Local Oscillations in Finite Difference Solutions of Hyperbolic Conservation Laws

Huazhong Tang

School of Mathematical Sciences
Peking University
Beijing 100871, P.R. China
Website: *dsec.pku.edu.cn/∼tanghz*

Joint work with *Jiequan Li & Lumei Zhang* (Capital Normal University, China)
Gerald Warnecke (Otto-von-Guericke-Universität of Magdeburg, Germany)

Outline

1. Motivation
2. Chequerboard modes in the initial discretization
3. Propagation of the chequerboard mode
4. Discrete Fourier analysis
5. Modified equation analysis
6. Conclusions

Outline

1. Motivation
2. Chequerboard modes in the initial discretization
3. Propagation of the chequerboard mode
4. Discrete Fourier analysis
5. Modified equation analysis
6. Conclusions

Motivation

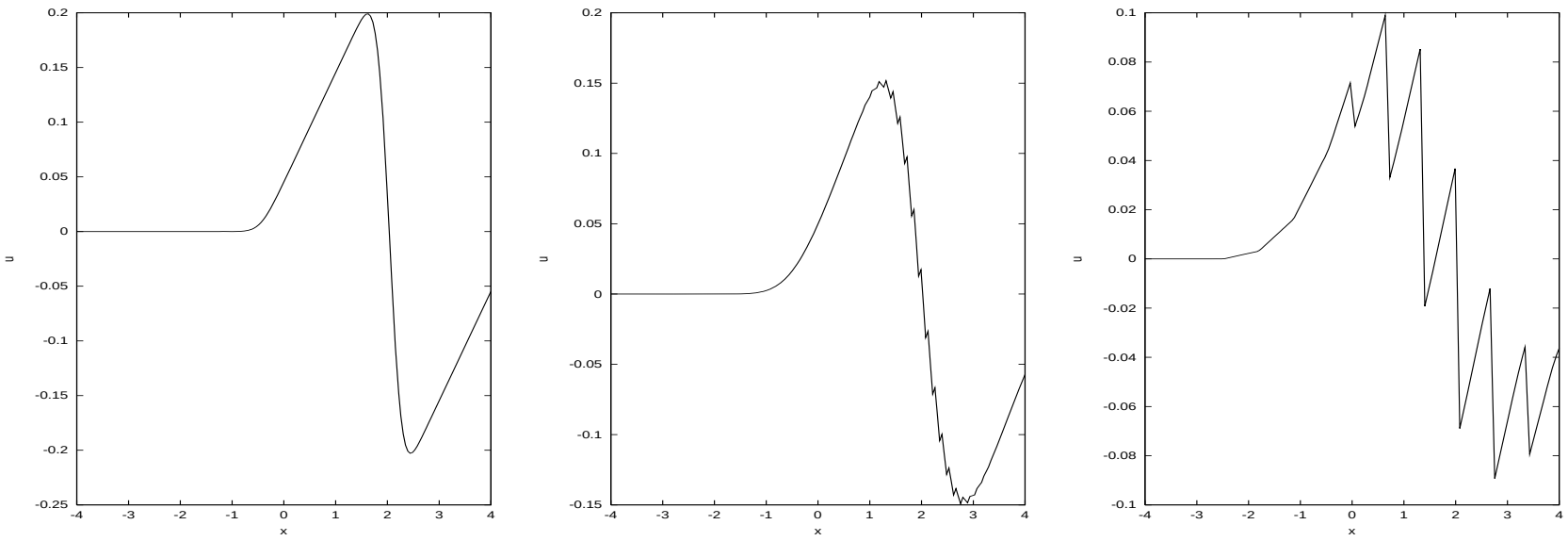

Figure: Numerical solutions of IVP of the linear advection equation $u_t + 0.2u_x = 0$ by MONOTONE schemes [H.Z. Tang & G. Warnecke, M2AN, 38 (2004)].

- **Why does the monotone scheme produce local /counterintuitive oscillations?**

Review of numerical schemes for HCL

- IVP of scalar HCL

$$
\begin{aligned}
u_t + f(u)_x &= 0, \quad x \in \mathbb{R}, \quad t > 0, \\
u(x,t) &= u_0(x), \quad x \in \mathbb{R}, \quad t = 0.
\end{aligned}
$$

- Integrating the above PDE and using the divergence theorem give

$$
\frac{d}{dt}\int_{\mathbb{R}} u \; dx + f|_{x=+\infty} - f|_{x=-\infty} = 0
$$

  which means that the time rate of change of u in the domain is equal to the net flux of u through its boundary.

- HCL plays an important role in gas dynamics etc.

Review of numerical schemes for HCL

- IVP of scalar HCL

$$
\begin{aligned}
u_t + f(u)_x &= 0, \quad x \in \mathbb{R}, \quad t > 0, \\
u(x,t) &= u_0(x), \quad x \in \mathbb{R}, \quad t = 0.
\end{aligned}
$$

- Integrating the above PDE and using the divergence theorem give

$$
\frac{d}{dt} \int_{\mathbb{R}} u \, dx + f|_{x=+\infty} - f|_{x=-\infty} = 0
$$

  which means that the time rate of change of u in the domain is equal to the net flux of u through its boundary.

- HCL plays an important role in gas dynamics etc.

Review of numerical schemes for HCL

- IVP of scalar HCL

$$
\begin{aligned}
u_t + f(u)_x &= 0, \quad x \in \mathbb{R}, \quad t > 0, \\
u(x,t) &= u_0(x), \quad x \in \mathbb{R}, \quad t = 0.
\end{aligned}
$$

- Integrating the above PDE and using the divergence theorem give

$$
\frac{d}{dt} \int_{\mathbb{R}} u \, dx + f|_{x=+\infty} - f|_{x=-\infty} = 0
$$

  which means that the time rate of change of u in the domain is equal to the net flux of u through its boundary.

- HCL plays an important role in gas dynamics etc.

Review of numerical schemes for HCL

- Lax-Wendroff theorem [CPAM, 1960]: If the solutions of the conservative scheme converge to a function u as $\max\{\tau, h\} \to 0$, then u is a weak solution of HCL.
- Conservative scheme in the viscous form

$$\begin{aligned} u_j^{n+1} =& u_j^n - \frac{\nu}{2}\big(f(u_{j+1}^n) - f(u_{j-1}^n)\big) + \frac{Q_{j+1/2}^n}{2}(u_{j+1}^n - u_j^n) \\ & - \frac{Q_{j-1/2}^n}{2}(u_j^n - u_{j-1}^n), \end{aligned}$$

where $\nu = \tau/h$, and τ & h are time and space step sizes, respectively.

Review of numerical schemes for HCL

- Lax-Friedrichs scheme: $Q_{j+1/2} = 1$; Lax-Wendroff scheme: $Q_{j+1/2} = |\nu a_{j+1/2}|^2$; Upwind scheme: $Q_{j+1/2} = |\nu a_{j+1/2}|$.

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

- Generalized LxF scheme: $Q_{j+1/2} = q$ ($0 < q < 1$ constant)
- Monotone scheme: If $u_j^n \geq v_j^n$ for each j, then $u_j^{n+1} \geq v_j^{n+1}$ for each j.
- TVD (Total variation diminishing): $TV(u^{n+1}) \leq TV(u^n) := \sum_j |u_{j+1} - u_j|$.
- Monotonicity-preserving: If $\{u_j^n\}$ is monotone w.r.t. j, so does $\{u_j^{n+1}\}$.

Review of numerical schemes for HCL

- Some well-known facts:

  (i) The monotone scheme is TVD, while the TVD scheme is monotonicity-preserving;

  (ii) The TVD scheme as well as the monotonicity-preserving scheme is (essentially) non-oscillatory;

  (iii) The solution of the monotone scheme converges to unique entropy solution of HCL;

  (iv) The GLF & LxF schemes are monotone under their CFL conditions.

[1] M. G. Crandall & A. Majda, Math. Comp., 34(1980), 1-21.

[2] A. Harten, J. M. Hyman, & P.D. Lax, Comm. Pure Appl. Math. 29(1976), 297-321.

