

Chap 3: Finite difference methods

for PDEs

3.1 Basic notions of discretization

while we are interested in the continuum limit of PDEs, discretization of some kind is required for most numerical approaches

$$\begin{array}{ccc} X_0 & & Y_0 \\ U & & V \\ L : X & \xrightarrow{\quad} & Y \\ (u & \mapsto & Lu = S) \\ \downarrow D_h^X & & \downarrow D_h^Y \\ L_h : X_h & \longrightarrow & Y_h \end{array}$$

continuum problem: $\boxed{Lu = S}$ continuum PDE
typically: $X_0 = C^0(\Omega), \Omega \subseteq \mathbb{R}^n$
 $X = \{u \in C^2(\Omega) \mid \sup_{\Omega} |Lu| < \infty\} \subset X_0$

regularity depends on differential operator L and the type of solution to be found

$$Y \subseteq Y_0 = C^0(\Omega)$$

Examples: $L = \Delta, \square, \partial_t - \partial_x f(\cdot), \dots$

discretized problem:

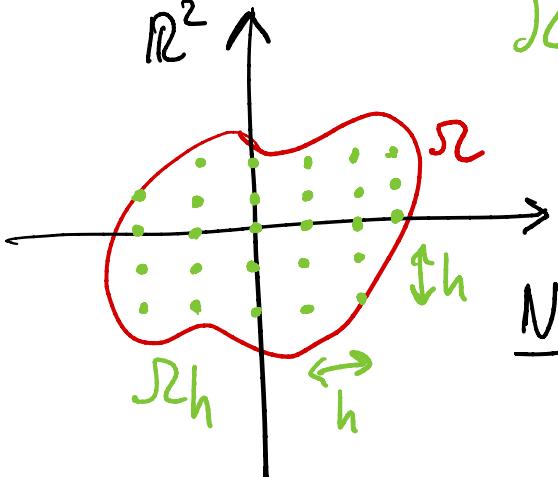
$$L_h u_h = S_h$$

discretized PDE

typically: $X_h = \{\text{grid functions}\}$

$$= \{v : \Omega_h \rightarrow \mathbb{R}^n\}$$

$$\Omega_h = \text{"grid"}, \text{e.g.: } \Omega \cap h\mathbb{Z}^n$$



$h = \text{discretization parameter}$

Notation:

$$x \rightarrow x_i \in \{x_1, \dots, x_{N_x}\}$$

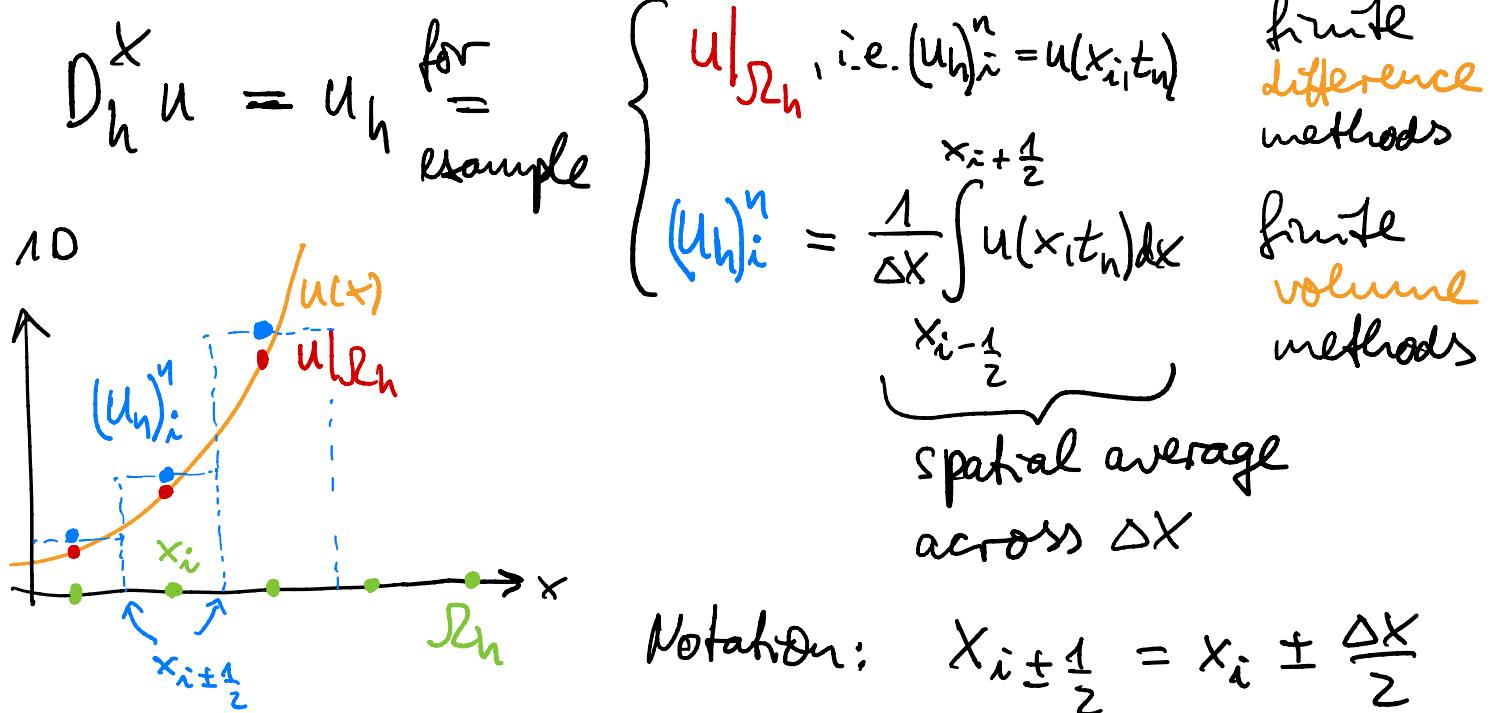
$$t \rightarrow t_n \in \{t_1, \dots, t_{N_t}\}$$

$$v \in X_h : v_i^n = v(x_i, t_n)$$

D_h^X : discretization operator

$$D_h^X : X \rightarrow X_h$$

$$u \mapsto D_h^X u \equiv u_h$$



Example: $\Omega = \mathbb{R} \times (0, \infty)$, $h = \Delta x = \Delta t$

$$\mathcal{S}_h = \{(h_i, h_n) \mid i \in \mathbb{Z}, n \in \mathbb{N}\}$$

Notation: $(x_i, t_n) \equiv (h_i, h_n)$

$$v_i^n \equiv v(x_i, t_n), \quad v \in X_h$$

3.2 Finite difference approximations

Discretize continuum PDE by discrete grid:

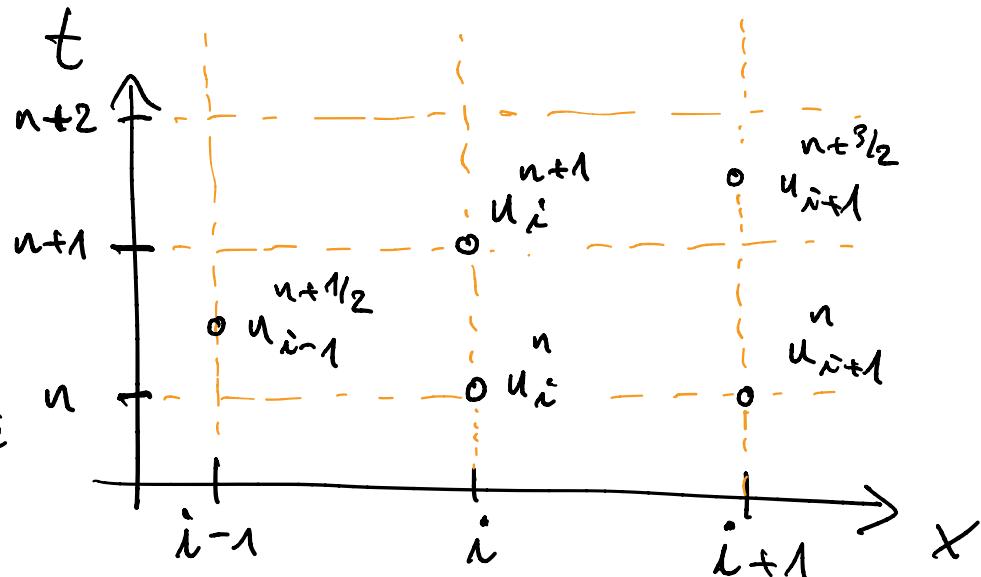
- **finite differences**: point values at grid points (or staggered)
- **finite volume**: discrete values $\hat{=}$ averages over finite volumes
(and later)

