Computers and Fluids, 230 (22), 2011, pp. 8010–8027.

Positivity-Preserving Flux-Limited Method for Compressible Fluid Flow

Bernard Parent*

The extension of a flux discretization method to second-order accuracy can lead to some difficulties in maintaining positivity preservation. While the MUSCL-TVD scheme maintains the positivity preservation property of the underlying 1st-order flux discretization method, a flux-limited-TVD scheme does not. A modification is here proposed to the flux-limited-TVD scheme to make it positivity-preserving when used in conjunction with the Steger-Warming flux vector splitting method. The proposed algorithm is then compared to MUSCL for several test cases. Results obtained indicate that while the proposed scheme is more dissipative in the vicinity of contact discontinuities, it performs significantly better than MUSCL when solving strong shocks in hypersonic flowfields: the amount of pressure overshoot downstream of the shock is minimized and the time step can be set to a value typically two or three times higher. While only test cases solving the one-dimensional Euler equations are here presented, the proposed scheme is written in general form and can be extended to other physical models.

1. Introduction

ADDITIONALLY to conservation, to monotonicity, and to the non-violation of the second-law of thermodynamics, a desirable characteristic of numerical methods for compressible flow is positivity preservation. Positivity preservation refers to the capability of an algorithm to guarantee the positivity of the internal energy and the density on all nodes as the solution advances in time, provided the initial density and internal energy are positive.

Positivity preservation of the internal energy can be especially problematic when the flow speed is in the hypersonic range with the kinetic energy composing the quasi-totality of the total energy. Indeed, since the internal energy is determined as the difference between the total and kinetic energy, a small over-estimation of the kinetic energy (or under-estimation of the total energy) by the flow solver can lead to a negative internal energy. Even with proper “clipping” of the pressure and temperature in order to keep the internal energy positive, it can then be very difficult or even impossible for the flow solver to continue the time integration process because the solution has been directed towards non-physical states.

Much attention has hence been given to determine whether the most commonly used flux discretization methods are positivity-preserving in order to assess the robustness of existing CFD codes for high speed flow. In Refs. [1, 2], it is shown that while the Godunov scheme [3], the Van Leer scheme [4] and the Steger-Warming scheme [5] are positivity-preserving under a CFL-like condition, the Roe scheme [6] and the HLL schemes [7] are not. To make the Roe and HLL schemes positivity-preserving, some modifications to the original discretization stencils are proposed in Refs. [8, 1].

The extension of the flux discretization method to second-order accuracy can also lead to some difficulties in maintaining positivity preservation, although some success has been reported when using Total Variation Diminishing (TVD) algorithms. Indeed, a MUSCL-TVD scheme [9] is shown in Ref. [10] to maintain the positivity preservation property of the underlying first-order method, and an attempt is made in Ref. [11] to show that a minmod-flux-limited-TVD scheme can also maintain positivity at least when the limiter is asymmetric (that is, the limiter achieves second-order accuracy for rightward traveling waves, but first-order accuracy for leftward traveling waves). However, for a symmetric minmod flux limiter achieving second-order accuracy for both leftward and rightward traveling waves, numerical tests indicate that the flux-limited-TVD scheme does not generally maintain positivity.

Because a flux-limited scheme is advantaged over MUSCL by not requiring a reconstruction of the vector of conserved variables (hence resulting in reduced computing effort), a novel modification to the flux-limiters is here presented to attain

*Faculty Member, Dept. of Aerospace Engineering, Pusan National University, Busan 609-735, Korea, http://www.bernardparent.com.

positivity preservation. A comparison is then made between the proposed algorithm and MUSCL for several test cases. To enable a fair comparison, the same underlying limiters (minmod) and the same underlying first-order monotonic method (Steger Warming Flux-Vector-Splitting) are used in both cases.

2. Physical Model

While the discretization methods outlined hereafter are written in general form and can be applied to multidimensional viscous and reacting flows, it is deemed sufficient at this stage to compare the performance of the different numerical schemes in solving the one-dimensional inviscid and non-reacting Euler equations for a perfect gas:

$$\frac{\partial U}{\partial t}+\frac{\partial F}{\partial x}=0 \tag{1}$$

In the latter, U is the vector of conserved variables:

$$U=\left[\begin{array}{c}\rho\\ \rho u\\ \rho c_v T+\dfrac{\rho u^2}{2}\end{array}\right]=\left[\begin{array}{c}\rho\\ \rho u\\ \dfrac{\rho a^2}{\gamma(\gamma-1)}+\dfrac{\rho u^2}{2}\end{array}\right] \tag{2}$$

and F is the convective flux vector:

$$F=\left[\begin{array}{c}\rho u\\ \rho u^2+P\\ \rho u c_p T+\dfrac{\rho u^3}{2}\end{array}\right]=\left[\begin{array}{c}\rho u\\ \rho u^2+\rho a^2/\gamma\\ \dfrac{\rho u a^2}{\gamma-1}+\dfrac{\rho u^3}{2}\end{array}\right] \tag{3}$$

where ρ, a, u, T, and P are the density, sound speed, velocity, temperature, and pressure, while c_v, c_p, and γ are the specific heats and the specific heat ratio. Defined as $\partial F/\partial U$, the convective flux Jacobian corresponds to:

$$A=\left[\begin{array}{ccc}0 & 1 & 0\\ \dfrac{\gamma-3}{2}u^2 & -(\gamma-3)u & \gamma-1\\ \dfrac{\gamma-2}{2}u^3-\dfrac{ua^2}{\gamma-1} & \left(\dfrac{3}{2}-\gamma\right)u^2+\dfrac{a^2}{\gamma-1} & \gamma u\end{array}\right] \tag{4}$$

The convective flux Jacobian can also be written as $A=L^{-1}\Lambda L$ with Λ the eigenvalue matrix equal to $[u,\ u+a, u-a]^{\rm D}$ and L^{-1} the right eigenvectors matrix equal to:

$$L^{-1}=\left[\begin{array}{ccc}1 & 1 & 1\\ u & u+a & u-a\\ \dfrac{u^2}{2} & \dfrac{u^2}{2}+ua+\dfrac{a^2}{\gamma-1} & \dfrac{u^2}{2}-ua+\dfrac{a^2}{\gamma-1}\end{array}\right] \tag{5}$$

3. Rule of the Positive Coefficients

The rule of the positive coefficients [12] is an approach widely used to determine the time step restrictions and to craft discretization stencils that maintain positivity when solving the heat equation or the scalar equations of incompressible flow. It can be summarized as follows. Consider a discrete equation solving the scalar property $u_{\rm A}$ in terms of the properties on the nearby nodes $u_{\rm B}$, $u_{\rm C}$, etc:

$$c_{\rm A}u_{\rm A}=c_{\rm B}u_{\rm B}+c_{\rm C}u_{\rm C}+c_{\rm D}u_{\rm D}+... \tag{6}$$

Then, as long as the coefficients $c_{\rm A}$, $c_{\rm B}$, etc are all positive, and as long as the flow property u on all the nearby nodes is positive, $u_{\rm A}$ will also be positive.

Interestingly, a similar rule also exists for the coupled equations of compressible flow. The rule can be summarized as follows. Consider the following equation determining the vector $U_{\rm A}$ as a function of the vector of conserved variables on the nearby nodes ($U_{\rm B}$, $U_{\rm C}$, etc):

$$C_{\rm A} U_{\rm A} = C_{\rm B} U_{\rm B} + C_{\rm C} U_{\rm C} + C_{\rm D} U_{\rm D} + ... \tag{7}$$

where $C = L^{-1} D L$ with D a diagonal matrix and L the left eigenvectors of the convective flux Jacobian. Then, the internal energy and density associated with vector $U_{\rm A}$ will be positive as long as (i) the internal energy and density associated with the vectors $U_{\rm B}$, $U_{\rm C}$, etc are positive, and (ii) the coefficients $C_{\rm A}$, $C_{\rm B}$, $C_{\rm C}$, etc are "positive". A coefficient is here considered positive if the following two conditions are met:

1. The eigenvectors associated with the coefficient $C_{\rm A}$ must correspond to the eigenvectors of the convective flux Jacobian $A_{\rm A} \equiv \partial F_{\rm A} / \partial U_{\rm A}$. Similarly, the eigenvectors associated with $C_{\rm B}$, $C_{\rm C}$, etc, must correspond to the eigenvectors of the convective flux Jacobians $A_{\rm B}$, $A_{\rm C}$, etc.

2. The eigenvalues associated with the coefficient $C_{\rm A}$, $C_{\rm B}$, etc must all be positive, but do not necessarily correspond to the eigenvalues of the respective flux Jacobian.

It can be shown that the latter is true as long as the specific heat ratio is in the range:

$$1 < \gamma < 2 \tag{8}$$

which is the case for all gases.

A proof of the rule of the positive coefficients is given in Appendix A.

