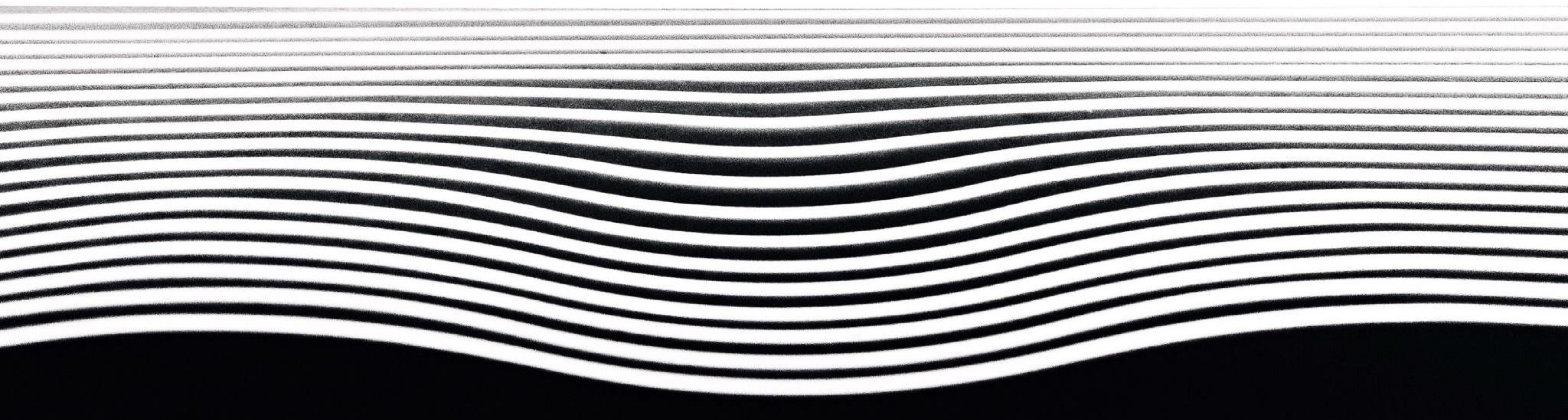
Numerical methods for conservation laws
9: Error analysis of finite-difference schemes



Recall the definition of our grid.

$$x_j = j \cdot h \qquad j = 0, 1, \ldots, N, \qquad N \cdot h = L$$

$$t^n = n \cdot k, \qquad n = 0, 1, \ldots, K, \qquad K \cdot k = T$$

Suppose that U^n and u are the numerical and the exact solution, respectively. What about

$$|u(x_j, t^n) - U_j^n| \xrightarrow[h,k \searrow 0]{} 0 \quad ?$$

that is, u has no jumps

Let's assume u is defined pointwise for demonstrative purposes.

We define a sequence based on the exact solution

$$u(t^n)_j := u(x_j, t^n)$$

and consider

$$u(t^{n+1})_j = G(u(t^n))_j + k\,T_j^n$$

Here, we define the truncation error

$$T_j^n = \frac{1}{k}\left(u(t^{n+1})_j - G(u(t^n))_j \right)$$

We want the truncation error to vanish as $h, k \searrow 0$.

For the error analysis,

1. Assume that u exists and is suff. smooth
2. Obtain T_j^n via Taylor expansion of u.

Example: Transport Equation

$$u_t + c u_x = 0, \qquad c > 0, \qquad u(x,0) = u_0(x)$$

Consider the FTBS scheme

$$U_j^{n+1} = G(U^n)_j = U_j^n - \frac{k}{h} c \left(U_j^n - U_{j-1}^n \right)$$

We perform Taylor expansions around (x_j, t^n)

$$u(x_j, t^{n+1}) = u(x_j, t^n) + k\, \partial_t u(x_j, t^n) + \frac{k^2}{2} \partial_{tt}^2 u(x_j, t^n) + \mathcal{O}(k^3)$$

$$u(x_{j-1}, t^n) = u(x_j, t^n) - h\, \partial_x u(x_j, t^n) + \frac{h^2}{2} \partial_{xx}^2 u(x_j, t^n) + \mathcal{O}(h^3)$$

Example: Transport Equation

Consequently,

$$T^n_j = \frac{1}{k}\Big(u(t^{n+1})_j - G(u(t^n))_j \Big)$$

$$= \frac{1}{k}\Big(u + k\,\partial_t u + \frac{k}{2}\partial_{tt}^2 u + \mathcal{O}(k^3)$$
$$- u + \frac{k}{h}c\big(u - u + h\,\partial_x u - \frac{h^2}{2}\partial_{xx}^2 u + \mathcal{O}(h^3)\big)\Big)$$

$$= \frac{1}{k}\Big(k\underbrace{\big(\partial_t u + c\,\partial_x u\big)}_{=0} + \frac{k^2}{2}\partial_{tt}^2 u - khc\,\partial_{xx}^2 u + \mathcal{O}(k^3 + h^3)\Big)$$

Example: Transport Equation

Hence $T_j^h = \frac{k}{2} \partial_{tt}^2 u - c h \partial_{xx}^2 u + \text{H.O.T.}$

$=: O(h + k)$

The truncation error for FTBS goes to zero as h and k vanish. This error analysis requires additional smoothness of u in time and space.

