Journal of Computational Physics, Vol. 361, Pages 391–411, 2018.

Positivity-Preserving Dual Time Stepping Schemes for Gas Dynamics

Bernard Parent*

A new approach at discretizing the temporal derivative of the Euler equations is here presented which can be used with dual time stepping. The temporal discretization stencil is derived along the lines of the Cauchy-Kowalevski procedure resulting in cross differences in spacetime but with some novel modifications which ensure the positivity of the discretization coefficients. It is then shown that the so-obtained spacetime cross differences result in changes to the wave speeds and can thus be incorporated within Roe or Steger-Warming schemes (with and without reconstruction-evolution) simply by altering the eigenvalues. The proposed approach is advantaged over alternatives in that it is positivity-preserving for the Euler equations. Further, it yields monotone solutions near discontinuities while exhibiting a truncation error in smooth regions less than the one of the second- or third-order accurate backward-difference-formula (BDF) for either small or large time steps. The high resolution and positivity preservation of the proposed discretization stencils are independent of the convergence acceleration technique which can be set to multigrid, preconditioning, Jacobian-free Newton-Krylov, block-implicit, etc. Thus, the current paper also offers the first implicit integration of the time-accurate Euler equations that is positivity-preserving in the strict sense (that is, the density and temperature are guaranteed to remain positive). This is in contrast to all previous positivity-preserving implicit methods which only guaranteed the positivity of the density, not of the temperature or pressure. Several stringent reacting and inert test cases confirm the positivity-preserving property of the proposed method as well as its higher resolution and higher computational efficiency over other second-order and third-order implicit temporal discretization strategies.

1. Introduction

First outlined in Payret and Taylor [1] and subsequently used to simulate the unsteady Euler equations in conjunction with finite-volume schemes by Jameson [2] and others [3, 4], dual time stepping consists of adding a pseudotime derivative to the physical model and performing subiterations in pseudotime until the residual (consisting of the sum of the spatial and temporal derivatives) becomes small. Such a strategy is advantaged over alternate implicit integration methods by permitting the use of physical time steps that can be as high or even orders of magnitude larger than the physical time scales. Indeed, other implicit integration strategies that make use of a locally frozen Jacobian [5, 6, 7, 8] induce considerable linearization error in flow regions where the Jacobian changes significantly from one physical time level to the next [9] as often occurs for problems with high variations in mesh spacing or with a large disparity of the time scales (due to chemical reactions for instance). Linearization error can then become predominant unless excessively small integration steplengths (often several orders of magnitude below the CFL condition) are specified. Because it is not subject to linearization error, dual time stepping can be used with larger integration steplengths for problems involving a time scale discrepancy hence resulting in significant gains in computational efficiency. For this reason, a dual-time integration strategy is often the preferred strategy to simulate chemically-reacting compressible flows where the time scales are disparate (see for instance Refs. [10, 11, 12, 13, 14]).

It is emphasized that "dual time stepping" is here defined as a time-accurate method which minimizes the residual once at each physical time level through subiterations. Thus, dual time stepping is here not limited to the use of pseudotime stepping but would also include any other type of convergence acceleration such as Jacobian-Free Newton-Krylov, preconditioning, multigrid, etc as long as such convergence acceleration aims to reduce the magnitude of the residual at each physical time step with the

*Associate Professor, Dept. of Aerospace Engineering, Pusan National University, Busan 609-735, Korea, https://bernardparent.ca.

residual corresponding to the sum of the source terms, of the discretized spatial derivatives, and of the discretized temporal derivative.

Whereas several second-order-or-higher stencils for the temporal derivative are available for dual time stepping, none is monotone. Indeed, second-order or higher backward difference formula (BDF) and the trapezoidal rule introduce oscillations in the vicinity of shockwaves or other discontinuities [13] leading not only to a loss of accuracy but also to convergence hangs. Further, neither the trapezoidal rule nor the BDF functions (except for the first order BDF) is positivity-preserving even when used in conjunction with positivity-preserving spatial discretization stencils, hence leading to difficulties in reaching convergence for many problems involving vacuums or strong shocks due to the density or pressure becoming negative. It is noted that positivity-preservation is here meant in the most strict sense as the capability to maintain the positivity of all determinative properties with a determinative property being a property that must necessarily be positive to be physically meaningful. Thus, when solving the Euler equations of gas dynamics, the determinative properties would include the density, the pressure, and the temperature, and a positivity-preserving scheme guarantees that all such properties are always maintained positive.

To overcome the deficiencies of existing stencils, a new approach is here proposed for discretizing the temporal derivative within dual time stepping. The approach is derived along the lines of the Cauchy-Kowalevski procedure leading to cross differences in spacetime but with several novel modifications that ensure that the resulting discretization equation in space-time has all-positive coefficients. Not only does this lead to a monotone solution free of spurious oscillations near discontinuities, but this also leads to a positivity-preserving algorithm. Indeed, should a positive coefficient be defined as a matrix that has positive eigenvalues, having positive coefficients is a sufficient condition to guarantee positivity-preservation of all determinative properties at least for the Euler equations of gas dynamics (see proof in the Appendix of Ref. [15]).

Because the method proposed is strictly a discretization stencil, it can be used with any convergence acceleration technique such as block-implicit schemes, Jacobian-Free Newton-Krylov, preconditioning, etc, and such will not affect in any way its positivity-preserving and monotonicity-preserving properties. Thus, the present paper also offers the first implicit integration of the time-accurate Euler equations that is positivity-preserving. Indeed, no previous implicit method maintains the positivity of all determinative properties when applied to the Euler equations. In previous work related to positive implicit schemes (see for instance Refs. [16, 17, 18, 19]), the implicit schemes presented only guarantee to maintain the positivity of a transported property, not of all determinative properties. Thus, when applied to the Euler equations, such would maintain the positivity of the density (because the latter is a transported property) but would not necessarily maintain the positivity of the temperature or pressure (because the latter are not transported properties). In contrast, the discretization stencil proposed herein guarantees that the positivity of all determinative properties is maintained whether or not the determinative properties are transported properties and whether or not an implicit convergence acceleration technique is used.

This paper is divided as follows. First, a discretization equation in space-time applicable to dual time stepping strategies is obtained through the Cauchy-Kowalevski procedure applied to the advection equation. Second, some novel modifications to the Cauchy-Kowalevski procedure are outlined to obtain a discretization equation in space-time with all-positive coefficients. Third, it is demonstrated how the latter can be incorporated within positivity-preserving FDS (Flux Difference Splitting) and FVS (Flux Vector Splitting) spatial discretization schemes for the Euler equations. The gains in resolution and positivity over the conventional stencils for the temporal derivative (trapezoidal rule, BDF2, BDF3, TVD-RK2, etc) are then assessed through some stringent test cases.

2. Cauchy-Kowalevski Discretization of the Advection Equation

Using the Cauchy-Kowalevski procedure [20, Section 19.3], it is here desired to determine an adequate discretization equation for the advection equation supplemented by a pseudotime derivative. Thus, first add a pseudotime derivative to the advection equation as follows:

$$\frac{\partial u}{\partial \tau} + \frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0 \tag{1}$$

with τ the pseudotime, t the time coordinate, x the space coordinate, and u the conserved variable. The pseudotime derivative will here play the same role in our derivation of discretization stencils for the unsteady Euler equations as a time derivative does in the derivation of stencils for the steady Euler equations. Such will eventually lead to a pseudotime step which, when replaced by a characteristic pseudotime step (i.e., an expression for the pseudotime step which does not involve pseudotime), will affect the monotonicity of the discretization stencils for the spatial and temporal derivatives. This follows the idea first proposed by Huang and Lerat [21] who obtained stencils for the steady-state Euler equations by retaining the time derivative through the Cauchy-Kowalevski process and replacing the physical time step by a characteristic time step to improve the monotonicity of the spatial stencils.

Let us now take the derivative of Eq. (1) with respect to τ, the derivative of Eq. (1) multiplied by a with respect to x, and the derivative of Eq. (1) with respect to t. After subtracting the last two resulting equations from the first, we obtain:

$$\frac{\partial^2 u}{\partial \tau^2} = a^2\frac{\partial^2 u}{\partial x^2} + 2a\frac{\partial^2 u}{\partial t\,\partial x} + \frac{\partial^2 u}{\partial t^2} \tag{2}$$

Previous methods making use of a characteristic timestep to improve the resolution of steady-state stencils in $x - y$ imposed a first-order forward stencil on the time derivative [21, 22]. Here because we seek to improve the resolution of the stencils in $x - t$ rather than $x - y$, we treat the time derivative similarly to the spatial derivative and rather impose a first-order forward stencil on the pseudotime derivative. Thus, when the pseudotime derivative is discretized through a first-order forward stencil, the following can be shown through Taylor series expansion centered on the pseudotime level m:

$$\frac{\partial u}{\partial \tau} = \frac{u_i^{m+1} - u_i^m}{\Delta\tau} - \frac{\Delta\tau}{2}\frac{\partial^2 u}{\partial \tau^2} + O(\Delta\tau^2) \tag{3}$$

Substitute $\partial_\tau^2 u$ from Eq. (2) and substitute the resulting equation in the advection equation, Eq. (1):

$$\frac{u_i^{m+1} - u_i^m}{\Delta\tau} - \frac{\Delta\tau}{2}\left(a^2\frac{\partial^2 u}{\partial x^2} + 2a\frac{\partial^2 u}{\partial x \partial t} + \frac{\partial^2 u}{\partial t^2}\right) + a\frac{\partial u}{\partial x} + \frac{\partial u}{\partial t} = O(\Delta\tau^2) \tag{4}$$

Express the spatial derivatives with the following centered stencils:

$$\frac{\partial^2 u}{\partial x^2} = \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{\Delta x^2} + O(\Delta x^2) \tag{5}$$

$$\frac{\partial u}{\partial x} = \frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} + O(\Delta x^2) \tag{6}$$

and express the time derivative using backward stencils to allow for a computationally-efficient time-marching scheme to be used:

$$\frac{\partial u}{\partial t} = \frac{\frac{3}{2}u_i^n - 2u_i^{n-1} + \frac{1}{2}u_i^{n-2}}{\Delta t} + O(\Delta t^2) \tag{7}$$

$$\frac{\partial^2 u}{\partial t^2} = \frac{u_i^n - 2u_i^{n-1} + u_i^{n-2}}{\Delta t^2} + O(\Delta t) \tag{8}$$

In the latter, i and n are the grid indices along the x and t coordinates, respectively. As for the cross derivatives, we will use stencils that do not make use of any "future" value in time, as follows:

$$\frac{\partial^2 u}{\partial t \partial x} = \frac{u_{i+1/2}^n - u_{i-1/2}^n - u_{i+1/2}^{n-1} + u_{i-1/2}^{n-1}}{\Delta x \Delta t} + O(\Delta x) \tag{9}$$

where $\delta_x u$ refers to the discretization of $\partial_x u$ at a certain time level and where the subscript $i + 1/2$ refers to the interface located between node i and node $i + 1$. How $\delta_x u$ is specified within the cross-derivative terms has a significant impact on the monotonicity of the solution and will be determined subsequently.