4. Steger-Warming Method

When solving the Euler equations on a uniformly spaced mesh using a first-order explicit time stepping, the Steger-Warming flux vector splitting scheme yields the following discrete equation:

$$\frac{U_i^{n+1} - U_i}{\Delta t} + \frac{(F_i^+ + F_{i+1}^-) - (F_{i-1}^+ + F_i^-)}{\Delta x} = 0 \tag{9}$$

where $F^\pm \equiv L^{-1} \Lambda^\pm L U$ with $\Lambda^\pm = \frac{1}{2}(\Lambda \pm |\Lambda|)$. Making use of the property $F = AU = L^{-1} \Lambda L U$, the latter can also be rearranged in the form:

$$C_i^{n+1} U_i^{n+1} = C_{i-1} U_{i-1} + C_i U_i + C_{i+1} U_{i+1} \tag{10}$$

with the coefficients equal to:

$$C_i^{n+1} = \frac{\Delta x}{\Delta t} I \tag{11}$$

$$C_i = \frac{\Delta x}{\Delta t} I - L_i^{-1} \Lambda_i^+ L_i + L_i^{-1} \Lambda_i^- L_i \tag{12}$$

$$C_{i-1} = L_{i-1}^{-1} \Lambda_{i-1}^+ L_{i-1} \tag{13}$$

$$C_{i+1} = -L_{i+1}^{-1} \Lambda_{i+1}^- L_{i+1} \tag{14}$$

In order to maintain the internal energy and the density positive, the coefficients must have positive eigenvalues. This is the case for all coefficients except C_i which may not have positive eigenvalues if the time step is set to a too high value. The Steger-Warming scheme is hence positivity preserving under the condition of a small time step. The minimum time step for positivity preservation is derived below in Section 6.1.

For convenience, the discrete equation can be rewritten in conservative form:

$$\frac{U_i^{n+1} - U_i}{\Delta t} + \frac{F_{i+1/2} - F_{i-1/2}}{\Delta x} = 0 \tag{15}$$

where the flux at the interface is equal to:

$$F_{i+1/2} = F_i^+ + F_{i+1}^- \tag{16}$$

4.1. Second-Order Extension with MUSCL-TVD

Following the so-called MUSCL-TVD approach outlined in Ref. [9], the Steger-Warming scheme can be extended to second-order accuracy by rewriting the flux at the interface in the following form:

$$F_{i+1/2} = F^{+}\left(U^{\rm L}_{i+1/2}\right) + F^{-}\left(U^{\rm R}_{i+1/2}\right) \tag{17}$$

where $U^{\rm L}_{i+1/2}$ and $U^{\rm R}_{i+1/2}$ are vectors reconstructed from extrapolated primitive variables. For instance, the temperature, velocity, and density needed to reconstruct $U^{\rm L}_{i+1/2}$ are extrapolated using a minmod limiter with a leftward bias such as:

$$T^{\rm L}_{i+1/2} = T_i + \frac{1}{2}(T_i - T_{i-1}) \max\left(0,\ \min\left(1,\ \frac{T_{i+1} - T_i}{T_i - T_{i-1}}\right)\right) \tag{18}$$

On the other hand, the temperature, velocity, and density needed to reconstruct the vector $U^{\rm R}_{i+1/2}$ are extrapolated using a minmod limiter with a rightward bias such as:

$$T^{\rm R}_{i+1/2} = T_{i+1} + \frac{1}{2}(T_{i+1} - T_{i+2}) \max\left(0,\ \min\left(1,\ \frac{T_i - T_{i+1}}{T_{i+1} - T_{i+2}}\right)\right) \tag{19}$$

The latter has been shown to be positivity-preserving [10].

4.2. Second-Order Extension with Flux-Limited-TVD

Alternately, the Steger-Warming scheme can be extended to second-order accuracy through the use of a Total Variation Diminishing (TVD) scheme applied directly to the fluxes. For instance, using the minmod limiter, the flux at the interface becomes:

$$\begin{aligned} [F_{i+1/2}]_r = {} & \left[F_i^{+}\right]_r + \frac{1}{2}\left[F_i^{+} - F_{i-1}^{+}\right]_r \max\left(0,\ \min\left(1,\ \frac{\left[F_{i+1}^{+} - F_i^{+}\right]_r}{\left[F_i^{+} - F_{i-1}^{+}\right]_r}\right)\right) \\ & + \left[F_{i+1}^{-}\right]_r + \frac{1}{2}\left[F_{i+1}^{-} - F_{i+2}^{-}\right]_r \max\left(0,\ \min\left(1,\ \frac{\left[F_i^{-} - F_{i+1}^{-}\right]_r}{\left[F_{i+1}^{-} - F_{i+2}^{-}\right]_r}\right)\right) \end{aligned} \tag{20}$$

In the latter, the flux limiter is applied component-wise, with each flux being limited independently. Because the flux-limited scheme does not require a reconstruction of the vector of conserved variables, it requires less computing effort than MUSCL. However, as can be easily shown through numerical tests, the latter is not positivity preserving and can yield negative internal energies. This is found to be the case especially for strong shockwaves occurring within hypervelocity flows.

5. New Positivity-Preserving Flux-Limited Scheme

Making use of the rule of the positive coefficients, it is now desired to find a second-order accurate discretization stencil which does not involve reconstruction of the fluxes while being positivity-preserving. A second order accurate extension of the Steger-Warming scheme can be written as:

$$\begin{aligned} \frac{U_i^{n+1} - U_i}{\Delta t} = {} & -\frac{F_i^{+} + \frac{1}{2}\phi^{+}_{i+1/2}(F_i^{+} - F_{i-1}^{+}) + F_{i+1}^{-} + \frac{1}{2}\phi^{-}_{i+1/2}(F_{i+1}^{-} - F_{i+2}^{-})}{\Delta x} \\ & + \frac{F_{i-1}^{+} + \frac{1}{2}\phi^{+}_{i-1/2}(F_{i-1}^{+} - F_{i-2}^{+}) + F_i^{-} + \frac{1}{2}\phi^{-}_{i-1/2}(F_i^{-} - F_{i+1}^{-})}{\Delta x} \end{aligned} \tag{21}$$

where $\phi = 0$ would yield a piecewise-constant distribution of the fluxes while $\phi = 1$ would yield a piecewise linear distribution. Making use of the property $F = AU = L^{-1}\Lambda L U$, the latter can also be rearranged in the form:

$$C_i^{n+1} U_i^{n+1} = C_{i-2} U_{i-2} + C_{i-1} U_{i-1} + C_i U_i + C_{i+1} U_{i+1} + C_{i+2} U_{i+2} \tag{22}$$

with the discretization coefficients equal to:

$$C_i^{n+1} = \frac{\Delta x}{\Delta t} I \tag{23}$$

$$C_i = \frac{\Delta x}{\Delta t} I - \left(1 + \frac{1}{2}\phi^+_{i+1/2}\right) L_i^{-1} \Lambda_i^+ L_i + \left(1 + \frac{1}{2}\phi^-_{i-1/2}\right) L_i^{-1} \Lambda_i^- L_i \tag{24}$$

$$C_{i-1} = \left(1 + \frac{1}{2}\phi^+_{i+1/2} + \frac{1}{2}\phi^+_{i-1/2}\right) L_{i-1}^{-1} \Lambda_{i-1}^+ L_{i-1} \tag{25}$$

$$C_{i+1} = -\left(1 + \frac{1}{2}\phi^-_{i+1/2} + \frac{1}{2}\phi^-_{i-1/2}\right) L_{i+1}^{-1} \Lambda_{i+1}^- L_{i+1} \tag{26}$$

$$C_{i-2} = -\frac{1}{2}\phi^+_{i-1/2} L_{i-2}^{-1} \Lambda_{i-2}^+ L_{i-2} \tag{27}$$

$$C_{i+2} = \frac{1}{2}\phi^-_{i+1/2} L_{i+2}^{-1} \Lambda_{i+2}^- L_{i+2} \tag{28}$$

According to the rule of the positive coefficients, the eigenvalues of the discretization coefficients must all be positive in order to maintain the internal energy and the density positive. Unfortunately, the coefficients C_{i-2} and C_{i+2} are always negative, unless $\phi^-_{i+1/2}$ and $\phi^+_{i-1/2}$ are set to zero. Then, the scheme would be first-order accurate. Indeed a second order scheme that respects the rule of the positive coefficients is not possible if the discrete equation is written as in Eq. (22).