Consider $u = u(x, t)$ find $\mathcal{S}_h = \{(i\Delta x, n\Delta t)\}_{i, n \in \mathbb{Z}\}}$

Notation:

$$u(x_i, t^n) = u_j^n$$

$$u(x_i + \frac{\Delta x}{2}, t^n) = u_{i+\frac{1}{2}}^n$$



Note: in a finite difference approximation (FDA) of a PDE, different functions can be defined at different grid locations in which case the grid is called "staggered".

and "staggered grid function"

e.g. u defined at $i + \frac{\Delta x}{2}$ is $u_{i+\frac{1}{2}}^n$

3.2.1 Partial derivatives / differential operators

Consider Taylor expansion:

$$u_{i+1}^n = u_i^n + \left. \frac{\partial u}{\partial x} \right|_i \Delta x + \frac{1}{2} \left. \frac{\partial^2 u}{\partial x^2} \right|_i (\Delta x)^2 + O((\Delta x)^3)$$

$$\left[u(x_i + \Delta x) = u(x_i) + \sum_{k=1}^n \frac{(\Delta x)^k}{k!} u^{(k)}(x_i) + O((\Delta x)^{n+1}) \right]$$

Solve for $\frac{\partial u}{\partial x}$:

$$\left. \frac{\partial u}{\partial x} \right|_j = \frac{u_{i+1}^n - u_i^n}{\Delta x} - \frac{1}{2} \left. \frac{\partial^2 u}{\partial x^2} \right|_i \Delta x + O((\Delta x)^2)$$

(*)

1D

$$D_1^+ u = \frac{u_{i+1} - u_i^n}{\Delta x}$$

"forward difference approximation"

"stencil size": max distance to neighboring grid points involved

accuracy / truncation error:

$$D_1^+ u - \frac{\partial u}{\partial x} \stackrel{(*)}{=} \frac{1}{2} \frac{\partial u^2}{\partial x^2} \Big|_i \Delta x + O((\Delta x)^2) = O(\Delta x)$$

→ first-order accurate

Similarly:

$$D_1^- u = \frac{u_i^n - u_{i-1}^n}{\Delta x}$$

"backward difference approximation"

improve accuracy:

$$\Delta x_i^+ \equiv x_{i+1} - x_i, \quad \Delta x_i^- \equiv x_i - x_{i-1}$$

(can differ in general)

$$\Rightarrow \Delta x_i \equiv \frac{1}{2} (\Delta x_i^+ + \Delta x_i^-)$$

$$\text{forward: } u_{i+1}^n = u_i^n + \frac{\partial u}{\partial x}\Big|_i \Delta x_i^+ + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}\Big|_i (\Delta x_i^+)^2 + \dots$$

$$\text{backward: } u_{i-1}^n = u_i^n + \frac{\partial u}{\partial x}\Big|_i (-\Delta x_i^-) + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}\Big|_i (\Delta x_i^-)^2 + \dots$$

↓ subtract

$$u_{i+1}^n - u_{i-1}^n = \frac{\partial u}{\partial x}\Big|_i (\Delta x_i^+ + \Delta x_i^-) + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}\Big|_i [(\Delta x_i^+)^2 - (\Delta x_i^-)^2] + \dots$$

$$\text{MD } \frac{\partial u}{\partial x}\Big|_i = \frac{u_{i+1}^n - u_{i-1}^n}{2 \Delta x_i} - \frac{1}{2} \frac{\partial^2 u}{\partial x^2}\Big|_i \frac{(\Delta x_i^+)^2 - (\Delta x_i^-)^2}{2 \Delta x_i} + O((\Delta x_i)^2)$$

On a uniform grid:

$$D_h^0 u = \frac{u_{i+1}^n - u_{i-1}^n}{2 \Delta x} + O((\Delta x)^2)$$

"centered difference approximation"

Remark: By virtue of Taylor expansion we assume that the function u is continuously differentiable. Note that the limit

$$\lim_{\Delta x \rightarrow 0} \frac{u_{i+1}^n - u_{i-1}^n}{2 \Delta x}$$

can exist even if u is not differentiable
 (e.g. $u(x,t) = |x|$ and $\lim_{\Delta x \rightarrow 0} D_1^0 u = 0$)
 we shall explore discontinuous solutions
 later.

Higher-order derivatives:

intuitively: repeatedly apply 1st order derivatives

$$\begin{aligned}
 \text{Examples: } \frac{\partial^2 u}{\partial x^2} &\approx D_1^+ D_1^- u = \frac{1}{\Delta x} \left(D_1^+ u_i^n - D_1^- u_{i-1}^n \right) \\
 &= \frac{1}{\Delta x} \left(\frac{u_{i+1}^n - u_i^n}{\Delta x} - \frac{u_i^n - u_{i-1}^n}{\Delta x} \right) \\
 &= \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{(\Delta x)^2} \\
 &\equiv D^2 u
 \end{aligned}$$

$$\frac{\partial^2 u}{\partial x^2} \approx D_1^- D_1^+ u = \dots = D^2 u$$

$$\frac{\partial^2 u}{\partial x^2} \approx D_{1/2}^0 D_{1/2}^0 u = \frac{1}{\Delta x} D_{1/2}^{1/2} \left(u_{i+\frac{1}{2}}^n - u_{i-\frac{1}{2}}^n \right)$$

$$= \frac{1}{\Delta x} \left(\frac{u_{i+1}^n - u_i^n}{\Delta x} - \frac{u_i^n - u_{i-1}^n}{\Delta x} \right)$$

$$= D^2 u$$

General approach:

goal: compute approximation to $u^{(k)} = \frac{\partial^k u}{\partial x^k}$

based on a stencil of $n \geq k+1$ points

x_1, \dots, x_n

Example: want one-sided approximation

to $\frac{\partial u}{\partial x}|_i$ based on $u_i^n, u_{i-1}^n, u_{i-2}^n$

→ Taylor expansion

$$\rightarrow u_{i-1} = u_i - u^{(1)}|_i \Delta x + \frac{1}{2} u^{(2)}|_i (\Delta x)^2 - \frac{1}{6} u^{(3)}|_i (\Delta x)^3 + O(\Delta x^4)$$

$$\rightarrow u_{i-2} = u_i - u^{(1)}|_i (2\Delta x) + \frac{1}{2} u^{(2)}|_i (2\Delta x)^2 - \frac{1}{6} u^{(3)}|_i (2\Delta x)^3 + O(\Delta x^4)$$

Ausatz:

$$D_2 u|_i = c_1 u_i + \underline{c_2 u_{i-1}} + \underline{c_3 u_{i-2}}$$

$$= c_1 u_i + \underline{c_2 u_i - c_2 u^{(1)} \Delta x + c_2 \frac{1}{2} u^{(2)} (\Delta x)^2}$$

$$\underline{- c_2 \frac{1}{6} u^{(3)} (\Delta x)^3}$$

$$+ \underline{c_3 u_i - c_3 u^{(1)} (2\Delta x) + c_3 \frac{1}{2} u^{(2)} (2\Delta x)^2}$$

$$\underline{- c_3 \frac{1}{6} u^{(3)} (\Delta x)^3} + \dots$$

$$= (c_1 + c_2 + c_3) u_i - (c_2 + 2c_3) \Delta x u^{(1)}$$

$$+ \frac{1}{2} (c_2 + 4c_3) (\Delta x)^2 u^{(2)} - \frac{1}{6} (c_2 + 8c_3) (\Delta x)^3 u^{(3)}$$