Now, substitute Eqs. (5), (7), (8), (9), (6) in Eq. (4):

$$\begin{aligned}\frac{u_i^{m+1} - u_i^m}{\Delta\tau} &- a^2\frac{\Delta\tau}{2}\frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{\Delta x^2} - a\Delta\tau\frac{u_{i+1/2}^n - u_{i-1/2}^n - u_{i+1/2}^{n-1} + u_{i-1/2}^{n-1}}{\Delta x \Delta t} \\ &- \frac{\Delta\tau}{2}\frac{u_i^n - 2u_i^{n-1} + u_i^{n-2}}{\Delta t^2} + a\frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} + \frac{3u_i^n - 4u_i^{n-1} + u_i^{n-2}}{2\Delta t} \\ &= O(\Delta t\Delta\tau,\ \Delta x^2,\ \Delta t^2,\ \Delta\tau^2,\ \Delta\tau\Delta x,\ \Delta\tau\Delta x^2\Delta t^{-1})\end{aligned} \tag{10}$$

Because we seek flux functions at the interfaces of cells written in finite-volume form so they can be integrated using a dual time stepping approach, we wish to express the discretization equation in finite-volume form as follows:

$$\frac{u_i^{m+1} - u_i^m}{\Delta\tau} + \frac{f_{i+1/2}^n - f_{i-1/2}^n}{\Delta x} + \frac{u_i^{n+1/2} - u_i^{n-1/2}}{\Delta t} = 0 \tag{11}$$

To do so, recast Eq. (10) so that its terms can be readily regrouped within flux functions at the different interfaces:

$$\begin{aligned}\frac{u_i^{m+1} - u_i^m}{\Delta\tau} &- a^2\frac{\Delta\tau}{2}\frac{u_{i+1}^n - u_i^n}{\Delta x^2} + a^2\frac{\Delta\tau}{2}\frac{u_i^n - u_{i-1}^n}{\Delta x^2} + a\frac{u_{i+1}^n + u_i^n}{2\Delta x} - a\frac{u_i^n + u_{i-1}^n}{2\Delta x} - \beta_{xt}a\Delta\tau\frac{u_{i+1/2}^n - u_{i+1/2}^{n-1}}{\Delta x \Delta t} \\ &+ \beta_{xt}a\Delta\tau\frac{u_{i-1/2}^n - u_{i-1/2}^{n-1}}{\Delta x \Delta t} - (1 - \beta_{xt})a\Delta\tau\frac{u_{i+1/2}^n - u_{i-1/2}^n}{\Delta x \Delta t} \\ &+ (1 - \beta_{xt})a\Delta\tau\frac{u_{i+1/2}^{n-1} - u_{i-1/2}^{n-1}}{\Delta x \Delta t} - \frac{\Delta\tau}{2}\frac{u_i^n - u_i^{n-1}}{\Delta t^2} + \frac{\Delta\tau}{2}\frac{u_i^{n-1} - u_i^{n-2}}{\Delta t^2} + \frac{3u_i^n - u_i^{n-1}}{2\Delta t} \\ &- \frac{3u_i^{n-1} - u_i^{n-2}}{2\Delta t} = O(\Delta t\Delta\tau,\ \Delta x^2,\ \Delta t^2,\ \Delta\tau^2,\ \Delta\tau\Delta x,\ \Delta\tau\Delta t^2\Delta x^{-1})\end{aligned} \tag{12}$$

A comparison of Eq. (12) with Eq. (11) yields the following flux functions in finite-volume form:

$$f^n_{i+1/2} = a\frac{u^n_{i+1}+u^n_i}{2} - a^2\frac{\Delta\tau^n_{i+1/2}}{2}\frac{u^n_{i+1}-u^n_i}{\Delta x} - \beta_{xt}a\Delta\tau^n_{i+1/2}\frac{u^n_{i+1/2}-u^{n-1}_{i+1/2}}{\Delta t} \tag{13}$$

$$u^{n+1/2}_i = \frac{3u^n_i - u^{n-1}_i}{2} - \frac{\Delta\tau^{n+1/2}_i}{2}\frac{u^n_i-u^{n-1}_i}{\Delta t} - (1-\beta_{xt})a\Delta\tau^{n+1/2}_i\frac{u^n_{i+1/2}-u^n_{i-1/2}}{\Delta x} \tag{14}$$

The newly-introduced parameter β_{xt} is used to split the cross-derivative terms among the space and time dimensions. Its usefulness will become clear when the flux functions are written in terms of a characteristic pseudotime step where it will be apparent how a change in β_{xt} can impact the monotonicity of the solution. However, because a characteristic pseudotime step has not yet been introduced, note that β_{xt} is such that it can be set to any value and Eq. (12) will collapse to the discretization equation previously outlined in Eq. (10).

3. Enforcement of Positivity on the Discretization Coefficients

The flux functions outlined in Eqs. (13)-(14) were derived through the standard Cauchy-Kowalevski procedure and are second-order accurate in space-time. However, when the physical time step is set to moderate-to-large values, they result in a solution that is non-monotone with spurious oscillations appearing in the vicinity of discontinuities. To prevent spurious oscillations from forming, we here propose some novel alterations to the flux functions that are such that they result in a discretization equation with all positive coefficients.

The first modification consists of substituting the pseudotime steps appearing within the flux functions by a *characteristic* pseudotime step. For the flux function (13) to yield positive coefficients, the coefficient multiplying the u^n_{i+1} term must be less than or equal to zero and the coefficient multiplying the u_i term must be greater or equal to zero. It can be shown that this would lead to the condition $\Delta\tau^n_{i+1/2} \geq \frac{\Delta x}{|a|}$. The least dissipative stencil would occur when the magnitude of the second derivatives is the least, which would occur when $\Delta\tau^n_{i+1/2}$ is as small as possible. Therefore, we here set the pseudotime step at the spatial interface to a characteristic pseudotime step equal to:

$$\Delta\tau^n_{i+1/2} = \frac{\Delta x}{|a|} \tag{15}$$

Similarly, the coefficient multiplying the u^{n-1}_i term within the flux function (14) must be greater or equal to zero, leading to the condition $\Delta\tau^{n+1/2}_i \geq \Delta t$. For the smallest amount of dissipation, the second derivatives in time should be as small as possible. Consequently, the pseudotime step should be as small as possible while satisfying the latter condition. This yields:

$$\Delta\tau^{n+1/2}_i = \Delta t \tag{16}$$

It is noted that Eqs. (15) and (16) correspond to the definitions of Δt and Δx for a given pseudotime step $\Delta\tau = \Delta\tau^{n+1/2}_i = \Delta\tau^n_{i+1/2}$. Therefore, the latter substitutions of the pseudotime step by a characteristic pseudotime step does not result in a loss in accuracy as long as Δx and Δt are such that they make the latter two equations equal to each other. Should Δx and Δt be specified differently, the scheme will still have positive coefficients and lead to monotone solutions, but its truncation error will be more significant. After substituting the characteristic pseudotime steps from Eqs. (15) and (16) in the fluxes outlined in Eqs. (13) and (14) we obtain:

$$f^n_{i+1/2} = a\frac{u^n_{i+1}+u^n_i}{2} - |a|\frac{u^n_{i+1}-u^n_i}{2} - \beta_{xt}\frac{a}{|a|}\frac{\Delta x}{\Delta t}\left(u^n_{i+1/2} - u^{n-1}_{i+1/2}\right) \tag{17}$$

$$u^{n+1/2}_i = u^n_i - (1-\beta_{xt})a\frac{\Delta t}{\Delta x}\left(u^n_{i+1/2} - u^n_{i-1/2}\right) \tag{18}$$

However, substituting the pseudotime step by a characteristic pseudotime step only ensures that the normal difference terms, and not the cross-difference terms, lead to positive coefficients. Thus, the second modification consists of splitting the cross-difference terms so that they lead to a discretization equation with positive coefficients as well. As will be demonstrated, this can be accomplished through a proper specification of β_{xt} and $u^n_{i+1/2}$.

We here set β_{xt} such that the magnitude of the cross-derivative coefficient remains small compared to the one of the normal-derivative coefficient, which was a strategy that was shown in Refs. [22, 23] to be effective in improving the monotonicity of the cross-differences. Thus, following Ref. [22], we here set β_{xt} to:

$$\beta_{xt} = \frac{|a|}{|a| + \frac{\Delta x}{\Delta t}} \tag{19}$$

and we upwind u at the interface instead of taking an arithmetic average:

$$u^n_{i+1/2} = \frac{1}{2}\left(u^n_{i+1} + u^n_i\right) - \frac{|a|}{2a}\left(u^n_{i+1} - u^n_i\right) \tag{20}$$

The choice of β_{xt} and $u^n_{i+1/2}$ is not unique and there may be other formulations that would also lead to a positive coefficients. Nonetheless, the other formulations that have been tried were noticed to result in more restrictive conditions on obtaining positive coefficients, eventually leading to a lower resolution of the solution. Let us now substitute the latter 2 expressions into the flux functions outlined in (17)-(18):

$$f^n_{i+1/2} = a\frac{u^n_{i+1} + u^n_i}{2} - |a|\frac{u^n_{i+1} - u^n_i}{2} - \frac{(a-|a|)u^n_{i+1} + (a+|a|)u^n_i - (a-|a|)u^{n-1}_{i+1} - (a+|a|)u^{n-1}_i}{2|a|\frac{\Delta t}{\Delta x} + 2} \tag{21}$$

$$u^{n+1/2}_i = u^n_i - \frac{(a-|a|)u^n_{i+1} + (a+|a|)u^n_i - (a-|a|)u^{n-1}_{i+1} - (a+|a|)u^{n-1}_i}{2|a| + 2\frac{\Delta x}{\Delta t}} \tag{22}$$

It can further be shown that, when substituted back in the discretized equation (11), the following fluxes would yield the same discretized equation as the latter fluxes:

$$f^n_{i+1/2} = a\frac{u^n_{i+1} + u^n_i}{2} - |a|\frac{u^n_{i+1} - u^n_i}{2} - \frac{(a-|a|)u^n_{i+1} + (a+|a|)u^n_i - (a-|a|)u^{n-1}_{i+1} - (a+|a|)u^{n-1}_i}{|a|\frac{\Delta t}{\Delta x} + 1} \tag{23}$$

$$u^{n+1/2}_i = u^n_i \tag{24}$$

Further, multiply the cross-difference terms by the limiter ϕ as follows:

$$f^n_{i+1/2} = a\frac{u^n_{i+1} + u^n_i}{2} - |a|\frac{u^n_{i+1} - u^n_i}{2} - \frac{2}{|a|\frac{\Delta t}{\Delta x} + 1}\cdot\phi\left(\frac{|a|\Delta t}{\Delta x}\right)\cdot\left((a-|a|)\frac{u^n_{i+1} - u^{n-1}_{i+1}}{2} + (a+|a|)\frac{u^n_i - u^{n-1}_i}{2}\right) \tag{25}$$

where it is desired to set the limiter ϕ to a value as close to 1 as possible while ensuring that the discretization equation has positive coefficients. This can be accomplished with the following expression (see proof in Appendix A):

$$\phi(b) \equiv \frac{b+1}{2\max(b,1)} \tag{26}$$

Substitute the latter in the former and simplify:

$$f^n_{i+1/2} = \underbrace{a\frac{u^n_{i+1} + u^n_i}{2} - |a|\frac{u^n_{i+1} - u^n_i}{2}}_{\text{terms originating from the spatial derivative}} \underbrace{-\frac{1}{\max\left(1,\ |a|\frac{\Delta t}{\Delta x}\right)}\cdot\left((a-|a|)\frac{u^n_{i+1} - u^{n-1}_{i+1}}{2} + (a+|a|)\frac{u^n_i - u^{n-1}_i}{2}\right)}_{\text{terms originating from the temporal derivative}} \tag{27}$$

When substituted back in Eq. (11), and in contrast to previous stencils obtained through the Cauchy-Kowalevski procedure, the latter flux functions lead to a discretization equation with strictly positive coefficients. As such, the stencils here proposed yield monotone solutions free of spurious oscillations even in the vicinity of strong shocks or other discontinuities.

The discretization stencil here proposed for the temporal derivative may seem first-order accurate at first because the discretized temporal derivative becomes $(u^n_i - u^{n-1}_i)/\Delta t$ (see Eq. (24)). However, such is misleading. Because the stencil was obtained through Cauchy-Kowalevski, there are terms added to the spatial flux $f_{i+1/2}$ that are multidimensional (in space and time). Thus, the time and space derivatives can not be separated when assessing order of accuracy. When the sum of the space and time derivatives is considered together, the order of accuracy in space-time approaches 2 when the time step Δt approaches $\Delta x/|a|$. In other words, when (23) et (24) are substituted in (11), the discretization equation will have an order of accuracy ranging between 1 and 2, despite using a first-order backward stencil for the time derivative.

It can also be convenient to write Eqs. (27)-(24) in terms of modified wave speeds as follows:

$$f^n_{i+1/2} = \lambda^n_{i+1/2}\frac{u^n_{i+1} + u^n_i}{2} - |\lambda|^n_{i+1/2}\frac{u^n_{i+1} - u^n_i}{2} \tag{28}$$

where the modified wave speeds correspond to:

$$\lambda^n_{i+1/2} \equiv a - \frac{a}{\max\left(1,\ |a|\frac{\Delta t}{\Delta x}\right)} \cdot \psi\left(\frac{u^n_i - u^{n-1}_i}{u^n_i},\ \frac{u^n_{i+1} - u^{n-1}_{i+1}}{u^n_{i+1}}\right) \tag{29}$$

$$|\lambda|^n_{i+1/2} \equiv |a| - \frac{|a|}{\max\left(1,\ |a|\frac{\Delta t}{\Delta x}\right)} \cdot \psi\left(\frac{u^n_i - u^{n-1}_i}{u^n_i},\ \frac{u^n_{i+1} - u^{n-1}_{i+1}}{u^n_{i+1}}\right) \tag{30}$$

where the ψ function must return a value that has a magnitude less than either argument for the scheme to yield positive coefficients. One function that does satisfy the latter condition and that has been verified to yield a particularly high resolution as well as good convergence characteristics is the following:

$$\psi(c,d) = \text{minmod}(c,d) \cdot \min\left(1,\ \xi\frac{\max(|c|,|d|)}{\min(|c|,|d|)}\right) \tag{31}$$

where the minmod function returns the argument that has the lowest magnitude if both arguments have the same sign but zero otherwise. Also, ξ is a user-adjustable parameter which is typically set to 0.5. The higher the value given to ξ, the more compressive the scheme becomes.

When $\lambda = a$ and $|\lambda| = |a|$, Eq. (28) is clearly an upwinded scheme. Thus, the method presented here can be incorporated within a FVS or FDS solver applicable to systems simply by adding the second terms on the RHS of Eqs. (29)-(30) to the eigenvalues, as is shown next.

