Understanding of Numerical Oscillations in terms of Modified Equation

Jiequan Li ()

School of Mathematical Sciences, Beijing Normal University

Email:jiequan@bnu.edu.cn

Homepage: http://math.bnu.edu.cn/∼lijiequan/

2011 Int'l Workshop on Numerical Methods and Modeling
for Compressible Multimaterial Flows and Mixing

IAPCM, Sept. 12–15, 2011

Based on:

1. Local Oscillations in Finite Difference Solutions of Hyperbolic Conservation Laws

   Jiequan Li , Huazhong Tang, Gerald Warnecke and Lumei Zhang, *Math. Comp*, 78, 1997-2018, 2009.

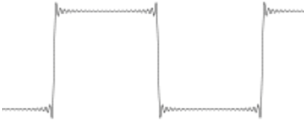
2. Heuristic Modified Equation Analysis on Oscillations in Numerical Solutions of Conservation Laws
   Jiequan Li and Zhicheng Yang, SIAM J. Numer. Anal., in press, 2011.

3. The von Neumann Analysis and Modified Equation Approach for Finite Difference Schemes
   Jiequan Li and Zhicheng Yang, preprint, 2011.

Ubiquity of Numerical Oscillations

Gibbs Phenomenon. The *n*–th partial sum of the Fourier series has large oscillations near the jump, which might increase the maximum of the partial sum above that of the function itself. The overshoot does not die out as the frequency increases, but approaches a finite limit[1].

(a) Functional approximation of square wave using 5 harmonics



(b) Functional approximation of square wave using 25 harmonics

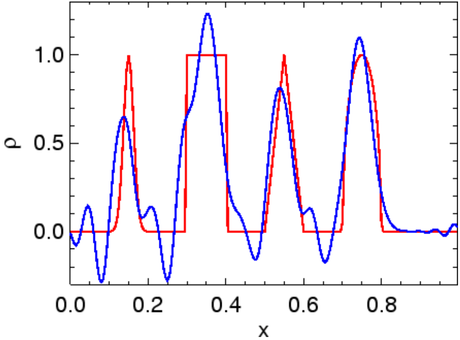
[1] H. S. Carslaw (1930): Introduction to the theory of Fourier's series and integrals.

Lax–Wendroff Method

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

$$u_j^{n+1} = u_j^n - \frac{\nu}{2}[u_{j+1}^n - u_{j-1}^n] + \frac{\nu^2}{2}[u_{j+1}^n - 2u_j^n + u_{j-1}^n], \quad \nu = c\tau/h.$$

In this lecture, τ is the time step size and h is the grid spacing.



The Lax–Wendroff method has accuracy of second order both in space and time.

P. Lax further showed that[2]

when a discontinuous initial value problem for a scalar hyperbolic equation in one space variable is approximated by a difference scheme that is **more than first order accurate**; it leads to overshoots analogous to the Gibbs phenomenon when discontinuous functions are approximated by sections of Fourier series.

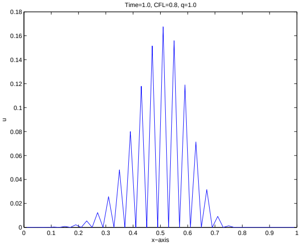
Lax's clarification is very intuitive and algebraic.

It is seemingly impressed that oscillations are present near jumps only when high order accurate schemes are used to approximate discontinuous solutions.

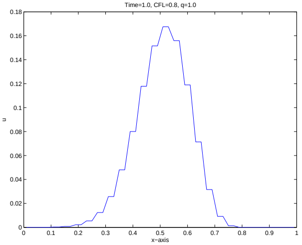
[2] P. D. Lax, Gibbs phenomena, J. Sci. Comput., Vol. 28, Nos. 2/3 (2006), 445–449.

Local Oscillations in Lax–Friedrichs Solutions[3][4]

Numerical results for the advection equation $u_t + u_x = 0$, 50 grid points are used, $\nu = 0.8$



(c) $u_j^0 = 1$ for $j = 25$ and $u_j^0 \equiv 0$ otherwise



(d) $u_j^0 = 1$ for $j = 25, 26$, and $u_j^0 \equiv 0$ otherwise

[3] H.-Z. Tang and G. Warnecke, A note on $(2K+1)$-point conservative monotone schemes, M2AN Math. Model. Numer. Anal., 38(2004), 345–357.

