Finite difference schemes for scalar linear hyperbolic PDE in 1-D

Praveen. C
praveen@math.tifrbng.res.in

Tata Institute of Fundamental Research
Center for Applicable Mathematics
Bangalore 560065
http://math.tifrbng.res.in/~praveen

February 4, 2013

Simplest hyperbolic PDE

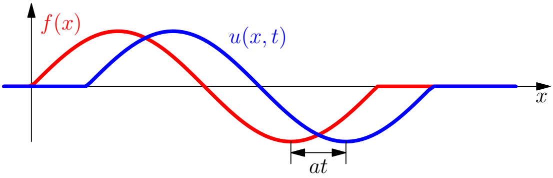
Linear, scalar convection/advection equation (Initial value problem)

$$\begin{aligned} u_t + a u_x &= 0 \qquad x \in \mathbb{R}, \quad t > 0 \\ u(x,0) &= f(x) \qquad x \in \mathbb{R} \end{aligned} \tag{1}$$

Exact solution

$$u(x,t) = f(x - at)$$

Initial condition is convected with speed a without change of form.



Hence the extrema of the solution do not change with time. Also the L^2-norm of the solution does not change with time. If $E(t)$ is the solution operator

$$u(x,t) = E(t) f(x) \qquad \Longrightarrow \qquad \|E(t)u\| = \|u\|$$

in both sup-norm and L^2-norm.

Method of characteristics

How does the solution vary along the space-time curve $x = x(t)$ with

$$\frac{\mathrm{d}}{\mathrm{d}t} x(t) = a$$

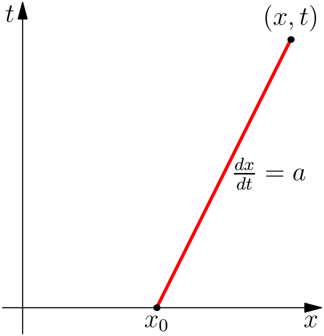
Variation of u along this curve

$$\begin{aligned} & \frac{\mathrm{d}}{\mathrm{d}t} u(x(t), t) \\ = \quad & \frac{\partial u}{\partial t}(x(t), t) + \frac{\mathrm{d}}{\mathrm{d}t} x(t) \cdot \frac{\partial u}{\partial x}(x(t), t) \\ = \quad & u_t + a u_x \\ = \quad & 0 \end{aligned}$$

Solution is constant along the characteristic curve.

Backward characteristic through (x, t) hits initial time $t = 0$ at

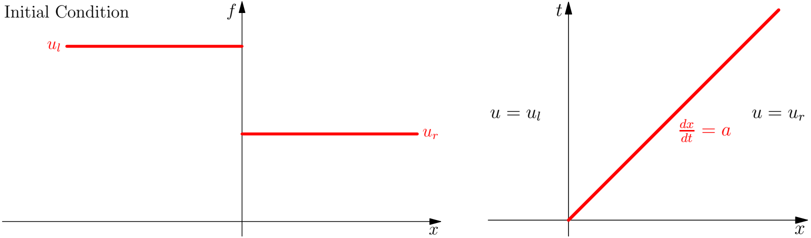
$$x_0 = x - at$$



$$u(x, t) = u(x_0, 0) = f(x - at)$$

Discontinuous solutions

If the initial condition is discontinuous, then we can still talk of a solution, even though the PDE is not satisfied everywhere. The MOC method is still valid and gives the solution at any later time. The solution is again obtained by translating the initial condition. The discontinuity propagates along the characteristic curve.



Such a discontinuity is called a contact discontinuity/wave. The charactersitics on either side of the discontinuity curve are parallel to it.

Finite difference scheme

Forward time, backward space (FTBS)

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + a\frac{u_j^n - u_{j-1}^n}{h} = 0 \quad \Rightarrow \quad u_j^{n+1} = (1 - a\lambda)u_j^n + a\lambda u_{j-1}^n$$

Forward time, central space (FTCS)

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + a\frac{u_{j+1}^n - u_{j-1}^n}{2h} = 0 \quad \Rightarrow \quad u_j^{n+1} = u_j^n + \frac{a\lambda}{2}(u_{j-1}^n - u_{j+1}^n)$$

Forward time, forward space (FTFS)

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + a\frac{u_{j+1}^n - u_j^n}{h} = 0 \quad \Rightarrow \quad u_j^{n+1} = (1 + a\lambda)u_j^n - a\lambda u_{j+1}^n$$

$$\lambda = \frac{\Delta t}{h}$$

Maximum principle

- $a > 0$: FTBS scheme is stable in maximum norm if $a\lambda \leq 1$, others are unstable.
- $a < 0$: FTFS scheme is stable in maximum norm if $|a|\lambda \leq 1$, others are unstable.
- FTCS scheme is never stable in maximum norm.
- Upwind scheme: switch between backward and forward difference

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + a^+ \frac{u_j^n - u_{j-1}^n}{h} + a^- \frac{u_{j+1}^n - u_j^n}{h} = 0$$

Stable in maximum norm for any a provided $|a|\lambda \leq 1$ is satisfied.

Courant-Friedrichs-Levy (CFL) number, CFL condition

$$\mathrm{CFL} = |a|\lambda = \frac{|a|\Delta t}{h}, \qquad \mathrm{CFL} \leq 1, \qquad \Delta t = \mathcal{O}(h)$$

Numerical example

- PDE
$$u_t + au_x = 0, \qquad x \in (0,1)$$
- with initial condition
$$u(x,0) = \sin(2\pi x)$$
- and periodic boundary conditions.
- Use $a = 1$. Try backward and forward difference scheme with CFL=0.5. The scheme is implemented in the matlab program `lin_hyp_1d_periodic.m`
- Backward difference: `lin_hyp_1d_periodic(100, 0.5, 'bd')`
- Forward difference: `lin_hyp_1d_periodic(100, 0.5, 'fd')`

Lax-Friedrichs (LF) scheme

- Forward time, central space (FTCS)

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + a\frac{u_{j+1}^n - u_{j-1}^n}{2h} = 0$$

  is not stable in maximum norm. We can modify it in following way.