Review of numerical schemes for HCL

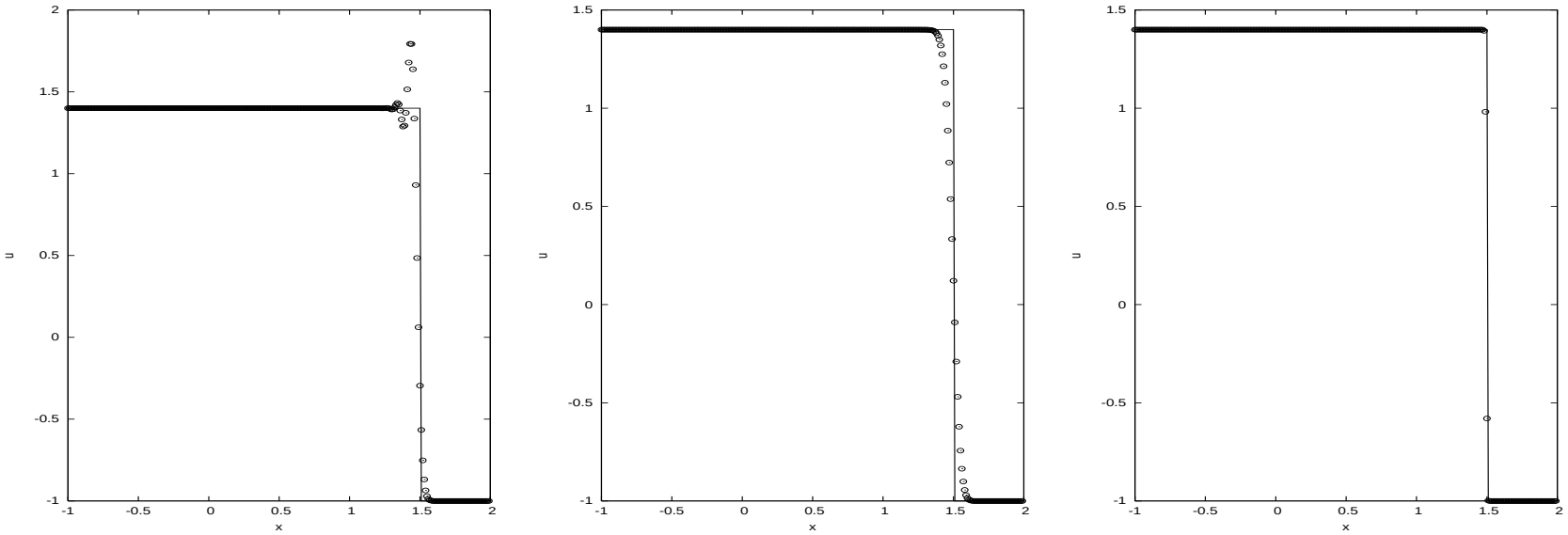

Figure: Numerical results for the linear advection equation $u_t + u_x = 0$. From left to right: 2nd order LW, 1st order upwind, and 2nd order TVD.

- **The LW scheme is 2nd order accurate in time and space, and not monotone & TVD as well as monotonicity-preserving, thus produces numerical oscillations.**

Glimpse of local oscillations in the GLF scheme

Consider the generalized LxF schemes (GLF)

$$u_j^{n+1} = u_j^n - \frac{\nu}{2}\big(f(u_{j+1}^n) - f(u_{j-1}^n)\big) + \frac{q}{2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n), \quad (1)$$

and understand local oscillations in solutions of (1) by Fourier analysis and modified equation methods, even though GLF is monotone and TVD under a certain restriction.

A glimpse: Taking the highest frequency mode, a chequerboard mode, as the initial data, then we have the solution of GLF (1)

$$u_j^n = (1-2q)^n(-1)^j.$$

It shows that the solution of (1) is still a chequerboard mode, except for the modified LxF scheme with $q = 1/2$. The amplitude of the solution is diminishing if $0 < q < 1$ but keeps invariant for the LxF scheme with $q = 1$ or for the unstable central scheme with $q = 0$.

Example 1. Linear Advection Equation

First, we present several examples to display local oscillations in the solution of CL by GLF.

Consider the LAE

$$u_t + u_x = 0, \quad x \in [0, 1],$$

by using the LxF scheme (i.e. $q = 1$). Take the grid points $M = 50$, $\nu = \tau/h = 0.8$, and use periodic B.C. just for simplicity.

We first look at the impulsive initial data

$$u_j^0 = \begin{cases} 1, & j = M/2, \\ 0, & \text{otherwise}, \end{cases}$$

and the distributed square pulse initial data

$$u_j^0 = \begin{cases} 1, & j = M/2, M/2 + 1, \\ 0, & \text{otherwise}. \end{cases}$$

Example 1. Linear Advection Equation

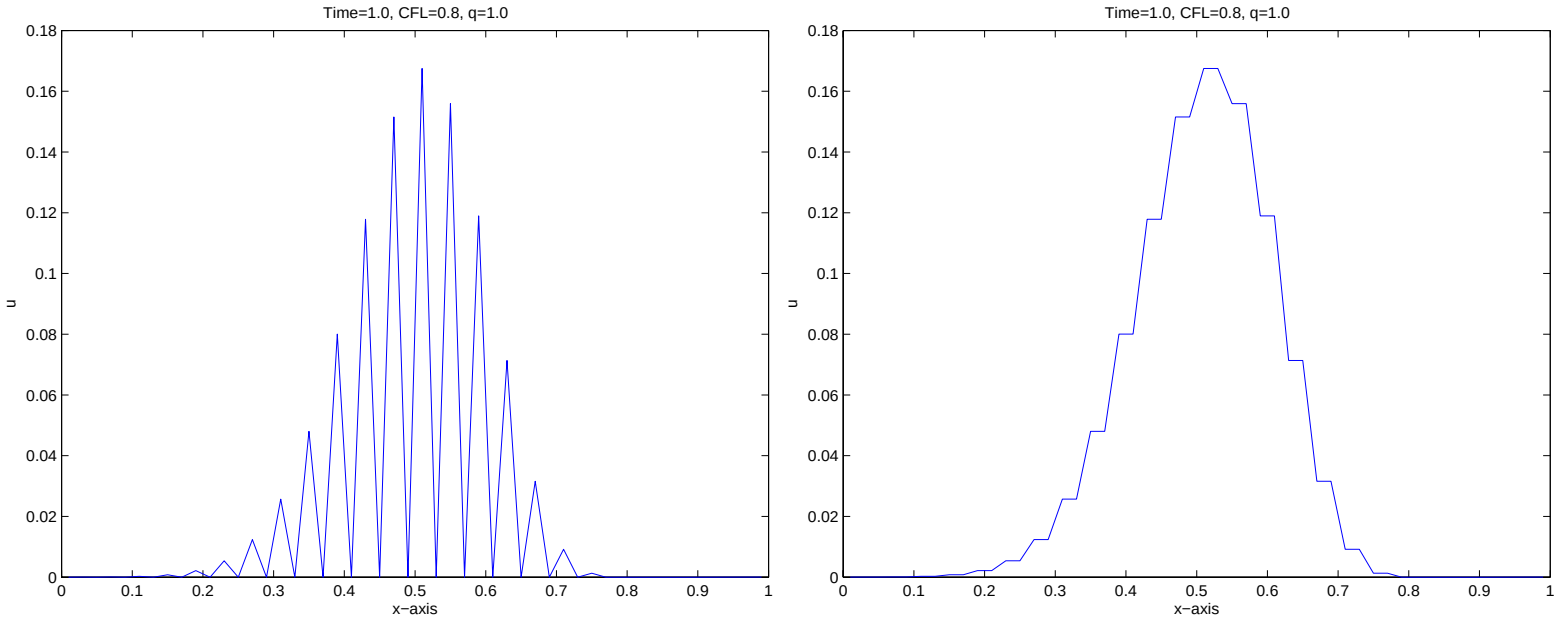

Figure: Numerical results for the advection equation with 50 grid points and $\nu = 0.8$. Left: impulsive data; right: square pulse data.

The left plot displays clear oscillations. Note that the TV keeps invariant 2. The right one shows exactly the opposite behavior (No oscillation is present).