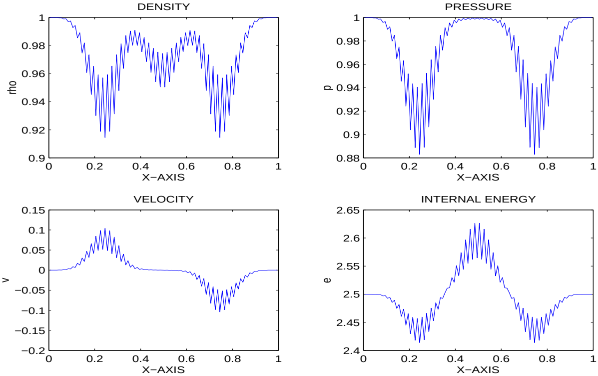
[4] M. Breuss, The correct use of the Lax-Friedrichs method, M2AN Math. Model. Numer. Anal. 38(2004), 519–540.

L-F Solutions of Compressible Euler Equations[5]

$$\rho_t + (\rho v)_x = 0, \quad (\rho v)_t + (\rho v^2 + p)_x = 0, \quad (\rho E)_t + (v(\rho E + p))_x = 0.$$

$T = 0.25$, 100 grid points, $CFL = 0.6$.

Initial data: $(\rho_j^0, u_j^0, p_j^0) = (0.125, 0, 0.1)$ for $j = 49$, 50, 51, and $(\rho_j^0, u_j^0, p_j^0) = (1, 0, 1)$ otherwise.



[5] J. L. Li, H. T. Tang, G. Warnecke and L. M. Zhang, local oscillations in finite difference solutions of hyperbolic conservation laws, Math. Comp., (2009), Vol. 78, pp. 1997–2018.

Remarks

1. The Lax–Friedrichs scheme is first order accurate, has TVD property and the maximum numerical viscosity in the class of TVD schemes. The large dissipation SHOULD suppresses local oscillations!!! This oscillatory phenomenon is very counter–intuitive and misleading.[6]

2. Oscillations in solutions by higher order schemes usually violate the maximum principle, and they are substantially different from those by TVD schemes.

[6] Nevertheless, the oscillations are just local and do not contradict to the TVD property because the latter is a global definition while the oscillations are of local nature. The preservation of the TVD property is due to the compensation by strong decrease in solution amplitude.

Purposes

Understanding of numerical oscillations in solutions by finite difference schemes, finite volume, spectral $\cdots$

Tools: Modified Equation Analysis, plus Discrete Fourier Analysis

Heuristic Observation

Discrete Fourier Decomposition of Initial Data

The initial datum is a square signal

$$u_0(x) = \begin{cases} 1, & 0 < x_1 < x_2 < 1, \\ 0, & \text{Otherwise}, \end{cases}$$

and the interval $[x_1, x_2]$ are partitioned with M grid points.

• Discretization with an odd number of grid points (j_1 and p are parameters related with the grid point indices, $\xi = 2kh$)

$$u_j^0 = (-1)^{j+j_1+M/2}h + ph + \sum_{k\neq 0, M/2} \frac{(-1)^k e^{i\xi(j+j_1)}(1 - e^{-i\xi p})}{M(1 - e^{-i\xi})}. \qquad \text{(C)}$$

• Discretization with an even number of grid points

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

Highest frequency modes $(-1)^j$ are included in discrete initial data (C) if M is odd.

Chequerboard modes from improper numerical boundary treatment

Example The leap-frog scheme for $v_t + v_x = 0$, $x \in (0, 1)$,

$$u_j^{n+1} = u_j^{n-1} - \nu(u_{j+1}^n - u_{j-1}^n).$$

We need numerical boundary conditions both at $x = 0$ and $x = 1$. At $x = 1$ ($j = M$), if the (improper) boundary formula is used [7]

$$u_M^{n+1} = \frac{1}{2}(u_{M-2}^n + u_M^n), \quad n \geq 1.$$

This occurs to many central schemes!!!

---

[7] It is well-known to result in instability. An easy remedy is to use

$$u_M^{n+1} = \frac{1}{2}(u_{M-1}^n + u_M^n)$$

Chequerboard Mode Solutions

• Generalized Lax–Friedrichs (GLF) scheme with initial datum $u_j^0 = (-1)^j$

$$u_j^{n+1} = u_j^n - \frac{\nu}{2}[u_{j-1}^n - u_{j+1}^n] + \frac{q}{2}[u_{j+1}^n - 2u_j^n + u_{j-1}^n], \quad 0 < q_{\min} \le q \le 1.$$

1. As $q = 1$, it is the Lax–Friedrichs scheme, which has the chequerboard mode solution:

   $$u_j^n = (-1)^{j+n}.$$

2. As $0 < q_{\min} \le q < 1$, there is a damped solution

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

   Note that it is the Lax–Wendroff scheme as $q = \nu^2$.

The same holds true for nonlinear problems:

$$u_j^{n+1} = u_j^n - \frac{\nu}{2}[f(u_{j-1}^n) - f(u_{j+1}^n)] + \frac{1}{2}[u_{j+1}^n - 2u_j^n + u_{j-1}^n].$$

- GLF solution with the square signal initial datum

1. Odd discretization case:

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

2. Even discretization case:

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

The highest frequency mode is persistent and not damped, which may be the cause of (local) oscillations.

