MATHEMATICS OF COMPUTATION
VOLUME 62, NUMBER 205
JANUARY 1994, PAGES 119-131

MAXIMUM PRINCIPLE ON THE ENTROPY AND SECOND-ORDER KINETIC SCHEMES

BRAHIM KHOBALATTE AND BENOIT PERTHAME

ABSTRACT. We consider kinetic schemes for the multidimensional inviscid gas dynamics equations (compressible Euler equations). We prove that the discrete maximum principle holds for the specific entropy. This fixes the choice of the equilibrium functions necessary for kinetic schemes. We use this property to perform a second-order oscillation-free scheme, where only one slope limitation (for three conserved quantities in 1D) is necessary. Numerical results exhibit stability and strong convergence of the scheme.

INTRODUCTION

We consider the gas dynamics equations in one or two space dimensions,

$$(1)\qquad \begin{cases} \partial_t \rho + \operatorname{div}(\rho u) = 0, \\ \partial_t \rho u_j + \operatorname{div}(\rho u_j u) + \partial_{x_j} p = 0, \qquad j = 1, 2, \\ \partial_t E + \operatorname{div}[(E+p)u] = 0, \end{cases}$$

where $x = (x_1, x_2)$, $u = (u_1, u_2)$, and the total energy $E = \rho|u|^2/2 + \rho T/(\gamma - 1)$ is related to the pressure by the relation $p = \rho T$, $1 < \gamma \leq 2$ in dimension 2, $1 < \gamma \leq 3$ in dimension 1.

It is known that, because of shock waves, an entropy inequality has to be added to (1) (see Lax [3] for instance),

$$(2)\qquad \partial_t \rho S + \operatorname{div}(\rho u S) \leq 0,$$

where the specific entropy can be chosen as

$$(3)\qquad S = \rho / T^{1/(\gamma-1)}.$$

As was proved by Tadmor [10], the combination of (1) and (2) yields that S satisfies the maximum principle

$$(4)\qquad S(x, t+h) \leq \max\{S(y, t);\ |y - x| \leq \|u\|_\infty h\},$$

and, in 1D, Godunov and Lax-Friedrichs schemes preserve this property at the discretized level because they solve exactly the system (1). A reason why (4) should hold is that S satisfies the (meaningless) equation

$$\partial_t S + u \cdot \nabla_x S \leq 0.$$

Received by the editor March 10, 1992 and, in revised form, January 7, 1993.
1991 *Mathematics Subject Classification.* Primary 35L65, 76N10, 65M06, 76P05.
Key words and phrases. Compressible Euler equations, upwind schemes, kinetic schemes, entropy property, second-order schemes.

0025-5718/94 $1.00 + $.25 per page

The purpose of this paper is to show that the property (4) is also satisfied for kinetic schemes in one or two dimensions (we do not consider higher dimensions here, but the extension is straightforward). This requires one to choose the equilibrium function in an appropriate way, in the class introduced by Perthame [6, 7], and to interpret the scheme as a discretization of a transport equation. Then, the property (4) follows from a variational principle. It is remarkable that the appropriate equilibrium function is not the Maxwellian distribution.

It is natural to try to extend this property to second-order accurate schemes. It then appears that a conservative second-order reconstruction, following the method introduced by Van Leer [12], has to increase the specific entropy, and we can only impose the maximum principle up to a second-order error. This is achieved in reconstructing a second-order approximation $\varphi_{i+1/2}$ of $\varphi(x_{i+1/2})$ for $\varphi = \rho$, u, or S. To do so, we use *centered* predictions of $\Delta\varphi$, and we impose both

$$0 \le S_{i+1/2}, \quad S_{i-1/2} \le \max(S_i, S_{i+1}, S_{i-1}), \quad \text{and} \quad \rho_{i+1/2} \ge 0 \tag{5}$$

and the conservation of the quantities $\Psi = \rho$, ρu, E, i.e.,

$$2\Psi_i = \Psi_{i+1/2} + \Psi_{i-1/2}. \tag{6}$$

In practice, to realize (6), we have to relax (5) up to second order.

Numerical tests show that this limitation (5) alone is enough to prevent much of the oscillations in the fully second-order scheme, at least for some classical tests. This is somewhat surprising since nonoscillatory schemes usually require as many limitations as conserved quantities, even though the ENO theory ([2, 9] and the references therein) shows that some flexibility in the reconstruction is possible.

We would like to point out that the conservative entropy inequality (2) is well understood at the discrete level (Osher [5], Tadmor [11]) for general hyperbolic systems. But the maximum principle for the specific entropy (3) is not a consequence of (2) alone, and it holds only for the particular case of gas dynamics (and related systems); therefore, it requires a specific proof.

The paper is organized as follows. In the first section, we consider the 1D case; the 2D case, for a general mesh (rectangular, triangular, dual type), is treated in §2. Then, the second-order scheme, how the limitation (4) is used, and numerical tests are discussed in §3.

1. The 1D case

The general form of a conservative scheme for (1) can be written

$$U_i^{n+1} - U_i^n + \sigma(F_{i+1/2}^n - F_{i-1/2}^n) = 0, \tag{7}$$

where $U_i^n = (\rho_i^n, (\rho u)_i^n, E_i^n)^t$ is the average, on the mesh $(x_{i-1/2}, x_{i+1/2})$ with uniform size Δx, of the vector $(\rho(x, n\Delta t), \rho u(x, n\Delta t), E(x, n\Delta t))$. The time step Δt is related to σ by

$$\sigma = \Delta t/\Delta x. \tag{8}$$

The class of kinetic schemes which we are going to consider is given by a flux splitting

$$F_{i+1/2}^n = F^+(U_i^n) + F^-(U_{i+1}^n), \tag{9}$$

(10)
$$F^+(U) = \rho \int_{v\geq 0} v \left[\left(1, v, \frac{v^2}{2}\right)^t \chi\left(\frac{v-u}{\sqrt{T}}\right) + (0,0,T)^t \zeta\left(\frac{v-u}{\sqrt{T}}\right) \right] dv/\sqrt{T},$$

and F^- is obtained by integrating over $v \leq 0$ rather than $v \geq 0$. This flux is consistent as soon as $F^-(U) + F^+(U) = (\rho u, \rho u^2 + \rho T, (E+p)u)^t$, which is achieved when χ, ζ are even, nonnegative functions satisfying (see [7])

$$(11) \qquad \int_{\mathbb{R}} (1, w^2)\chi(w)\, dx = (1,1), \qquad \int_{\mathbb{R}} \zeta(w)\, dw = \lambda := \tfrac{1}{2}(3-\gamma)/(\gamma-1).$$

