Effects of Numerical Dissipation on the Speed of Combustion Waves

by

David Lindström

Report No. 174/1996

UPPSALA UNIVERSITET
Institutionen för
teknisk databehandling

UPPSALA UNIVERSITY
Department of
Scientific Computing

Effects of Numerical Dissipation on the Speed of Combustion Waves

David Lindström[1]
Uppsala University
Department of Scientific Computing

April 3, 1996

[1]Research supported by NFR, Swedish Natural Science Research Council, grant F - FU 6329 - 303.

Abstract

Numerical methods for solving hyperbolic conservations laws usually involves addition of extra dissipation. When the conservations laws are coupled to a source term representing chemical reactions, this extra dissipation causes the reaction fronts to move with an incorrect speed of propagation. It is studied how the propagation speed depends on the order of approximation of the numerical method. Remedies are presented which reduce the error in the speed.

Contents

1 Introduction

In computations of hypersonic flow, as well as in computations of supersonic combustion problems, the chemical reactions must be modeled along with the fluid dynamics. The compressible Navier-Stokes equations, coupled to a chemical kinetics equation, can be used as a model of the flow. For inviscid flow we can use the Euler equations, coupled to a chemical source term. For simpler models, see e.g. [9].

In the systems of equations which approximates the flow, the chemical time scales usually are orders of magnitudes faster than the fluid dynamical time scales. By using an implicit ODE solver for the chemical term, a solution can be computed with time steps appropriate for the slower time scales. However, in [2], Colella, Majda and Roytburd observed that for large $K\Delta x$ (K the reaction rate, Δx the space step) the reaction waves often move at the non-physical wave speed of one grid point per time step. LeVeque and Yee explained that these spurious waves, in which all the chemical energy is released, are initiated by the few grid points in the fluid dynamical shocks, [7].

The use of intermediate states in shocks can be avoided for one dimensional equations by using the subcell resolution technique of Harten, [5]. This technique is however not successful for multi dimensional problems. Other attempts to overcome the problem of using intermediate states in the shocks are the projection technique and the extrapolation technique, [3], [4], by Engquist and Sjögreen. A modification of the extrapolation technique for shocks is introduced in Section 3.3 of this article.

Front-tracking and adaptive mesh refinement can be used to reduce the error in the numerical wave speed. An estimate of the error of the local numerical shock speed for an one-dimensional equation was derived by Klingenstein in [6].

Non-physical behavior can also be seen in strong deflagration waves for the Navier-Stokes system of equations. We have observed that the relative error in the numerical speed of propagation can be more than 100% for large $K\Delta x$. See Section 5.3 for some illustrative plots.

In [11], Tegnér investigated the effect of dissipation on the speed of reaction waves. Under certain requirements on $g(u)$ and $\Psi(u)$ he showed that the speed of a traveling wave solution to the model equation

$$u_t + g(u)u_x = -K\Psi(u) + \nu u_{xx}$$

depends on the non negative product νK. If $\nu K \ll 1$, then the speed of propagation is $O(\nu K)$ from the speed of the waves in the equation with $\nu = 0$. If $\nu K \gg 1$, then the speed is proportional to $\sqrt{\nu K}$.

In Section 2 of this article we consider the more general model equation

$$u_t + g(u)u_x = -K\Psi(u) + \nu u_x^{(p)},$$

where $u_x^{(p)}$ is the x-derivative of u of order p. The high order term $u_x^{(p)}$ is thought on as a truncation error of order Δx^{p-1} in a numerical approximation of the equation $u_t+g(u)u_x = -K\Psi(u)$. When this term is added to the equation, the speed of propagation changes. For shock waves, the change in the speed of propagation is of order $K\Delta x$. For reversed boundary conditions, the change in the speed of propagation is of order $(K\Delta x)^{p-1}$.

In some numerical experiments in Section 3 it is once again seen that the order of accuracy in the speed of propagation for a shock wave is independent of the order of accuracy of the numerical method. The accuracy of the wave speed can instead be substantially increased by using the extrapolation technique which is described in this section.

On the other hand, for the smooth traveling waves which corresponds to reversed boundary conditions, the order of accuracy in the speed of propagation depends on the order of accuracy of the numerical method. This conclusion is supported by the results of some numerical experiments in Section 4. Hence, the accuracy of the wave speed can be significantly improved if the order of the numerical method is increased. An increase of the order of accuracy in space can be done by either a change of the numerical approximation of the convective term, or by a modification of the source term. It is discussed how to modify the source term to cancel the major part of the truncation error in space.

A deflagration wave solution to the reactive Navier-Stokes equations is studied in Section 5. The numerical experiments indicates that the order of accuracy in the speed of propagation is equal to the order of accuracy of the numerical method.

2 Introduction of a model problem

2.1 Description of problem, relation to physics problem

We will use the following problem as a model for the equations of combustion in compressible fluid flow. The equation is the inviscid Burger's equation, with a source term,

$$\begin{array}{lcll} u_t + f(u)_x &=& \Psi(u), & -\infty < x < \infty \quad , \quad t \geq 0, \\ f(u) &=& u^2/2, & \\ \Psi(u) &=& -K(u+1)(u-0.1)(u-1.1) & , \quad K > 0, \\ u(x,0) &\in& [-1, 1.1]. & \end{array} \tag{1}$$

The source term of this equation can be made arbitrarily stiff by increasing the parameter K.

When solving the model problem numerically we will encounter some of the difficulties which could arise if we were solving a more complicated system of equations, such as the Euler equations of inviscid reacting gas flow.

In Section 2.3 we will analyze the effect of the truncation error in space on the propagation speed of a traveling wave. We get evidence for the fact that the order of the method is not important in the shock case, but with reversed boundary conditions, convergence is much faster with a high order method once the wave is resolved. This hypothesis is then validated numerically in Section 3 and 4.

In Section 3 we will study some numerical solutions to equation (1) with initial data

$$u_0(x) = \begin{cases} 1.1, & \text{if } x < 0, \\ -1, & \text{otherwise.} \end{cases} \tag{2}$$

For this initial data, the exact solution of (1) is the shock,

$$u(x,t) = u_0(x - 0.05t),$$

with the shock speed obtained from the Rankine-Hugoniot condition. When solving the model problem numerically, numerical dissipation will create some grid points in the shock. If the source term is evaluated in this smeared shock, the speed of the shock will be wrong. A similar thing would happen with a detonation wave solution of a more complicated system of equations.

In Section 4 we will consider the model problem (1) again, but now with initial data

$$u_0(x) = \begin{cases} -1, & \text{if } x < 0, \\ 1.1, & \text{otherwise.} \end{cases} \tag{3}$$

With this initial data, the exact solution of (1) converges with time into a smooth traveling wave solution,

$$u(x,t) = w(x - st), \qquad s = 0.1.$$

The shape of this traveling wave is given by the solution to the following ODE,

$$\begin{aligned} w'(\xi) &= -K(w+1)(w-1.1), \\ \lim_{\xi \to -\infty} w(\xi) &= -1, \\ \lim_{\xi \to \infty} w(\xi) &= 1.1. \end{aligned}$$

In the limit $K \to \infty$, the solution is a discontinuity which does not satisfy the entropy condition.

When studying this model problem in Section 4, we will get some hints on what could happen in computations of smooth deflagration waves.

2.2 Expected performance for high order schemes

When solving any equation on the form

$$u_t + g(u)u_x = -K\Psi(u) \tag{4}$$

numerically, a truncation error is introduced. Let Δx denote the distance in space between two grid points. If the order of approximation of the numerical method is p, except for an artificial dissipation term approximatively equal to $\Delta x^{p-1} u_x^{(p)}$, then the truncation error in space is approximatively equal to $\nu u_x^{(p)}$, where ν is proportional to Δx^{p-1}. We will study how the dissipative term $\nu u_x^{(p)}$ influence the speed of a traveling wave solution of (4) in Section 2.3. This is done by considering the model equation

$$u_t + g(u) u_x = -K\Psi(u) + \nu u_x^{(p)}, \tag{5}$$

where K is a positive constant and $p \geq 2$.

The special case of (5) with $p = 2$ was studied by J. Tegnér in [11]. Under the boundary conditions

$$\lim_{x \to -\infty} u(x,t) = u_L, \qquad \lim_{x \to \infty} u(x,t) = u_R,$$

where

$$u_L = u_{\max} > u_R = u_{\min}, \tag{6}$$

or

$$u_L = u_{\min} < u_R = u_{\max}, \tag{7}$$

and the following conditions on Ψ and g,

$$\begin{aligned}
&\Psi(u_{\min}) = 0, \\
&\Psi(u) > 0, \qquad \forall u \in (u_{\min}, u_0), \\
&\Psi(u_0) = 0, \\
&\Psi(u) < 0, \qquad \forall u \in (u_0, u_{\max}), \\
&\Psi(u_{\max}) = 0, \\
&\Psi'(u_{\min}) > 0, \quad \Psi'(u_0) < 0, \quad \Psi'(u_{\max}) > 0, \\
&g(u) \text{ strictly increasing}, \\
&\lim_{u \to u_0} \frac{\Psi(u)}{g(u) - g(u_0)} \neq 0 \text{ and bounded},
\end{aligned} \tag{8}$$

Tegnér showed that it exists a unique traveling wave solution to the model problem (5) with $p = 2$. For small νK, the wave speed of this traveling wave is equal to $s_0 + O(\nu K)$, where s_0 is the wave speed for the problem when $\nu K \to 0$. For $\nu K \gg 1$, the speed of the wave is proportional to $\sqrt{\nu K}$.

Hence, when solving equation (4) with a numerical method of order one in space, a traveling wave solution can be obtained, which move with an incorrect propagation speed if $\nu K \gg 1$.

In the next section we will investigate how the speed of propagation depends on K and Δx for a general p.

2.3 Scaling properties for hyper-viscosity

We note that a traveling wave solution of (5),

$$u(x,t) = w(\xi), \qquad \xi = x + st,$$

must satisfy the equation

$$(s + g(w(\xi)))w'(\xi) = -K\Psi(w(\xi)) + \nu w^{(p)}(\xi). \tag{9}$$

By making the transformation of variables $\xi = \alpha\eta$, where α is a scalar, we get

$$(s + g(\tilde{w}(\eta)))\tilde{w}'(\eta) = -K\alpha\Psi(\tilde{w}(\eta)) + \frac{\nu}{\alpha^{p-1}}\tilde{w}^{(p)}(\eta), \tag{10}$$

where $\tilde{w}(\eta) = w(\alpha\eta)$.

2.3.1 The case $u_L > u_R$.

By using the scaling factor $\alpha = \nu^{1/(p-1)}$, equation (10) becomes

$$(s + g(\tilde{w}(\eta)))\tilde{w}'(\eta) = -\delta\Psi(\tilde{w}(\eta)) + \tilde{w}^{(p)}(\eta), \tag{11}$$

where $\delta = \nu^{1/(p-1)}K$. If $\delta = 0$, equation (11) implies that

$$\lim_{N\to\infty}\int_{\tilde{w}(-N)}^{\tilde{w}(N)}(s + g(\tilde{w}))d\tilde{w} = \lim_{N\to\infty}\int_{-N}^{N}\tilde{w}^{(p)}(\eta)d\eta.$$

Since $\lim_{N\to\infty}\tilde{w}^{(p-1)}(\pm N) = 0$ for $p \geq 2$,

$$s_0 = -\frac{\int_{u_{\min}}^{u_{\max}} g(u)du}{u_{\max} - u_{\min}}. \tag{12}$$

If the asymptotic expansions $s = s_0 + \delta s_1 + \delta^2 s_2$, $\tilde{w} = \tilde{w}_0 + \delta\tilde{w}_1 + \delta^2\tilde{w}_2$ can be justified, the speed of the traveling wave solution to equation (11) is

$$s = s_0 + O(\delta) = s_0 + O(K\cdot\nu^{1/(p-1)}) = s_0 + O(K\Delta x),$$

where s_0 is the propagation speed given by (12).

2.3.2 The case $u_L < u_R$.

By using the scaling factor $\alpha = 1/K$, equation (10) becomes

$$(s + g(\tilde{w}(\eta)))\tilde{w}'(\eta) = -\Psi(\tilde{w}(\eta)) + \delta\tilde{w}^{(p)}(\eta), \tag{13}$$

where $\delta = \nu K^{p-1}$. If $\delta = 0$, equation (13) implies that

$$\int_{\tilde{w}(\eta_1)}^{\tilde{w}(\eta_2)} \frac{s + g(\tilde{w})}{\Psi(\tilde{w})} d\tilde{w} = -\int_{\eta_1}^{\eta_2} d\eta, \tag{14}$$

where η_1 and η_2 are arbitrary. We choose η_1 and η_2 such that they are bounded and such that $u_0 \in [\tilde{w}(\eta_1), \tilde{w}(\eta_2)]$. Since the right hand side of (14) is bounded, the integrand on the left hand side must be integrable. Hence

$$s = -g(u_0),$$

because of assumption (8).

If the asymptotic expansions $s = s_0 + \delta s_1 + \delta^2 s_2$, $\tilde{w} = \tilde{w}_0 + \delta \tilde{w}_1 + \delta^2 \tilde{w}_2$ can be justified, the speed of the traveling wave solution to equation (13) is

$$s = s_0 + O(\delta) = s_0 + O((K\Delta x)^{p-1}),$$

where $s_0 = -g(u_0)$.

2.3.3 Both cases.

We will now analyze what will happen with the wave speed when the truncation error in space is large, or more exactly, when $|\nu| \cdot K^{p-1} \gg 1$. By using the scaling factor $\alpha = \sqrt[p]{\nu/K}$, equation (10) becomes

$$\delta(s + g(\tilde{w}(\eta)))\tilde{w}'(\eta) = -\Psi(\tilde{w}(\eta)) + \tilde{w}^{(p)}(\eta), \tag{15}$$

where $\delta = \nu^{-1/p} K^{1/p-1}$. By integrating (15), we obtain the relation

$$\delta s I_1 + \delta I_2 = -I_3,$$

where

$$I_1 = \lim_{N\to\infty} (\tilde{w}(N) - \tilde{w}(-N)),$$

$$I_2 = \lim_{N\to\infty} \int_{\tilde{w}(-N)}^{\tilde{w}(N)} g(\tilde{w}) d\tilde{w}$$

and

$$I_3 = \int_{-\infty}^{\infty} \Psi(\tilde{w}(\eta)) d\eta.$$

If a solution of (5) exist, then it has the propagation speed

$$s = -\frac{1}{\delta} \frac{I_3 + \delta I_2}{I_1}.$$

If $I_3 \neq 0$ and $\delta \approx 0$, i.e., if $|\nu| \cdot K^{p-1} \gg 1$, then $s = O(1/\delta)$, which is proportional to $(K\Delta x)^{(p-1)/p}$. Hence the error in propagation speed gets worse for large $K\Delta x$ if the order of the method is increased.

2.3.4 Conclusion

We have seen that the scaling properties of equation (5) are such that the following is likely to hold for general p.

For $K\Delta x \ll 1$, the speed of the wave is equal to $s_0+O(K\Delta x)$ if the boundary condition is (6), and equal to $s_0 + O((K\Delta x)^{p-1})$ if the boundary condition is (7). For $K\Delta x \gg 1$, the speed of the wave is proportional to $(K\Delta x)^{(p-1)/p}$ for both boundary conditions.

We remark that in the shock case, where the error in the propagation speed is of order $K\Delta x$, the order of the method is not important. However, with reversed boundary conditions, convergence is much faster with a high order method once the layer is resolved.

3 Numerical study of the shock case

3.1 The numerical method

To solve the model problem (1) numerically, we discretize in space and time. We use an equidistant grid, where the distance in space between two grid points is Δx. The time step is denoted Δt. We define the operators I, E, Δ_-, Δ_+, and Δ_0 by

$$Iu_j = u_j, \quad Eu_j = u_{j+1},$$

$$\Delta_- = (I - E^{-1}), \quad \Delta_+ = (E - I), \quad \Delta_0 = (\Delta_- + \Delta_+)/2.$$

We will also use the notations

$$D_- = \Delta_-/\Delta x, \quad D_+ = \Delta_+/\Delta x, \quad D_0 = \Delta_0/\Delta x.$$

Moreover we denote the approximative solution of $u(x,t)$ at $x = x_j = x_0 + j\Delta x$ and $t = t_n = n\Delta t$, by u_j^n. The numerical solution is calculated using a finite difference method as follows.