- LF scheme

$$\frac{u_j^{n+1} - \frac{1}{2}(u_{j-1}^n + u_{j+1}^n)}{\Delta t} + a\frac{u_{j+1}^n - u_{j-1}^n}{2h} = 0$$

$$\Downarrow$$

$$u_j^{n+1} = \frac{1}{2}(1 + a\lambda)u_{j-1}^n + \frac{1}{2}(1 - a\lambda)u_{j+1}^n$$

  Stable in maximum norm if $|a|\lambda \leq 1$.

Lax-Wendroff (LW) scheme

Taylor's formula

$$u(x_j, t^{n+1}) = u(x_j, t^n) + \Delta t u_t(x_j, t^n) + \frac{1}{2}\Delta t^2 u_{tt}(x_j, t^n) + \mathcal{O}\left(\Delta t^3\right)$$

Use the PDE

$$u_t = -a u_x \qquad u_{tt} = a^2 u_{xx}$$

to get

$$u(x_j, t^{n+1}) = u(x_j, t^n) - a\Delta t u_x(x_j, t^n) + \frac{1}{2}a^2\Delta t^2 u_{xx}(x_j, t^n) + \mathcal{O}\left(\Delta t^3\right)$$

Approximate u_x and u_{xx} by central differences

$$u_j^{n+1} = u_j^n - a\Delta t \frac{u_{j+1}^n - u_{j-1}^n}{2h} + \frac{1}{2}a^2\Delta t^2 \frac{u_{j-1}^n - 2u_j^n + u_{j+1}^n}{h^2} + \mathcal{O}\left(\Delta t^3\right)$$

LW scheme

$$u_j^{n+1} = u_j^n - \frac{1}{2}a\lambda(u_{j+1}^n - u_{j-1}^n) + \frac{1}{2}a^2\lambda^2(u_{j-1}^n - 2u_j^n + u_{j+1}^n)$$

This scheme is not bounded in maximum norm but is stable in L^2 norm.

More schemes

- Leapfrog scheme (Three level scheme)

$$\frac{u_j^{n+1} - u_j^{n-1}}{2\Delta t} + a\frac{u_{j+1}^n - u_{j-1}^n}{2h} = 0$$

  Not self-starting, need different scheme for first time step.

- Backward time, central space, BTCS (Backward Euler implicit scheme)

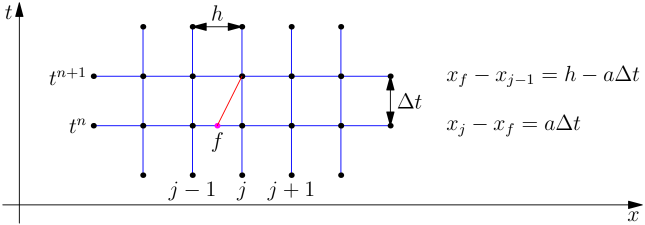
$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + a\frac{u_{j+1}^{n+1} - u_{j-1}^{n+1}}{2h} = 0$$

Schemes via method of characteristics, CFL condition

Exact solution is

$$u(x_j, t^{n+1}) = u(x_f, t^n)$$

Assume $a > 0$ and $a\lambda \leq 1$, i.e., $a\Delta t \leq \Delta x$



- Linear interpolation between x_{j-1} and x_j

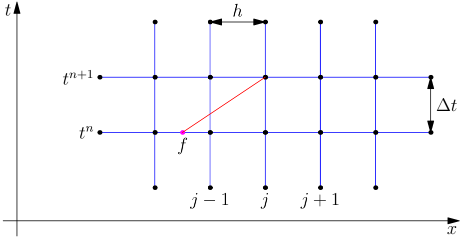
$$u_j^{n+1} = \frac{x_j - x_f}{h} u_{j-1} + \frac{x_f - x_{j-1}}{h} u_j \Longrightarrow u_j^{n+1} = a\lambda u_{j-1}^n + (1 - a\lambda) u_j^n$$

which is the FTBS (upwind) scheme.

Schemes via method of characteristics, CFL condition

- Quadratic interpolation between x_{j-1}, x_j, x_{j+1} leads to the Lax-Wendroff scheme. Valid procedure for $a > 0$ and $a < 0$.

If $a\lambda > 1$, then foot of backward characteristic x_f intersects $t = t^n$ outside the interval (x_{j-1}, x_{j+1}).



Interpolation using the values u_{j-1}, u_j, u_{j+1} would be wrong in this case. CFL condition ensures correct domain of dependance

Exact domain of dependance $\subset$ Numerical domain of dependance

Consistency and accuracy

Local truncation error : e.g. FTCS scheme

$$\tau_j^n = \frac{u(x_j, t^{n+1}) - u(x_j, t^n)}{\Delta t} + a\frac{u(x_{j+1}, t^n) - u(x_{j-1}, t^n)}{2h}$$

Truncation error

$$\tau(\Delta t, h) = \max_{j,n} |\tau_j^n|$$

Consistency

The numerical scheme is consistent if

$$\tau(\Delta t, h) \to 0 \qquad \text{as} \qquad \Delta t, h \to 0$$

Accuracy

The numerical scheme is accurate of order p in time and to the order q in space, if for a sufficiently regular solution of the exact problem, we have

$$\tau(\Delta t, h) = \mathcal{O}\left(\Delta t^p + h^q\right)$$

Consistency and accuracy

Scheme	$\tau(\Delta t, h)$
FTCS	$\mathcal{O}\left(\Delta t + h^2\right)$
Upwind	$\mathcal{O}\left(\Delta t + h\right)$
Lax-Friedrichs	$\mathcal{O}\left(\frac{h^2}{\Delta t} + \Delta t + h^2\right)$
Lax-Wendroff	$\mathcal{O}\left(\Delta t^2 + h^2 + h^2 \Delta t\right)$

Convergence in maximum norm

The scheme is convergent in the maximum norm if

$$\lim_{\Delta t, h \to 0} \max_{j,n} |u(x_j, t^n) - u_j^n| = 0$$

Remark: We will also consider weaker norms such as $\|\cdot\|_{1,h}$ and $\|\cdot\|_{2,h}$

$$\|u\|_{1,h} = h \sum_j |u_j|, \qquad \|u\|_{2,h} = (h \sum_j |u_j|^2)^{\frac{1}{2}}$$

Stability

Definition: Stable scheme (finite time stability)

A numerical scheme is stable if for each time T, there exists a constant $C_T > 0$ such that for each $h > 0$, there exists $\delta_0 > 0$ (possibly dependent on h) such that for any $0 < \Delta t < \delta_0$, we have

$$\|u^n\| \leq C_T \left\|u^0\right\|$$

for each n such that $n\Delta t \leq T$, and for each initial condition u^0. The constant C_T cannot depend on h.