Several choices of χ, ζ are possible, but to fit with the general theory, they should yield an entropy in cell property. This is achieved, for instance, by the classical Maxwellians $\alpha e^{-w^2/2}$ as in Deshpande [1] for the single macroscopic entropy $\rho \ln \frac{\rho^{\gamma-1}}{T}$. Other choices, as in [6], are also possible but always associated with one single macroscopic entropy. On the other hand, it is very clear in the proof given in [10] that infinitely many entropies are necessary for the maximum principle. The only choice of χ, ζ which meets this requirement is

$$(12) \qquad \chi(w) = \alpha(1 - w^2/\beta)_+^{\lambda}, \qquad \zeta(w) = \delta^{(\gamma-1)/(\gamma-3)}[\chi(w)]^{(\gamma+1)/(3-\gamma)},$$

where α, β, δ are chosen to satisfy (11), i.e.,

$$(13) \qquad \begin{cases} \alpha = \left[2\sqrt{\beta} \displaystyle\int_0^{\pi/2} \cos^{2/(\gamma-1)} \theta\, d\theta \right]^{-1}, \\ \beta = \displaystyle\int_0^{\pi/2} \cos^{2/(\gamma-1)} \theta\, d\theta \Big/ \int_0^{\pi/2} \sin^2\theta \cos^{2/(\gamma-1)} \theta\, d\theta, \\ \delta^{(\gamma+1)/(\gamma-3)} = \dfrac{\lambda\beta}{2(\lambda+1)} \alpha^{-1/\lambda}. \end{cases}$$

Indeed, with this choice we can obtain a family of singular entropy inequalities. They correspond to the generalized convex functions $\rho\Pi(U)$, where Π is parametrized by $\eta > 0$:

$$(14) \qquad \Pi(U) = \begin{cases} 0 & \text{if } \rho^{\gamma-1}/T < \eta, \\ 1 & \text{if } \rho^{\gamma-1}/T = \eta, \\ +\infty & \text{otherwise}, \end{cases}$$

which are obtained as the limit as p tends to $+\infty$ of the convex entropies $\rho(\rho^{\gamma-1}/T\eta)^p$. The corresponding conservative entropy inequality has a flux splitting form

$$(15) \qquad (\rho\Pi)_i^{n+1} \leq (\rho\Pi)_i^n - \sigma G^+(U_i^n) + \sigma G^-(U_i^n) + \sigma G^+(U_{i-1}^n) - \sigma G^-(U_{i+1}^n),$$

where the entropy fluxes $G^{\pm}$ depend on η,

$$(16) \qquad G^{\pm}(U) = F_\rho^{\pm}(U)\Pi(U),$$

and where $F_\rho^{\pm}$ is the mass flux in (7), (9). It has to be noted that $G^{\pm}$ has the sign $\pm$ and $\rho_i^n - \sigma F_\rho^+(U_i^n) + \sigma F_\rho^-(U_i^n) \geq 0$ with the above CFL condition. Therefore, the right-hand side of (15) is the sum of three nonnegative terms depending, respectively, on U_i^n, U_{i-1}^n, U_{i+1}^n. We use the convention that the right-hand side is $+\infty$ whenever one of these three terms is $+\infty$.

We can now state our main result.

Theorem 1. *With the choice* (12), *the kinetic scheme* (7)–(10) *satisfies*
(i) $\rho_i^{n+1} \geq 0$, $T_i^{n+1} \geq 0$ *whenever* ρ_i^n, $T_i^n \geq 0$,
(ii) *the conservative entropy inequalities* (15) *for any* $\eta > 0$,
(iii) *the maximum principle on the specific entropy*

$$S_i^{n+1} := \rho_i^{n+1}/(T_i^{n+1})^{1/(\gamma-1)} \leq \max(S_i^n, S_{i+1}^n, S_{i-1}^n) \tag{17}$$

under the CFL condition $(|u_i^n| + \sqrt{\beta T_i^n})\sigma \leq 1$ *for all* i.

Remarks. (1) For $\gamma = 1.4$, we find $\beta = 7$, and thus our CFL condition is stricter then the classical one. But in practice we can use the classical one.

(2) The theory of kinetic formulation of the isentropic system, developed in Lions, Perthame, and Tadmor [4], requires the same χ-function in (12), but there, the entropies are much simpler than those developed in the proof below. The exact relation between the two theories is unclear to the authors.

Proof of Theorem 1. *First step*: *The kinetic level.* We first introduce the discretized transport equations

$$\bar{f}_i(v) - f_i^n(v) + \sigma[v_+ f_i^n(v) - v_- f_{i+1}^n(v) - v_+ f_{i-1}^n(v) + v_- f_i^n(v)] = 0, \tag{18}$$

$$\bar{g}_i(v) - g_i^n(v) + \sigma[v_+ g_i^n(v) - v_- g_{i+1}^n(v) - v_+ g_{i-1}^n(v) + v_- g_i^n(v)] = 0, \tag{19}$$

where $v_+ = \max(0, v)$, $v_+ - v_- = v$, and

$$f_i^n(v) = \rho_i^n \chi[(v - u_i^n)/\sqrt{T_i^n}], \qquad g_i^n(v) = \rho_i^n \sqrt{T_i^n}\, \zeta[(v - u_i^n)/\sqrt{T_i^n}]. \tag{20}$$

As usual [1, 7], the finite difference scheme (7), (10) is deduced by multiplying (18) by the vector $(1, v, v^2/2)^t$ and adding to it (19) multiplied by $(0, 0, 1)^t$ and integrating against dv.