Example 2. Burgers equation $u_t + (u^2/2)_x = 0$

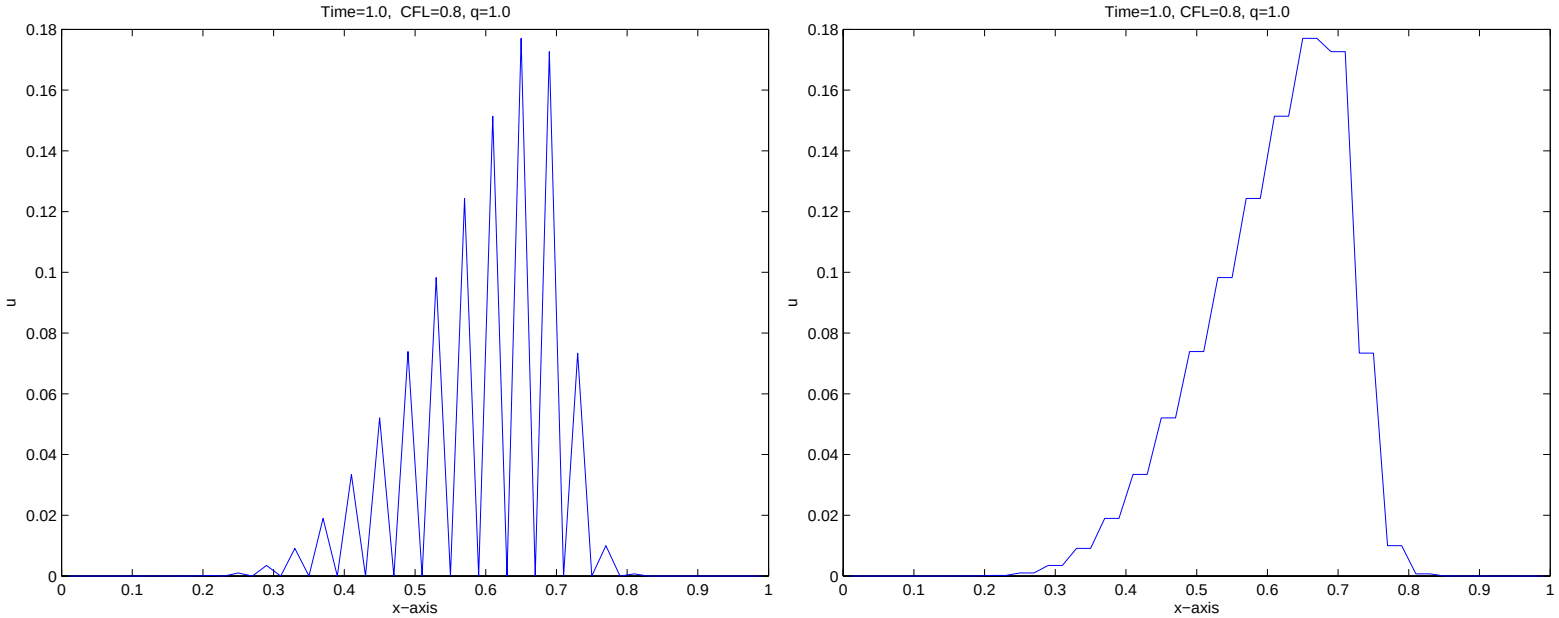

Figure: Same as the last figure except for the Burgers equation. Left: impulsive data; right: square pulse data.

Similar numerical phenomena are observed. Therefore, the oscillations are not connected to the nonlinearity, just for chequerboard mode.

Example 3. Compressible Euler equations

This example is for nonlinear systems

$$\frac{\partial U}{\partial t} + \frac{\partial F(U)}{\partial x} = 0,$$

with $U = (\rho, \rho u, E)^\top$ and $F(U) = (\rho u, \rho u^2 + p, u(E+p))^\top$. We take the equation of state $p = (\gamma - 1)\rho e$ for polytropic gases with $\gamma = 1.4$. Similar to the scalar case, we still use odd and even points to discretize a square-shaped signal initial data such as

$$(\rho_j^0, u_j^0, p_j^0) = \begin{cases} (0.125, 0, 0.1), & j = 49, 50, 51, \\ (1, 0, 1), & \text{otherwise}, \end{cases}$$

and

$$(\rho_j^0, u_j^0, p_j^0) = \begin{cases} (0.125, 0, 0.1), & j = 50, 51, \\ (1, 0, 1), & \text{otherwise}. \end{cases}$$

Example 3. Compressible Euler equations

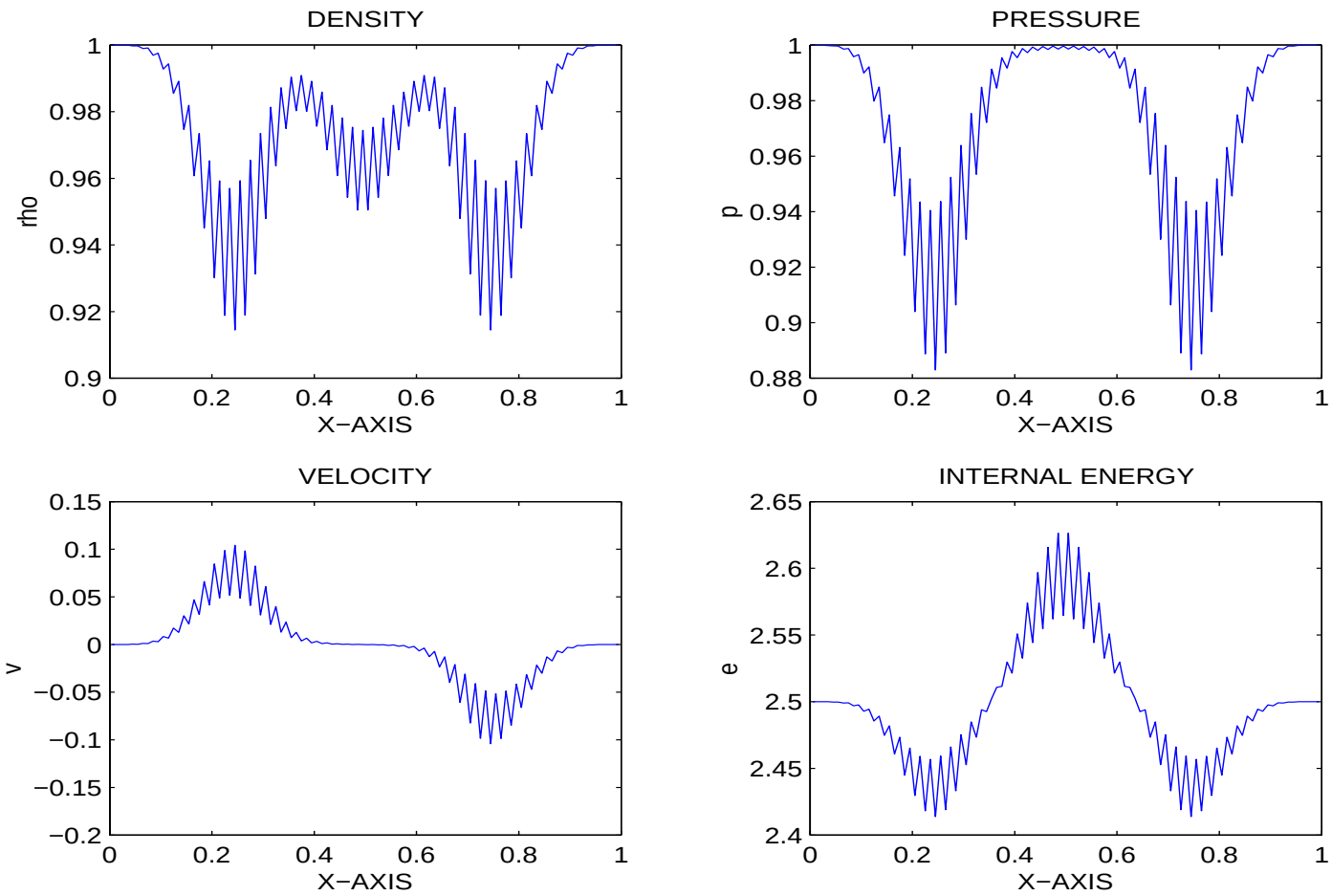

Figure: $t = 0.25$, 100 grid points, $CFL = 0.6$, and $q = 1$. The obvious oscillations are generated by the LxF scheme.