4. Extension to the Euler Equations

The monotone flux functions outlined in Eqs. (28)-(24) were derived for a scalar advection equation. Here, we seek to extend these fluxes to the unsteady Euler equations which can be written as:

$$\frac{\partial U}{\partial t} + A\frac{\partial U}{\partial x} = 0 \tag{32}$$

where U is the vector of conserved variables and A is the convective flux Jacobian equal to $A \equiv \partial F/\partial U$ with F the convective flux vector. It is desired that the discretization equation is such that it can be written in finite-volume form in terms of flux functions at the interfaces as follows:

$$\frac{U_i^{n+1/2} - U_i^{n-1/2}}{\Delta t} + \frac{F^n_{i+1/2} - F^n_{i-1/2}}{\Delta x} = 0 \tag{33}$$

We here focus on 4 commonly-used variants: (i) Flux Difference Splitting (FDS), (ii) FDS with reconstruction-evolution, (iii) Flux Vector Splitting (FVS), and (iv) FVS with reconstruction-evolution.

To do so, first note that when combined with the temporal discretization scheme outlined herein, any finite-volume spatial discretization scheme (such as FDS or FVS) as well as its second-or-higher order extensions can be written in the following form:

$$F^n_{i+1/2} = L_i^{-1}\Lambda_i^+ L_i U_i + L_{i+1}^{-1}\Lambda_{i+1}^- L_{i+1} U_{i+1} \tag{34}$$

$$U_i^{n+1/2} = U_i^n \tag{35}$$

where L and L^{-1} are the left and right eigenvector matrices of the convective flux Jacobian A and where the diagonal matrices $\Lambda_i^{\pm}$ vary depending on the scheme. Note that when the time level is not indicated, the time level n is implied. Interestingly, a discretization equation will maintain the positivity of the density and pressure as long as the matrix coefficients are positive [15, 24], with a positive matrix here defined as a matrix having all-positive eigenvalues and sharing the eigenvectors of the convective flux jacobian evaluated at the corresponding node. This can here be achieved by ensuring that the flux functions are such that Λ^+ and Λ^- in Eq. (35) are composed strictly of positive and negative diagonal elements, respectively.

4.1. Flux Vector Splitting

The positive and negative eigenvalues $\Lambda^{\pm}$ in Eq. (34) can be obtained for a flux vector splitting scheme of Steger-Warming type as follows. First add the spacetime wavespeeds to the Steger-Warming node velocities as follows:

$$\left[V_i^{\pm}\right]_{r,r} = \frac{1}{2}\left[\Lambda_i \pm |\Lambda_i| - \Delta^{xt}_{i\pm1/2}\left(\Lambda_i^{xt} \pm |\Lambda_i^{xt}|\right)\right]_{r,r} \tag{36}$$

with the spacetime eigenvalues set to:

$$\left[\Lambda_i^{xt}\right]_{r,r} = \frac{[\Lambda_i]_{r,r}}{\max\left(1,\ \frac{\Delta t}{\Delta x}\left|[\Lambda_i]_{r,r}\right|\right)} \tag{37}$$

and the interface spacetime gradients defined as:

$$\left[\Delta_{i\pm1/2}^{xt}\right]_{r,r} = \psi\left(\frac{\left[L_i(U_i - U_i^{n-1})\right]_r}{[L_i U_i]_r},\ \frac{\left[L_{i\pm1}(U_{i\pm1} - U_{i\pm1}^{n-1})\right]_r}{[L_{i\pm1}U_{i\pm1}]_r}\right) \tag{38}$$

In the latter, Λ corresponds to the eigenvalues of the convective flux Jacobian A, and I is the identity matrix. As well, the subscript r denotes the rth row of a vector, while the subscript r,r denotes the rth element of a diagonal matrix. Also, the ψ function is obtained from Equation (31). Simply setting the eigenvalues to their respective wave speeds (i.e. $\Lambda^{\pm} = V^{\pm}$) would result in a non-positivity-preserving scheme except for steady-state problems for which Λ^{xt} is set to zero. This can be remedied by splitting the wave speeds such that Λ_i^+ contains strictly positive elements and Λ_i^- contains strictly negative elements. The resulting discretization equation then would have all-positive coefficients and the scheme would be positivity-preserving. This can be accomplished as follows:

$$\left[\Lambda_i^-\right]_{r,r} = \min\left(0,\ \left[V_i^-\right]_{r,r}\right) + \min\left(0,\ \left[V_i^+\right]_{r,r}\right) \tag{39}$$

$$\left[\Lambda_i^+\right]_{r,r} = \max\left(0,\ \left[V_i^-\right]_{r,r}\right) + \max\left(0,\ \left[V_i^+\right]_{r,r}\right) \tag{40}$$

The latter can be extended to second-order accuracy through the use of flux limiters that retain the positive eigenvalues of the first order FVS scheme, such as those outlined in Ref. [24].

4.2. Flux Difference Splitting

We now incorporate the proposed temporal discretization stencils within FDS solvers written in general matrix form as in [25, 26]. Although the stencils could also be incorporated within FDS solvers not written in terms of the convective flux jacobian eigenvectors and eigenvalues (see Dubroca's scheme for instance [27, 28]), such would require more algebra and would be cumbersome. Thus, for a flux difference splitting algorithm in general matrix form, the left and right node wave speeds V_i^+ and V_{i+1}^- become function of the interface properties as follows:

$$\left[V_i^{\pm}\right]_{r,r} = \frac{1}{2}\left[\Lambda_i - \Lambda_i^{xt}\Delta_{i\pm1/2}^{xt}\right]_{r,r} \pm \frac{1}{2\,[L_i U_i]_r}\left[L_i L_{i\pm1/2}^{-1}\left(\left|\Lambda_{i\pm1/2}\right| - \left|\Lambda_{i\pm1/2}^{xt}\right|\Delta_{i\pm1/2}^{xt}\right)L_{i\pm1/2}U_i\right]_r \tag{41}$$

Should the $\Lambda^{\pm}$ eigenvalues be set to $V^{\pm}$ as specified in Eq. (41), the flux function outlined in Eq. (34) would become exactly the Roe FDS solver at steady-state. However, such is not positivity-preserving with the effect that negative densities or temperatures appear sometimes in the solution. This can be remedied following Ref. [26] by splitting the wave speeds between the left and right nodes such that Λ^+ has strictly positive elements and Λ^- has strictly negative elements:

$$\left[\Lambda_{i+1}^-\right]_{r,r} = \min\left(0,\ \left[V_{i+1}^-\right]_{r,r}\right) + \min\left(0,\ \left[V_i^+\right]_{r,r}\right) \tag{42}$$

$$\left[\Lambda_i^+\right]_{r,r} = \max\left(0,\ \left[V_{i+1}^-\right]_{r,r}\right) + \max\left(0,\ \left[V_i^+\right]_{r,r}\right) \tag{43}$$

The latter can be extended to second-order accuracy through TVD limiters applied to the characteristic variables that maintain Λ^- negative and Λ^+ positive, as shown in Section 5 in Ref. [26]. Although Eqs. (42)-(43) are recommended for inviscid flows, they introduce significantly more dissipation than the Roe scheme within high Reynolds number boundary layers. A solution to this problem is to implement an iterative procedure such that Λ^+ and Λ^- are as close as possible to V^+ and V^- while being composed, respectively, of strictly positive and negative elements (see Section 4.1 in Ref. [26]). In doing-so, the positivity-preserving FDS can capture with as high resolution laminar and turbulent boundary layers as the non-positivity Roe scheme as long as the boundary layer is refined sufficiently (with at least $\sim$5 nodes spanning the boundary layer height).

4.3. Flux Vector Splitting with Reconstruction-Evolution

As outlined in [29], the reconstruction-evolution technique consists of first finding left and right states obtained through a second-order-or-higher scheme (such as TVD-MUSCL or WENO) and then using these left and right states as the left and right node properties within a FVS scheme. However, the reconstruction-evolution technique can not be readily deployed here because

Table 1: Ratio of CPU time of proposed scheme and of CPU time of standard method.[a,b]

Physical system	Relaxation method		
	Explicit Euler	Block ADI	Block DDADI [30, 31]
Navier-Stokes perfect gas	1.43	1.31	1.37
Favre-Reynolds 4-species with $k\omega$ turbulence model	1.16	1.15	1.18

[a] The proposed stencils for the temporal derivative are used in conjunction with a WENO 5th order scheme reconstructed over a positivity-preserving FDS (scheme in Section 4.4).

[b] The standard method corresponds to a BDF2 temporal discretization used in conjunction with a WENO 5th order scheme reconstructed over the Roe non-positivity-preserving FDS.

some of the eigenvectors must be evaluated at the node itself rather than at the reconstructed state in order to yield a positivity-preserving algorithm. Thus, the flux function must consist of a mix of reconstructed properties (the properties denoted with a tilde) and of node properties (the properties not denoted with a tilde) as follows:

$$\left[V_i^{\pm}\right]_{r,r} = \frac{1}{2\left[L_i U_i\right]_r}\left[L_i \widetilde{L}_i^{-1}\left(\widetilde{\Lambda}_i \pm \left|\widetilde{\Lambda}_i\right| - \Delta^{xt}_{i\pm1/2}\left(\widetilde{\Lambda}^{xt}_i \pm \left|\widetilde{\Lambda}^{xt}_i\right|\right)\right)\widetilde{L}_i\widetilde{U}_i\right]_r \tag{44}$$

The latter was derived by setting the flux vector at the node i, $F_i = L_i^{-1} V_i L_i U_i$, equal to the reconstructed flux vector $\widetilde{F}_i = \widetilde{L}_i^{-1}(\cdots)\widetilde{L}_i\widetilde{U}_i$ and then isolating the node wave speeds V_i. Note that the notation here used is different from the standard notation used in reconstruction-evolution with U_L and U_R referring to the left and right states respectively. Rather, we here denote the left properties U_L by $\widetilde{U}_i$ and the right properties U_R by $\widetilde{U}_{i+1}$. Also the notation $\widetilde{L}_i$ stands for the left eigenvector matrix reconstructed from $\widetilde{U}_i$, and similarly for the reconstructed right eigenvectors and eigenvalues. After obtaining the positive and negative eigenvalues from the wave speeds using Eq. (42) and (43), and substituting such eigenvalues back in the flux function outlined in Eq. (34), we obtain a positivity-preserving FVS scheme of arbitrary order of accuracy in space through the use of reconstruction-evolution.

4.4. Flux Difference Splitting with Reconstruction-Evolution

The reconstruction-evolution strategy is not limited to flux vector splitting and can also be deployed to any monotone scheme that determines the flux at the interface from a left and right state, such as flux difference splitting schemes for instance. Denoting the reconstructed properties (obtained through a TVD or WENO scheme) by a tilde, the wave speeds of a positivity-preserving FDS solver with reconstruction-evolution take the form:

$$\left[V_i^{\pm}\right]_{r,r} = \frac{1}{2\left[L_i U_i\right]_r}\left[L_i\left(\widetilde{F}_i \pm \widetilde{L}^{-1}_{i\pm1/2}\left(\left|\widetilde{\Lambda}_{i\pm1/2}\right| - \left|\widetilde{\Lambda}^{xt}_{i\pm1/2}\right|\Delta^{xt}_{i\pm1/2}\right)\widetilde{L}_{i\pm1/2}\widetilde{U}_i\right)\right]_r - \frac{1}{2}\left[\widetilde{\Lambda}_i^{xt}\Delta^{xt}_{i\pm1/2}\right]_{r,r} \tag{45}$$

The wave speeds are then substituted in the expressions for the eigenvalues outlined in Eqs. (42) and (43).

Although the latter stencil can capture viscous layers with significantly less dissipation than a flux vector splitting method, the amount of dissipation introduced can be excessive for laminar boundary layers at high Reynolds number. Such dissipation can be reduced significantly through an iterative procedure applied to the positive and negative eigenvalues such as the one outlined in Section 4.1 in Ref. [26]. Through this iterative procedure, the reconstruction-evolution positivity-preserving FDS scheme presented here yields a skin friction coefficient and heat flux at the surfaces that is essentially the same as the one of the Roe FDS solver as long as the boundary layer is properly resolved (with typically $\sim$5 nodes or more). This iterative procedure within the FDS solver does not require as much computational effort as one may expect: as shown in Table 1, a positivity-preserving FDS solver (with 3 subiterations to capture boundary layers with high resolution) including the proposed temporal discretization requires only a 15-30% increase in CPU time per iteration over the conventional Roe FDS solver with a BDF2 temporal discretization, with the differences in CPU time becoming smaller as the number of equations within the system is increased.

