



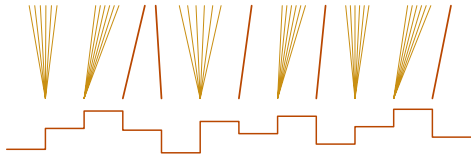
Computer Science &
Artificial Intelligence
Laboratory

Conservation Laws II

Conservative finite volume schemes and timestepping (Sections 8.3–8.4)

Johannes Herter

MIT CSAIL · Numerical Methods



Lecture I: the Cauchy problem

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) = 0, \quad u(x, 0) = u_0(x), \quad (8.2.7)$$

in $\mathbb{R} \times]0, T[$. Weak solutions may be discontinuous; the **entropy solution** is the physically meaningful one.

Structure of entropy solutions (Sect. 8.2.7):
comparison principle, L^∞ -stability,
 L^1 -contractivity, finite speed of propagation, no
new extrema.

Today.

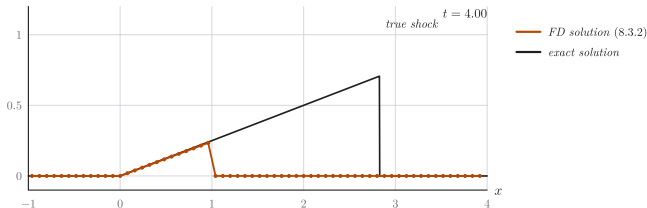
- 8.3** Discretizations that respect conservation: finite volumes in **conservation form**, numerical flux functions (central, Lax–Friedrichs, upwind, Godunov), **monotone schemes**.
- 8.4** Timestepping: the **CFL-condition**, linear (von Neumann) stability, convergence and its limits.

8.3 *Conservative finite volume discretization*

Example 8.3.1: a naive finite difference scheme

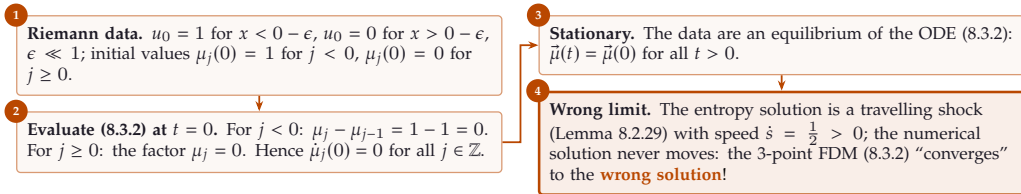
Burgers (8.1.46) by the product rule: $\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0$, linear advection with velocity $v = u$. For $u_0 \geq 0$, $u \geq 0$ (Thm. 8.2.41): transport to the right; upwind (backward) differences on $x_j = hj$ (Sect. 7.3.1):

$$\dot{\mu}_j(t) + \mu_j \frac{\mu_j - \mu_{j-1}}{h} = 0, \quad j \in \mathbb{Z}, \quad 0 < t < T. \quad (8.3.2)$$



“Box” data of Ex. 8.2.39, $h = 0.08$. OK for the rarefaction wave, but the scheme cannot capture the speed of the shock!

Example 8.3.1: why the shock does not move



Next. An approach to discretization with built-in safeguards against such failures: conservation form.

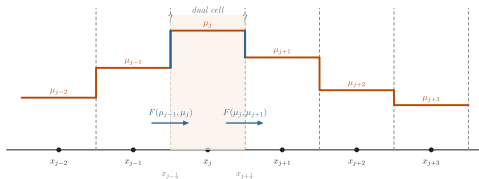
8.3.1 Semi-discrete conservation form

Spatial semi-discretization of (8.2.7) on an infinite equidistant mesh, a method of lines approach (Sect. 6.1.4):

$$\mathcal{M} := \{]x_{j-1}, x_j[: x_j := jh, j \in \mathbb{Z}\}. \quad (8.3.3)$$

Finite volume interpretation of the unknowns: $\mu_j \leftrightarrow$ conserved quantity in the **dual cells** $]x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}[, x_{j-\frac{1}{2}} := \frac{1}{2}(x_j + x_{j-1})$:

$$\mu_j(t) \approx \frac{1}{h} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, t) dx. \quad (8.3.4)$$



$\vec{\mu}(t) = (\mu_j(t))_{j \in \mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}} \longleftrightarrow$ the piecewise constant function

$$u_N(x, t) = \sum_{j \in \mathbb{Z}} \mu_j(t) \chi_{]x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}[}(x), \quad (8.3.5)$$

discontinuous at the dual cell boundaries $x_{j+\frac{1}{2}}$.