In order to obtain a second order scheme that respects the rule of the positive coefficients, it is necessary to first rewrite the discrete equation as:

$$C_i^{n+1} U_i^{n+1} = C_i U_i + C''_{i+1} U_{i+1} + C''_{i-1} U_{i-1} + C'_{i+1} U_{i+1} + C'_{i-1} U_{i-1} \tag{29}$$

with C_i^{n+1} and C_i as defined previously, but with the coefficients C''_{i-1} and C''_{i+1} now corresponding to:

$$C''_{i-1} = \frac{1}{2}\phi^+_{i+1/2} L_{i-1}^{-1} \Lambda_{i-1}^+ L_{i-1} \tag{30}$$

$$C''_{i+1} = -\frac{1}{2}\phi^-_{i-1/2} L_{i+1}^{-1} \Lambda_{i+1}^- L_{i+1} \tag{31}$$

and with $C'_{i-1} U_{i-1}$ and $C'_{i+1} U_{i+1}$ equal to:

$$C'_{i-1} U_{i-1} = \left(1 + \frac{1}{2}\phi^+_{i-1/2}\right) L_{i-1}^{-1} \Lambda_{i-1}^+ L_{i-1} U_{i-1} - \frac{1}{2}\phi^+_{i-1/2} L_{i-2}^{-1} \Lambda_{i-2}^+ L_{i-2} U_{i-2} \tag{32}$$

$$C'_{i+1} U_{i+1} = -\left(1 + \frac{1}{2}\phi^-_{i+1/2}\right) L_{i+1}^{-1} \Lambda_{i+1}^- L_{i+1} U_{i+1} + \frac{1}{2}\phi^-_{i+1/2} L_{i+2}^{-1} \Lambda_{i+2}^- L_{i+2} U_{i+2} \tag{33}$$

In the rewritten discrete equation the only coefficients that are not positive or that cannot be made positive by choosing a small enough time step are C'_{i-1} and C'_{i+1}. Then, to obtain a positivity-preserving scheme, it is sufficient to determine under which conditions these two coefficients are positive. To do so, first note that Eq. (33) can be rewritten as:

$$L_{i+1}^{-1} D_{i+1}^+ L_{i+1} U_{i+1} = -\left(1 + \frac{1}{2}\phi^-_{i+1/2}\right) L_{i+1}^{-1} \Lambda_{i+1}^- L_{i+1} U_{i+1} + \frac{1}{2}\phi^-_{i+1/2} L_{i+2}^{-1} \Lambda_{i+2}^- L_{i+2} U_{i+2} \tag{34}$$

where D_{i+1}^+ is a diagonal matrix with positive values. Regroup the terms including U_{i+1} together, divide through by the scalar $\phi^-_{i+1/2}$ and multiply both sides by the matrix L_{i+1}:

$$\frac{\left[D_{i+1}^+ + (1 + \frac{1}{2}\phi^-_{i+1/2})\Lambda_{i+1}^-\right]}{\frac{1}{2}\phi^-_{i+1/2}} L_{i+1} U_{i+1} = L_{i+1} L_{i+2}^{-1} \Lambda_{i+2}^- L_{i+2} U_{i+2} \tag{35}$$

Because D^+ and Λ^- are diagonal matrices we can write the latter in tensor form as:

$$\frac{\left[D_{i+1}^+\right]_{r,r} + (1 + \frac{1}{2}\phi^-_{i+1/2})\left[\Lambda_{i+1}^-\right]_{r,r}}{\frac{1}{2}\phi^-_{i+1/2}} \left[L_{i+1} U_{i+1}\right]_r = \left[L_{i+1} L_{i+2}^{-1} \Lambda_{i+2}^- L_{i+2} U_{i+2}\right]_r \tag{36}$$

Now, divide through by $\left[L_{i+1} U_{i+1}\right]_r$ and isolate $\left[D_{i+1}^+\right]_{r,r}$:

$$\left[D_{i+1}^+\right]_{r,r} = \frac{1}{2}\phi^-_{i+1/2} \frac{\left[L_{i+1} L_{i+2}^{-1} \Lambda_{i+2}^- L_{i+2} U_{i+2}\right]_r}{\left[L_{i+1} U_{i+1}\right]_r} - \left(1 + \frac{1}{2}\phi^-_{i+1/2}\right) \left[\Lambda_{i+1}^-\right]_{r,r} \tag{37}$$

But, since all the values within matrix D^+_{i+1} must be positive, it follows that:

$$\frac{1}{2}\phi^-_{i+1/2}\frac{\left[L_{i+1}L^{-1}_{i+2}\Lambda^-_{i+2}L_{i+2}U_{i+2}\right]_r}{\left[L_{i+1}U_{i+1}\right]_r}-\left(1+\frac{1}{2}\phi^-_{i+1/2}\right)\left[\Lambda^-_{i+1}\right]_{r,r}\geq 0 \tag{38}$$

Then, after defining G^- as:

$$G^-\equiv\Lambda^-LU \tag{39}$$

substituting it in the former, and isolating $\phi^-_{i+1/2}$, we obtain:

$$\phi^-_{i+1/2}\leq 2\frac{-\left[G^-_{i+1}\right]_r}{\left[G^-_{i+1}\right]_r-\left[L_{i+1}L^{-1}_{i+2}G^-_{i+2}\right]_r}\quad\text{if the RHS}\geq 0 \tag{40}$$

The latter can be rewritten as:

$$\phi^-_{i+1/2}\leq 2\Bigg/\max\left(\theta,\ \frac{\left[L_{i+1}L^{-1}_{i+2}G^-_{i+2}\right]_r-\left[G^-_{i+1}\right]_r}{\left[G^-_{i+1}\right]_r}\right) \tag{41}$$

where θ is a small positive value approaching zero.

To ensure monotonicity, the flux at the interface must be further limited with a TVD limiter. Although any limiter satisfying the TVD condition can be used, it is here chosen to use the minmod function. This results in the additional conditions:

$$\phi^-_{i+1/2}\leq\max\left(0,\ \min\left(1,\ \frac{\left[F^-_i\right]_r-\left[F^-_{i+1}\right]_r}{\left[F^-_{i+1}\right]_r-\left[F^-_{i+2}\right]_r}\right)\right) \tag{42}$$

The proposed flux-limited positivity-preserving scheme can also be written in conservative form. Then, the flux at the interface can be shown to correspond to:

$$F_{i+1/2}=F^+_i+\frac{1}{2}\phi^+_{i+1/2}\left(F^+_i-F^+_{i-1}\right)+F^-_{i+1}+\frac{1}{2}\phi^-_{i+1/2}\left(F^-_{i+1}-F^-_{i+2}\right) \tag{43}$$

where $\phi^-_{i+1/2}$ corresponds to the combination of both conditions outlined in (41) and (42):

$$\phi^-_{i+1/2}\leq\max\left(0,\ \min\left(1,\ \underbrace{\left(\left[F^-_i\right]_r-\left[F^-_{i+1}\right]_r\right)\Big/\left(\left[F^-_{i+1}\right]_r-\left[F^-_{i+2}\right]_r\right)}_{\text{flux limiter for monotonicity preservation}},\ \underbrace{\frac{2}{\max\left(\theta,\ \left(\left[L_{i+1}L^{-1}_{i+2}G^-_{i+2}\right]_r-\left[G^-_{i+1}\right]_r\right)/\left[G^-_{i+1}\right]_r\right)}}_{\text{additional limiter for positivity preservation}}\right)\right) \tag{44}$$

and with $\phi^+_{i+1/2}$ determined in a similar manner:

$$\phi^+_{i+1/2}\leq\max\left(0,\ \min\left(1,\ \underbrace{\left(\left[F^+_{i+1}\right]_r-\left[F^+_i\right]_r\right)\Big/\left(\left[F^+_i\right]_r-\left[F^+_{i-1}\right]_r\right)}_{\text{flux limiter for monotonicity preservation}},\ \underbrace{\frac{2}{\max\left(\theta,\ \left(\left[L_iL^{-1}_{i-1}G^+_{i-1}\right]_r-\left[G^+_i\right]_r\right)/\left[G^+_i\right]_r\right)}}_{\text{additional limiter for positivity preservation}}\right)\right) \tag{45}$$

In the latter, the flux limiter ϕ corresponds to the minimum over all components, θ is a small positive value approaching zero, and $G^\pm$ is defined as:

$$G^\pm\equiv\Lambda^\pm LU \tag{46}$$

By comparing Eq. (20) and Eqs. (43-45), it can be seen that a flux limited stencil can be made positivity-preserving through two modifications: (i) the limiter function must have the same value for all flux components instead of being determined independently for each component, and (ii) a positivity limiter must be added to the monotonicity limiter. The limiter function must have the same value for all components because it is assumed to be a scalar (i.e., ϕ is assumed constant over all flux components) when deriving the condition for positivity preservation starting with Eq. (35).

6. Time Step Restriction for Positivity Preservation

The maximum time step that can be used while maintaining positivity is now determined. This can be done by imposing the condition of positive eigenvalues to the discretization coefficients in which the time step appears.

6.1. First-Order Steger-Warming

For the first-order Steger-Warming scheme, the time step appears only within the coefficient C_i outlined in Eq. (12):

$$C_i = \frac{\Delta x}{\Delta t} I - L_i^{-1} \Lambda_i^{+} L_i + L_i^{-1} \Lambda_i^{-} L_i \tag{47}$$

We can determine under which condition the coefficient can be considered positive by rewriting it as:

$$C_i = \frac{\Delta x}{\Delta t} L_i^{-1} I L_i - L_i^{-1} \Lambda_i^{+} L_i + L_i^{-1} \Lambda_i^{-} L_i \tag{48}$$

with I the identity matrix. Then, after regrouping the 3 terms on the RHS, the following can be obtained:

$$C_i = L_i^{-1} \times \left[\frac{\Delta x}{\Delta t} I - |\Lambda_i| \right] \times L_i \tag{49}$$

From the latter, it is easy to show that the eigenvalues would all be positive only if

$$\Delta t_i \le \frac{\Delta x}{|u_i| + a_i} \tag{50}$$

The rule of the positive coefficient hence yields the CFL condition. That is, the Courant number should be less or equal to 1.