4.5. Flux Difference Splitting with Reconstruction-Evolution (Non Positivity-Preserving)

The present scheme can also be used within the "standard" non-positivity-preserving Roe FDS through the use of reconstruction-evolution. This is achieved by first redefining the eigenvalues of the FDS solver as follows:

$$\left[\Lambda_i^{\pm}\right]_{r,r} = \left[V_i^{\pm}\right]_{r,r} \tag{46}$$

with the wave speeds $V_i^{\pm}$ as outlined in the previous subsection. The eigenvalue matrices $\Lambda^{\pm}$ are then substituted back in Eq. (34). Note that because the terms within Λ^+ are not necessarily all positive and the terms within Λ^- are not necessarily negative, there is no guarantee that the scheme will maintain the positivity of the determinative properties such as pressure, density, etc. However, such has the advantage of reverting at steady state to the standard FDS solver which can capture within one cell a contact discontinuity.

5. Eigenvalue Conditioning

To prevent carbuncle phenomena or other aphysical phenomena from forming, it is necessary to condition the eigenvalue matrices $\Lambda^{\pm}$ appearing in the fluxes outlined in Eq. (34). Simply adding some dissipation would prevent carbuncles but would also lead to much reduced resolution in other parts of the domain. Thus, the approach followed here is to introduce dissipation only in the vicinity of large pressure gradients (i.e., near shocks) and close to none elsewhere. This can be achieved by redefining the eigenvalues at the $i + 1/2$ interface as follows:

$$\left[\Lambda_i^+\right]_{r,r} \rightarrow \min\left(q_{\text{ref}} + \frac{a_{\text{ref}}}{\zeta_2},\ \left[\Lambda_i^+\right]_{r,r}\right) + \zeta_1 \left(q_{\text{ref}} + a_{\text{ref}}\right) \frac{P_{\max} - P_{\min}}{P_{\min}} \tag{47}$$

$$\left[\Lambda_{i+1}^-\right]_{r,r} \rightarrow \max\left(-q_{\text{ref}} - \frac{a_{\text{ref}}}{\zeta_2},\ \left[\Lambda_{i+1}^-\right]_{r,r}\right) - \zeta_1 \left(q_{\text{ref}} + a_{\text{ref}}\right) \frac{P_{\max} - P_{\min}}{P_{\min}} \tag{48}$$

where q_{ref} is a reference flow speed set to the maximum of the flow speeds at the nodes left and right of the interface and a_{ref} is a reference sound speed set to the maximum of the sound speeds at the nodes left and right of the interface. As well, $P_{\min}$ and $P_{\max}$ are the minimum and maximum pressure on all nodes neighboring the interface. For structured meshes, the number of nodes neighboring the interface corresponds to 2 in 1D, to 6 in 2D, and to 10 in 3D. The user-adjustable parameter ζ_1 is typically set between 0 and 0.3 while ζ_2 is typically set between 0.1 and 0.5. Because the dissipation added is proportional to $(P_{\max} - P_{\min})/P_{\min}$ and because the latter is only significant within 2 or 3 nodes of shockwaves, there is essentially no dissipation added in flow regions that are distant from shocks by 3 nodes or more. Thus, the latter eigenvalue conditioning affects the solution negligibly away from shocks and can thus be used in high-Reynolds-number boundary layers or with high-order schemes without an associated loss in resolution.

6. Convergence Acceleration by Preconditioning

Local pseudotime stepping, or letting the pseudotime step vary such that the waves propagate through an equal number of nodes at each iteration, is well known to result in significant gains in convergence acceleration for many problems. However, setting the pseudotime step following the CFL condition as is common practice is found to lead to particularly slow convergence in the vicinity of strong shocks or other discontinuities especially when using FDS schemes. This slow convergence can be remedied by substituting the CFL-condition-based pseudotime step by the maximum local pseudotime step that maintains the positivity of the solution. Within the context of the proposed time-accurate stencils presented in Section 4 above, such can be obtained by substituting $F_{i+1/2}^n$ and $U_i^{n+1/2}$ from Eqs. (34)-(35) in the discretization equation (33), adding a pseudotime derivative, and then imposing the condition that all coefficients within the discretization equation should be positive. The following condition involving the pseudotime step is thus obtained:

$$\frac{1}{\Delta\tau_i} I - \frac{1}{\Delta t} I - \frac{1}{\Delta x}\Lambda_i^+ + \frac{1}{\Delta x}\Lambda_i^- > 0 \tag{49}$$

Because the sum of the terms on the LHS is a matrix, its eigenvalues should be all-positive. After some algebra, it can be shown that such would occur for the following range of $\Delta\tau$ at the ith node:

$$\Delta\tau_i = \text{CFL}\frac{\Delta t}{1 + \frac{\Delta t}{\Delta x}\max_r\left[\Lambda_i^+ - \Lambda_i^-\right]_{r,r}} \tag{50}$$

where $\Lambda^{\pm}$ varies depending on the scheme as specified within Section 4 and where CFL is a user-defined parameter which should be set to a value below 1 to yield a discretization equation with coefficients that are positive. For a first-order-accurate FVS scheme, or for a first-order-accurate FDS scheme in regions with uniform properties, it can be shown that the local pseudotime step outlined here would have a form similar to the one obtained from the CFL condition. However, it does depart from the CFL condition when the schemes are turned second-order accurate or, in the case of the FDS, in regions with property gradients. In fact, in the vicinity of shocks or discontinuities, the local pseudotime step proposed in Eq. (50) can differ by several orders of magnitude from the CFL condition when used with FDS schemes.

The formulation for $\Delta\tau$ presented in Eq. (50) is such that when used with an explicit pseudotime stepping approach, positivity-preservation throughout the iteration process is guaranteed for a CFL set to a value less than 1. However, regardless of the CFL value and of the pseudotime integration strategy used, positivity-preservation of the converged solution at each physical time level is guaranteed as long as convergence can be attained. Thus, when used with an implicit pseudotime stepping approach with the CFL set to large values, although there is no guarantee that the solution will remain positivity-preserving through the iterative process, the solution will be positivity-preserving when converged sufficiently. For this reason, we here recommend not to use Eq. (50) because such would lead to too slow convergence for certain problems where there is a large variation of the magnitude of the eigenvalues. Rather, instead of using a local pseudotime step, we recommend a preconditioner with the diagonal terms of the preconditioning matrix Γ set to:

$$[\Gamma]_{r,r} = \frac{1}{\sqrt{\Delta\tau_{\text{ref}} \cdot \max\left(0.01\Delta\tau_{\text{ref}},\ \min\left(100\Delta\tau_{\text{ref}},\ \left[\Lambda^+ - \Lambda^-\right]_{r,r}\right)\right)}} \tag{51}$$

with $\Delta\tau_{\text{ref}}$ some reference time step fixed to:

$$\Delta\tau_{\text{ref}} = \text{CFL}\left(\frac{\Delta t}{1 + \frac{\Delta t}{\Delta x}\max\limits_r\left[\Lambda^+ - \Lambda^-\right]_{r,r}}\right)^{1-\sigma}\left(\frac{\Delta t}{1 + \frac{\Delta t}{\Delta x}\min\limits_r\left[\Lambda^+ - \Lambda^-\right]_{r,r}}\right)^{\sigma} \tag{52}$$

where σ is a user-defined parameter typically given a value of 0.3.

Numerous numerical experiments of stringent problems (blast waves, vacuum generation, combustion) indicate that an implicit approach combined with the latter preconditioning does converge reliably with Γ as specified in Eq. (51) with the CFL parameter set to values varying from 1 to 10.

Nonetheless, it is emphasized that the positivity-preserving and monotone properties of the proposed discretization stencils shown in the previous section do not depend in any way on the convergence acceleration by preconditioning presented here. The preconditioner shown here is strictly a convergence acceleration technique used to decrease the residual as quickly as possible at each physical time level and does not alter the converged solution. Thus, other types of convergence acceleration techniques (such as Jacobian-free Newton-Krylov for instance) could be used and would not alter the positivity-preserving or monotone properties of the proposed stencil.

7. Test Cases

Some test cases are now presented to assess the gains in resolution of the proposed temporal derivative discretization stencil which we denote as the CDF+ scheme (Cross Difference Formula Positivity-Preserving) over the BDF and Trapezoidal stencils commonly used in dual time stepping. We here use the same BDF1, BDF2, and BDF3 discretization coefficients as those outlined in Table 1 in Ref. [32] and the second-order-accurate Trapezoidal scheme outlined in [33]. For some test cases, the present method will also be compared to the second-order-accurate Total Variation Diminishing Runge-Kutta scheme (TVD-RK2) outlined in [34]. Unless otherwise specified, the test cases presented solve for the calorically-perfect and thermally-perfect Euler equations of gas dynamics with the ratio of the specific heats set to 7/5 and with the eigenvalues and eigenvectors as specified in the Appendix of Ref. [24]. This choice of eigenvectors is not trivial as the schemes proposed herein require that the eigenvectors have the property $LU \neq 0$, as is the case for the eigenvectors specified in Ref. [24] but is not the case for eigenvector sets in common use in compressible flow codes for which at least one element of the LU vector is zero. Further, unless otherwise specified, the results are obtained by combining the various temporal derivatives with a fifth-order WENO [35] reconstructed over either a FDS flux [25] or a FDS+ flux [26], with the latter chosen over the former when the temporal discretization is set to CDF+. The WENO stencil is here chosen because it is one of the preferred methods currently used to solve compressible flows either in RANS or LES formulations. Although the flux functions presented in the previous section are in Cartesian coordinates, they are here extended to generalized curvilinear coordinates simply by setting $\Delta x = \Delta y = 1$ and substituting the Cartesian eigenvector and eigenvalue matrices by their curvilinear analogues.

Because a dual time stepping algorithm strategy is employed, the time-accurate solutions can be made effectively independent of the convergence acceleration techniques used to minimize the residual at every physical time level. We here choose to accelerate the convergence through a block ADI algorithm [36, 6], the preconditioner as specified in Section 6, and the "multizone cycle" [37]. The latter consists of separating the computational domain in a large number of zones of equal size and of performing iterations only on the zones for which the residual is above a certain user-defined threshold. Although the multizone cycle does not reduce the number of subiterations, it does lead to a two- to five-fold reduction in computational effort and is such that, when convergence is attained, the residual of all nodes is guaranteed to lie below the user-defined threshold. As such, and similarly to the block-implicit integration strategy and preconditioning, it does not affect the accuracy of the solution (within the error associated with the residual threshold) when used in conjunction with a dual time stepping algorithm and is hence strictly a convergence acceleration technique.

Table 2: Relative error assessment in solving Sod's test case at $t = 0.4$ ms.[a,b]

		Average relative error					
		$\frac{1}{L}\int_{-L/2}^{L/2} \frac{\lvert\rho-\rho_{\text{exact}}\rvert}{\rho}\,\mathrm{d}x$			$\frac{1}{L}\int_{-L/2}^{L/2} \frac{\lvert T-T_{\text{exact}}\rvert}{T}\,\mathrm{d}x$		
Temporal Discretization	Spatial Discretization	$\Delta t = 5\ \mu$s	1.25 μs	0.3125 μs	$\Delta t = 5\ \mu$s	1.25 μs	0.3125 μs
TVD RK2 (explicit)	FDS TVD2	–	0.0092	0.0088	–	0.0082	0.0080
BDF2	FDS TVD2	0.0209	0.0083	0.0087	0.0197	0.0077	0.0080
Trapezoidal	FDS TVD2	0.0092	0.0085	0.0087	0.0083	0.0079	0.0080
CDF+ (present)	FDS+ TVD2	0.0145	0.0086	0.0090	0.0103	0.0080	0.0084
TVD RK2 (explicit)	FDS WENO5	–	0.0066	0.0066	–	0.0057	0.0058
BDF1	FDS WENO5	0.0370	0.0158	0.0090	0.0350	0.0157	0.0084
BDF2	FDS WENO5	0.0340	0.0069	0.0067	0.0270	0.0063	0.0059
BDF3	FDS WENO5	–	0.0067	0.0066	–	0.0061	0.0059
Trapezoidal	FDS WENO5	0.0240	0.0066	0.0066	0.0180	0.0060	0.0059
CDF+ (present)	FDS+ WENO5	0.0110	0.0059	0.0064	0.0065	0.0046	0.0055

[a] The domain length L is set to 1.2 m, the number of nodes is set to 481, and the spatial flux discretization is either a fifth-order WENO or a second-order TVD reconstructed over a FDS/FDS+ flux.

[b] For $\Delta t = 5\mu$s, 1.25 μs, and 0.3125 μs, the CFL is at the most 2.4, 0.6, and 0.15, respectively.

