5M: Advanced Numerical Methods
Assessed Coursework 2
Hyperbolic PDEs

Yana Staneva – 2019862s

Friday 16th December, 2016

Consider the linear advection equation in 1-D, given by

$$\frac{\partial u}{\partial t} + a(x,t)\frac{\partial u}{\partial x} = 0, \tag{1}$$

which is the simplest example of a 1-D hyperbolic PDE. The analytic solution of this equation is obtained via the method of characteristics, which relies on solving the Monge Equations, as studied in 4H PDEs, namely

$$\frac{dt}{d\tau} = 1, \qquad \frac{dx}{d\tau} = a(x,t), \qquad \frac{du}{d\tau} = 0.$$

In order to solve the PDE numerically, we need to discretize the continuous plane (x,t) by a grid with uniform spacing Δx in the x direction and time step Δt in the t direction, i.e. $x_j = x_0 + j\Delta x$ and $t_n = t_0 + n\Delta t$. Thus, we evaluate the function U at the discretized points, i.e. $U(x_j, t_n) = U(x_0 + j\Delta x, t_0 + n\Delta t) \equiv U_j^n$.

The simplest and easiest to implement scheme, which is also stable, is the *upwind scheme*, discussed in Lectures. We now set $a =$ constant, until specified otherwise. The scheme itself is given by

$$U_j^{n+1} = \begin{cases} U_j^n - a\dfrac{\Delta t}{\Delta x}\left[U_{j+1}^n - U_j^n\right] & \text{for } a < 0, \\ \\ U_j^n - a\dfrac{\Delta t}{\Delta x}\left[U_j^n - U_{j-1}^n\right] & \text{for } a > 0. \end{cases}$$

We now list several features of the upwind scheme to be used later on as a point of reference for comparison:

- The upwind scheme relies on computing the current value of the function U_j^n via linear interpolation, as described in Morton & Mayers.
- The *amplification factor* of the scheme is given by $\lambda(k) = 1 - \nu(1 - e^{-jk\Delta x})$, where $\nu := |a|\Delta t/\Delta x$.
- The CFL condition gives the correct stability restrictions, i.e. $|\lambda(k)| \leq 1$, $\forall k$, if $0 \leq \nu \leq 1$.
- The *amplitude error per time step* is of order ξ^2, where $\xi = k\Delta x$.
- The *global amplitude error* is of order ξ.
- The phase of the numerical mode is given by $\arg\lambda \sim -\nu\xi\left[1 - \dfrac{1}{6}(1-\nu)(1-2\nu)\xi^2 + \cdots\right]$.
- The *relative phase error* is of order ξ^2, with the sign depending on the value of ν and is equal to 0 when $\nu = \dfrac{1}{2}$.

Regardless of the small phase error, which correctly approximates the speed of the travelling wave, the upwind scheme produces solutions with severe damping, which may result in incorrect calculations of the amplitudes of the waves. Instead, we consider a scheme, which relies on quadratic interpolation using three points, as described in Morton & Mayers, see Fig. 4.5, namely the *Lax–Wendroff scheme*, given by

$$U_j^{n+1} = \frac{1}{2}\nu(1+\nu)U_{j-1}^n + (1-\nu^2)U_j^n - \frac{1}{2}\nu(1-\nu)U_{j+1}^n,$$

assuming a straight characteristic with slope ν.

We now perform the usual Fourier analysis to study the amplification of the Lax–Wendroff scheme. Substitute with the ansatz $U_j^n = \lambda^n e^{ikj\Delta x}$ to obtain

$$\lambda^{n+1} e^{ikj\Delta x} = \frac{1}{2}\nu(1+\nu)\lambda^n e^{ik(j-1)\Delta x} + (1-\nu^2)\lambda^n e^{ikj\Delta x} - \frac{1}{2}\nu(1-\nu)\lambda^n e^{ik(j+1)\Delta x}.$$

Upon cancellation of the $\lambda^n e^{ikj\Delta x}$ term we obtain

$$\begin{aligned}
\implies \lambda &= \frac{1}{2}\nu(1+\nu)e^{-ik\Delta x} + (1-\nu^2) - \frac{1}{2}\nu(1-\nu)e^{ik\Delta x} && \text{use Euler's formula to substitute with trig functions,} \\
&= \nu^2\cos(k\Delta x) - i\nu\sin(k\Delta x) + (1-\nu^2) && \text{grouping by trig functions,} \\
&= 1 - i\nu\sin(k\Delta x) - 2\nu^2\sin^2\left(\frac{k\Delta x}{2}\right) && \text{using } \cos(2\alpha) = 1 - 2\sin^2(\alpha).
\end{aligned}$$

Next we want to analyze the damping of the Lax–Wendroff scheme. We proceed by deriving an expression for $|\lambda|^2$ as follows

$$\begin{aligned}
|\lambda|^2 &= \left(1 - 2\nu^2\sin^2\left(\frac{k\Delta x}{2}\right)\right)^2 + \nu^2\sin^2(k\Delta x) && \text{grouping and squaring,} \\
&= 1 - 4\nu^2\sin^2\left(\frac{k\Delta x}{2}\right) + 4\nu^4\sin^4\left(\frac{k\Delta x}{2}\right) + \nu^2\sin^2(k\Delta x) && \text{expanding the brackets,} \\
&= 1 - 4\nu^2\sin^2\left(\frac{k\Delta x}{2}\right) + 4\nu^4\sin^4\left(\frac{k\Delta x}{2}\right) + 4\nu^2\sin^2\left(\frac{k\Delta x}{2}\right)\cos^2\left(\frac{k\Delta x}{2}\right) && \text{by } \sin(2\alpha) = 2\sin(\alpha)\cos(\alpha), \\
&= 1 - 4\nu^2\sin^2\left(\frac{k\Delta x}{2}\right) + 4\nu^4\sin^4\left(\frac{k\Delta x}{2}\right) + 4\nu^2\sin^2\left(\frac{k\Delta x}{2}\right)\left[1 - \sin^2\left(\frac{k\Delta x}{2}\right)\right] && \text{Pythagorean trig identity,} \\
&= 1 - \cancel{4\nu^2\sin^2\left(\frac{k\Delta x}{2}\right)} + 4\nu^4\sin^4\left(\frac{k\Delta x}{2}\right) + \cancel{4\nu^2\sin^2\left(\frac{k\Delta x}{2}\right)} - 4\nu^2\sin^4\left(\frac{k\Delta x}{2}\right) && \text{cancelations,} \\
&= 1 - 4\nu^2\left(1-\nu^2\right)\sin^4\left(\frac{k\Delta x}{2}\right).
\end{aligned}$$