+ ... ↑ note: coefficients are of the form

$$\frac{1}{(l-1)!} \sum_{m=1}^n c_m (x_m - x_i)^{l-1}$$

Need to agree with $u^{(l)}|_i$ to highest order possible:

$$\text{and } \left. \begin{array}{l} c_1 + c_2 + c_3 = 0 \\ -(c_2 + c_3) \Delta x = 1 \\ \frac{1}{2} (c_2 + 4c_3) (\Delta x)^2 = 0 \end{array} \right\} \quad \left. \begin{array}{l} c_1 = \frac{3}{2\Delta x} \\ c_2 = -\frac{3}{\Delta x} \\ c_3 = \frac{1}{2\Delta x} \end{array} \right.$$

(Note: higher coefficients 0 would lead to over-determined system)

$$\text{and } D_2 u|_i = \frac{1}{2\Delta x} (3u_i - 4u_{i-1} + u_{i-2})$$

Truncation error:

$$\begin{aligned} D_2 u|_{\bar{x}} - u^{(4)}|_{\bar{x}} &= -\frac{1}{6} (c_2 + 8c_3) (\Delta x)^3 u^{(3)}|_{\bar{x}} + \dots \\ &= -\frac{1}{6} \left(-\frac{2}{\Delta x} + \frac{4}{\Delta x} \right) (\Delta x)^3 u^{(3)}|_{\bar{x}} + \dots \\ &= -\frac{1}{3} (\Delta x)^2 u^{(3)}|_{\bar{x}} + \dots = \Theta(\Delta x^2) \end{aligned}$$

and 2nd order accurate

General procedure: Consider stencil $\{x_e\}_{e=1,\dots,n}$ around point $\bar{x} = x_i$ (may or may not be part of stencil), $n \geq k+1$. Consider n Taylor series $e=1,\dots,n$:

$$(*) \frac{u(x_e) - \underbrace{\tilde{u}(x_i)}_{\text{known}}}{\underbrace{\Delta x}_k} = u^{(1)}|_{\bar{x}} (x_e - x_i) + \dots + \frac{1}{k!} \underbrace{(x_e - x_i)^k u^{(k)}|_{\bar{x}}}_{\text{arrow}} + \Theta(\Delta x^k)$$

and linear system of n equations in k unknowns

$$\text{Ansatz: } u^{(k)}|_{\bar{x}} = c_1 u(x_1) + \dots + c_n u(x_n) + \Theta(\Delta x^k)$$

know that we can choose (see above)

$$(***) \frac{1}{(l-1)!} \sum_{m=1}^n c_m (x_m - x_i)^{l-1} = \begin{cases} 1, & l-1=k \\ 0, & \text{otherwise} \end{cases}$$

$$l=1,\dots,n$$

If points $\{x_l\}_{l=1,\dots,n}$ are distinct, (1) is an $n \times n$ non-singular system and thus has a unique solution. Can write:

$$\left(\begin{array}{cccc|c} 1 & \cdots & \cdots & 1 \\ \vdots & & & \vdots \\ (a_{lm})_{\substack{l=2,\dots,n \\ m=1,\dots,n}} & = & \frac{1}{(l-1)!} (x_m - x_i)^{l-1} \\ \vdots & & \vdots & \vdots \\ 1 & \cdots & \cdots & 1 \end{array} \right) \left(\begin{array}{c} c_1 \\ \vdots \\ c_n \end{array} \right) = \left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{array} \right)$$

$\equiv A \in \text{Mat}(n \times n) \quad c \quad b$

cancellation of lower order terms
 ← $k+1$
 cancellation of higher order terms

Remarks: 1) if $n \leq k$ there are too few points in the stencil, then RHS b and solution c both zero.

2) **Accuracy:** RHS and in Taylor expansions (2):

$$\left(\sum_{m=1}^n c_m (x_m - x_i)^{l-1} \right) u^{(l-1)}|_i = 0$$

$\nearrow l-1 \leq k$: necessary
 to cancel lower order terms
 and get $O(\Delta x)$ accuracy
 $\searrow l-1 > k$.

cancel higher
order terms and
obtain higher than
1st order accuracy

→ procedure is at least

$$O((\Delta x)^p), p = n-k \text{ accurate}$$

$p > n-k$ achievable if additional higher
order terms cancel (e.g. centered differences)

General: increasing stencil size increases
accuracy

3.2.2 Sample Discretizations

1) 1D wave equation, standard $\mathcal{O}(\Delta x^2)$

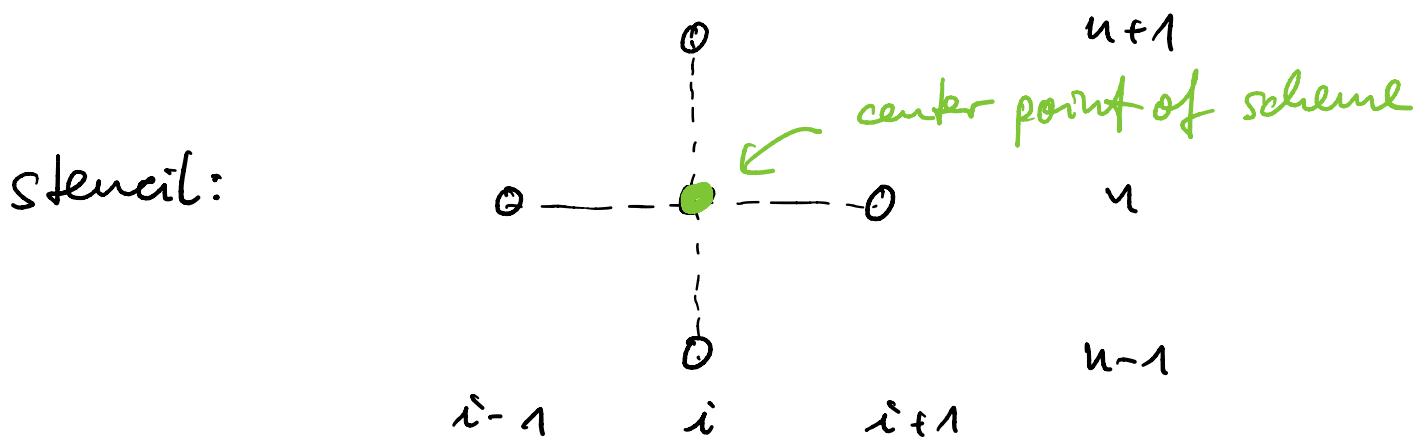
IBVP

$$\left\{ \begin{array}{l} u_{tt} - c^2 u_{xx} = 0, \quad 0 \leq x \leq 1, \quad t \geq 0 \\ u(x, 0) = u_0(x) \\ u_t(x, 0) = v_0(x) \\ u(0, t) = u(1, t) = 0 \end{array} \right. \begin{array}{l} \text{initial data} \\ \text{fixed (Dirichlet)} \\ \text{boundary conditions} \end{array}$$

Consider standard centered $\mathcal{O}(h^2)$ discretization
in space & time

$$\frac{\partial^2 u}{\partial t^2} \Big|_i^n = \frac{u_i^{n+1} - 2u_i^n + u_i^{n-1}}{(\Delta t)^2} + \mathcal{O}(\Delta t^2)$$

$$\frac{\partial^2 u}{\partial x^2} \Big|_i^n = \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{(\Delta x)^2} + \mathcal{O}(\Delta x^2)$$



1D FEA approximation to $O(\Delta x^2, \Delta t^2)$:

$$\frac{u_i^{n+1} - 2u_i^n + u_i^{n-1}}{(\Delta t)^2} = c^2 \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{(\Delta x)^2}$$

$$\left. \begin{array}{l} u_i^0 = u(x_i, 0) \\ u_i^1 = v(x_i, 0) \Delta t \end{array} \right\} \text{discretized initial data}$$

$$\left. \begin{array}{l} u_1^n = u_I^n = 0 \\ u_{x=0} = u_{x=1} \end{array} \right\} \text{discretized boundary condition}$$

explicitly solve for u^{n+1} :

$$u_i^{n+1} = 2u_i^n - u_i^{n-1} + \underbrace{\left(\frac{c \Delta t}{\Delta x} \right)^2}_{=\lambda} (u_{i+1}^n - 2u_i^n + u_{i-1}^n)$$