Example 3. Compressible Euler equations

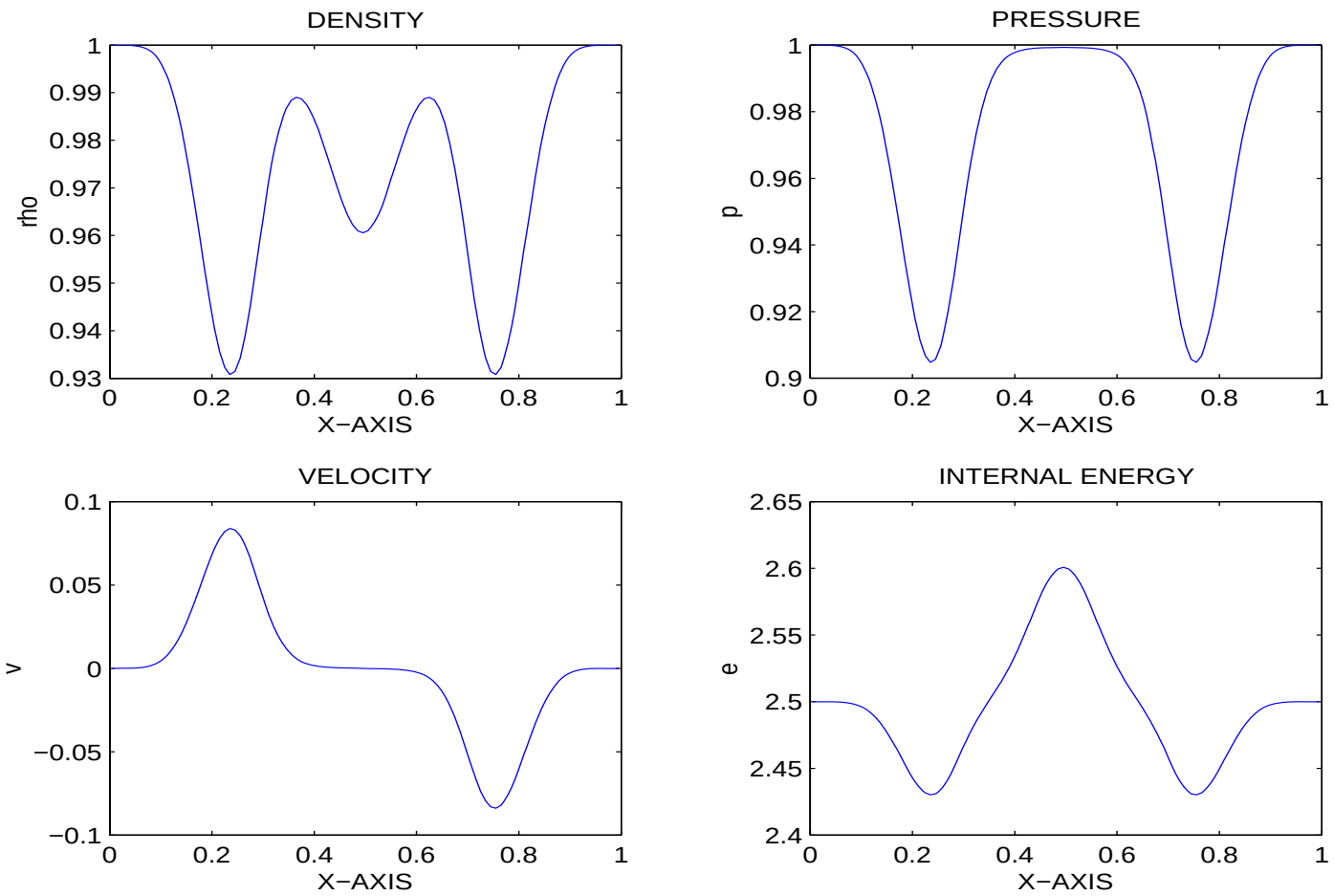

Figure: Same as the last Figure except with $q = 0.9$. The solution is non-oscillatory.

Example 3. Compressible Euler equations

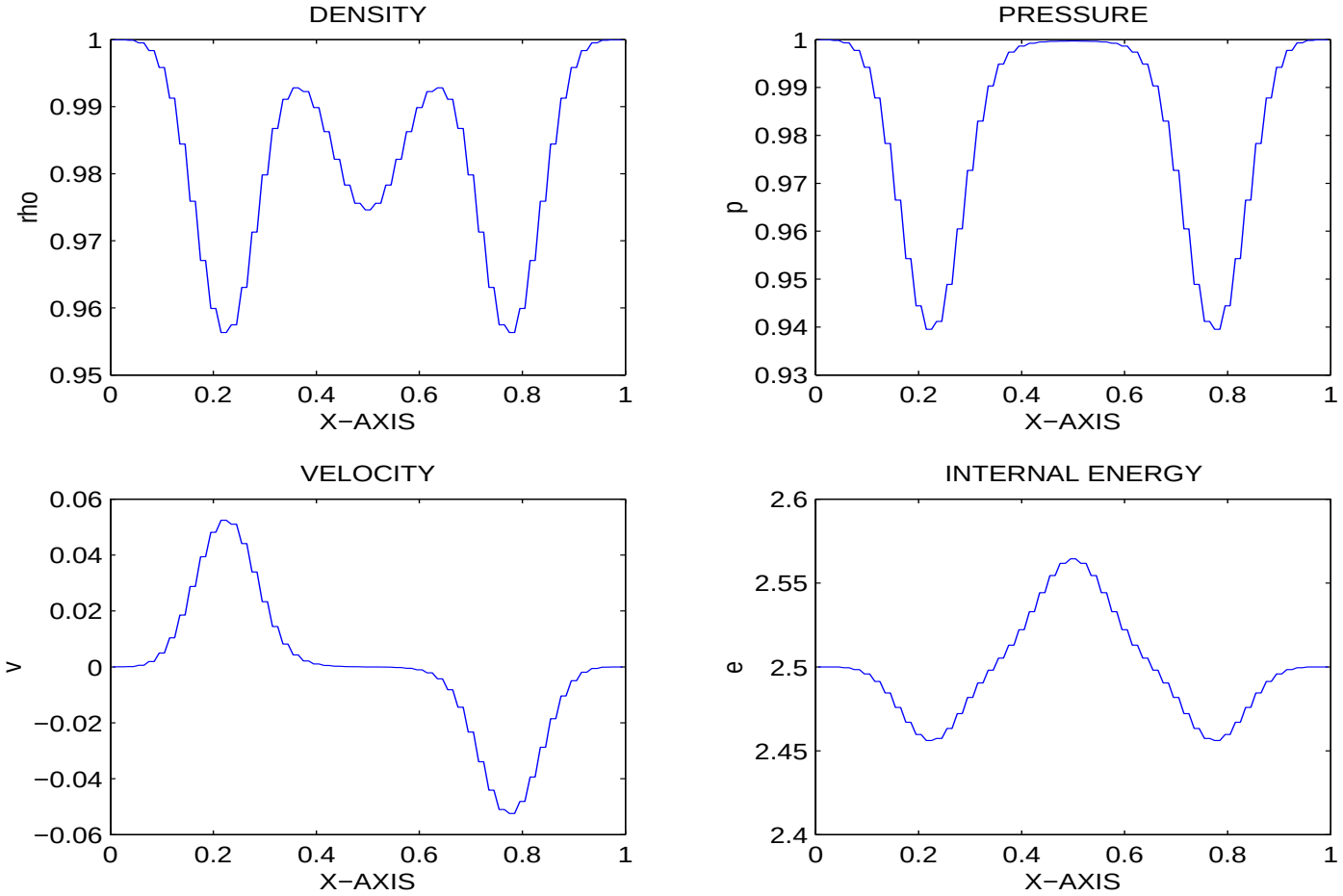

Figure: Same as before except that even points are used to discretize the initial square-shaped signal. $q = 1$. No oscillation is present.

Local oscillations in the solutions by the GLF scheme

Remarks:

- The presence of oscillations is not only related to the ways in which the initial data are approximated, but also to the numerical viscosity coefficient q.
- For the LxF scheme, the local oscillations are very strong although the TV property is not violated.
- Those examples show that the different ways of initial discretization lead to distinct solution behaviors. This motivates the analysis of the discretization of initial data.

Outline

1. Motivation
2. Chequerboard modes in the initial discretization
3. Propagation of the chequerboard mode
4. Discrete Fourier analysis
5. Modified equation analysis
6. Conclusions

Chequerboard modes in the initial discretization

As observed previously and also in [M. Breuss, M2AN, 38(2004), 519-540], *the numerical solutions display very distinct behaviors if the initial data are discretized in different ways*. This motivates us to discuss the discretization of initial data

$$u(x,0) = u_0(x), \quad x \in [0,1], \tag{2}$$

with M grid points and $h = 1/M$. For simplicity, we assume that M is even, the initial function is periodic, i.e. $u_0(0) = u_0(1)$ and its value at the grid point x_j is denoted by u_j^0. We express grid function $\{u_j^0\}$ by using the usual discrete Fourier sums with scaled wave number $\xi = 2\pi kh$

$$u_j^0 = \sum_{k=-M/2+1}^{M/2} c_k^0 e^{i\xi j}, \quad i^2 = -1, \quad j = 0, 1, \cdots, M-1, \tag{3}$$

Chequerboard modes in the initial discretization

where the coefficients c_k^0 are, in turn, expressed as

$$c_k^0 = \frac{1}{M} \sum_{j=0}^{M-1} u_j^0 e^{-i\xi j}, \qquad k = -M/2+1, \cdots, M/2. \tag{4}$$

We first pay special attention to the particular case that

$$c_k^0 = \begin{cases} 1, & \text{if } k = M/2, \\ 0, & \text{otherwise,} \end{cases}$$

i.e. the initial data are taken to be just the highest Fourier mode which is a single chequerboard mode

$$u_j^0 = e^{i2\pi \frac{M}{2} jh} = e^{i\pi j} = (-1)^j.$$

Single square signal case.

We start with the simple initial data of a square signal such as

$$u_0(x) = \begin{cases} 1, & 0 < x^{(1)} < x < x^{(2)} < 1, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

and take the following two ways to approximate the step function (5) as a grid function: One uses an ODD number of grid points to take the value one of the square signal and the other uses the next smaller EVEN number of grid points. They could be seen as approximations to some given fixed interval with end points not represented on the mesh, which we do not explicitly specify here.

Single square signal case.

We use the discrete Fourier sum to clarify their difference.