Next we wish to find a bound for ν, which satisfies $|\lambda| \leq 1$. Hence,

$$\begin{aligned}
|\lambda| \leq 1 &\iff \sqrt{1 - 4\nu^2\left(1-\nu^2\right)\sin^4\left(\frac{k\Delta x}{2}\right)} \leq 1 \\
&\iff 0 \leq \nu^2\left(1-\nu^2\right)\sin^4\left(\frac{k\Delta x}{2}\right) && \text{after cancelations and rearranging} \\
&\iff \nu^2 \leq 1 && \text{as } \sin^4(\alpha) \in [0,1] \\
&\iff |\nu| \leq 1.
\end{aligned}$$

Thus, we conclude that the Lax–Wendroff scheme is stable for values of $\nu \in [-1, 1]$, which is much less restrictive than the stability allowed by the upwind scheme, where we require $\nu \in [0, 1]$. However, there is still some damping as $|\lambda| < 1$. Also the *amplitude error per time step* is of order ξ^4, which is much better compared to the one of the upwind scheme, which is of order ξ^2. Let us now examine the dispersion. Hence, for the argument of λ we obtain

$$\arg\lambda = -\tan^{-1}\left[\frac{\nu\sin(\xi)}{1 - 2\nu^2\sin^2\left(\frac{\xi}{2}\right)}\right], \qquad \text{where} \quad \xi = k\Delta x.$$

Using Wolfram Alpha, we compute the Taylor expansion of the term in the brackets, which yields

$$\frac{\nu\sin(\xi)}{1 - 2\nu^2\sin^2\left(\frac{\xi}{2}\right)} = \nu\xi + \frac{1}{6}\nu(3\nu^2 - 1)\xi^3 + O(\xi^5).$$

Hence, using Lemma 4.1 from Morton & Mayers, we obtain that

$$\tan^{-1}\left[\frac{\nu\sin(\xi)}{1-2\nu^2\sin^2\left(\frac{\xi}{2}\right)}\right] \sim \nu\xi + \left(\frac{1}{6}\nu(3\nu^2-1) - \frac{\nu^3}{3}\right)\xi^3 + O(\xi^5)$$
$$\sim \nu\xi + \frac{1}{6}\nu(\nu^2-1)\xi^3 + O(\xi^5).$$

We finally substitute back in the expression for the argument of λ to obtain

$$\arg\lambda \sim -\nu\xi\left[1 - \frac{1}{6}(1-\nu^2)\xi^2 + O(\xi^4)\right].$$

Hence, the *relative phase error* of the Lax–Wendroff scheme is of order ξ^2, which is exactly the same as the upwind scheme for values of $\nu \approx 0$. However, the Lax–Wendroff scheme produces some phase error, meaning that the wave does not travel with the right speed and thus, we observe some lag. Nevertheless, the much smaller damping of the Lax–Wendroff scheme usually makes it the more preferable method, as the damping of the upwind method results in more significant loss of information about the amplitude of the waves.

From now on we consider Equation (1), with speed given by $a(x,t) = \dfrac{1+x^2}{(1+x^2)^2+2xt}$. Then the first Monge equation yields that $t = \tau$ by imposing the initial conditions, and hence, using the second equation $\dfrac{dx}{dt} = a(x,t)$ we get

$$\left((1+x^2)^2\right)\frac{dx}{dt} + 2xt\frac{dx}{dt} = (1+x^2) \qquad \text{by common denominator and expanding.}$$

Next, we divide both sides by $(1+x^2)^2$ to obtain

$$\frac{dx}{dt} = -\frac{2xt}{(1+x^2)^2}\frac{dx}{dt} + \frac{1}{1+x^2} \qquad \text{observe that the RHS is the total time derivative of } \frac{t}{(1+x^2)},$$

$$\int\frac{dx}{dt} = \int\frac{d}{dt}\left(\frac{t}{(1+x^2)}\right) \qquad \text{integrate both sides,}$$

$$x_k = \frac{t}{(1+x_k^2)} + x_{k_0} \qquad \text{where } x_{k_0} \text{ – constant, determined by IC.}$$

Examining the initial profile of the wave, we see that at $t = 0$, $x_b = 0.1 = x_{b_0}$ and $x_f = 0.3 = x_{f_0}$. Thus,

$$x_b(t) = \frac{t}{1+x_b^2(t)} + 0.1 \qquad \text{and} \qquad x_f(t) = \frac{t}{1+x_f^2(t)} + 0.3.$$

Next we discuss the Lax–Wendroff method for non-constant speed $a(x,t)$. The scheme is given by

$$U_j^{n+1} = U_j^n - a_j^n\Delta t\frac{\Delta_{0x}U_j^n}{\Delta x} + \frac{1}{2}(\Delta t)^2\left[-(a_t)_j^n\frac{\Delta_{0x}U_j^n}{\Delta x} + a_j^n\frac{\delta_x(a_j^n\delta_x U_j^n}{(\Delta x)^2}\right],$$

with notation as in Morton & Mayers, and is implemented in the two separate MATLAB files `laxWendroff002.m` and `laxWendroff001.m` for the two different spatial steps specified $\Delta x = 0.02$ and $\Delta x = 0.01$ respectively. Note that we choose $\Delta t = \Delta x$ in both cases, which makes sense since the speed $a(x,t) < 1$ for all values of x and t and, thus, the CFL condition is satisfied. Below we superimpose the numerical and exact solutions of $u(x,t)$ as a function of the space x at times $t = 0.1$, $t = 0.2$, $t = 0.5$, and $t = 1$ with the two different spatial steps in Figure 1. The exact solution is obtained via the `roots` command for the specified quadratic polynomial, derived by the following two iterative formulas for the position of the back and the front of the square waves,

$$x_b^3 - x_b^2 + x_b - t - 0.1 = 0, \qquad \text{and} \qquad x_f^3 - x_f^2 + x_f - t - 0.3 = 0.$$

