On spurious numerics in solving reactive equations

By D.V. Kotov, H.C. Yee, W. Wang, and C.-W. Shu

1. Motivation and objectives

Consider 3D reactive Euler equations of the form

\[ U_t + F(U)_x + G(U)_y + H(U)_z = S(U), \]  \hspace{1cm} (1.1)

where \( U, F(U), G(U), H(U) \) and \( S(U) \) are vectors. Here, the source term \( S(U) \) is restricted to be homogeneous in \( U \); that is, \( (x, y, z) \) and \( t \) do not appear explicitly in \( S(U) \). If physical viscosities are present, viscous flux derivative should be added. If the time scale of the ordinary differential equation (ODE) \( U_t = S(U) \) for the source term is orders of magnitude smaller than the time scale of the homogeneous conservation law \( U_t + F(U)_x + G(U)_y + H(U)_z = 0 \), then the problem is said to be stiff due to the source terms. In combustion or high speed chemical reacting flows the source term represents the chemical reactions which may be much faster than the gas flow, leading to problems of numerical stiffness. Insufficient spatial/temporal resolution may cause an incorrect propagation speed of discontinuities and nonphysical states for standard numerical methods that were developed for non-reacting flows. See Wang et al. (2012) for a comprehensive overview of the last two decades of development. Schemes designed to improve the prediction of propagation speed of discontinuities for systems of stiff reacting flows remain a challenge for algorithm development (Wang et al. 2012). Wang et al. also proposed a new high order finite difference method with subcell resolution for advection equations with stiff source terms for a single reaction for (1.1) to overcome this difficulty. Research for multi-species (or more species and multi-reactions) is forthcoming.

The objective of this study is to gain a deeper understanding of the behavior of high order shock-capturing schemes for problems with stiff source terms and discontinuities and on corresponding numerical prediction strategies. The studies by Yee et al. (2012) and Wang et al. (2012) focus only on solving the reactive system by the fractional step method using the Strang splitting (Strang 1968). It is a common practice by developers in computational physics and engineering simulations to include a cut off safeguard if densities are outside the permissible range. Here we compare the spurious behavior of the same schemes by solving the fully coupled reactive system without the Strang splitting vs. using the Strang splitting. Comparison between the two procedures and the effects of a cut off safeguard is the focus the present study. The comparison of the performance of these schemes is largely based on the degree to which each method captures the correct location of the reaction front for coarse grids. Here “coarse grids” means standard mesh density requirement for accurate simulation of typical non-reacting flows of similar problem setup. It is remarked that, in order to resolve the sharp reaction front, local refinement beyond standard mesh density is still needed.

For reacting flows there are different ways in formulating (1.1). The present study considers the following two commonly used formulations. These are using all the species variables vs. using the total density and \( N_s - 1 \) number of species variables (\( N_s \) is the total number of species).
2. Overview of Two Recently Developed High Order Shock-Capturing Schemes

Here we briefly describe recently developed high order methods with subcell resolution (Wang et al. 2012) and their nonlinear filter counterparts (Yee & Sjögreen 2007, 2010).

2.1. High Order Finite Difference Methods with Subcell Resolution for Advection Equations with Stiff Source Terms (Wang et al. 2012)

The general fractional step approach is based on Strang-splitting (Strang 1968) for the 3D reactive Euler equations (1.1). The numerical solution at time level $t_{n+1}$ is approximated by

$$U^{n+1} = A \left( \frac{\Delta t}{2} \right) R(\Delta t) A \left( \frac{\Delta t}{2} \right) U^n. \quad (2.1)$$

The reaction operator $R$ is over a time step $\Delta t$ and the convection operator $A$ is over $\Delta t/2$. Except the first and last time step, the two half-step reaction operations over adjacent time steps can be combined to save cost. The convection operator $A$ is defined to approximate the solution of the homogeneous part of the problem on the time interval, i.e.,

$$U_t + F(U)_x + G(U)_y + H(U)_z = 0, \quad t_n \leq t \leq t_{n+1}. \quad (2.2)$$