Integrate over the dual cells (the control volumes of (8.2.1)) and apply the fundamental theorem of calculus:

$$\frac{d}{dt} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, t) dx + f(u(x_{j+\frac{1}{2}}, t)) - f(u(x_{j-\frac{1}{2}}, t)) = 0, \quad j \in \mathbb{Z}, \quad (8.3.6)$$

$$\stackrel{(8.3.4)}{\implies} \frac{d\mu_j}{dt}(t) + \frac{1}{h} \left(\underbrace{f(u_N(x_{j+\frac{1}{2}}, t))}_{?} - \underbrace{f(u_N(x_{j-\frac{1}{2}}, t))}_{?} \right) = 0. \quad (8.3.7)$$

Problem: owing to the jumps of $u_N(t)$ the values $u_N(x_{j\pm\frac{1}{2}}, t)$ are ambiguous (a similar situation as for upwind quadrature, Sect. 7.2.2.1).

Abstract “solution”: approximate

$$f(u_N(x_{j+\frac{1}{2}}, t)) \approx f_{j+\frac{1}{2}}(t) := F(\mu_{j-m_l+1}(t), \dots, \mu_{j+m_r}(t))$$

with a **numerical flux function** $F : \mathbb{R}^{m_l+m_r} \rightarrow \mathbb{R}$, $m_l, m_r \in \mathbb{N}_0$. If $f = f(u)$, the same numerical flux function is usually used for all dual cells.

Plugging the numerical fluxes into (8.3.7) gives a (formally infinite) system of ODEs:

Finite volume semi-discrete evolution for (8.2.7) in conservation form

$$\frac{d\mu_j}{dt}(t) = -\frac{1}{h} \left(F(\mu_{j-m_l+1}(t), \dots, \mu_{j+m_r}(t)) - F(\mu_{j-m_l}(t), \dots, \mu_{j+m_r-1}(t)) \right), \quad j \in \mathbb{Z}. \quad (8.3.9)$$

Special case of a **2-point numerical flux** ($m_l = m_r = 1$), $F = F(v, w)$, $v \hat{=}$ left state, $w \hat{=}$ right state:

$$\frac{d\mu_j}{dt}(t) = -\frac{1}{h} \left(F(\mu_j(t), \mu_{j+1}(t)) - F(\mu_{j-1}(t), \mu_j(t)) \right), \quad j \in \mathbb{Z}. \quad (8.3.10)$$

Assumption on numerical flux functions: F Lipschitz-continuous in each argument. Time integration: high order explicit Runge–Kutta methods (Sect. 8.4).

8.3.2 Discrete conservation property

(8.3.14) Preservation of constant data. With $f_{j+\frac{1}{2}}(t) := F(\mu_{j-m_l+1}(t), \dots, \mu_{j+m_r}(t))$:

$$\mu_j(0) = \mu_0 \quad \forall j \Rightarrow \mu_j(t) = \mu_0 \quad \forall j, t \quad (8.3.15)$$

well-balanced discretization.

Definition 8.3.18 (Consistent numerical flux)

$F : \mathbb{R}^{m_l+m_r} \rightarrow \mathbb{R}$ is consistent with $f : \mathbb{R} \rightarrow \mathbb{R}$, if $F(u, \dots, u) = f(u) \quad \forall u \in \mathbb{R}$.