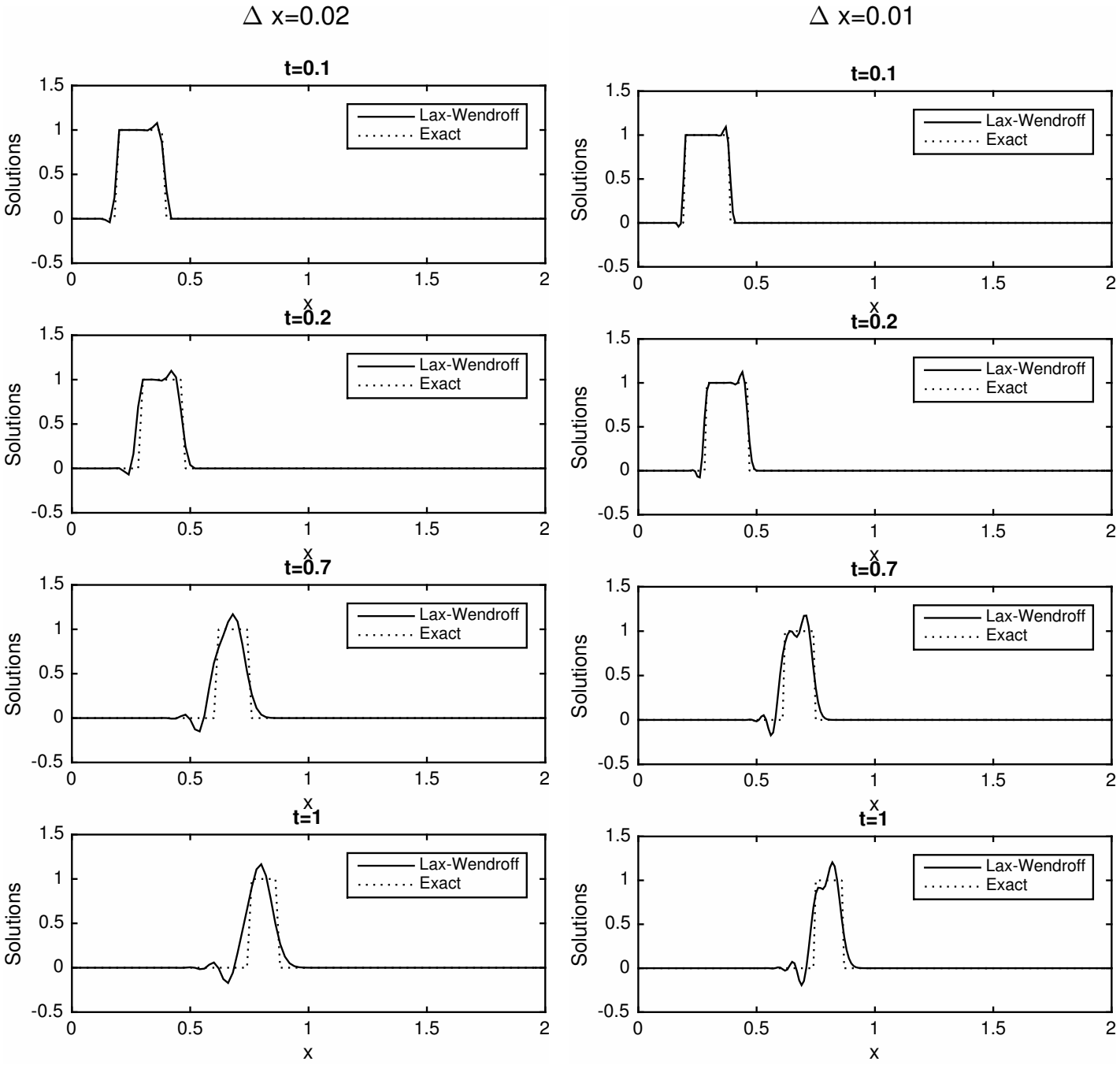

Figure 1: Numerical and Exact Solutions at different times with two different spatial steps.

Figure 1 illustrates both the maintenance of the height and width of the travelling wave and the small oscillations produced by the Lax–Wendroff scheme. When the spatial step is set as $\Delta x = 0.02$, the resulting plot is a bit less accurate than the one produced by $\Delta x = 0.01$. Thus, a decrease in the spatial step results in a more precise approximation of the pulse. Further, the oscillations occur slightly before the discontinuities of the square wave, which is a clear illustration of the lag discussed previously. Also, we observe that the oscillations are indeed the result of the weighted average of the three values, calculated on the previous time level, or in other words, the quadratic interpolation.

Below in Figure 2 we plot $u(x = 0.5, t)$ as a function of time t for the specific spatial position $x = 0.5$ with both spatial steps $\Delta x = 0.02$ and $\Delta x = 0.01$.

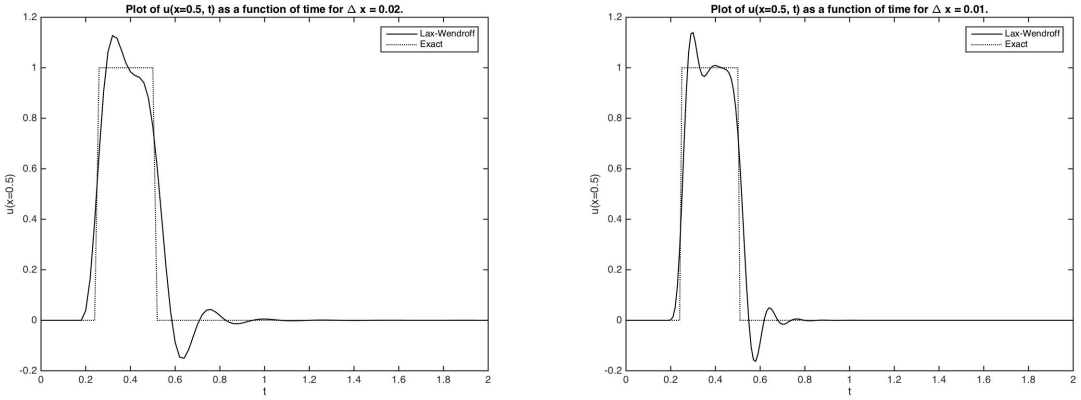

Figure 2: Plot of $u(x = 0.5, t)$ as a function of time.

The plots above in Figure 2 confirm our previous findings. Decreasing the spatial step results in a more accurate approximation, the oscillations occur due to the discontinuities, and the method preserves the height and width of the pulse more accurately compared to the upwind method. Hence, the Lax–Wendroff method outweighs the upwind method, regardless of the disadvantageous oscillations.