(i) Discretization with an <u>ODD</u> number of grid points. Take $j_1, j_2 \in \mathbb{N}$ such that $j_1 + j_2$ is an even number. We set $x^{(1)} = (\frac{M}{2} - j_1)h$ and $x^{(2)} = (\frac{M}{2} + j_2)h$. Discretize the square signal (5) with $p := j_1 + j_2 + 1$ nodes, i.e. an odd number of grid points, such that

$$u_j^0 = \begin{cases} 1, & \text{if } \ j = M/2 - j_1, \cdots, M/2 + j_2, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Substituting them into (4), we obtain by simple calculation,

$$c_k^0 = h \sum_{j=0}^{M-1} u_j^0 e^{-i\xi j} = \begin{cases} \frac{(-1)^k e^{i\xi j_1}(1-e^{-i\xi p})}{M(1-e^{-i\xi})}, & \text{for } k \neq 0, \\ ph, & \text{for } k = 0. \end{cases} \tag{7}$$

Single square signal case.

Pay special attention to the term

$$c^0_{M/2} = (-1)^{j_1+M/2} h,$$

since M is even and p is odd. Hence the initial data (6) can be expressed in the form

$$u^0_j = (-1)^{j+j_1+M/2} h + ph + \sum_{k\neq 0, M/2} \frac{(-1)^k e^{i\xi(j+j_1)}(1-e^{-i\xi p})}{M(1-e^{-i\xi})}. \tag{8}$$

It turns out that chequerboard modes are present and will affect the solutions if the initial data contains a square signal and are discretized with an odd number of grid points.

Single square signal case.

(ii) Discretization with an <u>EVEN</u> number of grid points.

Rather than (i) above, we use $p := j_1 + j_2$ even number of grid points to express the square signal in (5) as follows

$$u_j^0 = \begin{cases} 1, & \text{if } j = M/2 - j_1 + 1, \cdots, M/2 + j_2, \\ 0, & \text{otherwise.} \end{cases}$$

Then we substitute these initial data into (4) to obtain

$$c_k^0 = \begin{cases} \frac{(-1)^k e^{i\xi(j_1-1)}[1-e^{-i\xi(p-1)}]}{M(1-e^{-i\xi})}, & \text{for } k \neq 0, \\ (p-1)h, & \text{for } k = 0, \end{cases}$$

and $c_{M/2}^0 = 0$. The initial data can be written by using the discrete Fourier sums as

$$u_j^0 = 0 \times (-1)^j h + (p-1)h + \sum_{k \neq 0, M/2} \frac{(-1)^k e^{i\xi(j+j_1-1)}[1 - e^{-i\xi(p-1)}]}{M(1 - e^{-i\xi})}. \tag{9}$$

Step function initial data case

For more general piecewise constant initial data $u_0(x)$ there are analogous discrete Fourier sum expressions. We divide the computational domain $[0,1]$ into L subintervals I_l $(l = 1, 2, \cdots, L)$, $\bigcup_{l=1}^{L} I_l = [0,1]$, the number of the discrete points of a subinterval I_l is M_l, $M_1 + M_2 + \cdots + M_L = M$, and the initial data (2) are expressed as

$$u_j^0 = \sum_{l=1}^{L} \bar{U}_0^l \cdot \chi_l(j), \tag{10}$$

where $\bar{U}_0^l$ are constants, $\chi_l(j)$ is the characteristic function on I_l

$$\chi_l(j) = \begin{cases} 1, & \text{if } x_j \in I_l, \\ 0, & \text{otherwise .} \end{cases}$$

Note that (10) can be regarded as the superposition of several single square signals of the form (5).

Step function initial data case

Then we express (10) as a discrete Fourier sum of the form (3) with c_k^0. For $k \neq 0,\ M/2$, we have

$$
\begin{aligned}
c_k^0 =& \frac{1}{M}\left[\bar{U}_0^1 \sum_{j=0}^{M_1-1} e^{-i\xi j} + \bar{U}_0^2 \sum_{j=M_1}^{M_1+M_2-1} e^{-i\xi j} + \cdots + \bar{U}_0^L \sum_{j=p_{L-1}}^{M-1} e^{-i\xi j}\right] \\
=& \frac{1}{M(1-e^{-i\xi})} \sum_{l=1}^{L} \bar{U}_0^l (e^{i\xi M_l} - 1) e^{-i\xi p_l}, \qquad (11)
\end{aligned}
$$

where $p_l = M_1 + \cdots + M_l$; and for $k = 0,\ M/2$, we have

$$
c_0^0 = \tfrac{1}{M}(\bar{U}_0^1 M_1 + \bar{U}_0^2 M_2 + \cdots + \bar{U}_0^L M_L),
$$

$$
c_{M/2}^0 = \tfrac{1}{M} \sum_{l=1}^{L} \bar{U}_0^l \left[\sum_{j=0}^{M-1} \chi_l(j)(-1)^j\right].
$$

Chequerboard modes in the initial discretization

Thus, the initial data are expressed as

$$u_j^0 = \frac{1}{M}\sum_{l=1}^{L}\bar{U}_0^l\left[\sum_{m=0}^{M-1}\chi_l(m)(-1)^m\right](-1)^j + \frac{1}{M}(\bar{U}_0^1 M_1 + \bar{U}_0^2 M_2 + \cdots$$
$$+ \bar{U}_0^L M_L) + \frac{1}{M}\sum_{k\neq 0, M/2}\frac{1}{(1-e^{-i\xi})}\sum_{l=1}^{L}\bar{U}_0^l(e^{i\xi M_l}-1)e^{-i\xi p_l}. \quad (12)$$

Similar to the case of a single square signal, it depends on $c_{M/2}^0$ whether there is a chequerboard mode in the discrete initial data. Therefore, we have three cases here too.

(i) If the gridpoint number M_l in each I_l is odd, $c_{M/2}^0$ is

$$c_{M/2}^0 = \frac{1}{M}\sum_{l=1}^{L}\bar{U}_0^l\left[\sum_{j=0}^{M-1}\chi_l(j)(-1)^j\right] = \frac{1}{M}\sum_{l=1}^{L}\bar{U}_0^l(-1)^{l+1};$$

Chequerboard modes in the initial discretization

(ii) If the gridpoint number M_l in each I_l is even, $c^0_{M/2}$ vanishes,

$$c^0_{M/2} = \frac{1}{M} \sum_{l=1}^{L} \bar{U}^l_0 \left[\sum_{j=0}^{M-1} \chi_l(j)(-1)^j \right] = 0.$$

(iii) If the gridpoint number in some I_l is odd while in the others it is even, $c^0_{M/2} = \frac{1}{M} \sum_{l=1}^{L} \bar{U}^l_0 \phi(l)$, where

$$\phi(l) = \begin{cases} 0, & \text{if } I_l \text{ is in the even case,} \\ 1 \text{ or } (-1), & \text{if } I_l \text{ is in the odd case.} \end{cases}$$

Chequerboard modes in the initial discretization

Thus, there is no chequerboard mode for Case (ii). For Case (i), the summation may be zero when the factors cancel. However, since this summation is taken in the global sense and the chequerboard mode exists in each subinterval, the solution may still contain oscillations due to the finite propagation speed property of the scheme.

Chequerboard modes in the initial discretization

We summarize all of the above analysis as follows.

Proposition

Suppose that the initial data

$$u(x,0) = u_0(x)$$

are given as a step function. We can approximate them as the superposition of several single square signals. For each square signal we have two different types of discretizations. If they are approximated with an ODD number of grid points, the chequerboard (i.e. highest frequency) mode is present. In contrast, if they are discretized with an EVEN number of grid points, there is no chequerboard mode.

Outline

1. Motivation
2. Chequerboard modes in the initial discretization
3. Propagation of the chequerboard mode
4. Discrete Fourier analysis
5. Modified equation analysis
6. Conclusions

Propagation of the chequerboard mode propagation

In this part we simply look at the GLF scheme for the linear advection equation, $f(u) = au$,

$$u_j^{n+1} = u_j^n - \frac{\nu a}{2}(u_{j+1}^n - u_{j-1}^n) + \frac{q}{2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n), \quad (13)$$

and catch a glimpse of the resolution of high frequency modes, where $|\nu a| \leq q \leq 1$.

As usual for stability analysis, the solution to (13) is expressed analogously to (3) in the standard form of a discrete Fourier sum using $\xi = 2\pi kh$

$$u_j^n = \sum_{k=-M/2+1}^{M/2} c_k^n e^{i\xi j}. \quad (14)$$

The coefficients c_k^n are obtained successively and expressed as,

$$c_k^n = (1 + q(\cos\xi - 1) - i\nu a \sin\xi)^n \, c_k^0. \quad (15)$$

Corresponding to the two kinds of discretization of a single square signal, c_j^0 have different expressions, and the solutions become:

(i) Odd discretization case. With the initial data (8), the solution of (13) is

$$u_j^n = \frac{1}{M}(1-2q)^n(-1)^{j+j_1+M/2} + \sum_{k=-M/2+1}^{M/2-1} c_k^n e^{i\xi j}. \tag{16}$$

(ii) Even discretization case. With the initial data (9), we have

$$u_j^n = 0 \times (1-2q)^n(-1)^j + \sum_{k=-M/2+1}^{M/2-1} c_k^n e^{i\xi j}. \tag{17}$$

Comparing them, we see that in the odd case the chequerboard mode does not vanish if it exists initially, unless we have the MLF scheme $q = 1/2$, although it decays with a rate of $|2q-1|$ at each time step. So, a proper discretization of initial data would be important to avoid these possible oscillations in solving CL.