Indeed, this clearly follows from the identities

$$\begin{aligned} U_i^n &= \int_{\mathbb{R}} \left(f_i^n, v f_i^n, \frac{v^2}{2} f_i^n + g_i^n \right)^t dv, \\ U_i^{n+1} &:= \int \left(\bar{f}_i, v \bar{f}_i, \frac{v^2}{2} \bar{f}_i + \bar{g}_i \right)^t dv, \\ F^{\pm}(U_i^n) &= \pm \int v_{\pm} \left(f_i^n, v f_i^n, \frac{v^2}{2} f_i^n + g_i^n \right)^t dv, \end{aligned} \tag{21}$$

which follow from (10) and the consistency relations (11). Now, we have $\bar{f}_i \geq 0$, $\bar{g}_i \geq 0$, under the condition $\sigma|v| \leq 1$ for all v such that $f_i^n(v) \neq 0$, and this is exactly the CFL condition of Theorem 1. This proves (i).

Second step: *The maximum principle.* We notice that $h = (f^{\gamma+1} g^{\gamma-3})^{1/(\gamma-1)}$ is a convex function of f, g. Since $\bar{f}_i$ and $\bar{g}_i$ are also convex combinations of f_i^n, f_{i+1}^n, f_{i-1}^n and g_i^n, g_{i+1}^n, g_{i-1}^n (whenever σ satisfies the CFL condition), we thus have

$$\bar{h}_i \leq h_i^n(1 - \sigma v_+ - \sigma v_-) + h_{i+1}^n \sigma v_- + h_{i-1}^n \sigma v_+. \tag{22}$$

Now ,with the choice (12) of χ, ζ, the function h is just given by

$$h_i^n = \delta (S_i^n)^2 \mathbf{1}_{\{|v-u_i^n|^2 \leq \beta T_i^n\}}, \tag{23}$$

and thus we obtain

$$\bar{h}_i \leq \Sigma := \delta \max(S_i^n, S_{i+1}^n, S_{i-1}^n)^2. \tag{24}$$

At this level we need the following lemma, which is similar to those of [6] and whose proof uses simple calculus of variation and is thus skipped.

Lemma 2. *Let* $e = \rho\tau/(\gamma-1)$ *be such that*

$$e = \min\left\{\int_{\mathbb{R}}\left[\frac{v^2}{2}f(v)+g(v)\right]dv\,;\ f(v)\geq 0,\ g(v)\geq 0,\right.$$
$$\left.\int_{\mathbb{R}}(1,v)f(v)\,dv=(\rho,0)\ (\rho\geq 0),\ f^{\gamma+1}g^{\gamma-3}\leq\Sigma^{\gamma-1}\right\}.$$

Then $\rho/\tau^{1/(\gamma-1)} = \sqrt{\Sigma/\delta}$, *and the minimum is uniquely achieved by* $f = \frac{\rho}{\sqrt{\tau}}\chi(\frac{v-u}{\sqrt{\tau}})$, $g = \rho\sqrt{\tau}\zeta(\frac{v-u}{\sqrt{\tau}})$. □

Returning to the proof of Theorem 1, we apply the lemma with $f = \bar{f}_i(v-u_i^{n+1})$, $g = \bar{g}_i(v-u_i^{n+1})$, so that the constraints in the minimization problem are realized with $\rho = \rho_i^{n+1}$ and Σ given in (24). We thus have

$$(\gamma-1)\int\left(\frac{v^2}{2}f+g\right)=\rho_i^{n+1}T_i^{n+1}\geq e(\gamma-1)=\rho_i^{n+1}\left(\rho_i^{n+1}/\sqrt{\frac{\Sigma}{\delta}}\right)^{\gamma-1},$$

which exactly means $S_i^{n+1}\geq\sqrt{\Sigma/\delta}$ and (iii) is proved.

Third step: *Entropy inequality.* As in the second step, let us introduce, for a fixed positive number η and for $p\geq 1$, the function

$$k = f\left[\frac{(f^{\gamma+1}g^{\gamma-3})^{1/(\gamma-1)}}{\eta^2}\right]^p.$$

Since it is a convex function of f and g, we also have

$$\bar{k}_i(v)\leq k_i^n(v)(1-\sigma v_+-\sigma v_-)+k_{i+1}^n(v)\sigma v_-+k_{i-1}^n(v)\sigma v_+.$$

We need now the following lemma.

Lemma 3. *The minimization problem*

$$\min\left\{\int_{\mathbb{R}}f(f^{\gamma+1}g^{\gamma-3})^{p/(\gamma-1)}\,dv\,;\ f\geq 0,\ g\geq 0,\right.$$
$$\left.\int_{\mathbb{R}}(1,v)f\,dv=(\rho,0),\ \int_{\mathbb{R}}\frac{v^2}{2}f+g=\rho T/(\gamma-1)\right\}\tag{25}$$

admits a unique minimum

$$F_p=\frac{\rho\alpha_p}{\sqrt{T}}\left(1-\frac{v^2}{\beta_pT}\right)_+^{(1+2p\lambda)/2p},\qquad G_p=T\delta_pF_p\cdot\left(1-\frac{v^2}{\beta_pT}\right)_+,$$

where α_p, β_p, *and* δ_p *are such that the constraints in* (25) *are satisfied.* □

Again we skip the proof of this lemma which consists in writing the Euler-Lagrange equations associated with (25). As before, we use this lemma with $\rho=\rho_i^{n+1}$, $T=T_i^{n+1}$, $f=\bar{f}_i(v-u_i^{n+1})$, $g=\bar{g}_i(v-u_i^{n+1})$, and the corresponding minimizers F_p, G_p thus satisfy

$$\int_{\mathbb{R}}F_p\left[\frac{(F_p^{\gamma+1}G_p^{\gamma-3})^{1/(\gamma-1)}}{\eta^2}\right]^p\leq\int_{\mathbb{R}}\bar{k}_i(v)\,dv$$
$$\leq\int_{\mathbb{R}}[k_i^n(v)(1-\sigma v_+-\sigma v_-)+k_{i+1}^n(v)\sigma v_-+k_{i-1}^n(v)\sigma v_+]\,dv.\tag{26}$$