Next we proceed by quickly analyzing the positions of the front and the back of the waves. Figure 3 below illustrates the numerical estimation and analytic prediction of the back and front of the wave with a spatial step $\Delta x = 0.01$, where we have isolated points $|u(x,t)| > \varepsilon$ for four different tolerances, specified in the title of each subplot, produced by changing the value of the error and the title of the figure in the MATLAB file `frontBackEstimate.m` (Line 76).

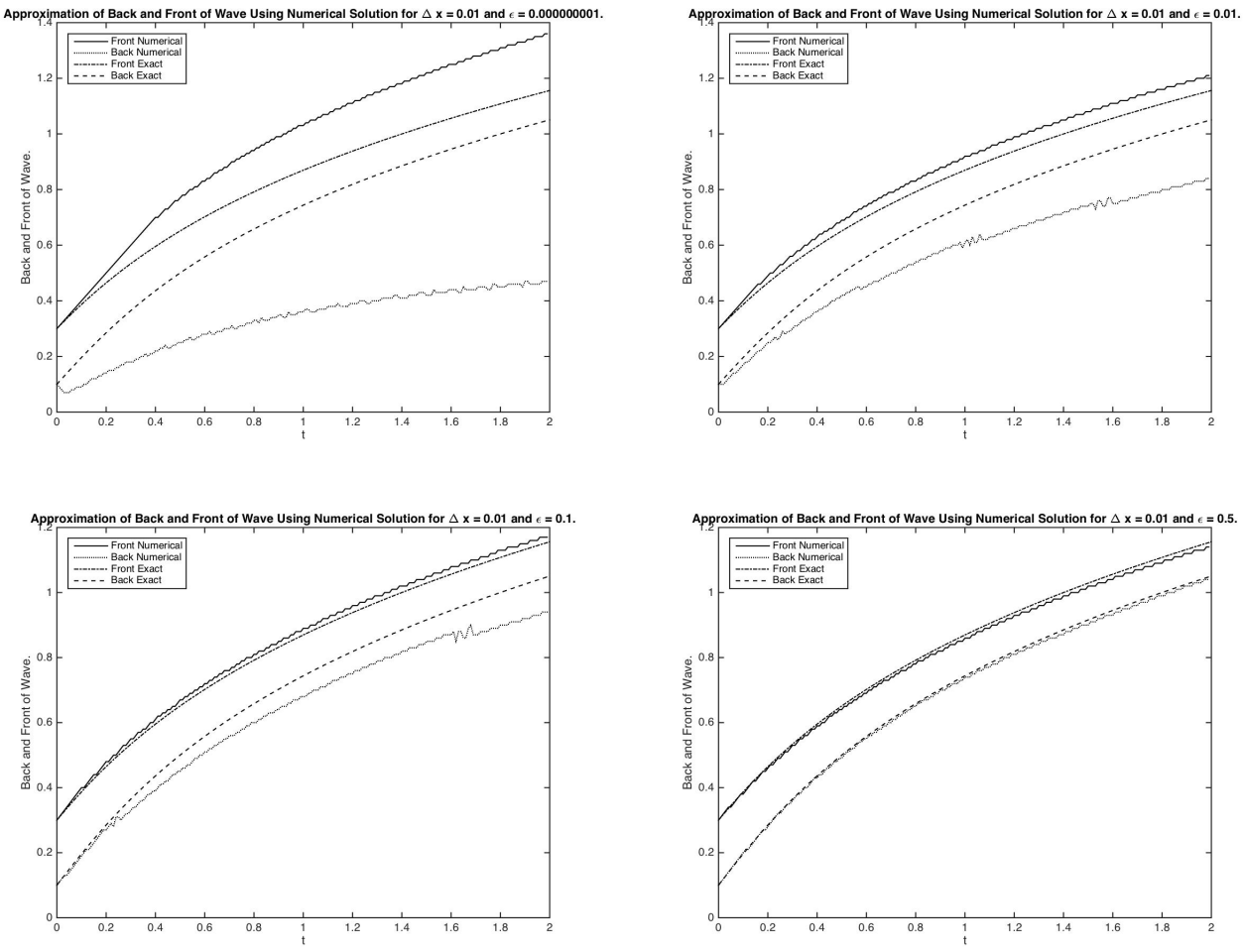

Figure 3: Numerical and Analytic Predictions for the Back and Front for $\Delta x = 0.01$.

Consider the plots in Figure 3 where $\Delta x = 0.01$. When the specified tolerance is $\varepsilon \sim 0.00000001$, the numerical solution yields a large deviation from the analytic prediction. As we increase the specified tolerance, $\varepsilon = 0.1$ we get a much better approximation. Ultimately, when the tolerance is $\varepsilon = 0.5$, our numerical approximations of the front and the back are quite satisfactory. Hence, we observe that when specifying a larger tolerance for the value $u(x,t)$, we get a better approximation of the location of the front and the back of the wave. This makes sense as by specifying a larger error tolerance, we actually consider more information from the pulse with values larger than 0. In other words, when we specify $\varepsilon = 0.5$, we consider the part of the peak of the wave with amplitude larger than the specified value. Thus, a higher tolerance leads to a more precise estimation as we consider more information about the non-trivial behaviour of the travelling wave.