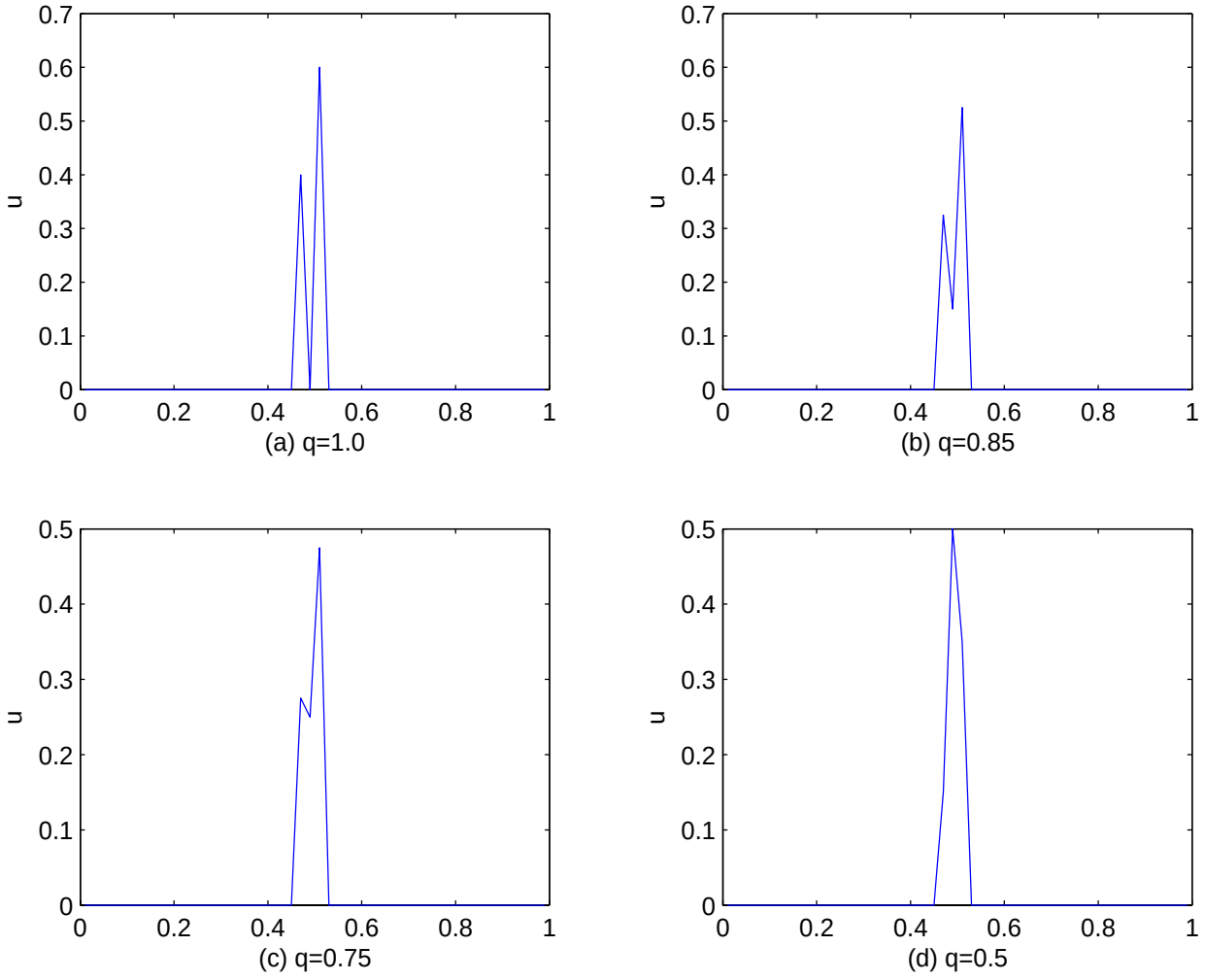

Figure: The decay of oscillations as the parameter q decreases: The initial data are impulse signal, the CFL number is 0.2 and only one time step is taken.

Remark 1. The solution of LxF $u_j^n = (-1)^{n+j}$ oscillates between 1 and -1 alternately if the chequerboard mode initial data are taken. The large numerical dissipation does not have any effect.

Remark 2. In case $0 < q < 1$, the chequerboard mode is damped out quickly. In particular, for the MLF $q = 1/2$ the chequerboard mode is eliminated and has no influence on the solution at all.

In the following we attempt to analyze the numerical dissipation and phase error mechanisms of GLF, particularly on high frequency modes and explain the phenomenon of oscillations caused by high frequency modes. We apply discrete Fourier analysis and the method of modified equation analysis. Both of the methods give complimentary results, which are consistent. We show that as $0 \leq q \leq 1$ is away from $1/2$, the damping on high frequency modes becomes weak. In particular, there is no damping effect in the LxF scheme (i.e. $q = 1$) and the unstable central scheme (i.e. $q = 0$).

Outline

1. Motivation
2. Chequerboard modes in the initial discretization
3. Propagation of the chequerboard mode
4. Discrete Fourier analysis
5. Modified equation analysis
6. Conclusions

Discrete Fourier analysis

We use the discrete Fourier analysis to discuss the dissipation and phase error mechanism of GLF for $f = au$, which is monotone under $0 < |\nu a| < q \leq 1$. We are particularly concerned with the phase accuracy of Fourier modes. Denote a Fourier mode by $e^{i\xi}$ with $\xi = 2\pi kh$. Then using it as initial data for a linear FDS results in the solution at $t = n\tau$

$$u_k^n = \lambda_k^n e^{i\xi} = (\lambda(k))^n \, e^{i\xi}, \quad i^2 = -1, \tag{18}$$

where λ_k^n is the amplitude. The ratio $\lambda(k) = \lambda_k^{n+1}/\lambda_k^n$ is the *amplitude* of the mode for one time step.

Discrete Fourier analysis

For GLF we have in particular

$$\lambda(k) = 1 + q(\cos\xi - 1) - i\nu a \sin\xi, \qquad \nu = \tau/h, \tag{19}$$

and its modulus

$$|\lambda(k)|^2 = 1 + 4(a^2\nu^2 - q)\sin^2(\xi/2) + 4(q^2 - a^2\nu^2)\sin^4(\xi/2). \tag{20}$$

For $q = \nu|a|$ the last term at RHS vanishes, whereas for $q = \nu^2 a^2$ the 2nd term vanishes.

Discrete Fourier analysis

Also we see that for GLF, under the conditions $0 < \nu^2 a^2 \leq q \leq 1$, we have from (20) the estimate of the modulus of *amplitude*

$$|\lambda(k)|^2 = 1+4(a^2\nu^2-q)\left(\sin^2(\xi/2) - \sin^4(\xi/2)\right)+4q(q-1)\sin^4(\xi/2) \leq 1. \tag{21}$$

Thus these conditions imply that the schemes are linearly stable.

In fact, these conditions are necessary and sufficient for stability.

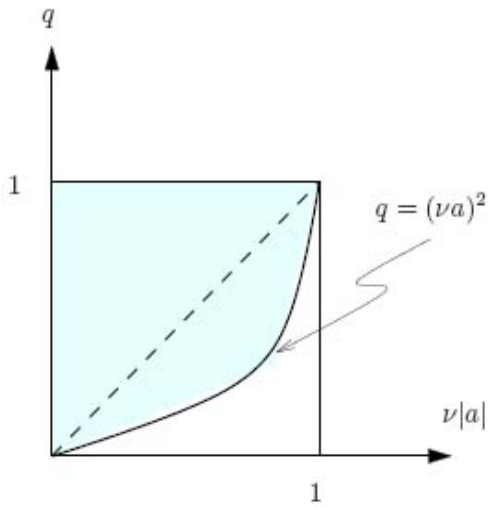

Figure: Range of stability of parameter q over CFL number νa.

