$

for  $l = 1, 2, 3$ , where in general  $\phi_{i+1/2}^{l,n+1}$  depends on  $W_i^n, W_{i+1}^n, W_i^{n+1}, W_{i+1}^{n+1}$ : the values  $(W^l)_i^{n+1} (1 \leq l \leq 3)$  are computed from the values  $W_i^n$  by solving at each time step either a nonlinear or a linear discrete problem (see e.g. [7], [12] for such implicit versions of the most classical upwind Euler schemes).

Then, we construct an implicit multi-component scheme of type (C) by adding the following discrete species equation:

$$\frac{(\rho Y)_i^{n+1} - (\rho Y)_i^n}{\Delta t} + \frac{\phi_{i+1/2}^{4,n+1} - \phi_{i-1/2}^{4,n+1}}{\Delta x} = 0, \quad (4.11)$$

where we take:

$$\phi_{i+1/2}^{4,n+1} = \phi_{i+1/2}^{1,n+1} \times \begin{cases} Y_i^{n+1} & \text{if } \phi_{i+1/2}^{1,n+1} > 0, \\ Y_{i+1}^{n+1} & \text{if } \phi_{i+1/2}^{1,n+1} < 0. \end{cases} \quad (4.12)$$

Therefore the solution of a linear system is required at each time step to evaluate the mass fractions  $Y_i^{n+1}$ .

The next lemma shows that this scheme still has the property (3.10), without any restriction on the time step.

**Lemma 5:**

For any value of the time step  $\Delta t > 0$ , the scheme (4.10)-(4.12) preserves the maximum principle for the mass fraction  $Y$ : for all  $i$  and  $n \geq 0$ :

$$\min_j Y_j^0 \leq Y_i^n \leq \max_j Y_j^0. \bullet \quad (4.13)$$

PROOF: The mass fractions  $Y_i^{n+1}$  at time level  $n + 1$  are obtained by solving the tridiagonal system:

$$\sum_j A_{i,j}^{n+1} Y_j^{n+1} = \sum_j B_{i,j}^n Y_j^n, \quad (4.14)$$

where  $B_{i,j}^n = \rho_i^n \delta_{i,j}$  ( $\delta_{i,j}$  is the Kronecker delta), and:

$$\begin{cases} A_{i,i}^{n+1} = \rho_i^{n+1} + \lambda \max(\phi_{i+1/2}^{1,n+1}, 0) - \lambda \min(\phi_{i-1/2}^{1,n+1}, 0), \\ A_{i,i+1}^{n+1} = \lambda \min(\phi_{i+1/2}^{1,n+1}, 0), \\ A_{i,i-1}^{n+1} = -\lambda \max(\phi_{i-1/2}^{1,n+1}, 0). \end{cases} \quad (4.15)$$

It is clear from these expressions that the matrix  $A = (A_{i,j}^{n+1})$  is an M-matrix, i.e. that:

$$\begin{cases} A_{i,i}^{n+1} > 0 \quad \forall i, \\ A_{i,j}^{n+1} < 0 \quad \forall i \neq j, \\ \sum_j A_{i,j}^{n+1} > 0 \quad \forall i. \end{cases} \quad (4.16)$$

These properties imply that the matrix  $A^{-1}$  has only positive elements (see e.g. [29]):  $(A^{-1})_{i,j} \geq 0$  for all  $i$  and  $j$ . Thus, writing  $C = A^{-1}B$  (where  $B = (B_{i,j}^n)$ ), we see that:

$$C_{i,j} \geq 0 \text{ for all } i \text{ and } j . \quad (4.17)$$

Denoting now  $U$  the vector whose components are all equal to 1, we see from (4.11)-(4.12) and (4.10) written for  $l = 1$  that  $AU = BU$ , whence  $CU = U$ , that is:

$$\sum_j C_{i,j} = 1 \text{ for all } i . \quad (4.18)$$

Since (4.14) implies that  $Y_i^{n+1} = \sum C_{i,j} Y_j^n$ , (4.18) and (4.19) show that  $Y_i^{n+1}$  is again a linear convex combination of the old mass fraction values  $Y_j^n$ , which ends the proof. •

**Remark 14:** The preceding proof is slightly too vague: we have not taken any boundary condition into account. In practice, the matrix  $A$  and the vector  $U$  are of finite size, and one has to check that the discrete maximum principle (4.13) holds with the actual discrete boundary conditions. •

