INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2006; **50**:1029–1061
Published online 15 November 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/fld.1104

Second-order entropy diminishing scheme for the Euler equations

F. Coquel[1], P. Helluy[2,*,†] and J. Schneider[2]

[1]*Laboratoire d'Analyse Numérique, Université Paris VI, France*
[2]*Laboratoire ANAM/MNC, Université de Toulon, France*

SUMMARY

In several papers of Bouchut *et al*., Coquel and Le Floch (*Math. Comput.* 1996; **65**(216):1439–1461; *Numer. Math.* 1996; **74**(01):1–34), a general methodology has been developed to construct second-order finite volume schemes for hyperbolic systems of conservation laws satisfying the entire family of entropy inequalities. This approach is mainly based on the construction of an *entropy diminishing projection*. Unfortunately, the explicit computation of this projection is not always easy. In the first part of this paper, we carry out this computation in the important case of the Euler equations of gas dynamics. In the second part, we present several numerical applications of the projection in the context of finite volume schemes. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: Godunov scheme; entropy; second order

0. INTRODUCTION

In this paper, a methodology to build second-order generalized Godunov schemes satisfying all the entropy inequalities is described. Our goal is to adapt the Godunov's original idea, which leads to first-order schemes, in order to obtain a second-order scheme. The construction is done completely in the case of the Euler equations.

Let us recall that in the classical first-order scheme, each time-step of the time-marching procedure is made up of two stages:

1. Exact (or high order) resolution of the system of conservation laws starting from the cell averages of the previous time-step. This resolution is performed during a short time τ so that the Riemann problems of the two sides of each cell do not interact.

*Correspondence to: P. Helluy, Laboratoire ANAM/MNC, ISITV, BP 56, 83162 La Valette CEDEX, France.
†E-mail: helluy@univ-tln.fr

Received 26 November 2002
Revised 12 January 2005
Accepted 18 August 2005

2. The result of the previous stage is no longer constant in the cells. A projection is then realized in order to recover a piecewise constant approximation. For a first-order scheme, the only conservative projection is cell averaging.

This description is of course theoretical. In practice, a numerical flux can be associated to this procedure so that the two stages become transparent in the implementation of the scheme.

The first scheme designed in this way is described by Godunov in Reference [1]. We shall call in the sequel this kind of scheme, based on an exact or approximate Riemann solver, a *Godunov scheme*. This terminology is developed in the review paper of Harten *et al.* [2]. Most of the conservative schemes can be put in this framework as the HLL scheme [2], HLLC scheme of Toro *et al.* [3], Roe scheme [4], Engquist–Osher scheme [5], kinetic schemes [6–8], and many others.

The original Godunov scheme has the important property that it respects on the discrete level the decrease of all the Lax entropies of the approximated hyperbolic system. This is an important property which is closely linked to the stability of the scheme and with the possibility of non-physical waves if the scheme is not entropic. For example, the HLL, HLLC and Engquist–Osher schemes are entropic whereas the original Roe scheme is not entropic. With a simple modification, it can be made entropic (for example, see the fix of Harten and Hyman [9]).

Another important feature emerging from the construction of the Godunov schemes is that the cell values are high-order approximations of the exact mean values. They are also second-order approximations of the exact values in the centre of the cells. Unfortunately, because they are only first-order approximations on the cell sides, an error of order one is committed in the computation of the numerical flux.

These remarks have been used to improve the precision of Godunov schemes. The most widespread improvement consists of reconstructing a more precise approximation of the solution by using cell averages and Taylor expansions. This reconstruction allows one to compute more accurately the fluxes, but has to be corrected by a limitation procedure in order to avoid oscillations or non-physical values.

The main criterion in these methods is a total variation diminishing (TVD) criterion. Many works deal with second-order Godunov schemes (we can cite for example the works of Van Leer [10] or Harten [11]) which give good numerical results in various practical computations. Anyway, since the work of Rauch [12] it is well known that the TVD criterions are inadequate in the theoretical study of systems of conservation laws in higher dimensions. There is no hope to prove convergence results in a general framework with these limitations.[‡]

In this paper, another approach is followed, which is closer to the Godunov's original idea. On the one hand, we shall consider the resolution of a generalized Riemann problem with piecewise linear (instead of constant) initial data. The solution will then have to be projected back onto a set of piecewise linear functions. On the other hand, in the projection step, we substitute for the TVD criterion a mean entropy decreasing criterion which seems more adequate for systems. An important feature is that this *stability criterion will be verified for the whole family of entropies of the system*. Such an approach has been initiated for scalar conservation laws by Bouchut *et al.* in Reference [13] and has been extended to 2×2 systems

[‡]It should also be noticed that the TVD criterion is limited to cartesian grids in the multi-d case, even for scalar equations.

of conservation laws and to the 3×3 system of Lagrangian gas dynamics by Coquel and Le Floch in Reference [14]. Our paper deals with the case of the Euler system.

The first part is devoted to a mathematical study of the projection stage. The problem can be stated in the following way: the three conservative variables, after the Riemann problem step are, generally speaking, piecewise regular functions in each cell of the finite volume method. The goal of the projection is to recover linear variables in the cells in order to pursue the resolution. Because the mean values of these variables are given by the conservation property, it is sufficient to compute three slopes. The projection can then be seen as the operator which gives these slopes from the original piecewise regular conservatives variables in the cells. It is known that this operator cannot be linear. Indeed, it can be proved that a general linear projection (such as the classical L^2 projection) will not always respect the positivity of the projected density and pressure. It is of course linked to the fact that there is no second-order linear three point scheme which is also TVD for the scalar conservation laws (see Reference [15]). Because we are working on systems we will require that the projection operator is entropic. By entropic, we mean that the mean value in each cell of the entropy of the projection is smaller than the mean entropy of the original conservative variables. It is very interesting that *this entropy diminishing property gives also the positivity of the projected variables*. This result is proved in Remark 2.

Here, we are able to prove the existence of a non-linear projection for the Euler equations and provide explicit formulas. The construction is based on several ingredients:

- First, we recall the theory of second-order entropic projections of Reference [13]. This theory is based on the definition of an approximate derivative (Definition 5) and a sufficient condition in order to have an entropic projection (Theorem 1). A very nice feature of this theory is that, when applied to a scalar conservation law, the entropic projection is very similar to a classical minmod limiter.
- This sufficient condition is not exploitable as is for the Euler system, and it does not give easily an explicit formula for the slopes of the projection. Thus we propose to seek the projection as the composition of two non-linear operators. To build the first operator, we work with special variables (density, momentum and a particular entropy) for which the sufficient condition of Theorem 1 can be computed. We then reduce the first step of the projection to the solution of *a triangular set of inequalities* (16)–(18) for the three slopes.
- The first operator is not conservative for the energy. The mean value of the projected energy has thus to be corrected. We prove that the correction is still entropy diminishing and thus does not impede the whole process. This fact can also be used by to built simple entropic schemes for the Euler equations based on the intermediate solution of the entropy conservation law.
- Finally, we provide formulas that explicitly solve the fundamental triangular set of inequalities for the slopes. These formulas are summed up in Theorem 2.

The second part of the paper is then devoted to several numerical experiments with the previously constructed projection. One approach could have been to develop a generalized Riemann solver as in Reference [14] for the gas dynamics equations in Lagrangian coordinates. The theory of the generalized Riemann problem can be found, for example, in the papers of Bourgeade *et al.* [16], Ben Artzi and Falcovitz [17]. It is also sketched in the book of Godlewski and Raviart [15]. The problem is that the implementation of this solver is very

complex. So we prefer to follow here much simpler approaches. We present two kinds of numerical results:

- The first are obtained with a second-order kinetic scheme. We use the Boltzmann kinetic interpretation of the Euler equations in order to construct an approximate second-order Riemann solver. For the computations, we choose a compactly supported Maxwellian proposed by Perthame [7]. But instead of taking a piecewise constant density function in the microscopic scheme, we take a piecewise linear density function. The free transport Boltzmann equation can then be solved exactly. A second-order approximate Riemann solver is then obtained by taking the moments of the resulting microscopic solution. After the Boltzmann step, the solution is piecewise polynomial in each cell and can be computed explicitly. We are then in a position to apply the results of the first part of the paper and provide some numerical experiments which validate the whole procedure.
- The previous construction is quite complicated. So we propose also another approach. We do not try to construct a high order Riemann solver but reconstruct a high order approximation of the solution from its cell averages. The cell averages are computed with a standard Godunov flux. Without limitations, the scheme would present oscillations. We then apply the entropic projection to the reconstruction in order to evaluate the damping of the oscillations.

We conclude the paper with some comments about possible extensions and applications of the entropic projection.

1. ENTROPY SOLUTION OF EULER EQUATIONS

1.1. Euler equations

In the present paper, we focus our attention on the numerical approximation of the discontinuous solutions of the Euler system for polytropic gases. With standard notation, this system reads:

$$\begin{aligned} \partial_t w + \partial_x f(w) = 0, \quad t > 0, \ \ x \in R \\ w(0,x) = w_0(x) \end{aligned} \tag{1}$$

where

$$w = \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}, \quad f(w) = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ (E+p)u \end{pmatrix}$$

and

$$E = \rho\varepsilon + \frac{(\rho u)^2}{2\rho}$$

$$p = (\gamma - 1)\rho\varepsilon, \quad \gamma > 1$$

It is well known that this system is strictly hyperbolic over the phase space

$$\Omega = \left\{ w \in R^3, \rho > 0, \rho u \in R, E - \frac{(\rho u)^2}{2\rho} > 0 \right\}$$

1.2. Entropy condition

Generally, the weak solution of (1) is not unique. Assuming that theoretical results for scalar conservation laws extend to systems, an entropy condition has to be added to the Euler system in order to recover uniqueness.

Definition 1
Let $U(w)$ be a convex function of w, and let $F(w)$ be a function such that the following additional conservation law holds whenever w is a strong solution

$$\partial_t U(w) + \partial_x F(w) = 0$$

(U,F) is then called a Lax entropy pair for the system (1) (U is the entropy and F the entropy flux).

Definition 2
A weak solution $w(t,x)$ of (1) is an entropy solution if and only if for every entropy pair (U,F), the following inequality holds:

$$\partial_t U(w) + \partial_x F(w) \leqslant 0, \quad t > 0, \ x \in R$$

The previous notions have been introduced by Lax in Reference [18] for general systems of conservation laws. The mathematical existence and uniqueness of the entropy solution is still an open problem. In this paper this well-posedness result is supposed to hold.