Discrete Fourier analysis

The exact solution of the Fourier mode $e^{i\xi}$ for $x = h$ after one time step τ is $e^{i(\xi - 2\pi a k \tau)} = e^{-i2\pi a k \tau} e^{i\xi} = \lambda_{exact}(k) e^{i\xi}$. The exact amplitude $\lambda_{exact}(k)$ has modulus 1. We see from (20) that the *amplification error*, i.e. the error in amplitude modulus, is of order $\mathcal{O}(\xi)$ for the monotone schemes and order $\mathcal{O}(\xi^2)$ for the LW scheme. If the modulus of $\lambda(k)$ is less than one, the effect of the multiplication of a solution component with $\lambda(k)$ is called *numerical dissipation* and then the amplification error is called *dissipation error*. If the modulus is larger than 1, this leads to the amplification of the Fourier mode, i.e. instability of any solution containing it. Further, comparing the exponents of $\lambda(k)$ and $\lambda_{exact}(k)$ there is a phase error $\arg \lambda(k) - (-2\pi a k \tau)$.

Discrete Fourier analysis

The relative phase error is then defined as

$$E_p(k) := \frac{\arg \lambda(k)}{-2\pi a k \tau} - 1 = -\frac{\arg \lambda(k)}{\nu a \xi} - 1.$$

A mode is a low frequency mode if $\xi \approx 0$ and a high frequency mode if $\xi \approx \pi$.

We first look at the low frequency modes ($\xi \approx 0$)

$$(U^s)_j^n := \lambda_k^n e^{i\xi j}.$$

For $k = \xi = 0$ we have $\lambda(k) = 1$. From (20) we obtain

$$\frac{d(|\lambda(k)|^2)}{dq} = 2(1 + q(\cos\xi - 1))(\cos\xi - 1) < 0, \qquad (22)$$

for fixed $\xi \in]0, \pi/2]$. This implies that the dissipation becomes weaker as q decreases. The LxF scheme with $q = 1$ has the largest numerical dissipation for low frequency modes.

Discrete Fourier analysis

The phase of the low frequency modes is approximated by Taylor expansion at $\xi = 0$

$$\arg\lambda = \arctan\left(\frac{-\nu a \sin\xi}{1 + q[\cos\xi - 1]}\right) \approx -\nu a \xi \left(1 + \frac{3q - 1 - 2\nu^2 a^2}{6}\xi^2 + \cdots\right).$$

This phase has a relative error $E_p(k)$ of order $\mathcal{O}(\xi^2)$. For the LW scheme, this phase error causes oscillations, which cannot be suppressed by the weaker dissipation of order $\mathcal{O}(\xi^2)$, compared to the dissipation error $\mathcal{O}(\xi)$ of the upwind scheme.

Discrete Fourier analysis

For high frequency modes (18), $\xi \approx \pi$, the situation is very different. We introduce the decomposition $\xi = \pi + \xi'$, i.e. $\xi' = 2\pi k' h$ with $kh = 1/2 + k'h$, and thus $\xi' \approx 0$. We write the modes in the form

$$(U^h)_j^n = \lambda_k^n e^{i\xi j} = \lambda_k^n e^{i(\pi+\xi')j} = (-1)^{j+n} \lambda_{k'}^n e^{i\xi' j}, \qquad (23)$$

with $\lambda_{k'}^n = (-1)^{j+n} e^{i\pi j} \lambda_k^n$ and set

$$(U^o)_j^n := \lambda_{k'}^n e^{i\xi' j}.$$

The factor $(U^o)_j^n$ can be regarded as a perturbation amplitude of the chequerboard modes $(e^{i\pi})^{j+n} = (-1)^{j+n}$. The dissipation (amplitude error) depends only on $\lambda_{k'}^n$. Then substituting $(U^h)_j^n$ into (13) yields

$$\lambda' := \lambda_{k'}^{n+1} / \lambda_{k'}^n = -1 + q(1 + \cos \xi') - i\nu a \sin \xi'.$$

Therefore, we have

$$
\begin{aligned}
|\lambda'|^2 =& (1 - q(1 + \cos\xi'))^2 + \nu^2 a^2 \sin^2 \xi' \\
=& 1 + 4(a^2\nu^2 - q)\cos^2(\xi'/2) + 4(q^2 - a^2\nu^2)\cos^4(\xi'/2). \qquad (24)
\end{aligned}
$$

It is consistent with a shift of π in the variable ξ in (20). Regarding the high frequency modes, for all schemes the amplitude error is $\mathcal{O}(1)$. At $\xi' = 0$ we have

$$
|\lambda'|^2 = 1 - 4q(1 - q), \qquad (25)
$$

so we have the lowest amplitude error for $q = 1$ or near zero, the highest for the modified LxF scheme with $q = 1/2$. Obviously, for small ξ', $|\lambda'|^2$ is an increasing function of q if $q > 1/2$ because

$$
d(|\lambda'|^2)/dq = -2[1 - q(1 + \cos(\xi'))](1 + \cos\xi') > 0. \qquad (26)
$$

That is, GLF becomes much more dissipative for high frequency modes as the parameter q decreases, which is in sharp contrast with the situation for low frequency modes, see (22).

Furthermore, let us look at the relative phase error. We compute

$$\begin{aligned}\arg\lambda' =& \tan^{-1}\left(\frac{-\nu a\sin\xi'}{-1+q(1+\cos\xi')}\right)\\ =&\frac{-\nu a\xi'}{2q-1}-\frac{\nu a}{3(2q-1)^2}\left[\frac{q+1}{2}-\frac{\nu^2a^2}{2q-1}\right]\xi'^3+\mathcal{O}(\xi'^5). \qquad (27)\end{aligned}$$

Then for the high frequency modes $(U^h)_j^n$, we have by recalling that $\xi=\pi+\xi'$

$$\begin{aligned}(U^h)_j^n =&(-1)^{j+n}\lambda_k'^n e^{i\xi' j}\\ =&|\lambda'|^n e^{i(j\xi-2\pi kan\tau)}\cdot e^{in(-\pi+\arg\lambda'+\nu a\xi)}.\end{aligned}$$

So, the relative phase error of high frequency modes after each time step is

$$
\begin{aligned}
E_p(k) &= -\frac{-\pi + \arg\lambda' + \nu a\xi}{\nu a\xi} \\
&= -\frac{\pi(1-\nu a)}{\nu a\xi} + \frac{2(q-1)\nu a\xi'}{(2q-1)\nu a\xi} - \frac{1}{3(2q-1)^2\xi}\left[\frac{q+1}{2} - \frac{\nu^2 a^2}{2q-1}\right]\xi'^3 \\
&\quad + \mathcal{O}(\xi'^5).
\end{aligned} \tag{28}
$$

Because $\xi \approx \pi$, the relative phase error has $\mathcal{O}(1)$. This error is huge, and strong numerical dissipation is needed to suppress it.

We summarize the above Fourier analysis in the proposition.

Proposition

We distinguish low and high frequency Fourier modes $u_j^n = \lambda_k^n e^{ij\xi}$, $\xi = 2\pi kh$, *and they behave differently.*
(i) For the low frequency modes ($\xi \sim 0$), *the relative phase error is of order* $\mathcal{O}(\xi^2)$, *and the amplitude error (dissipation) becomes smaller as the parameter* q *decreases. The order of amplitude error is* $\mathcal{O}(\xi)$ *for the monotone schemes and* $\mathcal{O}(\xi^2)$ *for the LW scheme.*
(ii) For the high frequency modes ($\xi \sim \pi$), *the relative phase error is of order* $\mathcal{O}(1)$, *the amplitude error becomes larger as the parameter* q *is closer to* $1/2$.

Outline

1. Motivation
2. Chequerboard modes in the initial discretization
3. Propagation of the chequerboard mode
4. Discrete Fourier analysis
5. Modified equation analysis
6. Conclusions

Modified equation analysis for linear cases

As we know, the amplitude error and relative phase error of the Fourier modes have a correspondence with dissipation and phase error mechanisms displayed by related PDEs. Here we use the modified equation to further investigate the mechanisms of dissipation and phase error of GLF. Particularly, we want to see how the dissipation offsets the large phase error of high frequency modes. The modified equation analysis was originally introduced for low frequency modes[1]. Here it is especially used for high frequency modes, as its usefulness was clearly shown in [K.W. Morton and D.F. Mayers, Numerical Solution of Partial Differential Equations, Cambridge Univ Press, 2005].

The modified equation is derived by first expanding each term of a difference scheme in a Taylor series and then eliminating time derivatives higher than first order by certain algebraic manipulations.

Modified equation analysis for linear cases

We begin here with the linear case, and still use notation $\nu = \tau/h$. As in [Morton & Mayers, 2005, P173], we will consider a smooth solution $(U^s)_j^n$ and an oscillatory solution $(U^h)_j^n$, respectively. The oscillatory solution $(U^h)_j^n$ is written as

$$(U^h)_j^n = (-1)^{j+n}(U^o)_j^n, \tag{29}$$

where $(U^o)_j^n$ is viewed as the perturbation amplitude of the chequerboard mode.