The reaction operator $R$ is defined to approximate the solution on a time step of the reaction problem:

$$\frac{dU}{dt} = S(U), \quad t_n \leq t \leq t_{n+1}. \quad (2.3)$$

Here, the convection operator consists of, e.g., WENO5 with Roe flux (Jiang & Shu 1996) and RK4 for time discretization. If there is no smearing of discontinuities in the convection step, any ODE solver can be used as the reaction operator. However, all the standard shock-capturing schemes will produce a few transition points in the shock when solving the convection equation. These transition points are usually responsible for causing incorrect numerical results in the stiff case. Thus, a direct application of a standard ODE solver at these transition points will create incorrect shock speed. To avoid this, the Harten’s subcell resolution technique Harten (1989) in the reaction step is employed. The general idea is as follows. If a point is considered a transition point of the shock, information from neighboring points that are deemed not transition points will be used instead. In multidimensional case the subcell resolution procedure is applied dimension by dimension. Here only two mixture components are considered, so that $U^T = (\rho_1, \rho_2, \rho u, \rho v, \rho w, E)$, $\rho = \rho_1 + \rho_2$ and the mass fraction $z = \rho_2/\rho$ is selected for the shock indicator. The reaction operator is applied to the solution obtained after applying the subcell resolution technique.

In an earlier study Wang et al. reported that, in general, a regular $CFL = 0.1$ using the explicit Euler to solve the reaction operator step can be used in the subcell resolution scheme to produce a stable solution. But the solution in the reaction zone is not resolved enough both in space and time. In order to obtain more accurate results in the reaction zone, one reaction step can be evolved via $N_r$ sub steps, i.e.,

$$U^{n+1} = A \left( \frac{\Delta t}{2} \right) R \left( \frac{\Delta t}{N_r} \right) \cdots R \left( \frac{\Delta t}{N_r} \right) A \left( \frac{\Delta t}{2} \right) U^n \quad (2.4)$$

which have been used in some numerical examples studied in Wang et al. (2012).
2.2. Well-Balanced High Order Filter Schemes for Reacting Flows (Yee & Sjögreen 2007, 2010; Sjögreen & Yee 2004; Wang et al. 2011)

The high order nonlinear filter scheme of Yee & Sjögreen (2007, 2010); Sjögreen & Yee (2004), if used in conjunction with a dissipative portion of a well-balanced shock-capturing scheme as the nonlinear numerical flux, is a well-balanced scheme, i.e. they are able to exactly preserve specific steady-state solutions of the governing equations (Wang et al. 2011). The well-balanced high order nonlinear filter scheme for reacting flows consists of three steps.

(1) Preprocessing Step: Before the application of a high order non-dissipative spatial base scheme, in order to improve stability the pre-processing step performs splitting of inviscid flux derivatives of the governing equation(s) via the Ducros et al. splitting (Ducros et al. 2000)

(2) Base Scheme Step: A full time step is advanced using a high order non-dissipative (or very low dissipation) spatially central scheme on the split form of the governing partial differential equations (PDEs). Summation-by-parts (SBP) boundary operator Olsson (1995); Sjögreen & Yee (2007) and matching order conservative high order free stream metric evaluation for curvilinear grids (Vinokur & Yee 2002) are used. High order temporal discretization such as the third-order or fourth-order Runge-Kutta (RK3 or RK4) is used. It is remarked that other temporal discretizations can be used for the base scheme step. Numerical experiments only focused on RK4 using Roe’s approximate Riemann solver.