Similarly, $O(h + k)$ for FTFS and $O(h^2 + k)$ for FTCS.

We have seen that this is not enough. For example, FTFS quickly blows up due to the lack of upwinding.

Recall that

$$U_j^{n+1} = G(U^n)_j \qquad u(t^{n+1}, x_j) = G(u(t^n)_j) + kT_j^n$$

We define the local solution error, and express it in terms of the truncation error

$$\mathcal{E}_j^n := u(x_j, t^n) - U_j^n$$

Let us assume that G is linear in U. Via iteration we compute

$$\mathcal{E}_j^0 = u(x_j, 0) - U_j^0 = 0$$

$$\mathcal{E}_j^1 = G(\mathcal{E}^0)_j + kT_j^0$$

$$\mathcal{E}_j^2 = G(\mathcal{E}^1)_j + kT_j^1$$

$$= G^2(\mathcal{E}^0)_j + k(G(T^0)_j + T_j^1)$$

More generally, the following pattern emerges

$$\varepsilon_j^n = G^n(\varepsilon_j^0) + k\sum_{i=0}^{n-1} G^{n-i-1}(T^i)_j$$

Suppose that we have some norm ||·|| in which we want to measure the error. Then

$$\|\varepsilon^n\| \leq \|G^n(\varepsilon^0)\| + k\sum_{i=0}^{n-1}\|G^{n-i-1}(T^i)\|$$

$$\leq |||G^n||| \cdot \|\varepsilon^0\| + \underbrace{kn}_{\leq T} \cdot \max_i |||G^{n-i-1}||| \cdot \|T^i\|$$

Here

$$|||G||| = \sup_{U \neq 0} \frac{\|G(U)\|}{\|U\|} \quad \text{(induced operator norm)}$$

Consequently, the total error vanishes if

1) $\|\varepsilon^0\| \longrightarrow 0$ for $h \to 0$

The initial data are approximated

2) $\|T^i\| \longrightarrow 0$ for $k, h \to 0$

The truncation error vanishes

3) $|||G^n||| \leq C$

The scheme is stable.

Remarks:

- We say that the scheme has <u>order</u> (p, q) if

$$\| T^i \| = \mathcal{O}(h^p + k^q)$$

- A sufficient condition for stability is

$$|||G||| \leq 1 + \alpha k$$

Then:

$$|||G^n||| \leq |||G|||^n \leq (1 + \alpha k)^n$$

$$\leq e^{\alpha k n} \leq e^{\alpha T} := C$$

Example: monotone schemes are stable

$$U_j^{n+1} = G(U^n)_j \qquad u(t^{n+1}, x_j) = G(u(t^n)_j) + k T_j^n$$

$$\mathcal{E}_j^{n+1} = G(u(t^n)_j) - G(U^n)_j + k T_j^n$$

We measure the error in the discrete L^1-norm

$$\| \mathcal{E}^{n+1} \| = \sum_j h \cdot | \mathcal{E}_j^{n+1} |$$

$$\| \mathcal{E}^{n+1} \| \leq \| G(u(t^n)_j) - G(U^n)_j \| + k \cdot \| T_j^n \|$$

$$\|\varepsilon^{n+1}\| \leq \| G(u(t^n)_j) - G(U^n)_j \| + k \cdot \|T_j^n\|$$

$$\leq \underbrace{\| u(t^n)_j - U_j^n \|}_{= \|\varepsilon^n\|} + k \cdot \|T_j^n\|$$

In that sense, monotone schemes are stable.

We find easily, like before

$$\begin{aligned}
\|\varepsilon^n\| &\leq \|\varepsilon^{n-1}\| + k\|T_j^{n-1}\| \\
&\leq \|\varepsilon^{n-2}\| + k\|T_j^{n-2}\| + k\|T_j^{n-1}\| \\
&\leq \dots \\
&\leq \|\varepsilon^0\| + T \cdot \max_{0 \leq l \leq n-1} \|T_j^l\|
\end{aligned}$$

While monotone schemes have many desirable properties, they have an important drawback: they at most of first order.

Gudonov Barrier Theorem (1959)
A linear monotonicity-preserving scheme is at most first order.
A monotone scheme (possibly nonlinear) is at most first order.

$$\begin{tikzcd}
\text{Monotone} \arrow[r, Rightarrow] & L^1 \text{ contraction} \arrow[r, Rightarrow] & \text{TVD} \arrow[r, Rightarrow] & \text{MP} \arrow[lll, bend left, "\text{Linear}"]
\end{tikzcd}$$

Consequence: if we want stable higher-order schemes, then we need to study nonlinear schemes that are not monotone.