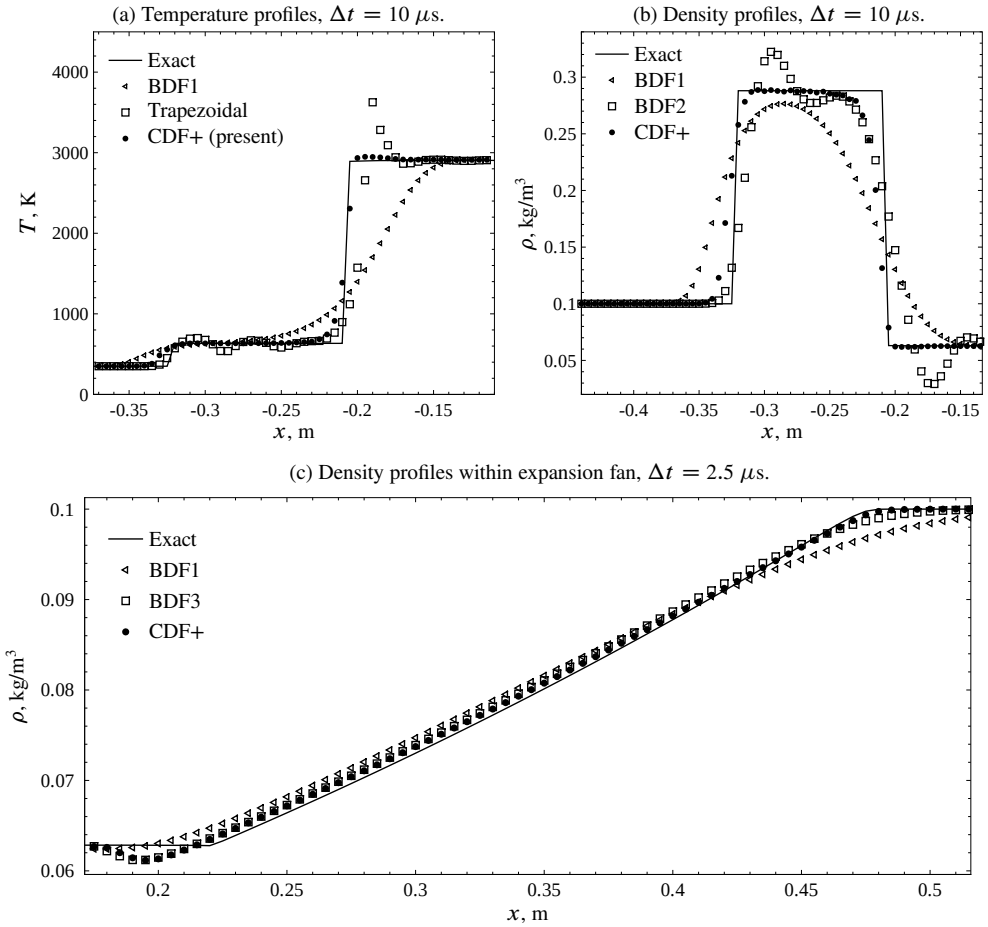

FIGURE 1. Comparison between different discretization stencils for the time derivative for the Sod test case at $t = 0.4$ ms using a grid of 241 nodes. For (a) and (b), the physical time step is such that the maximum CFL is 2.4. For (c), the physical time step is such that the maximum CFL is 0.6.

7.1. Sod Case

The first test case consists of Sod's problem in which gas lies at rest in a shocktube with two different states separated from each other through a diaphragm. At the time $t = 0$ the diaphragm is ruptured and, because of the different pressures between the two states, a shockwave, an expansion fan, and a contact discontinuity are formed and subsequently propagate through the one-dimensional shocktube. The initial conditions are such that the pressure is set to 10 kPa and 100 kPa leftward and rightward of the diaphragm respectively, whereas the density is set to 0.1 kg/m^3 within both initial states.

Such a problem can be difficult to capture accurately with the stencils commonly used to discretize the temporal derivative within dual time stepping. For instance, as is apparent from the density and temperature profiles shown in Fig. 1, the 2nd-order backward difference formula (BDF2) and the Trapezoidal rule introduce spurious overshoots and undershoots of the properties near the discontinuities. The spurious oscillations are so severe that a solution using a third-order backward difference formula (BDF3) could not be obtained for the largest time step size as it lead to negative densities and temperatures appearing after a few physical time levels. The spurious oscillations could be avoided through the use of a first-order backward formula, but at the expense of a significant loss in resolution.

The scheme proposed here does remedy this problem by yielding a solution that is as monotone as the one obtained through BDF1 while being of significantly higher resolution (see Fig. 1) not only near discontinuities but also within the expansion fan. Further, as outlined in Table 2, the present approach exhibits an error on either the density or temperature that is lower than the second-order BDF (BDF2), the third order BDF (BDF3), the Trapezoidal rule, and the explicit TVD RK2, independently of the physical time step size specified. Such is particularly noteworthy because not only does the present scheme yield a more accurate solution than all second-order schemes at high physical time steps, but it also does so for small time steps lower than the CFL condition. This is true not only in the vicinity of discontinuities but also within smooth regions. For instance, consider the density profiles within the expansion fan shown in Fig. 1c that were obtained using a time step size such that the CFL is less than 0.6. In this case, the proposed CDF+ scheme yields a solution closer to the exact solution than a third-order BDF3 stencil throughout the expansion fan. Although not shown here, a similar increase in resolution is observed over the TVD RK2 and the trapezoidal rule, and an even larger increase in resolution is observed over the BDF2.

Further, the present discretization stencil is advantaged by having a compact 2-node bandwidth in time whereas the BDF2 and BDF3 require a 3-node and 4-node bandwidth stencil in time, respectively. The higher resolution of the present scheme despite its smaller stencil bandwidth (and hence lower storage requirements) is due to the BDF assuming that the spatial and temporal derivatives can be discretized independently of each other. Such leads to excessive dissipation within flow regions where the stencils used to discretize the spatial derivatives are significantly upwinded. On the other hand, by deriving the temporal and spatial derivatives together through the novel modification of the Cauchy-Kowalevski procedure presented here, the stencil exhibits a much higher resolution despite having a smaller bandwidth and requiring less storage.

7.2. Sedov Case

One case that is commonly used to test the positivity-preserving capability of schemes is the Sedov blast wave case. The 2D Sedov case consists of calorically and thermally perfect air (with a specific heat ratio of 1.4 and gas constant of 286 J/kgK) initially at rest with a density of 1 kg/m^3 and a pressure of 106 Pa. The physical time step Δt is set to 0.02 s, the mesh is composed of $129^2 - 257^2$ nodes spanning a domain of 2.5×2.5 m^2 while the spatial discretization is set to a 5th-order WENO reconstructed over a positivity-preserving FDS. The initial pressure at the origin is set to 0.3917056 Pa·m^2/A_{cell} with A_{cell} the area of the cell at the origin. Due to the large initial pressure difference between the cell at the origin and the neighboring cells, a cylindrical shockwave of significant strength is formed and reaches a radius of 1 m after 1 second with a maximum density peak of 6 kg/m^3. The shock is followed by a rarefaction wave which eventually leads to low densities at the center of the domain. This proves to be difficult to capture with standard stencils. For instance, when using the BDF3 or Trapezoidal approaches for the time derivative, negative densities eventually appear and can not be suppressed independently of the value given to the CFL parameter. On the other hand, the BDF2, BDF1 as well as the present approach do yield a solution at all physical time levels free of negative densities and pressures as long as they are used in conjunction with spatial flux discretization schemes that have positive coefficients. As can be seen through the density profiles shown in Fig. 2, the present method leads to a significant gain in resolution over the BDF2 scheme in the vicinity of the shock as well as within the expansion fan region.

7.3. Shock Density Interaction Case

The third test case consists of the interaction between a shock and a flow with a density varying proportionally to a sine function, as specified in Ref. [38] with the parameter ϵ fixed to 0.2. Such a problem is representative of wave propagation in turbulent flows and thus constitutes a good test case to assess the efficiency of the proposed method when used for Large Eddy Simulations. The initial conditions and computational domain are as specified in [38] with the grid being composed of 340 nodes. In order to focus on the error originating from the temporal discretization scheme, the error originating from the spatial discretization must be kept low. For this reason, we here choose to discretize the spatial derivatives using the adaptive 3rd-5th order WENO proposed in Ref. [39] which, when used in conjunction with a FDS flux, has a significantly higher resolution than the standard

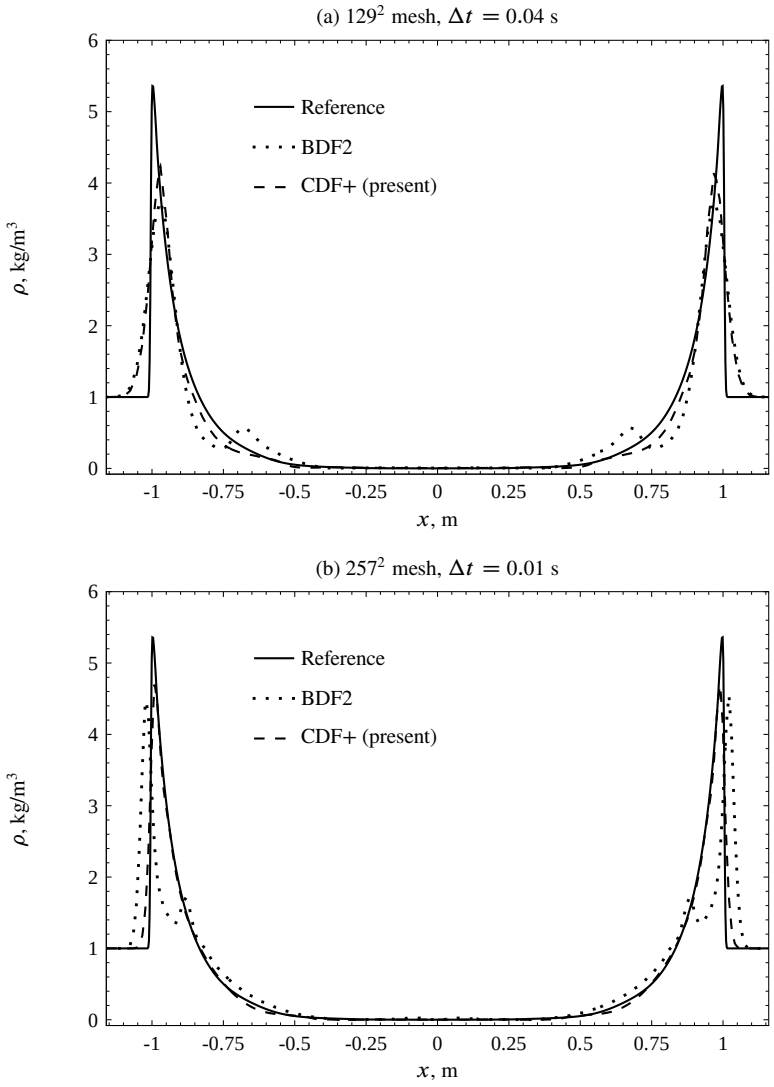

FIGURE 2: Effect of the temporal discretization stencil on the density profile at $t = 1$ s for the Sedov test case.

WENO for this problem. Further, it is ensured that the 340-node grid is fine enough so that, in the limit of an infinitely small time step, the solution obtained is very close to the reference solution.

The density is measured at a time of 1.8 seconds and is plotted in Fig. 3 for various temporal discretization schemes. The reference solution shown is obtained using a mesh of 1360 nodes and a second-order explicit TVD Runge-Kutta method. Only the density is plotted here because this is the property that experiences the largest changes and is the hardest to predict accurately. The different schemes are compared by using various time steps varying between 1.3 and 10.6 ms. At its lowest value of 1.3 ms, the time step has been verified to lead to a CFL number less than 0.25 and is hence small enough to be used in conjunction with an explicit method without leading to instabilities. At the lowest time step size considered, the scheme that exhibits the highest resolution is the proposed CDF+ which exhibits an error 10% to 40% less than any second-order scheme including the TVD Runge-Kutta (see Table 3). But an even more significant increase in resolution is displayed by the proposed method when the time step is increased and approaches the maximum physical time scales (i.e. $\Delta t = 5.3$ ms): then, the proposed method exhibits an error 2 to 5 times less than the second-order methods. What leads to this large reduction in error is the fact that, contrarily to all other temporal discretization schemes here considered, the CDF+ scheme shows little dependence on the time step used as long as the CFL is less than 1.