\Rightarrow linear system for unknowns $\{u_i^{n+1}\}_{i=1, \dots, I}$

write:

$$A u^{n+1} = b$$

$$\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 0 & \\ 0 & & & 1 \end{pmatrix} \begin{pmatrix} u_1^{n+1} \\ \vdots \\ u_I^{n+1} \end{pmatrix} = \begin{pmatrix} 2u_1^n - u_1^{n+1} + \lambda^2 () \\ \vdots \\ 2u_I^n - u_I^{n+1} + \lambda^2 () \end{pmatrix}$$

accuracy: $\Theta(\Delta x^2, \Delta t^2)$

A: diagonal and explicit scheme
 (will see: stable if $\lambda = \frac{c\Delta t}{\Delta x} \leq 1$ CFL condition)

2) 1D diffusion equation, Crank-Nicholson

IBVP

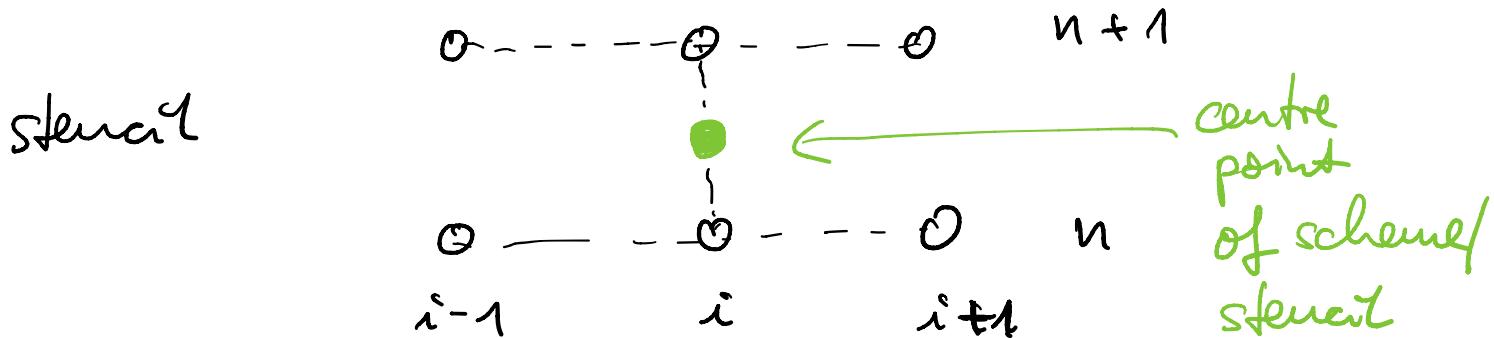
$$\left\{ \begin{array}{l} u_t - \sigma u_{xx} = 0, \quad 0 \leq x \leq 1, \quad t \geq 0 \\ u(x, 0) = u_0(x) \\ u(0, t) = u(1, t) = 0 \end{array} \right.$$

Consider: centered in time & space

(minimizes truncation error,
 minimizes instabilities)

$$u_t \Big|_i^{n+\frac{1}{2}} = \frac{u_i^{n+1} - u_i^n}{\Delta t} + \theta(\Delta t^2)$$

$$u_{xx} \Big|_i^n = \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{(\Delta x)^2} + \theta(\Delta x^2)$$



need to apply time averaging on RHS:

$$T^{n+\frac{1}{2}}(v) = \frac{1}{2}(v^{n+1} + v^n) = v^{n+\frac{1}{2}} + O(\Delta t^2)$$

and FDA approximation to $O((\Delta x)^2, (\Delta t)^2)$:

at $t = n + \frac{1}{2}$

$$\begin{aligned} \frac{u_i^{n+1} - u_i^n}{\Delta t} &= \sigma T^{n+1}(u_{xx}|_i) \quad i = 2, \dots, I \\ &= \sigma \frac{1}{2} \left[\frac{u_{i+1}^{n+1} - 2u_i^{n+1} + u_{i-1}^{n+1}}{(\Delta x)^2} \right. \\ &\quad \left. - \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{(\Delta x)^2} \right] \end{aligned}$$

and linear system for $\{u_i^{n+1}\}_{i=2, \dots, I-1}$

write as

$$A u^{n+1} \equiv a_+ u_{i+1}^{n+1} + a_0 u_i^{n+1} + a_- u_{i-1}^{n+1} = b, \quad i = 2, \dots, I-1$$

$$\begin{pmatrix} \frac{1}{\Delta t^2} + \frac{1}{\Delta x^2} & -\frac{1}{\Delta x^2} & & & \\ -\frac{1}{\Delta x^2} & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & -\frac{1}{\Delta x^2} \\ & & \ddots & \ddots & \\ 0 & & & -\frac{1}{\Delta x^2} & \frac{1}{\Delta t^2} + \frac{1}{\Delta x^2} \end{pmatrix} \begin{pmatrix} 0 \\ u_2^{n+1} \\ \vdots \\ u_{I-1}^{n+1} \end{pmatrix} = \begin{pmatrix} \left(\frac{1}{\Delta t^2} + \frac{1}{\Delta x^2} \right) u_i^n + \\ \frac{1}{2(\Delta x)^2} (u_{i+1}^n + u_{i-1}^n) \end{pmatrix}$$

and tridiagonal matrix and implicit scheme,
couples unknowns

u_{i+1}, u_i, u_{i-1} at
advanced time level
 $n+1$!

accuracy: $\theta(\Delta x^2, \Delta t^2)$

3) 1D advection equation

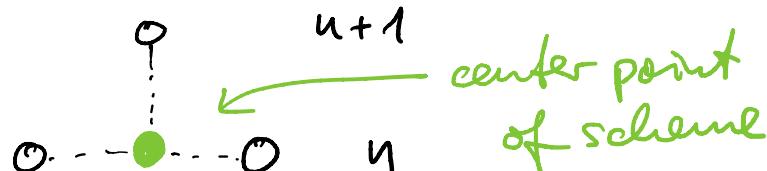
IBVP

$$\begin{cases} u_t + a u_x = 0 & 0 \leq x \leq 1, t \geq 0 \\ u(x, 0) = u_0(x) \\ u(0, t) = u(1, t) = 0 \end{cases}$$

(i) Consider simple forward in time, centered in space FDA:

$$u_x|_i^n \approx \frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x}, \quad u_t|_i^n = \frac{u_i^{n+1} - u_i^n}{\Delta t}$$

stencil



$i-1 \quad i \quad i+1$

and

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + \frac{a}{2\Delta x} (u_{i+1}^n - u_{i-1}^n) = 0$$

$$\Leftrightarrow u_i^{n+1} = u_i^n - \underbrace{\frac{a\Delta t}{2\Delta x}}_{=\lambda} (u_{i+1}^n - u_{i-1}^n)$$

linear system for u^{n+1} : $A u^{n+1} = b$

$$\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} \begin{pmatrix} u_1^{n+1} \\ \vdots \\ u_I^{n+1} \end{pmatrix} = \begin{pmatrix} u_1^n - \lambda (u_{i+1}^n - u_{i-1}^n) \\ \vdots \\ u_I^n - \lambda (u_{i+1}^n - u_{i-1}^n) \end{pmatrix}$$

and explicit scheme $\Theta(\Delta t, \Delta x^2)$

(will see: unstable for any fixed $\frac{\Delta t}{\Delta x}$)

(ii) Lax-Friedrichs method

$$\text{replace } u_i^n \rightarrow \frac{1}{2} (u_{i-1}^n + u_{i+1}^n)$$

$$\text{and } u_i^{n+1} = \frac{1}{2} (u_{i+1}^n + u_{i-1}^n) - \frac{a\Delta t}{2\Delta x} (u_{i+1}^n - u_{i-1}^n)$$

accuracy: $\Theta(\Delta t, \Delta x^2)$

(will see: stable if $\left| \frac{a\Delta t}{\Delta x} \right| \leq 1$)

3.3 Consistency, stability, convergence

General problem: Does the solution to the discrete problem provide an approximation to the continuous problem?