**Remark 15:** In practice, implementing the implicit scheme (4.10)-(4.12) may be far less simple than implementing the explicit schemes of type (C). Indeed, the implicit integration of the Euler equations often uses a linearized implicit flux (see e.g. [7], [12]); that is, the fluxes  $\phi_{i+1/2}^{l,n+1}$  in (4.10) is given as:

$$\phi_{i+1/2}^{l,n+1} = \Phi^l(W_i^n, W_{i+1}^n) + \frac{\partial \Phi^l}{\partial W_i} (W_i^{n+1} - W_i^n) + \frac{\partial \Phi^l}{\partial W_{i+1}} (W_{i+1}^{n+1} - W_{i+1}^n) , \quad (4.19)$$

which leads to solving at each time step a linear system for computing  $\rho_i^{n+1}$ ,  $(\rho u)_i^{n+1}$ ,  $E_i^{n+1}$  for all  $i$ . Once these values are known, one has to use them to evaluate the fluxes  $\phi_{i+1/2}^{1,n+1}$  given by (4.19) before one can solve the species equation using the species fluxes (4.12) (see [21] for more details). •

**Remark 16:** In some circumstances, such as for highly subsonic flows, it may be of interest to integrate explicitly the species equation while treating implicitly the Euler equations (see [9]). Then, to construct a semi-implicit multi-component scheme of type (C), we simply replace (4.12) by:

$$\phi_{i+1/2}^{4,n+1} = \phi_{i+1/2}^{1,n+1} \times \begin{cases} Y_i^n & \text{if } \phi_{i+1/2}^{1,n+1} > 0 , \\ Y_{i+1}^n & \text{if } \phi_{i+1/2}^{1,n+1} < 0 . \end{cases} \quad (4.20)$$

This scheme again preserves the maximum principle (4.13) for the mass fraction. •

**Remark 17:** Consider again the flow of  $N$  species, with  $N > 2$ . Then, the matrix  $A$  appearing in the linear system which is to be solved at each time step for updating the mass fraction values is the same for all species equations; moreover,  $A$  is an M-matrix, which makes it easier to solve the linear systems •

**Remark 18:** Remark 11 also holds for implicit schemes. When the species equation has the form (3.16), the maximum principle (3.17) is preserved if the implicit approach (C) is used in conjunction with classical explicit or implicit approximations of the diffusive and reactive terms (more precisely, the maximum principle is preserved for any value of the time step if all terms are treated implicitly, whereas it is preserved under classical stability conditions if the diffusive or the reactive term is integrated explicitly). •

### 4.3. Multidimensional flows

All above methods which have been presented so far in a one-dimensional context can be extended to multidimensional flows. We simply recall here the basic features of this extension to two-dimensional geometries; to make the approach very general, we consider a two-dimensional unstructured triangular mesh (see Figure 4), following the ideas of Dervieux and Fezoui (see [7], [11]); but all approaches (A), (B) and (C) can of course be used on structured quadrilateral finite-volume meshes.

Let us therefore consider the two-dimensional multi-component Euler equations  $\hat{W}_t + \hat{F}_x + \hat{G}_y = 0$ , with:

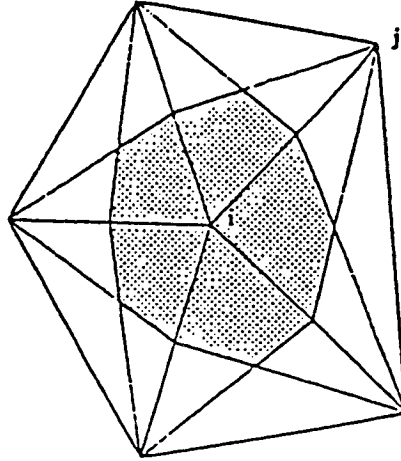
$$\hat{W} = \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ E \\ \rho Y \end{pmatrix}, \quad \hat{F} = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ u(E + p) \\ \rho u Y \end{pmatrix}, \quad \hat{G} = \begin{pmatrix} \rho v \\ \rho uv \\ \rho v^2 + p \\ v(E + p) \\ \rho v Y \end{pmatrix}, \quad (4.21)$$

with:

$$p = (\gamma - 1) \left[ E - \frac{1}{2} \rho (u^2 + v^2) \right], \quad (4.22)$$

where  $\gamma$  is again the local specific heat ratio (1.3) of the mixture.

Then, we consider a dual partition of the domain in control volumes or cells: a cell  $C_i$  is constructed around each vertex  $S_i$  by means of the medians of the neighbouring triangles, as shown on Figure 4.