First we approximate $-f(u(x,t_n))_x$ at $x = x_j$, with $L(u^n)_j$ in a conservative way, i.e., with a difference of two numerical flux functions,

$$-f(u(x_j,t_n))_x \approx L(u)_j = -\frac{h_{j+\frac{1}{2}} - h_{j-\frac{1}{2}}}{\Delta x}.$$

These numerical flux functions depend linearly on r consecutive function values, r the order of the scheme in space. They are chosen such that either $L(u)_j$ becomes a central difference approximation, or in an ENO fashion, given by the central biased ENO scheme by Lindström, [8]. For order 1 and 2 in space the central biased ENO scheme [8] is identical to the standard ENO scheme of Shu and Osher, [10]. For higher orders it uses more central fluxes for stability reasons.

For the time discretization we use the following Runge-Kutta scheme, which is described in [4],

$$\begin{aligned} u_j^{(1)} &= u_j^n + \Delta t L(u^n)_j + \Delta t \Psi(u_j^n) \\ u_j^{(2)} &= \tfrac{3}{4}u_j^n + \tfrac{1}{4}u_j^{(1)} + \tfrac{1}{4}\Delta t L(u^{(1)})_j - \tfrac{3}{4}\Delta t \Psi(u_j^{(1)}) + \Delta t \Psi(u_j^{(2)}) \\ u_j^{n+1} &= \tfrac{1}{3}u_j^n + \tfrac{2}{3}u_j^{(2)} + \tfrac{2}{3}\Delta t L(u^{(2)})_j + \tfrac{2}{3}\Delta t \Psi(u_j^{n+1}). \end{aligned} \tag{16}$$

In this scheme, the source term is added to the third order TVD Runge-Kutta scheme constructed by Shu and Osher, [10]. It is added in such a way that the resulting method, (16), is A-stable when $f(u)_x = 0$. Furthermore, the method is stable for scalar linear problems when the convective term satisfies a CFL condition.

We will use the third order Runge-Kutta method in all numerical experiments, to be able to compare the results obtained for different approximations of $L(u)_j$. We will refer to (16) as the implicit Runge-Kutta scheme of third order. When we refer to the explicit Runge-Kutta scheme of third order, we have replaced $\Psi(u_j^{(2)})$ with $\Psi(u_j^{(1)})$ in stage 2 and $\Psi(u_j^{n+1})$ with $\Psi(u_j^{(2)})$ in stage 3.

3.2 Higher accuracy gives no improvement in shock speed

To solve the model equation (1) with initial data (2), we let the computational domain contain $\max(100, 1/\Delta x)$ grid points. Thus we domain is extended for large Δx. This is done to permit the wave to approach u_L and u_R at the endpoints. The same number of grid points will be used in Section 4, where the boundary conditions are reversed. In all numerical experiments in Section 3.2 and 3.3 the time step is set to $\Delta t = \Delta x/2$.

For the model problem, where $\Psi(u) = -100(u+1)(u-0.1)(u-1.1)$,

$$\frac{d\Psi}{du}(u) \in [-231, 110], \quad \forall u \in [-1, 1.1].$$

Thus we can use the explicit Runge-Kutta scheme in the numerical experiments, where $\Delta t \leq 0.01$.

To measure the speed of the shock we do as follows. We start with the initial data given by (2) and take 3000 time steps to make the wave stabilize. Then we take another 1000 time steps and calculate the shock speed as

$$s^* = \sum_{j=1}^{N} \frac{(u_j^{4000} - u_j^{3000}) \cdot \Delta x}{(u_L - u_R) \cdot 1000 \cdot \Delta t}, \tag{17}$$

where $u_L = 1.1$ and $u_R = -1$ for the initial data (2).

To prevent the wave from traveling out of the computational domain, the wave is

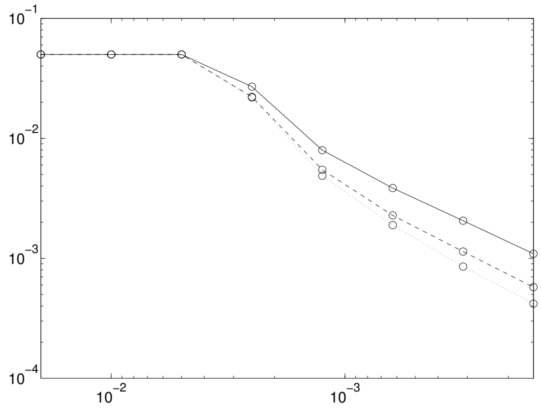


Figure 1: *The absolute error in shock speed as a function of* Δx. *Numerical results obtained when* f_x *is approximated by a central difference of order two are displayed with a drawn line. Results obtained with central difference approximations of* f_x *of order four and six are displayed with a dashed line and a dotted line respectively.*

moved such that

$$\begin{cases} u_j^n > \dfrac{u_L + u_R}{2}, & \forall j \leq N/2, \\ u_{N/2+1}^n \leq \dfrac{u_L + u_R}{2}, \end{cases}$$

is fulfilled after every time step.

In Figure 1 we show the numerical results obtained by approximating $f(u)_x$ with central differences of order 2, 4 and 6, e.g., $f(u)_x \approx D_0 f_j$. To prevent that the overshoots, which are created at the discontinuity, are traveling into the smooth part of the solution, the artificial viscosity term $-\Delta_+^3 \Delta_-^2 u_j/80$ is added to every numerical flux function, $h_{j+1/2}$. For the time integration the explicit version of the third order Runge-Kutta method is used. We see that the order of convergence in shock speed is one for all three central difference approximations of $f(u)_x$. For $\Delta x \geq 1/200$ the relative error in the numerical shock speed is 100%.

In Figure 2 we display the error in the shock speed obtained by approximating $f(u)_x$ with an ENO approximation of order 1, 2, 3, 4 and 5 respectively, given by the central biased ENO method [8]. Numerical solutions at $t = 0.5$ are displayed in Figure 3. For convex functions $f(u)$, the central biased ENO scheme [8] need to be entropy corrected whenever

$$\frac{df(u_j)}{du} < 0 < \frac{df(u_{j+1})}{du}.$$

For the model problem (1), this condition is equivalent to $u_j < 0 < u_{j+1}$, which never happens when the initial data is chosen according to (2). Since no major overshoots are created at the discontinuity when $f(u)_x$ is approximated by an ENO approximation, we

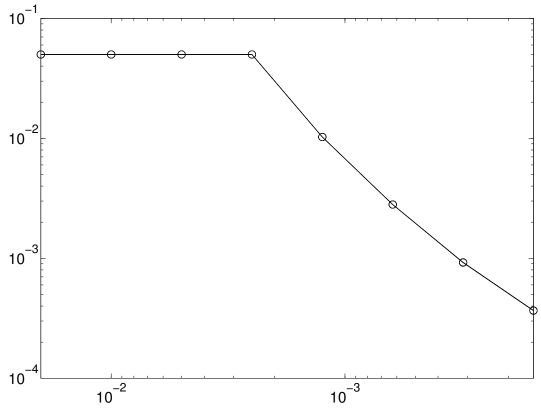


Figure 2: *The absolute error in shock speed as a function of Δx. ENO approximation of f_x of order 1, 2, 3, 4 and 5. The difference between the shock speeds given by the different methods is so small that it can not be seen in this plot.*

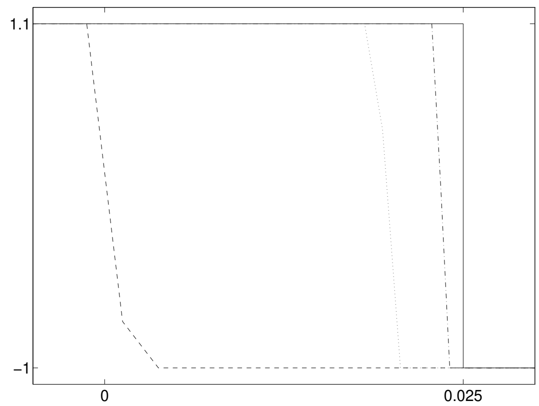


Figure 3: *Numerical solutions at $t = 0.5$ obtained with $\Delta x = 1/400$ (dashed), $\Delta x = 1/800$ (dotted) and $\Delta x = 1/1600$ (dot-dashed). ENO approximation of f_x of order 2. The exact solution is displayed with a drawn line.*

do not add any artificial viscosity term to the numerical approximation of $f(u)_x$. Once again the order of convergence in the speed of the shock is one for all approximations of $f(u)_x$.

3.3 Extrapolation of the source term

When $u(x)$ is discontinuous it is not meaningful to talk about order of approximation of the numerical method. We have just seen that the order of convergence in shock speed is one for the model problem (1), independently of the order of the numerical method. The accuracy in the shock speed can however be increased considerably by using the following extrapolation procedure.

In every stage of the Runge-Kutta method, if

$$s_j > \max(s_{j-1}, s_{j+1}) + \epsilon_s, \tag{18}$$

where s_j is defined by

$$s_j = \min(|\Delta_- u_j|, |\Delta_+ u_j|), \tag{19}$$

then extrapolate Ψ_j according to

$$\Psi_j = \begin{cases} 2\Psi(u_{j-1}) - \Psi(u_{j-2}) & \text{if } |\Delta_- u_j| < |\Delta_+ u_j|, \\ 2\Psi(u_{j+1}) - \Psi(u_{j+2}) & \text{otherwise,} \end{cases} \tag{20}$$

or

$$\Psi_j = \begin{cases} \Psi(2u_{j-1} - u_{j-2}) & \text{if } |\Delta_- u_j| < |\Delta_+ u_j|, \\ \Psi(2u_{j+1} - u_{j+2}) & \text{otherwise.} \end{cases} \tag{21}$$

The small positive constant ϵ_s in (18) is included to prevent small oscillations in a smooth solution from activating the extrapolation. In all numerical computations in this section, ϵ_s is set to 10^{-6}. Other orders of extrapolation could also be used in modification (20) and (21) of the source term, e.g., $\Psi_j = \Psi(u_{j\pm1})$.

If the numerical method is of a higher order of approximation than the order of approximation in the extrapolation (20) or (21), the extrapolation will destroy the accuracy in the smooth part of the solution if it is applied there. To prevent that the extrapolation is used in a deflagration wave, we also demand that

$$a_{j-1/2} > a_{j+1/2} + \epsilon_a$$

before the modification (20) or (21) of the source term is done. Here the local shock speed, $a_{j+1/2}$, is given by

$$a_{j+1/2} = \begin{cases} \dfrac{f_{j+1} - f_j}{u_{j+1} - u_j}, & \text{if } u_{j+1} \neq u_j, \\ f'(u_j), & \text{otherwise} \end{cases} \tag{22}$$

and ϵ_a is a small positive constant. In all numerical computations in this section, ϵ_a is set to $0.5 \cdot 10^{-6}$.

By approximating $f(u)_x$ by an ENO approximation of order 3, given by the central biased ENO method [8] and extrapolating the source term according to (20) or (21), the speed of the shock is exact for $\Delta x = 1/100, 1/200, \ldots, 1/6400$. The shock speed is measured between time level t_{60000} and t_{61000} in these numerical experiments to allow the wave to stabilize.

3.4 Source term extrapolation in two space dimensions.

For two dimensional problems we generalize the extrapolation of the source term as follows. First we define the logical variables $e_x(i,j)$ and $e_y(i,j)$. For all i, j, $e_x(i,j)$ is true if and only if

$$s_x(i,j) > \max(s_x(i-1,j), s_x(i+1,j)) + \epsilon_s$$

and

$$a_{i-1/2,j} > a_{i+1/2,j} + \epsilon_a,$$

where

$$s_x(i,j) = \min(|\Delta_{+i} u_{i,j}|, |\Delta_{-i} u_{i,j}|), \quad \forall i, j$$

and

$$a_{i+1/2,j} = \begin{cases} \dfrac{f_{i+1,j} - f_{i,j}}{u_{i+1,j} - u_{i,j}}, & \text{if } u_{j+1} \neq u_j, \\ f'(u_{i,j}), & \text{otherwise.} \end{cases}$$

Here ϵ_s and ϵ_a are small positive constants. In the numerical test in this section we set $\epsilon_s = \epsilon_a = 10^{-6}$.

In the same way, $e_y(i,j)$ is true if and only if

$$s_y(i,j) > \max(s_y(i,j-1), s_y(i,j+1)) + \epsilon_s$$

and

$$a_{i,j-1/2} > a_{i,j+1/2} + \epsilon_a,$$

where

$$s_y(i,j) = \min(|\Delta_{+j} u_{i,j}|, |\Delta_{-j} u_{i,j}|).$$

If $e_x(i,j)$ is true, but not $e_y(i,j)$, then the source term is calculated by an extrapolation procedure in the smoothest x-direction, i.e.,

$$\Psi_{i,j} = \begin{cases} 2\Psi(u_{i-1,j}) - \Psi(u_{i-2,j}) & \text{if } |\Delta_{-i} u_{i,j}| < |\Delta_{+i} u_{i,j}|, \\ 2\Psi(u_{i+1,j}) - \Psi(u_{i+2,j}) & \text{otherwise.} \end{cases}$$

If only $e_y(i,j)$ is true, then the source term is calculated by an extrapolation procedure in the smoothest y-direction, and if both $e_x(i,j)$ and $e_y(i,j)$ are true, then the source term is extrapolated in the smoothest diagonal direction, i.e.,

$$\Psi_{i,j} = 2\Psi(u_{i+p,j+q}) - \Psi(u_{i+2p,j+2q}), \tag{23}$$

where

$$p = \text{sign}(|\Delta_{-i} u_{i,j}| - |\Delta_{+i} u_{i,j}|),$$

and

$$q = \text{sign}(|\Delta_{-j} u_{i,j}| - |\Delta_{+j} u_{i,j}|).$$

To test the extrapolation algorithm (23) above, we consider the following model problem, which is a generalization of the one dimensional model problem (1) to two space dimensions.

$$
\begin{array}{lcl}
u_t + f(u)_x + g(u)_y & = & \Psi(u), \qquad -\infty < x < \infty, \qquad -\infty < y < \infty, \qquad t \geq 0, \\
f(u) & = & u^2/2, \\
g(u) & = & u^2/2, \\
\Psi(u) & = & -K(u+1)(u-0.1)(u-1.1), \qquad K > 0, \\
u(x,y,0) & = & \begin{cases} 1.1, & \text{if } x^2+y^2 \leq 0.25, \\ -1, & \text{otherwise.} \end{cases}
\end{array}
$$

The source term in this model problem does not have any effect on the exact solution, but if K is large it will affect any approximative numerical solution heavily.

In Figure 4 we display numerical solutions at $t = 4$ obtained by approximating $f(u)_x$ with an ENO approximation of order 2, given by the central biased ENO method [8]. 201×201 grid points are used to resolve the solution, with $\Delta x = \Delta y = 1/200$. For the time integration the explicit version of the third order Runge-Kutta method is used, with the time step set to $\Delta t = \Delta x/3.3$. At the boundaries the exact solution is enforced. The results in Figure 4(a) are obtained for $K = 0$. The displayed contour lines of the numerical solution are $u = -1, -0.9, \ldots, 1.1$. $K = 100$ gives the results displayed in Figure 4(b) and 4(c). The results in Figure 4(c) are obtained by modifying the source term according to algorithm (23) above. In Figure 4(b) no extrapolation was made. The stiffness of the source term prevents the shock from propagating. This phenomenon is analyzed in [3]. When the source term is extrapolated according to (23), the correct shock speed is obtained.

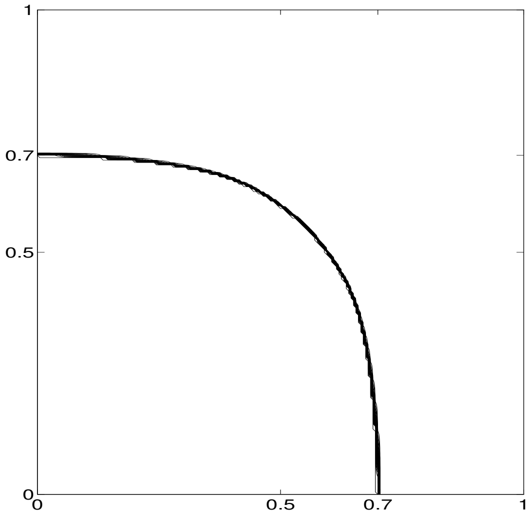
4 Numerical study of the case with reversed boundary conditions

In this section we will study the model problem (1) with (3) as initial data. The analytical solution of this model problem converges with time into a smooth traveling wave solution,

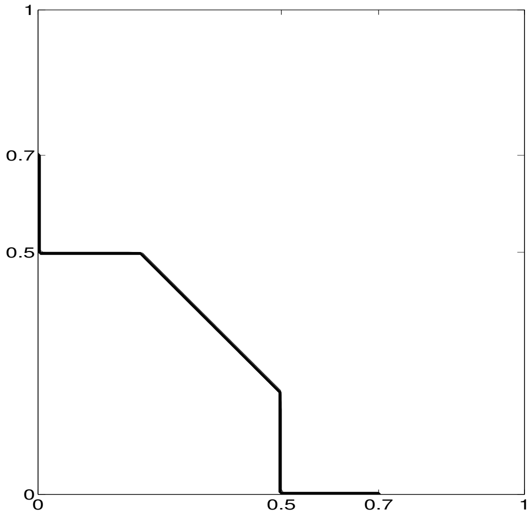
$$u(x,t) = w(x - st), \qquad s = 0.1.$$

In Section 4.1, we will investigate the effect of the resolution and the order of the numerical method on the wave speed. We conclude that the order of convergence in the wave speed is equal to the order of approximation of the numerical method. For small Δx, the accuracy of the numerical wave speed is significantly increased when the order of the method is increased. However, on a coarse mesh, the accuracy of the resulting wave speed is as bad for the higher order methods as for the numerical methods of lower order.