$$\|u - u_h\| \xrightarrow{h \rightarrow 0} 0 \quad \text{for some appropriate norm } \|\cdot\|?$$

no need to understand how u_h varies with h

no need basic concepts of numerical analysis

Problem setting: Consider numerical scheme

of the form

$$u^{n+1} = G(S_+, S_-) u^n$$

column vector
containing sufficient
unknowns to write
problem in 1st order
in time form

update
operator,
polynomial
in S_+, S_- ,
not necessarily
linear

$$S_+ u_i = u_{i+1}$$
$$S_- u_i = u_{i-1}$$

Note: a large class of problems can be written in this form

Example 1: $u_t + au_x = 0$

$$\text{with } u_i^{n+1} = u_i^n - a \frac{\Delta t}{\Delta x} (u_i^n - u_{i-1}^n)$$

$$\text{and } G(S_+, S_-) = \left(1 - a \frac{\Delta t}{\Delta x}\right) \mathbb{I} + \frac{a \Delta t}{\Delta x} S_-$$

Example 2: rewriting multi-level schemes

$$u_{tt} - c^2 u_{xx} = 0$$

$$\text{with: } u_i^{n+1} = 2u_i^n - u_i^{n-1} + \lambda^2 (u_{i+1}^n - 2u_i^n + u_{i-1}^n)$$

(3-level scheme) $\lambda = \frac{c \Delta t}{\Delta x}$

and define auxiliary variables

$$u_i^{n+1} = 2u_i^n - v_i^n + \lambda^2 (-)$$

$$v_i^{n+1} = u_i^n$$

$$\text{and } u^n = (u_1^n, v_1^n, \dots, u_I^n, v_I^n)$$

Example 3: Implicit schemes are of the form $L(S_+, S_-) u^{n+1} = Q u^n$

$$\text{and consider: } u^{n+1} = L^{-1} Q u^n \equiv Gu^n$$

Definition (Stability): Consider above setting, let $T > 0$ be fixed. The numerical scheme is called stable wrt. the norm $\|\cdot\|$ if there exists constants $c(T), \beta, \tau$ such that

$$\|u^n\| \leq c(T) e^{\beta \frac{n\Delta t}{T}} \|u^0\| \quad \forall 0 \leq \Delta t < \tau, n \leq \frac{T}{\Delta t}$$

Definition (truncation error, consistency):

Consider above setting. The numerical scheme is said to be consistent of order (q,p) wrt. the norm $\|\cdot\|$ if for any exact solution v of the PDE

$$\|e^n\| = \frac{1}{\Delta t} \|v(\cdot, t^{n+1}) - G v(\cdot, t^n)\| = [\mathcal{O}(\Delta t^q) + \mathcal{O}(\Delta x^p)]$$

e^n : "local truncation error"

Definition (convergence): Consider above setting. The numerical scheme is convergent of order (q,p) wrt. to the norm $\|\cdot\|$ if

for any exact solution v of the PDE

$$\|E^n\| = \|v(\cdot, t^n) - u^n\| = \Theta(\Delta t^q) + \Theta(\Delta x^p)$$

uniformly for all $n \in \mathbb{N}$.

$$E^n = v^n - u^n: \text{"global error"}$$

Theorem (Lax theorem):

(Lax & Richtmyer 1956, Comm. Pure Appl. Math. 9, 267)

Consider above setting and assume that G is linear in u^n . If the numerical scheme is stable and consistent of order (q, p) and the initial data are approximated consistently, i.e. $\|u^0 - v(\cdot, 0)\| = \Theta(\Delta x^p)$, then the scheme is convergent of order (q, p) , i.e.

$$\|u^n - v(\cdot, t^n)\| = \Theta(\Delta t^q) + \Theta(\Delta x^p)$$

uniformly $\forall n \leq \frac{T}{\Delta t}$ and v exact solution to the PDE.

Remark: the above definition & the last theorem require the IVP to be **well-posed**, i.e. that a solution v exists and that it is unique. More on this later.

Proof: let $t^n = n\Delta t \leq T$.

stability: $\|u^n\| = \|G^n u^0\| \leq c(T) e^{\beta \frac{n\Delta t}{T}} \|u^0\|$

$$\rightarrow \|G^n\| \leq c(T) e^{\beta \frac{n\Delta t}{T}}$$

consistency:

$$v^n = v(\cdot, t^n) = G v^{n-1} + \Delta t \tilde{v}^n$$

$$\text{with } \|\tilde{v}^n\| = O(\Delta x^p) + O(\Delta t^q)$$

no global error:

$$E^n = u^n - v^n = G u^{n-1} - G v^{n-1} - \Delta t \tilde{v}^n$$

$$\text{linearity} \Rightarrow E^n = G E^{n-1} - \Delta t \tilde{v}^n$$

$$= G^2 E^{n-2} - \Delta t G \tilde{v}^{n-1} - \Delta t \tilde{v}^n$$

$$= G^n E^0 - \Delta t \sum_{j=0}^{n-1} (G)^j \tilde{v}^j$$

↓ consistency

of initial data approx. $\|E^0\| = O(\Delta x^p)$

$$\begin{aligned}
 \text{and } \|E^n\| &\leq \Delta t \sum_{j=0}^{n-1} \|(\mathcal{G})^j z^j\| + O(\Delta x^p) \\
 \|z\| = \max_{j \leq n-1} \|z^j\| &\quad \xrightarrow{\text{green arrow}} \leq \Delta t \|z\| \sum_{j=0}^{n-1} c(T) e^{\beta \frac{j \Delta t}{T}} + O(\Delta x^p) \\
 &\leq \|z\| \underbrace{\Delta t}_T \underbrace{n}_{=T} c(T) e^{\beta \frac{n \Delta t}{T}} + O(\Delta x^p) \quad \hookrightarrow 1 + (\beta \frac{n \Delta t}{T}) + \frac{1}{2} \left(\dots \right)^2 \\
 &= O(\Delta t^q) + O(\Delta x^p)
 \end{aligned}$$

□

- Remarks:
- 1) The converse is also true (and "equivalence theorem"), but requires more work.
For practical purposes we will not need the converse statement.
 - 2) For non-linear operators \mathcal{G} the statement is false, in general.

3.4 Stability analysis & CFL cond.

Problem: how to tell a scheme converges?

most often don't know exact solution
which is why a numerical approach
is chosen in the first place!

we use Lax theorem, consistency &
stability easier to show, but need
sufficient criterion for stability

Theorem (von Neumann stability condition):

A finite difference scheme

$$\text{vector of } u^{n+1} = G(S_+, S_-) u^n$$

linear in u is stable wrt. to the L^2 -norm if
and only if there exist constants α, γ
such that

$$|g(\zeta)| \leq 1 + \alpha \Delta t$$

for all $\xi \in \mathbb{R}$ and $0 \leq \omega t \leq T$. Here, $g(\xi)$ is the amplification factor ("symbol of G ")

$$g(\xi) \equiv G(e^{-i\xi}, e^{i\xi}),$$

Proof: Consider discrete Fourier transform of $u = (u_j)_{j \in \mathbb{N}}$:

$$\hat{u}(\xi) = \sum_j u_j e^{ij\xi} \quad 0 \leq \xi \leq 2\pi$$

Observe:

$$\begin{aligned} \hat{S}_+ u(\xi) &= \sum_j S_+ u_j e^{ij\xi} = \sum_j u_{j+1} e^{i(j+1)\xi} \\ &\stackrel{\substack{l=j+1 \\ \cong}}{=} \sum_l u_l e^{i(l-1)\xi} = e^{-i\xi} \sum_j u_j e^{ij\xi} \\ &= e^{-i\xi} \hat{u}(\xi) \end{aligned}$$

$$\hat{S}_- u(\xi) = e^{i\xi} \hat{u}(\xi)$$

$$\begin{aligned} \Rightarrow \hat{u}^{n+1} &= G(e^{-i\xi}, e^{i\xi}) \hat{u}^n \\ &= g(\xi) \hat{u}^n \end{aligned} \quad (\star)$$