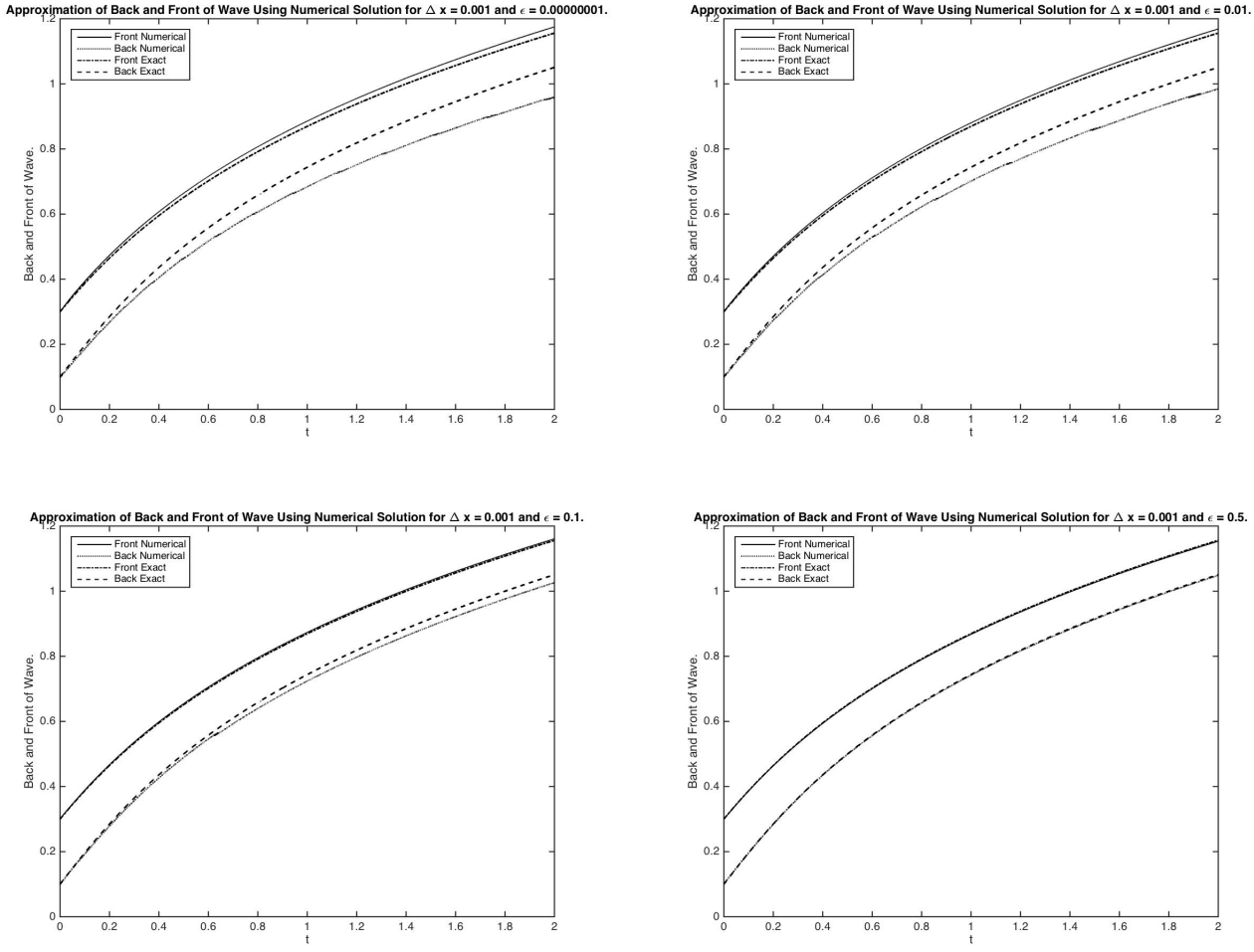

Figure 4: Numerical and Analytic Predictions for the Back and Front for $\Delta x = 0.001$.

Similar analysis can be applied to the plots in Figure 4. In this case we are considering a smaller spatial step $\Delta x = 0.001$, which further enhances our calculations. We see that for the same specified tolerances, we obtain a much more accurate approximation of the back and the front of the pulse. Again, as we increase the tolerance, we get closer to the exact analytic predictions. On close inspection of the last subplot, where the tolerance is set to be $\varepsilon = 0.5$, we see that the two estimations for the front, i.e. the numerical and the analytic, and the two for the back, actually coincide. Thus, in conclusion, a higher error tolerance ε such that $|u(x,t)| > \varepsilon$ leads to better numerical approximations of the front and the back of the pulse. Furthermore, a smaller spatial step also improves the numerical predictions.

In practice, hyperbolic equations are usually of the form

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = 0,$$

which is known as the *conservation form*. This can be easily rewritten in the form of Equation (1), where $a = \dfrac{\partial f}{\partial u}$. The Lax–Wendroff scheme can then be expressed as

$$U_j^{n+1} = U_j^n - \frac{1}{2}\frac{\Delta t}{\Delta x}\left\{\left[1 - A_{j+1/2}^n \frac{\Delta t}{\Delta x}\right]\Delta_{+x} F_j^n + \left[1 + A_{j-1/2}^n \frac{\Delta t}{\Delta x}\right]\Delta_{-x} F_j^n\right\}, \tag{2}$$

in the notation of Morton & Mayers. As we have seen, however, the Lax–Wendroff scheme is susceptible to producing errors because of the singularities occurring due to the intersection of two or more characteristics. Analytically this problem is tackled by the introduction of *shocks*, as covered in 4H PDEs. The Lax–Wendroff method performs quite well away from the shock, but once the shock occurs, the scheme yields large oscillations. On the contrary, the upwind scheme provides a reasonable approximation after the introduction of the shock. Hence, in some specific cases, it is reasonable to alternate between the two methods.

The Lax–Wendroff method can be easily generalized to tackle a system of equations. A necessary condition for the stability of the scheme is given by the higher dimensional analogue of the CFL condition, namely

$$\frac{\rho \Delta t}{\Delta x} \leq 1,$$

where ρ is the *spectral radius* of the constant matrix A, satisfying $\mathbf{f}(\mathbf{u}) = A\mathbf{u}$. This is another advantage of the scheme, since the condition involves the computation of the magnitude of the eigenvalues of A, rather than any sign considerations, in contrast with the upwind method. Ultimately, the Lax–Wendroff scheme can be easily extended to a two-step method, without the need to calculate the Jacobian A.

In general, the most commonly used and preferred methods for solving hyperbolic equations are the class of *finite volume schemes*. The name stems from the fact that a small volume surrounding each point on the discretized mesh contributes to the approximation of the function. The small volume is approximated by some form of quadrature.

As discussed above the Lax–Wendroff method produces oscillatory solutions. Finite volume schemes provide a way to somewhat overcome this problem, namely by controlling the *total variation* of the solution, defined by

$$TV(U^n) \coloneqq \sum_{j=1}^{J-1} |U_{j+1}^n - U_j^n| \equiv \sum_{j=1}^{J-1} |\Delta_{+x} U_j^n|,$$

where J is the number of partitions of the finite domain $[0, X]$. The key observation about the total variation of a conservation law is that it is a non-increasing function of time t. Hence, we arrive at the following key definition.

Definition 1. A scheme is said to be *total variation diminishing (TVD)* if the following condition is satisfied:

$$TV(U^{n+1}) \leq TV(U^n).$$

Next we need an important theorem.