1

(8.3.16) Exact balance (cf. (6.1.3)), (8.3.17):

$$\frac{d}{dt} \int_a^b u \, dx = -(f(u(b, t)) - f(u(a, t))).$$

2

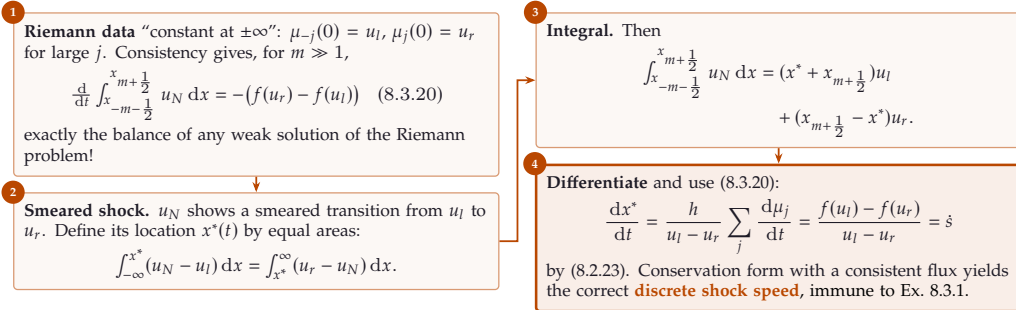
Telescoping sum of (8.3.9) over $k \leq l \leq m$, with (8.3.5):

$$\begin{aligned} \frac{d}{dt} \int_{x_{k-\frac{1}{2}}}^{x_{m+\frac{1}{2}}} u_N \, dx &= h \sum_{l=k}^m \frac{d\mu_l}{dt} \\ &= -(f_{m+\frac{1}{2}} - f_{k-\frac{1}{2}}). \end{aligned}$$

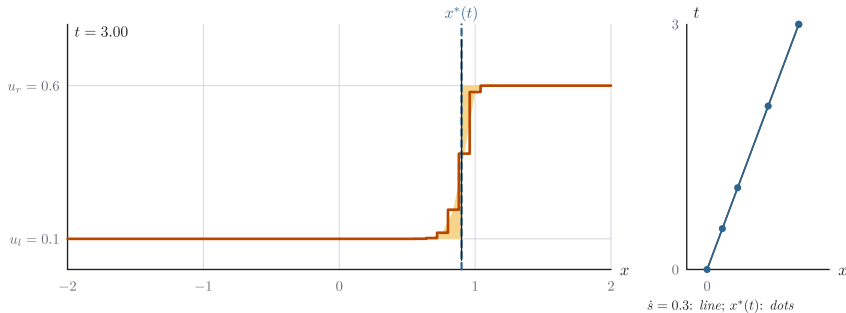
3

Discrete conservation. On unions of dual cells, with numerical fluxes, $u_N(t)$ satisfies a balance law of the same structure as a weak solution of (8.2.7).

(8.3.19) Discrete shock speed



The discrete shock moves at the right speed



Traffic flow, $u_l = 0.1 < u_r = 0.6$ (a shock), Lax–Friedrichs flux (Sect. 8.3.3.2), $h = 0.08$. The transition is smeared over a few cells, but the equal-area location $x^*(t)$ (yellow areas equal) travels exactly with $\dot{s} = \frac{f(u_l) - f(u_r)}{u_l - u_r} = 0.3$.

8.3.3 Numerical fluxes: 8.3.3.1 central flux

Concrete consistent (Def. 8.3.18) 2-point fluxes $F = F(v, w)$; finding good ones is not straightforward: both *instability* and solutions *violating the entropy condition* must be avoided.

Central numerical fluxes (arithmetic averaging)

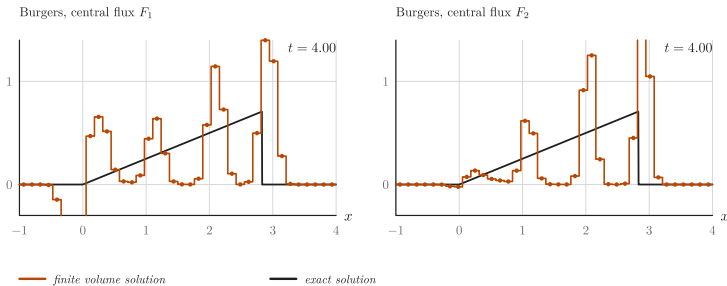
$$F_1(v, w) := \frac{1}{2} (f(v) + f(w)), \quad F_2(v, w) := f\left(\frac{1}{2}(v + w)\right). \quad (8.3.21)$$