**Figure 4:** The control volume  $C_i$ .

Integrating the system  $\hat{W}_t + \hat{F}_x + \hat{G}_y = 0$  on the control volume  $C_i$ , we get:

$$\int \int_{C_i} \hat{W}_t + \int_{\partial C_i} (\hat{F}\nu_i^x + \hat{G}\nu_i^y) = 0, \quad (4.23)$$

where  $\vec{\nu}_i = (\nu_i^x, \nu_i^y)$  is the outward unit normal on  $\partial C_i$ . In fact, we rewrite (4.23) as:

$$\int \int_{C_i} \hat{W}_t + \sum_{j \in \mathcal{K}(i)} \int_{\partial C_{ij}} (\hat{F}\nu_i^x + \hat{G}\nu_i^y) = 0, \quad (4.24)$$

where  $\mathcal{K}(i)$  is the set of neighbouring nodes of  $S_i$ , and where  $\partial C_{ij} = \partial C_i \cap \partial C_j$ . Then, the flux integrals in (4.24) are approximated using a 1D-like procedure, as explained below:

The fact that the system  $\hat{W}_t + \hat{F}(\hat{W})_x + \hat{G}(\hat{W})_y = 0$  is a nonlinear hyperbolic system of conservation laws means that, for any  $(\alpha, \beta) \in \mathbb{R}^2$ , the matrix  $\alpha \frac{\partial \hat{F}}{\partial \hat{W}} + \beta \frac{\partial \hat{G}}{\partial \hat{W}}$  has five real eigenvalues:

$$\begin{cases} \lambda_1 = \alpha u + \beta v - \sqrt{\alpha^2 + \beta^2} c, \\ \lambda_2 = \lambda_3 = \lambda_4 = \alpha u + \beta v, \\ \lambda_5 = \alpha u + \beta v + \sqrt{\alpha^2 + \beta^2} c, \end{cases} \quad (4.25)$$

with  $c = \sqrt{\frac{\gamma p}{\rho}}$ , and a complete set of real eigenvectors. Thus, we can extend all approximations defined in Section 2 for the one-dimensional flux vector  $F$  to the flux

vector  $\alpha\hat{F} + \beta\hat{G}$  (in other words, we use here the rotational invariance of the Euler equations). Consider for instance the Steger and Warming approximation: given two values  $\hat{W}_L$  and  $\hat{W}_R$ , and a vector  $\vec{\eta} = (\eta^x, \eta^y)$ , we define a numerical flux function  $\Phi$  by:

$$\Phi(\hat{W}_L, \hat{W}_R, \vec{\eta}) = (A_\eta)_+(\hat{W}_L)\hat{W}_L + (A_\eta)_-(\hat{W}_R)\hat{W}_R, \quad (4.26)$$

where  $A_\eta$  is the matrix  $\eta^x \frac{\partial \hat{F}}{\partial \hat{W}} + \eta^y \frac{\partial \hat{G}}{\partial \hat{W}}$ ; thus, we define the vector  $\vec{\nu}_{ij} = (\nu_{ij}^x, \nu_{ij}^y)$  by:

$$\nu_{ij}^x = \int_{\partial C_{ij}} \nu_i^x, \quad \nu_{ij}^y = \int_{\partial C_{ij}} \nu_i^y, \quad (4.27)$$

and we obtain a first-order accurate explicit upwind approximation of (4.24) by writing:

$$\frac{\hat{W}_i^{n+1} - \hat{W}_i^n}{\Delta t} \text{area}(C_i) + \sum_{j \in \mathcal{K}(i)} \Phi(\hat{W}_i^n, \hat{W}_j^n, \vec{\nu}_{ij}) = 0. \quad (4.28)$$

This extension to two space dimensions can of course be used for method (C). Then, the first four components of  $\Phi(\hat{W}_i, \hat{W}_j, \vec{\nu}_{ij})$  are evaluated using an ‘‘Euler scheme’’, and the fifth component is given by:

$$\Phi^5(\hat{W}_i, \hat{W}_j, \vec{\nu}_{ij}) = \Phi^1(\hat{W}_i, \hat{W}_j, \vec{\nu}_{ij}) \times \begin{cases} Y_i & \text{if } \Phi^1(\hat{W}_i, \hat{W}_j, \vec{\nu}_{ij}) > 0, \\ Y_j & \text{if } \Phi^1(\hat{W}_i, \hat{W}_j, \vec{\nu}_{ij}) < 0. \end{cases} \quad (4.29)$$

It is straightforward to check that this two-dimensional extension of method (C) still preserves the maximum principle for the mass fraction. Also, transforming a multi-dimensional Euler code into a multi-dimensional multi-component code using method (C) is very easy and cheap.