- MacCormack scheme with initial data $u_j^0 = (-1)^j$

$$\begin{aligned} u_j^* &= u_j^n - \lambda\left[f(u_{j+1}^n) - f(u_j^n)\right], \\ u_j^{n+1} &= \frac{1}{2}\left[u_j^n + u_j^* - \lambda\big(f(u_j^*) - f(u_{j-1}^*)\big)\right]. \end{aligned}$$

1. As $f(u)$ is an even function, the solution is [8]

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

2. As $f(u)$ is an odd function, the solution is

$$u_j^{n+1} = (-1)^{j+n} b\left[1 - 2\lambda^2 \frac{f(b)}{b} f'(\xi)\right], \quad \xi \in (b, b + 2\lambda f(b))$$

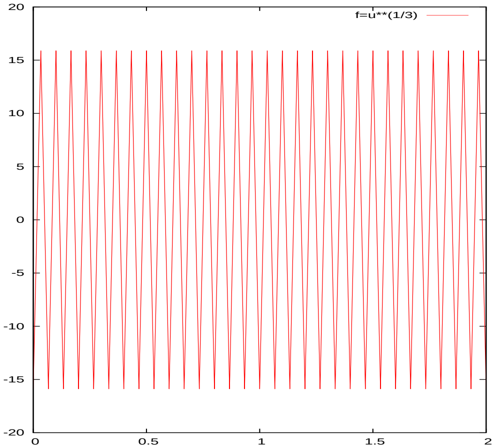
if $u_j^n = (-1)^{j+n} b$. Amplification factor: $1 - 2\lambda^2 \frac{f(b)}{b} f'(\xi)$.

The amplitude of solution is amplified if $|1 - 2\lambda^2 \frac{f(b)}{b} f'(\xi)| > 1$.

[8] Private communication with Huazhong Tang

- **MacCormack scheme for** $u_t + (u^\alpha)_x = 0,\ 0 < \alpha < 1$

The initial data is $u_j^0 = (-1)^j$. $\alpha = \frac{1}{3}$, $t = 1.19$, CFL=0.9.



Oscillations are crazily amplified.

Brief Summary

1. High frequency modes are often present in data, which may contaminate solutions.
2. High frequency modes may be persistent in solutions, not damped or be even amplified if there is no sufficiently strong dissipation.

Question: How do we quantify dissipation effects of numerical schemes so that the cause of oscillations could be well understood?

Discrete Fourier Analysis on Linear Problems (GLF)

$$u_j^n = \lambda_k^n e^{ij\xi}, \quad i^2 = -1, \xi = kh.$$

- **Low Frequency Modes** $\xi \approx 0$ The amplitude λ_k is

$$\lambda(k) = 1 + q(\cos\xi - 1) - i\nu a \sin\xi,$$

and

$$|\lambda_k|^2 = |\lambda(k)|^2 = 1 + 4(\nu^2 - q)\sin^2(\xi/2) + 4(q^2 - \nu^2)\sin^4(\xi/2).$$

The amplitude error is

$$|\lambda_k| - 1 = \begin{cases} \mathcal{O}(\xi), & \text{for monotone schemes} \\ \mathcal{O}(\xi^2), & \text{for L-W } q = \nu^2. \end{cases}$$

The relative phase error is

$$E_p(k) := -\frac{\arg\lambda(k)}{\nu\xi} - 1 = \mathcal{O}(\xi^2).$$

Common understanding: The numerical dissipation of monotone schemes $\mathcal{O}(\xi)$ suppresses relative phase errors, while the dissipation of L–W is too weaker $\mathcal{O}(\xi^2)$ and oscillations are triggered.

Moreover, for discontinuous L–W solutions, it can be shown that

$$\hat{u}(\xi) \sim 1/|\xi|, \qquad \text{as} \;\; \xi \to 0,$$

where $\hat{u}$ is the Fourier representation of u.

This implies that low frequency modes can cause oscillations in the L-W solutions.

- **High Frequency Modes $\xi \approx \pi$.**

Introduce $\xi' = k'h \approx 0$ such that $\xi = \pi + \xi'$

$$u_j^n = \lambda_k^n e^{ij\xi} = \lambda_k^n e^{i(\pi+\xi')j} = (-1)^{j+n}\lambda_{k'}^n e^{i\xi' j}.$$

λ_k and $\lambda_{k'}$ have the same amplitude modulus:

$$|\lambda_{k'}|^2 = 1 + 4(\nu^2 - q)\cos^2(\xi'/2) + 4(q^2 - \nu^2)\cos^4(\xi'/2)$$

The amplitude error at $\xi' = 0$ is

$$|\lambda_k|^2 - 1 = -4q(1-q),$$

which vanishes at $q = 0$. GLF becomes much less dissipative for high frequency modes as the parameter q increases.
The relative phase error at $\xi = \pi$ is

$$E_p(k) := -\frac{\pi(1-\nu)}{\nu\xi} + \frac{2(q-1)\nu\xi'}{(2q-1)\nu\xi} + \cdots = \mathcal{O}(1).$$

The relative phase errors of high frequency modes $\mathcal{O}(1)$ are huge and are difficult to be suppressed by the numerical dissipation.