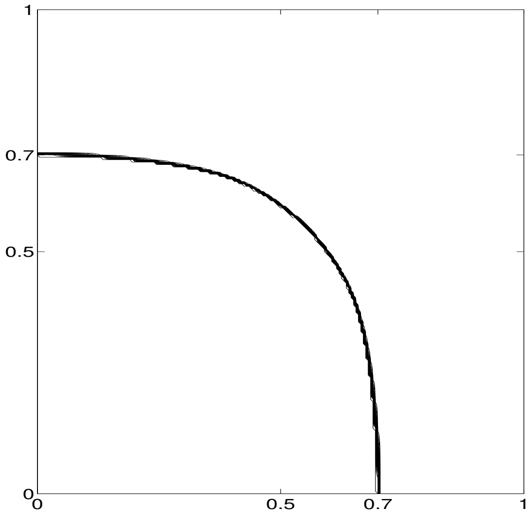
In Section 4.2 and 4.3, we introduce two different techniques for reducing the error in the numerical wave speed when the resolution is low, the artificial grid point technique,



(a) $K = 0$.



(b) $K = 100$.



(c) $K = 100$, *modified source term.*

Figure 4: *Numerical solutions of a scalar model problem at* $t = 4$.

and the artificial viscosity technique. The artificial viscosity technique turns out to be the most effective.

In Section 4.4, the accuracy of the numerical wave speed is enhanced by a tuning procedure. In this section the source term is modified according to $(I + \alpha\Delta_-^p\Delta_+^q)\Psi(u_j^n)$ and $\Psi((I + \alpha\Delta_-^p\Delta_+^q)u_j^n)$ respectively. α is the parameter which is tuned. For sufficiently small Δx, the optimal tuning corresponds to a cancelation of the truncation error in space.

The origin of the truncation error in space is the numerical approximation of the convective term. However, it might not always be feasible to change the numerical approximation of this term. Instead, a modification of the source term can be an alternative approach. We have studied different ways to modify the source term to cancel the truncation error in space in Section 4.5.

If the choice of the numerical flux functions in a shock capturing method is made in the local characteristic variables, then the amount of computational work will increase significantly if the order of approximation of the convective term is enhanced. In Section 5.3 it is shown that for the reactive Navier-Stokes equations, the major part of the truncation error can be canceled by a modification of the source term in the physical variables, to a lower cost.

4.1 Higher accuracy improves the accuracy of the wave speed

To calculate a numerical solution to the model problem, which is a conservation equation with a source term,

$$u(x,t)_t + f(u(x,t))_x = \Psi(u(x,t)), \tag{24}$$

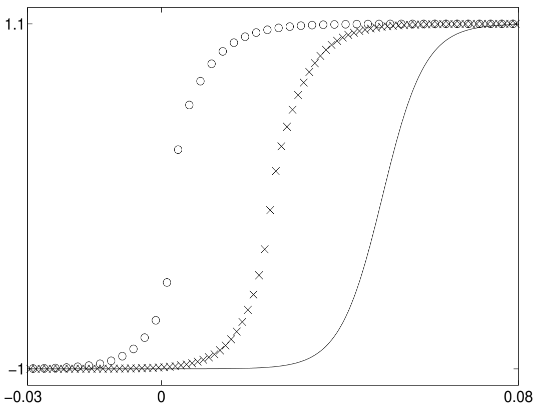
we first construct a semi discrete approximation of the equation,

$$\frac{d}{dt}u_j(t) = L(u(t))_j + \Psi(u_j(t)), \tag{25}$$

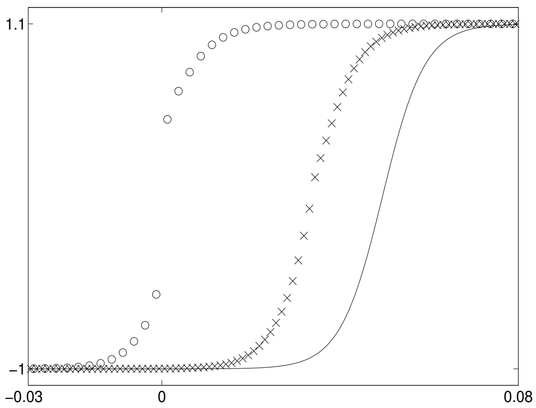
where the approximation of $f(u(x_j,t))_x$ is written in a conservative way, i.e., as $-L(u)_j = (h_{j+1/2} - h_{j-1/2})/\Delta x$. The approximation (25) is then integrated in time with the explicit version of the third order Runge-Kutta method, with the time step $\Delta t = \Delta x/2$, if not stated differently.

To obtain an entropy correct solution we have to modify the numerical flux function $h_{j+1/2}$, if df/du change sign between u_j and u_{j+1}. This correction is done by adding a dissipation term of high order to $h_{j+1/2}$. For the model problem we have to do this entropy correction of $h_{j+1/2}$ if $u_j < 0 < u_{j+1}$. Since the traveling wave solution of the model problem (1) with initial data (3) is smooth, we can add a high order dissipation term to all numerical fluxes to damp small oscillations.

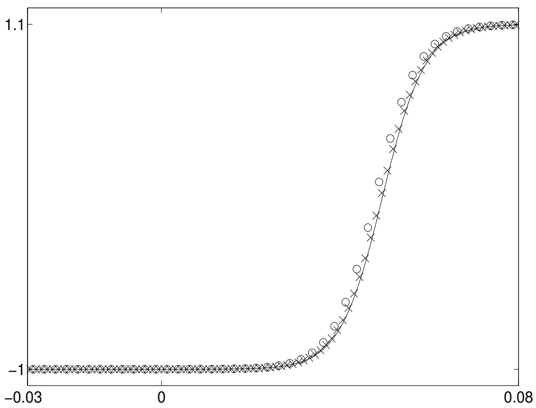
We display numerical solutions to the model equation (1) at $t = 0.5$ in Figure 5. These solutions are obtained by approximating f_x by the ENO approximation [10] of order one



(a) *ENO 1*



(b) *ENO 2*



(c) *Central 4*

Figure 5: *Numerical solutions at $t = 0.5$. Results obtained with $\Delta x = 1/400$ and $\Delta x = 1/800$ are displayed with circles and crosses respectively. The exact solution is displayed with a drawn line.*

and two, and by a central difference approximation of order four respectively. To the approximations of f_x the artificial viscosity term $-\Delta_-^3\Delta_+^3 u_j/(80\Delta x)$ is added. Though the resolution is high, the position of the wave is poor for the low order ENO methods. By using a fourth order method, the accuracy in wave position is increased considerably. Compare with the results presented in Figure 6 and 7.

In Figure 6 we display the obtained error in the numerical wave speed when f_x is approximated by a central difference approximation of order two, four and six respectively. To this approximation the artificial viscosity term $-\Delta_+^3\Delta_-^3 u_j/(80\Delta x)$ is added. The numerical results are in agreement with the analytical results from Section 2.3. For $\Delta x \approx 1/6400$, the order of convergence in wave speed is two and four for the central difference approximations of order two and four respectively. The results obtained with

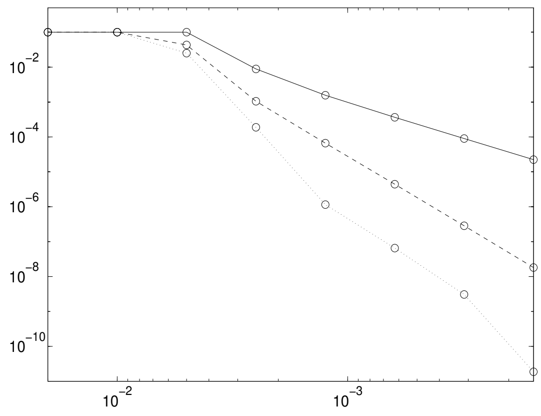


Figure 6: *The absolute error in wave speed as a function of Δx. Numerical results obtained when f_x is approximated by a central difference of order two are displayed with a drawn line. Results obtained with central difference approximations of f_x of order four and six are displayed with a dashed line and a dotted line respectively. Compare with Figure 1.*

the central difference approximation of order six are affected by the truncation error descending from the third order Runge-Kutta method.

The numerical wave speed is measured as the average wave speed between time level t_{3000} and t_{4000}, which is given by (17). For $\Delta x = 1/3200$ and $\Delta x = 1/6400$, the wave stabilizes during the first 1000 and 2000 time steps respectively. Hence 3000 time steps are sufficient to stabilize the waves for all $\Delta x \geq 1/6400$. After the stabilizing time steps, the numerical wave speed varies between subsequent time steps in an oscillatory manner. For example, when $\Delta x = 1/400$ and f_x is approximated with a central difference of order 4, the error in the numerical wave speed oscillates around the mean value $1.0508 \cdot 10^{-3}$. The amplitude of these oscillations is $0.0013 \cdot 10^{-3}$ and the period is $\sim 20\Delta t$. As a consequence the average wave speed between time level t_{2000} and t_{3000} is almost the same as the average wave speed between time level t_{3000} and t_{4000}.

In Figure 7 we display the error in wave speed obtained with the central biased ENO scheme, [8], of order 1, 2, 3, 4 and 5. The artificial viscosity term $-\Delta_+^3\Delta_-^2 u_j/80$ is added to every numerical flux. By studying the numerical results we see that the order of convergence in wave speed is equal to the order of approximation of the numerical method.

4.2 Modification of the source term using artificial grid points

In the last section it was seen that the order of convergence in wave speed is equal to the order of approximation of the numerical method. However, on a sufficiently coarse mesh,

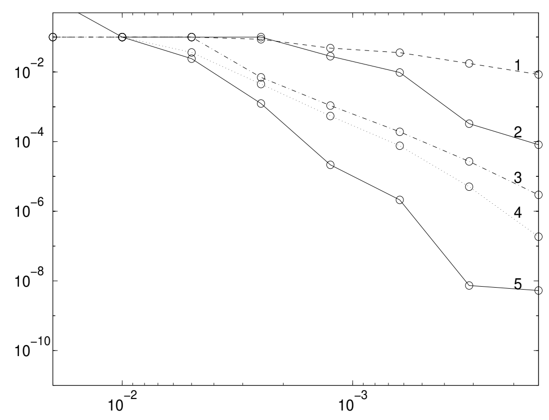
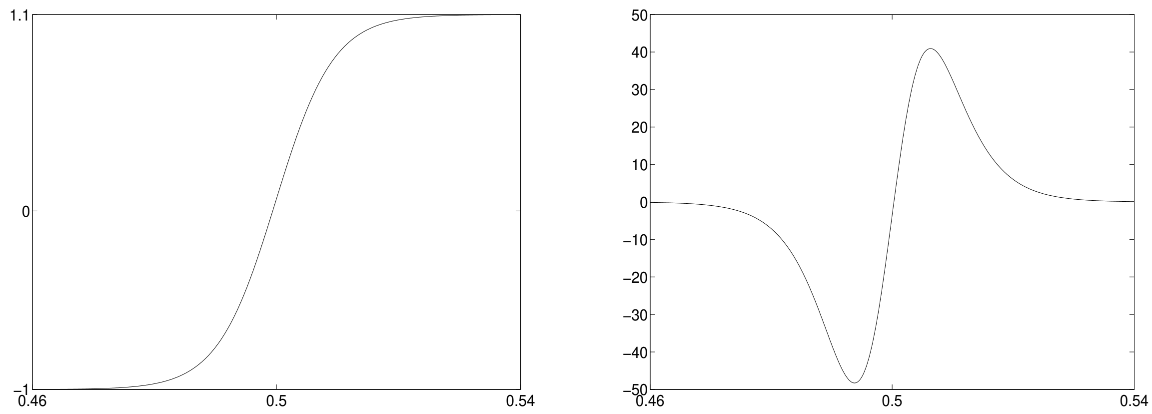


Figure 7: *The absolute error in wave speed as a function of Δx. ENO approximation of f_x of order 1, 2, 3, 4 and 5. Compare with Figure 2.*



(a) $w(\xi)$.

(b) $\Psi(w(\xi))$.

Figure 8: *The wave profile, and the source term as a function of this profile.*

the accuracy in the resulting wave speed is as bad for the higher order methods as for the numerical methods of lower order. See Figure 6 and 7.

When the resolution is low, $\sum_{j=1}^{N} \Psi(u_j(t)) \cdot \Delta x$ will be a poor approximation of $\int_{x_{1/2}}^{x_{N+1/2}} \Psi(u(x,t))dx$. This is easy to see in Figure 8, where the traveling wave, $u(x,t) = w(x-st)$, and the source term, $\Psi(u(x,t))$ are displayed. What makes the integral approximation especially poor, is the fact that the magnitude of Ψ is large in comparison with $\int_{x_{1/2}}^{x_{N+1/2}} \Psi(u(x,t))dx$, because of cancelation effects. The poor approximation of $\int_{x_{1/2}}^{x_{N+1/2}} \Psi(u(x,t))dx$ will automaticly result in a large error in the numerical wave speed. See the discussion below for the details.

By creating artificial grid points in the wave profile by an interpolation procedure, it may be possible to increase the accuracy of the numerical wave speed. Because, in this way, more information can be obtained about the source term throughout the wave, and this information can then be used to increase the accuracy in the approximation of $\int_{x_{1/2}}^{x_{N+1/2}} \Psi(u(x,t))dx$.

To show that an error in the numerical wave speed is a consequence of a poor approximation of the integral of the source term, we first notice that a traveling wave solution of (24), $u(x,t) = w(x - st)$, has the wave speed

$$s = \frac{f(u_L) - f(u_R)}{u_L - u_R} + \frac{\int_{-\infty}^{\infty} \Psi(u(x,t))dx}{u_L - u_R}. \tag{26}$$

Furthermore, in our model problem, where the leftmost grid values, $u_1, u_2, \ldots, u_p$ are arbitrarily close to u_L and the rightmost grid values, $u_N, u_{N-1}, \ldots u_{N-q}$ are arbitrarily close to u_R, the semi discrete approximation (25) gives us

$$\frac{d}{dt}\sum_{j=1}^{N} u_j(t) \cdot \Delta x = f(u_L) - f(u_R) + \sum_{j=1}^{N} \Psi(u_j(t)) \cdot \Delta x,$$

because the numerical flux functions we use, $h(u_{j-k}, \ldots, u_{j+k+1})$, are consistent with the flux function, $f(u)$, in the sense that $h(u, \ldots, u) = f(u)$.

Hence, the instantaneous numerical wave speed, which is consistent with (17), i.e.,

$$\frac{d}{dt}\sum_{j=1}^{N} u_j(t) \cdot \Delta x \Big/ (u_L - u_R),$$

is equal to

$$\frac{f(u_L) - f(u_R)}{u_L - u_R} + \frac{\sum_{j=1}^{N} \Psi(u_j(t)) \cdot \Delta x}{u_L - u_R}. \tag{27}$$

By comparing equations (26) and (27) we conclude that an error in the instantaneous numerical wave speed can only be a consequence of a poor approximation of the integral of the source term. This fact was observed by LeVeque and Yee in the shock case, [7], but of course it also holds for reversed boundary conditions.