## 5. NUMERICAL RESULTS

Several numerical experiments, involving the first- or second-order accurate schemes of type (A), (B) and (C) have been done by D. Chargy, R. Abgrall, L. Fezoui and the author, and detailed comparisons are presented in [5]. We simply illustrate here the main features of methods (A), (B) and (C) by showing one numerical example.



We consider the well-known shock tube of Sod [25], with two different species. More precisely, we solve the Riemann problem (3.1), with:

$$\begin{cases} \rho_L = 1, & \rho_R = 0.125, \\ u_L = 0, & u_R = 0, \\ p_L = 1, & p_R = 0.1, \\ Y_L = 1, & Y_R = 0, \end{cases} \quad (5.1)$$

with  $\gamma_1 = 1.4$ ,  $\gamma_2 = 1.2$  and  $C_{v1} = C_{v2}$ ). We have used 101 equally spaced mesh points in the interval  $[-0.5, 0.5]$ , and  $\Delta t$  is chosen at each time level according to the classical CFL condition:

$$\max_i (|u_i + c_i|, |u_i|, |u_i - c_i|) \frac{\Delta t}{\Delta x} = CFL \leq 1 \quad (5.2)$$

(we take  $CFL = 0.75$ ).

Figure 5 shows the mass fraction profiles obtained at time  $t = 0.21$  using the first-order accurate explicit Roe schemes of types (A), (B) and (C). Clearly the results of methods (B) and (C) are far superior to those of approach (A), which gives very bad mass fraction values. Moreover, the computed mass fractions are slightly better with approach (C) than with approach (B): notice indeed that values of  $Y$  which are slightly out of the interval  $[0, 1]$  appear in case (B), where  $\max_i Y_i^n = 1 + 1.5 \cdot 10^{-6}$  (the error in this value is small but significant, since the computation has been made using double-precision real numbers; in other situations, negative values of  $Y$  of the order of  $10^{-2}$  appear when schemes of type (B) are used). Therefore, schemes of type (C), which preserve the maximum principle, actually give more accurate results for the mass fraction.