Now we let p go to $+\infty$ and we find exactly the entropy inequality (15), since the right-hand side of (26) goes to $\rho_i^{n+1}\Pi(U_i^{n+1})$ and

$$\int_{\mathbb{R}} k_i^n(v)(1-\sigma v_+ - \sigma v_-)\,dv \to (\rho_i^n - \sigma F_\rho^+(U_i^n) + \sigma F_\rho^-(U_i^n))\Pi(U_i^n),$$

$$\int k_{i\pm1}^n(v) v_\pm\,dv \to \pm F_\rho^\pm(U_i^n)\Pi(U_i^n).$$

This concludes the proof of Theorem 1. □

We end this section with some remarks on the entropy. First, notice that the choice

$$\chi = \alpha_p\left(1-\frac{w^2}{\beta_p}\right)_+^{(1+2p\lambda)/2p}, \qquad \varphi = \delta_p\chi\cdot\left(1-\frac{w^2}{\beta_p}\right)$$

in the scheme (7)–(10) leads to an entropy inequality (for a regular entropy now) of the form (15) with

$$\Pi(U) = \mu_p\rho(\rho/T^{1/(\gamma-1)})^{2p}, \qquad G^+ = \int_{v\geq 0} vF_p(F_p^{\gamma+1}G_p^{\gamma-3})^{p/(\gamma-1)}\,dv,$$

with F_p, G_p defined in Lemma 3 and some appropriate constant μ_p. The proof of this, as well as the proof of (ii) in Theorem 1, follows in fact that of [6], but here we have a more general approach dealing with two functions f, g rather than two kinetic variables v, I as in Deshpande [1]. Also we would like to emphasize that an exact entropy inequality is necessary to get a maximum principle on the specific entropy, and it is an open question if the proofs of Osher [5] or Tadmor [11] could be extended to get, for Roe or Osher schemes, a maximum principle, or for kinetic schemes the entropy inequality.

2. The 2D case

We show here that our results can be naturally extended to the 2D equations discretized on an unstructured mesh. Our motivations and notations follow those introduced in Perthame and Qiu [8].

Consider a grid as shown in Figure 1, where cells C_i have $L(i)$ edges $E_1, \dots, E_L$ ($L = 3$ for triangles, 4 for rectangles, and depends on i for dual type grids). We call ν_l the unit outward normal to E_l, $|E_l|$ the length of E_l, $|C_i|$ the area of C_i, and $j(l)$ the index of the cell $C_{j(l)}$ neighboring C_i along E_l ($j(l)$ also depends on i, but we omit this dependence for simplicity).

We now set $U_i^n = (\rho, \rho u_1, \rho u_2, E)_i^n$ and we consider numerical schemes for the equations (1) of the form

$$\text{(27)}\quad \begin{cases} U_i^{n+1}|C_i| = U_i^n|C_i| - \Delta t\displaystyle\sum_{l=1}^{L(i)}|E_l|F_l\cdot\nu_l, \\ F_l\nu_l = F^+(U_i^n,\nu_l) + F^-(U_{j(l)}^n,\nu_l), \\ F^\pm(U,\nu) = \pm\rho\displaystyle\int_{\mathbb{R}^2}(v\cdot\nu)_\pm\left[\left(1, v, \frac{|v|^2}{2}\right)^t\chi\left(\frac{v-u}{\sqrt{T}}\right)\right. \\ \qquad\qquad \left. + (0,0,T)^t\varphi\left(\frac{v-u}{\sqrt{T}}\right)\right]dv/T. \end{cases}$$

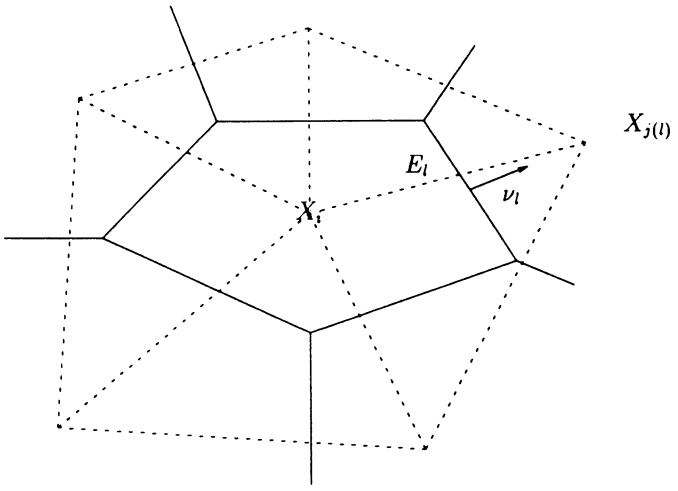

FIGURE 1. An example of unstructured mesh

The consistency relations are now that the nonnegative even functions χ, φ satisfy

$$\int_{\mathbb{R}^2} (1, w_k w_l)\chi(w)\,dw = (1, \delta_{kl}), \qquad \int_{\mathbb{R}^2} \varphi(w)\,dw = \lambda := \frac{2-\gamma}{2(\gamma-1)}; \tag{28}$$

the general value of λ is $(2 + N - N\gamma)/2(\gamma - 1)$ in N dimensions, $1 < \gamma \leq (N+2)/N$. We now choose

$$\chi(w) = \alpha\left(1 - \frac{|w|^2}{\beta}\right)_+^{\lambda}, \qquad \varphi(w) = \delta\left(1 - \frac{|w|^2}{\beta}\right)_+^{\lambda+1}, \tag{29}$$

where again α, β, and δ are the only constants which yield (28). We obtain the following theorem.