Example: Transport Equation

$$u_t + c u_x = 0, \qquad c > 0, \qquad u(x,0) = u_0(x)$$

Notice that

$$u_{tt} = -c u_{xt} = -c u_{tx} = c^2 u_{xx}$$

More generally,

$$\partial_t^r u = (-c)^r \partial_x^r u$$

What order are linear monotone schemes?

$$U_j^{n+1} = G(U^n)_j = \sum_\ell c_\ell U_{j+\ell}^n$$

Recall the Taylor expansions

$$u(x_j, t^{n+1}) = u + k\,\partial_t u + \frac{k^2}{2}\partial_{tt}^2 u + \dots \mathcal{O}(k^3)$$

$$u(x_{j+\ell}, t^n) = u + h\ell \cdot \partial_x u + \frac{(h\ell)^2}{2}\partial_{xx}^2 u + \dots \mathcal{O}(h^3)$$

Hence

$$u(x_j, t^{n+1}) - G(u(t^n))_j =: kT_j^n$$

$$= u - ck\,\partial_x u + \frac{ck^2}{2}\partial_{xx}^2 u + \dots \mathcal{O}(k^3)$$

$$- \left(\sum c_\ell\right) u - \left(\sum h\ell c_\ell\right)\partial_x u - \frac{1}{2}\left(\sum (\ell h)^2 c_\ell\right)\partial_{xx}^2 u - \dots \mathcal{O}(h^3)$$

$$= \left(1 - \sum_\ell c_\ell\right) u + \left(-ck - h\sum_\ell \ell c_\ell\right)\partial_x u$$

$$+ \left(\frac{c^2k^2}{2} - \frac{h^2}{2}\sum_\ell \ell^2 c_\ell\right)\partial_{xx}^2 u + \dots \mathcal{O}(h^3 + k^3)$$

If we want a quadratic order scheme, then we need the leading coefficients to be zero. That is,

$$\sum_\ell c_\ell = 1, \qquad \sum_\ell \ell \cdot c_\ell = -c\frac{k}{h}, \qquad \sum_\ell \ell^2 c_\ell = \left(\frac{ck}{h}\right)^2$$

Now suppose that the scheme is monotone / MP. Then $c_\ell \geq 0$.

We write $c_\ell = \gamma_\ell^2$, using that we can take square roots. Now

$$\left(\sum_\ell \gamma_\ell^2\right)\left(\sum_\ell \ell^2 \gamma_\ell^2\right) = \left(\sum_\ell \ell \gamma_\ell^2\right)^2$$

$$\langle a, b\rangle \cdot \langle b, b\rangle = \langle a, b\rangle^2$$

The equation

$$\langle a, a \rangle \langle b, b \rangle = \langle a, b \rangle^2$$

is an instance of the Cauchy-Schwarz inequality with equality [sic!].

But that means the vectors a and b are colinear, that is, $b = J \cdot a$.
Concretely, this means

$$\mathcal{T}_e \,/\, \ell \cdot \mathcal{T}_e = J \quad (\text{const})$$

Clearly, this can't be.

Similar calculations apply to general (nonlinear) monotone schemes.

As a coda to this discussion, we revisit the CFL condition.

Example: FTBS scheme for the transport equation

$$U_j^{n+1} = U_j^n - \underbrace{c\frac{k}{h}}_{=:\lambda}\left(U_j^n - U_{j-1}^n\right) = (1-\lambda)U_j^n + \lambda U_{j-1}^n$$