At a first glance it looks like that the error in the speed of propagation could be decreased by increasing the order of accuracy in the integral approximation of Ψ, because $\sum_{j=1}^{N} \Psi(u_j(t))$ is only a second order approximation of $\int_{x_{1/2}}^{x_{N+1/2}} \Psi(u(x,t))dx$. However, the order of accuracy in the approximation

$$\int_{x_{1/2}}^{x_{N+1/2}} \Psi(u(x,t))dx \approx \sum_{j=1}^{N} \Psi(u_j(t)) \cdot \Delta x$$

is very high. The reason is that the leftmost grid values are arbitrarily close to u_L, the rightmost grid values are arbitrarily close to u_R and $\Psi(u_L) = \Psi(u_R) = 0$. As a matter of this fact, $\sum_{j=1}^{N} \Psi_j$ can be written as a mean value of highly accurate numerical quadrature formulas, e.g. Simpson's rule,

$$\begin{aligned}
\sum_{j=1}^{N} \Psi(u_j) \cdot \Delta x \;&\approx\; \Big(\frac{1}{6}\Psi(u_0) + \frac{5}{6}\Psi(u_1) + \sum_{j=2}^{N-1} \Psi(u_j) + \frac{5}{6}\Psi(u_N) + \frac{1}{6}\Psi(u_{N+1})\Big) \cdot \Delta x = \\
&=\; \frac{\Delta x}{6}(\Psi(u_0) + 4\Psi(u_1) + 2\Psi(u_2) + 4\Psi(u_3) + \ldots + \Psi(u_{N+1})) + \\
&+\; \frac{\Delta x}{6}(\Psi(u_1) + 4\Psi(u_2) + 2\Psi(u_3) + 4\Psi(u_4) + \ldots + \Psi(u_N)) = \\
&=\; \frac{1}{2}\int_{x_0}^{x_{N+1}} \Psi(u(x,t))dx + O(\Delta x^4) + \frac{1}{2}\int_{x_1}^{x_N} \Psi(u(x,t))dx + O(\Delta x^4) \approx \\
&\approx\; \int_{x_{1/2}}^{x_{N+1/2}} \Psi(u(x,t))dx + O(\Delta x^4).
\end{aligned}$$

In the same way it is easy to show that

$$\sum_{j=1}^{N} \Psi(u_j) \cdot \Delta x \approx \int_{x_{1/2}}^{x_{N+1/2}} \Psi(u(x,t))dx + O(\Delta x^p),$$

for any p, if a sufficient number of the out-most grid values are approximatively equal to u_L and u_R respectively.

Clearly, the order of accuracy in the integral approximation of Ψ is high, but anyhow the error in wave speed comes from a poor approximation of $\int_{x_{1/2}}^{x_{N+1/2}} \Psi(u(x,t))dx$. There are two significant sources of error in this approximation. First, we have already assumed that the wave is poorly resolved. Second, the source term is a function of $u(x,t)$, but the numerical wave profile is only an approximation of $u(x,t)$. In the beginning of this section, we presented the possibility to increase the accuracy of the wave speed by creating artificial grid values, u_k, $x_k \in (x_j, x_{j+1})$, by an interpolation procedure, and use the corresponding function values, $\Psi(u_k)$, to enhance the accuracy in the integral approximation of Ψ. The method of introducing artificial grid values can be used to reduce the first source of error.

To test this idea in practice we change the numerical method (16) as follows. In all Runge-Kutta stages we calculate the artificial grid values $u_{j+m/4}$, $m = 1, 2, 3$, for all j. These artificial grid values, u_k, $k \in (j, j+1)$, are set to $P_j(x_k)$, where $P_j(x)$ is the interpolation polynomial of order five which satisfies $P_j(x_l) = u_l$, $l = j-2, \ldots, j+3$. Finally we replace the source term $\Psi_j = \Psi(u_j)$ with the weighted average

$$\Psi_j = \sum_{m=-4}^{4} \alpha_m \cdot \Psi(u_{j+m/4}), \qquad \sum_{m=-4}^{4} \alpha_m = 1, \tag{28}$$

where the constants α_m are chosen as

$$(\alpha_{-4}, \alpha_{-3}, \alpha_{-2}, \ldots, \alpha_4) = \frac{1}{128}(7, -26, 16, 58, 18, 58, 16, -26, 7).$$

With this choice, the approximation of the source term, (28), satisfies

$$\Psi_j = \Psi(u_j) + O(\Delta x^6),$$

and the order of approximation of the numerical method is not reduced. Furthermore, by summing together the modified source terms, Ψ_j, we see that

$$\sum_{j=1}^{N} \Psi_j = \sum_{j=1}^{N} \sum_{m=0}^{3} \beta_{j+m/4} \Psi(u_{j+m/4}),$$

where

$$\beta_{j+m/4} = 1/4, \quad m = 0, \ldots, 3, \quad \forall j, j \not\approx 1, j \not\approx N,$$

i.e., the different parts of Ψ contribute equally to the integral approximation of the source term.

By using the modification (28) of the source term, the accuracy of the numerical wave speed is marginally increased for $\Delta x \approx 1/200$ when f_x is approximated by central differences of order four. It is also observed that the difference in wave speed with and without modification (28) is of $O(\Delta x^6)$. Hence, the positive effect of the modification is neglectable for $\Delta x \leq 1/400$. Similar results are obtained when f_x is approximated by central differences of order two and six.

With or without modification (28), for sufficiently small Δx, the grid values, u_j, will travel from level u_R to level u_L, and experience the different parts of $\Psi(u), u \in [u_L, u_R]$. Because different grid points visit different parts of Ψ, the accuracy of the wave speed is significantly increased when Δx is reduced so much that the wave starts to move.

A modification similar to (28) can be used to reduce the negative effects of a stiff source term, i.e., to get the wave moving. However, the artificial viscosity technique which will be presented in the next section has shown to be a more effective means for this purpose.

What makes the higher order methods approximating the wave speed better for sufficiently small Δx is the fact that they approximate the shape of $u(x)$ more accurately. Hence the moving grid values experience the different parts of $\Psi(u)$ in a way that is more fair in an average way of view. However, the influence of the source term on the numerical solution will vary wildly from one time step to another, especially when the wave is poorly resolved. To reduce the effect of this oscillatory behavior on the numerical results, we measure the average value of the wave speed during 1000 time steps.

4.3 More viscosity increases the accuracy of the wave speed

When grid points move from level u_R to level u_L they experience different parts of $\Psi(u)$. As a consequence the average wave speed will be a relatively good approximation of the

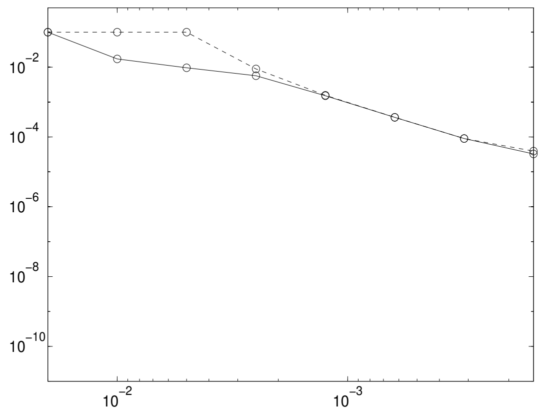


Figure 9: *The absolute error in wave speed as a function of Δx. Central difference approximation of f_x of order two modified with artificial viscosity terms of order six of different magnitude. The results obtained when the constant in the artificial viscosity term is equal to $-1/15$ are displayed with a drawn line. Those obtained when the constant is equal to $-1/80$ are displayed with a dashed line.*

exact wave speed. However, for sufficiently large Δx the source term is so dominant that no grid point will ever move from level u_R to level u_L. However, by adding a large quantity of artificial viscosity to the numerical method we can make the wave move, making the accuracy of the wave speed increase drastically. The price we have to pay is an increase in the number of time steps. When using a large amount of artificial viscosity we have to reduce the CFL number to keep the numerical method stable. We also pay by making the overall approximation of u less accurate.

In Figure 9 we show the numerical wave speeds which are obtained by adding the artificial viscosity terms $-\Delta_+^3\Delta_-^3 u_j/(80\Delta x)$ and $-\Delta_+^3\Delta_-^3 u_j/(15\Delta x)$ respectively to the second order accurate central difference approximation of f_x, $D_0 f_j$. The third order explicit Runge-Kutta scheme is used for the time integration, with the time step set to $\Delta t = \Delta x/10$.

For large Δx the accuracy of the wave speed is increased considerably when more artificial viscosity is added to the numerical scheme. For small Δx however the change in wave speed is neglectable, because the artificial viscosity term is of order Δx^5 and the error in the approximation of f_x is of order Δx^2.

By comparing the results above with the results in Figure 6, where the time step is set to $\Delta t = \Delta x/2$, we conclude that it is not the reduction in Δt that makes the accuracy of the wave speed increase.

We have observed that the artificial viscosity term $4c \cdot \Delta^{2r} u_j$ gives a more accurate wave speed than the modification $-c \cdot \Delta^{2r+2} u_j$ for large Δx. Since one of these two modifications of the flux functions reduces the CFL-number as much as the other, we conclude that one should not use an artificial viscosity term of higher order of convergence than the numerical approximation of f_x.

4.4 Enhancement in the accuracy of the wave speed by a tuning procedure

In [1], Bermudez and Vazquez studied the scalar advection equation

$$\begin{aligned} u_t + u_x &= \Psi(u), & -\infty < x < \infty, \quad t \geq 0, \\ \Psi(u) &= -\eta u(u-1)(u-0.5), \quad \eta > 0, & \\ u(x,0) &= \begin{cases} 1, & x \leq 0.3, \\ 0, & \text{otherwise}, \end{cases} & \end{aligned}$$

which has the exact solution $u(x,t) = u(x-t,0)$. They discretized u_x using the first order upwind approximation. For the time integration they used the forward Euler method. Since the grid points in the shock were heavily affected by the source term, the resulting scheme did not capture the right position of the discontinuity. However, if the source term was upwinded according to

$$\Psi_j^n = (I + \alpha\Delta_-)\Psi(u_j^n), \quad \alpha < 0,$$

or

$$\Psi_j^n = \Psi((I + \alpha\Delta_-)u_j^n), \quad \alpha < 0,$$

the position of the discontinuity was more accurate.

The observation by Bermudez and Vazquez lead us to the hypothesis that if the source term in our model problem is modified according to

$$\Psi_j^n = (I + \alpha\Delta_-^p\Delta_+^q)\Psi(u_j^n), \quad p \geq 0, \quad q \geq 0,$$

or

$$\Psi_j^n = \Psi((I + \alpha\Delta_-^p\Delta_+^q)u_j^n), \quad p \geq 0, \quad q \geq 0,$$

then the speed of the wave will be more accurately computed.

First we study the case when f_x is approximated by $D_0 f_j$. The artificial viscosity term $-\Delta_+^3\Delta_-^2 u_j/80$ is added to all fluxes, $h_{j+1/2}$. Since the order of convergence in wave speed is two for this scheme, we make a second order modification of the source term. This modification will affect the numerical solution with an $O(\Delta t \cdot \Delta x^2)$-term in every time step. If the numerical scheme is convergent, every grid point at the last time level will therefore be modified with a term of order $O(\Delta x^2)$ and the modification of the average wave speed, s^*, will be of order $O(\Delta x^2)$. By tuning the parameter α in the modified source term we can eliminate the error of order $O(\Delta x^2)$ in s^* and in this way improve the accuracy of the wave speed.

Since the traveling wave solution to our model equation with $u_L < u_R$ is smooth, we do not make the modification of the source term in the upwind direction as Bermudez and Vazquez did with the shock wave. Instead we use one of the following centered approximations,

$$\Psi_j^n = (I + \alpha\Delta_+\Delta_-)\Psi(u_j^n), \tag{29}$$

or

$$\Psi_j^n = \Psi((I + \alpha\Delta_+\Delta_-)u_j^n). \tag{30}$$

The first modification of the source term, (29), leaves the contribution of the source term on the average wave speed unaffected in the first stage of the first Runge-Kutta time step. This is true, because $\Delta_+\Delta_-\Psi(u_j) = 0$ at the boundaries, which means that the sum of the added source terms over all grid points is not affected. However, the modification (29) modifies the shape of the numerical approximation of $u(x,t)$ and hence has an effect on the contribution of the source term on the wave speed in the remaining Runge-Kutta stages. The modification (30) of the source term has a similar effect on the shape of the numerical approximation of $u(x,t)$.

To investigate if our hypothesis is true, we approximate f_x by $D_0 f_j - \Delta_+^3\Delta_-^3 u_j/(80\Delta x)$ and modify the source term according to (29) or (30). By tuning the parameter α, the error in the wave speed can be reduced by 100% with both modification (29) and (30). When using one of these modifications, the numerical wave speed will depend on both α and Δx, i.e.,

$$s^* = s^*(\alpha, \Delta x).$$

To find α_{opt}, the optimal α for a given Δx, for which the numerical wave speed is exact,

$$s^*(\alpha_{opt}, \Delta x) = s,$$

we compute s^* for different α. An approximation, α_{opt}^*, of the optimal α is then calculated by linear extrapolation,

$$\alpha_{opt}^*(\alpha, \Delta x) = \alpha \cdot \frac{s - s^*(0, \Delta x)}{s^*(\alpha, \Delta x) - s^*(0, \Delta x)}.$$

We show α_{opt}^* as a function of α and Δx for the modifications (29) and (30) of the source term in Table 1 and 2. The results in Table 1 belong to modification (29), those in Table 2 to modification (30). When α is chosen too large for a given Δx, the resulting method becomes unstable. These cases are marked with a star in the following tables. The accuracy of the wave speed is increased if and only if $\alpha_{opt}^* > \alpha/2$. The closer α_{opt}^* is to α, the better approximation of the wave speed.

We note that α_{opt}^* is approximatively constant for a given Δx if Δx is sufficiently small. Hence the change in the numerical wave speed is approximatively proportional

Δx	$\alpha^*_{opt}(0.1, \Delta x)$	$\alpha^*_{opt}(0.2, \Delta x)$	$\alpha^*_{opt}(0.3, \Delta x)$	$\alpha^*_{opt}(0.4, \Delta x)$
1/100	∞	0.2846	0.2491	0.0341
1/200	0.1361	0.2115	0.2872	0.3625
1/400	0.2024	0.2411	0.2754	0.3062
1/800	0.2603	0.2680	0.2752	0.2820
1/1600	0.2735	0.2751	0.2767	0.2783
1/3200	0.2764	0.2768	0.2771	0.2775
1/6400	0.2770	0.2771	0.2772	0.2773

Table 1: α^*_{opt} *as a function of* α *and* Δx *for the modification (29) of the source term.*

Δx	$\alpha^*_{opt}(0.3, \Delta x)$	$\alpha^*_{opt}(0.5, \Delta x)$	$\alpha^*_{opt}(0.7, \Delta x)$	$\alpha^*_{opt}(0.9, \Delta x)$
1/100	0.0060	0.0051	0.0027	0.0037
1/200	0.3556	0.0343	0.0052	0.0030
1/400	0.5357	0.6490	0.0015	0.0007
1/800	0.7486	0.7777	2.2597	*
1/1600	0.8117	0.8184	0.8249	-0.0669
1/3200	0.8269	0.8285	0.8301	0.8317
1/6400	0.8306	0.8310	0.8314	0.8318

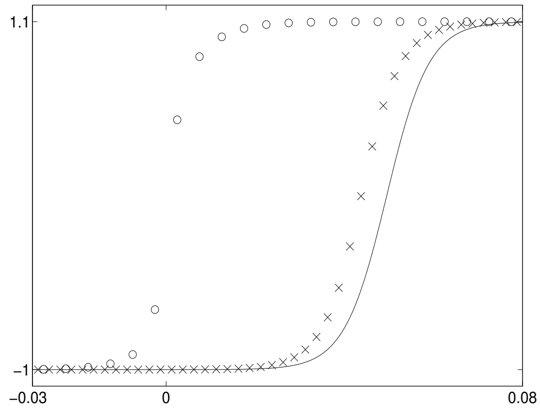
Table 2: α^*_{opt} *as a function of* α *and* Δx *for the modification (30) of the source term.*

to the change in α. This fact gives that α^*_{opt} is a good approximation of α_{opt} if Δx is sufficiently small. It also seems as if α_{opt} converges to a constant value when $\Delta x \to 0$. This constant value is approximatively equal to 0.2772 for the modification (29) of the source term and 0.8317 for the modification (30).