Theorem 4. *The scheme* (27)–(29) *satisfies*

(i) $\rho_i^{n+1} \geq 0$, $T_i^{n+1} \geq 0$ *whenever* $\rho_i^n \geq 0$, $T_i^n \geq 0$,

(ii) *the singular family of conservative entropy inequalities*

$$\begin{aligned}(\rho\Pi)_i^{n+1}|C_i| \leq &\left[(\rho\Pi)_i^n|C_i| - \Delta t\sum_l |E_l|G^+(U_i^n, \nu_l)\right] \\ &- \Delta t\sum_l |E_l|G^-(U_{j(l)}^n, \nu_l),\end{aligned} \tag{30}$$

(iii) *the maximum principle on the specific entropy*

$$S_i^{n+1} \leq \max(S_i^n, S_{j(1)}^n, \ldots, S_{j(l)}^n), \tag{31}$$

under the CFL condition $\Delta t\sum_l |E_l|(|u_i^n| + \sqrt{\beta T_i^n}) \leq |C_i|$ *for all* i. □

In (30), Π is defined as before by formula (14) and

$$G^{\pm}(U, \nu_l) = F_\rho^{\pm}(U, \nu_l)\Pi(U). \tag{32}$$

Notice that the notation $\pm$ here differs from that of the 1D case, because of the introduction of the normals, and because we have no natural orientation for

an unstructured grid. Again, the right-hand side of (30) is composed of $L+1$ nonnegative terms, and we use the convention that it is $+\infty$ whenever one of those $L+1$ terms is $+\infty$.

We skip the proof of Theorem 4, which is a straightforward extension of that in §1. The only new point is to introduce, following [8], the kinetic scheme

$$\bar{f}_i(v)|C_i| = f_i^n(v)\left(|C_i| - \Delta t \sum_l (v\cdot\nu_l)_+|E_l|\right) + \sum_l (v\cdot\nu_l)_-|E_l| f_{j(l)}^n(v)$$

(with similar formulae acting on g), together with the conditions

$$f_i^n = \rho_i^n \chi\left(\frac{v-u_i^n}{\sqrt{T_i^n}}\right)\Big/ T_i^n, \qquad g_i^n = \rho_i^n \varphi\left(\frac{v-u_i^n}{\sqrt{T_i^n}}\right). \tag{33}$$

Then the exponents in (29) are uniquely recovered by the requirement that, $(fg^{\gamma-2})^{1/(\gamma-1)}$ being homogeneous to $S\mathbf{1}_{\{\dots\}}$, the minimum in Lemma 2, with the constraint $fg^{\gamma-2} \le \Sigma^{\gamma-1}$, be achieved for our choice of χ, φ in (29).

3. Minimal limitations for second-order schemes

We return to the 1D case and consider second-order schemes in space and time obtained using slope reconstruction (see [12]) together with a Runge-Kutta scheme in time. Our purpose is to show that only few oscillations appear (see Figure 2) with the above kinetic scheme, using *centered* slopes on ρ, u, T and limited so as to preserve the nonnegativity of ρ and T as in [7]. Moreover, an additional limitation ensuring the maximum principle on the specific entropy up to second order is enough to damp all oscillations (see Figure 3 on p. 128). This amounts to a single limitation of min-mod type, combining $D\rho$, DT for three quantities. The results are more accurate than with a min-mod limitation on the three quantities, as is shown in Figure 4 (p. 128).

3.1. **The second-order scheme.** Denote by $U_i^{n,\pm}$ the inner approximations in the mesh i of $U^n(x_{i+1/2} \pm \Delta x/2)$. The construction of ΔU is discussed later.

Then, the second-order, in space and time, scheme we use is

$$\begin{cases} \tilde{U}_i - U_i^n + \sigma(F^+(U_i^{n,+}) + F^-(U_{i+1}^{n,-}) - F^+(U_{i-1}^{n,+}) - F^-(U_i^{n,-})) = 0, \\ \hat{U}_i - \tilde{U}_i + \sigma(F^+(\tilde{U}_i^+) + F^-(\tilde{U}_{i+1}^-) - F^+(\tilde{U}_{i-1}^+) - F^-(\tilde{U}_i^-)) = 0, \\ U_i^{n+1} = (U_i^n + \hat{U}_i)/2. \end{cases} \tag{34}$$

This particular Runge-Kutta scheme will preserve nonnegativity, while we would be unable to prove it with other schemes. The reason is that $\tilde{U}$ and $\hat{U}$ will have nonnegative density and temperature, and then a convex combination of them, as U_i^{n+1}, will also, since ρ and ρT are concave functions of U.

3.2. **Nonnegativity of ρ, T and limitations.** We now prove that the scheme (34), with light limitations on a centered prediction of the derivatives, preserves nonnegativity. We use the variables ρ, u, and $\Sigma = \rho^\gamma/p = S^{\gamma-1}$, and set

$$\begin{cases} \Delta\rho_i = \operatorname{sgn}(\rho_{i+1} - \rho_{i-1})\min(|\rho_{i+1} - \rho_{i-1}|/4,\ \rho_i), \\ \Delta u_i = \operatorname{sgn}(u_{i+1} - u_{i-1})\min(|u_{i+1} - u_{i-1}|/4,\ \sqrt{T_i/(\gamma-1)}), \\ \Delta\Sigma_i = \operatorname{sgn}(\Sigma_{i+1} - \Sigma_{i-1})\min(|\Sigma_{i+1} - \Sigma_{i-1}|/4,\ \Sigma_i/4). \end{cases} \tag{35}$$