Discrete L^2 norms:

$$\|u^n\|_2^2 = \Delta x \sum_j (u_j^n)^2$$

$\xleftarrow[2\pi]{j}$

periodicity

$$\|\hat{u}^n\|_2^2 = \int_0^{2\pi} |\hat{u}^n(\xi)|^2 d\xi$$

Parseval property:

$$\begin{aligned} \|\hat{u}\|_2^2 &= \int_0^{2\pi} |\hat{u}(\xi)|^2 d\xi = \int_0^{2\pi} \left| \sum_j u_j e^{i j \xi} \right|^2 d\xi \\ &\stackrel{\substack{(e^{i j \xi}) \\ \text{orthogonal} \\ j \in \mathbb{N}}}{=} \int_0^{2\pi} \sum_j |u_j e^{i j \xi}|^2 d\xi \stackrel{|e^{i j \xi}|=1}{=} \int_0^{2\pi} \sum_j u_j^2 d\xi \\ &= 2\pi \sum_j u_j^2 = \frac{2\pi}{\Delta x} \|u\|_2^2 \end{aligned}$$

① assume $g(\xi) \leq 1 + \alpha \Delta t$:

$$\text{and } \|u^{n+1}\|_2^2 = \frac{\Delta x}{2\pi} \int_0^{2\pi} |\hat{u}^{n+1}(\xi)|^2 d\xi$$

$\underbrace{\quad}_{\|\hat{u}^{n+1}\|_2^2}$

$$\stackrel{(*)}{=} \frac{\Delta x}{2\pi} \int_0^{2\pi} |g(\xi)|^2 |\hat{u}^n(\xi)|^2 d\xi$$

$$\leq \frac{\Delta x}{2\pi} (1 + \alpha \Delta t)^2 \int_0^{2\pi} |\hat{u}^n(\xi)|^2 d\xi$$

$\underbrace{\quad}_{\|\hat{u}^n\|_2^2}$

$$\begin{aligned}
 &= (1 + \alpha \Delta t)^2 \|u^n\|_2^2 \\
 &\leq (e^{\alpha \Delta t})^2 \|u^n\|_2^2
 \end{aligned}$$

$$\Rightarrow \|u^n\|_2^2 \leq (e^{\alpha \Delta t})^{2n} \|u^0\|_2^2$$

scheme is stable!

- ② the reverse: assume $|g(\xi)| \leq 1 + \alpha \Delta t$ violated,
show scheme cannot be stable

\rightsquigarrow for any $\alpha > 0$, $T > 0$ $\exists \xi$ and $0 \leq \Delta t < T$
such that $|g(\xi)| > 1 + \alpha \Delta t$
for a finite region $\xi \in I_\alpha = [\xi_{\min}, \xi_{\max}]$

Consider $u^n = G u^{n-1}$ for $1 \leq n \leq \frac{T}{\Delta t}$

and without loss of generality consider
initial conditions that vanish outside I_α

i.e. $\hat{u}^0 = 0$ on $\mathbb{R} \setminus I_\alpha$.

$\rightsquigarrow |\hat{u}^n(\xi)| = |g^n(\xi) \hat{u}^0(\xi)| > (1 + \alpha \Delta t)^n |\hat{u}^0(\xi)|$
on I_α

Then:

$$\|G^n\| \|u^0\|_2 \geq \|G^n u^0\|_2 = \|u^n\|_2 \xrightarrow{\text{Parseval property}} (1 + \alpha \Delta t)^n \|u^0\|_2$$

$$\Rightarrow \|G^n\| \geq (1 + \alpha \Delta t)^n$$

Now assume scheme is stable

$$\Rightarrow \|G^n\| \leq C_0 \quad \forall 1 \leq n \leq \frac{T}{\Delta t}$$

Thus for all $\alpha, \tau > 0$ $\exists \Delta t \in (0, \tau)$ such that

$$(1 + \alpha \Delta t)^n \leq \|G^n\| \leq C_0 \quad \forall 0 \leq n \leq \frac{T}{\Delta t}$$

Now choose $n = \frac{T}{\Delta t}$

$$\Rightarrow \left(1 + \alpha \frac{T}{n}\right)^n \leq C_0$$

choose α sufficiently large such that $e^{\frac{\alpha T}{2}} > C_0$
choose τ sufficiently small such that

$$\left(1 + \frac{\alpha T}{n}\right)^n > \frac{2}{3} e^{\alpha T} \quad \text{for } n = \frac{T}{\Delta t}, \Delta t \in (0, \tau)$$

(n sufficiently large)

$$\left(\text{remember: } \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x\right)$$

Then:

$$\frac{1}{2} e^{\alpha T} > \left(1 + \frac{\alpha T}{n}\right)^n \geq \frac{2}{3} e^{\alpha T}$$

contradiction.

□

Remarks:

- 1) Note the theorem makes use of the fact that boundedness of the L^2 -norm of u^n follows from boundedness of L^2 -norm of \hat{u}^n via Parseval's identity. The latter is much easier to show since while all $\{u_j^n\}_{j \in \mathbb{N}}$ are coupled via the difference equations, each $\hat{u}(\xi)$ satisfies (A) and is thus decoupled from all other wavenumbers ξ !

- 2) One can make the replacement

$$|g(\xi)| \leq e^{\alpha \Delta t}$$

and the theorem still holds.

Note: $|g(\xi)| \leq (1 + \alpha \Delta t)^n = \left(1 + \frac{\alpha T}{n}\right)^n$

$$\xrightarrow[n \rightarrow \infty]{(\Delta t \rightarrow 0)} e^{\alpha T} \quad (\text{for } T > 0 \text{ fixed})$$

- 3) The stability bound allows for exponential growth of modes as, for example, required for $u_t + \alpha u_x = u$

but sometimes it is not appropriate, e.g., for

$$u_t + \alpha u_x = \varepsilon u_{xx}, \quad \varepsilon > 0$$

(all modes are damped)

Central in space, forward in time discretization

$$\text{leads to: } |g| \leq 1 + \frac{1}{2} \frac{\alpha^2}{\varepsilon} \Delta t$$

no concept of "strict/strong stability:

If $|\hat{u}(\xi, t + \Delta t)| \leq e^{\kappa \Delta t} |\hat{u}(\xi, t)| \quad \forall \xi$ (cont. problem),

require $|g(\xi)| \leq e^{\kappa \Delta t} \quad \forall \xi$ (discrete problem)

- 4) If one wants $\|u^n\| \leq c \|u^0\|$ (cf. 3))
 for stability (cf. Def. in Sec. 3.3),
 the requirement on $g(\xi)$ in the
 theorem sharpens to

$$|g(\xi)| \leq 1 \quad \forall \xi$$

5) alternative estimation of amplification factor

1) \Rightarrow it suffices to consider a arbitrary single wave number?

$$u_j^n = e^{i\zeta x_j}$$

to evaluate amplification factor.

6) 3+ - level schemes: either

(i) use method of 5) \rightarrow e.g. leapfrog scheme

(ii) amplification matrix analysis:

convert multi-level scheme to effective 2-level scheme as discussed before

$$\mathbf{U}^{n+1} = \mathbf{G} \mathbf{U}^n$$

all unknowns \rightarrow

$$\begin{pmatrix} u^{n+1} \\ v^{n+1} \end{pmatrix} = \mathbf{G} \begin{pmatrix} u^n \\ v^n \end{pmatrix}$$

$$\begin{pmatrix} \hat{u}^{n+1}(\zeta) \\ \hat{v}^{n+1}(\zeta) \end{pmatrix} = \mathbf{G} (e^{-i\zeta}, e^{i\zeta}) \begin{pmatrix} \hat{u}^n(\zeta) \\ \hat{v}^n(\zeta) \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} g_{11}(\xi) & g_{12}(\xi) \\ g_{21}(\xi) & g_{22}(\xi) \end{pmatrix}}_{g(\xi) = \text{amplification matrix}} \begin{pmatrix} \hat{u}^n \\ \hat{v}^n \end{pmatrix}$$

Assume that $g(\xi)$ has a complete set of (orthonormal) eigenvectors $g_\ell(\xi)$ and corresponding eigenvalues $\lambda_\ell(\xi)$, and suppose there exist constants $\alpha, \gamma > 0$ such that

$$|\lambda_\ell(\xi)| \leq (1 + \alpha \Delta t), \quad \ell = 1, 2$$

for all $\xi \in \mathbb{R}$ and $0 \leq \Delta t \leq \tilde{T}$.