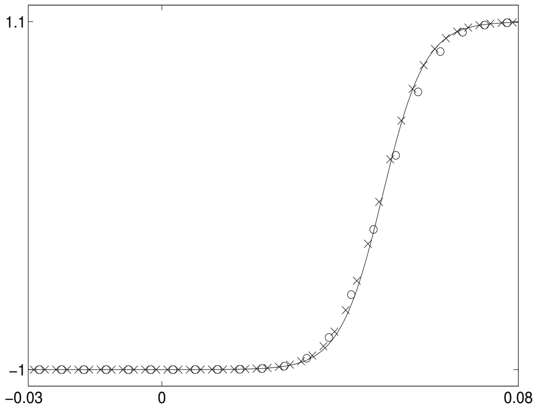
In Figure 10 we show numerical solutions to the model equation (1) at $t = 0.5$. In the computation of these, the convective term was approximated by $D_0 f_j - \Delta_+^3 \Delta_-^3 u_j/(80\Delta x)$ and the source term was modified according to (29). The resolution in the computations was $\Delta x = 1/200$ and $\Delta x = 1/400$ respectively. In Figure 10(a), $\alpha = 0$, i.e., the source term is unmodified. In Figure 10(b), $\alpha = 0.2772$. Notice the impressive effect of the modification. The reason is that $\alpha_{opt} \approx 0.2772$ for every space step shown in Table 1, i.e., $\Delta x = 1/100, 1/200, \ldots$. For modification (30), $\alpha_{opt} \approx 0.8317$ for $\Delta x = 1/1600, 1/3200, \ldots$, but on a coarser grid, $\alpha_{opt} \ll 0.8317$.

When the order of approximation of f_x is four, we can use one of the following centered approximations of the source term to enhance the accuracy of the numerical wave speed,

$$\Psi_j^n = (I - \alpha \Delta_+^2 \Delta_-^2)\Psi(u_j^n), \tag{31}$$



(a) $\alpha = 0$.



(b) $\alpha = 0.2772$.

Figure 10: *Numerical solutions at* $t = 0.5$. *Results obtained with* $\Delta x = 1/200$ *and* $\Delta x = 1/400$ *are displayed with circles and crosses respectively. The exact solution is displayed with a drawn line. In Figure 10(b), the modification (29) is used to enhance the accuracy of the numerical solution.*

or

$$\Psi_j^n = \Psi((I - \alpha\Delta_+^2\Delta_-^2)u_j^n), \tag{32}$$

In a numerical experiment, we approximate f_x by $D_0(I-\frac{1}{6}\Delta_+\Delta_-)f_j-\Delta_+^3\Delta_-^3u_j/(80\Delta x)$ and use the explicit version of the third order Runge-Kutta scheme for the time integration. The time step is chosen such that $\Delta t = \Delta x/40$. If the time step is set to $\Delta t = \Delta x/2$, the improvement in wave speed (which is of order Δx^4) will be neglectable in comparison with the truncation error of order Δt^3 from the time integration method for sufficiently small Δx. For $\Delta t = \Delta x/2$ this is the case when $\Delta x \leq 1/1600$. Since we reduce Δt by a factor 20, we also have to increase the number of time steps by a factor 20, to give the wave enough time to stabilize. Hence the average wave speed is measured between time level t_{60000} and t_{80000} in this experiment. The resulting α_{opt}^* for the modified source term (32) is shown in Table 3. We note that $\alpha_{opt} \to\sim 0.31$ when $\Delta x \to 0$.

When $-f(u)_x$ is approximated by $-D_0f_j$, a truncation error is introduced which is equal to $(\Delta_0\Delta_+\Delta_-f_j)/(6\Delta x)+O(\Delta x^4)$. In modification (29), the source term is modified with the term $\alpha\Delta_+\Delta_-\Psi(u_j)$. When $\Delta x = 1/6400$ the optimal α is approximatively equal to 0.2772 for this modification of the source term.

In Figure 11(a) we show both the approximative truncation error $(\Delta_0\Delta_+\Delta_-f_j)/(6\Delta x)$ and the correction $0.2772\Delta_+\Delta_-\Psi(u_j)$. In this figure we choose u_j as the final state in a computation of the model problem, where f_x is approximated by a central difference of sixth order and $\Delta x = 1/6400$. In the interior part of the wave, $(\Delta_0\Delta_+\Delta_-f_j)/(6\Delta x) \approx 0.2772\Delta_+\Delta_-\Psi(u_j)$. This fact indicates that we have canceled the truncation error in space by modifying the source term.

Δx	$\alpha^*_{opt}(0.1, \Delta x)$	$\alpha^*_{opt}(0.2, \Delta x)$	$\alpha^*_{opt}(0.3, \Delta x)$	$\alpha^*_{opt}(0.4, \Delta x)$
1/100	0.0012	0.0004	0.0001	0.0001
1/200	0.1169	*	0.0001	0.0000
1/400	0.2140	0.1267	0.0000	0.0000
1/800	0.2828	0.2862	*	0.0000
1/1600	0.3015	0.3018	0.3020	-0.0002
1/3200	0.3088	0.3089	0.3089	0.3089
1/6400	0.3123	0.3123	0.3123	0.3123

Table 3: α^*_{opt} *as a function of* α *and* Δx *for the modification (32) of the source term.*

Further evidence of this fact are given in Figure 11(b) and 11(c). The modification (30) of the source term, $\Psi((I + \alpha\Delta_+\Delta_-)u_j^n) - \Psi(u_j^n)$, is compared with the truncation error of the second order approximation $-f(u)_x \approx -D_0 f_j$ in Figure 11(b). In Figure 11(c) the modification (32) is compared with the truncation error of the fourth order approximation $-f(u)_x \approx -D_0(I - \frac{1}{6}\Delta_+\Delta_-)f_j$, which is equal to $-(\Delta_0\Delta_+^2\Delta_-^2 f_j)/(30\Delta x) + O(\Delta x^6)$.

We believe that it is easier to cancel the truncation error in space directly by Taylor expansion instead of tuning α in the modified source terms (29), (30), (31) and (32). This idea will be investigated further in the next section.

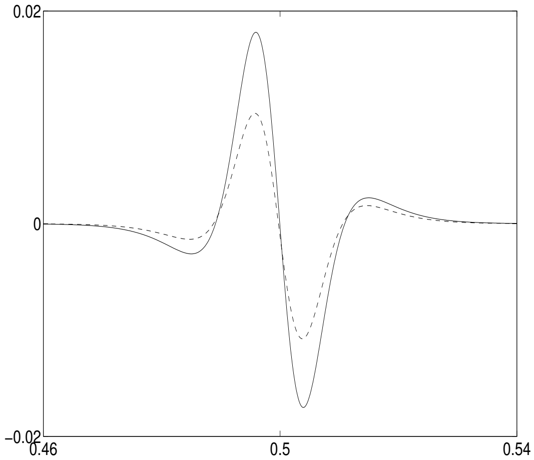
4.5 Cancelation of the truncation error in space

In the numerical experiments in the previous section we saw that if the source term is implemented as $\Psi((I + \alpha\Delta_+\Delta_-)u_j)$ or $(I + \alpha\Delta_+\Delta_-)\Psi_j$, a change in α will change the wave speed with an amount which is approximatively proportional to α. There were also evidences for that the optimal way of tuning α is to let the correction of the source term correspond to the truncation error of the numerical approximation of f_x. In this section we will investigate this assumption further.

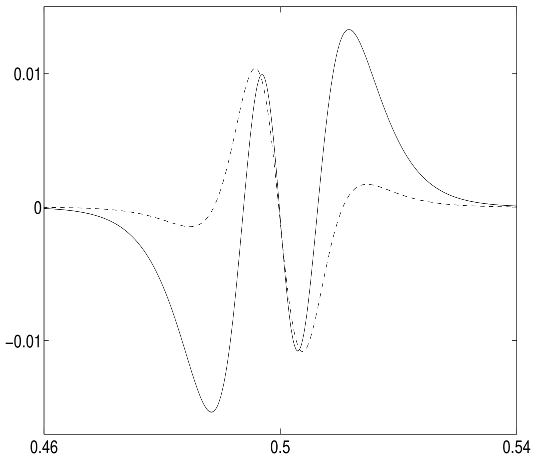
To investigate if we can enhance the accuracy of the wave speed by modifying the source term to cancel the truncation error, we return to the model problem (1). When using the second order central difference $D_0 f_j$ to approximate $f(u_j(t))_x$, the truncation error in space is, by Taylor expansion,

$$f(u_j(t))_x - D_0 f_j = -\frac{\Delta x^2}{6} f(u_j(t))_{xxx} + O(\Delta x^4).$$

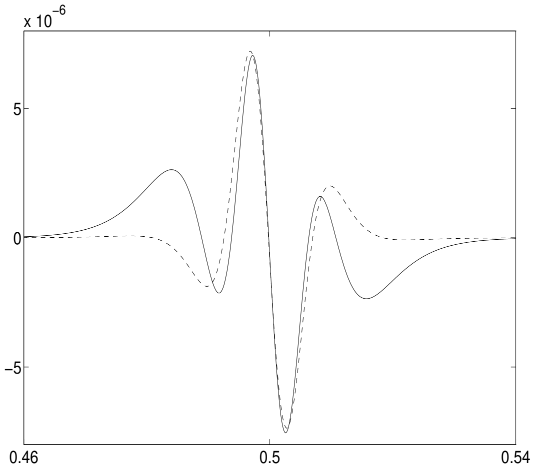
By using the third order accurate Runge-Kutta method for the time integration, the truncation error in time is neglectable. We verified this statement in a numerical test. If f_x is approximated by $D_0 f_j$, the wave speed is approximatively constant when Δt is reduced for a fixed Δx.



(a) *The approximative truncation error* $(\Delta_0\Delta_+\Delta_- f_j)/(6\Delta x)$ *is displayed with a dashed line. Displayed with a drawn line, the correction* $0.2772\Delta_+\Delta_-\Psi(u_j)$.



(b) *The approximative truncation error* $(\Delta_0\Delta_+\Delta_- f_j)/(6\Delta x)$ *is displayed with a dashed line. Displayed with a drawn line, the correction* $\Psi((I + 0.8317\Delta_+\Delta_-)u_j^n) - \Psi(u_j^n)$.



(c) *The approximative truncation error* $-(\Delta_0\Delta_+^2\Delta_-^2 f_j)/(30\Delta x)$ *is displayed with a dashed line. Displayed with a drawn line, the correction* $\Psi((I - 0.3123\Delta_+^2\Delta_-^2)u_j^n) - \Psi(u_j^n)$.

Figure 11: *Approximative truncation errors and optimal corrections of the source term.*

To cancel the second order part of the truncation error in space, we replace Ψ_j with an approximation of $\Psi(u_j) + \frac{\Delta x^2}{6} f(u_j)_{xxx}$. The most straightforward way of modifying the source term is to set it equal to $\Psi(u_j) + \frac{\Delta x^2}{6} D_0 D_+ D_- f_j$. However, for the explicit version of the Runge-Kutta scheme, this change of the source term is equivalent with replacing the second order approximation $D_0 f_j$ with a central difference approximation of fourth order, $f_x \approx (I - \frac{1}{6}\Delta_+\Delta_-)D_0 f_j$. We have already seen that this change enhance the accuracy of the wave speed, see Figure 6.

We will now modify the source term in a non linear way, similar to modification (30). To be able to cancel the major part of the truncation error in space when f_x is approximated by $D_0 f_j$, we want to find an ϵ such that the modified source term $\Psi(u + \epsilon)$ is approximatively equal to $\Psi(u) + \frac{\Delta x^2}{6} f_{xxx}$. Making a Taylor expansion of $\Psi(u + \epsilon)$,

$$\Psi(u + \epsilon) = \Psi(u) + \epsilon \Psi'(u) + O(\epsilon^2),$$

we see that an approximation of

$$\epsilon = \frac{\Delta x^2}{6\Psi'} f_{xxx}$$

is a good choice for the modified source term $\Psi(u + \epsilon)$. In the numerical method we approximate f_{xxx} with $D_0 D_+ D_- f_j$, i.e., we use the modified source term

$$\Psi(u_j + \frac{\Delta x^2}{6\Psi'(u_j)} D_0 D_+ D_- f_j) = \Psi(u_j) + \frac{\Delta x^2}{6} f_{xxx}(u_j) + O(\Delta x^4). \tag{33}$$

For stability reasons the modified source term has to be limited whenever $\Psi'(u_j) \approx 0$. Therefore we restrict the input to the modified source term (33), $u_j + \epsilon_j$, to the interval $[u_{j-1}, u_{j+1}]$. When Δx is sufficiently small, this restriction of the source term will take place only in those parts of the wave where $\Psi'(u_j) \approx 0$.

By using this restriction of ϵ_j, we allow the source term to be modified with an $O(\Delta x)$-term whenever $\Psi'(u_j) \approx 0$. On the other hand, the limit which is the natural extension of modification (30), i.e., $|\epsilon_j| \leq |\Delta_+\Delta_- u_j|$, is found to be too restrictive, not allowing the modified source term to cancel the whole part of the second order truncation error in the approximation of $f(u)_x$.

The error in the average wave speed obtained using the modified source term (33) and the approximation $f(u_j)_x \approx D_0 f_j - \Delta_+^3 \Delta_-^3 u_j/(80\Delta x)$ is displayed in Figure 12. For comparison we include results obtained without modifying the source term. For small Δx, the numerical scheme where f_x is approximated by a central difference of second order and the source term is modified according to (33), gives approximatively the same accuracy of the wave speed as the scheme where f_x is approximated with a central difference of order four.

When using the implicit Runge-Kutta scheme (16), the modification of the source term, (33), will force us to replace the Newton iteration on u_j, $\forall j$, with a Newton iteration

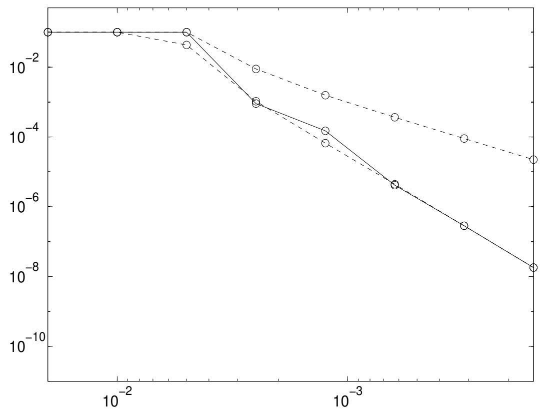


Figure 12: *The absolute error in wave speed as a function of* Δx. *Numerical results obtained when* f_x *is approximated by a central difference of order two and the source term is modified to cancel the truncation error, are displayed with a drawn line. Results obtained with central difference approximations of* f_x *of order two and four respectively, when the source term is not modified, are displayed with dashed lines.*

on a system of $u_1, u_2, \ldots, u_N$, because the modification (33) makes u_j depend on u_{j-1} and u_{j+1}. However, with either of the following modifications of the source term,

$$\Psi(u_j + \frac{\Delta x^2}{6\Psi'} D_-^3 f_j) = \Psi(u_j) + \frac{\Delta x^2}{6} f_{xxx} + O(\Delta x^3), \tag{34}$$

$$\Psi(u_j + \frac{\Delta x^2}{6\Psi'} D_-^3 (I + \frac{3}{2}\Delta_-) f_j) = \Psi(u_j) + \frac{\Delta x^2}{6} f_{xxx} + O(\Delta x^4), \tag{35}$$

a scalar Newton iteration can be used to determine u_1, u_2, u_3 and so on. The modification (34) gives a less accurate wave speed than the modification (33) or (35) does, because with modification (34) the truncation error in space is of order Δx^3 instead of order Δx^4.

The truncation error in a five point central difference approximation of $f(u_j)_x$ of fourth order is $\Delta x^4 f^{(5)}/30$. Hence, with this approximation of $f(u_j)_x$, we can increase the accuracy of the wave speed by using the modified source term

$$\Psi(u_j - \frac{\Delta x^4}{30\Psi'(u_j)} D_0 D_+^2 D_-^2 f_j). \tag{36}$$

In a numerical experiment where f_x was approximated by the following central difference approximation of fourth order,

$$(I - \frac{1}{6}\Delta_+\Delta_-) D_0 f_j - \Delta_+^3 \Delta_-^3 u_j / (80\Delta x)$$

and the source term was modified according to (36), we obtained a wave speed which was as accurate as the wave speed obtained when f_x is approximated by a central difference of order six.