The practical computation of the Lax entropies for the Euler system is given for example (among many others) in the PhD thesis of Croisille [19] or in the book of Raviart and Godlewski [20]. It appears that a family of regular entropies can be constructed in the following way: let us introduce the following quantity:

$$S = (\rho\varepsilon)^{1/\gamma}$$

It can be checked that $(-S, -uS)$ is a Lax entropy pair of the Euler system. We now consider the family (U, uU) defined by

$$U = \rho G\left(\frac{S}{\rho}\right)$$

where G is a C^2 function on R^{+*} such that

$$G' < 0 \quad \text{and} \quad G'' > 0$$

It is proved in Reference [19] that this construction gives all the C^2 entropies of the Euler system of the form (U,F) with $F = uU$. Another expression of the entropies is given by

$$U = \rho H\left(\frac{\rho}{S}\right)$$

where H is a C^2 function on R^{+*} such that

$$H'(x) + xH''(x) > 0, \quad x \in R^{+*}$$

2. GENERALIZED GODUNOV SCHEME

The framework of generalized second-order Godunov schemes approximating (1) can now be stated.

A time-step $\tau > 0$, and a space-step $h > 0$ being given, we consider the following discretization in time $t_n = n\tau$ ($n \geqslant 0$) and space $x_i = ih$ ($i \in \mathbf{Z}$). The cell i is defined by $C_i =]x_{i-1/2}, x_{i+1/2}[$.

We start with an approximation of $w(t_n, x)$ at time t_n

$$w^n(x) \approx w(t_n, x)$$

which is supposed to be piecewise linear with

$$w^n(x) = w_i^n + s_i^n(x - x_i), \quad \forall x \in C_i$$

(in the classical Godunov scheme $s_n^i = 0$).

Remark 1
To be more general, we can also consider another set of variables given by a regular transformation:

$$\phi = \phi(w), \quad w \in \Omega$$

and suppose that ϕ is piecewise linear:

$$w^n(x) = \phi^{-1}(\phi_i^n + d_i^n(x - x_i)) \quad \forall x \in C_i$$

This approach will be developed in the next part.

In order to compute a new $w^{n+1}(x)$, two steps are then performed:

- First, the Euler equations and the entropy conditions are exactly solved during a short period of time τ:

$$\partial_t v + \partial_x f(v) = 0 \quad t > 0, \quad x \in R$$
$$\partial_t U(v) + \partial_x F(v) \leqslant 0$$
$$v(0, x) = w^n(x)$$

  and this is done for all Lax entropy pairs (U, F). The solution $v(\tau, x)$ will be denoted by $w^{n+1,-}(x)$ and is generally no longer linear in each cell.
- $w^{n+1,-}(x)$ has to be approximated again by a piecewise linear function. Let P_h^1 be the space of piecewise (per cell) linear functions, then we look for a (possibly non-linear) operator $\Pi : L^\infty(R) \to P_h^1$ such that

$$w^{n+1}(x) = \Pi\, w^{n+1,-}(x)$$

  Π is also supposed to be a projection in the sense that $\Pi w = w$ whenever $w \in P_h^1$.

In the case of the classical first-order Godunov scheme the projection consists simply of cell-averaging

$$\Pi w(x) = \bar{w} = \frac{1}{h} \int_{C_i} w, \quad x \in C_i, \quad i \in Z$$

If the projection is an approximation of order 2 (with respect to h) of $w^{n+1,-}$, then the resulting scheme is of order 2 in space (and of order 1 in time) in the sense of consistency. But good precision will be achieved only under some stability conditions. We shall now describe a possible set of stability conditions.

3. ENTROPY DIMINISHING PROJECTION

As mentioned in the introduction, we wish to generalize the Godunov idea to second-order schemes. The first step consists of solving exactly (or at least with given precision) generalized Riemann problems between two cells starting from a piecewise linear approximation of the solution. Actually, in order to be more general, we shall consider in each cell the case of an initial condition in a finite dimensional manifold M of regular functions defined on the cell. It is clear that after the exact resolution step, the solution is, generally speaking, no longer in M. A regular approximation in the cell has to be recovered, with the decrease of entropies.

A second-order Riemann solver will be constructed later on, and in this part we focus only on the projection step of the Godunov method.

3.1. Second-order entropy diminishing projections

Because we are looking for a projection defined locally, i.e. on each cell C_i, we can suppose, for simplicity, that $C_i =]0, 1[$. The projection problem can then be presented in the following abstract setting (introduced in Reference [13, 14]):

Definition 3
Let M be a finite dimension manifold included in $C^\infty([0,1], \Omega)$. Let $w = (\rho, \rho u, E)^{\mathrm{T}}$ be an element of $L^\infty \cap BV([0,1], \Omega)$, and let Π be a (generally non-linear) projection from $L^\infty \cap BV([0,1], \Omega)$ into M which satisfies the following property: for all Lax entropy pairs (U, F),

$$\int_0^1 U(\Pi w(x))\,\mathrm{d}x \leqslant \int_0^1 U(w(x))\,\mathrm{d}x \tag{2}$$

Such an operator will be denoted as an entropy diminishing projection on M.

Let us note that condition (2) implies a conservation property. Indeed, when applied to the following degenerate entropies:

$$U(w) = \pm \rho, \quad U(w) = \pm \rho u, \quad U(w) = \pm E$$

we get

$$\bar{w} := \int_0^1 \Pi w(x)\,\mathrm{d}x = \int_0^1 w(x)\,\mathrm{d}x$$

In the classical Godunov method, M is the set of constant states which are in Ω and the projection is given by

$$\Pi w = \bar{w}$$

Inequality (2) then holds thanks to the Jensen inequality. The problem here is that we are looking for an approximation of the function w which is of order two when rescaled to an interval of size h. Let us define precisely this second-order property.

Definition 4
Let Π be an entropy diminishing projection from $L^\infty \cap BV([0,1],\Omega)$ into M. Let $\tilde{w}$ be an element of $C^2([0,h],\Omega)$ and $w(x) = \tilde{w}(xh),\ x \in [0,1]$. After projection of w we can define

$$\tilde{\Pi}\tilde{w}(t) := \Pi w\left(\frac{t}{h}\right), \quad t \in [0,h]$$

then Π is said to be a second-order entropy diminishing projection if

$$\left|\tilde{\Pi}\tilde{w}(t) - \tilde{w}(t)\right| \leqslant \lambda h^2 \tag{3}$$

where λ is a constant which depends only on the C^2 norm of $\tilde{w}$.

As we have seen before, in the case of first-order Godunov scheme, the first-order entropy diminishing projection is necessarily unique. On the other hand, there exist many second-order entropy diminishing projections with various choices of manifolds M. In the sequel, we recall a general framework which allows the construction of second-order entropy diminishing projections. We then use special features of the Euler equations to build a simple explicit projection for this hyperbolic system.

3.2. Pseudo-derivative

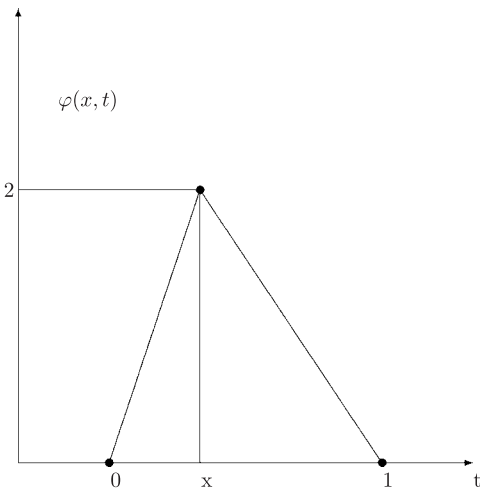
We are going to derive a sufficient condition on Π in order to satisfy the very important inequality (2). It is based on the following notion of (what we decided to call) pseudo-derivative which was introduced first by Bouchut *et al.* in Reference [13] in the case of a scalar equation. It was also studied by Coquel and Le Floch in Reference [14] for systems.

Definition 5
Let w be an element of $L^\infty \cap BV([0,1],\Omega)$. The *pseudo-derivative of* w is the continuous function $N(w) \in C(]0,1[,R^3)$ defined by

$$N(w)(x) = \frac{2}{1-x}\int_x^1 w(t)\,\mathrm{d}t - \frac{2}{x}\int_0^x w(t)\,\mathrm{d}t$$

The linear operator N satisfies the following properties:

- For all constants C, $N(w+C)=N(w)$.
- If w is regular enough, $N(w)(x)=\int_0^1 \varphi(x,t)w'(t)\,\mathrm{d}t$. Where the graph of $t\to\varphi(x,t)$ is given below



- If w is regular, for all $x\in]0,1[$ there is a $\theta(x)\in]0,1[$ such that

$$N(w)(x)=w'(\theta(x)) \tag{4}$$

- If $\bar{w}=\int_0^1 w(t)\,\mathrm{d}t=0$ then

$$-\frac{x(1-x)}{2}N(w)(x)=\int_0^x w(t)\,\mathrm{d}t$$

- Finally, if $\bar{w}=\int_0^1 w(t)\,\mathrm{d}t=0$, the following integration by parts formula holds:

$$\int_0^1 v(t)w(t)\,\mathrm{d}t=\int_0^1 \frac{t(1-t)}{2}N(w)(t)v'(t)\,\mathrm{d}t$$

A sufficient condition for the decrease of the mean entropy can then be stated.

Theorem 1
Let Π be a projection from $L^\infty\cap BV([0,1],\Omega)$ into M. If for all $w\in L^\infty\cap BV([0,1],\Omega)$, all $x\in]0,1[$, and all entropies U,

$$d_x\Pi w\cdot U''(\Pi w)(x)(N(w)(x)-d_x\Pi w)\geqslant 0 \tag{5}$$

then, Π is an entropy diminishing projection.

Proof 1
This is a consequence of the basic convex inequality

$$\begin{aligned}\int_0^1 U(w) - U(\Pi w) &\geqslant \int_0^1 U'(\Pi w)(w - \Pi w) \\ &= \int_0^1 U'(\Pi w)(\bar{w} - \overline{\Pi w}) \\ &\quad + \int_0^1 \frac{t(1-t)}{2} \mathrm{d}_x \Pi w \cdot U''(\Pi w)(t)(N(w)(t) - d_x \Pi w)\,\mathrm{d}t \end{aligned} \tag{6}$$

At this stage, condition (5) would provide us with a practical definition of Πw if we were studying a scalar equation instead of the Euler system (see Reference [13]). Using the convexity of entropies U, the previous condition reduces to

$$d_x \Pi w \cdot (N(w)(x) - d_x \Pi w) \geqslant 0 \tag{7}$$

Define the minmod function of a set of reals as

$$\operatorname{minmod} E = \begin{cases} \inf E & \text{if } E \subset R^+ \\ \sup E & \text{if } E \subset R^- \\ 0 & \text{otherwise} \end{cases}$$

then a possible choice for Πw is the linear function $\Pi w(x) = \bar{w} + d_x \Pi w(x - 1/2)$ where the real number $d_x \Pi w$ is defined by

$$d_x \Pi w = \operatorname{minmod}\{N(w)(x), \quad x \in]0,1[\} \tag{8}$$

This defines a second-order entropy diminishing projection thanks to property (4).