(3) Post-Processing (Nonlinear Filter Step): After the application of a non-dissipative high order spatial base scheme on the split form of the governing equation(s), to further improve nonlinear stability from the non-dissipative spatial base scheme, the post-processing step of Yee & Sjögreen (2007, 2010); Sjögreen & Yee (2004) nonlinearly filters the solution by a dissipative portion of a high order shock-capturing scheme with a local flow sensor. The flow sensor provides locations and amounts of built-in shock-capturing dissipation that can be further reduced or eliminated. The idea of these nonlinear filter schemes for turbulence with shocks is that, instead of solely relying on very high order high-resolution shock-capturing methods for accuracy, the filter schemes Yee et al. (1999, 2000); Sjögreen & Yee (2004); Yee & Sjögreen (2007); Yee & Sjögreen (2008) take advantage of the effectiveness of the nonlinear dissipation contained in good shock-capturing schemes as stabilizing mechanisms (a post-processing step) at locations where needed. The nonlinear dissipative portion of a high-resolution shock-capturing scheme can be any shock-capturing scheme. Unlike standard shock-capturing and/or hybrid shock-capturing methods, the nonlinear filter method requires one Riemann solve per dimension per time step, independent of time discretizations. The nonlinear filter method is more efficient than its shock-capturing method counterparts employing the same order of the respective methods. See Yee & Sjögreen (2010) for the recent improvements of the work Yee et al. (1999, 2000); Sjögreen & Yee (2004); Yee & Sjögreen (2007) that are suitable for a wide range of flow speed with minimal tuning of scheme parameters.

The nonlinear filter counterpart of the subcell resolution method (denoted by WENO5fi/SR or WENO7fi/SR) employing, e.g., WENO5 or WENO7 as the dissipative portion of the filter numerical flux (WENO5fi or WENO7fi) can be obtained by replacing the convection operator in 2.1 by the nonlinear filter scheme. Nonlinear filter schemes that include the Ducros et al. splitting of the governing equation preprocessing step are denoted by “split” such as “WENO5fi+split” and “WENO5fi/SR+split”. This Ducros et al. splitting
3. Numerical Results

The well known stiff detonation test case consisting of the Arrhenius 1D Chapman-Jouguet (C-J) detonation wave Helzel et al. (1999); Tosatto & Vigevano (2008) is considered. This is the same test case studied in Wang et al. (2012); Yee et al. (2011, 2012). Consider a 1D inviscid reactive flow containing two species, burned and unburned gas. Let \( \rho_b \) be the density of burned gas, \( \rho_u \) - the density of unburned gas, \( u \) - the mixture velocity. The mass fraction of the unburnt gas is \( z = \rho_u/\rho \), the mixture density is \( \rho = \rho_b + \rho_u \), \( p = (\gamma - 1) [E - \frac{1}{2} \rho u^2 - q_0 z] \) and \( q_0 \) is the chemical heat released. The reaction rate \( K(T) \) is modeled by an Arrhenius law \( K(T) = K_0 \exp \left(-\frac{T_{ign}}{T} \right) \), where \( K_0 \) is the reaction rate constant and \( T_{ign} \) is the ignition temperature. The initial values consist of totally burnt gas on the left-hand side and totally unburnt gas on the right-hand side. The dimensionless density, velocity, and pressure of the unburnt gas are given by \( \rho_u = 1 \), \( u_u = 0 \) and \( p_u = 1 \). The initial state of the burnt gas is calculated from the C-J condition:

\[
\begin{align*}
  p_b &= -b + (b^2 - c)^{1/2}, \\
  \rho_b &= \frac{\rho_u[p_b(\gamma + 1) - p_u]}{\gamma p_b}, \\
  S_{CJ} &= \left[ \rho_u u_u + (\gamma p_b \rho_b)^{1/2} \right]/\rho_u, \\
  u_b &= S_{CJ} - (\gamma p_b/\rho_b)^{1/2},
\end{align*}
\]

where

\[
\begin{align*}
  b &= -p_u - \rho_u q_0(\gamma - 1), \\
  c &= p_u^2 + 2(\gamma - 1)p_u \rho_u q_0/(\gamma + 1).
\end{align*}
\]

The heat release \( q_0 = 25 \) and the ratio of specific heats is set to \( \gamma = 1.4 \). The ignition temperature \( T_{ign} = 25 \) and \( K_0 = 16,418 \). The computation domain is \([0, 30]\). Initially, the discontinuity is located at \( x = 10 \). At time \( t = 1.8 \), the detonation wave has moved to \( x = 22.8 \). The reference solution is computed by the regular WENO5 scheme with 10,000 uniform grid points and \( CFL=0.05 \).