Remark: An equivalent statement of the above condition is

$$\|u^n\| \leq (1 + C\Delta t) \left\|u^{n-1}\right\|, \qquad C \geq 0$$

where the constant C cannot depend on Δt or h. Then

$$\|u^n\| \leq (1 + C\Delta t)^n \left\|u^0\right\| \leq \mathrm{e}^{CT} \left\|u^0\right\|$$

Stability

Strongly stable scheme

A numerical scheme is said to be strongly stable if

$$\|u^n\| \leq \left\|u^{n-1}\right\|$$

i.e., if $C_T = 1$.

Remark: For hyperbolic problems, one is often interested in solutions after a long time ($T \gg 1$). In such cases, it is desirable to have strong stability since then the numerical solutions are guaranteed to be bounded for each value of T.

Stability in $\|\cdot\|_{1,h}$ (Upwind, LF and LW schemes)

If the CFL condition $|a|\lambda \leq 1$ is satisfied, then the upwind, Lax-Friedrichs and Lax-Wendroff schemes are strongly stable in the norm $\|\cdot\|_{1,h}$.

Proof: (1) Consider the upwind scheme with $a > 0$

$$u_j^{n+1} = (1 - a\lambda)u_j^n + a\lambda u_{j-1}^n$$

Then

$$\left\|u^{n+1}\right\|_{1,h} \leq h \sum_j |(1 - a\lambda)u_j^n| + h \sum_j |a\lambda u_{j-1}^n|$$

and since $a\lambda > 0$ and $(1 - a\lambda) > 0$, we get

$$\left\|u^{n+1}\right\|_{1,h} \leq h \sum_j (1 - a\lambda)|u_j^n| + h \sum_j a\lambda |u_{j-1}^n| = \|u^n\|_{1,h}$$

(2) For the Lax-Friedrichs scheme

$$u_j^{n+1} = \frac{1}{2}(1 + a\lambda)u_{j-1}^n + \frac{1}{2}(1 - a\lambda)u_{j+1}^n$$

with $(1 \pm a\lambda) \geq 0$ due to CFL condition, so that

$$\begin{aligned}
\left\|u^{n+1}\right\|_{1,h} &\leq \frac{1}{2}h\sum_j |(1+a\lambda)u_{j-1}^n| + \frac{1}{2}h\sum_j |(1-a\lambda)u_{j+1}^n| \\
&= \frac{1}{2}h\sum_j (1+a\lambda)|u_{j-1}^n| + \frac{1}{2}h\sum_j (1-a\lambda)|u_{j+1}^n| \\
&= \|u^n\|_{1,h}
\end{aligned}$$

(3) For the Lax-Wendroff scheme, the proof is analogous.

Remark: We have already shown the upwind and LF schemes to be stable in maximum norm. The LW scheme is NOT bounded in maximum norm.

$$u_j^{n+1} = \frac{1}{2}a\lambda(1+a\lambda)u_{j-1}^n + (1-a^2\lambda^2)u_j^n + \frac{1}{2}a\lambda(-1+a\lambda)u_{j+1}^n$$

However this does not mean that it is unstable in maximum norm.

Stability in $\|\cdot\|_{2,h}$ (BTCS, FTCS)

(1) The BTCS scheme is strongly stable in the norm $\|\cdot\|_{2,h}$ with no restriction on Δt. (2) The FTCS is never strongly stable. However, it is stable with constant $C_T = \mathrm{e}^{T/2}$ provided we assume that Δt satisfies the following condition (which is more restrictive than the CFL condition)

$$\Delta t \leq \left(\frac{h}{a}\right)^2$$

<u>Proof</u>: (1) <u>BTCS (implicit Euler scheme)</u>

$$u_j^{n+1}\left(\frac{u_j^{n+1}-u_j^n}{\Delta t}+a\frac{u_{j+1}^{n+1}-u_{j-1}^{n+1}}{2h}\right)=0$$

which can re-arranged as

$$(u_j^{n+1})^2+(u_j^{n+1}-u_j^n)^2=(u_j^n)^2-a\lambda(u_{j+1}^{n+1}-u_{j-1}^{n+1})u_j^{n+1}$$

Summing up over all the cells, the last term vanishes (telescopic collapse) to yield

$$\left\|u^{n+1}\right\|_{2,h}^2 = \|u^n\|_{2,h}^2 - h\sum_j (u_j^{n+1} - u_j^n)^2 \le \|u^n\|_{2,h}^2$$

(2) FTCS scheme

$$u_j^n \left(\frac{u_j^{n+1} - u_j^n}{\Delta t} + a\frac{u_{j+1}^n - u_{j-1}^n}{2h} \right) = 0$$

which can be rearranged as

$$(u_j^{n+1})^2 = (u_j^n)^2 + (u_j^{n+1} - u_j^n)^2 - a\lambda(u_{j+1}^n - u_{j-1}^n)u_j^n$$

so that

$$\left\|u^{n+1}\right\|_{2,h}^2 = \|u^n\|_{2,h}^2 + h\sum_j (u_j^{n+1} - u_j^n)^2 \ge \|u^n\|_{2,h}^2$$

and the *FTCS scheme is not strongly stable*. However if we write the scheme

$$u_j^{n+1} - u_j^n = -\frac{a\lambda}{2}(u_{j+1}^n - u_{j-1}^n)$$

and use it in above equation, we get

$$(u_j^{n+1})^2 = (u_j^n)^2 + \left(\frac{a\lambda}{2}\right)^2 (u_{j+1}^n - u_{j-1}^n)^2 - a\lambda(u_{j+1}^n - u_{j-1}^n)u_j^n$$