The high resolution exhibited by the CDF+ scheme at large time steps approaching the CFL condition leads to a very high computational efficiency, significantly higher than the one of all other implicit schemes. Because both an increase in work and an increase in error lead to a decrease in efficiency, computational efficiency is here defined as being proportional to the inverse of the product between the work and the error. The work itself is defined as the sum of the residual updates and the implicit

relaxation iterations. Such is a fair measure of the work required in this case because the CPU time necessary for one implicit relaxation sweep is more or less the same as the one needed for one residual update. To enable a fair comparison, all dual time stepping schemes use the same convergence acceleration techniques (i.e., multizone decomposition, block-ADI, and the preconditioner presented in Section 6). This has been verified to yield optimal or near-optimal computational efficiency for every scheme. As outlined in Table 3, the proposed CDF+ scheme is typically 50% more efficient than the trapezoidal method, and 2 times more efficient than BDF2 for both low and high values of the physical time step. Further, the efficiency of the CDF+ is at least half of the one of the explicit Runge-Kutta scheme at the time steps where the RK2 is the most efficient. This is a particularly strong performance of the CDF+ here because explicit methods are well known to be the most efficient time

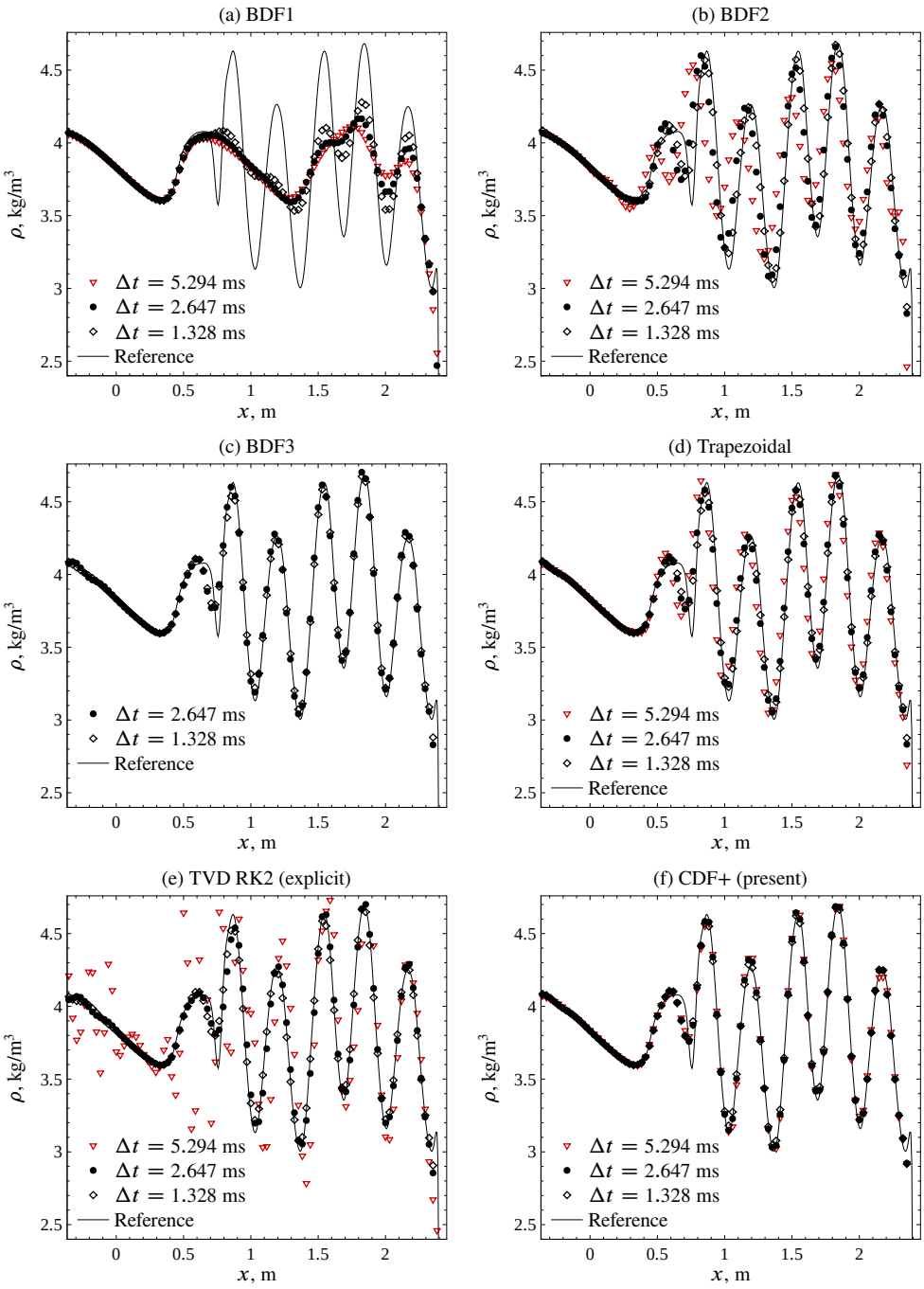

FIGURE 3. Effect of various temporal discretization schemes on density for the shock density interaction test case. The results are obtained using a 340 node grid and a spatial discretization based on either the FDS or FDS+ solver and WENO5 interpolation.

TABLE 3. Work, error, and efficiency comparison of several temporal discretization schemes in solving the shock density interaction test case with a grid of 340 nodes.

Temporal Discretization[a]	Δt, ms	Implicit Relaxation Iterations	Residual Updates	Work[b]	Error[c]	Efficiency[d]
BDF1	2.647	1480	2433	3913	0.0202	0.11
BDF1	0.662	4057	7416	11473	0.0134	0.05
BDF1	0.165	14658	27711	42369	0.0082	0.02
BDF2	5.294	911	1425	2336	0.0220	0.17
BDF2	2.647	1528	2535	4063	0.0123	0.17
BDF2	1.324	2600	4464	7064	0.0072	0.17
BDF3	5.294	–	–	–	–	–
BDF3	2.647	2048	3194	5242	0.0064	0.26
BDF3	1.324	4126	6282	10408	0.0058	0.14
TVD RK2 (explicit)	5.294	0	678	678	0.0280	0.45
TVD RK2 (explicit)	2.647	0	1358	1358	0.0063	1.00
TVD RK2 (explicit)	1.324	0	2718	2718	0.0057	0.55
Trapezoidal	10.588	659	924	1583	0.0248	0.22
Trapezoidal	5.294	797	1274	2071	0.0130	0.31
Trapezoidal	2.647	1253	2134	3387	0.0080	0.31
Trapezoidal	1.324	2174	3852	6026	0.0065	0.22
CDF+ (present)	10.588	1335	1713	3048	0.0095	0.30
CDF+ (present)	5.294	1372	2022	3394	0.0052	0.49
CDF+ (present)	2.647	1723	2756	4479	0.0051	0.37
CDF+ (present)	1.324	2618	4455	7073	0.0052	0.23

[a] All results are obtained with the spatial discretization stencil set to an adaptive third/fifth order WENO interpolation reconstructed over a FDS solver.

[b] The work is here defined as the sum of the residual updates and of the implicit relaxation iterations, and closely scales with CPU time.

[c] The error is computed as $L^{-1}\int_{-L/2}^{L/2}\rho^{-1}\,|\rho-\rho_{\text{ref}}|\;\mathrm{d}x$ with the domain length L set to 10 m and the reference density ρ_{ref} calculated using the TVD RK2 scheme using a grid of 1360 nodes.

[d] The efficiency is here calculated as 8.56/(Work × Error).

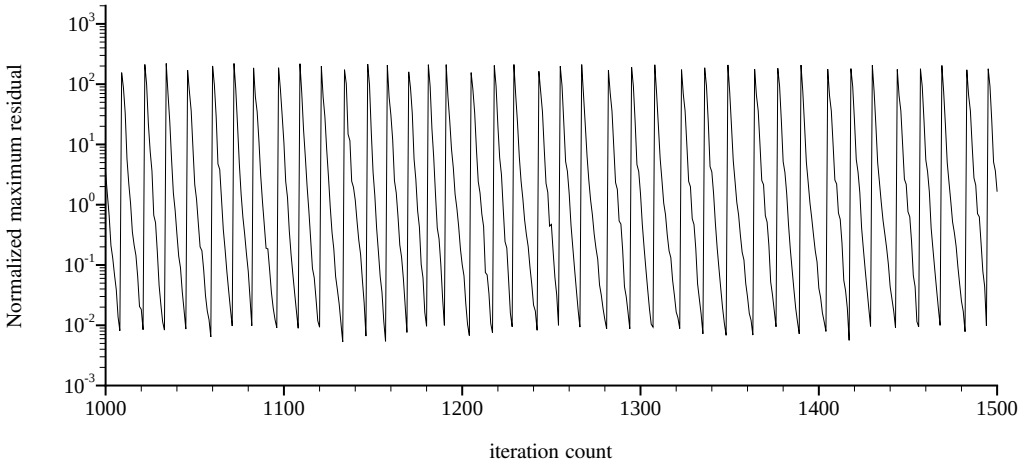

FIGURE 4. Residual convergence history for the proposed CDF+ scheme when solving the shock density interaction case with $\Delta t = 2.647$ ms. The user-specified CFL parameter is set to 6 at the first subiteration and decreased linearly such that it reaches 1 after 15 subiterations.

integration techniques to solve problems where the physical time scales don't vary much (as is the case for this inviscid non-reactive flow on a uniformly-spaced mesh). The fact that an implicit dual time stepping method has a computational efficiency so close to the one of an explicit Runge-Kutta scheme for this test case is stunning. The reason for the high performance of the CDF+ is fourfold: first, the preconditioner outlined above in Section 6 permits the residual to be converged 4 orders in a mere 12 subiterations (see Fig. 4); second, the multizone convergence acceleration [37] reduces the work fivefold by not updating the

solution in the flow regions where the residual is already below the convergence threshold effectively reducing the number of subiterations per time step from 12 to less than 3; third, the CDF+ scheme has a resolution about 20% higher than the TVD RK2 at moderate values of the time step; and fourth, because CDF+ has a lower order of accuracy, the error increases much less as the time step is increased, and this leads to an error several times lower than second-order-or-higher schemes at higher values of the time step approaching the CFL condition.

For this problem, the explicit TVD RK2 is the most computationally efficient method and the implicit CDF+ is a close second, far superior to any other implicit scheme tested. Such high performance of the TVD RK2 is however limited to simple problems (such as this one) without a high discrepancy of the physical time scales. When solving viscous flows or reactive flows where the time scales associated with the viscous terms or the chemical reactions can be orders of magnitude smaller than the one associated with the convection terms, the explicit TVD RK2 offers a computational efficiency orders of magnitude lower than the one of the implicit CDF+ as will be demonstrated in the next subsection.

7.4. Combustion Vacuum Case

The fourth test case consists of the simulation of hydrogen-air finite-rate combustion taking place within a flow in which a vacuum forms. The physical model consists of the multispecies Navier-Stokes in which the enthalpies and specific heats are determined using the NASA Glenn polynomials [40]. The chemical solver is the 9-species 20-reaction Jachimowsky model [41] with the species consisting of H_2, O_2, H, O, OH, H_2O, HO_2, H_2O_2, and N_2. The computational domain consists of a 1 m $\times$ 0.5 m block rotated 15° counterclockwise with the lower left quadrant blanked (see Fig. 5). All the boundaries are set to slip walls whereas the initial conditions consist of a mixture of air and hydrogen in near stoichiometric conditions (i.e. the oxygen, nitrogen, and hydrogen mass fractions are set to 0.235, 0.735, and 0.03 respectively) with a pressure of 10 kPa, an above-ignition temperature of 1700 K, and a Mach number along x of 4. The mesh is composed of 60 $\times$ 40 equally-spaced nodes and the flux discretization scheme is set to a 5th order WENO reconstructed over a positivity-preserving FDS, with the eigenvalue conditioning factors set as $\zeta_1 = 0.1$ and $\zeta_2 = 0.5$.

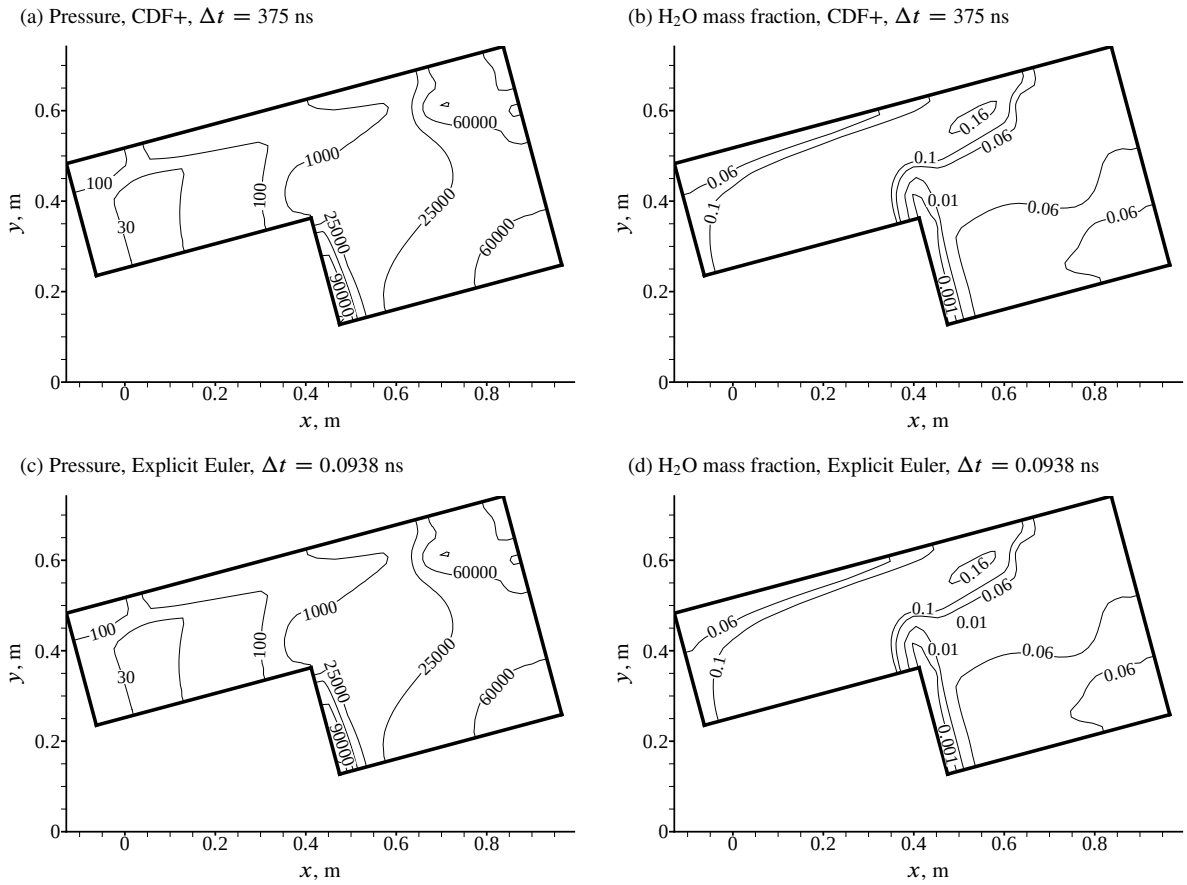

FIGURE 5. Pressure contours in Pascals and water vapor mass fractions for the combustion vacuum case at a time of 3×10^{-4} s. All results were obtained using a FDS+ WENO5 spatial discretization.