Obviously consistent. The semi-discrete schemes (8.3.10):

$$F_1 : \frac{d\mu_j}{dt} = -\frac{1}{2h} (f(\mu_{j+1}) - f(\mu_{j-1})), \quad F_2 : \frac{d\mu_j}{dt} = -\frac{1}{h} \left(f\left(\frac{\mu_j + \mu_{j+1}}{2}\right) - f\left(\frac{\mu_j + \mu_{j-1}}{2}\right) \right).$$

Exp. 8.3.22: central flux, Burgers

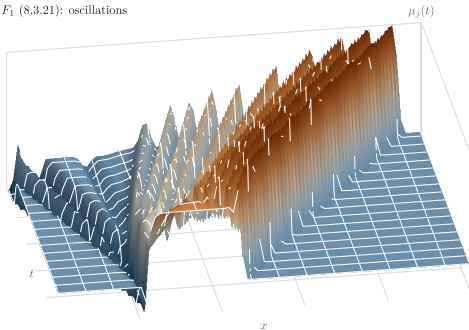
Box data of Ex. 8.2.39, conservation form (8.3.9) with (8.3.21), adaptive explicit Runge–Kutta timestepping (here RK4 with small steps), $h = 0.13$:



Observation: massive spurious oscillations utterly pollute the numerical solution.

Central flux: the space-time surface

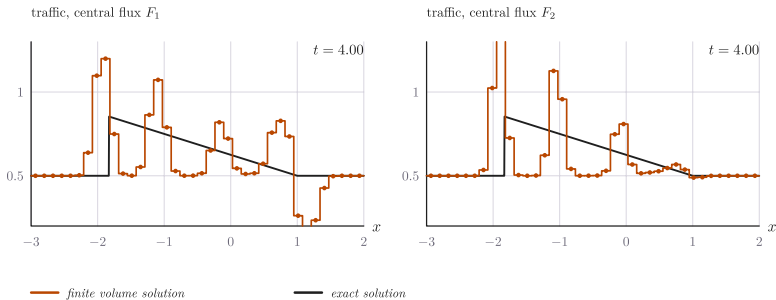
central flux F_1 (8.3.21): oscillations



The cell averages $\mu_j(t)$ of Exp. 8.3.22 over the x - t plane (cf. the 3D output of the script's finite volume code): spurious oscillations grow into ridges and troughs that run along the characteristics.

Exp. 8.3.23: central flux, traffic flow

Box data of Ex. 8.2.40 ($u_0 = 1$ on $[0, 1]$, 0.5 elsewhere), same discretization:



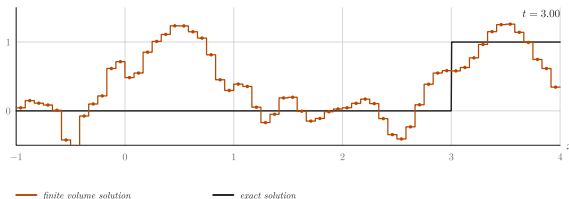
Observation: again massive spurious oscillations.

Exp. 8.3.24: central flux, advection

Is this an inherent weakness of central fluxes? Linear advection (8.1.10), $c = 1$, $f(u) = cu$: both central fluxes give

$$\frac{d\mu_j}{dt}(t) = -\frac{c}{2h} (\mu_{j+1}(t) - \mu_{j-1}(t)), \quad j \in \mathbb{Z}. \quad (8.3.25)$$

linear advection $c = 1$, central flux (8.3.25), $h = 0.083$



Remark 8.3.26. (8.3.25) = linear FE Galerkin semi-discretization of (7.3.7), prone to spurious oscillations (Ex. 7.3.4): this also explains the failure for Burgers/traffic flow.

8.3.3.2 (8.3.27) *Fighting oscillations with diffusion*

Linear advection models heat transport in a moving fluid: extrema are carried along; with heat conductivity $\kappa > 0$ they decay. For $c = \kappa = 1$:

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} - \underbrace{\frac{\partial^2 u}{\partial x^2}}_{\text{diffusive term}} = 0, \quad u_0(x) = \sin(x) \Rightarrow u(x, t) = e^{-t} \sin(x - t). \quad (8.3.28)$$

Add diffusion of strength $\kappa > 0$: $\frac{\partial u}{\partial t} + c \frac{\partial u}{\partial x} - \kappa \frac{\partial^2 u}{\partial x^2} = 0$ (8.3.29), a conservation law with flux $f(u) = cu - \kappa \frac{\partial u}{\partial x}$ (8.3.30), Rem. 8.2.2. Discretize each part:

Central + diffusive numerical flux

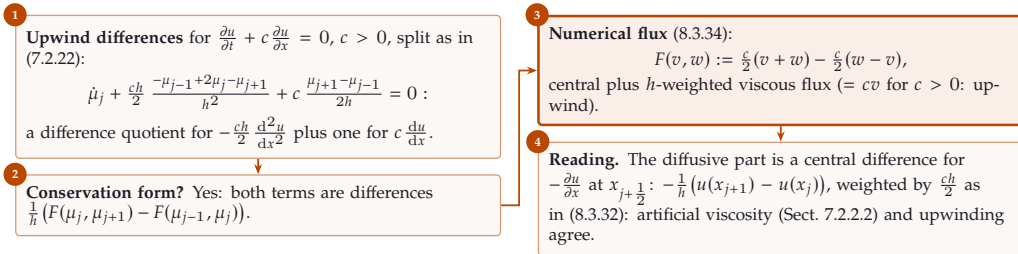
$$F(v, w) = \underbrace{\frac{c}{2}(v + w)}_{\text{central}} - \underbrace{\kappa \frac{w - v}{h}}_{\text{diffusive}}$$

$$\dot{\mu}_j + c \frac{\mu_{j+1} - \mu_{j-1}}{2h} + \kappa \frac{-\mu_{j+1} + 2\mu_j - \mu_{j-1}}{h^2} = 0 \quad (8.3.31)$$

= linear FE Galerkin MOL-ODE for (8.3.29).

Caution: the extra diffusion must vanish for $h \rightarrow 0$, $\kappa = \kappa(h)$; guideline: $\kappa = \frac{ch}{2}$ (8.3.32).

Rem. 8.3.33: upwinding = central + viscosity



General scalar conservation laws are linked with linear advection by

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) = \frac{\partial u}{\partial t} + \underbrace{f'(u)}_{\leftrightarrow c} \frac{\partial u}{\partial x} = 0 : \quad (8.3.35)$$

the local speed of transport $f'(u)$ varies, so the artificial diffusion should vary too. Choose it by (8.3.32), large enough for the maximal local velocity:

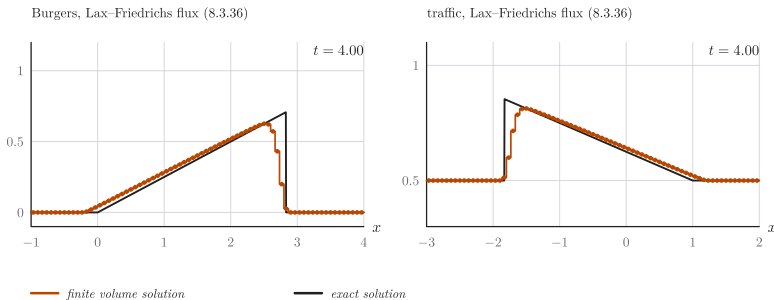
$$\kappa = \frac{h}{2} \max\{|f'(u)| : \min\{v, w\} \leq u \leq \max\{v, w\}\}.$$

(Local) Lax–Friedrichs / Rusanov flux

$$F_{\text{LF}}(v, w) = \frac{1}{2} (f(v) + f(w)) - \frac{1}{2} (w - v) \cdot \max_{\min\{v, w\} \leq u \leq \max\{v, w\}} |f'(u)|. \quad (8.3.36)$$

Examples 8.3.37 / 8.3.38: Lax–Friedrichs flux

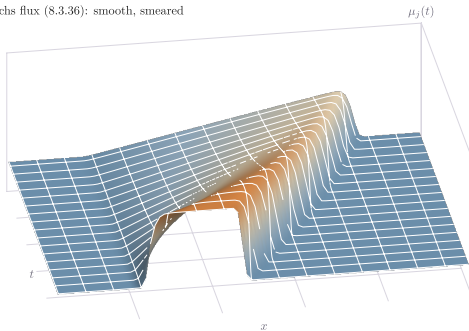
Same settings as Exps. 8.3.22, 8.3.23; numerical flux (8.3.36), $h = \frac{1}{15}$:



Observation: spurious oscillations are suppressed completely; qualitatively good resolution of shock and rarefaction. Effect of artificial diffusion: *smearing of the shock* (cf. Ex. 7.2.31).