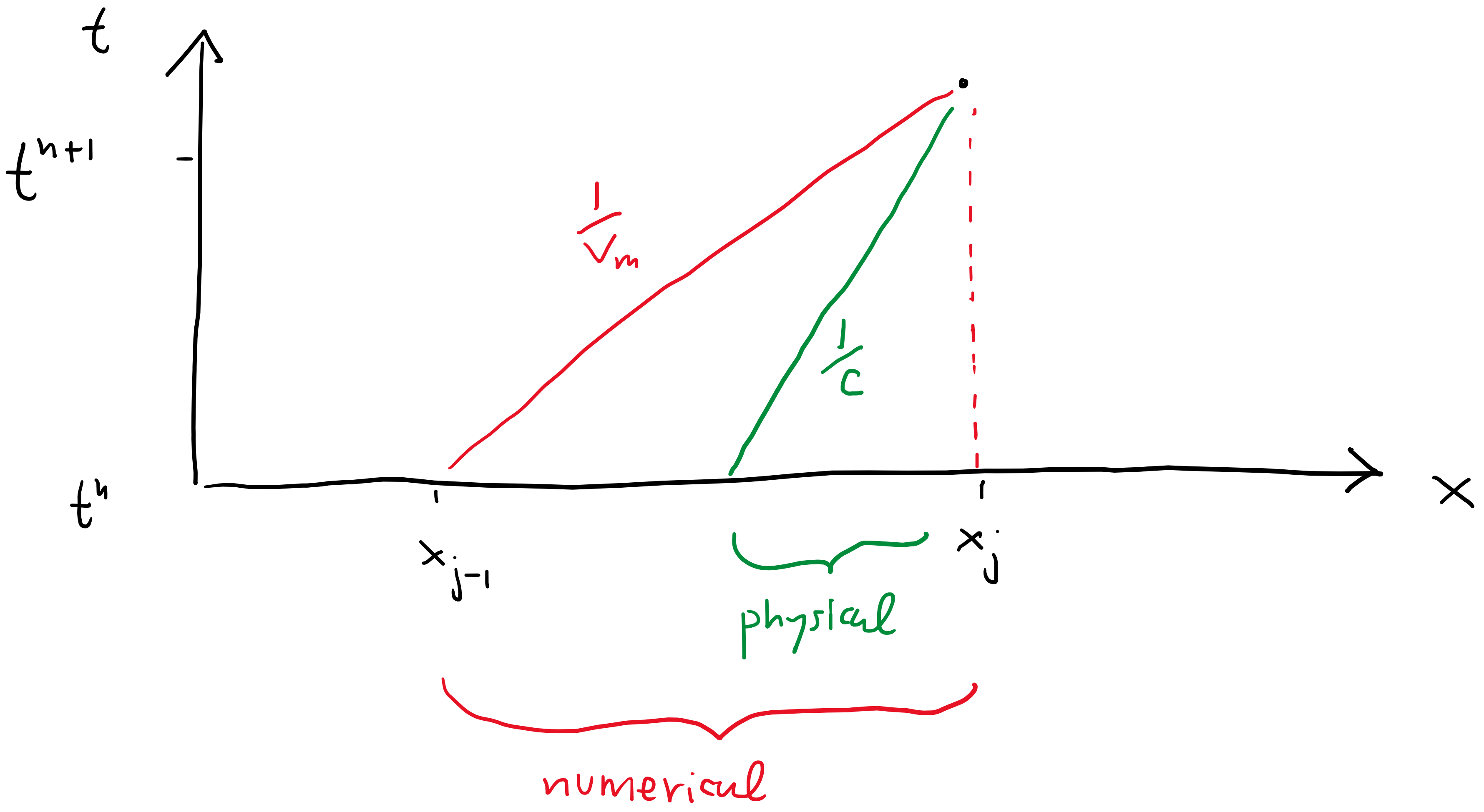
Hence

$$\left|U_j^{n+1}\right| \leq \left|U_j^n\right| \cdot |1-\lambda| + \left|U_{j-1}^n\right| \cdot |\lambda| \leq \|U^n\|_\infty(|1-\lambda| + |\lambda|)$$

The scheme is a contraction if $\lambda \in [0,1]$. We define the mesh velocity $v_m = h/k$

For that we need $0 \leq c \leq v_m$

Graphically, the physical domain of dependence must lie within the numerical domain of dependence.



Example: FTCS scheme for the transport equation

$$U_j^{n+1} = U_j^n - c\,\frac{k}{2h}\left(U_{j+1}^n - U_{j-1}^n\right)$$

Consider a modification:

$$U_j^{n+1} = \frac{U_{j+1}^n + U_{j-1}^n}{2} - c\,\frac{k}{2h}\left(U_{j+1}^n - U_{j-1}^n\right)$$

This is a conservative scheme with flux

$$F_{j+\frac{1}{2}}^n = \frac{c}{2}\left(U_{j+1}^n + U_j^n\right) - \frac{k}{2h}\left(U_{j+1}^n - U_j^n\right)$$

when $|c/v_m| \leq 1$.

Example: Lax-Friedrichs flux and variations

This is a conservative scheme with flux

$$F(U,V) = \frac{1}{2}\left(f(U) + f(V) \right) - \frac{\alpha}{2}(V-U)$$

LF

$$F(U,V) = \frac{1}{2}\left(f(U) + f(V) \right) - \frac{h}{2k}(V-U)$$

LF with upper estimate

$$F(U,V) = \frac{1}{2}\left(f(U) + f(V) \right) - \frac{\tilde{\alpha}}{2}(V-U)$$

local LF / Rusanov

Here, $\alpha = \max|f'|$ and $\tilde{\alpha} = \max\{|f'(U)|, |f'(V)|\}$.

Monotone when $\alpha \leq \frac{k}{h}$ and $\tilde{\alpha} \leq \frac{k}{h}$.

$\tilde{\alpha} \leq \alpha$