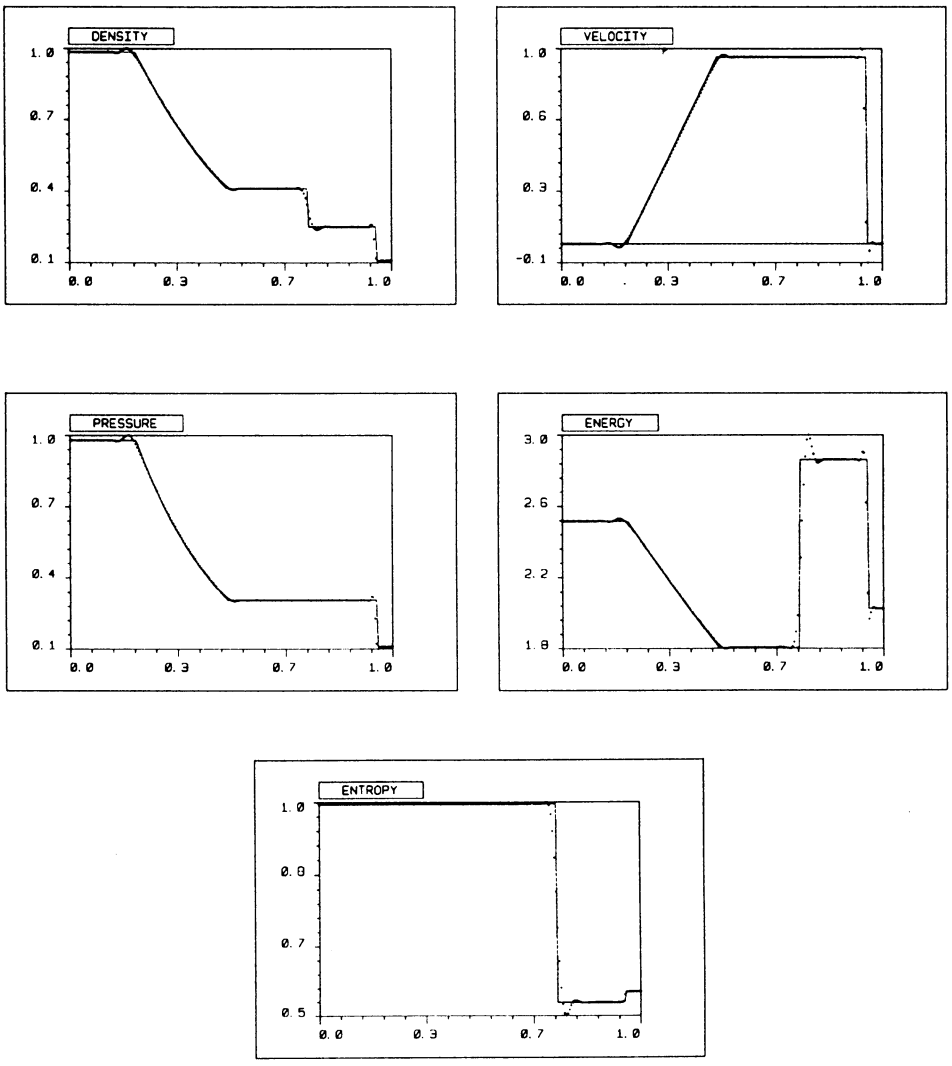

FIGURE 2. Sod shock tube, 200 points, second-order scheme with centered, nonlimited slopes

Then, following the idea introduced in [7], we set (dropping the exponent n)

$$\rho_i^{\pm} = \rho_i \pm \Delta\rho_i\,, \quad u_i^{\pm} = \underline{u}_i \pm \Delta u_i\,, \quad \Sigma_i^{\pm} = \underline{\Sigma}_i \pm \Delta\Sigma_i\,, \tag{36}$$

where $\underline{u}_i$ and $\underline{\Sigma}_i$ are computed for conservation of momentum and energy by

$$\rho_i^+ u_i^+ + \rho_i^- u_i^- = 2\rho_i u_i\,, \qquad \frac{\rho_i^+ u_i^{+2}}{2} + \frac{\rho_i^{+\gamma}}{\Sigma_i^+(\gamma-1)} + \frac{\rho_i^- u_i^{-2}}{2} + \frac{\rho_i^{-\gamma}}{\Sigma_i^-(\gamma-1)} = 2E_i.$$

This is readily achieved for the *second-order modifications* of u_i, Σ_i given by

$$\underline{u}_i = u_i - \mu\Delta u_i\,, \qquad \mu = \Delta\rho_i/\rho_i\,, \tag{37}$$

and with $\underline{\Sigma}_i$ being the largest root of the polynomial

$$C\underline{\Sigma}^2 - (A+B)\underline{\Sigma} + (B-A)\Delta\Sigma_i - C\Delta\Sigma_i^2 = 0\,, \tag{37'}$$