6.2. Second-Order Steger-Warming

For a second-order accurate scheme, the maximum allowable time step is further restricted. Indeed, the coefficient function of the time step for a second-order scheme (see Eq. (24)) becomes:

$$C_i = \frac{\Delta x}{\Delta t} L_i^{-1} I L_i - \left(1 + \frac{1}{2}\phi^{+}_{i+1/2}\right) L_i^{-1} \Lambda_i^{+} L_i + \left(1 + \frac{1}{2}\phi^{-}_{i-1/2}\right) L_i^{-1} \Lambda_i^{-} L_i \tag{51}$$

After regrouping the terms on the RHS, the following can be obtained:

$$C_i = L_i^{-1} \times \left[\frac{\Delta x}{\Delta t} I - \left(1 + \frac{1}{4}\phi^{+}_{i+1/2} + \frac{1}{4}\phi^{-}_{i-1/2}\right) |\Lambda_i| - \left(\frac{1}{4}\phi^{+}_{i+1/2} - \frac{1}{4}\phi^{-}_{i-1/2}\right) \Lambda_i \right] \times L_i \tag{52}$$

The most restrictive condition on the time step occurs when the limiter functions are maximum, that is when $\phi^{+}_{i+1/2} = \phi^{-}_{i-1/2} = 1$. Then, it can be shown that the eigenvalues would all be positive only if

$$\Delta t_i \le \frac{2}{3} \times \frac{\Delta x}{|u_i| + a_i} \tag{53}$$

The time step hence needs to be set such that the Courant number is less or equal than $\frac{2}{3}$ to ensure positivity preservation of the second-order accurate Steger-Warming scheme.

7. Entropy Correction

The discretization stencils outlined in the previous sections do not necessarily satisfy the second law of thermodynamics. For instance, the schemes could allow the formation of entropy increasing aphysical phenomena such as expansion shocks. To minimize the possibility of aphysical phenomena from occurring, it is recommended [13] to redefine the eigenvalues as follows:

$$[\Lambda \pm |\Lambda|]_{r,r} \rightarrow [\Lambda]_{r,r} \pm \sqrt{[\Lambda]_{r,r}^2 + \delta a^2} \tag{54}$$

with δ a user-specified constant typically set to 0.1. This can be easily shown through the use of the positive coefficients not to affect the positivity preservation of the proposed scheme, albeit reducing slightly the maximum allowable time step. To ensure that aphysical phenomena do not form, the entropy correction needs only be applied to the acoustic waves: that is, only the $u + a$ and $u - a$ eigenvalues need to be corrected while the u eigenvalue does not require any correction. However, numerical tests indicate that applying the entropy correction to all eigenvalues increases significantly the robustness of the Steger-Warming scheme by permitting the use of higher time steps while not affecting the accuracy of the solution. For this reason, it is preferred to here apply the entropy correction to all the eigenvalues.

8. Test Cases

Several test cases were run on a digital computer to assess the performance of the positivity-preserving flux-limited scheme compared to the other methods. The cases include vacuum generation, strong shock reflection, and shocktube problems. In all cases, the Steger-Warming, MUSCL, and the proposed positivity-preserving flux-limited scheme (POSFL) maintain the positivity of the internal energy and density. On the other hand, the non-corrected flux-limited TVD scheme is seen to sometimes yield negative internal energy.

Although not all test cases necessitate the use of the entropy correction, it is preferred to set the entropy correction factor δ to 0.1 for all problems. This is done to confirm that the POSFL scheme is positivity preserving when used along with the entropy correction, because the latter is generally required for more complex flowfields.

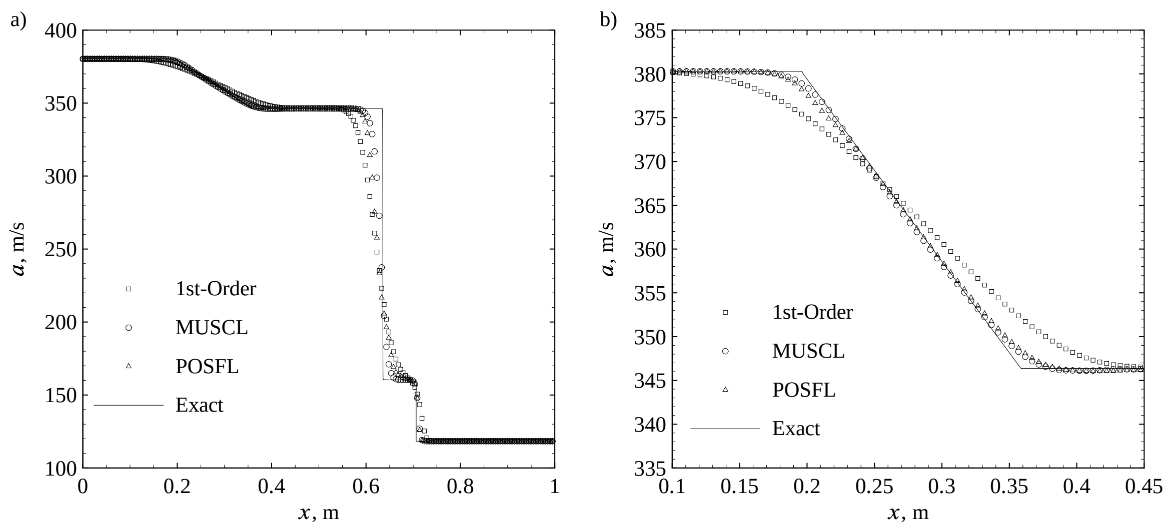

FIGURE 1. Test Case 1. Comparison of the speed of sound profiles obtained with the proposed positivity-preserving flux-limited scheme (POSFL) and the MUSCL scheme.

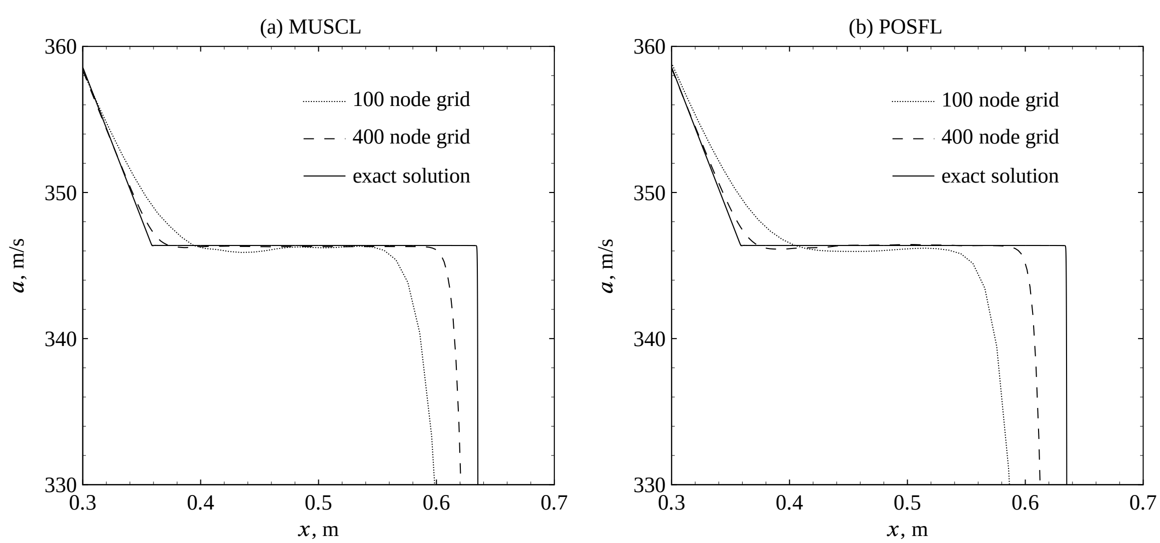

FIGURE 2. Test Case 1. Grid convergence study of the sound speed within the expansion fan and contact discontinuity for (a) MUSCL and (b) POSFL.