Theorem 1 (Harten)**.** A scheme is TVD if it can be written in the form

$$U_j^{n+1} = U_j^n - C_{j-1}\Delta_{-x} U_j^n + D_j \Delta_{+x} U_j^n, \tag{3}$$

where C_j and D_j are any functions of the set of variables $\{U_j^n\}$, satisfying

$$C_j \geq 0, \quad D_j \geq 0, \quad \text{and} \quad C_j + D_j \leq 1, \quad \forall j.$$

Harten's theorem is of great use as it tells us that any monotonic scheme is TVD and a TVD scheme preserves monotonicity. Consider the *Roe first order upwind scheme*, which uses mid-point quadrature, namely

$$U_j^{n+1} = U_j^n - \frac{\Delta t}{\Delta x}\left(F_{j+1/2}^{n+1/2} - F_{j-1/2}^{n+1/2}\right),$$

where the *fluxes* are defined by

$$F_{j+1/2}^{n+1/2} = \frac{1}{2}\left[\left(1 + \operatorname{sign} A_{j+1/2}^n\right) F_j^n + \left(1 - \operatorname{sign} A_{j+1/2}^n\right) F_{j+1}^n\right]$$

$$F_{j-1/2}^{n+1/2} = \frac{1}{2}\left[\left(1 + \operatorname{sign} A_{j-1/2}^n\right) F_{j-1}^n + \left(1 - \operatorname{sign} A_{j-1/2}^n\right) F_j^n\right].$$

Now comparing the expression above with Equation (2) and with Equation (3), we set

$$C_{j-1} = \frac{1}{2}\frac{\Delta t}{\Delta x}\left(1 + \operatorname{sign} A^n_{j-1/2}\right) A^n_{j-1/2} \quad \text{and} \quad D_j = -\frac{1}{2}\frac{\Delta t}{\Delta x}\left(1 - \operatorname{sign} A^n_{j+1/2}\right) A_{j+1/2}.$$

It is straightforward to see that both $C_{j-1} > 0$ and $D_j > 0$, thus, satisfying the first two conditions of Harten's theorem. Let us now compute the sum of C_j and D_j as follows:

$$\begin{aligned} C_j + D_j &= \frac{1}{2}\frac{\Delta t}{\Delta x}\left[\left(1 + \operatorname{sign} A^n_{j+1/2}\right) A^n_{j+1/2} + \left(1 - \operatorname{sign} A^n_{j+1/2}\right) A^n_{j+1/2}\right] \\ &= \frac{1}{2}\frac{\Delta t}{\Delta x}\left[A^n_{j+1/2} + \cancel{A^n_{j+1/2}(\operatorname{sign} A^n_{j+1/2})} + A^n_{j+1/2} - \cancel{A^n_{j+1/2}(\operatorname{sign} A^n_{j+1/2})}\right] \\ &= \frac{1}{2}\frac{\Delta t}{\Delta x}\left[2A^n_{j+1/2}\right] = \frac{\Delta t}{\Delta x}\left|A^n_{j+1/2}\right|, \end{aligned} \tag{4}$$

which is the CFL number. Thus, we conclude that the last condition of the Harten theorem is equivalent to the CFL stability condition, and the Roe first order upwind scheme is indeed TVD for appropriate choices of Δt.

Before discussing numerical schemes any further, we consider the analytic solution to one of the fundamental hyperbolic PDEs. Consider the 1-D inviscid Burgers' equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0,$$

with ICs and BCs $u(\pm L, t) = 0$, $u(x, 0) = \operatorname{sech}(mx)$, $x \in [-10/m, 10/m]$, for $m > 0$. We wish to obtain the analytic solution of the equation above. Proceeding via the Monge equations, we obtain

$$\frac{dt}{d\tau} = 1, \quad \frac{dx}{d\tau} = u, \quad \frac{du}{d\tau} = 0.$$

The third equations gives $u =$ constant, i.e. independent of τ, and hence

$$t = \tau, \qquad x = ut + \sigma,$$

where σ is the characteristic parameter. Now since $u = F(\sigma)$ is independent of t by the third Monge equation, we get that $u = u(\sigma, 0) = u_0(\sigma)$. Thus, the solution is given implicitly by

$$u(x, t) = u_0(\sigma), \qquad \text{where } x = u_0(\sigma)t + \sigma = \operatorname{sech}(m\sigma)t + \sigma.$$

In other words, the value of u on each characteristics is fixed by the initial condition at $t = 0$, and in turn, the ICs determine the characteristics. In the (x, t)-plane each of the characteristics is a straight line, passing through $(\sigma, 0)$ with gradient $\dfrac{1}{u_0(\sigma)}$. The value of u at each point on the characteristic for σ is $u_0(\sigma)$. Let $F(\sigma) = \operatorname{sech}(m\sigma)$. Differentiate the expression for u with respect to x to obtain

$$u_x = u_0'(\sigma)\sigma_x \quad \text{and} \quad 1 = F'(\sigma)\sigma_x t + \sigma_x.$$

Eliminating the σ_x in both equations, we get

$$u_x = \frac{u_0'(\sigma)}{1 + tF'(\sigma)} \to \infty \qquad \text{as} \qquad 1 + tF'(\sigma) \to 0.$$

Hence, a vertical tangent first appears at

$$t_m = \min_\sigma \frac{-1}{F'(\sigma)} = \min_\sigma G(\sigma).$$

Now in our specific case we have

$$F(\sigma) = \operatorname{sech}(m\sigma) \implies F'(\sigma) = -m\tanh(m\sigma)\operatorname{sech}(m\sigma) \qquad \text{and} \qquad G(\sigma) = -\frac{1}{F'(\sigma)} = \frac{1}{m\tanh(m\sigma)\operatorname{sech}(m\sigma)}.$$

In order to minimize $G(\sigma)$ we compute its first time derivative to obtain

$$G'(\sigma) = \cosh(m\sigma) - \coth(m\sigma)\operatorname{csch}(m\sigma) = 0 \implies \sigma = \frac{\sinh^{-1}(1)}{m} = \tau^*.$$

Hence, finally we get an expression for the breaking time t_m, namely,

$$t_m = G(\tau^*) = \frac{1}{m\tanh\left(m\dfrac{\sinh^{-1}(1)}{m}\right)\operatorname{sech}\left(m\dfrac{\sinh^{-1}(1)}{m}\right)} = \frac{1}{m\dfrac{1}{\sqrt{2}}\dfrac{1}{\sqrt{2}}} = \frac{2}{m}.$$

Thus, we have computed the breaking time, which depends on the value of m, namely, $t_m = \dfrac{2}{m}$. To illustrate the characteristics, we use the MATLAB function `wvburg`, created by R. Knobel, which can be accessed here.