Brief Summary for GLF

We distinguish low and high frequency Fourier modes $u_j^n = \lambda_k^n e^{ij\xi}$, $\xi = kh$. They behave differently.

1. For the low frequency modes ($\xi \approx 0$), the relative phase error is of order $\mathcal{O}(\xi^2)$, and the amplitude error (dissipation) becomes smaller as the parameter q decreases.
   1.1 The order of amplitude error is $\mathcal{O}(\xi)$ for the monotone schemes and thereby offsets the relative phase errors.
   1.2 The order of amplitude error is $\mathcal{O}(\xi^2)$ for the LW scheme and the relative phase errors cannot be offset. Hence oscillations are observable.
2. For the high frequency modes ($\xi \approx \pi$), the relative phase error is of order $\mathcal{O}(1)$, the amplitude error becomes larger as the parameter q is closer to $1/2$.
   In particular, as $q = 1$ (LF), the amplitude error vanishes and there is no dissipation on the highest frequency mode. This explains why oscillations in LF solutions are observable.

Application: Composite Scheme [9]

$$u^{n+1} = LF \circ (LW)_m u^n.$$

LF has large dissipation on low frequency modes; in contrast, LW has large dissipation on high frequency modes. When they are combined to form a composite scheme, their merits are preserved and some shortcomings are canceled out.

[9] R. Liska and B. Wendroff, Composite schemes for conservation laws. SIAM J. Numer. Anal. 35 (1998), no. 6, 2250 – 2271.

- **Remark on Discrete Fourier Analysis**

  1. Merits: Most often–used approach to investigate the stability of a scheme
  2. Limitation: Linear problems with constant coefficients; difficult to extend to nonlinear problems or even problems with variable coefficients.

Modified Equation Approach

1. Linear Part: Traditional modified equation at any wave number

   References:

   (a) F. Warming and B. J. Hyett, The modified equation approach to the stability and accuracy analysis of finite difference methods, J. Comp. Phys. 14, 159-179, 1974.
   (b) Jiequan Li and Zhicheng Yang, The von Neumann Analysis and Modified Equation Approach for Finite Difference Schemes, preprint, 2011.

2. Nonlinear Part: Heuristic modified equations (not rigorously)

   Reference: Jiequan Li and Zhicheng Yang, Heuristic modified equation analysis on oscillations in numerical solutions of conservation laws, SINUM, in press, 2011.

Linear Modified Equations

1. Time-dependent PDE with constant coefficients

$$u_t = \mathcal{L}_x(u), \quad \mathcal{L}_x(u) = \sum_{s=1}^{K} \alpha_s \partial_x^s u,$$

where α_s, $s = 1, \cdots, K$ are constant.

2. Two-level schemes take the form

$$Bu^{n+1} = Au^n, \quad \sum_{p=-N_1}^{N_2} B_p u_{j+p}^{n+1} = \sum_{p=-M_1}^{M_2} A_p u_{j+p}^n,$$

where B is invertible, the numerical solution u_j^n approximates the PDE solution $u(x_j, t_n)$,

$$u_j^n \sim u(x_j, t_n), \quad x_j = jh, t_n = n\tau.$$

Warming & Hyett's Approach (1974)

The modified equation is the **actual** partial differential equation that the difference scheme solves ($u_j^n \to u(x_j, t_n)$). The Taylor series expansion yields the modified equation

$$\partial_t u = \sum_s \beta_s(h, \tau) \partial_x^s u,$$

which implies <u>the consistency</u> to the equation if

$$\beta_s = \alpha_s, \quad s = 1, \cdots, K;$$

and <u>the accuracy of order q</u> if

$$\begin{array}{ll} (i) & \beta_s = \alpha_s, \quad s = 1, \cdots, K; \\ (ii) & \beta_s = 0, \quad s = K+1, \cdots, K+q-1; \\ (iii) & |\beta_{K+p}(h, \tau)| \le Ch^q, \quad C \text{ is constant.} \end{array}$$

Modified equation from discrete Fourier analysis (Classical!!!)