Again bound $\|u^{n+1}\|_2$ by $\|\hat{u}^{n+1}\|_2$ using Parseval's property:

$$\begin{aligned} \hat{u}^n &= g^n(\xi) \hat{u}^0 \\ &= g^n(\xi) \sum_{\ell=1}^2 c_\ell^0 g_\ell(\xi) \\ &= \sum_{\ell} c_\ell^0 \lambda_\ell^n g_\ell(\xi) \end{aligned}$$

$$\begin{aligned}
 \text{and } \|u^n\|_2^2 &= \frac{\Delta x}{2\pi} \int_0^{2\pi} |\hat{u}^n(\xi)|^2 d\xi \\
 &\stackrel{\text{see above}}{=} \frac{\Delta x}{2\pi} \int_0^{2\pi} \left| \sum_{i=1}^n \lambda_i^n c_i^0 g_i(\xi) \right|^2 d\xi \quad \|\hat{u}^0\|_2^2 \\
 &\leq \frac{\Delta x}{2\pi} (1 + \alpha \Delta t)^{2n} \int_0^{2\pi} |\hat{u}^0(\xi)|^2 d\xi \\
 &= (1 + \alpha \Delta t)^{2n} \|u^0\|_2^2 \\
 &\leq (e^{\alpha \Delta t})^{2n} \|u^0\|_2^2 \quad \text{stable!}
 \end{aligned}$$

Example: stability of advection equation
with scheme:

$$\begin{aligned}
 u_j^{n+1} &= u_j^n - a\lambda \left[\theta u_{j+1}^n + (1-2\theta)u_j^n - (1-\theta)u_{j-1}^n \right] \\
 &= \left[(1-a\lambda(1-2\theta))I - a\lambda\theta S_+ + a\lambda(1-\theta)S_- \right] u_j^n \\
 &\equiv G(S_+, S_-) u_j^n, \quad \lambda = \frac{\Delta t}{\Delta x}
 \end{aligned}$$

$\theta \in [0, 1]$, where

$$\left. \begin{array}{l} \theta = 0 : \text{backward} \\ \theta = \frac{1}{2} : \text{centered} \\ \theta = 1 : \text{forward} \end{array} \right\} \text{difference in space}$$

$$w\ g(\xi) = Q(e^{-i\xi}, e^{i\xi})$$

$$\begin{aligned}
 &= 1 - a\lambda(1-2\theta) - a\lambda\theta e^{-i\xi} + a\lambda(1-\theta)e^{i\xi} \\
 &= \left[1 - a\lambda(1-2\theta) + a\lambda(1-2\theta)\cos\xi \right] \\
 &\quad + i(a\lambda\sin\xi)
 \end{aligned}$$

$$\Rightarrow |g(\xi)|^2 = 1 - 2a\lambda(1-\cos\xi) \left\{ 1 - 2\theta - \lambda a [1 + 2\theta(\theta-1) \times (1-\cos\xi)] \right\}$$

Assume $a > \theta$.

$\text{w/ } \theta = 1:$

$$\begin{aligned}|g(\xi)|^2 &= 1 - 2\alpha\lambda(1-\cos\xi) \{-1-\lambda a\} \\&= 1 + 2\alpha\lambda \underbrace{(1-\cos\xi)}_{\in [0,2]} (1+\lambda a) \\&\leq 1 + 4\alpha\lambda(1+\lambda a) \rightarrow \underline{\text{unstable}}\end{aligned}$$

$\theta = \frac{1}{2}:$ $|g(\xi)|^2 = 1 + 2\lambda^2 a^2 (1-\cos\xi)$

$$\underbrace{\left\{1 - \frac{1}{2}(1-\cos\xi)\right\}}_{\in [0,1]}$$

$$\leq 1 + \lambda^2 a^2 \xrightarrow{\in [0,1]} \underline{\text{unstable}}$$

$\theta = 0:$ $|g(\xi)|^2 = 1 - \alpha\lambda \underbrace{(1-\cos\xi)}_{\in [0,2]} (1-\lambda a)$

$$\leq 1 - 2\alpha\lambda(1-\lambda a)$$

$\rightarrow \underline{\text{stable if }} 1-\lambda a > 0$

$$C = \alpha\lambda = \frac{\alpha\Delta t}{\Delta x} < 1$$

CFL
condition

Courant
number

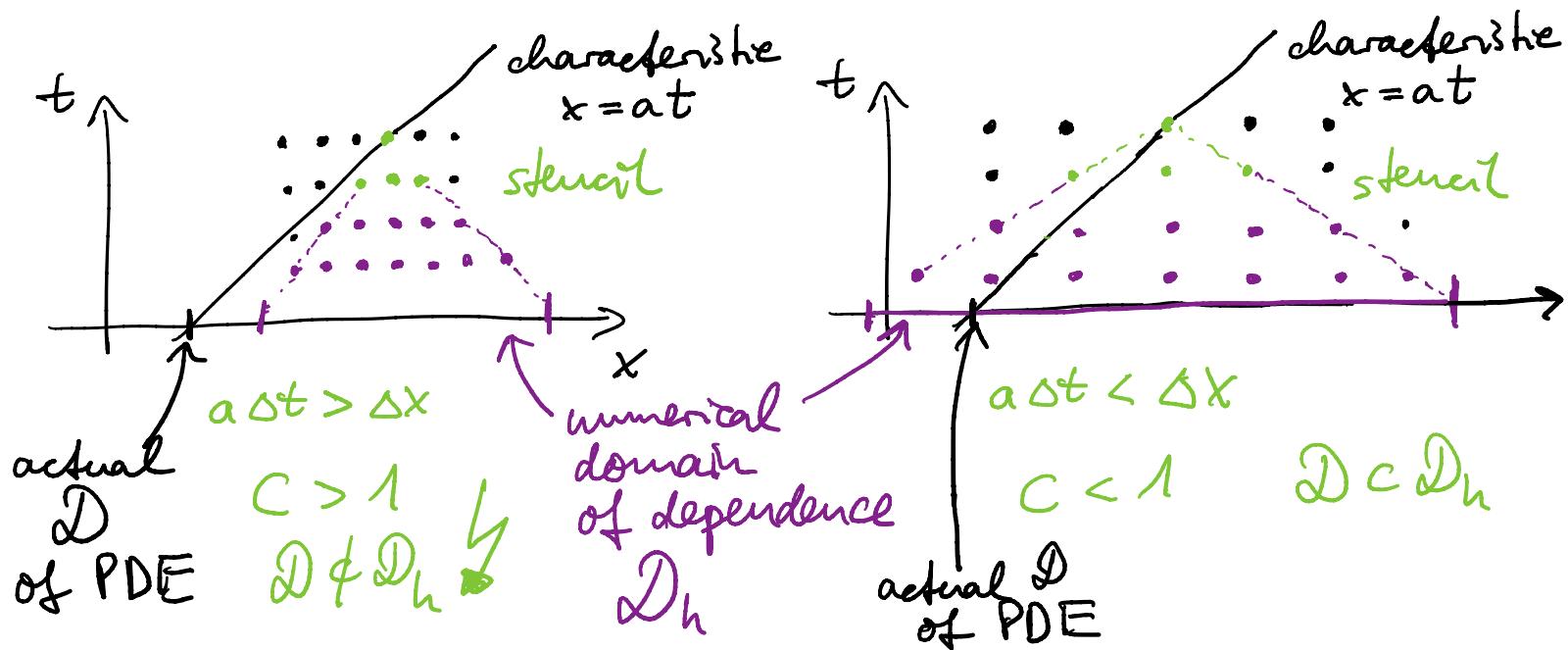
(Courant, Friedrichs, Lewy)

Remark: The previous result $\frac{a\Delta t}{\Delta x} < 1$ is one special case of a more general principle, a necessary but not sufficient condition for stability & convergence:

The CFL condition: A numerical method can be convergent only if its numerical domain of dependence contains the true domain of dependence of the PDE, at least in the limit as Δt and Δx go to zero.

(Courant, Friedrichs & Lewy, Math. Ann. 100, 32 (1928))

Advection equation:



Remarks: 1) For linear hyperbolic systems with m eigenvalues/wavespeeds $\lambda^1, \dots, \lambda^m$ (cf. discussion on domain of dependence):

$$C = \frac{\Delta t}{\Delta x} \max_p |\lambda^p| \leq 1$$

Typically use explicit methods with C fixed.