The left subfigure of Fig. 1 shows the density comparison among the standard TVD, WENO5 and WENO7 schemes using 50 uniform grid points and \( CFL = 0.05 \) for the same stiffness \( K_0 = 16,418 \) used in Yee et al. (2011). The right subfigure of Fig. 1 shows the density comparison among the less dissipative WENO5/SR, WENO5fi and WENO5fi+split schemes using the same 50 uniform grid points. All of the computations employ Strang splitting and the cut off safeguard procedure by RK4. For this particular problem and grid size, all standard TVD WENO5 and WENO7 exhibit wrong shock speed of propagation with the lower order and more dissipative schemes exhibiting the largest error. WENO5fi+split compares well with WENO5/SR for the computed pressure solution. WENO5/SR and WENO5fi+split can capture the correct structure using fewer grid points than those in Helzel et al. (1999) and Tosatto & Vigevano (2008). A careful examination of the 50 coarse grid mass fraction solutions indicates that WENO5fi+split is 0.7 grid point ahead of WENO5/SR at the discontinuity location when compared to the reference solution. Since WENO5fi+split is less dissipative than WENO5, the restriction of the shock-capturing dissipation using the wavelet flow sensor helps to improve the
On spurious numerics in solving reactive equations

Figure 1. 1D C-J detonation problem, Arrhenius case for the original stiffness $K_0$ at $t = 1.8$.
Left: Density comparison with the Reference solution (line 1) of three standard shock-capturing methods: TVD (line 2), WENO5 (line 3) and WENO7 (line 4) using 50 uniform grid points with $CFL = 0.05$. Right: Density comparison with the Reference solution (line 1) of standard high order shock-capturing methods and low dissipative methods: WENO5 (line 2), WENO5/SR (line 3), WENO5fi+split (line 4) and WENO5fi/SR+split (line 5). All of the computations use RK4, Strang splitting and the cut off safeguard procedure.

Wrong propagation speed of discontinuities without the subcell resolution procedure. For the same Strang splitting and the cut off safeguard procedure and CFL value, as the stiffness coefficient increases to $100K_0$ and $1000K_0$, the error in the prediction of the shock location increases. The spurious behavior of the studied schemes as a function of CFL for the same three stiffness coefficients and three grids 50, 150 and 300 indicates complex spurious behavior. See Fig. 2 for the less dissipative schemes behavior. Note that the error is measured in number of grid points, so that the absolute error for the grid 300 is less than for other two grids. For more details and additional test cases, see Yee et al. (2012). All of the results shown above are by RK4 temporal discretization. Previous studies Yee et al. (2012) indicated that RK4 and RK3 exhibit a similar trend but with slight variation in solution behavior for the 1D detonation problem.

3.1. Solving Fully Coupled Reactive Equations vs. Strang Splitting of the Reactive Equations

Studies show that solving the fully coupled reactive equations in conjunction with the safeguard procedure or without are very unstable for standard shock-capturing schemes as well as for their high order filter counterparts. Using a very small CFL for $K_0$, RK4, and the same three grids and CFL range, a similar wrong propagation speed of discontinuities is observed by standard shock-capturing schemes for all considered CFL (with the exception of one grid point error for WENO7 using a 50 grid), see Yee et al. (2012) for details. However, WENO5fi+split and WENO7fi+split are able to obtain the correct shock speed using the same small CFL. For stiffness coefficients $100K_0$ and $1000K_0$ using the same three grids, no stable solutions are obtained except in the case of $100K_0$ and 300 grid points using CFL $= 6.316455696 \times 10^{-3}$ (a wrong speed solution is obtained). Figures 3 and 4 summarize the comparison among (a) Strang/Safeguard, $N_r = 10$, (b) Strang/No-safeguard, $N_r$, (c) No-Strang/Safeguard and (d) No-Strang/No-safeguard for
Figure 2. 1D C-J detonation problem, Arrhenius case at $t = 1.8$. Number of grid point away from the reference shock solution ($Err$) as a function of the CFL number (128 discrete CFL values with $6.316455696 \times 10^{-3}$ equal increment) for low dissipative shock-capturing methods using 50, 150, 300 uniform grid points (across) and for stiffness $K_0, 100K_0, 1000K_0$ (top to bottom): WENO5 (line 1), WENO5/SR (line 2), WENO5fi+split (line 3) and WENO5fi/SR+split (line 4). All of the computations use Strang splitting and the cut off safeguard procedure by RK4. Note the difference in $Err$ plotting ranges.