Write

$$\lambda_k = e^{i\omega(k)\tau}$$

where $\omega(k)$ is the frequency. Expand $\omega(k)$ at $k = 0$,

$$i\omega(k) = i\sum_{s=0}^{\infty} \frac{\omega^{(s)}(0)}{s!} k^s = \frac{1}{\tau}\sum_{s=0}^{\infty} \frac{[\ln \lambda(0)]^{(s)}}{s!} k^s, \quad \omega(0) = 0.$$

Formally transform $ik \to \partial_x$, $i\omega \to \partial_t$ to obtain the modified equation at $k = 0$,

$$\partial_t u = \frac{1}{\tau}\sum_{s=0}^{\infty} \frac{i^{-s}}{s!}[\ln \lambda(0)]^{(s)} \partial_x^s u.$$

Note that (e.g., for explicit scheme)

$$\lambda^{(s)}(0) = i^s h^s \sum_p p^s A_p.$$

Can write out $\beta_s(h, \tau)$ explicitly only in terms of $\lambda(0)$.

Example: Linear Advection Equation

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

Consistency conditions are

$$\sum_p A_p = 1, \quad \sum_p pA_p = \nu,$$

which implies that the modified equation should be

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = \frac{1}{\tau}\sum_{s=2}^{\infty}\frac{i^{-s}}{s!}[\ln\lambda(0)]^{(s)}\partial_x^s u.$$

The term of second order derivative is the *numerical viscosity*, which vanishes for L-W.

L-W:

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = -\frac{h^2}{6}(1-\nu^2)\frac{\partial^3 u}{\partial x^3} + \cdots .$$

L-F:

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = \frac{h^2}{2\tau}(1-\nu^2)\frac{\partial^2 u}{\partial x^2} + \cdots .$$

Modified equation at any wave-number

Motivated by the von Neumann analysis, we make an ansatz,

$$u_j^n = \alpha_j^n \tilde{u}_j^n, \quad \alpha_j^n = e^{i(n\omega_0\tau + jk_0h)}, \quad \omega_0 = \omega(k_0).$$

Substituting it into the (explicit) scheme yields

$$e^{i\omega_0\tau}\tilde{u}_j^{n+1} = \sum_p A_p \alpha_p^0 \tilde{u}_{j+p}^n, \qquad \alpha_p^0 = e^{ipk_0h}.$$

Replace $\tilde{u}_j^n$ by the value of a smooth function $\tilde{u}(x_j, t_n)$ to obtain

$$e^{i\omega_0\tau}\tilde{u}(x_j, t_{n+1}) = \sum_p A_p \alpha_p^0 \tilde{u}(x_{j+p}, t_n).$$

Then we can obtain the modified equation for $\tilde{u}(x, t)$

Theorem. The modified equation for $\tilde{u}$ at k_0 is

$$\partial_t \tilde{u} = \frac{\ln|\lambda(k_0)|}{\tau}\tilde{u} + \frac{1}{\tau}\sum_{s=1}^{\infty}\frac{i^{-s}}{s!}[\ln\lambda(k_0)]^{(s)}\partial_x^s \tilde{u}, \quad i^2 = -1.$$

Interpretation of modified equation at any wave number

1. Connection between $\tilde{u}$ and u. At $k = 0$, $\tilde{u} = u$. Their amplitudes are the same.
2. Full connection with the von Neumann analysis.

$$\tilde{\omega} = \tilde{\omega}(k_0 + \tilde{k}) = \sum_{s=1}^{\infty} \frac{\tilde{\omega}^{(s)}(k_0)}{s!} \tilde{k}^s.$$

   Then we transform $i\omega \to \partial_t$ and $i\tilde{k} \to \partial_x$ to obtain formally the modified equation.

3. The propagation of wave packets and dispersion. Introduce the group velocity

$$G(k) = -\frac{d\omega(k)}{dk} = \frac{i}{\tau} \frac{d \ln \lambda(k)}{dk}.$$

   Then the modified equation can be rewritten as

$$\partial_t \tilde{u} + G(k_0) \partial_x \tilde{u} = \frac{\ln |\lambda(k_0)|}{\tau} \tilde{u} - \sum_{s=1}^{\infty} \frac{i^{-(s-1)}}{s!} G^{(s-1)}(k_0) \partial_x^s \tilde{u}.$$

   This gives the PDE interpretation of group velocity in terms of modified equation.