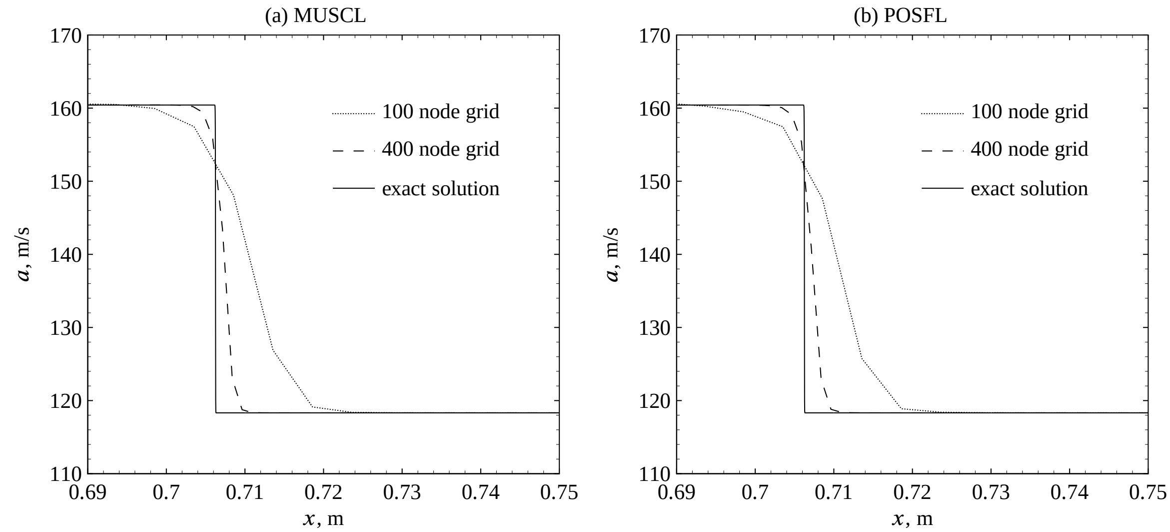

FIGURE 3: Test Case 1. Grid convergence study of the sound speed within the shockwave for (a) MUSCL and (b) POSFL.

8.1. Test Case 1: Riemann Problem in Air at Rest

The first test case consists of a Riemann problem with the initial conditions set to:

$$\rho = 1\ \text{kg/m}^3, \quad u = 0\ \text{m/s}, \quad P = \begin{cases} 103300\ \text{Pa} & \text{for } x < 0.5\ \text{m} \\ 10000\ \text{Pa} & \text{for } x \geq 0.5\ \text{m} \end{cases} \tag{55}$$

Due to the initial pressure difference, the latter generates a rightward traveling shock and a leftward traveling expansion fan separated by a contact discontinuity. Since the shock strength and velocities are moderate, all schemes considered could preserve positivity of the density and internal energy. The grid is made up of 200 nodes equally spaced and the time step is such that the Courant number is at the most 0.3. The entropy correction factor δ is set to 0.1 for all schemes. Although not shown, the entropy correction is here noticed to have negligible impact on the properties.

A comparison between the MUSCL scheme and the proposed positivity-preserving flux limited (POSFL) scheme is shown in Fig. 1 on the basis of the speed of sound. The proposed scheme is seen to perform as well as MUSCL in the expansion fan and shock regions. However, POSFL is seen to be more dissipative in the vicinity of the contact discontinuity, with more nodes required to achieve the same resolution. Additional tests indicate that this is not due to the additional limiter needed to ensure positivity preservation. Rather, it is attributed to the limiter function being the same for all flux components instead of being determined independently for each component.

This is confirmed by a grid convergence study (see Figs. 2 and 3). The grid convergence study shows that POSFL exhibits essentially the same order of accuracy as MUSCL in the vicinity of the shockwave and within the expansion fan, while being more dissipative in the vicinity of the contact discontinuity.

8.2. Test Case 2: Riemann Problem in Air Moving at Hypervelocities

The second test case consists of a Riemann problem with the initial conditions set to:

$$\rho = 1\ \text{kg/m}^3, \quad u = 1600\ \text{m/s}, \quad P = \begin{cases} 1033000\ \text{Pa} & \text{for } x < 0.5\ \text{m} \\ 10000\ \text{Pa} & \text{for } x \geq 0.5\ \text{m} \end{cases} \tag{56}$$

As for the previous test case, the grid is composed of 200 equally-spaced nodes and the entropy correction factor δ is set to 0.1. The entropy correction is verified not to impact noticeably the properties.

This test case is significantly more difficult to solve than the previous one due to the strong shock and expansion fan taking place on a flow already moving at hypervelocities. In fact, when using the non-corrected flux-limited scheme, it is not possible to obtain a solution due to negative internal energies appearing in the first few iterations, independently of the time step size.

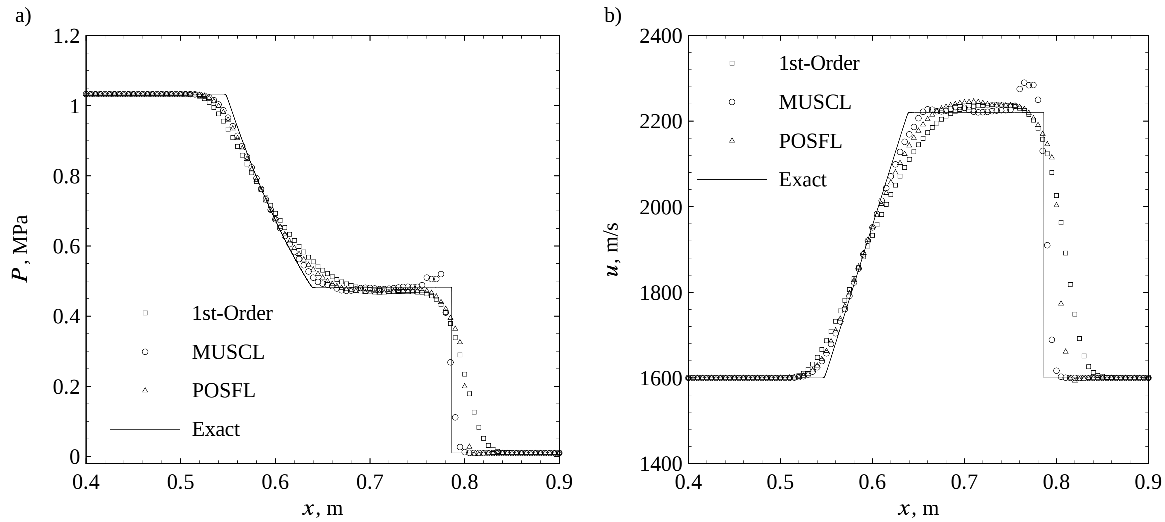

FIGURE 4. Test Case 2. Comparison of the pressure and velocity profiles obtained with the proposed positivity-preserving flux-limited scheme (POSFL) and the MUSCL scheme.

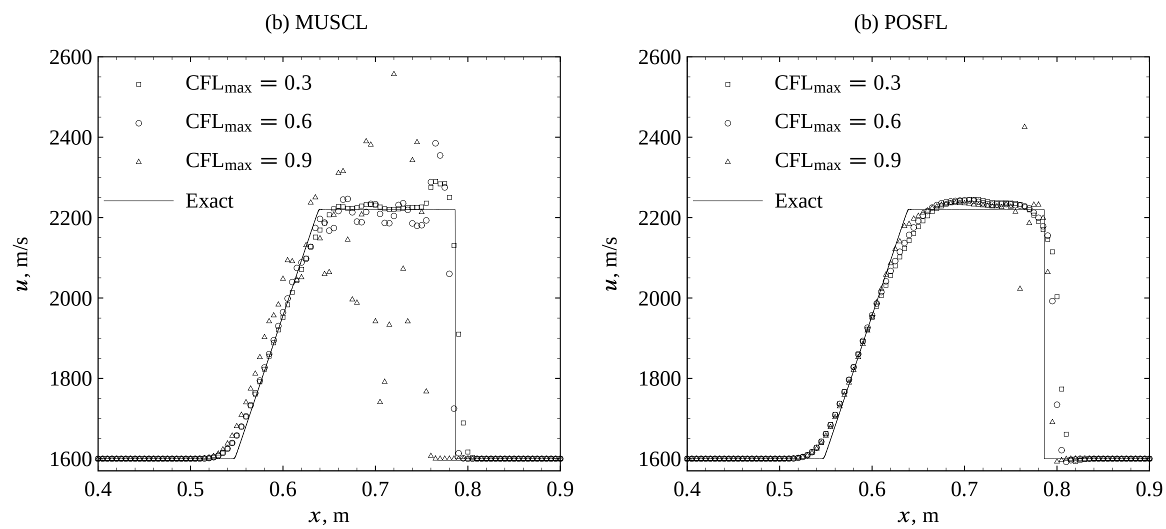

FIGURE 5. Test Case 2. Effect of the Courant number (here denoted by the acronym CFL) on the velocity profiles of (a) the MUSCL scheme and (b) the proposed positivity-preserving flux-limited scheme (POSFL).

When using the positivity-preserving MUSCL and POSFL schemes, it is possible to obtain a solution as long as the Courant number is less than 0.9.

A first comparison between the two schemes is performed at a maximum Courant number of 0.3. In Fig. 4 the pressure and velocity profiles indicate that POSFL is slightly more diffusive through the expansion fan but captures the shockwave without introducing an overshoot. The difference between the results obtained by both methods is not so significant.