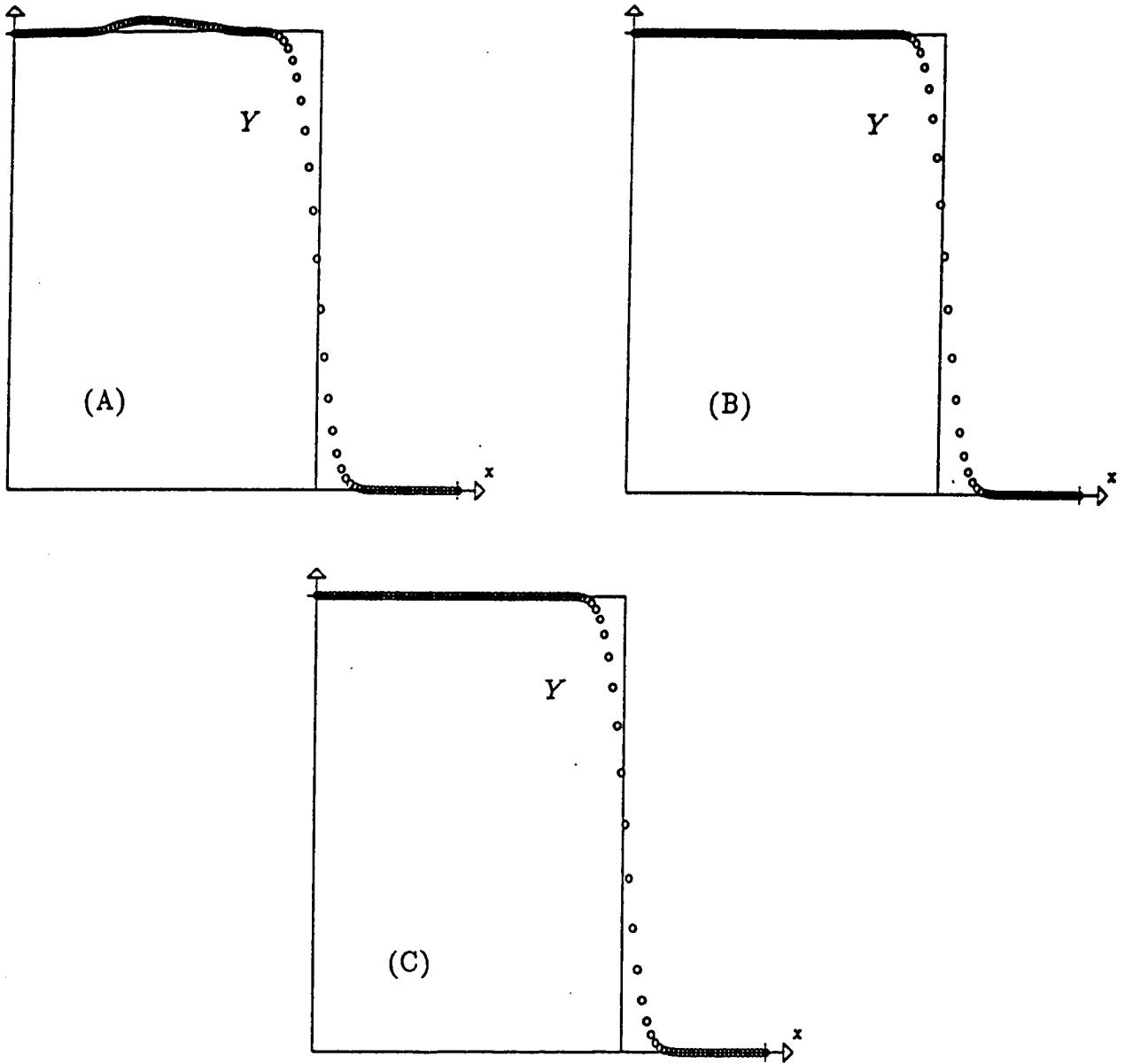
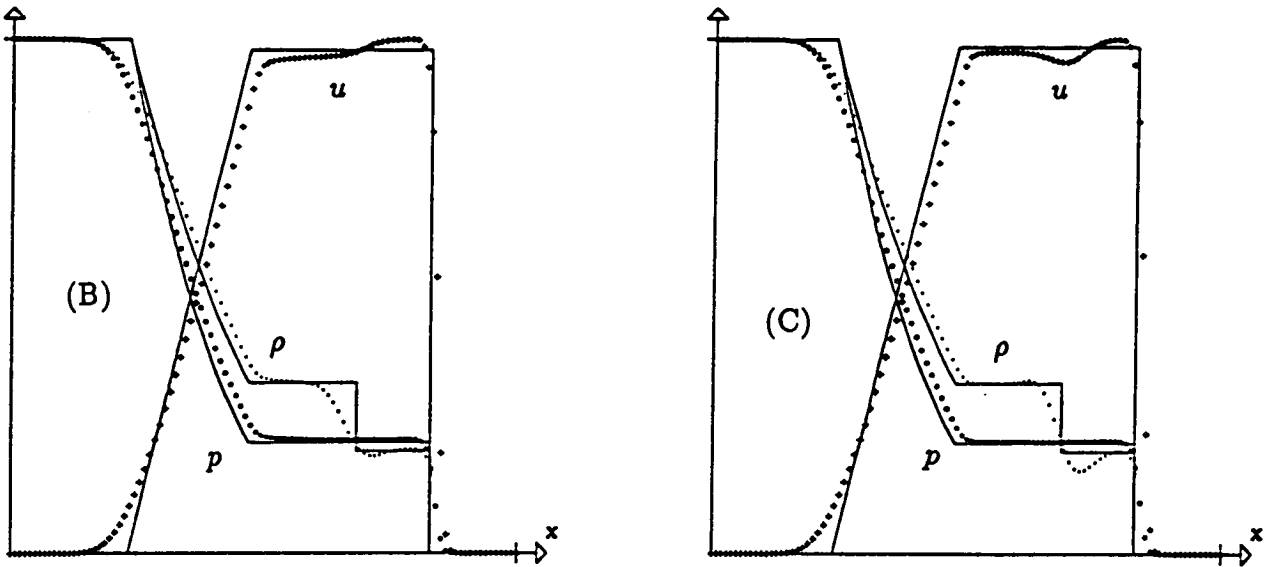


Figure 5: Mass fraction profiles: computed and exact solutions.

On the other hand, it appears on Figure 6 that the values of the hydrodynamical variables  $\rho, u, p$  are slightly less accurate in case (C) than in case (B). This could be expected since approach (C) uses a frozen  $\gamma$  for evaluating the Euler fluxes  $\phi^1, \phi^2, \phi^3$ , whereas the fully coupled approach (B) takes into account the variations of  $\gamma$  (the partial derivatives  $\frac{\partial \gamma}{\partial W^i}$  appear in the Roe matrix  $\bar{A}$  in (2.8)). However, we need to add here that this drawback of method (C) with respect to method (B) almost completely vanishes when the Osher scheme is used instead of the Roe scheme (see [5]).