Table 4: Work needed by several temporal discretization schemes to solve the combustion vacuum test case to a similar level of accuracy.[a]

Temporal Discretization[b]	Optimal Δt, ns[c]	Average Subiterations per Time Level[d]	Implicit Relaxation Iterations	Residual Updates	Work[e]
Explicit Euler	0.0938	1	0	3,200,000	3,200,000
TVD RK2 (explicit)	0.188	2	0	3,200,000	3,200,000
CDF+ (present)	375.0	8.28	6,621	6,621	13,242

[a] All cases listed here have been verified to yield essentially the same solution at a time of 0.3 ms; computational efficiency is hence only function here of iteration count and hence work.

[b] All schemes are used in conjunction with a positivity-preserving variant of the Roe FDS solver [26] with WENO5 interpolation.

[c] For each method, the optimal Δt is found through trial and error as the one that results in the least work to obtain essentially the same solution as the one shown in Fig. 5.

[d] With CDF+, the number of subiterations is not fixed but allowed to vary such that the residual converges by 3 orders at each time level.

[e] The work is here defined as the sum of the residual updates and of the implicit relaxation iterations, and closely scales with CPU time.

As can be seen from Fig. 5 from the water vapor mass fraction contours, significant chemical reactions occur throughout the computational domain either in the zones of high density or in the zones where a vacuum forms such as near the left boundary. Thus, such is a good test case to verify whether the proposed method is positivity-preserving when simulating chemically reacting real-gas flows in the vicinity of and within a vacuum. The solution obtained with the present CDF+ scheme can be seen in Fig. 5 to be essentially the same as the one obtained with an explicit Euler temporal discretization. This is as expected because both schemes use the same spatial discretization stencils and, as such, should yield the same solution when using a small-enough time step. Because both approaches are positivity-preserving when used with a positivity-preserving FDS, no negative pressure or temperature was encountered throughout the iterative process. In the case of the dual time stepping CDF+ scheme, some negative mass fractions did occur occasionally but such took place while obtaining convergence at a particular time level. Once convergence was attained (after the residual was driven down by 3 orders of magnitude or so), it was verified at every physical time level that the solution was free of negative mass fractions. This is in contrast to the second-order-or-higher BDF stencils or the Trapezoidal scheme which yielded negative mass fractions and temperatures within the first few time levels leading to divergence independently of the value given to the CFL parameter.

Not only is the CDF+ scheme advantaged over all other implicit integration strategies here tested by being positivity-preserving, it is advantaged over a positivity-preserving explicit Euler time stepping method by being considerably more efficient. For the explicit Euler time integration, the maximum time step size that can be specified is 0.0938 nanosecond, as any higher value leads to some determinative properties becoming negative and hence to divergence towards aphysical states. Such a time step restriction is several orders of magnitude lower than the one obtained from the CFL condition and is due to the very short sub-nanosecond time scales associated with the hydrogen-air chemical reactions. This leads to excessive computational effort for this problem as more than 3 million iterations are needed to reach a time of 0.3 ms. Because the proposed method is implicit, the stability restrictions are relieved and the time step can be set to values more than 3000 times higher while still remaining positivity-preserving and yielding essentially the same solution. Because only 8 subiterations are needed at each time level to converge the residual by 3 orders of magnitude, this leads to a more than two-hundredfold increase in computational efficiency (see Table 4). Converging the residual by only three orders and specifying a time step of 375 ns is optimal in this case as no discernible difference of the pressure and mass fraction contours could be observed when converging the residual more or lowering the time step size further.

8. Conclusions

A new positivity-preserving discretization of the temporal derivative of the Euler equations is here presented that can be used within dual time stepping and which is denoted as the cross difference formula (CDF+). Such is derived along the lines of the Cauchy-Kowalevski procedure resulting in cross differences in spacetime but with some novel modifications which ensure that the resulting discretization equation has all-positive coefficients and hence yields monotone solutions. It is then shown that the so-obtained spacetime cross differences result in changes to the wave speeds and can thus be incorporated within FVS and FDS solvers (with and without reconstruction-evolution) simply by altering the eigenvalues. This is accomplished in such a way that the discretization equation has positive matrix coefficients, with a positive matrix coefficient defined as a matrix with all-positive eigenvalues and with the same eigenvectors as those of the convective flux Jacobian evaluated at the corresponding node. At least for the Euler equations of gas dynamics, such is a sufficient condition for preserving the positivity of the density and temperature.

The discretization stencils proposed herein are independent of the relaxation schemes and can be used with any convergence acceleration technique such as alternate-direction-implicit, preconditioning, Jacobian-free Newton-Krylov, block-implicit

methods, multigrid, etc. Thus, the CDF+ method presented here constitutes the first implicit positivity-preserving scheme that preserves the positivity of both the density and the temperature when solving the Euler equations of gas dynamics.

Not only is the proposed approach advantaged over alternative implicit temporal discretization stencils by being positivity-preserving, it is also advantaged by a higher resolution and higher computational efficiency. At time steps close or exceeding the physical time scales, the BDFs or the Trapezoidal rule introduce spurious oscillations near discontinuities which are further amplified should a high-order stencil be used to discretize the spatial derivatives. In contrast, the proposed CDF+ method yields monotone solutions independently of the size of the physical time step. At lower time steps significantly below the CFL condition where the standard implicit temporal discretization stencils yield monotone solutions, the proposed CDF+ often exhibits higher resolution exceeding even the one of the third-order BDF3 both in smooth flow regions and in the vicinity of discontinuities. Because the higher resolution is obtained without sacrificing on the convergence rates, the present method exhibits a computational efficiency 30% to 2 times higher than the one of the BDF2, BDF3, or the Trapezoidal rule.

A third advantage is a more compact stencil: whereas the second-order BDF2 and third-order BDF3 stencils require a 3-node and 4-node bandwidth stencil along the time coordinate, the current stencil has a mere 2-node bandwidth. The lower resolution of the BDF despite its greater stencil bandwidth (and hence higher storage requirements) is attributed to the BDF discretizing the temporal derivative independently of the spatial derivatives, leading to excessive dissipation within regions where spatial upwinding occurs. Such dissipation is here reduced by deriving the temporal and spatial derivatives conjunctly in a genuinely multidimensional manner.

When simulating strongly non-linear problems with disparate time scales such as those occurring in reactive or plasma flows, an implicit integration strategy combined with dual time stepping is often necessary to relieve the stiffness of the system and obtain high computational efficiency. For such problems, the implicit temporal discretization outlined herein is highly recommended as a substitute to the BDF or Trapezoidal stencils as it yields advantages in positivity, monotonicity, resolution, storage, as well as computational efficiency without any associated drawback.

Acknowledgement

This research was supported by a two-year research grant of Pusan National University.

A. Conditions for Cross-Difference Terms to Yield Positive Coefficients

Recall the flux at the interface in Eq. (23):

$$f^n_{i+1/2} = a\frac{u^n_{i+1}+u^n_i}{2} - |a|\frac{u^n_{i+1}-u^n_i}{2} - \phi\frac{(a-|a|)u^n_{i+1}+(a+|a|)u^n_i-(a-|a|)u^{n-1}_{i+1}-(a+|a|)u^{n-1}_i}{|a|\frac{\Delta t}{\Delta x}+1} \tag{A.1}$$

In the latter, ϕ is a limiter function multiplying the cross-difference terms which can not be greater than 1 or less than 0 and which we desire to be as close to 1 as possible. Substitute the latter fluxes in the finite volume form of the governing equations, Eq. (11), and rewrite the resulting equation in coefficient form as follows:

$$\frac{1}{\Delta\tau}u^{m+1}_i = c^n_{i+1}u^n_{i+1} + c^n_i u^n_i + c^n_{i-1}u^n_{i-1} + c^{n-1}_{i-1}u^{n-1}_{i-1} + c^{n-1}_{i+1}u^{n-1}_{i+1} + c^{n-1}_i u^{n-1}_i \tag{A.2}$$

with the discretization coefficients equal to:

$$c^{n-1}_i = \frac{1}{\Delta t} - 2|a|\frac{\phi}{|a|\Delta t+\Delta x} \tag{A.3}$$

$$c^{n-1}_{i+1} = \frac{1}{\Delta x}\left(\phi\frac{\Delta x(|a|-a)}{|a|\Delta t+\Delta x}\right) \tag{A.4}$$

$$c^{n-1}_{i-1} = \frac{1}{\Delta x}\left(\phi\frac{\Delta x(|a|+a)}{|a|\Delta t+\Delta x}\right) \tag{A.5}$$

$$c^n_{i+1} = \frac{1}{\Delta x}\left(-\frac{a}{2}+\frac{|a|}{2}+\phi\frac{\Delta x(a-|a|)}{|a|\Delta t+\Delta x}\right) \tag{A.6}$$

$$c^n_{i-1} = \frac{1}{\Delta x}\left(\frac{a}{2}+\frac{|a|}{2}-\phi\frac{\Delta x(a+|a|)}{|a|\Delta t+\Delta x}\right) \tag{A.7}$$

$$c^n_i = \frac{1}{\Delta x}\left(-\frac{a}{2}-\frac{|a|}{2}+\frac{a}{2}-\frac{|a|}{2}+\phi\frac{\Delta x(|a|+a)}{|a|\Delta t+\Delta x}+\phi\frac{\Delta x(|a|-a)}{|a|\Delta t+\Delta x}\right)-\frac{1}{\Delta t}+\frac{1}{\Delta\tau} \tag{A.8}$$

First impose positivity on the first coefficient c_i^{n-1}:

$$\frac{1}{\Delta t} - \frac{2|a|\phi}{|a|\Delta t + \Delta x} \geq 0 \tag{A.9}$$

Isolate ϕ:

$$\phi \leq \frac{1}{2} + \frac{\Delta x}{2|a|\Delta t} \tag{A.10}$$

Second, note that because ϕ can not be negative, the second and third coefficients c_{i+1}^{n-1}, c_{i-1}^{n-1} are obviously always greater or equal to zero and hence do not restrict how large ϕ can be set.