Lax–Friedrichs: the space-time surface

Lax–Friedrichs flux (8.3.36): smooth, smeared



The same problem with the Lax–Friedrichs flux (8.3.36): no oscillations, a clean rarefaction ramp and a shock smeared over a few cells (the artificial diffusion).

8.3.3.3 Upwind flux

“Upwinding” = take information from where transport brings it (Sect. 7.2.2.1); but the slope $f'(u_N(x_{j+\frac{1}{2}}, t))$ of the characteristic through $(x_{j+\frac{1}{2}}, t)$ is ambiguous too.

Idea. There is a “velocity of propagation” even at discontinuities: the Rankine–Hugoniot speed (8.2.23),

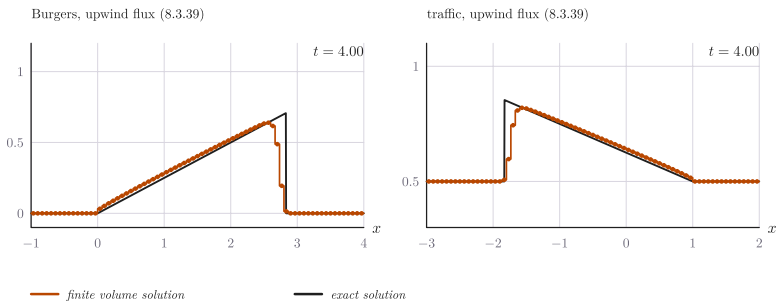
$$\text{local velocity of transport} = \begin{cases} f'(u), & \text{unique state } u = u_l = u_r, \\ \frac{f(u_r) - f(u_l)}{u_r - u_l}, & \text{at a discontinuity.} \end{cases}$$

Upwind numerical flux for a scalar conservation law with flux function f

$$F_{\text{uw}}(v, w) = \begin{cases} f(v), & \text{if } \dot{s} \geq 0, \\ f(w), & \text{if } \dot{s} < 0, \end{cases} \quad \dot{s} := \begin{cases} \frac{f(w) - f(v)}{w - v}, & v \neq w, \\ f'(v), & v = w. \end{cases} \quad (8.3.39)$$

Examples 8.3.40 / 8.3.41: upwind flux

Same settings; numerical flux (8.3.39). Traffic flow: the box solution comprises a stationary shock and a rarefaction fan, which merge eventually.



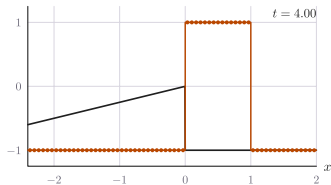
Observation: a satisfactory resolution of the shock and the rarefaction fan.

Ex. 8.3.42: upwind flux, transonic fan

Burgers with $u_0 = 1$ on $]0, 1[$, -1 elsewhere
(8.3.43)

traffic with $u_0 = 1$ on $]0, 1[$, 0 elsewhere (8.3.44)

Burgers, (8.3.43), upwind flux

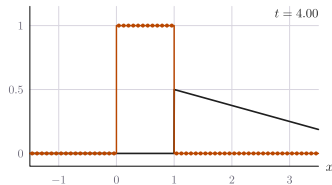


— finite volume solution

— exact solution

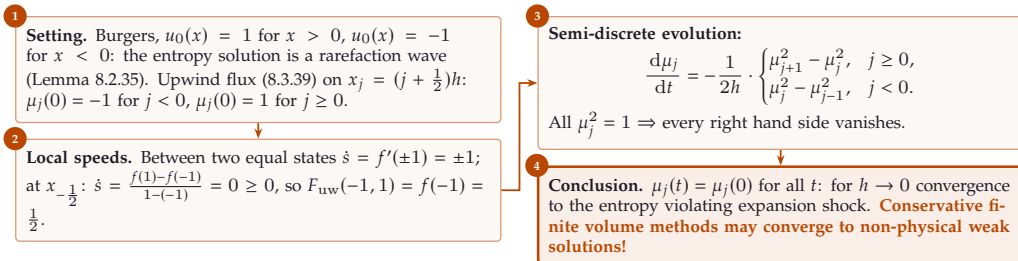
The entropy solutions feature a **transonic rarefaction**: a fan whose edges move in opposite directions.