**Figure 6:** Density, pressure and velocity profiles: computed and exact solutions.

To conclude these comparisons, we can already say that, clearly, method (C) should be in any case preferred to approach (A). The choice between methods (B) and (C) is less simple, unless the specific heat ratio  $\gamma$  is constant (which happens if  $\gamma_1 = \gamma_2$  in (1.3), or if  $Y$  is a variable which does not affect  $\gamma$ ; see Remark 2). Indeed, when  $\gamma$  is constant, it is readily seen that the Euler fluxes  $\phi^1, \phi^2, \phi^3$  are identical with all methods (A), (B) and (C). Then, approach (C), which is cheaper and preserves the maximum principle for the mass fractions, is clearly better than method (B). When  $\gamma$  is not constant, the competition remains open: schemes of type (C) present a drawback (less accurate values of the Euler variables), and two advantages (more accurate and in particular positive mass fraction values, smaller computational cost). One may think then that method

(C) should be used when the variations of  $\gamma$  are “not too important”, whereas approach (B) is preferable when  $\gamma$  has large variations.

## 6. CONCLUSIONS

We have proposed a simple and new way of approximating the convective terms of the species equations in compressible multi-component flows. The main idea in this new approach is to use the same discrete mass fluxes for the species equation as for the continuity equation. This new approximation has over the existing schemes the important advantage of preserving the maximum principle, and in particular the positivity of all mass fractions. This advantage may reveal useful for many applications (see the results in [5] and [21]), and in particular for reactive flows where the nonlinear chemical source terms precisely involve the mass fractions and where the appearance of negative mass fractions very often leads to nonlinear numerical instabilities.

## APPENDIX

As already said in Remark 12, we examine in this Appendix the extension of approach (C) to fully second-order accurate (in both space and time) schemes.

In comparison with Section 4.1, the temporal second-order accuracy will now be obtained by evaluating the fluxes at the half time step  $n + 1/2$  (see e.g. [11], [27]). Thus, the integration from time  $t^n$  to  $t^{n+1}$  is done using the five successive steps described below (the steps (a) and (b) are the same as in Section 4.1):

(a) At each time step, starting from the values  $W_i^n$ , one first evaluates slopes  $s_i^n$  for all variables which are chosen to be piecewise linear. Again, we take here  $Y$  (and not  $\rho Y$ ) as a piecewise linear variable.

(b) Slope limiters are then used in order to avoid the creation of new extrema; in particular (4.1) still holds:

$$\min_{|j-i|\leq 1} Y_j^n \leq Y_i^n \pm \frac{\Delta x}{2} s_i^n \leq \max_{|j-i|\leq 1} Y_j^n . \quad (A.1)$$

(c) The limited slopes are used to evaluate cell-interface values  $W_{i+1/2,\pm}^n$ , and the solution is advanced in time over a half time step using a centered predictor, setting:

$$W_i^{n+1/2} = W_i^n - \frac{\lambda}{2} \left[ F(W_{i+1/2,-}^n) - F(W_{i-1/2,+}^n) \right] . \quad (A.2)$$

(d) Next, we again use the same slopes  $s_i^n$  to evaluate cell interface values  $W_{i+1/2,\pm}^{n+1/2}$  at the half time step (in particular, we set  $Y_{i+1/2,-}^{n+1/2} = Y_i^{n+1/2} + \frac{\Delta x}{2} s_i^n$ ).

(e) Lastly, we complete the time step by evaluating  $W_i^{n+1}$ , using the upwind numerical flux function  $\Phi$ :

$$\frac{W_i^{n+1} - W_i^n}{\Delta t} + \frac{\phi_{i+1/2}^{n+1/2} - \phi_{i-1/2}^{n+1/2}}{\Delta x} = 0, \quad (\text{A.3})$$

where:

$$\phi_{i+1/2}^{n+1/2} = \Phi(W_{i+1/2,-}^{n+1/2}, W_{i+1/2,+}^{n+1/2}). \quad (\text{A.4})$$

Let us use this construction (a)-(e) with a numerical flux function of type (C), that is with the fourth component  $\phi^4$  of the flux evaluated using (3.8), and examine whether the maximum principle is preserved for the mass fraction under the following CFL-like condition (which now replaces (4.3)):

$$\frac{\Delta t}{\Delta x} \left[ \frac{\max(\phi_{i+1/2}^{1,n+1/2}, 0)}{\rho_i^n} - \frac{\min(\phi_{i+1/2}^{1,n+1/2}, 0)}{\rho_{i+1}^n} \right] \leq \frac{1}{2} \quad \forall i. \quad (\text{A.5})$$