In the case of a system of conservation laws, $U''(\Pi w)(x)$ are positive definite matrices which define a metric for all values of $x \in]0,1[$. If that metric was constant then we may define $d_x \Pi w$ as the vector in the convex envelope N° of $\{N(w)(x),\ x \in]0,1[\}$ that minimizes the norm associated to the metric i.e.

$$d_x \Pi w^{\mathrm{T}} \cdot U'' \cdot d_x \Pi w = \min_{v \in N^\circ} v^{\mathrm{T}} \cdot U'' \cdot v \tag{9}$$

But since the metric is changing with x, the practical use of condition (5) is not at all straightforward. The inequality can however be used to establish the existence of a solution (see Reference [14]) and to derive a family of projections that partially (that is to say up to a certain order n) fulfill it (see for an example Reference [14]).

In the next section, we adapt the previous method in order to build an exact second order entropy diminishing projection in the case of the Euler equations.

4. EXPLICIT COMPUTATION OF THE PROJECTION: EULER CASE

The starting point of the explicit projection procedure is an element

$$w_0 = (\rho_0, q_0 = \rho_0 u_0, E_0)^{\mathrm{T}} \in L^\infty \cap BV([0,1], \Omega)$$

In the sequel, we will denote by Π a second-order entropy diminishing projection. The projection $w_2 = \Pi w_0$ thus belongs to $C^\infty([0,1], \Omega)$. For practical reasons, w_2 will be obtained from an intermediate state, noted $w_1 = (\rho_1, q_1 = \rho_1 u_1, E_1)^{\mathrm{T}} \in C^\infty([0,1], \Omega)$. In other words, $w_1 = \Pi_1 w_0$, and our explicit projection Π is the composition of two operators Π_1 and Π_2:

$$\Pi = \Pi_2 \circ \Pi_1$$

Let us first describe operator Π_1. We have seen in the introduction that all the C^2 entropies of the Euler system have a rather simple form when expressed as a function of the quantity $S = (\rho\varepsilon)^{1/\gamma}$. We thus define

$$S_0 = (E_0 - \tfrac{1}{2}\rho_0 u_0^2)^{1/\gamma}$$

On the other hand, we wish w_1 to be C^∞. A simple possibility is to set

$$S_1 = \overline{S_0} + \mathrm{d}S_1(x - 1/2) \tag{10}$$

$$\rho_1 = \overline{\rho_0} + \mathrm{d}\rho_1(x - 1/2) \tag{11}$$

$$q_1 = \overline{q_0} + \mathrm{d}q_1(x - 1/2) \tag{12}$$

where, as before, $\bar{z}$ is a notation for the mean value of the quantity z on the interval $[0,1]$. The real numbers $\mathrm{d}S_1$, $\mathrm{d}\rho_1$, $\mathrm{d}q_1$ are to be guessed (and will be defined explicitly in the sequel). With our choice, the quantities S_1, ρ_1, q_1 are thus linear in the cell $[0,1]$ and have the same mean values as S_0, ρ_0, q_0.

Setting

$$w_1 = (\rho_1, q_1, E_1 = \tfrac{1}{2}\rho_1 u_1^2 + S_1^\gamma) \tag{13}$$

it is clear that w_1 is regular but not linear in the cell. It is also clear that, since $\overline{S_1} = \overline{S_0}$, the first projection operator Π_1 is not conservative in the sense that, in general, density and impulsion are conserved but not energy. Thus

$$\overline{q_1} = \overline{q_0}, \quad \overline{\rho_1} = \overline{\rho_0}, \quad \overline{E_1} \neq \overline{E_0}$$

This fact justifies the necessity of a correction $w_2 = \Pi_2 w_1$, if this approach is employed, in order to recover the conservation of energy.

The simplest way to obtain w_2 is to correct the energy by taking

$$\begin{aligned} \rho_2 &= \rho_1 \\ q_2 &= \rho_2 u_2 = q_1 \\ E_2 &= E_1 - \overline{E_1} + \overline{E_0} \end{aligned}$$

But with this choice, the operator $w_0 \to w_2$ is not a projection. For this reason, we prefer to take

$$E_2 = \tfrac{1}{2}\rho_1 u_1^2 + (K + S_1)^\gamma \tag{14}$$

where the constant K should be chosen in such a way that

$$\overline{E_2} = \overline{E_0} \tag{15}$$

We suggest then to define the approximation manifold M by:

Definition 6
M is the set of vector functions $w = x \to (\rho(x), q(x), (q(x)^2/2\rho(x)) + S(x)^\gamma)$ defined on $[0,1]$ such that the functions ρ, q, and S are linear and such that $\forall x \in [0,1]$, $w(x) \in \Omega$ (which is equivalent to $\forall x \in [0,1]$, $\rho > 0$, $S > 0$)

In other words, according to the notations of Remark 1, we have simply set $\phi = (\rho, \rho u, S)$.

The problem is now to verify that the correction (14) can be done with the decrease of any entropy U:

$$\int U(\rho_2, q_2, E_2) \leqslant \int U(\rho_1, q_1, E_1)$$

But the convexity of U with respect to the conservative variables gives

$$\int U(\rho_1, q_1, E_1) \geqslant \int U(\rho_2, q_2, E_2) + \int \frac{\partial U}{\partial E}(w_2)(E_1 - E_2)$$

On the other hand, we know that

$$\frac{\partial U}{\partial E}(w) = \rho G'\left(\frac{S}{\rho}\right)\frac{\partial S}{\partial E}$$

and

$$S = \left(E - \frac{q^2}{2\rho}\right)^{1/\gamma} \Rightarrow \frac{\partial S}{\partial E} = \frac{1}{\gamma} S^{1-\gamma} > 0$$

which implies that $\partial U/\partial E < 0$ on the phase space Ω. It is thus natural to require that, $E_1 - E_0 < 0$ on the cell, or, in an equivalent way, that

$$K \geqslant 0$$

Actually, it will be more convenient (and it is possible) to ask a little bit more. We know that the energy is the sum of two terms:

$$E = \frac{q^2}{2\rho} + S^\gamma$$

and we will require the decrease of these two terms separately:

$$\int \frac{q_1^2}{2\rho_1} \leqslant \int \frac{q_0^2}{2\rho_0} \tag{16}$$

and

$$\int C(S_1) \leqslant \int C(S_0) \quad \forall C \in C^2(R,R), \text{ convex} \tag{17}$$

In order that the global process $\Pi = \Pi_2 \circ \Pi_1$ be entropy decreasing, it is finally sufficient to require the last family of inequalities

$$\int \rho_1 H\left(\frac{S_1}{\rho_1}\right) \mathrm{d}x \leqslant \int \rho_0 H\left(\frac{S_0}{\rho_0}\right) \tag{18}$$

where H is any C^2 function on R^{+*} such that

$$H'(x) + xH''(x) > 0, \quad x \in R^{+*}$$

Remark 2
Inequality (17) automatically enforces $S_1 > 0$ on $]0,1[$. Indeed, consider a $C^2(R,R)$ convex function C satisfying

$$C(y) = 0 \quad \text{when} \quad y \geqslant 0$$

$$C(y) > 0 \quad \text{when} \quad y < 0$$

then, because $S_0 > 0$, we have $0 \leqslant \int C(S_1) \leqslant \int C(S_0) = 0$ and then $S_1 > 0$. In the same way, inequality (18) implies that $\rho_1 > 0$ if we suppose that $S_0 > 0$, $S_1 > 0$ and $\rho_0 > 0$. These properties are very important for the numerical approximation of the Euler equations. They ensure that the resulting second-order entropy diminishing scheme is also a positive scheme.

We sum up the previous construction in the following proposition:

Proposition 1
Let $w_0 = (\rho_0, q_0 = \rho_0 u_0, E_0)^{\mathrm{T}} \in L^\infty \cap BV([0,1],\Omega)$, and let $w_1 = (\rho_1, q_1 = \rho_1 u_1, E_1 = \frac{1}{2}\rho_1 u_1^2 + S_1^{\gamma})^{\mathrm{T}} \in C^\infty([0,1],\Omega)$ where ρ_1, q_1, and S_1 are linear functions defined by (10)–(12). Suppose that the slopes of ρ_1, q_1, and S_1 are computed in order to satisfy (16)–(18). Let finally $w_2 = (\rho_1, q_1, E_2)^{\mathrm{T}} \in C^\infty([0,1],\Omega)$, where E_2 is corrected according to (14) and (15). Then, the non-linear operator $\Pi : w_0 \to w_2$ is an entropy diminishing projection on the manifold M of Definition 6.

The construction of an entropy diminishing second-order projection is now reduced to the computation of three slopes $\mathrm{d}S_1$, $\mathrm{d}\rho_1$, and $\mathrm{d}q_1$ satisfying inequalities (16)–(18). The important fact is that we have now to solve a *triangular* set of inequalities. In practice, we will have first to find a $\mathrm{d}S_1$ satisfying (17). Then our choice of S_1 will be inserted in (18) in order to get $\mathrm{d}\rho_1$. Then, knowing S_1 and ρ_1, we are in a position to solve (16) and compute q_1.

The computation of S_1 is quite simple if we apply the computations leading to formula (8). We thus propose

$$\mathrm{d}S_1 = \mathrm{minmod}\{N(S_0)(x), \quad x \in [0,1]\} \tag{19}$$

For the evaluation of $\mathrm{d}\rho_1$ a longer computation has to be performed.

Lemma 1
Let ρ_0 and S_0 be two positive functions in $L^\infty \cap BV([0,1],R)$. Let ρ_1 and S_1 be two positive and linear functions defined on $[0,1]$:

$$S_1 = \overline{S_0} + \mathrm{d}S_1(x - 1/2) \tag{20}$$

$$\rho_1 = \overline{\rho_0} + \mathrm{d}\rho_1(x - 1/2) \tag{21}$$

Then, a sufficient condition on the slope $\mathrm{d}\rho_1$ in order to have (18) is that the following inequality holds on $[0,1]$:

$$\alpha \cdot \left(\frac{S_1}{\overline{S_0}} (\overline{S_0} N(\rho_0) - \overline{\rho_0} N(S_0)) - \frac{\overline{S_0} + N(S_0)(x - 1/2)}{\overline{S_0}} \alpha \right) \geqslant 0 \tag{22}$$

where α is defined by $\alpha = \overline{S_0}\mathrm{d}\rho_1 - \overline{\rho_0}\mathrm{d}S_1$.

Proof 2
It is easy to check that $(\rho, S) \to \rho H(S/\rho)$ is a convex function. Thus, using Theorem 1, a sufficient condition in order to have (18) is

$$(\mathrm{d}\rho_1, \mathrm{d}S_1) \begin{pmatrix} 1 & -\dfrac{\rho_1}{S_1} \\ -\dfrac{\rho_1}{S_1} & \left(\dfrac{\rho_1}{S_1}\right)^2 \end{pmatrix} \begin{pmatrix} N(\rho_0) - \mathrm{d}\rho_1 \\ N(S_0) - \mathrm{d}S_1 \end{pmatrix} \geqslant 0$$

or

$$(S_1\mathrm{d}\rho_1 - \rho_1\mathrm{d}S_1)(S_1 N(\rho_0) - \rho_1 N(S_0) - (S_1\mathrm{d}\rho_1 - \rho_1\mathrm{d}S_1)) \geqslant 0$$

using (20) and (21), and thanks to basic computations, we find (22).