However, when the time step size is raised, a more substantial difference between the two methods is observed. Figure 5 shows the impact of a raise in the maximum Courant number from 0.3 to 0.9 on the velocity profiles. While both schemes maintain the internal energy and the density positive, MUSCL exhibits much more severe even-odd node discoupling of the properties.

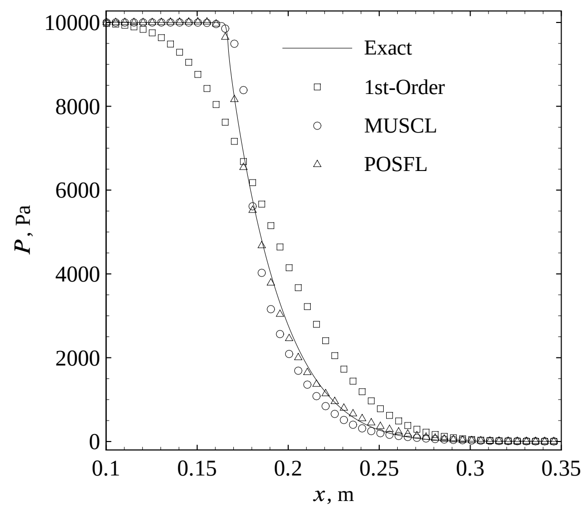

FIGURE 6. Test Case 3. Comparison of the pressure profiles obtained with the proposed positivity-preserving flux-limited scheme (POSFL) and the MUSCL scheme (every second node is not shown).

8.3. Test Case 3: Vacuum Generation

The third test case consists of the following initial conditions:

$$P = 10000 \text{ Pa}, \quad \rho = 1 \text{ kg/m}^3, \quad u = \begin{cases} -1000 \text{ m/s} & \text{for } x < 0.5 \text{ m} \\ 1000 \text{ m/s} & \text{for } x \geq 0.5 \text{ m} \end{cases} \tag{57}$$

Because the flow on the left of the domain moves leftward and the flow on the right of the domain moves rightward, two rarefaction waves are created, eventually resulting in the creation of a vacuum in the central portion of the domain. This test case (or a variant) is a particularly difficult problem to solve numerically because the pressure and density decrease by orders of magnitude within only a few nodes.

A comparison between the various discretization schemes is here done at a time of 0.3 ms, with the number of nodes fixed to 400 and the maximum Courant number fixed to 0.3. The pressure profiles obtained with the POSFL and MUSCL schemes are compared to the exact solution in Fig. 6. It is seen that both schemes remain positivity-preserving, with POSFL being slightly closer to the exact solution.

Several other test cases were computed in which a vacuum was created, including some in which the vacuum took place on a flow moving at hypervelocities. In all cases, both MUSCL and POSFL maintained the positivity of the density and the internal energy.

9. Conclusions

To help assess the positivity preservation property of numerical methods, or to help craft new positivity-preserving schemes for compressible flow, a simple rule called the "rule of the positive coefficients" is here presented. According to the latter, a scheme is guaranteed to preserve the positivity of the internal energy and density as long as the eigenvalues of all coefficients in the discretized equation are positive.

Using the rule of the positive coefficients, a new positivity-preserving flux-limited second-order accurate extension of the Steger Warming scheme is proposed. When compared to MUSCL, results obtained for several test cases indicate that the proposed scheme (POSFL) captures equally well expansion fans and shockwaves but is more dissipative in the vicinity of contact discontinuities. One advantage that POSFL does offer over MUSCL is more accurate results at a higher time step when solving hypersonic flows. For such cases, the new method (i) exhibits a significant reduction of pressure overshoot error downstream of strong shocks and (ii) allows the use of a time step generally two or three times higher while not introducing more discoupling of the properties between even and odd nodes.

The rule of the positive coefficients is also used to determine the maximum time step that maintains the positivity-preserving property. It is found that the maximum Courant number that preserves positivity drops from 1 to $\frac{2}{3}$ when the Steger-Warming scheme is turned second-order-accurate through flux limiters.

The method proposed in this paper is used solely in conjunction with the Steger-Warming flux discretization scheme and the minmod limiter to solve the one-dimensional Euler equations. Because the proposed method is written in general form, it can be used with any other TVD limiter while remaining positivity-preserving. However, when used in conjunction with other flux discretization schemes, the positivity-preserving property is maintained only if the discretized equation can be shown to obey the rule of the positive coefficients. Further, should a physical model other than the one-dimensional Euler equations be solved, positivity-preservation is only guaranteed as long as the rule of the positive coefficients can be extended to such physical model.

A. Rule of the Positive Coefficients

This section gives a proof of the "rule of the positive coefficients" for the one-dimensional Euler equations. The rule of the positive coefficients can be used to determine under what conditions a scheme is positivity-preserving and to craft a new positivity-preserving flux-limited stencil.

The rule can be summarized as follows. Consider the following equation determining the vector $U_{\rm A}$ as a function of vectors $U_{\rm B}$, $U_{\rm C}$, etc:

$$C_{\rm A} U_{\rm A} = C_{\rm B} U_{\rm B} + C_{\rm C} U_{\rm C} + C_{\rm D} U_{\rm D} + ... \tag{A.1}$$

where $CU = L^{-1} D L U$ with D a diagonal matrix and with L the left eigenvectors. The eigenvectors associated with the coefficient $C_{\rm A}$ correspond to the eigenvectors of the convective flux Jacobian $A_{\rm A} \equiv \partial F_{\rm A}/\partial U_{\rm A}$ with $F_{\rm A}$ determined from $U_{\rm A}$. Similarly, the eigenvectors associated with $C_{\rm B}$, $C_{\rm C}$, etc, correspond to the eigenvectors of the convective flux Jacobians $A_{\rm B}$, $A_{\rm C}$, etc.

According to the rule of the positive coefficients, the internal energy and the density of vector $U_{\rm A}$ are necessarily positive if (i) the internal energy and the density of vectors $U_{\rm B}$, $U_{\rm C}$, etc are positive and, (ii) if the elements on the diagonal matrices $D_{\rm A}$, $D_{\rm B}$, etc are all positive.

To prove the latter, it is sufficient to prove that the sign of the internal energy and density is conserved within the following equation:

$$(L^{-1} D^+ L)_{\rm A} U_{\rm A} = (L^{-1} D^+ L)_{\rm B} U_{\rm B} + (L^{-1} D^+ L)_{\rm C} U_{\rm C} \tag{A.2}$$

where $D^+_{\rm A,B,C}$ are positive diagonal matrices and $L^{-1}_{\rm A,B,C}$ are the right eigenvectors of the flux Jacobian matrices $A_{\rm A,B,C}$. To do so, notice that the flux $L^{-1} D^+ L U$ can be written as:

$$L^{-1} D^+ L U = \xi_1 L_1^{-1} + \xi_2 L_2^{-1} + \xi_3 L_3^{-1} \tag{A.3}$$

with L_1^{-1}, L_2^{-1}, and L_3^{-1} the right eigenvectors:

$$L_1^{-1} = \left[\begin{array}{c} 1 \\ u \\ \frac{1}{2}u^2 \end{array}\right] \quad L_2^{-1} = \left[\begin{array}{c} 1 \\ u+a \\ \frac{1}{2}u^2 + ua + \frac{a^2}{\gamma-1} \end{array}\right] \quad L_3^{-1} = \left[\begin{array}{c} 1 \\ u-a \\ \frac{1}{2}u^2 - ua + \frac{a^2}{\gamma-1} \end{array}\right] \tag{A.4}$$

and ξ_1, ξ_2, and ξ_3 some positive scalars:

$$\xi_1 = \frac{\rho D_1^+ (\gamma-1)}{\gamma} \qquad \xi_2 = \frac{\rho D_2^+}{2\gamma} \qquad \xi_3 = \frac{\rho D_3^+}{2\gamma} \tag{A.5}$$

The substitution of Eq. (A.3) in Eq. (A.2) yields the following vector equation:

$$\left(\xi_1 L_1^{-1} + \xi_2 L_2^{-1} + \xi_3 L_3^{-1}\right)_{\rm A} = \left(\xi_1 L_1^{-1} + \xi_2 L_2^{-1} + \xi_3 L_3^{-1}\right)_{\rm B} + \left(\xi_1 L_1^{-1} + \xi_2 L_2^{-1} + \xi_3 L_3^{-1}\right)_{\rm C} \tag{A.6}$$

The latter can be rewritten as:

$$\xi_{\rm A1} L_{\rm A1}^{-1} + \xi_{\rm A2} L_{\rm A2}^{-1} + \xi_{\rm A3} L_{\rm A3}^{-1} = \xi_{\rm B1} L_{\rm B1}^{-1} + \xi_{\rm B2} L_{\rm B2}^{-1} + \xi_{\rm B3} L_{\rm B3}^{-1} + \xi_{\rm C1} L_{\rm C1}^{-1} + \xi_{\rm C2} L_{\rm C2}^{-1} + \xi_{\rm C3} L_{\rm C3}^{-1} \tag{A.7}$$