We will assume that  $\phi_{i+1/2}^{1,n+1/2} > 0$  and  $\phi_{i-1/2}^{1,n+1/2} > 0$ . Then, we have  $\phi_{i\pm 1/2}^{4,n+1/2} = \phi_{i\pm 1/2}^{1,n+1/2} Y_{i\pm 1/2,-}^{n+1/2}$ , and we get:

$$Y_i^{n+1} = \frac{(\rho_i^n - \lambda \phi_{i+1/2}^{1,n+1/2}) \hat{Y} + \lambda \phi_{i-1/2}^{1,n+1/2} Y_{i-1/2,-}^{n+1/2}}{(\rho_i^n - \lambda \phi_{i+1/2}^{1,n+1/2}) + \lambda \phi_{i-1/2}^{1,n+1/2}}, \quad (\text{A.6})$$

where we now have set:

$$\hat{Y} = \frac{\rho_i^n Y_i^n - \lambda \phi_{i+1/2}^{1,n+1/2} Y_{i+1/2,-}^{n+1/2}}{\rho_i^n - \lambda \phi_{i+1/2}^{1,n+1/2}}. \quad (\text{A.7})$$

Let again  $a$  and  $b$  be two real numbers such that  $a \leq Y_j^n \leq b$  for all  $j$ : we have to examine if both values  $Y_{i-1/2,-}^{n+1/2}$  and  $\hat{Y}$  are in the interval  $[a, b]$ .

Let us introduce some notations. We set:

$$X_{i+1/2} = \lambda \frac{\phi_{i+1/2}^{1,n+1/2}}{\rho_i^n}, \quad (\text{A.8})$$

and:

$$Z_{i+1/2} = \lambda \frac{(\rho u)_{i+1/2,-}^n}{\rho_i^n}, \quad Z_{i-1/2} = \lambda \frac{(\rho u)_{i-1/2,+}^n}{\rho_i^n}. \quad (\text{A.9})$$

Then, we get:

$$\rho_i^{n+1/2} = \rho_i^n - \frac{1}{2} \rho_i^n (Z_{i+1/2} - Z_{i-1/2}), \quad (\text{A.10})$$

$$\rho_i^{n+1/2} Y_i^{n+1/2} = \rho_i^n Y_i^n - \frac{1}{2} \rho_i^n (Z_{i+1/2} Y_{i+1/2,-}^n - Z_{i-1/2} Y_{i-1/2,+}^n), \quad (\text{A.11})$$

and it is easy to check that:

$$Y_i^{n+1/2} = Y_i^n - \frac{\Delta x}{4} s_i^n \frac{Z_{i+1/2} + Z_{i-1/2}}{1 - \frac{1}{2}(Z_{i+1/2} - Z_{i-1/2})}, \quad (\text{A.12})$$

$$Y_{i+1/2,-}^{n+1/2} = Y_i^n + \frac{\Delta x}{2} s_i^n \left[ \frac{1 - Z_{i+1/2}}{1 - \frac{1}{2}(Z_{i+1/2} - Z_{i-1/2})} \right], \quad (\text{A.13})$$

$$\hat{Y} = Y_i^n - \frac{\Delta x}{2} s_i^n \left[ \frac{1 - Z_{i+1/2}}{1 - \frac{1}{2}(Z_{i+1/2} - Z_{i-1/2})} \right] \cdot \left( \frac{X_{i+1/2}}{1 - X_{i+1/2}} \right). \quad (\text{A.14})$$

We know from (A.5) and our assumptions on the sign of  $\phi_{i+1/2}^{1,n+1/2}$  that  $0 \leq X_{i+1/2} \leq \frac{1}{2}$ . Thus the term in parentheses in (A.14) is in the interval  $[0, 1]$ . If moreover we also have:

$$0 \leq Z_{j\pm 1/2} \leq \frac{1}{2} \quad \text{for all } j, \quad (\text{A.15})$$

then it is easy to conclude from (A.13), (A.14) and (A.1) that the term in brackets in (A.13)-(A.14) also is in the interval  $[0, 1]$ , whence  $Y_{i+1/2,-}^{n+1/2} \in [a, b]$  and  $\hat{Y} \in [a, b]$ . But these conclusions do not hold any longer if  $Z_{i-1/2} + Z_{i-1/2} < 0$  (the term in brackets in (A.13)-(A.14) then exceeds 1). Of course, since we have assumed that  $\phi_{i+1/2}^{1,n+1/2} > 0$  and  $\phi_{i-1/2}^{1,n+1/2} > 0$ , the quantities  $Z_{i+1/2}$  and  $Z_{i-1/2}$  are likely to be positive. Therefore, we can only conclude (but this conclusion is confirmed by the numerical experiments; see [5]) that this second-order accurate scheme preserves *in general* the maximum principle for the mass fraction.