The slope $\mathrm{d}q_1$ is computed in the same spirit.

Lemma 2
Let q_0 and ρ_0 be two functions in $L^\infty \cap BV([0,1],R)$, with $\rho_0 > 0$. Let q_1 and ρ_1 be two linear functions defined on $[0,1]$, with $\rho_1 > 0$.

$$\begin{aligned} q_1 &= \overline{q_0} + \mathrm{d}q_1(x - 1/2) \\ \rho_1 &= \overline{\rho_0} + \mathrm{d}\rho_1(x - 1/2) \end{aligned} \tag{23}$$

Then, a sufficient condition on the slope $\mathrm{d}q_1$ in order to have (16) is that the following inequality holds on $[0,1]$:

$$\beta \cdot \left(\frac{\rho_1}{\overline{\rho_0}} (\overline{\rho_0} N(q_0) - \overline{q_0} N(\rho_0)) - \frac{\overline{\rho_0} + N(\rho_0)(x - 1/2)}{\overline{\rho_0}} \beta \right) \geqslant 0 \tag{24}$$

where β is defined by $\beta = (\overline{\rho_0}\mathrm{d}q_1 - \overline{q_0}\mathrm{d}\rho_1)$.

The next theorem is devoted to the practical computation of the slopes $\mathrm{d}S_1, \mathrm{d}\rho_1, \mathrm{d}q_1$ in order to solve inequalities (16)–(18) and in order to maintain the second-order property.

Theorem 2
With the previous notations, consider:

$$\alpha = \overline{S_0}\mathrm{d}\rho_1 - \overline{\rho_0}\mathrm{d}S_1, \quad \beta = (\overline{\rho_0}\mathrm{d}q_1 - \overline{q_0}\mathrm{d}\rho_1)$$

$$g_\alpha(x) = \frac{\overline{S_0} + N(S_0)(x)(x-1/2)}{\overline{S_0}}, \quad g_\beta(x) = \frac{\overline{\rho_0} + N(\rho_0)(x)(x-1/2)}{\overline{\rho_0}}$$

$$h_\alpha(x) = \frac{S_1}{\overline{S_0}}(\overline{S_0}N(\rho_0)(x) - \overline{\rho_0}N(S_0)(x))$$

and

$$h_\beta(x) = \frac{\rho_1}{\overline{\rho_0}}(\overline{\rho_0}N(q_0)(x) - \overline{q_0}N(\rho_0)(x))$$

Then, if $\mathrm{d}S_1$ is defined by (19) and if α and β are defined by

$$\frac{1}{\alpha} = \begin{cases} \max\limits_{x\in[0,1]} \dfrac{g_\alpha(x)}{h_\alpha(x)} & \text{if } h_\alpha > 0 \\ \min\limits_{x\in[0,1]} \dfrac{g_\alpha(x)}{h_\alpha(x)} & \text{if } h_\alpha < 0 \\ \infty & \text{otherwise} \end{cases}$$

$$\frac{1}{\beta} = \begin{cases} \max\limits_{x\in[0,1]} \dfrac{g_\beta(x)}{h_\beta(x)} & \text{if } h_\beta > 0 \\ \min\limits_{x\in[0,1]} \dfrac{g_\beta(x)}{h_\beta(x)} & \text{if } h_\beta < 0 \\ \infty & \text{otherwise} \end{cases}$$

then, ρ_1, S_1, q_1 are second-order approximations of respectively ρ_0, S_0, q_0, and inequalities (17), (22), (24) are satisfied. In other words, with this choice of slopes, Π is a second-order entropy diminishing projection.

Proof 3
The case of $\mathrm{d}S_1$ has already been treated before. The inequality to solve for α is

$$\alpha(h_\alpha(x) - \alpha g_\alpha(x)) \geqslant 0$$

$\alpha = 0$ is a solution, but in order to achieve second-order, α has to be a first-order approximation of $h_\alpha(x)$. Thus, we can assume that α and h_α have the same signs. Then, suppose that $h_\alpha(x) > 0$ for all $x \in [0,1]$. If $\alpha > 0$, we have to solve $1/\alpha \geqslant g_\alpha(x)/h_\alpha(x)$, $x \in [0,1]$. We thus take $1/\alpha = \max_{[0,1]}(g_\alpha(x)/h_\alpha(x)) > 0$. In the same way, if $h_\alpha(x) < 0$ we choose

$1/\alpha = \min_{[0,1]}(g_\alpha(x)/h_\alpha(x)) < 0$. Finally, if h_α takes positive and negative values on $[0,1]$, we take $\alpha = 0$. The computation of β is completely similar.

5. FIRST APPLICATION: A SECOND-ORDER BOLTZMANN SCHEME

5.1. An approximate Riemann solver based on a kinetic interpretation

In this part, a possible computation of $w^{n+1,-}$ is presented. For practical reasons, we shall not use a classical Riemann solver, but instead, a resolution step based on the kinetic interpretation of the Euler equations. The kinetic interpretation that we will describe below has been proposed by Perthame in Reference [7]. His model has the property of being entropy diminishing for one particular entropy (see Reference [7]) and not necessarily for the other entropies. For simplicity, it will be exposed in the case $\gamma = 3$, where the computations are easier, but can be extented to other values of γ. Actually, it would be better to use the more recent model of Bouchut as described in Reference [21], which is entropy diminishing for all entropies.

Let us describe the original kinetic interpretation of Perthame. For this purpose, we introduce the following function (called in the sequel generalized Maxwellian).

$$M_w(v) = \frac{\rho}{2\sqrt{6\varepsilon}} Y\left(\frac{v-u}{\sqrt{6\varepsilon}}\right) \tag{25}$$

where

- Y is the characteristic function of $[-1,1]$:

$$Y(t) = \begin{cases} 1 & \text{if } |t| < 1 \\ 0 & \text{if } |t| \geqslant 1 \end{cases}$$

- v is the microscopic speed.
- $w = \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}$ is the macroscopic state of the gas.

It is straightforward to check that

$$\int_{v=-\infty}^{v=+\infty} M_w(v) \begin{pmatrix} 1 \\ v \\ v^2/2 \end{pmatrix} \mathrm{d}v = w$$

Let us now set

$$m(t,x,v) = M_{w(t,x)}(v)$$

It is also easy to check that

$$\int_{v=-\infty}^{v=+\infty} \partial_t m + v\partial_x m = \partial_t w + \partial_x f(w) \tag{26}$$

This fact leads to a numerical resolution known as a Boltzmann scheme. Many papers have been published on this subject. Without pretending to be exhaustive, we can cite for example the works of Deshpande [6], Bourdel *et al.* [8]. All these works are based on the physical Maxwellian. Each time step of a Boltzmann scheme is made of two substeps.

- Free transport step: $w^n(x)$ being given, the following evolution problem is solved during a time step:

$$\begin{gathered}\partial_t g + v\partial_x g = 0, \quad 0 \leqslant t \leqslant \tau \\ g(0,x,v) = M_{w^n(x)}(v)\end{gathered} \tag{27}$$

- Collision step: from the solution at time $t = \tau$, a new macroscopic state is recovered

$$w^{n+1,-}(x) = \int_{v=-\infty}^{v=+\infty} g(\tau,x,v)K(v)\,\mathrm{d}v$$

where

$$K(v) = \begin{pmatrix} 1 \\ v \\ v^2/2 \end{pmatrix}$$

It is important here to point out that the Boltzmann solver is only an approximate Riemann solver. This is due to the fact that during the free transport procedure (27) the equality

$$m(t,x,v) = g(t,x,v)$$

does not hold (the microscopic state is not a generalized Maxwellian). The collision step acts as a relaxation of the microscopic state to a Maxwellian state. Of course this solver tends to an exact solver when $\tau \to 0$.

On the other hand, despite its simplicity, this approach leads to very tedious computations for $w^{n+1,-}$. Therefore, we propose the following simplification: in the free transport step (27), we replace the initial condition by its linear interpolation on each cell

$$f(0,x,v) = \frac{x - x_{i-1/2}}{h} M_{w_{i,r}}(v) + \frac{x_{i+1/2} - x}{h} M_{w_{i,l}}(v) \quad \text{for } x \in C_i$$

where

$$w_{i,r} = w_i^n + \frac{h}{2} s_i^n \quad \text{and} \quad w_{i,l} = w_i^n - \frac{h}{2} s_i^n$$

It must be noticed that this approximation is conservative but unfortunately, it is also necessarily entropy increasing because we replace a Maxwellian state (which corresponds to a minimum of entropy) at each point of the cell by a linear approximation.

5.2. Computation of the approximate second-order Riemann solver

Let $x \in C_i$, we have

$$w^{n+1,-}(x) = \int_{v=-\infty}^{+\infty} f(0, x - v\tau, v) K(v)\, \mathrm{d}v$$

Consider then the speed $v_{\min}$ (respectively, $v_{\max}$) at which a particle in $x_{i+1/2}$ (respectively, $x_{i-1/2}$) at time $t = 0$ reaches x at time $t = \tau$

$$v_{\min} = \frac{x - x_{i+1/2}}{\tau} < 0$$

$$v_{\max} = \frac{x - x_{i-1/2}}{\tau} > 0$$

The time step τ is supposed to be smaller than hv^*, where v^* is greater than the biggest support of all the generalized Maxwellians plus the maximal speed of the flow. Thanks to this CFL condition, the computation of $w^{n+1,-}(x)$ can then be split into three parts: a contribution from the left cell C_{i-1}, the right cell C_{i+1}, and the middle cell C_i:

$$w^{n+1,-}(x) = A_l + A_m + A_r$$

$$A_l = \int_{v=v_{\max}}^{+\infty} \left[\frac{x - v\tau - x_{i-3/2}}{h} M_{w_{i-1,r}}(v) + \frac{x_{i-1/2} - x + v\tau}{h} M_{w_{i-1,l}}(v) \right] K(v)\, \mathrm{d}v$$

$$A_r = \int_{v=-\infty}^{v_{\min}} \left[\frac{x - v\tau - x_{i+1/2}}{h} M_{w_{i+1,r}}(v) + \frac{x_{i+3/2} - x + v\tau}{h} M_{w_{i+1,l}}(v) \right] K(v)\, \mathrm{d}v$$

$$A_m = \int_{v=v_{\min}}^{v_{\max}} \left[\frac{x - v\tau - x_{i-1/2}}{h} M_{w_{i,r}}(v) + \frac{x_{i+1/2} - x + v\tau}{h} M_{w_{i,l}}(v) \right] K(v)\, \mathrm{d}v$$

A more detailed expression of A_l, A_r, A_m is given in Appendix A. It can be checked that $w^{n+1,-}$ is piecewise polynomial of degree $\leqslant 4$ on each cell. The number of pieces is $\leqslant 22$.