Third, impose positivity on the coefficient c_{i+1}^{n}:

$$\phi \frac{\Delta x(a - |a|)}{|a|\Delta t + \Delta x} \geq \frac{a}{2} - \frac{|a|}{2} \tag{A.11}$$

But because $a - |a|$ can only be negative or zero, the latter becomes:

$$\phi \leq \frac{1}{2}\left(\frac{|a|\Delta t + \Delta x}{\Delta x}\right) \tag{A.12}$$

Fourth, impose positivity on the coefficient c_{i-1}^{n}:

$$\phi \frac{\Delta x(a + |a|)}{|a|\Delta t + \Delta x} \leq \frac{a}{2} + \frac{|a|}{2} \tag{A.13}$$

but $a + |a|$ can not be negative:

$$\phi \frac{\Delta x}{|a|\Delta t + \Delta x} \leq \frac{1}{2} \tag{A.14}$$

or

$$\phi \leq \frac{1}{2}\left(\frac{|a|\Delta t + \Delta x}{\Delta x}\right) \tag{A.15}$$

Fifth, proceed to determine the condition on $\Delta\tau$ that ensures that the last coefficient c_i^n is always positive. After setting $c_i^n \geq 0$, isolating $\Delta\tau$, and simplifying, we obtain the following:

$$\Delta\tau \leq \frac{1}{\frac{|a|}{\Delta x} - \frac{2\phi\frac{1}{\Delta t}\frac{|a|}{\Delta x}}{\frac{|a|}{\Delta x} + \frac{1}{\Delta t}} + \frac{1}{\Delta t}} \tag{A.16}$$

In the latter, the denominator on the RHS should remain positive for the condition to remain valid (if the denominator is negative, the latter would lead to a negative $\Delta\tau$). To determine under which condition this would occur, it is convenient to make the following substitution:

$$\frac{|a|}{\Delta x} \equiv \frac{b}{\Delta t} \tag{A.17}$$

After substituting the latter in the former and imposing positivity on the denominator, this leads to the following condition on ϕ:

$$-\frac{2\phi b}{(b+1)} \geq -b - 1 \tag{A.18}$$

But $b + 1$ is always positive because b is always positive. Consequently it follows that:

$$\phi \leq \frac{(b+1)^2}{2b} \tag{A.19}$$

We can regroup the four conditions on ϕ outlined in Eqs. (A.10), (A.12), (A.15), (A.19) and recall that the upper bound on ϕ is one to obtain the following:

$$\phi = \min\left(1,\ \frac{b+1}{2b},\ \frac{b+1}{2},\ \frac{(b+1)^2}{2b}\right) \tag{A.20}$$

where $b = |a|\Delta t/\Delta x$. But because b is necessarily positive, it follows that $(b+1)^2 > b+1$. Thus the latter can be simplified to:

$$\phi = \min\left(1,\ \frac{b+1}{2b},\ \frac{b+1}{2}\right) \tag{A.21}$$

Again, because b is positive, the latter simplifies to:

$$\phi = \min\left(1,\ \frac{b+1}{2\max(b,1)}\right) \tag{A.22}$$

But, the second term within the min function can never be higher than 1. Thus:

$$\phi = \frac{b+1}{2\max(b,1)} \tag{A.23}$$

where b is defined as:

$$b \equiv \frac{|a|\Delta t}{\Delta x} \tag{A.24}$$

B. Extension to Other Space-Time Systems

Not all systems of equations can be written in the form outlined in Eq. (32). For instance, when solving a fluid model of a weakly-ionized plasma, there is a computational advantage in casting the system in this form [42]:

$$Z\frac{\partial U}{\partial t} + A\frac{\partial U}{\partial x} = \text{diffusion and source terms} \tag{B.1}$$

where A is the convective flux Jacobian, which has the usual definition $A \equiv \partial F/\partial U$, and Z is a matrix that may or may not be entirely determined from the vector U because it may also depend on some electromagnetic fields that are solved in a separate module. To solve the latter using a dual time stepping method, add a pseudotime derivative as follows:

$$\frac{\partial U}{\partial \tau} + Z\frac{\partial U}{\partial t} + A\frac{\partial U}{\partial x} = \text{diffusion and source terms} \tag{B.2}$$

We seek a discretization of the latter in the following form:

$$\frac{U_i^{m+1} - U_i^m}{\Delta\tau} + Z_i^n\frac{U_i^{n+1/2} - U_i^{n-1/2}}{\Delta t} + \frac{F_{i+1/2}^n - F_{i-1/2}^n}{\Delta x} = \text{diffusion and source terms} \tag{B.3}$$

After performing the same steps as in Sections 2-5, it can be demonstrated that $U_i^{n+1/2}$ and $F_{i+1/2}^n$ can be specified as in Eqs. (34)-(35) with the positive and negative eigenvalues as shown in Section 4 and with the spacetime difference term Δ^{xt} as in Eq. (38), but with the spacetime eigenvalues rather defined as:

$$\left[\Lambda_i^{xt}\right]_{r,r} = \frac{[\Lambda(A_i)]_{r,r}}{\max\left(1,\ \dfrac{\Delta t\cdot\left|[\Lambda(A_i)]_{r,r}\right|}{\Delta x\cdot\max\left(0,[\Lambda(Z_i)]_{r,r}\right)}\right)} \tag{B.4}$$

where $\Lambda(A)$ and $\Lambda(Z)$ stand for the eigenvalues of the matrix A and Z respectively.

Acknowledgment

This research was supported by a 2-year Pusan National University Research Grant.

References

[1] PAYRET, R. AND TAYLOR, T., *Computational Methods for Fluid Flows*, Springer, NY, 1983.

[2] JAMESON, A., "Time Dependent Calculations Using Multigrid, with Applications to Unsteady Flows Past Airfoils and Wings," 1991, 10th Computational Fluid Dynamics Conference, Honolulu HI; AIAA Paper 91-1596.

[3] VENKATESWARAN, S. AND MERKLE, C. L., "Dual time-stepping and preconditioning for unsteady computations," 1995, 33rd Aerospace Sciences Meeting and Exhibit Reno, Nevada, USA.

[4] DeRango, S. and Zingg, D. W., "Improvements to a Dual-Time-Stepping Method for Computing Unsteady Flows," *AIAA Journal*, Vol. 35, No. 9, 1997, pp. 1548–1550.

[5] Pulliam, T. H. and Chaussee, D. S., "A Diagonal Form of an Implicit Approximate-Factorization Algorithm," *Journal of Computational Physics*, Vol. 39, 1981, pp. 347–363.

[6] Beam, R. and Warming, R. F., "An Implicit Finite-Difference Algorithm for Hyperbolic Systems in Conservation-Law Form," *Journal of Computational Physics*, Vol. 22, No. 1, 1976, pp. 87–110.

[7] Beam, R. and Warming, R. F., "An Implicit Factored Scheme for the Compressible Navier-Stokes Equations," *AIAA Journal*, Vol. 16, No. 4, 1978, pp. 393–402.

[8] Steger, J. L., "Implicit Finite-Difference Simulation of Flow about Arbitrary Two-Dimensional Geometries," *AIAA Journal*, Vol. 16, No. 7, 1978, pp. 679–686.

[9] Rumsey, C. L., Sanetrik, M. D., Biedron, R. T., Melson, N. D., and Parlette, E. B., "Efficiency and accuracy of time-accurate turbulent Navier-Stokes computations," *Computers & Fluids*, Vol. 25, No. 2, 1996, pp. 217–236.

[10] Hansen, M. A. and Sutherland, J. C., "Dual Timestepping Method for Detailed Combustion Chemistry," *Combustion Theory and Modelling*, Vol. 21, No. 2, 2017, pp. 329–345.

[11] Sankaran, V. and Oefelein, J. C., "Advanced Preconditioning Strategies for Chemically Reacting Flows," 2007, paper AIAA-2007-1432, 45th AIAA Aerospace Sciences and Exhibit, Reno, NV, USA.

[12] Hitch, B. D. and Lynch, E. D., "Use of Reduced, Accurate Ethylene Combustion Mechanisms for a Hydrocarbon-Fueled Ramjet Simulation," 2009, paper AIAA-2009-5384, 45th AIAA Joint Propulsion Conference and Exhibit, Denver, CO, USA.

[13] Housman, J., Barad, M., Kiris, C., and Kwak, D., "Space-Time Convergence Analysis of a Dual-Time Stepping Method for Simulating Ignition Overpressure Waves," *Computational Fluid Dynamics 2010*, edited by A. Kuzmin, Springer, St-Petersburg, Russia, 2011, pp. 645–652.

[14] Charest, M. R. J. and Groth, C. P. T., "A High-Order Central ENO Finite-Volume Scheme for Three-Dimensional Turbulent Reactive Flows on Unstructured Mesh," 2013, paper AIAA-2013-2567, 21st AIAA Computational Fluid Dynamics Conference, San Diego, CA, USA.

[15] Parent, B., "Positivity-Preserving Flux-Limited Method for Compressible Fluid Flow," *Computers & Fluids*, Vol. 44, No. 1, 2011, pp. 238–247.

[16] Batten, P., Leschziner, M. A., and Goldberg, U. C., "Average-State Jacobians and Implicit Methods for Compressible Viscous and Turbulent Flows," *Journal of Computational Physics*, Vol. 137, 1997, pp. 38–78.

[17] Moryossef, Y. and Levy, Y., "Unconditionally Positive Implicit Procedure for Two-Equation Turbulence Models: Application to $k-\omega$ Turbulence Models," *Journal of Computational Physics*, Vol. 220, 2006, pp. 88–108.

[18] Mor-Yossef, Y. and Levy, Y., "The Unconditionally Positive-Convergent Implicit Time Integration Scheme for Two-Equation Turbulence Models: Revisited," *Computers and Fluids*, Vol. 38, 2009, pp. 1984–1994.

[19] Kuzmin, D., *A Guide to Numerical Methods for Transport Equations*, http://www.mathematik.uni-dortmund.de/ kuzmin/Transport.pdf, 2010.

[20] Toro, E. F., *Riemann Solvers and Numerical Methods for Fluid Dynamics*, Springer, 3rd ed., 2009.

[21] Huang, Y. and Lerat, A., "Second-Order Upwinding Through a Characteristic Time-Step Matrix for Compressible Flow Calculations," *Journal of Computational Physics*, Vol. 142, 1998, pp. 445–472.

[22] Parent, B., "Multidimensional Flux Difference Splitting Schemes," *AIAA Journal*, Vol. 53, No. 7, 2015, pp. 1936–1948.

[23] Parent, B., "Multidimensional High Resolution Schemes for Viscous Hypersonic Flows," *AIAA Journal*, Vol. 55, No. 1, 2017, pp. 141–152.

[24] Parent, B., "Positivity-Preserving High-Resolution Schemes for Systems of Conservation Laws," *Journal of Computational Physics*, Vol. 231, No. 1, 2012, pp. 173–189.

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

[26] Parent, B., "Positivity-Preserving Flux Difference Splitting Schemes," *Journal of Computational Physics*, Vol. 243, No. 1, 2013, pp. 194–209.

[27] 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.

[28] Dubroca, B., "Solveur de Roe Positivement Conservatif," *Comptes Rendus de l'Académie des Sciences – Series I – Mathematics*, Vol. 329, No. 9, 1999, pp. 827–832.

[29] Anderson, W. K., Thomas, J. L., and Van Leer, B., "Comparison of Finite Volume Flux Vector Splittings for the Euler Equations," *AIAA Journal*, Vol. 24, 1986, pp. 1453–1460.

[30] Bardina, J. and Lombard, C. K., "Three Dimensional Hypersonic Flow Simulations with the CSCM Implicit Upwind Navier-Stokes Method," 1987, Proceedings of the 8th Computational Fluid Dynamics Conference, AIAA Paper 87-1114.

[31] MacCormack, R., "A New Implicit Algorithm for Fluid Flow," 1997, paper AIAA-97-2100.

[32] Gerolymos, G., Sénéchal, D., and Vallet, I., "Analysis of Dual-Time-Stepping for Advection-Diffusion-Type Equations with Explicit Subiterations," 2009, paper AIAA-2009-1608, 47th AIAA Aerospace Sciences Meeting, Orlando, FL, USA.

[33] Dwight, R. P., "Time-accurate Navier-Stokes calculations with approximately factored implicit schemes," *Computational Fluid Dynamics 2004*, edited by C. Groth and D. W. Zingg, Springer, Toronto, ON, Canada, 2006, pp. 211–217.

[34] Gottlieb, S. and Shu, C. W., "Total Variation Diminishing Runge Kutta Schemes," *Mathematics of Computation*, Vol. 67, 1998, pp. 73–85.

[35] Jiang, G. and Shu, C.-W., "Efficient Implementation of Weighted ENO Schemes," *Journal of Computational Physics*, Vol. 126, 1996, pp. 202–228.

[36] Briley, W. R. and McDonald, H., "On the Structure and Use of Linearized Block Implicit Schemes," *Journal of Computational Physics*, Vol. 34, 1980, pp. 54–73.

[37] Parent, B. and Sislian, J., "The Use of Domain Decomposition in Accelerating the Convergence of Quasi-Hyperbolic Systems," *Journal of Computational Physics*, Vol. 179, No. 1, 2002, pp. 140–169.

[38] Shu, C.-W. and Osher, S. J., "Efficient implementation of essentially non-oscillatory shock capturing schemes II," *Journal of Computational Physics*, Vol. 83, 1989, pp. 32–78.

[39] Balsara, D. S., Garain, S., and Shu, C. W., "An Efficient Class of WENO schemes with Adaptive Order," *Journal of Computational Physics*, Vol. 326, 2016, pp. 780–804.

[40] McBride, B. J., Zehe, M. J., and Gordon, S., "NASA Glenn Coefficients for Calculating Thermodynamic Properties of Individual Species," TP 211556, NASA, Sept. 2002.

[41] Jachimowsky, C. J., "An Analytical Study of the Hydrogen-Air Reaction Mechanism With Application To Scramjet Combustion," TP 2791, NASA, 1988.

[42] Parent, B., Macheret, S. O., and Shneider, M. N., "Electron and Ion Transport Equations in Computational Weakly-Ionized Plasmadynamics," *Journal of Computational Physics*, Vol. 259, 2014, pp. 51–69.