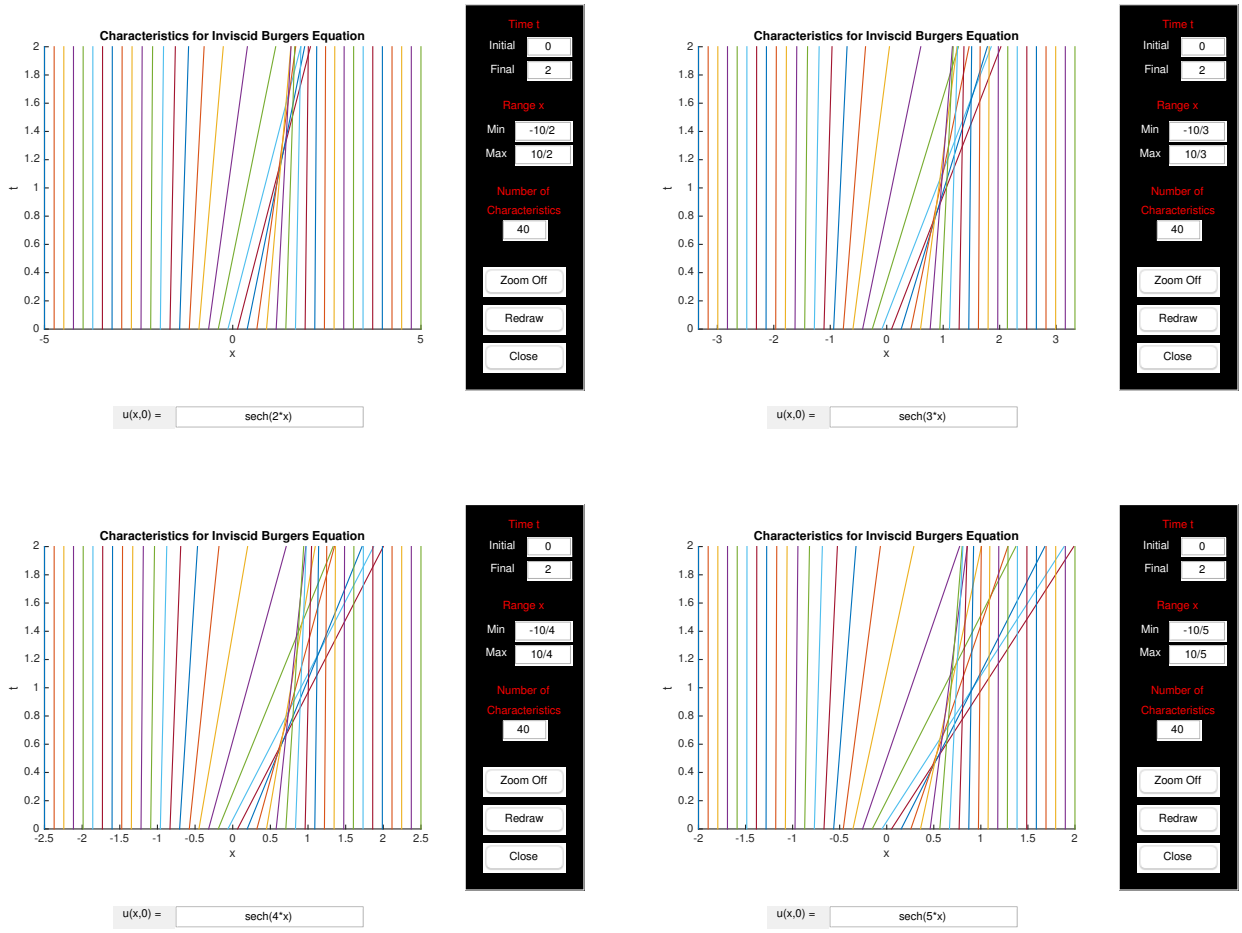

Figure 5: Plots of Characteristics in the (x,t)-Plane for Different Values of m.

Each of the plots in Figure 5 contains 40 different characteristics. In each plot we have changed the value of the parameter m in the initial condition $u_0 = \text{sech}(mx)$. This can be identified in the small box below each plot. In the top left subfigure $m = 2$ (we have changed the space domain as well $x \in [-5, 5]$, respectively), top right — $m = 3$ and $x \in [-10/3, 10/3]$, bottom left — $m = 4$ and $x \in [-10/4, 10/4]$, and bottom right — $m = 5$ and $x \in [-10/5, 10/5]$. The breaking times occur at $t = 1$, $t \approx 0.7$, $t \approx 0.5$, and $t \approx 0.4$, respectively for each plot as listed in the previous sentence, which makes sense as our analytic solutions yield $t_m = 1$, $t_m = 0.67$, $t_m = 0.5$, and $t_m = 0.4$ using $t_m = 2/m$. Hence, as we increase the value of the constant m, the breaking time t_m occurs earlier as it is inversely proportional to m.

In other words, by increasing the value of m, we actually decrease the spatial domain x, as $x \in [-10/m, 10/m]$. Hence, we get less grid points in the spatial interval x, and this results in an earlier shock formation. Now since we are solving the equation numerically, the earlier breaking time is not really in our favour. As soon as the characteristics intersect, we get a multivalued solution, and our numerical scheme simply fails to produce any sensible finite results.

Below in Figure 6 we provide a plot of the numerical solution of $u(x,t)$ as a function of x for four specific time instances. This will provide a clearer intuition about the problems we are facing once we reach the breaking time. We have set $\Delta t = \Delta x$. To see why, consider Equation (4). According to Harten's theorem, we require

$$\frac{\Delta t}{\Delta x}\left|A^n_{j+1/2}\right| \leq 1.$$

Now, recall that $A^n_{j+1/2} = \dfrac{F^n_{j+1} - F^n_j}{U^n_{j+1} - U^n_j}$, and in our example $A^n_{j+1/2} < 1$. Thus, if we choose any Δt such that $\Delta t \leq \Delta x$, our method is stable. Let us then discuss in greater detail the plot in Figure 6 produced by the MatLab file `finiteVolume.m`.