$K_0$ and 50, 150 and 300 grid points. Note that due to the wide range of the $Err$ values by the various schemes the subplots use different $Err$ plotting ranges.

These two figures show that the same computation without the cut off safeguard procedure using the Strang splitting is also very unstable (valid CFL range is very small). For $K_0$, and the same three grids and CFL range, a similar wrong propagation speed of discontinuities is observed by WENO5 for small CFL. However, WENO5/SR and WENO5fi+split are able to obtain the correct shock speed using the same small CFL. WENO5fi/SR+split is not able to obtain the correct shock speed for even the smallest considered CFL value (CFL= $6.316455696 \times 10^{-3}$). One of the possible causes might be due to the incompatibility of the combined Strang splitting using $N_r = 10$, and the nonlinear filter procedure. For stiffness coefficients $100K_0$ and $1000K_0$ using the same three grids, no stable solutions are obtained except in the case of $100K_0$ and 300 grid points using CFL= $6.316455696 \times 10^{-3}$ (a wrong speed solution is obtained). The solution behavior when solving the fully coupled reactive equations is different from the one using the Strang splitting without the cut off safeguard procedure (see the second, third and
On spurious numerics in solving reactive equations

Figure 3. Strang splitting vs. No-Strang splitting by standard schemes for the 1D C-J detonation problem, Arrhenius case at $t = 1.8$. Number of grid points away from the reference shock solution ($\text{Err}$) as a function of the CFL number (128 discrete CFL values with $6.316455696 \times 10^{-3}$ equal increments) using 50, 150, 300 uniform grid points (across) and for stiffness $K_0$: TVD (line 1), WENO5 (line 2) and WENO7 (line 3). All of the computations use RK4. Note the difference in $\text{Err}$ plotting ranges.

fourth rows of 3 and 4). The third and fourth rows of 3 and 4 indicate that there is no significant difference in solution behavior in using the cut off safeguard procedure or not when solving the fully coupled reactive equations without Strang splitting. To further examine the difference between the two procedures in solving the reactive equations, we compare the fully coupled solution procedure with the Strang splitting procedure using a 10,000 grid. Figure 5 indicates that for fine enough grid points, both procedures produce
Figure 4. Strang splitting vs. No-Strang splitting by improved schemes for the 1D C-J detonation problem, Arrhenius case at $t = 1.8$. Number of grid points away from the reference shock solution (Err) as a function of the CFL number (128 discrete CFL values with $6.316455696 \times 10^{-3}$ equal increments) using 50, 150, 300 uniform grid points (across) and for stiffness $K_{ic}$: WENO5 (line 1), WENO5/SR (line 2), ENO5fi (line 3), WENO5fi+split (line 4) and WENO5fi/SR+split (line 5). All of the computations use RK4. Note the difference in $Err$ plotting ranges.

the same result. In summary, the combination of Strang splitting and the use of safeguard procedure resulted in very complex spurious behavior.

A similar study comparing the tow ways in formulating 1.1 using all the $N_s$ species variables vs. using the total density and $N_s - 1$ number of species variables exhibit the similar behavior with a very slight variation in the $Err$ values (Figures not shown).
3.2. Effect of the $N_r$ parameter in Strang Splitting of the Reactive Equations

Figure 6 summarizes the comparison among the different values of $N_r = 1, 5, 10, 100$ for case (a) (Strang/Safeguard) using $K_0$ and 50, 150 and 300 grid points. The results indicate that a sufficient number of sub-reaction steps improves the overall accuracy and yields a reduction in spurious numerics. Further increase of $N_r$ does not show a significant improvement.