Summing up over all the equations

$$\begin{aligned}
\left\|u^{n+1}\right\|_{2,h}^2 &= \|u^n\|_{2,h}^2 + h\left(\frac{a\lambda}{2}\right)^2 \sum_j (u_{j+1}^n - u_{j-1}^n)^2 \\
&\leq \|u^n\|_{2,h}^2 + h\left(\frac{a\lambda}{2}\right)^2 2\sum_j [(u_{j+1}^n)^2 + (u_{j-1}^n)^2] \\
&= \|u^n\|_{2,h}^2 + a^2\lambda^2 \|u^n\|_{2,h}^2 \quad \text{assume} \quad a^2\lambda^2 \leq \Delta t \\
&\leq (1+\Delta t)\|u^n\|_{2,h}^2 \quad \text{if} \quad \Delta t \leq \left(\frac{h}{a}\right)^2
\end{aligned}$$

This yields

$$\|u^n\|_{2,h}^2 \leq (1+\Delta t)^n \left\|u^0\right\|_{2,h}^2 \leq \mathrm{e}^T \left\|u^0\right\|_{2,h}^2 \quad \forall n \quad \text{s.t.} \quad n\Delta t \leq T$$

which shows stability of FTCS scheme since $\|u^n\|_{2,h} \leq \mathrm{e}^{T/2}\left\|u^0\right\|_{2,h}$.

Remark: The proof in the last theorem requires the following algebraic identities: for any $A, B \in \mathbb{R}$

$$\begin{aligned}(B-A)B &= (B-A)^2 + (B-A)A \\ &= \frac{1}{2}[B^2 - A^2 + (B-A)^2] \\ &= \frac{1}{2}[(B-A)^2 + (B-A)(B+A)]\end{aligned}$$

and

$$(B-A)A = \frac{1}{2}[B^2 - A^2 - (B-A)^2]$$

Remark: The above proofs were possible due to the algebraic structure of the schemes. For other schemes, we may not be able to show stability in $\|\cdot\|_{2,h}$ using such algebraic techniques. In those cases, the Fourier or Von Neumann stability analysis will be useful. However this approach is limited to linear problems. Even for non-linear problems, Fourier stability is a necessary condition; from a computational viewpoint, it keeps roundoff errors under control.

Discrete Fourier transform

Periodic initial condition

$$f : [0, 2\pi] \to \mathbb{R}$$

Let N be an even integer. Consider a grid of $N+1$ points with grid point coordinates

$$x_j = jh, \quad j = 0, 1, \ldots, N \qquad \text{with} \qquad h = \frac{2\pi}{N}$$

We can approximate $f(x)$ on the grid by the discrete Fourier series

$$\tilde{f}(x) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \alpha_k e^{ikx}$$

Determine the coefficients α_k by interpolation $f_j = f(x_j) = \tilde{f}(x_j)$, i.e.,

$$f_j = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \alpha_k e^{ikjh}, \qquad j = 0, 1, \ldots, N-1$$

We ignore f_N since $f_N = f_0$ by periodicity.

Discrete Fourier transform

To determine α_k, multiply by $\mathrm{e}^{-\mathrm{i}ljh}$ and sum

$$\sum_{j=0}^{N-1} f_j \mathrm{e}^{-\mathrm{i}ljh} = \sum_{j=0}^{N-1} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \alpha_k \mathrm{e}^{\mathrm{i}kjh} \mathrm{e}^{-\mathrm{i}ljh} = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \alpha_k \sum_{j=0}^{N-1} \mathrm{e}^{\mathrm{i}(k-l)jh}$$

We can show that

$$\sum_{j=0}^{N-1} \mathrm{e}^{\mathrm{i}(k-l)jh} = \frac{2\pi}{h}\delta_{kl}, \qquad -\frac{N}{2} \leq k, l \leq \frac{N}{2} - 1$$

which gives the Fourier coefficient

$$\alpha_l = \frac{h}{2\pi} \sum_{j=0}^{N-1} f_j \mathrm{e}^{-\mathrm{i}ljh}$$

Parseval relation

We want to relate l^2 norm of a grid function with the l^2 norm of its Fourier coefficients.

$$f_j^2 = f_j^* f_j = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \alpha_k^* \mathrm{e}^{-\mathrm{i}kjh} f_j$$

Summing up over all the grid points

$$\sum_{j=0}^{N-1} f_j^2 = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \alpha_k^* \sum_{j=0}^{N-1} f_j \mathrm{e}^{-\mathrm{i}kjh} = \frac{2\pi}{h} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \alpha_k^* \alpha_k$$

which gives the discrete Parseval relation

$$\|f\|_{2,h}^2 = h \sum_{j=0}^{N-1} |f_j|^2 = 2\pi \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |\alpha_k|^2$$

Von Neumann stability analysis

Let us approximate the initial condition by the discrete Fourier series and apply any of the numerical schemes studied till now to this approximate initial condition. Then the solution at any time t^n can be written as

$$u_j^n = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \alpha_k (\gamma_k)^n \mathrm{e}^{\mathrm{i}kjh}$$

where $\gamma_k \in \mathbb{C}$ is called the *amplification coefficient* of the k-th frequency (or harmonic). If $|\gamma_k| > 1$ then the numerical solution is likely to blow up with time.

<u>FTCS scheme</u>: Consider first time step

$$\begin{aligned} u_j^1 &= u_j^0 + \frac{a\lambda}{2}(u_{j-1}^0 - u_{j+1}^0) \\ &= \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \alpha_k \mathrm{e}^{\mathrm{i}kjh} \underbrace{\left[1 + \frac{1}{2}a\lambda(\mathrm{e}^{-\mathrm{i}kh} - \mathrm{e}^{\mathrm{i}kh})\right]}_{\gamma_k} \end{aligned}$$

Von Neumann stability analysis

But

$$\gamma_k = 1 - \mathrm{i}a\lambda \sin(kh), \qquad |\gamma_k| = [1 + a^2\lambda^2 \sin^2(kh)]^{\frac{1}{2}} \geq 1$$

and hence the FTCS scheme is not strongly unstable.