4. Numerical Dissipation. We distinguish the dissipation as

4.1 Numerical damping: $\partial_t \tilde{u} \sim \dfrac{\ln|\lambda(k_0)|}{\tau}\tilde{u}$. Such solution behaves

$$\tilde{u}(t) \sim \tilde{u}(0)e^{\alpha t}, \quad \alpha = \frac{\ln|\lambda(k_0)|}{\tau}.$$

As $\alpha < 0$ ($\Longleftrightarrow |\lambda| < 1$), the solution decays exponentially. As $|\lambda| \leq 1 + C\tau$ (von Neumann stability condition), the solution is bounded

$$|\tilde{u}| \leq |\tilde{u}(0)|E^C.$$

4.2 Numerical diffusion: $\partial_t \tilde{u} \sim \epsilon\partial^2\tilde{u}$, $\epsilon = \frac{\tilde{\omega}''(k_0)}{2}$. The solution behaves

$$\tilde{u}(t) \sim \mathcal{O}(t^{-1/2}),$$

if $\epsilon > 0$. Note that:
(i) As $k_0 = 0$, $|\lambda(0)| = 1$ is the scheme is consistent. Then the numerical diffusion (viscosity) dominates the dissipation.
(ii) As $k_0 \neq 0$, generally speaking $|\lambda(k_0)| < 1$. Hence the numerical damping dominates the dissipation.

Nonlinear Modified Equations

- **Conservative Schemes**

$$u_j^{n+1} = u_j^n - \frac{\tau}{h}[F^n_{j+\frac{1}{2}} - F^n_{j-\frac{1}{2}}],$$

where numerical fluxes $F^n_{j+\frac{1}{2}} = F(u^n_{j-q}, \cdots, u^n_{j+p})$ is consistent with $f(u)$

$$F(u, \cdots, u) = f(u).$$

The classical modified equation of monotone schemes is (Harten, Hyman & Lax, CPAM, 1976),

$$u_t + f(u)_x = \tau\partial_x(\beta(u,\nu)\partial_x u), \quad \beta(u,\nu) > 0,$$

where the "red" term is the numerical viscosity, which suppresses oscillations near shocks.

Note that such dissipation just takes effect on oscillations caused by low frequency modes. Stronger dissipation is needed to suppress the oscillations caused by high frequency modes.

Ansatz for oscillatory modes [10]

$$u_j^n = (-1)^{j+n}\tilde{u}_j^n.$$

We obtain a heuristic modified equation for $\tilde{u}$,

$$\boxed{\mathcal{D}_t\tilde{u} + G(\tilde{u})_x = -2\frac{\Gamma(\tilde{u})}{\tau}\tilde{u} + \tau\cdot(\beta(\tilde{u},\lambda)\tilde{u}_x)_x + \tau\cdot\delta(\tilde{u},\lambda)\tilde{u}_x^2 + \mathcal{O}(\tau^2),}$$

where $\mathcal{D}_t$ is the forward difference operator in time $\mathcal{D}_t\tilde{u} = (\tilde{u}(t+\tau,\cdot) - \tilde{u}(t,\cdot))/\tau$, G, Γ, β and δ are all functions of $\tilde{u}$, $\lambda = \tau/h$.

The damping factor $\Gamma(\tilde{u})$ is computable

$$\boxed{\begin{aligned}\Gamma(\tilde{u}) :=\ & 1 + (-1)^{n+j+1}\lambda\left[F((-1)^{n+j-q+1}\tilde{u},\cdots,(-1)^{n+j+p}\tilde{u})\right.\\ & \left.-F((-1)^{n+j-q}\tilde{u},\cdots,(-1)^{n+j+p-1}\tilde{u})\right]/(2\tilde{u}).\end{aligned}}$$

[10] K.W. Morton and D.F.Mayers, Numerical solution of partial differential equations, 2nd edition, Cambridge University Press, 2005.

Numerical Damping Determines Dissipation!!!

$$\mathcal{D}_t \tilde{u} = -2\frac{\Gamma(\tilde{u})}{\tau}\tilde{u}.$$

1. $|\tilde{u}(t+\tau,\cdot)| < |\tilde{u}(t,\cdot)|$ if $0 < \Gamma(\tilde{u}) < 1$;
2. $|\tilde{u}(t+\tau,\cdot)| > |\tilde{u}(t,\cdot)|$ if $\Gamma(\tilde{u}) < 0$ or $\Gamma(\tilde{u}) > 1$;
3. $|\tilde{u}(t+\tau,\cdot)| = |\tilde{u}(t,\cdot)|$ if $\Gamma(\tilde{u}) = 0$ or $\Gamma(\tilde{u}) = 1$.

Stability Criterion from Heuristic Modified Equation

Proposition. The heuristic modified equation implies the criterion of the stability of the oscillatory solution:

(i) **Damping.** As $0 < \Gamma(\tilde{u}) < 1$, the term $-2\dfrac{\Gamma(\tilde{u})}{\tau}\tilde{u}$ plays the role of numerical damping that dissipates the oscillatory part and suppresses oscillations eventually.

(ii) **Neutrality.** As $\Gamma = 0$ or $\Gamma = 1$, the dissipation effect depends on the numerical viscosity (the term of second order derivative). For oscillatory modes, no dissipation effect exists and the oscillations persist.

(iii) **Amplification.** As $\Gamma < 0$ or $\Gamma > 1$, the factor $\Gamma(\tilde{u})$ amplifies the magnitude of the oscillatory modes, which may lead to the instability.