3.3. Positivity-Preserving High Order Methods

The newly developed positivity preserving flux limiters for general high-order schemes of Hu et al. (2012) keep the original scheme unchanged and detects critical numerical fluxes may lead to negative density and pressure, and then imposes a cut-off flux limitation to satisfy a positivity preserving condition. The Hu et al. (2012) methods appears to be a better strategy than the simple safeguard procedure considered above.

Figure 7 indicates that by using the Hu et al. (2012) positivity-preserving scheme in conjunction with the Strang splitting without the safeguard procedure can avoid wrong shock speed with a slightly larger CFL than in case of using standard WENO counterparts. Here the computations use RK3. On the other hand for the same computations using Zhang & Shu (2012) positive WENO scheme indicate less improvement (Figures not shown). One method to further improve the spurious behavior is to use variable time step control. Preliminary studies indicate a significant reduction of spurious behavior in some cases when checking the positivity after each RK stage and refining the timestep by a factor of 2 in case of failing the positivity criteria.

4. Summary

The obtained results illustrate spurious behavior of numerical solution for the problems with source terms and discontinuities. For the considered test case the principal observations are as follows:

- The subcell resolution schemes Wang et al. (2012) and the filter schemes Yee &
Figure 6. $N_r = 1, 5, 10, 100$ study using Strang splitting by improved schemes for the 1D C-J detonation problem, Arrhenius case at $t = 1.8$. Number of grid points away from the reference shock solution ($\text{Err}$) as a function of the CFL number (128 discrete CFL values with $6.316455696 \times 10^{-3}$ equal increments) using 50, 150, 300 uniform grid points (across) and for stiffness $K_0$: WENO5 (line 1), WENO5/SR (line 2), WENO5fi (line 3), WENO5fi+split (line 4) and WENO5fi/SR+split (line 5). All of the computations use RK4.

Sjögreen (2007, 2010); Sjögreen & Yee (2004) in certain cases can significantly improve the results in terms of reducing spurious numerics.

- The introduction of the adhoc safeguard procedure to the numerical scheme in combination with Strang splitting can extend the valid CFL range and obtain complex spurious behavior.
- Using fully coupled reactive equations with or without the safeguard procedures is constrained by a similar CFL range as using Strang splitting without the safeguards.
Figure 7. Strang splitting no safeguard by Hu et al. positivity-preserving schemes (top) and regular schemes using Lax-Friedrichs flux (bottom) for the 1D C-J detonation problem, Arrhenius case at $t = 1.8$. Number of grid points away from the reference shock solution ($\text{Err}$) as a function of the CFL number (128 discrete CFL values with $6.316455696 \times 10^{-3}$ equal increments) using 50, 150, 300 uniform grid points (across) and for stiffness $K_0$: WENO5 (line 1), WENO5/SR (line 2), WENO5fi (line 3), WENO5fi+split (line 4) and WENO5fi/SR+split (line 5). All of the computations use RK3, Strang splitting with $N_r = 10$.

- In the case of using Strang splitting, the increase of the subiterations number $N_r$ for the reactive operator step up to a certain level can improve the results, but further increase of $N_r$ does not make a significant improvement.

- Using the positivity preserving schemes Hu et al. (2012) instead of cutting off procedure can slightly extend the valid CFL range.

Acknowledgments

The support of the DOE/SciDAC SAP grant DE-AI02-06ER25796 and the collaboration with B. Sjögreen and A. Lani in developing the infrastructure of the nonequilibrium capability in the ADPDIS3D code used for this study are acknowledged. Insightful discussions throughout the course of this work with B. Sjögreen are gratefully acknowledged. The work was performed by the first author as a postdoc fellow at the Center for Turbulence Research, Stanford University. The financial support from the NASA Fundamental Aeronautics (Hypersonic) Program for the second author is gratefully acknowledged. The research of C.-W. Shu and W. Wang is partially supported by NASA grant NNX12AJ62A.

REFERENCES