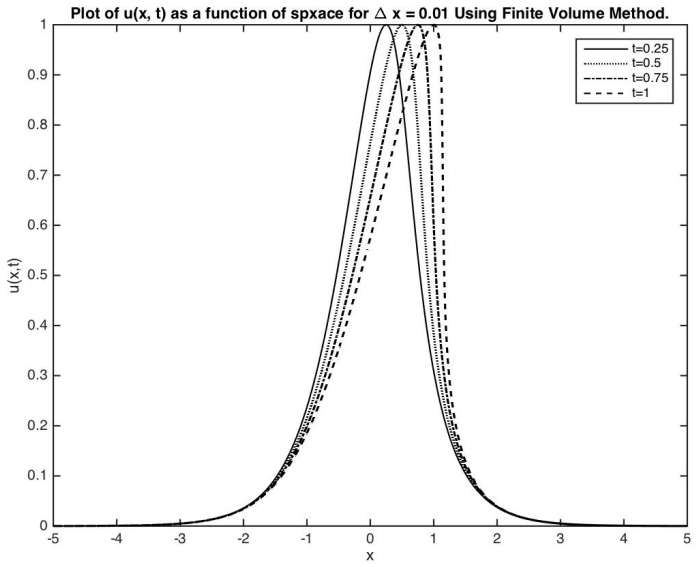

Figure 6: Numerical Solution of $u(x,t)$ via the Finite Volume Scheme for different time instances.

The plot in Figute 6 above illustrates numerous features of the implemented Roe upwind finite volume scheme. First, here we have set $m = 2$, which means that our spatial interval is $x \in [-5, 5]$, our IC is $u(x, 0) = \text{sech}(2x)$, and $\Delta x = \Delta t$, which clearly satisfies the stability condition as discussed above. Now, according to our analytic treatment of the equation, the breaking time occurs at $t_m = 1$. At this point we draw the attention of the reader to the graph where we observe that as $t \to t_m$, the wave 'tilts' to the right. In reality $t = t_m$ is the final instance at which we get any numerical approximations, since the multivalued solution occurs right after this point of time. If we go back to the characteristics setting, i.e. the (x, t) plane, this is the first time two characteristics cross. In the analytic setting we could introduce a shock, calculate its speed and apply the Equal Areas Rule to determine the shape of the wave. Unfortunately, we cannot do so via the chosen numerical scheme. Thus, we conclude that the numerical method considered provides a reasonable solution, provided the wave travels through space up to the breaking time t_m.

Next we proceed by analyzing the rate of change of the numerical solution $u(x, t)$ with respect to space. Figure 7 below illustrates the maximal slope of the wave as a function of t for $\Delta x = 0.01$ in the interval $[0, t_m]$ for four distinct values of m, namely, $m = 2$, $m = 3$, $m = 4$, and $m = 5$, implemented in the MATLAB file `maxSlopeDifferentM.m`.

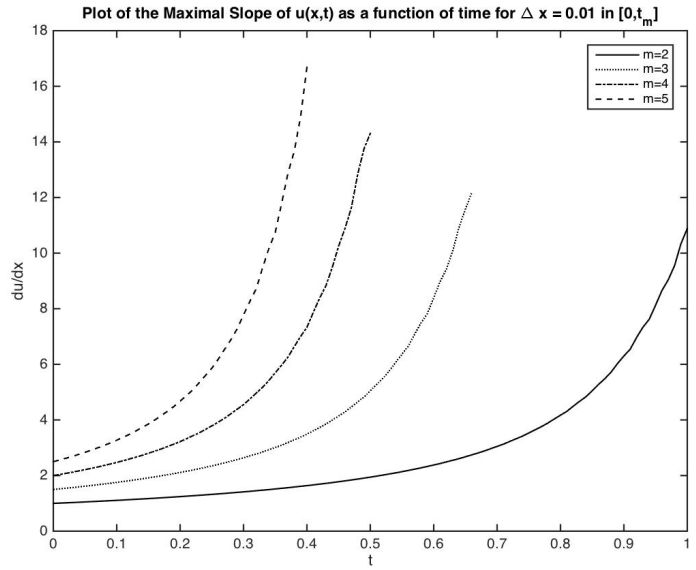

Figure 7: Maximal Slope of $u(x,t)$ as a Function of t for four values of m.

The formula used for the computation of the maximal slope is given by

$$\frac{|dU|}{|dx|} = \frac{\left|U_{j+1}^n - U_j^n\right|}{\Delta x}.$$

Let us compare the plots produced by the smallest and largest value of m. The graph produced when $m = 2$ is given by the solid black curve, and the one for $m = 5$ is the dashed curve. We see that the maximal slope of the numerical solution increases much more slowly for smaller values of m, whereas it becomes steep quite rapidly as we increase the value of m. As stated earlier, increasing the value of the parameter m leads to a smaller number of grid points in the space mesh, as we decrease the interval $x \in [-10/m, 10/m]$. This leads to an earlier shock formation, and the numerical solution evolves much faster to reach the breaking time. Hence, we confirm our previous findings, that given a smaller number of grid points, our numerical approximation reaches the earlier breaking time much faster, i.e. with a more rapidly increasing rate of change with respect to time. Now, theoretically, as we approach t_m, the slope of the wave increases, until it reaches infinity. However, since we are computing the values numerically, we are limited to produce a finite value of the maximal slope of the wave and we cannot compute any numerical values after the multivalued solution occurs.

Finally, we present a list of values for the maximal slope of the wave at the breaking time t_m, given different spatial steps Δx. The values are calculated using the MATLAB file `maxSlopeDeltaX.m`. See Table 1 below.

Δx	Maximal Slope
0.001	37.070
0.002	25.857
0.003	20.733
0.004	17.956
0.005	15.583
0.006	14.018
0.007	12.454
0.008	12.063
0.009	11.572
0.010	10.889
0.011	9.755
0.012	9.733
0.013	8.832
0.014	8.849
0.015	8.116
0.016	8.196
0.017	7.718
0.018	7.572
0.019	7.313
0.020	7.153

Table 1: Maximal Slope of $u(x, t_m)$ for Different Values of Δx and $m = 2$.

Upon inspecting the values presented in Table 1 we see that an increase of the spatial step leads to a reduction of the slope. In other words, if we choose a smaller spatial step and partition the spatial domain in more grid points, we obtain a much faster rate of change of the numerical solution with respect to space. This is attributed to the fact that a larger discretization of the spatial domain, meaning a smaller spatial step, leads to more precise computations of the value of $u(x, t)$ and thus, the wave reaches its breaking time more rapidly. Also as $\Delta x \to 0$, we see that $\max_{t_m} \dfrac{du}{dx} \to \infty$. This confirms our previous statement that as the wave approaches the breaking time t_m its slope tends to increase, until it finally blows up, i.e. becomes infinity in the theoretical sense. However, as before, we conclude that such a result cannot be obtained via the numerical methods used, and thus, indeed we observe the increase of the slope up to a finite value.