Monotone Schemes

$$u_j^{n+1} = H(u_{j-q}^n, \cdots, u_{j+p}^n) = u_j^n - \lambda[F_{j+\frac{1}{2}}^n - F_{j-\frac{1}{2}}^n],$$

where $F_{j+\frac{1}{2}}^n = F(u_{j-q+1}^n, \cdots, u_{j+p}^n)$. The monotonicity means that

$$\frac{\partial H}{\partial v_\ell}(v_{-q}, \cdots, v_\ell, \cdots, v_p) \geq 0,$$

for all $-q \leq \ell \leq p$.

Proposition. If the schemes are monotone, then there holds,

$$0 \leq \Gamma(\tilde{u}) \leq 1,$$

which implies that oscillatory modes are not amplified.

Specific Examples

1. **Generalized Lax–Friedrichs Schemes**

   The numerical flux $F^n_{j+\frac{1}{2}}$ takes the form

$$F^n_{j+1/2} = F(u^n_j, u^n_{j+1}) = \frac{f(u^n_{j+1}) + f(u^n_j)}{2} - \frac{Q}{2\lambda}(u^n_{j+1} - u^n_j),$$

   where $0 < Q_{\min} \leq Q \leq 1$. It is easy to calculate

$$\begin{aligned} \Gamma(\tilde{u}) = & \ 1 + (-1)^{n+1}\lambda \left[F((-1)^n\tilde{u}, (-1)^{n+1}\tilde{u}) \right. \\ & \left. -F((-1)^{n-1}\tilde{u}, (-1)^n\tilde{u})\right] / (2\tilde{u}) \\ = & \ 1 - Q. \end{aligned}$$

   Hence we have $0 \leq \Gamma(\tilde{u}) \leq 1$, and the generalized Lax-Friedrichs scheme does have the damping effect around the highest frequency mode if $0 < Q_{min} < Q < 1$.
   As $Q = 1$, it is the Lax–Friedrichs scheme and $\Gamma(\tilde{u}) = 1$. This explains why local oscillations in the solution obtained by the Lax–Friedrichs scheme are observed once the initial data is polluted by the oscillatory modes.

   Remark. This is consistent with Breuss's conclusion that schemes with a large viscosity coefficient are prone to oscillations at data extrema.

2. **The Richtmyer Scheme.** A two–step Lax–Wendroff scheme with

$$\begin{cases} F^n_{j+1/2} = F(u^n_j, u^n_{j+1}) = f(u^{n+1/2}_{j+1/2}), \\ u^{n+1/2}_{j+1/2} = \dfrac{1}{2}(u^n_j + u^n_{j+1}) - \dfrac{\lambda}{2}\left(f(u^n_{j+1}) - f(u^n_j)\right). \end{cases}$$

$$\begin{aligned} \Gamma(\tilde{u}) = \quad & 1 + (-1)^{n+1}\lambda\left\{f\left(-\tfrac{\lambda}{2}(f((-1)^{n+1}\tilde{u}) - f((-1)^n\tilde{u}))\right)\right. \\ & \left.-f\left(-\tfrac{\lambda}{2}(f((-1)^n\tilde{u}) - f((-1)^{n-1}\tilde{u}))\right)\right\}/(2\tilde{u}). \end{aligned}$$

- Odd flux case, $f(-u) = -f(u)$.

$$\Gamma(\tilde{u}) = 1 - \lambda\frac{f(\lambda f(\tilde{u}))}{\tilde{u}}.$$

- Even flux case, $f(-u) = f(u)$.

$$\Gamma(\tilde{u}) = 1.$$

- General cases, $f(u) = f_{odd}(u) + f_{even}(u)$.

$$\Gamma(\tilde{u}) = 1 - \lambda\frac{f_{odd}(\lambda f_{odd}(\tilde{u}))}{\tilde{u}}.$$

3. **Zwas–Abarbanel Third Order Scheme** A third order scheme with

$$
\begin{aligned}
u_j^{n+1} = \;& u_j^n - \lambda\left[\frac{1}{2}(f_{j+1}^n - f_{j-1}^n) - \frac{\delta}{12}(f_{j+2}^n - 2f_{j+1}^n + 2f_{j-1}^n - f_{j-2}^n)\right] \\
&+ \frac{\lambda^2}{2}\left[a_{j+\frac{1}{2}}^n(f_{j+1}^n - f_j^n) - a_{j-\frac{1}{2}}^n(f_j^n - f_{j-1}^n)\right] \\
&+ \frac{\epsilon\lambda^3}{12}\left[a_{j+1}^n(f_{j+2}^n - f_j^n) - 2a_j^n(f_{j+1}^n - f_{j-1}^n) + a_{j-1}^n(f_j^n - f_{j-2}^n)\right],
\end{aligned}
$$

As $\delta = \epsilon = 0$, back to L-W.