A non linear modification of the source term can also be used when f_x is approximated in an ENO way. An ENO method, such as the standard ENO method by Shu and Osher, [10], avoid centered differentiation across discontinuities, by approximating f_x with a difference of two numerical flux functions. These flux functions, $h_{j+1/2}$, switch the stencil of computation from grid point to grid point to prevent that they are calculated using information from two sides of a discontinuity. In the numerical method the flux functions $h_{j+1/2}$ are chosen from a set of functions, $h^{(1)}_{j+1/2}, h^{(2)}_{j+1/2}, \ldots, h^{(r+1)}_{j+1/2}$, where r is the order of the ENO method and $h^{(k)}_{j+1/2}$ is a linear combination of $f_{j+k-r}, \ldots, f_{j+k-1}$. For example, the standard ENO method of order two, [10], selects the numerical flux function $h_{j+1/2}$ from the following set of numerical fluxes,

$$\begin{aligned} h^{(1)}_{j+1/2} &= -1/2 f_{j-1} + 3/2 f_j, \\ h^{(2)}_{j+1/2} &= 1/2 f_j + 1/2 f_{j+1}, \\ h^{(3)}_{j+1/2} &= 3/2 f_{j+1} - 1/2 f_{j+2}. \end{aligned}$$

If we compare the approximation of $-f(u_j)_x$ in the second order ENO method above,

$$D_+ h_{j-1/2} = \left(h^{(k_{j+1/2})}_{j+1/2} - h^{(k_{j-1/2})}_{j-1/2} \right) / \Delta x$$

with a central difference approximation of $-f(u_j)_x$ of order four, $D_+ H_{j-1/2}$, where

$$H_{j-1/2} = (-f_{j-2} + 8 f_{j-1} + 8 f_j - f_{j+1})/12,$$

we see that

$$\Delta_+ h^{(k_{j-1/2})}_{j-1/2} - \Delta_+ H_{j-1/2} = \sum_{l=-1}^{1} c_l(k_{j-1/2}, k_{j+1/2}) \cdot \Delta_+ \Delta_- f_{j+l}, \tag{37}$$

where, e.g.,

$$c_1(1,2) = 5/12, \quad c_2(1,2) = 0, \quad c_3(1,2) = 1/12.$$

As an alternative we can expand $\Delta_+ h_{j-1/2} - \Delta_+ H_{j-1/2}$ as a linear combination of the function values $f_{j-2}, \ldots, f_{j+2}$,

$$\sum_{l=-2}^{2} d_l(k_{j-1/2}, k_{j+1/2}) \cdot f_{j+l}. \tag{38}$$

By modifying the source term into

$$\Psi(u_j + \frac{\Delta_+ h_{j-1/2} - \Delta_+ H_{j-1/2}}{\Psi'(u_j) \Delta x}), \tag{39}$$

where the input to the modified source term is restricted to the interval $[u_{j-1}, u_{j+1}]$, we obtain a method with high accuracy in the wave speed. To avoid overshoots at discontinuities the source term should not be modified at shock waves. In the next section we will construct a detector which tells us whether u_j belongs to a shock wave or not. We remark that no if-tests are needed to evaluate the modified source term (39), if the coefficients c_l in (37) (or d_l in (38)) are saved in a table.

In Figure 13 we show the wave speeds obtained with the central biased ENO scheme [8], of order 1, 2, 3, 4 and 5, when the source term is modified according to (39). In the computations $D_+H_{j-1/2}$ is set to central difference approximations of f_x of order 2, 4, 4, 6 and 6 respectively, e.g., $D_+H_{j-1/2} = D_0 f_j$. The entropy correction which is used, i.e.,

$$\Delta_+ h_{j-1/2} = \Delta_+ h^{ENO}_{j-1/2} - \frac{\Delta_+^3 \Delta_-^3 u_j}{80},$$

is not taken into account in the modification of the source term (39). For the time integration we use the explicit third order Runge-Kutta method, with the time step chosen such that $\Delta t = \Delta x/2$.

We remark that it is cheaper to calculate the numerical fluxes with a low order ENO method and modify the source term according to (39), than to use a high order ENO approximation directly.

We also remark that there are other ways to modify the source term to cancel the major part of the truncation error than (39). One way is to consider one more term in the Taylor expansion of $\Psi(u_j + \epsilon)$. For example, when $f(u_j)_x$ is approximated by $D_0 f_j$, we can choose ϵ such that

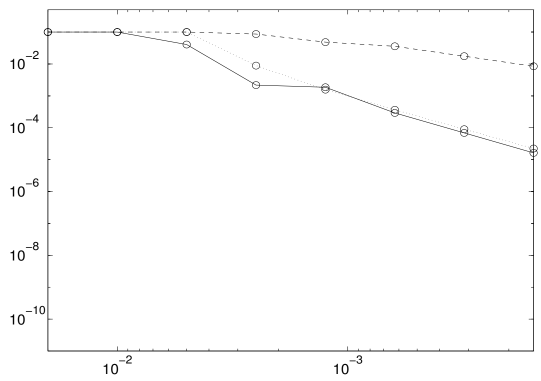
$$\epsilon \Psi'(u_j) + \frac{\epsilon^2}{2}\Psi''(u_j) = \frac{\Delta x^2}{6} D_0 D_+ D_- f_j. \tag{40}$$

With this choice there is no need to restrict ϵ, because $\Psi'(u_j) = \Psi''(u_j) = 0$ is impossible for the model problem (1).

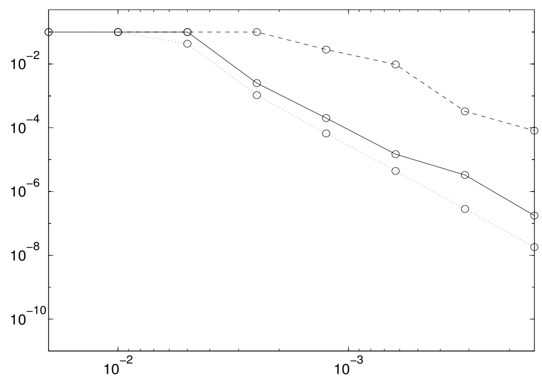
The approximation $D_0 f_j - \Delta_+^3 \Delta_-^3 u_j/(80\Delta x)$ of f_x was combined with the modified source term $\Psi(u_j + \epsilon)$, with ϵ defined by (40), in a numerical experiment. There are two roots of equation (40). We chose the real part of the one that is consistent with the previous modification of the source term, i.e., with

$$\epsilon = \frac{\Delta x^2}{6\Psi'(u_j)} D_0 D_+ D_- f_j.$$

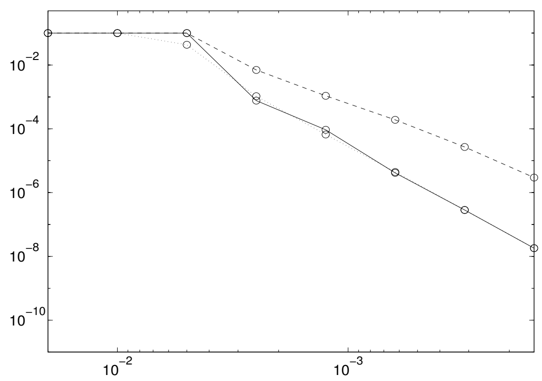
When plotting the error in the resulting numerical wave speed in a log-log-plot similar to Figure 12, no difference could be seen with the eye between these results and the results obtained with a fourth order central difference approximation of f_x. By considering more terms in the Taylor expansion of $\Psi(u_j + \epsilon)$, the correction of the source term becomes closer to the approximation of the truncation error.



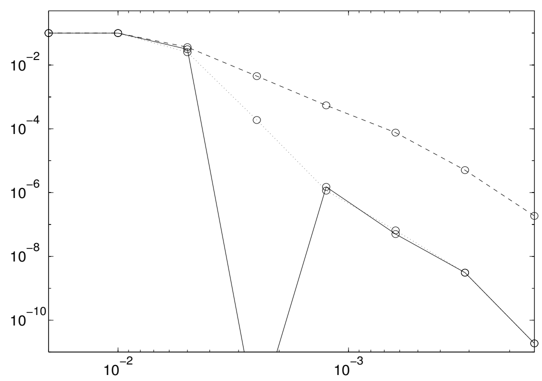
(a) *ENO 1, Central 2.*



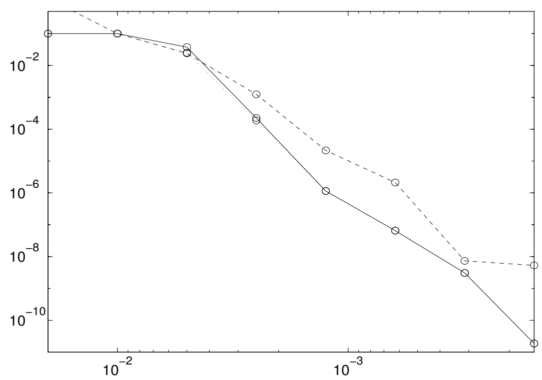
(b) *ENO 2, Central 4.*



(c) *ENO 3, Central 4.*

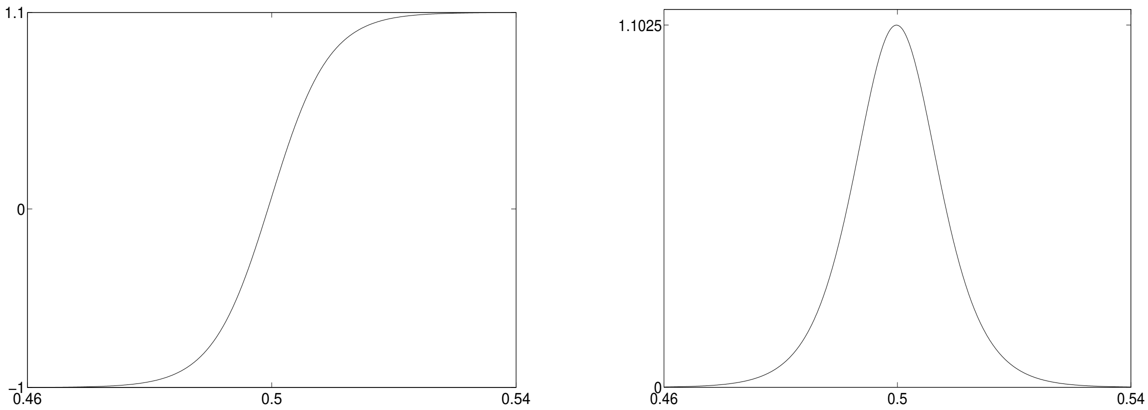


(d) *ENO 4, Central 6.*



(e) *ENO 5, Central 6.*

Figure 13: *The absolute error in wave speed as a function of* Δx*. Numerical results obtained with ENO approximations of* f_x *are displayed with dashed lines. Results obtained with ENO approximations of* f_x*, when the source term is modified to cancel the truncation error, are displayed with drawn lines. Results obtained with high order central difference approximations of* f_x *are displayed with dotted lines.*



(a) $w(\xi)$. (b) $w'(\xi)/K$.

Figure 14: *The appearance of the traveling wave.*

4.6 Switched techniques, a detector which finds the traveling waves

For the model problem (1) with initial data (3), we want to switch off the modification of the source term when $w \approx -1$ or $w \approx 1.1$, because for these w's the modification of the source term is of no significance.

In Section 2.1 we saw that a smooth traveling wave solution to this model problem, $u(x,t) = w(x - st)$, must satisfy the ODE

$$w'(\xi) = -K(w(\xi) + 1)(w(\xi) - 1.1).$$

The derivative of this solution satisfies

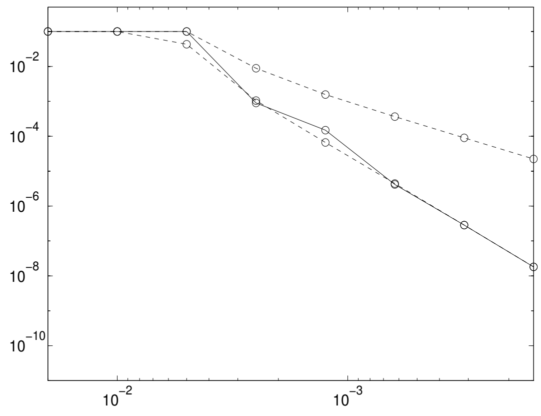
$$|w'(\xi)| \in [0, 1.1025 \cdot K],$$

see Figure 14. Since $|w'| \approx 0$ if and only if $w \approx -1$ or $w \approx 1.1$, this means that we can switch off the modification of the source term if

$$|D_0 u_j| \leq c \cdot K, \quad \text{where } c \in [0, 1.1025]. \tag{41}$$

By choosing $c \approx 1.1025$ we will obtain approximatively the same wave speeds as when the source term is not modified. By choosing $c \approx 0$ we will obtain approximatively the same results as when the source term is modified everywhere.

The modifications of the source term discussed earlier in this section should not be applied on shock waves. Therefore we also check that $a_{j-1/2} < a_{j+1/2}$ is fulfilled before such a modification is done. The local shock speed, $a_{j+1/2}$, is defined in (22). For model problem (1), the condition $a_{j-1/2} < a_{j+1/2}$ is equivalent to $u_{j-1} < u_{j+1}$.



(a) $c = 0.1$. (b) $c = 0.5$.

Figure 15: *The absolute error in wave speed as a function of* Δx. *Numerical results obtained when* f_x *is approximated by a central difference approximation of order two and the source term is modified to cancel the truncation error whenever the detector says so are displayed with a drawn line. Results obtained with central difference approximations of* f_x *of order two and four with unmodified source terms are displayed with dashed lines.*

To validate the detector (41) we repeat the numerical experiment in Figure 12 again, i.e., we approximate f_x with $D_0 f_j - \Delta_+^3 \Delta_-^3 u_j/(80\Delta x)$ and use the explicit version of the Runge-Kutta scheme (16) for the time integration. The time step is set to $\Delta t = \Delta x/2$ and the source term is modified according to (33) if $|D_0 u_j| > c \cdot K$. The numerical results of this experiment are displayed in Figure 15. By comparing with the results shown in Figure 12 we notice that the effect of the detector on the wave speed is marginal.

5 The Navier-Stokes system, some computations

5.1 The Navier-Stokes equations

Consider the compressible Navier-Stokes equations for a reacting mixture. Under the assumption that there are only two species present, the reactant and the product, and the reactant is converted to the product by a one-step irreversible chemical reaction governed by Arrhenius kinetics, that the specific heats at constant pressure, c_p, are equal and constant for the two mixtures, and that the molecular masses of the two mixtures are the same, these equations can be written

$$\begin{aligned} \rho_t + (\rho u)_x &= 0, \\ (\rho u)_t + (\rho u^2 + p)_x &= \mu u_{xx}, \\ e_t + ((e+p)u)_x &= \mu(u u_x)_x + \lambda T_{xx}, \\ (\rho z)_t + (\rho u z)_x &= -\rho z K(T) + D z_{xx}, \end{aligned} \tag{42}$$

after non-dimensionalization.

Here ρ, u, e and z are the density, velocity, energy and the fraction of unburnt gas respectively. The coefficients of viscosity, heat condition and species diffusion above, μ, λ and D, has been assumed to be constant.

The pressure and the temperature are given by the formulae

$$p = (\gamma - 1)(e - \frac{\rho u^2}{2} - q_0 \rho z)$$

and

$$T = p/\rho$$

respectively. The factor $K(T)$, which depends on the temperature, is defined by

$$K(T) = K_0 T^{\alpha} e^{-E/T},$$

where K_0 is a positive constant. In the rest of this article we will assume that $\alpha = 0$.

The dimension-less parameters appearing above, q_0, E, γ and K_0, correspond to chemical heat release, ignition temperature, c_p to c_v ratio and reaction rate respectively. We will set $q_0 = 50$, $E = 50$ and $\gamma = 1.2$ in the numerical experiments.

The compressible Euler equations for the reacting mixture are the special case when the coefficients of diffusion, μ, λ, and D are equal to zero.

5.2 Numerical computation of a deflagration wave

To find a traveling wave solution to (42), we introduce the variable $\xi = x + st$, where s is the propagation speed of the wave. Via a transformation of variables and an integration in ξ it is seen that a traveling wave solution of (42) must fulfill

$$\begin{aligned}
(s+u)\rho &= c_1, \\
(s+u)\rho u + p - \mu u_\xi &= c_2, \\
(s+u)e + up - \mu u u_\xi - \lambda T_\xi &= c_3, \\
c_1 z_\xi + \rho z K(T) - D z_{\xi\xi} &= 0.
\end{aligned} \tag{43}$$

The constants c_1, c_2 and c_3 are determined by the state at the left or right hand side of the wave. Since the first and second derivative of all variables is zero at the limit values $\xi = \pm\infty$ it follows that

$$c_1 = (s + u_L)\rho_L = (s + u_R)\rho_R, \tag{44}$$

$$c_2 = (s + u_L)\rho_L u_L + p_L = (s + u_R)\rho_R u_R + p_R, \tag{45}$$

and

$$c_3 = (s + u_L)e_L + u_L p_L = (s + u_R)e_R + u_R p_R. \tag{46}$$

Here we have denoted the variables at $\xi = -\infty$ with subindex L and the variables at $\xi = \infty$ with subindex R.