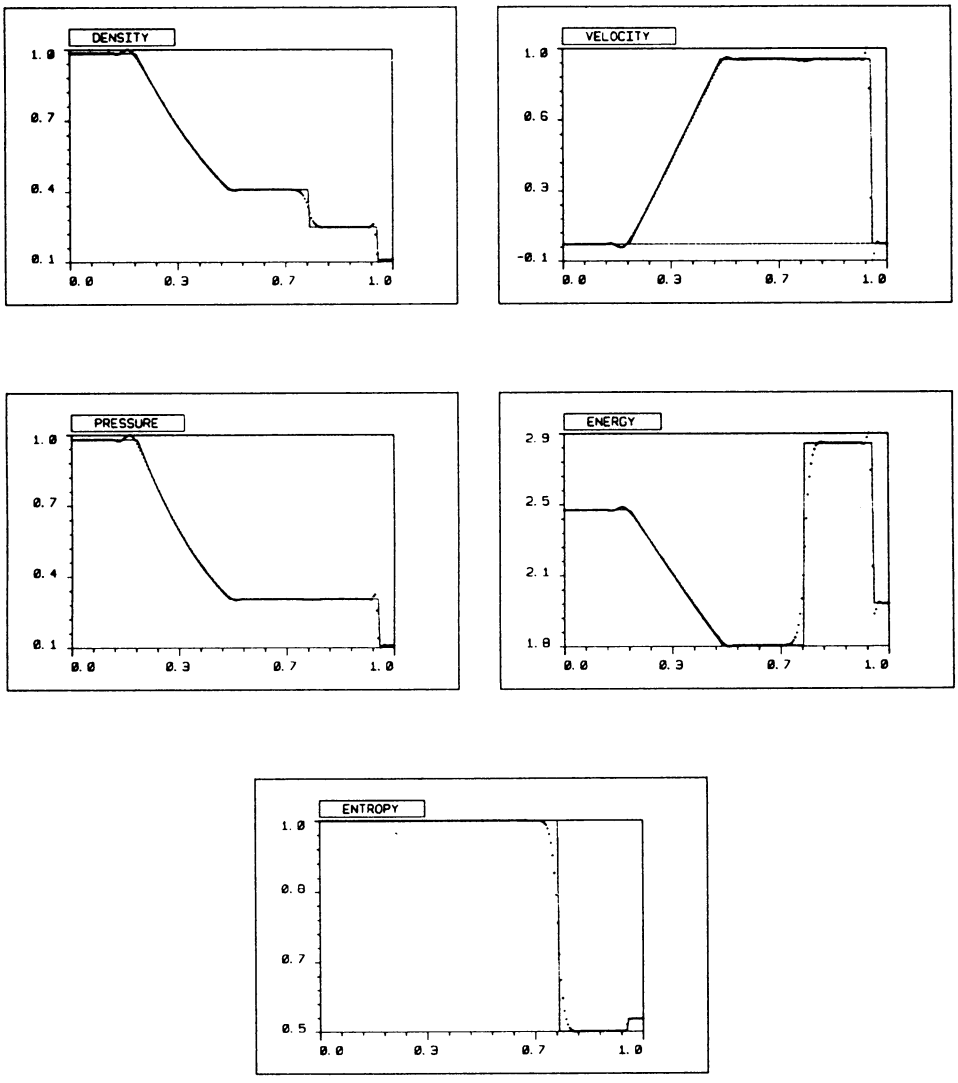

FIGURE 3. Sod shock tube, 200 points, second-order scheme with the only limitation on the entropy (6)

average order of the scheme type of reconstruction	energy	density	velocity
min-mod	.77	.87	.96
centered	.79	.87	.99
centered + limitation on entropy	.87	.91	1.00

FIGURE 4. L^1 error obtained for energy, density, and velocity computed with three types of construction of the slopes

where

$$A = (\rho_i - \Delta\rho_i)^\gamma, \quad B = (\rho_i + \Delta\rho_i)^\gamma, \quad C = 2\rho_i T_i + \tfrac{1}{2}(\mu^2 - 1)(\Delta u_i)^2/(\gamma - 1).$$

It is indeed easy to check that (37′) has a nonnegative discriminant, whatever $\Delta\Sigma_i$ is. Also, $\rho_i^{\pm} \geq 0$, and thus we obtain

Theorem 5. *The scheme* (34)–(37′) *preserves the positivity of* ρ *and* T, *under the CFL condition* $(|u_i^{n,\pm}| + \sqrt{\beta T_i^{n,\pm}})\sigma \leq 1/2$.

Proof. First of all, we show that $\tilde{U}$ has nonnegative density and temperature.

We use the following kinetic scheme:

$$\bar{f}_i - (f_i^{n,+} + f_i^{n,-})/2 + \sigma[v^+ f_i^{n,+} - v^- f_{i+1}^{n,-} - v^+ f_{i-1}^{n,+} + v^- f_i^{n,-}] = 0, \tag{38}$$

with the same equation for g, and

$$f_i^{n,\pm} = \rho_i^{n,\pm}\chi\left(\frac{(v - u_i^{n,\pm})}{\sqrt{T_i^{n,\pm}}}\right)\Big/\sqrt{T_i^{n,\pm}}, \qquad g_i^{n,\pm} = \rho_i^{n,\pm}\xi(\cdots)\sqrt{T_i^{n,\pm}}.$$

We claim that (38) yields, using the same combination as in §1, the scheme (34). Indeed, we just have to check that the second term of (38) gives U_i^n; this is true since

$$\int_{\mathbb{R}} [(1, v, v^2/2)(f_i^{n,+} + f_i^{n,-}) + (0, 0, 1)(g_i^{n,+} + g_i^{n,-})]\,dv$$
$$= U_i^{n,+} + U_i^{n,-} = 2U_i^n,$$

thanks to (36). Now to check the nonnegativity of $\bar{f}$, we need $1/2 \geq \sigma|v|$, which gives the CFL number of one half. □

At this level, the limitations involved in (35) are very light, but give few oscillations (Figure 2). Let us go one step further and consider the maximum principle on the entropy.

3.3. **Limitation by maximum of entropy.** We still denote $\Sigma = \rho^\gamma/p$ and we now require to have the maximum principle on S or Σ. Therefore, we impose the following additional limitation in (35), (37′):

$$|\Delta\Sigma_i| \leq \max(\Sigma_i, \Sigma_{i+1}, \Sigma_{i-1}) - \Sigma_i. \tag{39}$$

This implies a maximum principle on $\Sigma_{i\pm 1/2}$ up to a second-order term, because $\Sigma_{i\pm 1/2}$ is given through Σ_i and not Σ_i in (36). It seems impossible to perform second-order reconstruction satisfying the conservativity requirements (36) and the maximum principle on Σ or S.

In Figure 3, we show the numerical results obtained coupling the scheme (34)–(37) to the additional limitation (39); the oscillations around the contact discontinuity are damped completely and only an overshoot remains before and after the shock waves. This is also true for other tests problems: Lax shock tube, blast waves problem.