Modified equation analysis for linear cases

The smooth solution $(U^s)_j^n$ satisfies GLF, i.e.

$$(U^s)_j^{n+1} = (U^s)_j^n - \frac{\nu a}{2}((U^s)_{j+1}^n - (U^s)_{j-1}^n) + \frac{q}{2}((U^s)_{j+1}^n - 2(U^s)_j^n + (U^s)_{j-1}^n).$$

Then we derive a modified equation for this solution[2], and the notation $\widetilde{U}^s$ corresponds to the associated exact solution

$$\partial_t \widetilde{U}^s + a\partial_x \widetilde{U}^s = \frac{1}{2\tau}\left(qh^2 - a^2\tau^2\right)\partial_x^2 \widetilde{U}^s + a\left(-\frac{h^2}{6} + \frac{1}{2}qh^2 - \frac{1}{3}a^2\tau^2\right)\partial_x^3 \widetilde{U}^s + \cdots$$

It is evident that the numerical viscosity of GLF becomes stronger for low frequency modes as q is larger, and vice versa. Particularly, for the LW scheme the dissipation comes from the 4th order term and therefore is quite weak. This is consistent with the fact observed by the Fourier analysis that the dissipation becomes weaker as q decreases, provided that the scheme is stable.

[2] K.W. Morton & D.F. Mayers, CUP, 2005, p169

Modified equation analysis for linear cases

However, the numerical dissipation of GLF is very different for the high frequency modes. Let us discuss the perturbation $(U^o)_j^n$ of the oscillatory solution (29). Substituting (29) into GLF yields

$$\frac{(U^o)_j^{n+1}-(U^o)_j^n}{\tau}=\frac{q-2}{\tau}(U^o)_j^n-\frac{\nu a}{2\tau}[(U^o)_{j+1}^n-(U^o)_{j-1}^n]+\frac{q}{2\tau}[(U^o)_{j+1}^n+(U^o)_{j-1}^n]. \tag{30}$$

Compared to the difference equation for the low frequency modes, it contains an extra term $\frac{q-2}{\tau}(U^o)_j^n$, which plays the key role of damping on high frequency modes. We use the notation $\tilde{U}^o(jh,n\tau)$ to express $(U^o)_j^n$ inserted into the above equation and apply the standard approach. That is, taking the standard Taylor expansion yields

$$\mathcal{D}_{+t}\widetilde{U}^o+a\partial_x\widetilde{U}^o=\frac{2(q-1)}{\tau}\widetilde{U}^o+\frac{a^2q\tau}{2\nu^2}\partial_x^2\widetilde{U}^o-\frac{a^3\tau^2}{6\nu^2}\partial_x^3\widetilde{U}^o+\cdots, \tag{31}$$

where $D_{+t}=(e^{\tau\partial_t}-1)/\tau$.

Note that in (30) the term $\frac{q-2}{\tau}(\widetilde{U}^o)_j^n$ is unusual compared to classical modified equation analysis.

We may derive the modified equation for the oscillatory part

$$\partial_t \widetilde{U}^o + \frac{a}{2q-1}\partial_x \widetilde{U}^o = \frac{\ln|2q-1|}{\tau}\widetilde{U}^o + \frac{h^2}{2\tau}\frac{[q(2q-1)-\nu^2a^2]}{(2q-1)^2}\partial_x^2\widetilde{U}^o + \frac{ah^2}{6}\frac{[(q+1)(2q-1)-2\nu^2a^2]}{(2q-1)^3}\partial_x^3\widetilde{U}^o + \cdots .$$

Introducing for $q \neq 1/2$ a rescaling $x' = x(2q-1)$ and omitting the use of a primed variable gives

$$\partial_t \widetilde{U}^o + a\partial_x \widetilde{U}^o = \frac{\ln|2q-1|}{\tau}\widetilde{U}^o + \frac{h^2}{2\tau}[q(2q-1)-\nu^2a^2]\partial_x^2\widetilde{U}^o + \frac{ah^2}{6}[(q+1)(2q-1)-2\nu^2a^2]\partial_x^3\widetilde{U}^o + \cdots .$$

Unlike the modified equation (52) for the low frequency modes the numerical dissipation comes from two terms: Zero order term $\frac{\ln|2q-1|}{\tau}\widetilde{U}^o$ and the 2nd order term $\frac{a^2\tau}{2\nu^2a^2}[q(2q-1)-\nu^2a^2]\partial_x^2\widetilde{U}^o$. The former exerts more dominant dissipation than the latter.

Modified equation analysis for linear cases

The zero order term in the modified equation $\frac{\ln|2q-1|}{\tau}\widetilde{U}^o$ is called *a numerical damping* term and the 2nd order term $\frac{h^2}{2\tau}[q(2q-1)-\nu^2a^2]\partial_x^2\widetilde{U}^o$ a *numerical viscosity*. They play different dissipation roles in controlling the amplitude of high frequency modes.

Remarks: The modified equation of LxF for $\widetilde{U}^o$ is

$$\partial_t\widetilde{U}^o+a\partial_x\widetilde{U}^o=\frac{a^2\tau}{2\nu^2a^2}[1-\nu^2a^2]\partial_x^2\widetilde{U}^o+\frac{a^3\tau^2}{3\nu^2a^2}[1-\nu^2a^2]\partial_x^3\widetilde{U}^o+\cdots. \tag{32}$$

Although this part is dissipated through the numerical viscosity term if $|\nu a|<1$, this dissipation is still weak in comparison with the numerical damping term $\frac{\ln(|2q-1|)}{\tau}\widetilde{U}^o$. Thus the chequerboard mode is not perturbed and damped at all. This explains why the oscillations in the LxF scheme are observed. This was already highlighted above through the discrete Fourier analysis.

Modified equation analysis for linear cases

Remarks: As $0 < q < 1$, the strong damping term $\frac{\ln(|2q-1|)}{\tau}\widetilde{U}^o$ suppresses the oscillations well no matter how the viscosity term behaves. In particular, if $q = 1/2$, the oscillation is damped out immediately, by noting that

$$\lim_{q \to 1/2+0} \ln(|2q-1|) = -\infty. \tag{33}$$

So the damping becomes infinite for $q = 1/2$. This is consistent with two previous observations.

The high frequency modes are dissipated very quickly if $0 < q < 1$, unlike the LxF scheme with $q = 1$ or the unstable central scheme with $q = 0$, even though there is a chequerboard mode initially. This explains why the oscillations are less visible for $0 < q < 1$ than those in the LxF or the unstable central scheme for $q = 0$.

Outline

1. Motivation
2. Chequerboard modes in the initial discretization
3. Propagation of the chequerboard mode
4. Discrete Fourier analysis
5. Modified equation analysis
6. Conclusions

Conclusions

- The talk discussed the local oscillations in the GLF scheme. It has general implications into more extensive (monotone) schemes and multidimensional cases.
- The discrete Fourier analysis and the modified equation are applied to investigating the numerical dissipative and dispersive mechanisms as well as relative phase errors.
- The resolutions of the low and high frequency modes $u_j^n = \lambda_k^n e^{ij\xi}$, $\xi = 2\pi kh$ in numerical solutions are individually discussed.
- The presence of high frequency modes results from the initial/boundary conditions. The discretization may produce the chequerboard modes.

Conclusions

We summarize our results as follows.

- **Relative phase error**. For the low frequency modes, the error is of order $\mathcal{O}(\xi^2)$, while for high frequency modes the error is of order $\mathcal{O}(1)$ after each time step, which is generally independent of the parameter q.
- **Numerical dissipation**. For the low frequency modes, the dissipation is usually of order $\mathcal{O}(\xi)$ for the scheme GLF, which closely depends on the parameter q. As $q = \nu^2 a^2$, GLF becomes the LW scheme and it has the amplitude error $\mathcal{O}(\xi^2)$. For high frequency modes, the scheme usually has the numerical damping of order $\mathcal{O}(1)$ that becomes stronger as q is closer to $1/2$, unless it vanishes for the limit case ($q = 1$ or 0), in which the amplitude is dissipated via the numerical viscosity of 2nd order.

Conclusions

- In the LW scheme the oscillations are caused by the relative phase error of low frequency modes, while in LxF, oscillations are caused by the relative phase error of high frequency modes. To control the oscillations by high frequency modes, the strong numerical damping is necessary to add.
- Compared to LxF, the GLF $(0 < q < 1)$ introduces the numerical damping as well, which is stronger as q is closer to $1/2$. Hence, to control the oscillations caused by high frequency modes, *the numerical damping plays an important role*.

Conclusions

- Indeed, the GLF scheme is monotone and thus TVD under a certain restriction. The TVD property is proposed to describe the global property of solutions of CLs. The oscillations we are investigating is *local* and does not contradict to this *global* TVD property. As far as hyperbolic problems are concerned, local properties should be paid more attention because of finite propagation of waves.

Thank you for your attention!

J.Q. Li, H.Z. Tang, G. Warnecke,& L.M. Zhang, Math. Comp., 78(2009), 1997-2018.