Scheme	γ_k
FTCS	$1 - \mathrm{i}a\lambda \sin(kh)$
BTCS (imp. Euler)	$[1 + \mathrm{i}a\lambda \sin(kh)]^{-1}$
Upwind	$1 - \lvert a\rvert\lambda(1 - \mathrm{e}^{-\mathrm{i}kh})$
Lax-Friedrichs	$\cos(kh) - \mathrm{i}a\lambda \sin(kh)$
Lax-Wendroff	$1 - \mathrm{i}a\lambda \sin(kh) - a^2\lambda^2[1 - \cos(kh)]$

Von Neumann stability

If $\exists \beta \geq 0$ and a positive integer m such that, for suitable choices of Δt and h, we have $|\gamma_k| \leq (1+\beta\Delta t)^{\frac{1}{m}}$ for each k, then the scheme is stable with respect to $\|\cdot\|_{2,h}$ with a stability constant $C_T = \mathrm{e}^{\frac{\beta T}{m}}$. In particular, if $\beta = 0$ (so that $|\gamma_k| \leq 1$), then the scheme is strongly stable.

Proof: Using the Parseval relation, we get

$$\begin{aligned}
\|u^n\|_{2,h}^2 &= 2\pi \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |\alpha_k|^2 |\gamma_k|^{2n} \\
&\leq (1+\beta\Delta t)^{\frac{2n}{m}} 2\pi \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |\alpha_k|^2 = (1+\beta\Delta t)^{\frac{2n}{m}} \left\|u^0\right\|_{2,h}^2 \\
&\leq \mathrm{e}^{\frac{2\beta n\Delta t}{m}} \left\|u^0\right\|_{2,h}^2 \quad \text{since} \quad 1+\beta\Delta t \leq \mathrm{e}^{\beta\Delta t} \\
&\leq \mathrm{e}^{\frac{2\beta T}{m}} \left\|u^0\right\|_{2,h}^2 \qquad \forall n \quad \text{s.t.} \quad n\Delta t \leq T
\end{aligned}$$

The second part of the theorem follows easily.

Stability of standard schemes in $\|\cdot\|_{2,h}$

Scheme	Stable	Strongly stable	CFL	$\lvert\gamma_k\rvert \leq$
FTCS*	Yes	No	$\Delta t \leq \beta \frac{h^2}{a^2}$	$(1+\beta\Delta t)^{\frac{1}{2}}$
BTCS (imp. Euler)	Yes	Yes	-	1
Upwind	Yes	Yes	$\lvert a\rvert\lambda \leq 1$	1
Lax-Friedrichs	Yes	Yes	$\lvert a\rvert\lambda \leq 1$	1
Lax-Wendroff	Yes	Yes	$\lvert a\rvert\lambda \leq 1$	1

* For any $\beta > 0$.

Remark: Leapfrog scheme has

$$|\gamma_k| = 1, \qquad \text{if} \quad |a|\lambda \leq 1$$

so that it is neutrally stable.

Dissipation property

- The exact solution $u(x,t) = f(x - at)$ can be written as a Fourier series

$$u(x_j, t^n) = \sum_{k=-\infty}^{\infty} \alpha_k \mathrm{e}^{\mathrm{i}k(x_j - at_n)} = \sum_{k=-\infty}^{\infty} \alpha_k \mathrm{e}^{\mathrm{i}kjh} (g_k)^n, \qquad g_k = \mathrm{e}^{-\mathrm{i}ak\Delta t}$$

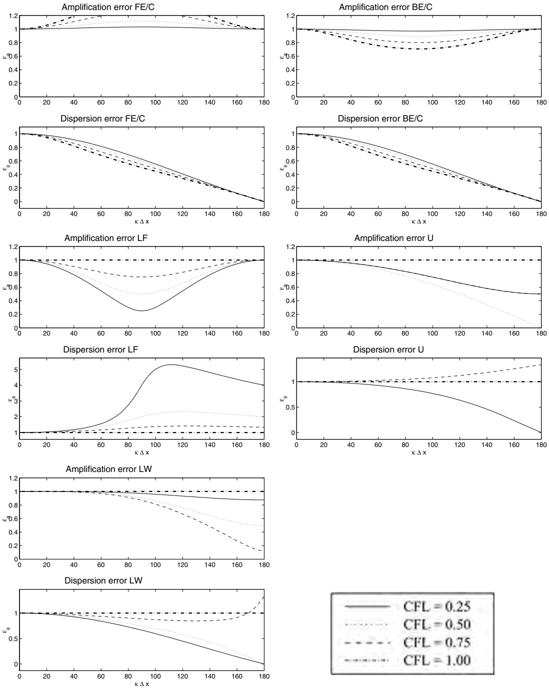
- Note that $|g_k| = 1$ which reflects the property that the exact solution is advected without change of shape (amplitude).
- In the numerical scheme, g_k is replaced by γ_k and for strong stability, we need $|\gamma_k| \leq 1$.
- If $|\gamma_k| < 1$ the amplitude of numerical solution will decrease with time, which is a sign of some dissipative mechanism in the numerical scheme.
- We can measure the amplification error by

$$\epsilon_a(k) = \frac{|\gamma_k|}{|g_k|} = \frac{\text{Numerical amplitude}}{\text{Exact amplitude}}$$

- Ideally, we would like to have $\epsilon_a(k) \approx 1$ for all wave numbers k that can be represented on the grid.