### ACKNOWLEDGEMENTS

I thank my colleagues R. Abgrall, D. Chagy and L. Fezoui for their help in this work.

## REFERENCES

- [1] R. ABGRALL, "Généralisation du schéma de Roe pour le calcul d'écoulements de mélanges de gaz à concentrations variables", *La Recherche Aéronautique*, pp. 31-43, 6, (1988).
- [2] 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", *La Recherche Aéronautique*, to appear.
- [3] 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, Dordrecht, (1988).
- [4] G. V. CANDLER & R. W. McCORMACK, "The computation of hypersonic flows in chemical and thermal nonequilibrium", Paper N° 107, Third National Aerospace Plane Technology Symposium, (1987).
- [5] D. CHARGY, R. ABGRALL, L. FEZOUÏ & B. LARROUTUROU, "Comparisons of several numerical schemes for multi-component one-dimensional flows", INRIA Report, to appear.
- [6] D. CHARGY, A. DERVIEUX & B. LARROUTUROU, "Upwind adaptive finite-element investigations of two-dimensional transonic reactive flows", *Int. J. Num. Meth. Fluids*, to appear.
- [7] 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).
- [8] J. A. DESIDERI, N. GLINSKY & E. HETTENA, "Hypersonic reactive flow computations", *Computers and Fluids*, to appear.
- [9] G. FERNANDEZ & H. GUILLARD, "Implicit schemes for subsonic combustion problems", "Numerical combustion", Dervieux Larroutrou eds., *Lecture Notes in Physics*, pp. 277-286, Springer Verlag, to appear.
- [10] G. FERNANDEZ & B. LARROUTUROU, "Hyperbolic schemes for multi-component Euler equations", *Nonlinear hyperbolic equations - Theory, numerical methods and applications*, Ballmann & Jeltsch eds., pp. 128-138, *Notes on Numerical Fluid Mechanics*, 24, Vieweg, Braunschweig, (1989).

[11] L. FEZOUÏ, "Résolution des équations d'Euler par un schéma de Van Leer en éléments finis", INRIA Report 358, (1985).

[12] L. FEZOUÏ & B. STOUFFLET, "A class of implicit upwind schemes for Euler simulation with unstructured meshes", to appear in J. Comp. Phys..

[13] 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).

[14] B. GROSSMANN & P. CINNELLA, "Upwind methods for flows with non-equilibrium chemistry and thermodynamics", "Numerical combustion", Dervieux Larrourou eds., Lecture Notes in Physics, pp. 323-337, Springer Verlag, to appear.

[15] A. HABBAL, A. DERVIEUX, H. GUILLARD & B. LARROUTUROU, "Explicit calculations of reactive flows with an upwind finite-element hydrodynamical code", INRIA Report 690, (1987).

[16] 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).

[17] K. KAILASANATH, E. S. ORAN & J. P. BORIS, "Numerical simulations of flames and detonations", "Numerical combustion", Dervieux Larrourou eds., Lecture Notes in Physics, pp. 82-97, Springer Verlag, to appear.

[18] B. LARROUTUROU, "Introduction to mathematical and numerical modelling in gaseous combustion", Springer Series in Computational Physics, to appear.

[19] B. LARROUTUROU & L. FEZOUÏ, "On the equations of multi-component perfect or real gas inviscid flow", "Non linear hyperbolic problems", Carasso Charrier Hanouzet Joly eds., Lecture Notes in Mathematics, Springer Verlag, Heidelberg, (1989).

[20] 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).

[21] B. N'KONGA, H. GUILLARD & B. LARROUTUROU, *in preparation*.

[22] S. OSHER & F. SOLOMON, "Upwind schemes for hyperbolic systems of conservation laws", Math. Comp., **38**, (158), pp. 339-374, (1982).

[23] W. RODI, "Turbulence models and their applications in hydraulics", AIRH, (1984).

[24] P. L. ROE, "Approximate Riemann solvers, parameter vectors and difference



schemes", J. Comp. Phys., **43**, pp. 357-372, (1981).

[25] G. A. SOD, "A survey of several finite-difference methods for systems of non-linear hyperbolic conservation laws", J. Comp. Phys., **27**, pp. 1-31, (1977).

[26] 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).

[27] 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).

[28] 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).

[29] R. S. VARGA, "Matrix iterative analysis", Prentice Hall, (1962).

2

3

4

5

6

7

8