For non-linear conservation equation

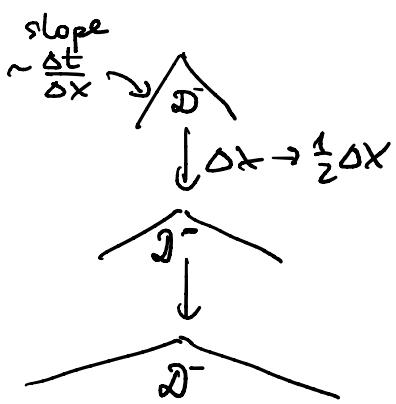
$$u_t + f(u)_x = 0$$

generalize to:

$$C = \sup \frac{\Delta t |f'(u)|}{\Delta x} < 1$$

- 2) A wider stencil \Rightarrow less restrictive CFL condition on Δt
- 3) Parabolic equations: diffusion equations
severe constraints on explicit method from CFL, since $D = (-\infty, \infty)$.
(infinite propagation speed)

and take $\Delta t = O(\Delta x^2)$ then
 explicit method captures \mathcal{D}
 in limit $\Delta t, \Delta x \rightarrow 0$



(as we refine $\Delta x \rightarrow \frac{1}{2} \Delta x \Rightarrow \Delta t \rightarrow \frac{1}{4} \Delta t$)
 \Rightarrow "slope of \mathcal{D} " widens: $\frac{\Delta t}{\Delta x} \rightarrow \frac{1}{2} \frac{\Delta t}{\Delta x}$
 $\xrightarrow{\Delta x \rightarrow 0} 0 \Rightarrow \mathcal{D}_h \xrightarrow{\Delta x \rightarrow 0} (-\infty, \infty)$

better: use implicit method

$$\rightarrow \mathcal{D}_h = \mathcal{D} = (-\infty, \infty). \text{ The}$$

Crank-Nicholson method, for example,
 satisfies CFL condition for any Δt
 (the tridiagonal matrix couples all
 points such that solution at a given
 point depends on all other points)

3.5 Diffusion & dispersion

Stability & accuracy are not the only aspects to be considered when evaluating a numerical scheme!

→ Discretization introduces additional terms not present in continuum PDE

Consider advection equation and let $v(x,t)$ denote an exact solution of the discretized eqn

Example 1: Upwind method

$$u_i^{n+1} = u_i^n - \frac{a\delta t}{\Delta x} (u_i^n - u_{i-1}^n), \quad a > 0.$$

↓
 v satisfies exactly

$$v(x_i, t + \delta t) = v(x_i, t) - \frac{a\delta t}{\Delta x} [v(x_i, t) - v(x_i - \Delta x, t)]$$

↓ Taylor series

$$\left(v_t + \frac{1}{2} \Delta t v_{tt} + \frac{1}{6} (\Delta t)^2 v_{ttt} + \dots \right) + a \left(v_x - \frac{1}{2} \Delta x v_{xx} + \frac{1}{6} (\Delta x)^2 v_{xxx} + \dots \right) = 0$$

$$\Leftrightarrow v_t + a v_x = \frac{1}{2} (a \Delta x v_{xx} - \Delta t v_{tt}) - \frac{1}{6} [a (\Delta x)^2 v_{xxx} + (\Delta t)^2 v_{ttt}] + \dots$$

different from
original continuum PDE!

PDE for $v(t,x)$

Take $\frac{\Delta t}{\Delta x} = \text{fixed}$.

1st order: recover original advection eqn

2nd order:

$$v_t + a v_x = \frac{1}{2} (a \Delta x v_{xx} - \Delta t v_{tt}) \quad (*)$$

$$\downarrow \partial_t$$

$$v_{tt} = -a v_{xt} + \frac{1}{2} (a \Delta x v_{xxt} - \Delta t v_{ttt})$$

$$\{ Q_x \}$$

$$v_{xt} = -a v_{xx} + \frac{1}{2} (a \Delta x v_{xxx} - \Delta t v_{ttx})$$

$$v_{tt} = a^2 v_{xx} + O(\Delta t)$$

$$\downarrow \text{in } (*)$$

advection-diffusion equation

$$v_t + a v_x = \frac{1}{2} a \Delta x \left(1 - \underbrace{\frac{a \Delta t}{\Delta x}}_{\lambda} \right) v_{xx}$$

λ : Courant number

Note: 1) v_i^n provide 2nd order accurate approximation to the true solution of this equation, but only 1st order accurate solution to advection eqn

2) the diffusion coefficient

$$v_d = \frac{1}{2} a \Delta x (1 - \lambda)$$

is called numerical diffusion. Observe:

$$v_d = 0 \Leftrightarrow \lambda = 1 \quad (\text{marginally stable})$$

$$v_d > 0 \Leftrightarrow \lambda < 1 \quad (\text{CFL condition})$$

Upwind method is
dissipative & stable!

$v_d < 0 \Leftrightarrow \lambda > 1$ ill-posed backward heat eqn with exponentially growing solutions
Upwind method unstable!

Note the relation between properties of modified eqn (PDE for v) and stability of continuum PDE

Example 2: Lax-Vendroff method

2nd order
accurate

2-level
3-point
method

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

↓ similar analysis in terms of exact solution $v(x,t)$ to discretized problem

$$v_t + av_x = -\frac{1}{6} a (\Delta x)^2 \left[1 - \left(\frac{a \Delta t}{\Delta x} \right)^2 \right] v_{xxx}$$

PDE for $v(x,t)$

Note: 1) u_n^i {3rd-order accurate sol. to this eqn.
2nd-order " " " " advection eq.

2) v_{xxx} term leads to dispersive behavior (see below)

General analysis: PDE dispersion relation

previous examples suggest PDE for exact solution of discretized problem of the general form

$$v_t + a_1 v_x + a_2 v_{xx} + a_3 v_{xxx} + a_4 v_{xxxx} + \dots = 0$$

\downarrow Fourier transformation $\hat{v}(\xi, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} v(x, t) e^{-ix\xi} dx$

$$\begin{aligned} \hat{v}_t(\xi, t) &= -a_1 i\xi \hat{v}(\xi, t) + a_2 \xi^2 \hat{v}(\xi, t) + a_3 i\xi^3 \hat{v}(\xi, t) \\ &\quad - a_4 \xi^4 \hat{v}(\xi, t) + \dots \end{aligned}$$

$$\Rightarrow \hat{v}(\xi, t) = e^{-i\omega t} \hat{v}(\xi) \quad (\hat{v}(\xi) = \hat{v}(\xi, 0))$$

with $\boxed{\omega(\xi) = a_1 \xi + i a_2 \xi^2 - a_3 \xi^3 - i a_4 \xi^4 - \dots}$

dispersion relation of the PDE

$$v(x, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{v}(\xi) e^{i(\xi x - \omega(\xi)t)} d\xi$$

Note: 1) dispersion relation can be found via ansatz $v(x, t) = e^{-i\omega t} e^{i\xi x}$

(different wave numbers ξ decouple)

2) $e^{-i\omega t} = e^{\underbrace{(\alpha_2 \xi^2 - \alpha_4 \xi^4 + \dots)t}_{\text{diffusive term}}} \times e^{\underbrace{i(\alpha_1 \xi - \alpha_3 \xi^3 + \dots)t}_{\text{dispersive term}}}$

(exponential growth or decay) (waves travel at different speed)

Upwind: diffusive

Lax-Wendroff: (purely) dispersive

In general: may be combination of both

3) purely dispersive $\Leftrightarrow \omega(\xi)$ real $\forall \xi \in \mathbb{R}$

$$\Rightarrow u(x,t) \stackrel{1)}{=} e^{i\xi(x - \frac{\omega}{\xi}t)}$$

phase velocity
(propagation speed of this mode)

$$C_p = \frac{\omega(\xi)}{\xi}$$

$$C_g = \frac{d\omega(\xi)}{d\xi}$$

group velocity

In general: $a \neq C_g \neq C_p$

Conclusion:

- schemes in which leading error term is composed of 2nd order spatial derivatives show numerical diffusion / numerical viscosity
- schemes in which leading error term is composed of 3rd order spatial derivatives introduce numerical dispersion