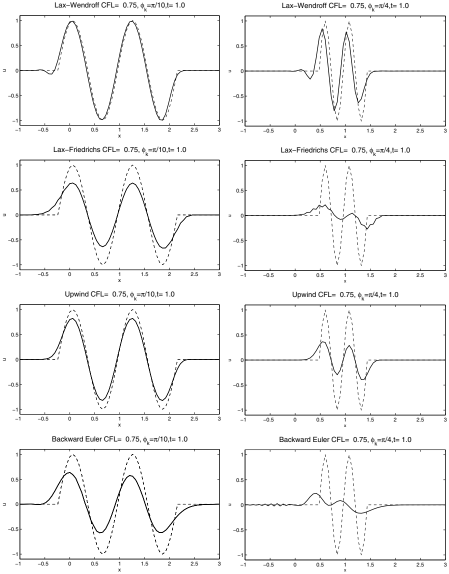
Dispersion property

- Let us set
$$\phi_k = kh \qquad \text{so that} \qquad g_k = \mathrm{e}^{-\mathrm{i}a\lambda\phi_k}$$
ϕ_k is called the phase angle of the k-th harmonic.
- The speed of propagation of k-th harmonic is a which is *independent* of the wave number k.
- Similarly for the scheme we can write
$$\gamma_k = |\gamma_k|\mathrm{e}^{-\mathrm{i}\omega_k \Delta t} = |\gamma_k|\mathrm{e}^{-\mathrm{i}\frac{\omega_k}{k}\lambda\phi_k}$$
- The quantity $\frac{\omega_k}{k}$ measures the numerical *propagation speed* of the k-th harmonic, which is not equal to the exact propagation speed which is a for all harmonics. This means that different harmonics can propagate with different speeds, which is known as dispersion error. This can be measured by
$$\epsilon_d(k) = \frac{\omega_k}{ak} = \frac{\omega_k h}{a\phi_k} =: \epsilon_d(\phi_k, \text{CFL}), \qquad -\pi \le \phi_k \le +\pi$$
- Ideally, we would like to have $\epsilon_d(k) \approx 1$ for all wave numbers k that can be represented on the grid.


Amplification error FE/C
Amplification error BE/C
Dispersion error FE/C
Dispersion error BE/C
Amplification error LF
Amplification error U
Dispersion error LF
Dispersion error U
Amplification error LW
Dispersion error LW
κ Δ x
ε_φ
CFL = 0.25
CFL = 0.50
CFL = 0.75
CFL = 1.00

$$l = 20h$$

$$\phi_k = \frac{\pi}{10}$$



$$l = 8h$$

$$\phi_k = \frac{\pi}{4}$$

Fig. 12.7. Numerical solution of the convective transport equation of a sinusoidal wave packet with different wavelengths ($l = 20h$ at the left, $l = 8h$ at the right) obtained with different numerical schemes. The numerical solution for $t = 1$ is displayed in solid line, while the exact solution at the same time instant is displayed in etched line

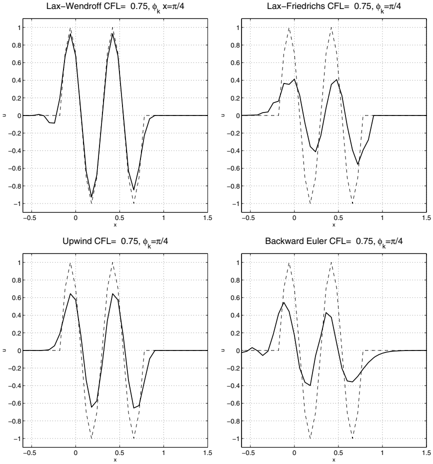


Fig. 12.8. Numerical solution of the convective transport of a packet of sinusoidal waves. The solid line represents the solution after 8 time steps. The etched line represents the corresponding exact solution at the same time level

Modified PDE/Equivalent equation

The numerical solution u_j^n is known at the grid points x_j and t^n only, and in general it is different from the exact solution $u(x,t)$

$$u_j^n \neq u(x_j, t^n)$$

The exact solution does not exactly satisfy the numerical scheme, since there is some truncation error. E.g., for FTBS scheme

$$\frac{u(x_j, t^{n+1}) - u(x_j, t^n)}{\Delta t} + a\frac{u(x_j, t^n) - u(x_{j-1}, t^n)}{h} = \tau_j^n \neq 0$$

Can we find a smooth function $v(x,t)$ such that

$$v(x_j, t^n) = u_j^n$$

and $v(x,t)$ exactly satisfies the numerical scheme ?

$$\frac{v(x_j, t^{n+1}) - v(x_j, t^n)}{\Delta t} + a\frac{v(x_j, t^n) - v(x_{j-1}, t^n)}{h} = 0$$

What is the PDE satisfied by $v(x,t)$?

MPDE for Upwind scheme

We begin with the Upwind scheme ($a > 0$)

$$\frac{v(x, t+\Delta t) - v(x,t)}{\Delta t} + a\frac{v(x,t) - v(x-h,t)}{h} = 0$$

and do a Taylor expansion about (x, t)

$$\frac{v(x, t+\Delta t) - v(x,t)}{\Delta t} = v_t + \frac{\Delta t}{2} v_{tt} + \frac{\Delta t^2}{6} v_{ttt} + \mathcal{O}\left(\Delta t^3\right)$$

$$a\frac{v(x,t) - v(x-h,t)}{h} = a v_x - \frac{ah}{2} v_{xx} + \frac{ah^2}{6} v_{xxx} + \mathcal{O}\left(h^3\right)$$

which leads to

$$v_t + a v_x = R + \mathcal{O}\left(\Delta t^3 + h^3\right)$$

with

$$R = \frac{1}{2}(ahv_{xx} - \Delta t v_{tt}) - \frac{1}{6}(ah^2 v_{xxx} + \Delta t^2 v_{ttt})$$

We want to write R in terms of x-derivatives. Differentiating wrt t and x

$$v_{tt} + a v_{xt} = R_t + \mathcal{O}\left(\Delta t^3 + h^3\right), \qquad v_{xt} + a v_{xx} = R_x + \mathcal{O}\left(\Delta t^3 + h^3\right)$$

MPDE for Upwind scheme

This gives

$$v_t + a v_x = \mu v_{xx} - \frac{1}{6}(a h^2 v_{xxx} + \Delta t^2 v_{ttt}) - \frac{\Delta t}{2}(R_t - a R_x) + \mathcal{O}\left(\Delta t^3 + h^3\right)$$

where

$$\boxed{\mu = \frac{1}{2} a h (1 - a\lambda)}, \qquad \mu = \mathcal{O}\left(h\right)$$

We now want to replace v_{ttt} term.