5.3. Numerical results

For the numerical results that are presented in this section, we decided to compute exactly $w^{n+1,-}$ given by the kinetic Riemann solver. This is done thanks to a C++ class of piecewise polynomial functions. It appears then that $S^{n+1,-}$ is not, in general, piecewise polynomial. We thus had to construct a piecewise polynomial approximation $\tilde{S}$ of $S^{n+1,-}$. This is done with a Tchebychev interpolation with three points on each piece of regularity of $w^{n+1,-}$. Then $N(\rho)$, $N(q)$, and $N(\tilde{S})$ can be computed exactly. The final limitation procedure has been performed numerically with a sampling of the functions we wanted to maximize or minimize on each cell.

We were able to verify that our projection operator acts at least as a classical minmod limiter. Indeed, the numerical results appear to be precise and present no oscillation on a 200 cells mesh.

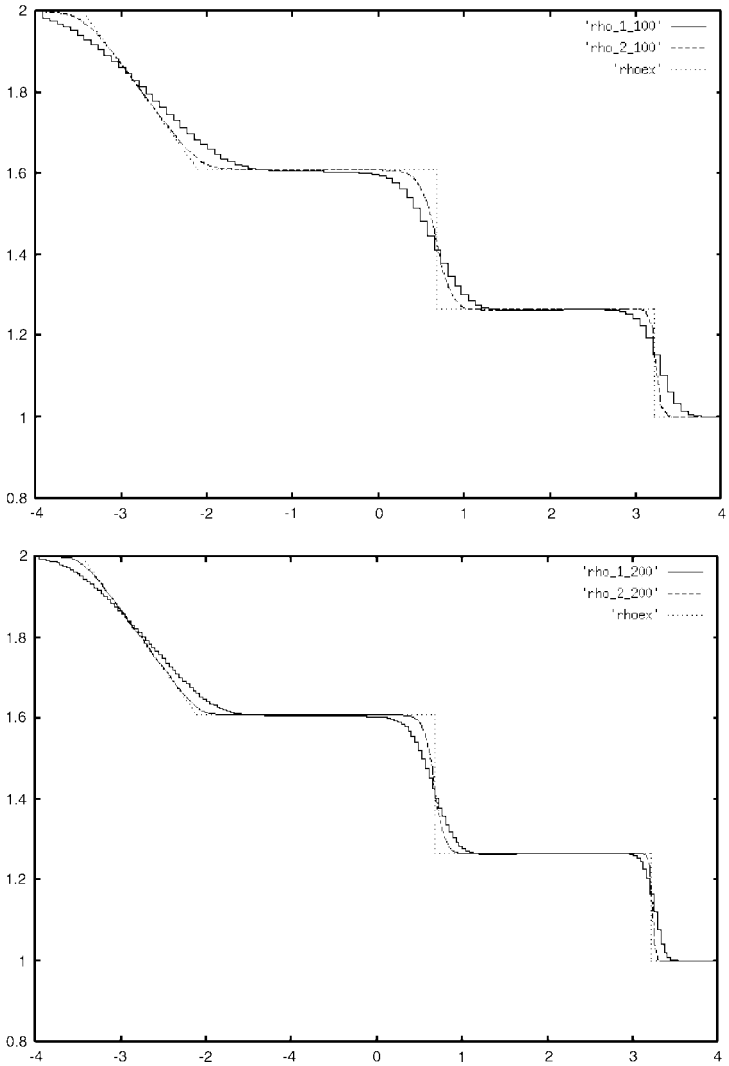


Figure 1. Density. First or second order. 100 or 200 cells.

We tested the scheme on the classical case of a shock tube problem with the following initial data:

Variable	Left state	Right state
Density	$\rho_L = 2$	$\rho_R = 1$
Velocity	$u_L = 0$	$u_R = 0$
Pressure	$p_L = 8$	$p_R = 2$

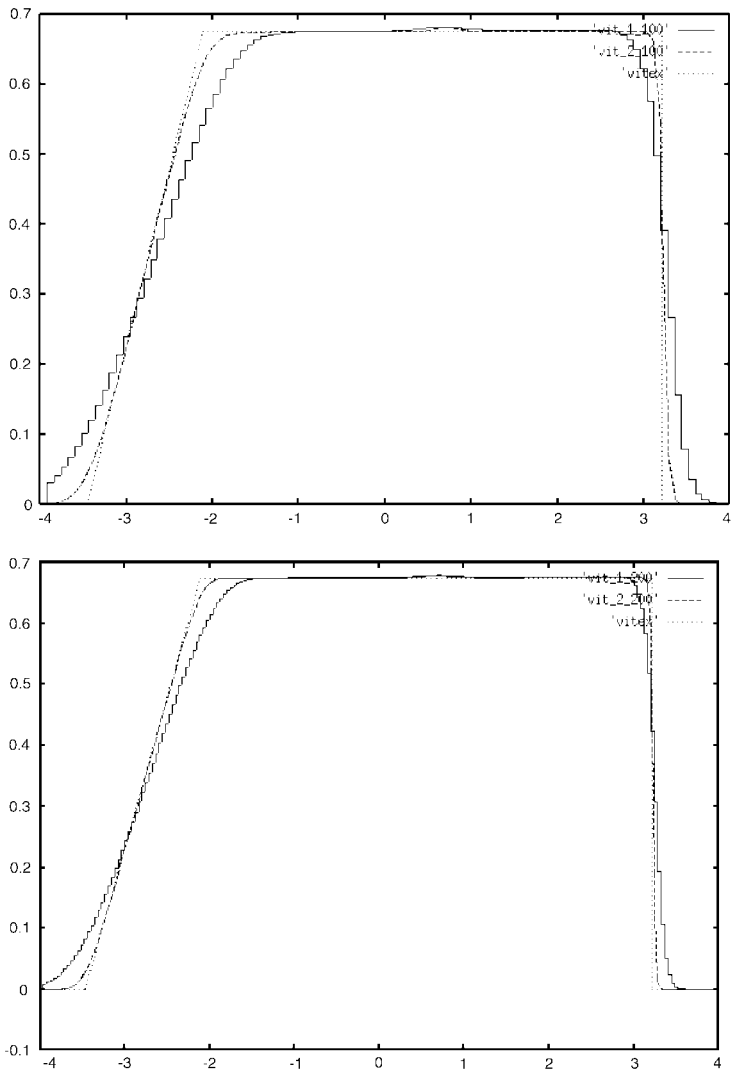


Figure 2. Velocity. First or second order. 100 or 200 cells.

In the Figures 1–3 a comparison is made between the first and the second-order schemes with 100 or 200 cells. The results are given at time $t = 1$. In the case of the first-order solution, the real (piecewise constant) mathematical solution has been plotted. Density, velocity and pressure are successively presented.

With a mesh refinement, oscillations start to appear. The phenomenon can be observed with a classical MUSCL scheme. This is due to the fact that the scheme is only first-order in time.

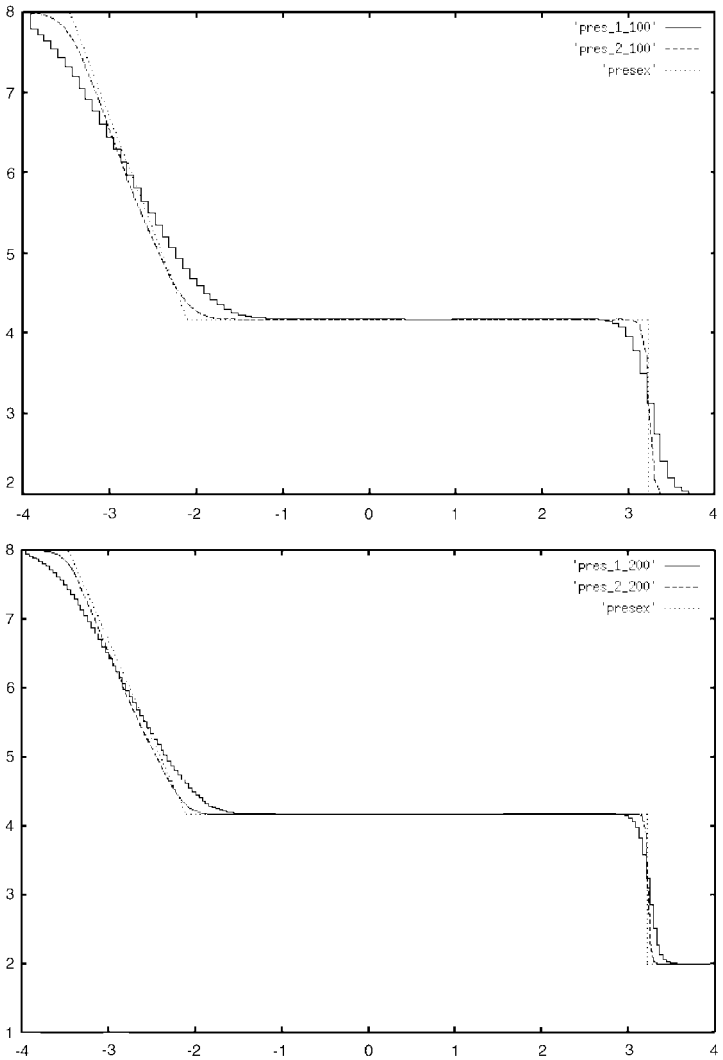


Figure 3. Pressure. First or second order. 100 or 200 cells.

The complexity of the scheme leaded us to abandon this approach. The next section is thus devoted to a more tractable application of the entropic projection.

6. SECOND APPLICATION: MEAN VALUE APPROACH

In this part, we envisage a simpler approach than the kinetic approach. We first build a polynomial interpolation of the solution, starting from cell averages, as in Harten's ENO

schemes [22]. We use an interpolation without upwinding or limitation in order to evaluate the effect of the entropic projection.

More precisely, suppose that we know the mean value w_i^n of the solution at time n in the cell C_i. A second order extension of the Godunov scheme reads

$$w_i^{n+1} = w_i^n - \frac{\tau}{h}(f_{i+1/2}^{n+1/2} - f_{i-1/2}^{n+1/2}) \tag{28}$$

The flux at interface $i+\frac{1}{2}$ and time $n+\frac{1}{2}$ is of the form

$$f_{i+1/2}^{n+1/2} = f(R(w_{i+1/2,-}^{n+1/2}, w_{i+1/2,+}^{n+1/2})) \tag{29}$$

The quantity $R(w_L, w_R)$ denotes the solution of the first-order exact Riemann problem at the interface between w_L and w_R. The quantity $w_{i+1/2,-}$ is the value of the reconstructed solution in the cell i at time $n+\frac{1}{2}$ and at the interface $i+\frac{1}{2}$. The choice $w_{i+1/2,-} = w_i^n$ corresponds to the classical first-order scheme. We focus now on the cell i. For simplicity, we suppose that $C_i =]0,1[$.