By eliminating u, p and e from (43) and introducing $\tau \equiv 1/\rho$ we obtain the following first order system

$$\begin{aligned}
\frac{d}{d\xi}\tau(\xi) &= \frac{1}{\mu c_1}\left(\frac{T(\xi)}{\tau(\xi)} + c_1^2\tau(\xi) - c_2 - sc_1\right), \\
\frac{d}{d\xi}T(\xi) &= \frac{c_1}{\lambda}\left(\frac{T(\xi)}{\gamma - 1} + q_0 z(\xi) - \frac{1}{2}c_1^2\tau^2(\xi) + (c_2 + sc_1)\tau(\xi) - \frac{c_3}{c_1} - \frac{sc_2}{c_1} - \frac{1}{2}s^2\right), \\
\frac{d}{d\xi}z(\xi) &= z'(\xi) \\
\frac{d}{d\xi}z'(\xi) &= \frac{1}{D}\left(c_1 z'(\xi) + \frac{z(\xi)}{\tau(\xi)}K(T(\xi))\right).
\end{aligned} \tag{47}$$

The endpoints of the wave are critical points, but the left endpoint only in an approximative sense. When using Arrhenius kinetics, $K(T_L)$ is only approximatively equal to zero. We assume that $z_L = 1$ and $z_R = 0$.

5.2.1 Analysis at the endpoints of the deflagration wave

By making a linearization around the critical point $(\tilde{\tau}, \tilde{T}, \tilde{z}, \tilde{z}')$, which is either of the two endpoints, we get

$$\begin{pmatrix} \tau \\ T \\ z \\ z' \end{pmatrix}' = \begin{pmatrix} (c_1^2 - \tilde{p}/\tilde{\tau})/(\mu c_1) & 1/(\mu c_1 \tilde{\tau}) & 0 & 0 \\ c_1\tilde{p}/\lambda & c_1/((\gamma-1)\lambda) & c_1 q_0/\lambda & 0 \\ 0 & 0 & 0 & 1 \\ -\tilde{z}K(\tilde{T})/(D\tilde{\tau}^2) & \tilde{z}EK(\tilde{T})/(D\tilde{\tau}\tilde{T}^2) & K(\tilde{T})/(D\tilde{\tau}) & c_1/D \end{pmatrix} \cdot \begin{pmatrix} \tau - \tilde{\tau} \\ T - \tilde{T} \\ z - \tilde{z} \\ z' - \tilde{z}' \end{pmatrix}. \tag{48}$$

In the derivation of this system of equations, we used the relation $\tilde{p} = c_2 + sc_1 - \tilde{\tau}c_1^2$, which is easily obtained from (44) and (45). It is assumed that the viscous constants are positive.

For the left endpoint to be a critical point, $K(T_L)$ must be neglectable. Furthermore, $z_R = 0$. Hence the eigenvalues, l, of the 4×4-matrix in (48) are given by

$$\begin{vmatrix} (c_1^2 - \tilde{p}/\tilde{\tau})/(\mu c_1) - l & 1/(\mu c_1\tilde{\tau}) \\ c_1\tilde{p}/\lambda & c_1/((\gamma-1)\lambda) - l \end{vmatrix} = 0 \tag{49}$$

and

$$\begin{vmatrix} -l & 1 \\ K(\tilde{T})/(D\tilde{\tau}) & c_1/D - l \end{vmatrix} = 0 \tag{50}$$

respectively.

The eigenvalues corresponding to (49) are given by

$$l_{1,2} = \frac{1}{2}\left(\frac{c_1}{\mu} - \frac{\tilde{p}}{c_1\mu\tilde{\tau}} + \frac{c_1}{\lambda(\gamma-1)}\right) \pm \frac{1}{2}\sqrt{\left(\frac{c_1}{\mu} - \frac{\tilde{p}}{c_1\mu\tilde{\tau}} - \frac{c_1}{\lambda(\gamma-1)}\right)^2 + \frac{4}{\mu\lambda}\frac{\tilde{p}}{\tilde{\tau}}}.$$

l_1 and l_2 are of different sign if $c_1^2 < \gamma\tilde{p}/\tilde{\tau}$. Since the Mach number

$$M_\infty \equiv \frac{u_R + s}{a_\infty} < 1$$

for a weak deflagration, and the speed of sound in an ideal gas is equal to

$$a_\infty = \sqrt{\frac{\gamma p_R}{\rho_R}},$$

we see that

$$[\rho_R(u_R + s)]^2 < \gamma p_R \rho_R,$$

i.e.,

$$c_1^2 < \frac{\gamma p_R}{\tau_R}.$$

It follows that $l_1 < 0 < l_2$ at both endpoints, because $p_R/\tau_R \leq p_L/\tau_L$ for a deflagration wave, see e.g. [12].

The eigenvectors corresponding to l_1 and l_2 are proportional to

$$\mathbf{e}_{1,2} = (\, c_1/(\gamma-1) - \lambda l_{1,2}, -c_1\tilde{p}, 0, 0\,)^T .$$

The non-zero elements of $\mathbf{e}_1$ are of different sign if $c_1 > 0$, since

$$\begin{aligned}\frac{c_1}{\gamma-1} - \lambda l_1 &= -\frac{1}{2}\left(\frac{\lambda}{\mu}c_1 - \frac{\lambda}{\mu}\frac{\tilde{p}}{c_1\tilde{\tau}} - \frac{c_1}{\gamma-1}\right) + \frac{1}{2}\sqrt{\left(\frac{\lambda}{\mu}c_1 - \frac{\lambda}{\mu}\frac{\tilde{p}}{c_1\tilde{\tau}} - \frac{c_1}{\gamma-1}\right)^2 + 4\frac{\lambda}{\mu}\frac{\tilde{p}}{\tilde{\tau}}} \\ &> 0.\end{aligned}$$

In the same way, the non-zero elements of $\mathbf{e}_2$ are of the same sign if $c_1 > 0$. It is obvious that $c_1 > 0$, because the reactant is converted into the product and not vice versa. Mathematically, this follows from the last equation of (43), i.e.,

$$c_1 z'(\xi) + \rho z K(T) - D z''(\xi) = 0.$$

Since $\rho z K(T) > 0$, except at the endpoints,

$$c_1 z'(\xi) < D z''(\xi).$$

Now $D \geq 0$. Furthermore, both $z'(\xi) < 0$ and $z''(\xi) < 0$ somewhere along the wave, because the traveling wave is smooth and $z''(\xi) = 0$ at the left endpoint. Hence $c_1 > 0$.

The eigenvalues corresponding to (50) are given by

$$l_{3,4} = \frac{c_1}{2D} \pm \frac{1}{2}\sqrt{\left(\frac{c_1}{D}\right)^2 + 4\frac{K(\tilde{T})}{D\tilde{\tau}}}.$$

If we assume that $K(T_L) = 0$, then $l_3 = 0$ and $l_4 = c_1/D$ at the left endpoint. At the right endpoint, $l_3 < 0 < l_4$.

The eigenvectors corresponding to these eigenvalues are proportional to

$$\mathbf{e}_{3,4} = (\, x_1, x_2, -1, -l_{3,4}\,)^T\,,$$

where

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{c_1 q_0}{\lambda N} \cdot \begin{pmatrix} -1/(\mu c_1 \tilde{\tau}) \\ c_1/\mu - \tilde{p}/(\mu c_1 \tilde{\tau}) - l_{3,4} \end{pmatrix}$$

and N is the determinant in equation (49).

For eigenvector $\mathbf{e}_3$, x_1 and x_2 have the same sign at the right endpoint only if c_1 is sufficiently small for a fixed K_0. For a sufficiently small c_1, $-\tilde{p}/(\mu c_1 \tilde{\tau})$ is the dominant term in the expression $c_1/\mu - \tilde{p}/(\mu c_1 \tilde{\tau}) - l_3$. However, for a fixed c_1, the positive term $-l_3$ becomes the dominant term in this expression when $K_0 \to \infty$. The larger K_0 is, the smaller must c_1 be, for x_1 and x_2 to have the same sign.

In computations of different deflagration waves with the left state $p_L = 1$, $\rho_L = 1$, $u_L = 0$, $z_L = 1$, we have seen that $s = O(\sqrt{K_0})$, i.e., that c_1 is increased when K_0 is increased. Hence it exists a constant $K_{0\max}$ for which x_1 and x_2 have different signs whenever $K_0 > K_{0\max}$ in this case.

The solution of the linearized problem is given by

$$\begin{pmatrix} \tau(\xi) \\ T(\xi) \\ z(\xi) \\ z'(\xi) \end{pmatrix} = \begin{pmatrix} \tilde{\tau} \\ \tilde{T} \\ \tilde{z} \\ \tilde{z}' \end{pmatrix} + \sum_{i=1}^{4} \mathbf{c_i} e^{l_i \xi},$$

where the vectors $\mathbf{c_i}$ are proportional to $\mathbf{e_i}$. At the right endpoint, the eigenvalues l_2 and l_4 are positive. Hence only eigenvector one and three can be used when linearizing around this point.

We know that T and τ increase through a deflagration wave, and that z and p decrease. However, we have just shown that the first two components of eigenvector one always have different signs. We also know that this is the case for eigenvector three at the right endpoint if $K_0 > K_{0\max}$. Hence it can not exist any deflagration wave solution to (42) with left state $p_L = 1$, $\rho_L = 1$, $u_L = 0$, $z_L = 1$, if K_0 is too large.

This is in correspondence with the results of Tegnér. In [11] he showed that it does not exist any detonation wave solution to (42) when $\mu = 1$, $\lambda = 0$ and $D = 0$, if $K_0 > K_{0\max}$.

In [11], Tegnér also showed that it exist a unique detonation wave solution to (42) when $\mu = 1$, $\lambda = 0$ and $D = 0$, if $K_0 \leq K_{0\max}$.

5.2.2 A numerical method to compute deflagration waves

To find a deflagration wave, $\mathbf{w}(\xi) = (\,\tau(\xi), T(\xi), z(\xi), z'(\xi)\,)^T$, satisfying (47), we discretize in space. $N+1$ grid points are distributed equally over the interval $[\xi_0, \xi_N]$. In the interior part of the wave, i.e., at $\xi_1, \xi_2, \ldots, \xi_{N-1}$, the differential operator $d/d\xi$ in (47) is approximated by the central difference operator D_0. By this approximation, we obtain the discretized system

$$\begin{cases} \dfrac{\tau_{j+1} - \tau_{j-1}}{2\Delta\xi} = f_1(\tau_j, T_j, z_j, z_j'), \quad \text{j=1,2,...,N-1}, \\ \quad\vdots \end{cases} \tag{51}$$

At the endpoints we use one sided second order approximations, e.g.,

$$\frac{-3\tau_0 + 4\tau_1 - \tau_2}{2\Delta\xi} = f_1(\tau_0, T_0, z_0, z_0').$$

We impose

$$\mathbf{l}_i^T \mathbf{w}_0 = \mathbf{l}_i^T \mathbf{w}_L,$$
$$\mathbf{l}_i^T \mathbf{w}_N = \mathbf{l}_i^T \mathbf{w}_R,$$

for the in-going waves at the left and the right boundary respectively. The left eigenvectors, $\mathbf{l}_i^T$, are defined by $\mathbf{l}_i^T \mathbf{r}_j = \delta_{ij}$. $\mathbf{r}_j$ are the right eigenvectors of the Jacobian matrix $\partial \mathbf{f}/\partial \mathbf{w}$, evaluated at the limiting states $\mathbf{w}_L$ and $\mathbf{w}_R$ respectively.

The system (51) is solved by Newton's method for the unknown (τ_j, T_j, z_j, z_j'). s is kept fixed. In general we do not obtain a full wave from $\mathbf{w}_L$ to $\mathbf{w}_R$, unless s is tuned to the correct wave speed. At the correct wave speed, however, the system becomes singular, due to the non-uniqueness of the wave. It can be translated arbitrarily. To overcome this problem, we let s be an additional unknown, and add the equation $z(0) = 0.5$ to the system of equations, so that the center of the wave front will appear at $\xi = 0$. The modified system, with s as an additional unknown, is solved by Newton iteration. The quadratic convergence is rapid and we obtain wave profiles with a large number of grid points in a few seconds of cpu-time.

In Figure 16 we display the obtained deflagration waves when the viscous constants are set to $\mu = 0.2$, $\lambda = 0.3$, $D = 0.2$, and K_0 is set to 500, 1000 and 2000 respectively. The left hand side of the wave is specified by $p_L = 1$, $\rho_L = 1$, $u_L = 0$, $z_L = 1$. The corresponding speeds of propagation for these choices of K_0 are $s \approx$ 0.085887, 0.11928 and 0.16390 respectively.

5.3 Some computations with the Navier-Stokes equations

Consider the Navier-Stokes equations for a reacting mixture, (42), when the source term is modeled by Arrhenius kinetics, i.e., $K(T) = K_0 T^{\alpha} e^{-E/T}$ and the parameters in the

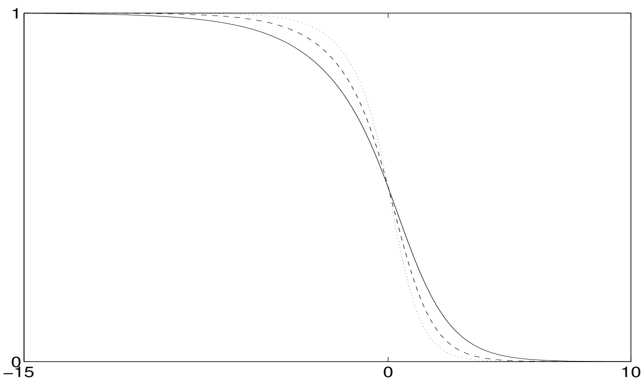


Figure 16: *The z-component of some deflagration waves. The deflagration wave corresponding to $K_0 = 500$ is displayed with a drawn line. The deflagration waves corresponding to $K_0 = 1000$ and 2000 are displayed with a dashed line and a dotted line respectively.*

system of equations are chosen as follows, $K_0 = 2000$, $\alpha = 0$, $E = 50$, $q_0 = 50$, $\gamma = 1.2$, $\mu = 0.2$, $\lambda = 0.3$ and $D = 0.2$.

An approximative deflagration wave solution to this system of equations was computed in the previous section, with $p = \rho = z = 1$ on the left hand side of the wave and $z = 0$ on the right hand side. The speed of the wave was determined to be approximatively equal to 0.163903. The error in this approximation is of the order 10^{-6}.

In this section we will follow the development of this traveling wave by using the explicit Runge-Kutta scheme of third order to integrate the Navier-Stokes equations in time. We will investigate if the accuracy of the obtained numerical wave speed is affected by the order of approximation of the numerical scheme. Specifically we want to see if the result for the scalar model equation still holds, i.e., if the error in the wave speed is proportional to Δx^r, where r is the order of approximation of the numerical method.

To measure the numerical speed of propagation we first take 40000 time steps to let the wave stabilize itself. Then we take another 40000 time steps and calculate the average wave speed as the change in z as follows,

$$s^* = -\sum_{j=1}^{N} \frac{\left(z_j^{80000} - z_j^{40000}\right) \cdot \Delta x}{(z_L - z_R) \cdot 40000 \cdot \Delta t},$$

where $z_L - z_R = 1$.

To prevent the wave from traveling out of the computational domain the wave is moved such that

$$\begin{cases} (\rho z)_j^n > (\rho z)_j^0, & \forall j \leq N/2, \\ (\rho z)_{N/2+1}^n \leq (\rho z)_j^0, & \end{cases}$$

is fulfilled after every time step.

For the convective part of the system of equations, the stability requirement is $\Delta t \leq a_1 \Delta x / |\lambda_c|$, where λ_c is the largest of the eigenvalues of the Jacobian matrix, $\partial \mathbf{f} / \partial \mathbf{w}$. The stability conditions for the reactive term is $\Delta t \leq \min(a_2 / |\min(0, \lambda_s)|)$, where λ_s is the nonzero eigenvalue of $\partial \mathbf{\Psi} / \partial \mathbf{w}$, i.e., $\lambda_s = K(T)((\gamma - 1) q_0 z E / T^2 - 1)$. If the dissipative term is implemented explicitly, then the CFL-number has to be reduced for small Δx, because the stability requirement for an explicit approximation of $u_t = u_{xx}$ is on the form $\Delta t \leq a_3 \Delta x^2$.

By implementing the dissipative term implicitly, it is possible to increase the size of the time steps considerably for small space steps. However, in this section we are only interested in the effect on the numerical wave speed of the order of approximation of the numerical scheme. Hence we have chosen to implement the dissipative term as simple as possible. We remark that the eigenvalues corresponding to the viscous part of the system of equations are 0, μ/ρ, $\lambda(\gamma - 1)/\rho$ and D/ρ respectively. Therefore the dissipative term would not affect the CFL-number if $\mu \approx \lambda \approx D \approx 0$.