$$\begin{aligned} v_{ttt} &= a^2 v_{xxt} + R_{tt} - a R_{xt} + \mathcal{O}\left(\Delta t^3 + h^3\right) \\ &= -a^3 v_{xxx} + a^2 R_{xx} + R_{tt} - a R_{xt} + \mathcal{O}\left(\Delta t^3 + h^3\right) \end{aligned}$$

Moreover

$$R_t = \frac{1}{2} a h v_{xxt} - \frac{\Delta t}{2} v_{ttt} - \frac{a h^2}{6} v_{xxxt} - \frac{\Delta t^2}{6} v_{tttt}$$

$$R_x = \frac{1}{2} a h v_{xxx} - \frac{\Delta t}{2} v_{ttx} - \frac{a h^2}{6} v_{xxxx} - \frac{\Delta t^2}{6} v_{tttx}$$

MPDE for Upwind scheme

We get

$$\begin{aligned} v_t + a v_x &= \mu v_{xx} - \frac{a h^2}{6}\left[1 - \frac{a^2 \Delta t^2}{h^2} - \frac{3 a \Delta t}{2h}\right] v_{xxx} \\ &+ \frac{\Delta t}{4}\left(\Delta t v_{ttt} - a h v_{xxt} - a \Delta t v_{ttx}\right) \\ &+ \frac{\Delta t}{12}\left(\Delta t^2 v_{tttt} - a \Delta t^2 v_{tttx} + a h^2 v_{xxxt} - a^2 h^2 v_{xxxx}\right) \\ &- \frac{a^2 \Delta t^2}{6} R_{xx} - \frac{\Delta t^2}{6} R_{tt} + \frac{a \Delta t^2}{6} R_{xt} + \mathcal{O}\left(\Delta t^3 + h^3\right) \end{aligned}$$

Due to CFL condition, $\Delta t = \mathcal{O}(h)$ and the terms in blue are $\mathcal{O}\left(h^3\right)$. The terms in red can be written as

$$\begin{aligned} v_{ttt} &= -a^3 v_{xxx} + \mathcal{O}(h) \\ v_{ttx} &= +a^2 v_{xxx} + \mathcal{O}(h) \\ v_{xxt} &= -a v_{xxx} + \mathcal{O}(h) \end{aligned}$$

MPDE for Upwind scheme

leads to the MPDE

$$\boxed{v_t + a v_x = \mu v_{xx} + \nu v_{xxx} + \mathcal{O}\left(h^3\right)}$$

where

$$\boxed{\nu = -\frac{ah^2}{6}(1 - 3a\lambda + 2a^2\lambda^2)}, \qquad \nu = \mathcal{O}\left(h^2\right)$$

MPDE: LF, LW, FTCS schemes

Upto $\mathcal{O}\left(h^2\right)$ we can write the MPDE as

$$v_t + av_x = \underbrace{\mu v_{xx}}_{\mathcal{O}(h)} + \underbrace{\nu v_{xxx}}_{\mathcal{O}(h^2)}$$

μ is called the ***numerical viscosity***.

- Lax-Friedrichs scheme

$$\mu = \frac{h^2}{2\Delta t}(1 - a^2\lambda^2), \qquad \nu = \frac{ah^2}{3}(1 - a^2\lambda^2)$$

- Law-Wendroff scheme

$$\mu = 0, \qquad \nu = \frac{ah^2}{6}(a^2\lambda^2 - 1)$$

- FTCS scheme

$$\mu = -\frac{a^2\Delta t}{2} < 0$$

Meaning of MPDE

The MPDE gives us information about dissipation and dispersion properties of the numerical scheme. If we consider

$$\begin{cases} v_t + a v_x = \mu v_{xx} + \nu v_{xxx}, & x \in \mathbb{R}, \quad t > 0 \\ v(x,0) = \mathrm{e}^{\mathrm{i}kx} \end{cases}$$

its solution is

$$v(x,t) = \mathrm{e}^{-\mu k^2 t} \mathrm{e}^{\mathrm{i}k[x-(a+\nu k^2)t]}$$

The exact solution of the original PDE ($\mu = 0, \nu = 0$) is

$$u(x,t) = \mathrm{e}^{\mathrm{i}k(x-at)} \qquad \text{(no dissipation, no dispersion)}$$

- If $\mu > 0$ then v is dissipated with time, indicating stability. For the numerical schemes like Upwind and LF, we see that $\mu > 0$ iff the CFL condition $|a|\lambda \leq 1$ is satisfied. LW scheme does not have any dissipation (It can have higher order dissipation, which would be much weaker.)
- If $\nu = 0$ then every harmonic travels with speed a. But if $\nu \neq 0$, then the k-th harmonic travels with speed $a + \nu k^2$ which leads to dispersion error.

MPDE example: Upwind scheme

PDE and MPDE for upwind scheme (we take $a = 1$)

$$u_t + au_x = 0, \qquad v_t + av_x = \mu v_{xx}, \quad \mu = \frac{1}{2}|a|h(1 - |a|\lambda)$$

Discontinuous initial condition (H is Heaviside function)

$$u_0(x) = 2[1 - H(x)]$$

Exact solution

$$u(x,t) = 2[1 - H(x - at)], \qquad v(x,t) = \text{erfc}\left(\frac{x - at}{\sqrt{4\mu t}}\right)$$

Numerical example

$$a = 1, \qquad h = 0.05, \qquad \Delta t = 0.04, \qquad |a|\lambda = 0.8, \qquad \mu = 5 \times 10^{-4}$$