The latter vector equation corresponds to the following three scalar equations (we here use the shorthand notation $\xi_{\rm A} \equiv \xi_{\rm A1} + \xi_{\rm A2} + \xi_{\rm A3}$, and similarly for $\xi_{\rm B}$ and $\xi_{\rm C}$):

$$\xi_{\rm A} = \xi_{\rm B} + \xi_{\rm C} \tag{A.8}$$

$$\xi_A u_A + (\xi_{A2} - \xi_{A3}) a_A = \xi_B u_B + (\xi_{B2} - \xi_{B3}) a_B + \xi_C u_C + (\xi_{C2} - \xi_{C3}) a_C \tag{A.9}$$

$$\begin{aligned}\frac{1}{2}\xi_A u_A^2 + (\xi_{A2} - \xi_{A3}) u_A a_A + \frac{(\xi_{A2} + \xi_{A3}) a_A^2}{\gamma - 1} &= \frac{1}{2}\xi_B u_B^2 + (\xi_{B2} - \xi_{B3}) u_B a_B + \frac{(\xi_{B2} + \xi_{B3}) a_B^2}{\gamma - 1} \\ &+ \frac{1}{2}\xi_C u_C^2 + (\xi_{C2} - \xi_{C3}) u_C a_C + \frac{(\xi_{C2} + \xi_{C3}) a_C^2}{\gamma - 1}\end{aligned} \tag{A.10}$$

After expanding Eq. (A.8), the following is obtained:

$$\underbrace{\left(2(\gamma-1) D_{A1}^+ + D_{A2}^+ + D_{A3}^+\right)}_{\text{positive if } \gamma>1} \times \rho_A = \underbrace{\left(2(\gamma-1) D_{B1}^+ + D_{B2}^+ + D_{B3}^+\right)}_{\text{positive if } \gamma>1} \times \rho_B + \underbrace{\left(2(\gamma-1) D_{C1}^+ + D_{C2}^+ + D_{C3}^+\right)}_{\text{positive if } \gamma>1} \times \rho_C \tag{A.11}$$

For $\gamma > 1$, and should ρ_B and ρ_C be positive, then ρ_A will also be positive since all the coefficients multiplying the densities are positive. Then, it can be stated that Eq. (A.2) preserves the sign of the density as long as $\gamma > 1$.

To determine the conditions under which the sign of the internal energy is preserved, first isolate u_A in Eq. (A.9) and substitute in Eq. (A.10). After some algebra, the following expression for the speed of sound is obtained:

$$\begin{aligned}\left(\frac{\xi_{A2} + \xi_{A3}}{\gamma - 1} - \frac{(\xi_{A2} - \xi_{A3})^2}{2\xi_A}\right) a_A^2 &= \left(\frac{\xi_{B2} + \xi_{B3}}{\gamma - 1} - \frac{(\xi_{B2} - \xi_{B3})^2}{2\xi_A}\right) a_B^2 + \left(\frac{\xi_{C2} + \xi_{C3}}{\gamma - 1} - \frac{(\xi_{C2} - \xi_{C3})^2}{2\xi_A}\right) a_C^2 \\ &+ \frac{((\xi_{B2} - \xi_{B3}) a_B - (\xi_{C2} - \xi_{C3}) a_C)^2}{2\xi_A} + \frac{\left(\xi_B^2 + \xi_B \xi_C + \xi_C^2\right)(u_B - u_C)^2}{2\xi_A} \\ &- \frac{(\xi_C (u_B - u_C) - (\xi_{B2} - \xi_{B3}) a_B)^2}{2\xi_A} - \frac{(\xi_B (u_B - u_C) + (\xi_{C2} - \xi_{C3}) a_C)^2}{2\xi_A}\end{aligned} \tag{A.12}$$

Since the square of the sound speed is proportional to the internal energy for a calorically perfect gas, it follows that to prove the positivity preservation of the internal energy, it is sufficient to prove the positivity preservation of the square of the sound speed. Consequently, let's proceed to find the conditions for which Eq. (A.12) preserves the sign of the square of the speed of sound. For this to occur, the sign of the RHS must be the same as the sign of the coefficient preceding a_A^2. Consider the case for which the RHS is positive. Then the coefficient preceding a_A^2 must be positive:

$$\frac{\xi_{A2} + \xi_{A3}}{\gamma - 1} - \frac{(\xi_{A2} - \xi_{A3})^2}{2\xi_A} > 0 \tag{A.13}$$

Then, recalling that $\xi_A = \xi_{A1} + \xi_{A2} + \xi_{A3}$, and recalling that $\xi_{A1} \geq 0$, $\xi_{A2} \geq 0$, and $\xi_{A3} \geq 0$, it follows that $\xi_A \geq \xi_{A2} + \xi_{A3}$. Then, the latter equation becomes:

$$\frac{\xi_{A2} + \xi_{A3}}{\gamma - 1} - \frac{(\xi_{A2} - \xi_{A3})^2}{2(\xi_{A2} + \xi_{A3})} > 0 \tag{A.14}$$

Further, since $\xi_{A2} + \xi_{A3} \geq \xi_{A2} - \xi_{A3}$:

$$\frac{\xi_{A2} + \xi_{A3}}{\gamma - 1} - \frac{(\xi_{A2} + \xi_{A3})^2}{2(\xi_{A2} + \xi_{A3})} > 0 \tag{A.15}$$

After a few simplifications, the following can be obtained:

$$\frac{3 - \gamma}{\gamma - 1} > 0 \tag{A.16}$$

which entails the γ range:

$$1 < \gamma < 3 \tag{A.17}$$

Having determined the γ range yielding a positive coefficient preceding the a_A^2 term in Eq. (A.12), let's proceed to determine the conditions for which the RHS of Eq. (A.12) is positive. To do so, split the RHS into three groups of terms:

$$\mathrm{RHS(A.12)} = \mathrm{RHS(A.12)}_1 + \mathrm{RHS(A.12)}_2 + \mathrm{RHS(A.12)}_3 \tag{A.18}$$

with:

$$\mathrm{RHS(A.12)}_1 = \left(\frac{\xi_{B2} + \xi_{B3}}{\gamma - 1} - \frac{(\xi_{B2} - \xi_{B3})^2}{2\xi_A}\right) a_B^2 + \frac{(\xi_C^2 + \frac{1}{2}\xi_B \xi_C)(u_B - u_C)^2}{2\xi_A} \tag{A.19}$$

$$-\frac{(\xi_{\rm C}(u_{\rm B}-u_{\rm C})-(\xi_{\rm B2}-\xi_{\rm B3})a_{\rm B})^2}{2\xi_{\rm A}}$$

$$\mathrm{RHS(A.12)}_2 = \left(\frac{\xi_{\rm C2}+\xi_{\rm C3}}{\gamma-1}-\frac{(\xi_{\rm C2}-\xi_{\rm C3})^2}{2\xi_{\rm A}}\right)a_{\rm C}^2+\frac{(\xi_{\rm B}^2+\frac{1}{2}\xi_{\rm B}\xi_{\rm C})(u_{\rm B}-u_{\rm C})^2}{2\xi_{\rm A}} \tag{A.20}$$

$$-\frac{(\xi_{\rm B}(u_{\rm B}-u_{\rm C})+(\xi_{\rm C2}-\xi_{\rm C3})a_{\rm C})^2}{2\xi_{\rm A}}$$

$$\mathrm{RHS(A.12)}_3 = \frac{((\xi_{\rm B2}-\xi_{\rm B3})a_{\rm B}-(\xi_{\rm C2}-\xi_{\rm C3})a_{\rm C})^2}{2\xi_{\rm A}} \tag{A.21}$$

Since $\mathrm{RHS(A.12)}_3$ is always positive due to $\xi_{\rm A}$ being always positive, it imposes no condition on the valid range of γ. Now proceed to determine when $\mathrm{RHS(A.12)}_1>0$:

$$\left(\frac{\xi_{\rm B2}+\xi_{\rm B3}}{\gamma-1}-\frac{(\xi_{\rm B2}-\xi_{\rm B3})^2}{2\xi_{\rm A}}\right)a_{\rm B}^2+\frac{(\xi_{\rm C}^2+\frac{1}{2}\xi_{\rm B}\xi_{\rm C})(u_{\rm B}-u_{\rm C})^2}{2\xi_{\rm A}}-\frac{(\xi_{\rm C}(u_{\rm B}-u_{\rm C})-(\xi_{\rm B2}-\xi_{\rm B3})a_{\rm B})^2}{2\xi_{\rm A}}>0 \tag{A.22}$$

To prove under which conditions the latter is true, first define ζ such that the following holds:

$$a_{\rm B}=\zeta|u_{\rm B}-u_{\rm C}| \tag{A.23}$$

After substituting the latter in the former and after doing a bit of algebra, the following is obtained:

$$\left(\frac{\xi_{\rm B2}+\xi_{\rm B3}}{\gamma-1}-\frac{(\xi_{\rm B2}-\xi_{\rm B3})^2}{\xi_{\rm A}}\right)\zeta^2(u_{\rm B}-u_{\rm C})^2+\frac{\xi_{\rm B}\xi_{\rm C}}{4\xi_{\rm A}}(u_{\rm B}-u_{\rm C})^2+\frac{\xi_{\rm C}(u_{\rm B}-u_{\rm C})(\xi_{\rm B2}-\xi_{\rm B3})\zeta|u_{\rm B}-u_{\rm C}|}{\xi_{\rm A}}>0 \tag{A.24}$$