The first step is to construct a high order approximation of w from the cell averages. We thus suppose, with the notations of Section 4, that ρ_0, q_0 and S_0 are second-order polynomials in $]0,1[$. We also suppose that the reconstruction is conservative

$$\int_j^{j+1} w_0(x)\,\mathrm{d}x = w_{i+j}^n, \quad j = -1, 0, 1 \tag{30}$$

with

$$w_0(x) = \left(\rho_0(x), q_0(x), \frac{q_0^2(x)}{2\rho_0(x)} + S_0(x)^\gamma\right)^{\mathrm{T}} \tag{31}$$

It is known that the resulting interpolation is not necessarily positive for the density and the pressure, even if all the mean values are positive. So we propose in Appendix B a simple procedure to correct the interpolation, if necessary.

Because ρ_0, q_0 and S_0 are now second-order polynomials, the computations described in Theorem 1 become almost explicit. It is then possible to compute the limited variables ρ_1, q_1 and S_1, the energy correction described in (14) and then the fluxes at cell interfaces. The second-order in time is achieved with the Hancock method. It uses the space slope estimate to compute a time derivative estimate thanks to the conservation laws: $w_t = -f(w)_x$. The time derivative estimate permits then to compute the approximation of w at time $n+\frac{1}{2}$.

We have tested the scheme on the Riemann problem whose data are given in Tables I and II. This case is chosen in the book of Toro [3]. It presents a sonic rarefaction wave and a shock.

Table I. Data of the Riemann problem.

Variable	Left state	Right state
Density	$\rho_L = 1$	$\rho_R = 0.125$
Velocity	$u_L = 0.75$	$u_R = 0$
Pressure	$p_L = 1$	$p_R = 0.1$

Table II. Computation characteristics.

Interval	$]-1/2, 1/2[$
Number of cells	200
CFL	0.8
Final time	$t = 0.2$
γ	1.4

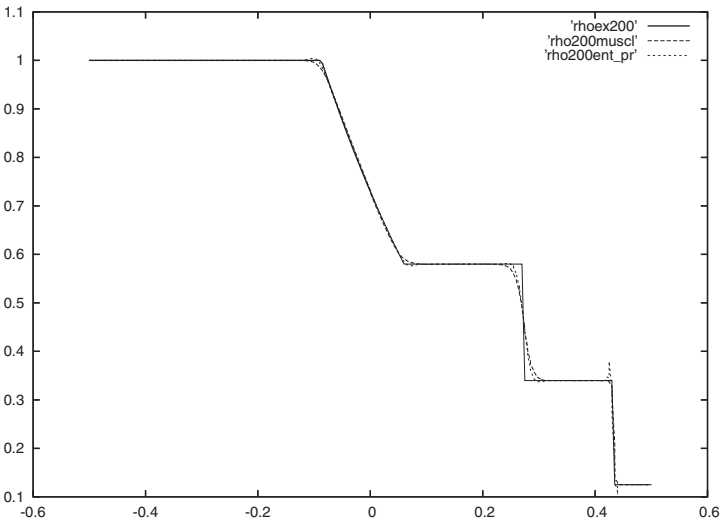


Figure 4. Comparison entropic scheme—MUSCL-Hancock scheme.

We compare in Figures 4–6 the results of the reconstruction-limitation scheme with a standard MUSCL-Hancock scheme (described in Reference [3]). We observe a slight improvement of the precision in the contact discontinuity, but also small overshoots and undershoots in the right shock. Without the correction described in Appendix B the computation would have not ended. It is necessary only on one cell in the first time step and only for the reconstruction of the density ρ near the contact discontinuity.

In the next numerical experiment, we evaluate the rate of convergence in the L^1 norm (for density ρ) of the reconstruction-limitation scheme and compared it in Figure 7 and Table III with the standard MUSCL-Hancock scheme. The rate is computed for a simple contact discontinuity whose values are given in Table IV. The characteristics of this computation are summed up in Table V. We observe that the projection scheme is more precise than the MUSCL scheme, but the asymptotic rates of convergence seems to be approximately the same. Recall that, for a simple contact discontinuity, the standard MUSCL 'second-order' scheme has a convergence rate of $\frac{2}{3} \simeq 0.66666$.

In the last numerical experiment, we evaluate the rate of convergence in the L^1 norm (for density ρ) of the reconstruction-limitation scheme and compared it in Figure 8 and Table VI

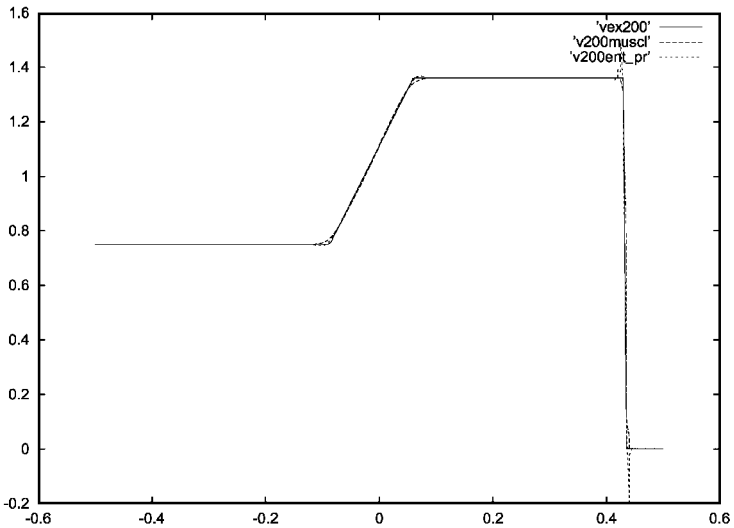


Figure 5. Comparison entropic scheme—MUSCL-Hancock scheme.

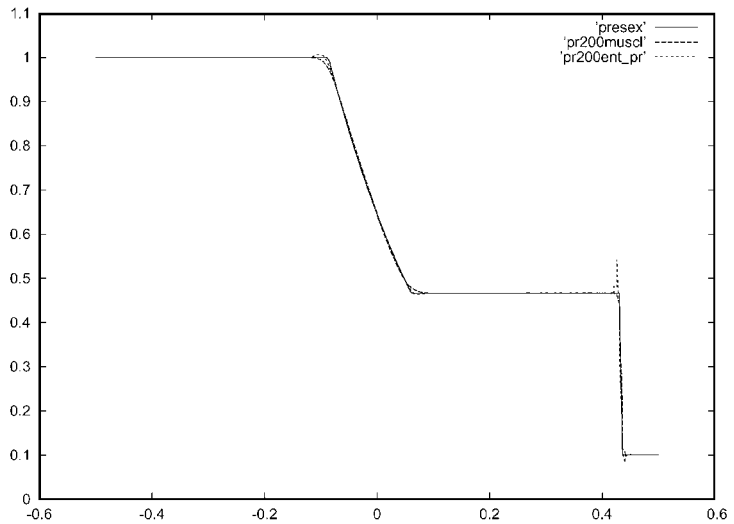


Figure 6. Comparison entropic scheme—MUSCL-Hancock scheme.

with the standard MUSCL-Hancock scheme. The rate is computed for a simple shock whose values are given in Table VII. The characteristics of this computation are summed up in Table VIII. We observe that the projection scheme is more precise than the MUSCL scheme,

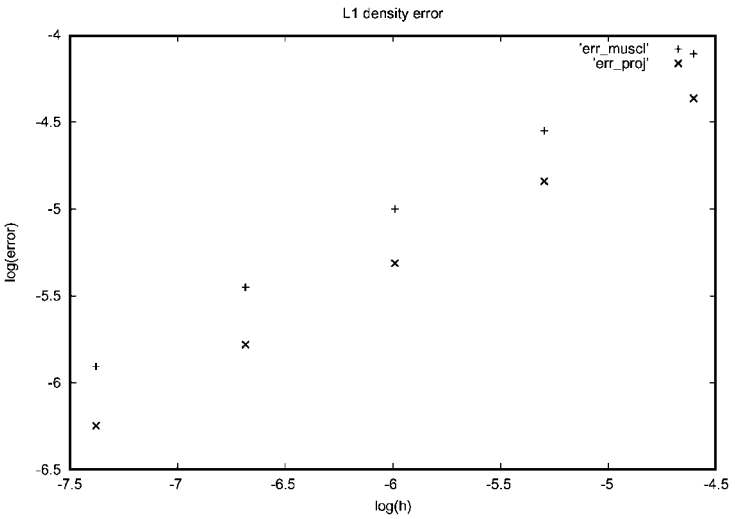


Figure 7. Comparison entropic scheme—MUSCL-Hancock scheme. Rate of convergence for a contact discontinuity.

Table III. Convergence test, contact discontinuity.

ln(h)	ln(error) MUSCL	ln(error) proj.	Rate MUSCL	Rate proj.
−4.60517	−4.10716	−4.36288	—	—
−5.29832	−4.55118	−4.84035	0.64058	0.68884
−5.99146	−4.99951	−5.31210	0.64681	0.68060
−6.68461	−5.45112	−5.77950	0.65153	0.67431
−7.37776	−5.90507	−6.24700	0.65491	0.67446

Table IV. Contact discontinuity.

Variable	Left state	Right state
Density	$\rho_L = 2$	$\rho_R = 1$
Velocity	$u_L = 1$	$u_R = 1$
Pressure	$p_L = 1$	$p_R = 1$

but the asymptotic rates of convergence seem to be approximately the same. Recall that, for a simple shock, the standard MUSCL 'second order' scheme has a convergence rate of 1.

The program we used for the numerical results of this section can be downloaded at `http://helluy/entropy/index.html`

Table V. Computation characteristics, contact.

Interval	$]-1/2, 1/2[$
Number of cells	200 to 1600
CFL	$\simeq 0.45$
Final time	$t = 0.2$
γ	1.4

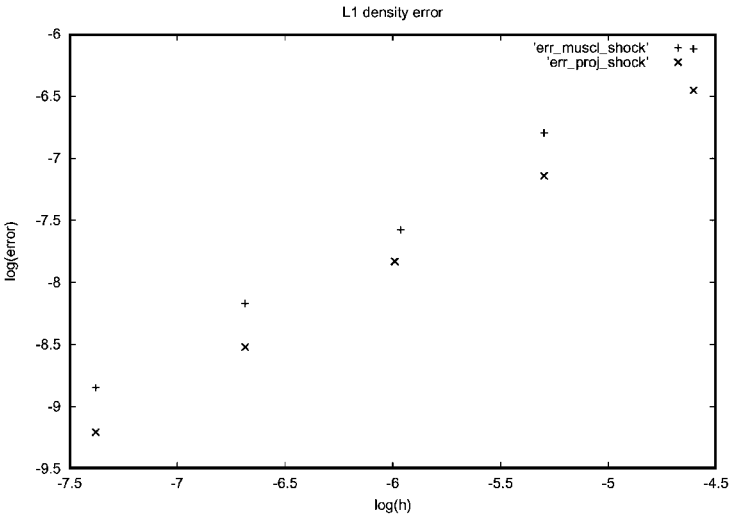


Figure 8. Comparison entropic scheme—MUSCL-Hancock scheme. Rate of convergence for a shock.