- Odd flux case.

$$\Gamma(\tilde{u}) = 1 - \lambda^2 a(0) f(\tilde{u})/\tilde{u}.$$

- Even flux case.

$$\Gamma(\tilde{u}) = 1.$$

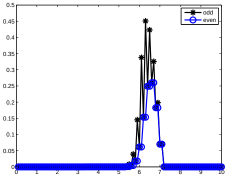
- General cases.

$$\Gamma(\tilde{u}) = 1 - \lambda^2 a(0) f_{odd}(\tilde{u})/\tilde{u}.$$

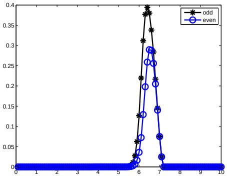
Numerical Examples

- **Damping.** The initial data

$$D_{odd}: \quad u_j^0 = \begin{cases} 1, & j = 49, 50, 51, \\ 0, & \text{otherwise.} \end{cases}$$



(e) The L–F scheme with $Q = 1.0$



(f) The generalized L–F scheme with $Q = 0.9$

Figure: The propagation of a single square signal. The flux $f(u) = \frac{u}{\sqrt{1+u^2}}$

4. Neutrality. The initial data

$$u_j^0 = C_j + 0.0001 \times (D_{odd})_j,$$

where C_j is the highest frequency mode $C_j = (-1)^j$.

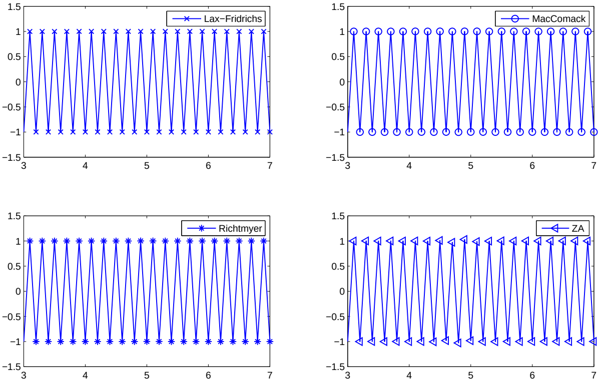


Figure: The neutrality of schemes on high frequency modes. The number of computation steps is 10. The Burgers equation is used.

5. Amplification.

$$\Gamma_{\mathrm{LxF}}(1) = 0, \quad \Gamma_{\mathrm{Mac}}(1) = 0.6994,$$
$$\Gamma_{\mathrm{Rich}}(1) = -0.9188, \quad \Gamma_{\mathrm{ZA}}(1) = -2.6204.$$

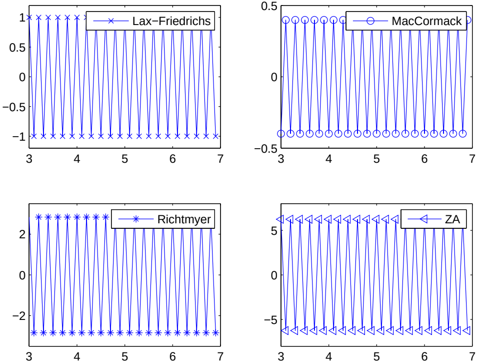


Figure: One first step computation: The flux function is $f(u) = \dfrac{u}{\sqrt{1+u^2}}$.

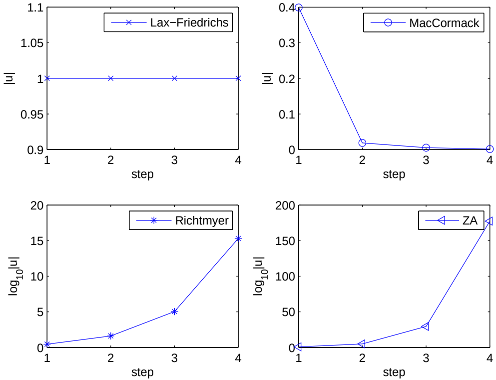


Figure: Amplification or damping trend of oscillatory modes.

Summary and Discussions

What we have done:

1. Both discrete Fourier analysis and modified equation approaches are used to analyze oscillatory modes. A heuristic modified equation is derived.
2. Numerical dissipation is distinguished as numerical damping ornumerical diffusion. Numerical damping plays a dominant role in suppressing oscillations caused by oscillatory modes.
3. A heuristic stability criterion is proposed to check the numerical damping effect on oscillatory modes.

What we want to do:

1. Rigorously derive the modified equation in order to analyze oscillatory modes.
2. Extend the results in a general setting, not only for scalar conservation laws.
3. Use the results to design new schemes.
4. $\cdots$

THANK YOU VERY MUCH