The most stringent condition would be when the last term on the LHS is negative. For such a case, (A.24) can be rewritten as:

$$\left(\frac{\xi_{\rm B2}+\xi_{\rm B3}}{\gamma-1}-\frac{(\xi_{\rm B2}-\xi_{\rm B3})^2}{\xi_{\rm A}}\right)\zeta^2(u_{\rm B}-u_{\rm C})^2+\frac{\xi_{\rm B}\xi_{\rm C}}{4\xi_{\rm A}}(u_{\rm B}-u_{\rm C})^2-\frac{\xi_{\rm C}(u_{\rm B}-u_{\rm C})^2|\xi_{\rm B2}-\xi_{\rm B3}|\zeta}{\xi_{\rm A}}>0 \tag{A.25}$$

Now, divide through by $(u_{\rm B}-u_{\rm C})^2$:

$$\left(\frac{\xi_{\rm B2}+\xi_{\rm B3}}{\gamma-1}-\frac{(\xi_{\rm B2}-\xi_{\rm B3})^2}{\xi_{\rm A}}\right)\zeta^2+\frac{\xi_{\rm B}\xi_{\rm C}}{4\xi_{\rm A}}-\frac{\xi_{\rm C}|\xi_{\rm B2}-\xi_{\rm B3}|\zeta}{\xi_{\rm A}}>0 \tag{A.26}$$

But, because $\xi_{\rm B2}+\xi_{\rm B3}\geq|\xi_{\rm B2}-\xi_{\rm B3}|$ and because $\xi_{\rm B}\geq\xi_{\rm B2}+\xi_{\rm B3}$, (A.26) becomes:

$$\left(\frac{\xi_{\rm B2}+\xi_{\rm B3}}{\gamma-1}-\frac{(\xi_{\rm B2}+\xi_{\rm B3})^2}{\xi_{\rm A}}\right)\zeta^2+\frac{(\xi_{\rm B2}+\xi_{\rm B3})\xi_{\rm C}}{4\xi_{\rm A}}-\frac{\xi_{\rm C}(\xi_{\rm B2}+\xi_{\rm B3})\zeta}{\xi_{\rm A}}>0 \tag{A.27}$$

After dividing through by $\xi_{\rm B2}+\xi_{\rm B3}$ and multiplying through by $\xi_{\rm A}$, the following is obtained:

$$\left(\frac{\xi_{\rm A}}{\gamma-1}-(\xi_{\rm B2}+\xi_{\rm B3})\right)\zeta^2+\frac{\xi_{\rm C}}{4}-\xi_{\rm C}\zeta>0 \tag{A.28}$$

But, because $\xi_{\rm B}\geq\xi_{\rm B2}+\xi_{\rm B3}$, (A.28) can also be written as:

$$\left(\frac{\xi_{\rm A}}{\gamma-1}-\xi_{\rm B}\right)\zeta^2+\frac{\xi_{\rm C}}{4}-\xi_{\rm C}\zeta>0 \tag{A.29}$$

Recall that $\xi_{\rm A}=\xi_{\rm B}+\xi_{\rm C}$ and rearrange:

$$\left(\frac{\xi_{\rm C}}{\gamma-1}+\frac{(2-\gamma)\xi_{\rm B}}{\gamma-1}\right)\zeta^2+\frac{\xi_{\rm C}}{4}-\xi_{\rm C}\zeta>0 \tag{A.30}$$

Because $\xi_{\rm B}$ and $\xi_{\rm C}$ are independent, (A.30) yields two conditions:

$$\frac{(2-\gamma)\xi_{\rm B}}{\gamma-1}\zeta^2>0 \quad\text{and}\quad \frac{\xi_{\rm C}}{\gamma-1}\zeta^2+\frac{\xi_{\rm C}}{4}-\xi_{\rm C}\zeta>0 \tag{A.31}$$

which simplifies to (noting that ξ_B and ξ_C are positive):

$$\frac{2-\gamma}{\gamma-1}\zeta^2 > 0 \quad \text{and} \quad \frac{1}{\gamma-1}\zeta^2 + \frac{1}{4} - \zeta > 0 \tag{A.32}$$

Both conditions yield a range of γ of:

$$1 < \gamma < 2 \tag{A.33}$$

As long as the latter is satisfied, it is guaranteed that $\text{RHS(A.12)}_1 > 0$. Similarly, it can be shown that the same restriction on the specific heat ratio guarantees that $\text{RHS(A.12)}_2 > 0$. It follows that Eq. (A.2) preserves the sign of the square of the sound speed (and hence, of the internal energy) as long as the specific heat ratio is in the range $1 < \gamma < 2$.

The "rule of the positive coefficients" can hence be summarized as follows. For a discrete equation of the form:

$$C_\text{A} U_\text{A} = C_\text{B} U_\text{B} + C_\text{C} U_\text{C} + C_\text{D} U_\text{D} + ... \tag{A.34}$$

The internal energy and density associated with vector U_A will be positive as long as (i) the specific heat ratio is in the range $1 < \gamma < 2$, (ii) the internal energy and density associated with the vectors U_B, U_C, etc are positive, and (iii) the coefficients C_A, C_B, C_C, etc are "positive". A coefficient is here considered positive if the following two conditions are met:

1. The eigenvectors associated with the coefficient C_A must correspond to the eigenvectors of the convective flux Jacobian $A_\text{A} \equiv \partial F_\text{A}/\partial U_\text{A}$. Similarly, the eigenvectors associated with C_B, C_C, etc, must correspond to the eigenvectors of the convective flux Jacobians A_B, A_C, etc.

2. The eigenvalues associated with the coefficient C_A, C_B, etc must all be positive, but do not necessarily correspond to the eigenvalues of the respective convective flux Jacobian.

Acknowledgment

This work was sponsored by a Grant from Pusan National University (Grant #20080762000).

References

[1] EINFELDT, B., MUNZ, C. D., ROE, P. L., AND SJOGREEN, B., "On Godunov-type Methods near Low Densities," *Journal of Computational Physics*, Vol. 92, 1991, pp. 273–295.

[2] GRESSIER, J., VILLEDIEU, P., AND MOSCHETTA, J.-M., "Positivity of Flux Vector Splitting Schemes," *Journal of Computational Physics*, Vol. 155, No. 1, 1999, pp. 199–220.

[3] GODUNOV, S. K., "A Finite Difference Method for the Numerical Computation of Discontinuous Solutions of the Equations of Fluid Dynamics," *Math. Sbornik*, Vol. 47, No. 3, 1959, pp. 271–306.

[4] VAN LEER, B., "Flux Vector Splitting for the Euler Equations," *Lecture Notes in Physics*, Vol. 170, 1982, pp. 505–512, Proceedings of the 8th International Conference of Numerical Methods for Fluid Dynamics.

[5] STEGER, J. L. AND WARMING, R. F., "Flux Vector Splitting of the Inviscid Gasdynamics Equations with Application to Finite Difference Methods," *Journal of Computational Physics*, Vol. 40, 1981, pp. 263–293.

[6] ROE, P. L., "Approximate Riemann Solvers, Parameter Vectors, and Difference Schemes," *Journal of Computational Physics*, Vol. 43, 1981, pp. 357–372.

[7] HARTEN, A., LAX, P. D., AND VAN LEER, B., "On Upstream Differencing and Godunov-type Schemes for Hyperbolic Conservation Laws," *SIAM Review*, Vol. 25, No. 1, 1983, pp. 35–61.

[8] DUBROCA, B., "Positively Conservative Roe's Matrix for Euler Equations," *Lecture Notes in Physics*, Vol. 515, 1998, pp. 272–277, Proceedings of the 16th International Conference of Numerical Methods for Fluid Dynamics.

[9] ANDERSON, W. K., THOMAS, J. L., AND VAN LEER, B., "Comparison of Finite Volume Flux Volume Flux Vector Splittings for the Euler Equations," *AIAA Journal*, Vol. 24, 1986, pp. 1453–1460.

[10] LINDE, T. AND ROE, P. L., "Robust Euler Codes," 1997, AIAA Paper 97-2098.

[11] WANG, C. AND LIU, J.-G., "Positivity Property of Second-Order Flux-Splitting Schemes for the Compressible Euler Equations," *Discrete and Continuous Dynamical Systems – Series B*, Vol. 3, No. 2, May 2003, pp. 201–228.

[12] PATANKAR, S. V., *Numerical Heat Transfer and Fluid Flow*, Taylor and Francis, 1980.

[13] LANEY, C. B., *Computational Gasdynamics*, Cambridge University Press, Cambridge, New-York, NY, 1998.