MPDE example: Upwind scheme

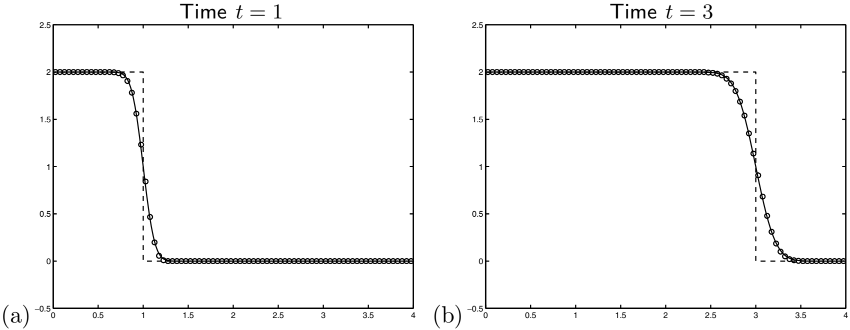


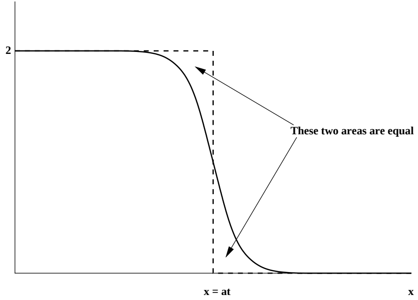
Fig. 8.3. Dashed line: exact solution to the advection equation. Points: numerical solution obtained with the upwind method. Solid line: exact solution to the modified equation (8.44). (a) At time $t = 1$. (b) At time $t = 3$. [book/chap8/modeqn]

Numerical solution is very close to the solution of MPDE !!!

MPDE example: Upwind scheme

Error in numerical solution can be approximated as

$$\begin{aligned} & \|u(t) - v(t)\|_1 \\ = & \int_{-\infty}^{\infty} |u(x,t) - v(x,t)| \mathrm{d}x \\ = & \int_{-\infty}^{at} |v(x,t) - 2| \mathrm{d}x + \int_{at}^{\infty} |v(x,t)| \mathrm{d}x \end{aligned}$$



Hence the 1-norm error is

$$\begin{aligned} \|v(t) - u(t)\|_1 &= 2 \int_{at}^{\infty} |v(x,t)| \mathrm{d}x \\ &= 2 \int_{at}^{\infty} \operatorname{erfc}\left(\frac{x - at}{\sqrt{4\mu t}} \right) \mathrm{d}x \\ &= 2\sqrt{4\mu t} \int_{0}^{\infty} \operatorname{erfc}(z) \mathrm{d}z \\ &= C_1 \sqrt{\mu t} \end{aligned}$$

MPDE example: Upwind scheme

for some constant C_1 independent of μ, Δx and t. Using the expression for μ we get the following estimate for the error

$$\|v(t) - u(t)\|_1 \approx C_2 t^{\frac{1}{2}} h^{\frac{1}{2}}$$

as $h \to 0$ with $\Delta t/h$ fixed. This indicates that the 1-norm of the error decays only like $h^{\frac{1}{2}}$ even though the method is formally *first order accurate* based on the local truncation error, which is valid for smooth solutions only. We also see that the thickness of the transition zone (shock structure) increases with time like $t^{\frac{1}{2}}$.

Remark:

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \mathrm{e}^{-z^2} \mathrm{d}z$$

Boundary conditions

- Most of the schemes we have studied till now can be written as

$$u_j^{n+1} = H(u_{j-1}^n, u_j^n, u_{j+1}^n)$$

  These are refered to as
  - two time level $(n, n+1)$
  - three point $(j-1, j, j+1)$

  schemes.
- Consider a grid of $N+1$ points indexed as $j = 0, 1, 2, \ldots, N$
- At $j = 0$ and/or $j = N$ we may need boundary conditions.

Periodic boundary condition

- At left boundary: $j = 0$

$$u_0^{n+1} = H(u_{N-1}^n, u_0^n, u_1^n)$$

- For $j = 1, 2, \ldots, N-1$

$$u_j^{n+1} = H(u_{j-1}^n, u_j^n, u_{j+1}^n)$$

- At right boundary: $j = N$

$$u_N^{n+1} = u_0^{n+1}$$

Matlab code is `lin_hyp_1d_periodic.m` and contains backward difference, forward difference and Lax-Wendroff schemes.

Dirichlet boundary condition

Appropriate boundary conditions must come from considerations of the well-posedness of the PDE problem.

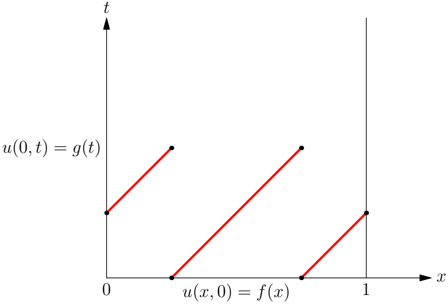
$$u_t + au_x = 0, \qquad a > 0, \qquad x \in (0,1)$$

<u>Method of characteristics</u>:

Need boundary conditions on

$$t = 0 \quad \text{and} \quad x = 0$$

for the solution to be fully determined.



Dirichlet BC: Numerical implementation

- At left boundary: $j = 0$

$$u_0^{n+1} = g(t^{n+1})$$

- For $j = 1, 2, \ldots, N-1$

$$u_j^{n+1} = H(u_{j-1}^n, u_j^n, u_{j+1}^n)$$

- At right boundary: $j = N$
  We cannot apply a three point scheme; use backward difference scheme

$$u_N^{n+1} = (1 - a\lambda)u_N^n + a\lambda u_{N-1}^n$$

  or do linear extrapolation of the solution from the interior

$$u_N^{n+1} = u_{N-1}^{n+1} + h\frac{u_{N-1}^{n+1} - u_{N-2}^{n+1}}{h} = 2u_{N-1}^{n+1} - u_{N-2}^{n+1}$$

Numerical Example for discontinuous solution

PDE

$$u_t + au_x = 0, \qquad x \in (0,1), \quad t \in (0, \frac{1}{4})$$

Discontinuous initial condition

$$u(x,0) = 2[1 - H(x - 1/2)]$$

Matlab code `lin_hyp_1d_disc.m`. Try with $N = 100$ and CFL=0.5

- Backward difference: `lin_hyp_1d_disc(100, 0.5, 'bd')`
- Forward difference: `lin_hyp_1d_disc(100, 0.5, 'fd')`
- Lax-Wendroff: `lin_hyp_1d_disc(100, 0.5, 'lw')`
- Also try with $N = 1000$ grid points

Remark: In this program, we do not update the solution at the first and last grid point. The final time $t = \frac{1}{4}$ is small enough that the discontinuity does not reach the boundary.