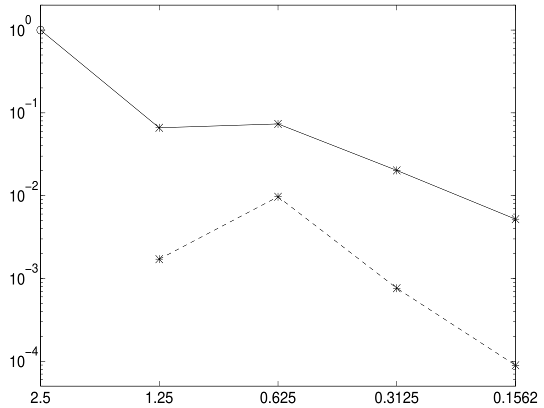
The approximative deflagration wave solution with 1601 grid points, which was calculated in the previous section, is used as initial data in the numerical experiments. The implicit Runge-Kutta scheme (16) is used for the time integration when $\Delta x > 5/4$. When $\Delta x \leq 5/4$, the explicit version of the Runge-Kutta scheme is used. In the first two experiments the convective term $\mathbf{f}(\mathbf{w})_x$ is approximated by central differences of order two and four respectively. One sided approximations are used at the boundaries. We impose

$$\mathbf{l}_i^T \mathbf{w}_0 = \mathbf{l}_i^T \mathbf{w}_L,$$

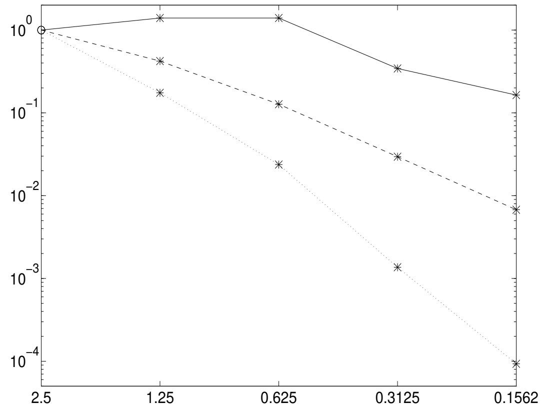
$$\mathbf{l}_i^T \mathbf{w}_N = \mathbf{l}_i^T \mathbf{w}_R,$$

for the in-going waves at the left and the right boundary respectively. The left eigenvectors $\mathbf{l}_i^T$ are evaluated at the far field. The viscous term is approximated explicitly by conservative central differences of order four. Because of stability reasons the viscous term is set to zero at the boundaries. This modification does not affect the order of the numerical method, since the viscous term is neglectable at the boundaries. The time step is set to $\Delta t = 0.1 \min(\Delta x, \Delta x^2)$. By choosing Δt proportional to Δx^2, the truncation error in the approximation of the time derivatives will be of order $O(\Delta x^6)$. Hence the order of approximation in the numerical methods is two and four respectively. The relative error in the resulting wave speeds is displayed in Figure 17(a). The error in the wave speed is of order two and four respectively.

In Figure 17(b) we show the relative error in the numerical wave speeds, which are obtained with the central biased ENO scheme [8] of order one, two and three. The selection of the numerical flux functions is made in the local characteristic variables. The eigenvalues and the right eigenvectors of the Jacobian matrix $\partial \mathbf{f} / \partial \mathbf{w}$ can be found in [4]. The z-component of some numerical solutions at $t = 100$ is plotted in Figure 18. In general, the numerical wave speeds obtained with a high order scheme are more accurate



(a) *Numerical results obtained when $\mathbf{f}_x$ is approximated by central differences of order two and four displayed with a drawn line and a dashed line respectively.*



(b) *Numerical results obtained when $\mathbf{f}_x$ is approximated by an ENO method of order one displayed with a drawn line. Results obtained with an ENO method of order two and three are shown with a dashed line and a dotted line respectively.*

Figure 17: *The relative error in wave speed as a function of Δx. The cases when the numerical wave speed, s^*, is greater than the exact wave speed, s, are marked with small stars. The cases when $s^* < s$ are marked with small circles.*

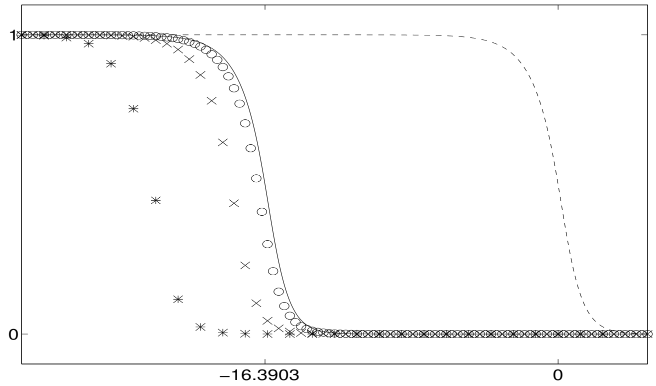
than the ones obtained with a low order scheme. For large Δx, the relative error in the numerical wave speed can be more than 100%. The error of the wave speed is once again of the same order as the numerical method.

In the following numerical experiment we show that the cancelation of truncation error technique from Section 4.5 also can be applied to the Navier-Stokes equations for a reacting mixture with success. In this experiment, the convective term is approximated by an ENO approximation of order two. The source term is then modified to cancel the major part of the truncation error in the fourth component. This is done by computing the fourth component of the difference between the ENO approximation and the fourth order central difference approximation of the convective term. This difference is then added to the source term. The relative error in the resulting wave speed is displayed in Figure 19(a). In Figure 19(b) we show the z-component of some numerical solutions at $t = 100$. Clearly, the modification increases the accuracy of the wave speed for all Δx, but it does not increase the order of the numerical method.

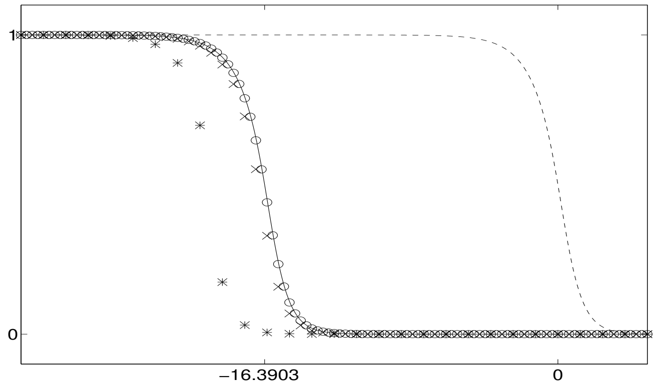
We remark that the cancelation of error technique can not be applied when the convective term is approximated by central differences of second order, since the resulting method will not be linearly stable.

Consider the convective part of the reactive Navier-Stokes equations,

$$\mathbf{w}_t + \mathbf{f}(\mathbf{w})_x = 0.$$

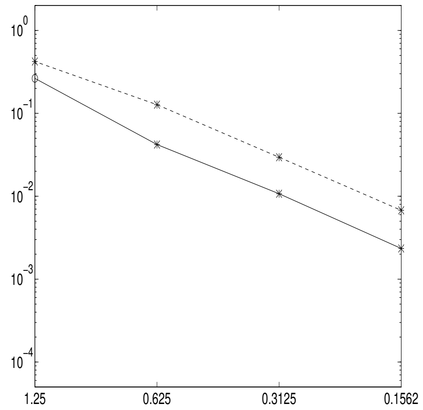


(a) *Numerical results obtained with the central biased ENO scheme of order two.*

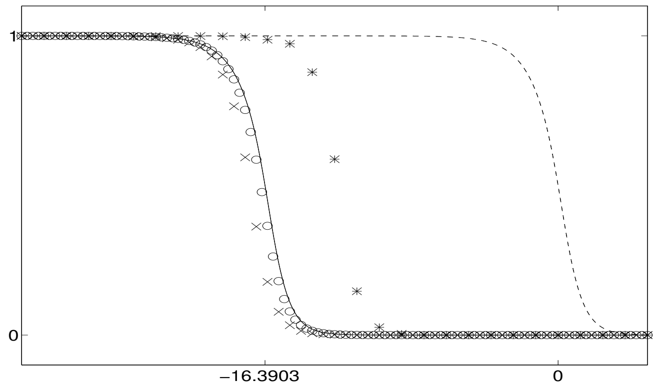


(b) *Numerical results obtained with the central biased ENO scheme of order three.*

Figure 18: *The z-component of some numerical solutions at $t = 100$. These solutions are obtained with the central biased ENO scheme of order two and three. The numerical results which corresponds to $\Delta x = 5/4$ are shown with small stars. The waves obtained when $\Delta x = 5/8$ and $\Delta x = 5/16$ are marked with crosses and circles respectively. The initial data is displayed with a dashed line. The exact solution at $t = 100$ is displayed with a drawn line.*



(a) *The relative error in wave speed as a function of Δx. Numerical results obtained when the source term is modified to cancel the truncation error are displayed with a drawn line. Results obtained when the source term is not modified are shown with a dashed line.*



(b) *The z-component of some numerical solutions at $t = 100$. The numerical results which corresponds to $\Delta x = 5/4$ are shown with small stars. The waves obtained when $\Delta x = 5/8$ and $\Delta x = 5/16$ are marked with crosses and circles respectively. Compare with Figure 18(a).*

Figure 19: *Numerical results obtained when the convective term is approximated by an ENO method of second order and the source term is modified to cancel the truncation error.*

The Jacobian $\mathbf{J} = \partial \mathbf{f}/\partial \mathbf{w}$ of this system of equations has real eigenvalues. Furthermore, in Fourier space the second order central difference operator D_0 corresponds to a multiplication with

$$Q_2(\omega) = i \sin \omega.$$

Moreover, the fourth order central difference operator corresponds to a multiplication with

$$Q_4(\omega) = \frac{i}{12}(8 \sin \omega - \sin 2\omega).$$

Therefore, the eigenvalues of $-J \cdot Q_2(\omega)$ and $-J \cdot Q_4(\omega)$ will always be imaginary, and hence the corresponding central difference methods are linearly stable.

On the other hand, the matrix which corresponds to the numerical method with central differences of different orders, i.e.,

$$-J \cdot \begin{pmatrix} Q_2(\omega) & & & \\ & Q_2(\omega) & & \\ & & Q_2(\omega) & \\ & & & Q_4(\omega) \end{pmatrix}$$

will have eigenvalues with positive real part at some places in the deflagration wave. Because of this fact, this numerical method will be unstable.

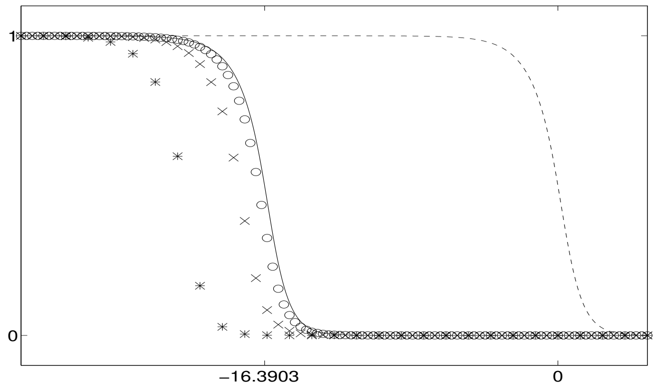
In a final experiment we investigate the effect of numerical viscosity on the numerical wave speed. An ENO method of order two is used to compute the numerical flux functions, $\mathbf{h}_{j+1/2}$. The dissipative term $\epsilon \Delta_+^2 \Delta_- \mathbf{w}_j$, where ϵ is a positive number, is then added to every numerical flux function. The z-component of some numerical solutions at $t = 100$ is shown in Figure 20. The more artificial viscosity being added to the numerical scheme, the higher is the accuracy in the numerical wave speed. To be able to add a dissipative term, $\epsilon \Delta_+^2 \Delta_- \mathbf{w}_j$, with $\epsilon > 0.1$, we have to reduce Δt.

6 Conclusions

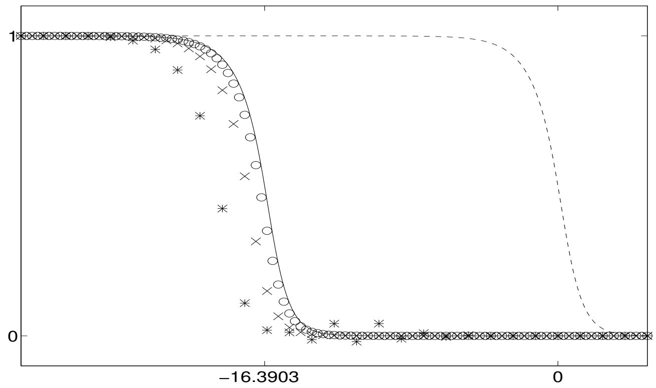
We have treated hyperbolic conservations laws which are coupled to a source term. For large $K\Delta x$, where K is the reaction rate, traveling wave solutions to these equations move at a totally wrong speed.

It is shown that for detonation waves the error in the numerical speed of propagation is of order $K\Delta x$, independently of the order of accuracy of the numerical method. The accuracy of the propagation speed can be significantly improved by using the extrapolation technique which is presented in this paper.

For deflagration waves, the error in the speed of propagation is of order $(K\Delta x)^r$, where r is the order of approximation of the numerical method. Hence, convergence is much faster with a high order method once the layer is resolved.



(a) $0.05\Delta_+^2\Delta_-\mathbf{w}_j$ *added to every numerical flux function.*



(b) $0.1\Delta_+^2\Delta_-\mathbf{w}_j$ *added to every numerical flux function.*

Figure 20: *The z-component of some numerical solutions at $t = 100$, obtained with the central biased ENO scheme of order two. An artificial viscosity term is added to every numerical flux function. The numerical results which corresponds to $\Delta x = 5/4$ are shown with small stars. The waves obtained when $\Delta x = 5/8$ and $\Delta x = 5/16$ are marked with crosses and circles respectively. Compare with Figure 18(a).*

An increase of the order of accuracy in space can be done by either a change of the numerical approximation of the convective term, or by a modification of the source term. We have discussed how to modify the source term to cancel the major part of the truncation error in space. To avoid overshoots at discontinuities the source term should only be modified in smooth deflagration waves. A detector has been constructed to tell which parts of the numerical solution belongs to a deflagration wave.

What makes the higher order methods approximating the wave speed better for deflagration waves, is the fact that they approximate the shape of the solution more accurately. Hence, when the grid values move through the wave, they experience the different parts of the source term function in a way that is more fair in an average way of view, and as a direct consequence the accuracy of the wave speed is increased. However, for large $K\Delta x$ the source term is so dominant that no grid point will ever move through the wave. The remedy here is to add a large amount of artificial viscosity to the numerical method, making the grid points move, and the accuracy of the wave speed increase drastically.

7 Acknowledgments

We thank Björn Sjögreen for the helpful discussions and the constructive comments on this work.

References

[1] A. Bermudez and E. Vazquez. Upwind methods for hyperbolic conservation laws with source terms. *Computers and Fluids*, 23(8):1049–1071, November 1994.

[2] P. Colella, A. Majda, and V. Roytburd. Theoretical and numerical structure for reacting shock waves. *SIAM Journal on Scientific and Statistical Computing*, 7(4):1059–1080, October 1986.

[3] B. Engquist and B. Sjögreen. Numerical approximation of hyperbolic conservation laws with stiff terms. Report CAM 89-07, Department of Mathematics, UCLA, 1989.

[4] B. Engquist and B. Sjögreen. Robust difference approximations of stiff inviscid detonation waves. Report CAM 91-03, Department of Mathematics, UCLA, 1991.

[5] A. Harten. Eno schemes with subcell resolution. *Journal of Computational Physics*, 83(1):148–184, July 1989.

[6] P. Klingenstein. Hyperbolic conservation laws with source terms: Errors of the shock location. Research Report No. 94-07, Eidgenössische Technische Hochschule, CH-8092 Zürich, Switzerland, August 1994.

[7] R. J. LeVeque and H. C. Yee. A study of numerical methods for hyperbolic conservation laws with stiff source terms. *Journal of Computational Physics*, 86(1):187–210, January 1990.

[8] D. Lindström. A central biased eno scheme of arbitrarily high accuracy. Report no. 158/1994, Uppsala University, Department of Scientific Computing, March 1994.

[9] A. Majda. A qualitative model for dynamic combustion. *SIAM Journal on Applied Mathematics*, 41(1):70–93, August 1981.

[10] C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes, ii. *Journal of Computational Physics*, 83(1):32–78, July 1989.

[11] J. Tegnér. *Properties of Detonation Waves.* PhD thesis, Royal Institute of Technology, Stockholm, Sweden, December 1992.

[12] F. A. Williams. *Combustion Theory.* The Benjamin/Cummings Publishing Company, Inc., second edition, September 1988.