Table VI. Convergence test, shock.

ln(h)	ln(error) MUSCL	ln(error) proj.	Rate MUSCL	Rate proj.
−4.60517	−6.12021	−6.45422	—	—
−5.29832	−6.79584	−7.14075	0.97472	0.99045
−5.99146	−7.48411	−7.83202	0.99297	0.99730
−6.68461	−8.17001	−8.52117	0.98954	0.99423
−7.37776	−8.84886	−9.20612	0.97937	0.98817

Table VII. Shock wave of velocity 1.

Variable	Left state	Right state
Density	$\rho_L = 1$	$\rho_R = \frac{3}{4}$
Velocity	$u_L = 0$	$u_R = -\frac{1}{3}$
Pressure	$p_L = 1$	$p_R = \frac{2}{3}$

Table VIII. Computation characteristics, shock.

Interval	$]-1/2, 1/2[$
Number of cells	200 to 1600
CFL	$\simeq 0.58$
Final time	$t = 0.2$
γ	1.4

7. CONCLUSION

In this paper, we have proposed a second-order generalization of the Godunov scheme for the Euler equations. In the first part, we have built a second-order entropic projection with explicit formulas. In the second part, we have numerically tested the projection. Because it is difficult to construct an exact second-order Riemann solver, we had to simplify the theoretical approach. We proposed two applications: the slope limitation applied to an approximate kinetic Riemann solver and the slope limitation applied to a polynomial reconstruction with cell averages.

Our approach is rigorous and gives a clear justification to the slope limitation procedure.

Many questions still remain. For example:

1. Is it possible to extend the scheme to higher dimensions?
2. Can we improve the efficiency of the computation?
3. Is it possible to extend the method to more general equation of state (EOS) than the perfect gas EOS?

The answer to the first question is yes. It can be done with at least two methods. The first method, which is the simplest could be to use an alternate direction method. Then the scheme is limited to cartesian grids. Another way would be to extend Theorem 1 to triangles or quadrilaterals. We do not know if the computation of the practical projection remains possible.

The answer to the second question is certainly yes. But we do not know if it is possible to find a simpler computation without abandoning an exact entropy decrease. By relaxing the exact entropy decrease or approximating the formula of Theorem 2, some schemes have already been designed for the Lagrangian equations in Reference [14].

The answer to the third question is: maybe yes. The main ingredient in the presented construction is the existence of an entropy whose hessian is degenerated. For a given pressure EOS

$$p = p(\tau = 1/\rho, \varepsilon) \tag{32}$$

the Lax entropies of the associated Euler equations are constructed, as described in Reference [23], from the concave solutions $s(\tau, \varepsilon)$ of

$$\frac{\partial s}{\partial \tau} - p \frac{\partial s}{\partial \varepsilon} = 0 \tag{33}$$

The Lax entropies are then $U = -\rho s$. The general concave solutions of (33) are of the form $s = A(s_0)$ where A is a function with monotony and convexity properties, and s_0 a particular solution. If the hessian of s_0 is degenerated, maybe that the previous construction can be generalized.

We would like to end this presentation by insisting on the fact that the second-order entropic projection can be used as a slope limiter for other numerical methods. The Galerkin discontinous method (see References [24–26]) could be a candidate for such a limiter. Indeed in this method a piecewise polynomial approximation at each time step has to be limited in order to avoid spurious oscillations.

APPENDIX A: COMPUTATIONS FOR THE BOLTZMANN SCHEME

We start with A_m:

$$A_m = \frac{x - x_{i-1/2}}{h} \int_{v=v_{\min}}^{v_{\max}} M_{w_{i,r}}(v) K(v)\, \mathrm{d}v$$

$$-\frac{\tau}{h} \int_{v=v_{\min}}^{v_{\max}} M_{w_{i,r}}(v) v K(v)\, \mathrm{d}v$$

$$+\frac{x_{i+1/2} - x}{h} \int_{v=v_{\min}}^{v_{\max}} M_{w_{i,l}}(v) K(v)\, \mathrm{d}v$$

$$+\frac{\tau}{h} \int_{v=v_{\min}}^{v_{\max}} M_{w_{i,l}}(v) v K(v)\, \mathrm{d}v$$

$$A_m = \frac{x - x_{i-1/2}}{h} \frac{\rho_{i,r}}{2\sqrt{6\varepsilon_{i,r}}} \begin{bmatrix} v \\ v^2/2 \\ v^3/6 \end{bmatrix}_{\max(v_{\min}, u_{i,r} - \sqrt{6\varepsilon_{i,r}})}^{\min(v_{\max}, u_{i,r} + \sqrt{6\varepsilon_{i,r}})}$$

$$-\frac{\tau}{h} \frac{\rho_{i,r}}{2\sqrt{6\varepsilon_{i,r}}} \begin{bmatrix} v^2/2 \\ v^3/3 \\ v^4/8 \end{bmatrix}_{\max(v_{\min}, u_{i,r} - \sqrt{6\varepsilon_{i,r}})}^{\min(v_{\max}, u_{i,r} + \sqrt{6\varepsilon_{i,r}})}$$

$$+\frac{x_{i+1/2}-x}{h}\,\frac{\rho_{i,l}}{2\sqrt{6\varepsilon_{i,l}}}\begin{bmatrix} v \\ v^2/2 \\ v^3/6 \end{bmatrix}_{\max(v_{\min},u_{i,l}-\sqrt{6\varepsilon_{i,l}})}^{\min(v_{\max},u_{i,l}+\sqrt{6\varepsilon_{i,l}})}$$

$$+\frac{\tau}{h}\,\frac{\rho_{i,l}}{2\sqrt{6\varepsilon_{i,l}}}\begin{bmatrix} v^2/2 \\ v^3/3 \\ v^4/8 \end{bmatrix}_{\max(v_{\min},u_{i,l}-\sqrt{6\varepsilon_{i,l}})}^{\min(v_{\max},u_{i,l}+\sqrt{6\varepsilon_{i,l}})}$$

In the same way:

$$A_l=\frac{x-x_{i-3/2}}{h}\,\frac{\rho_{i-1,r}}{2\sqrt{6\varepsilon_{i-1,r}}}\begin{bmatrix} v \\ v^2/2 \\ v^3/6 \end{bmatrix}_{\max(v_{\max},u_{i-1,r}-\sqrt{6\varepsilon_{i-1,r}})}^{u_{i-1,r}+\sqrt{6\varepsilon_{i-1,r}}}$$

$$-\frac{\tau}{h}\,\frac{\rho_{i-1,r}}{2\sqrt{6\varepsilon_{i-1,r}}}\begin{bmatrix} v^2/2 \\ v^3/3 \\ v^4/8 \end{bmatrix}_{\max(v_{\max},u_{i-1,r}-\sqrt{6\varepsilon_{i-1,r}})}^{u_{i-1,r}+\sqrt{6\varepsilon_{i-1,r}}}$$

$$+\frac{x_{i-1/2}-x}{h}\,\frac{\rho_{i-1,l}}{2\sqrt{6\varepsilon_{i-1,l}}}\begin{bmatrix} v \\ v^2/2 \\ v^3/6 \end{bmatrix}_{\max(v_{\max},u_{i-1,l}-\sqrt{6\varepsilon_{i-1,l}})}^{u_{i-1,l}+\sqrt{6\varepsilon_{i-1,l}}}$$

$$+\frac{\tau}{h}\,\frac{\rho_{i-1,l}}{2\sqrt{6\varepsilon_{i-1,l}}}\begin{bmatrix} v^2/2 \\ v^3/3 \\ v^4/8 \end{bmatrix}_{\max(v_{\max},u_{i-1,l}-\sqrt{6\varepsilon_{i-1,l}})}^{u_{i-1,l}+\sqrt{6\varepsilon_{i-1,l}}}$$

and

$$A_r=\frac{x-x_{i+1/2}}{h}\,\frac{\rho_{i+1,r}}{2\sqrt{6\varepsilon_{i+1,r}}}\begin{bmatrix} v \\ v^2/2 \\ v^3/6 \end{bmatrix}_{u_{i+1,r}-\sqrt{6\varepsilon_{i+1,r}}}^{\min(v_{\min},u_{i+1,r}+\sqrt{6\varepsilon_{i+1,r}})}$$

$$-\frac{\tau}{h}\frac{\rho_{i+1,r}}{2\sqrt{6\varepsilon_{i+1,r}}}\begin{bmatrix}v^2/2\\ v^3/3\\ v^4/8\end{bmatrix}_{u_{i+1,r}-\sqrt{6\varepsilon_{i+1,r}}}^{\min(v_{\min},u_{i+1,r}+\sqrt{6\varepsilon_{i+1,r}})}$$

$$+\frac{x_{i+3/2}-x}{h}\frac{\rho_{i+1,l}}{2\sqrt{6\varepsilon_{i+1,l}}}\begin{bmatrix}v\\ v^2/2\\ v^3/6\end{bmatrix}_{u_{i+1,l}-\sqrt{6\varepsilon_{i+1,l}}}^{\min(v_{\min},u_{i+1,l}+\sqrt{6\varepsilon_{i+1,l}})}$$

$$+\frac{\tau}{h}\frac{\rho_{i+1,l}}{2\sqrt{6\varepsilon_{i+1,l}}}\begin{bmatrix}v^2/2\\ v^3/3\\ v^4/8\end{bmatrix}_{u_{i+1,l}-\sqrt{6\varepsilon_{i+1,l}}}^{\min(v_{\min},u_{i+1,l}+\sqrt{6\varepsilon_{i+1,l}})}$$

It is easy to check that $w^{n+1,-}$ is piecewise polynomial and continuous on each cell. For instance, the practical computation of a term like

$$A=\begin{bmatrix}v\\ v^2/2\\ v^3/6\end{bmatrix}_{\max(v_{\min},v)}^{\min(v_{\max},V)}$$

where $v<V$, gives

$$A=\begin{bmatrix}V-v\\ V^2/2-v^2/2\\ V^3/6-v^3/6\end{bmatrix}\quad\text{if } v_{\min}<v \text{ and } v_{\max}>V, \text{ i.e. } x\in\,]V\tau+x_{i-1/2},v\tau+x_{i+1/2}[$$

$$A=\begin{bmatrix}V-v_{\min}\\ V^2/2-v_{\min}^2/2\\ V^3/6-v_{\min}^3/6\end{bmatrix}\quad\text{if } v_{\min}>v \text{ and } v_{\max}>V, \text{ i.e. } x\in\,]v\tau+x_{i+1/2},V\tau+x_{i+1/2}[$$

$$A=\begin{bmatrix}v_{\max}-v\\ v_{\max}^2/2-v^2/2\\ v_{\max}^3/6-v^3/6\end{bmatrix}\quad\text{if } v_{\min}<v \text{ and } v_{\max}<V, \text{ i.e. } x\in\,]v\tau+x_{i-1/2},V\tau+x_{i-1/2}[$$

$$A=0 \text{ if } x>V\tau+x_{i+1/2} \text{ or } x<v\tau+x_{i-1/2}$$

Remark
By the CFL condition the inequality $V\tau + x_{i-1/2} < v\tau + x_{i+1/2}$ necessarily holds.