In order to test other types of problems, we have run our method on the slow shock proposed in [13]. Again, an overshoot appears, which is immediately damped while an oscillation usually propagates with first-order solvers [13]. We have also tested the shock tube proposed by Einfeldt, Munz, Roe, and Sjögreen

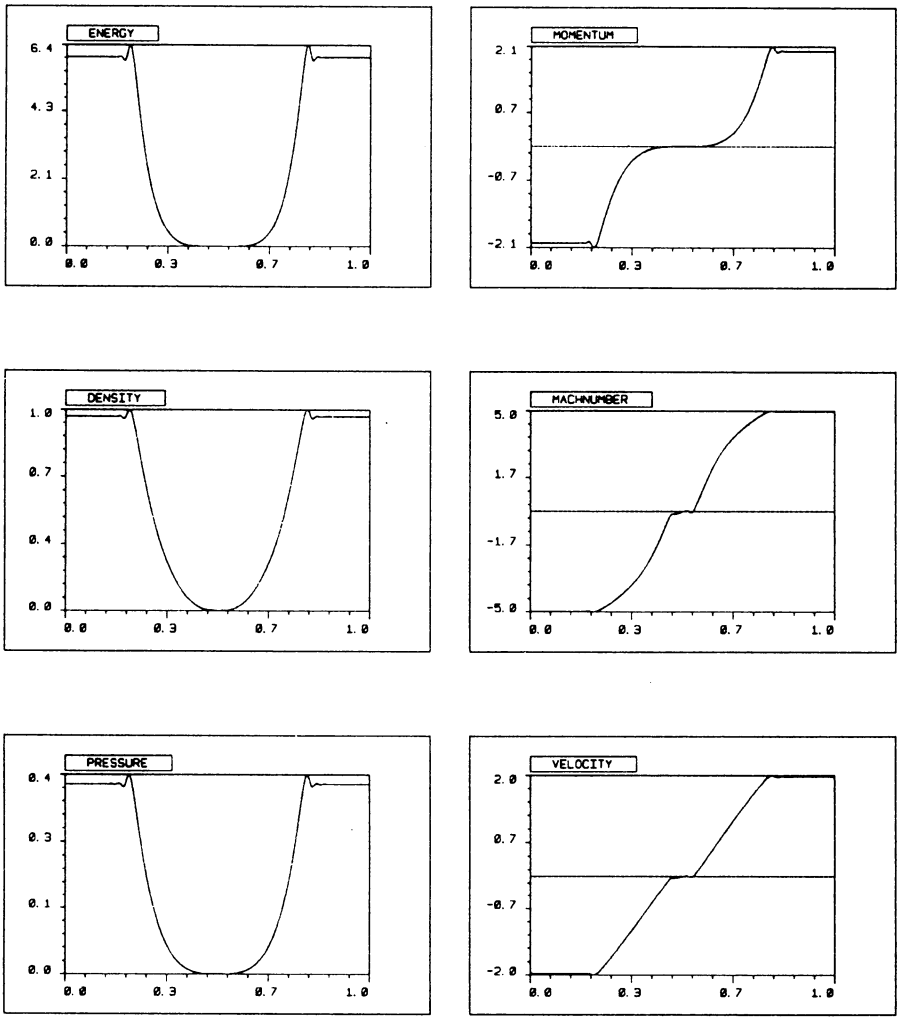

FIGURE 5. 200 points, second-order scheme with the only limitation on the entropy (6); problem $(1-2-0-3)$ at time $t = 0.1$ of [14]

[14], where a zero temperature point arises. As asserted by the mathematical study, the scheme is stable, and we obtain indeed second-order results (compare Figure 5 and [14]).

Remark. The choice of ρ and u as primitive variables for the reconstruction is somewhat arbitrary here. Only Σ plays a particular role. Let us only point out that they lead to particularly simple computations, and they are natural in the kinetic schemes.

ACKNOWLEDGMENT

The second author wishes to thank R. Sanders for valuable discussions.

BIBLIOGRAPHY

1. S. Deshpande, *A second order accurate, kinetic theory based, method for inviscid compressible flows*, NASA Langley Technical Paper no. 2613, 1986.
2. A. Harten, B. Engquist, S. Osher, and S. R. Chakravarthy, *Uniformly high order accurate essentially non-oscillatory schemes.* III, J. Comput. Phys. **71** (1987), 231–303.
3. P. D. Lax, *Hyperbolic systems of conservation laws and the mathematical theory of shock waves*, CBMS Regional Conf. Ser. in Appl. Math., vol. 11, SIAM, Philadelphia, PA, 1972.
4. P. L. Lions, B. Perthame, and E. Tadmor, *Kinetic formulation of the isentropic gas dynamics and P-system*, preprint.
5. S. Osher, *Riemann solvers, the entropy condition, and difference approximations*, SIAM J. Numer. Anal. **21** (1984), 217–235.
6. B. Perthame, *Boltzmann type schemes for gas dynamics and the entropy property*, SIAM J. Numer. Anal. **27** (1990), 1405–1421.
7. ——, *Second-order Boltzmann schemes for compressible Euler equations in one and two space dimensions*, SIAM J. Numer. Anal. **29** (1992), 1–19.
8. B. Perthame and Y. Qiu, *A new variant of Van Leer's method for multidimensional systems of conservation laws*, INRIA report no. 1562.
9. C.-W. Shu and S. Osher, *Efficient implementation of essentially non-oscillatory shock-capturing schemes.* II, J. Comput. Phys. **83** (1989), 32–78.
10. E. Tadmor, *A minimum entropy principle in the gas dynamics equations*, Appl. Numer. Math. **2** (1986), 211–219.
11. ——, *The numerical viscosity of entropy stable schemes for systems of conservation laws.* I, Math. Comp. **49** (1987), 91–103.
12. B. Van Leer, *Towards the ultimate conservative difference scheme.* V, *A second order sequel of Godunov's method*, J. Comput. Phys. **32** (1979), 101–136.
13. T. W. Roberts, *The behavior of flux difference splitting schemes near slowly moving shock waves*, J. Comput. Phys. **90** (1990), 141–160.
14. B. Einfeldt, C. D. Munz, P. L. Roe, and B. Sjögreen, *On Godunov-type methods near low densities*, J. Comput. Phys. **92** (1991), 273–295.

INRIA, CENTRE DE ROCQUENCOURT, PROJET MENUSIN BP. 105, 78173 LE CHESNAY CEDEX, FRANCE

LABORATOIRE D'ANALYSE NUMÉRIQUE, UNIVERSITÉ PARIS 6, TOUR 55-56, 5ième ÉTAGE, 4, PLACE JUSSIEU, F75252 PARIS CÉDEX 05 FRANCE