APPENDIX B: POSITIVE MEAN VALUE INTERPOLATION

This appendix is devoted to an algorithm in order to avoid negative values in the interpolation process. In all the numerical tests that we presented it was necessary to activate this correction only on a few cells. This method can be interesting for other purposes.

For this, we consider a scalar non-negative function f defined on the interval $[-1,2]$. We know the mean values of f on the sub-intervals $I_i = [i-1, i]$, $i = 0, 1, 2$

$$f_i = \int_{i-1}^{i} f(t)\,\mathrm{d}t \geqslant 0 \tag{B1}$$

A classical interpolation would be to find a second-order polynomial P satisfying

$$f_i = \int_{i-1}^{i} P(t)\,\mathrm{d}t \tag{B2}$$

but it is known that this interpolation can be negative in some point in the interval $[0,1]$, even if the three mean values f_i are positive. Instead, we will consider a constrained optimization problem. We consider a base (P_i) of the second-order polynomials satisfying $\int_{I_i} P_j = \delta_{ij}$ where δ_{ij} is the Kronecker symbol.

$$\begin{aligned} P_1(x) &= -x^2 + x + \tfrac{5}{6} \\ P_0(x) &= \frac{x^2}{2} - x + \frac{1}{3} \\ P_2(x) &= \frac{x^2}{2} - \frac{1}{6} \end{aligned} \tag{B3}$$

The polynomial P is searched under the form

$$P = g_0 P_0 + g_1 P_1 + g_2 P_2 \tag{B4}$$

In this way, the mean values of P on $[i-1, i]$ are g_i for $i = 0, 1, 2$. The conservation property imposes

$$g_1 = f_1 \tag{B5}$$

We then consider the functional $J(P) = \frac{1}{2}((g_0 - f_0)^2 + (g_2 - f_2)^2)$. We solve the optimization problem: find $P \geqslant 0$ such that $g_1 = f_1$ and $J(P)$ is minimal. Note that if the interpolation polynomial defined in (B2) is non-negative then it solves the minimization problem and then $J(P) = 0$.

Consider the Lagrangian $L(P, \mu) = J(P) - \langle \mu, P \rangle$, where μ is in the set $M^{1,+}$ of positive bounded measures on $[0,1]$. The optimization problem is equivalent to

$$\inf_{\substack{g_0, g_1, g_2 \\ g_1 = f_1}} \sup_{\mu \in M^{1,+}} L(P, \mu) \tag{B6}$$

The optimality condition classically reads

$$g_0 = f_0 + \langle \mu, P \rangle$$

$$g_2 = f_2 + \langle \mu, P \rangle$$

$$\forall x \in [0,1], \quad \mu(x)P(x) = 0$$

But a second-order polynomial has at most two roots. This means that μ is a linear combination of at most two Dirac masses. Due to the fact that P has to be non-negative, we can then distinguish the following cases:

1. P is positive, then $g_0 = f_0$, $f_1 = g_1$ and $g_2 = f_2$.
2. P is positive in $]0,1]$ and $P(0) = 0$ then $f_0 = g_0 + \mu_0 P_0(0)$, $f_1 = g_1$ and $f_2 = g_2$.
3. P is positive in $[0,1[$ and $P(1) = 0$ then $f_2 = g_2 + \mu_2 P_2(1)$, $f_1 = g_1$ and $f_0 = g_0$.
4. P is positive in $]0,1[$ and $P(0) = P(1) = 0$ then $f_0 = g_0 + \mu_0 P_0(0)$, $f_2 = g_2 + \mu_2 P_2(1)$ and $f_1 = g_1$.
5. P is positive in $[0,1] - x_0$, with $x_0 \in]0,1[$. Then, $f_0 = g_0 + \mu_0 P_0(x_0)$, $f_2 = g_2 + \mu_0 P_2(x_0)$, $f_1 = g_1$, $P(x_0) = 0$, $P'(x_0) = 0$.

Due to the positivity of P, case (4) never happens. The algorithm to compute P is then the following.

First, we have necessarily $f_1 = g_1$. Then the following cases are considered:

1. If $g_1 \geqslant 1/2(g_0 + g_2))$ take $f_0 = g_0$, $f_1 = g_1$, $f_2 = g_2$. The resulting polynomial is concave and > 0;
2. if $g_1 \leqslant 1/2(g_0 + g_2))$, try $f_0 = g_0$, $f_1 = g_1$, $f_2 = g_2$. The resulting polynomial is convex. It is solution if $P(0) \geqslant 0$, $P'(0) \geqslant 0$ or $P(1) \geqslant 0$, $P'(1) \leqslant 0$.
3. if $g_1 < 1/2(g_0 + g_2))$ and $7/2g_1 - 1/2g_2 \geqslant 0$ and $\mu_0 = 3/2g_2 - 15/2g_1 - 3g_0 \geqslant 0$ then the solution is given by $f_0 = g_0 + \mu_0/3$, $f_2 = g_2$;
4. if $g_1 < 1/2(g_0 + g_2))$ and $7/2g_1 - 1/2g_0 \geqslant 0$ and $\mu_2 = 3/2g_0 - 15/2g_1 - 3g_2 \geqslant 0$ then the solution is given by $f_0 = g_0$, $f_2 = g_2 + \mu_2/3$;
5. Finally, if $g_1 < 1/2(g_0 + g_2))$ in all the other cases, solve $f_0 = g_0 + \mu_0 P_0(x_0)$, $f_2 = g_2 + \mu_0 P_2(x_0)$, $f_1 = g_1$, $P(x_0) = 0$, $P'(x_0) = 0$ for x_0 and μ_0.

The algorithm, written in Maple and C++ languages, can be downloaded and tested at `http://helluy/entropy/index.html`.

REFERENCES

1. Godunov SK. A difference scheme for numerical computation of discontinuous solutions of equations of fluids mechanics. *Mathematics of the USSR-Sbornik* 1959; **47**:271–306.
2. Lax PD, Harten A, Van Leer B. On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. *SIAM Review* 1983; **25**:35–61.
3. Spruce M, Toro EF, Speares W. Restoration of the contact surface in the Hll–Riemann solver. *Shock Waves* 1994; **4**:25–34.
4. Roe PL. Approximate Riemann solvers, parameter vectors, and difference schemes. *Journal of Computational Physics* 1981; **43**(2):357–372.
5. Engquist B, Osher S. One-sided difference approximations for nonlinear conservation laws. *Mathematics of Computation* 1981; **36**(154):321–351.

6. Deshpande S, Mandal J. Kinetic theory based new upwind methods for inviscid compressible flows. *Proceedings of Euromekh Colloquium 224 on Kinetic Theory Aspects of Evaporation/Condensation Phenomena*, vol. 19, 1988; 3, 6, 9, 32–38.
7. Perthame B. Boltzmann type schemes for gas dynamics and the entropy property. *SIAM Journal on Numerical Analysis* 1990; **27**:1405–1421.
8. Bourdel F, Croisille J-P, Delorme P, Mazet P-A. On the approximation of K-diagonalizable hyperbolic systems by finite elements. Applications to the Euler equations and to gaseous mixtures. *Recherche Aérospatiale* 1989; **5**:15–34.
9. Harten A, Hyman JM. Self-adjusting grid methods for one-dimensional hyperbolic conservation laws. *Journal of Computational Physics* 1983; **50**(2):235–269.
10. Van Leer B. Towards the ultimate conservative difference scheme. a second order sequel to the Godunov's method. *Journal of Computational Physics* 1979; **32**:101–136.
11. Harten A, Osher S. Uniformly high order accurate nonoscillatory schemes, I. *SIAM Journal on Numerical Analyse* 1982; **24**:279–309.
12. Rauch J. BV estimates fail for most quasilinear hyperbolic systems in dimensions greater than one. *Communications in Mathematical Physics* 1986; **106**(3):481–484.
13. Bourdarias C, Bouchut F, Perthame B. A muscl method satisfying all the numerical entropy inequalities. *Mathematics of Computation* 1996; **65**(216):1439–1461.
14. Coquel F, LeFloch P. An entropy satisfying muscl scheme for systems of conservation laws. *Numerische Mathematik* 1996; **74**(01):1–34.
15. Godlewski E, Raviart P-A. *Numerical Approximation of Hyperbolic Systems of Conservation Laws*. Applied Mathematical Sciences, vol. 118. Springer: New York, 1996.
16. Bourgeade A, LeFloch Ph, Raviart P-A. An asymptotic expansion for the solution of the generalized Riemann problem. II. Application to the equations of gas dynamics. *Annales de l Institut Henri Poincaré-Analyse Non Linéaire* 1989; **6**(6):437–480.
17. Ben-Artzi M, Falcovitz J. An upwind second-order scheme for compressible duct flows. *SIAM Journal on Scientific and Statistical Computing* 1986; **7**(3):744–768.
18. Lax PD. Hyperbolic systems of conservation laws and the mathematical theory of shock waves. *CBMS Regional Conference Series in Applied Mathematics*, vol. 11. SIAM: Philadelphia, 1972.
19. Croisille JP. Contribution à l'étude théorique et à l'approximation par éléments finis du système hyperbolique de la dynamique des gaz multidimensionnelle et multiespèces. *Ph.D. Thesis*, Université Paris VI, 1991.
20. Godlewski E, Raviart PA. *Numerical Approximation of Hyperbolic Systems of Conservation Laws*. Springer: Berlin, 1996.
21. Bouchut F. Construction of BGK models with a family of kinetic entropies for a given system of conservation laws. *Journal of Statistical Physics* 1999; **95**(1–2):113–170.
22. Harten A. ENO schemes with subcell resolution. *Journal of Computational Physics* 1989; **83**(1):148–184.
23. Harten A, Lax PD, Levermore CD, Morokoff WJ. Convex entropies and hyperbolicity for general Euler equations. *SIAM Journal on Numerical Analysis* 1998; **35**(6):2117–2127.
24. Lesaint P, Raviart P-A. On a finite element method for solving the neutron transport equation. *Mathematical Aspects of Finite Elements in Partial Differential Equations* (*Proceedings of the Symposium*, *Mathematical Research Center*, University of Wisconsin, Madison, Wisconsin, 1974), Publication No. 33. Mathematical Research Center, University of Wisconsin-Madison. Academic Press: New York, 1974; 89–123.
25. Cockburn B, Lin SY, Shu C-W. TVB Runge–Kutta local projection discontinuous Galerkin finite element method for conservation laws. III. One-dimensional systems. *Journal of Computational Physics* 1989; **84**(1):90–113.
26. Dolejší V, Feistauer M, Schwab C. A finite volume discontinuous Galerkin scheme for nonlinear convection–diffusion problems. *Calcolo* 2002; **39**(1):1–40.