traffic, (8.3.44), upwind flux



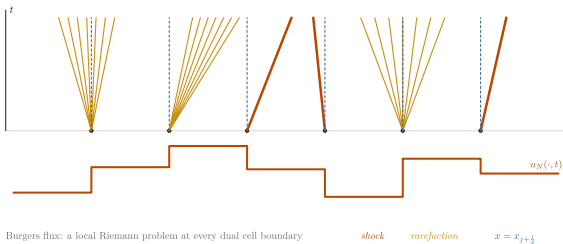
Upwind: a stationary **expansion shock** instead of the transonic rarefaction, a “non-physical shock”.

Ex. 8.3.45: convergence to an expansion shock



8.3.3.4 Godunov flux: local Riemann problems

Upwind: $F(v, w) = f(u^\downarrow(v, w))$, $u^\downarrow \in \{v, w\}$ the state at $x_{j+\frac{1}{2}}$ of an *all-shock* solution with data u_N : it may contain expansion shocks and misses rarefactions.



Idea: $u^\downarrow(v, w) = \psi(0)$ (8.3.47), $u = \psi(x/t)$ the *entropy* solution of the local Riemann problem with v for $x < 0$, w for $x \geq 0$ (8.3.48).

The local Riemann solutions (convex f)

f strictly convex & smooth (e.g. Burgers): the Riemann problem (8.3.48) has the entropy solution

- ❶ $v > w$: a **shock** (Lemma 8.2.29), $u = v$ for $x < \dot{s}t$, $u = w$ for $x > \dot{s}t$, $\dot{s} = \frac{f(v)-f(w)}{v-w}$;
(8.3.49)
- ❷ $v \leq w$: a **rarefaction** (Lemma 8.2.35): v for $x < f'(v)t$, $g(\frac{x}{t})$ for $f'(v) \leq \frac{x}{t} \leq f'(w)$, w for $x > f'(w)t$, $g := (f')^{-1}$ (8.3.50)

All weak Riemann solutions are **similarity solutions** $u = \psi(x/t)$ (cf. §8.2.34), with ψ

- piecewise constant with a jump at $\dot{s} = \frac{f(w)-f(v)}{w-v}$ for a shock (8.3.49),
- the continuous function

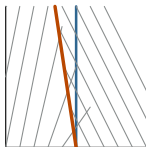
$$\psi(\xi) = \begin{cases} v, & \xi < f'(v), \\ (f')^{-1}(\xi), & f'(v) < \xi < f'(w), \\ w, & \xi > f'(w), \end{cases}$$

if $w > v$: the rarefaction (8.3.50).

(8.3.51) The intermediate state: five cases

For a convex flux function f :

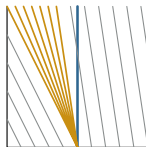
$$u^\downarrow(v, w) = \begin{cases} w, & \text{if } v > w \wedge \dot{s} < 0 \text{ (shock ❶) or } v < w \wedge f'(w) < 0 \text{ (rarefaction ❷),} \\ v, & \text{if } v > w \wedge \dot{s} > 0 \text{ (shock ❸) or } v < w \wedge f'(v) > 0 \text{ (rarefaction ❹),} \\ (f')^{-1}(0), & \text{if } v < w \wedge f'(v) \leq 0 \leq f'(w) \text{ (rarefaction ❺).} \end{cases} \quad (8.3.51)$$



❶ subsonic shock

$$u^\downarrow = w$$

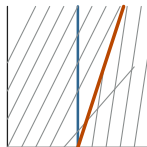
$$v = 0.4, w = -1$$



❷ subsonic rarefaction

$$u^\downarrow = w$$

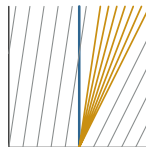
$$v = -1, w = -0.3$$



❸ supersonic shock

$$u^\downarrow = v$$

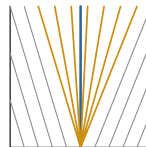
$$v = 1, w = 0.3$$



❹ supersonic rarefaction

$$u^\downarrow = v$$

$$v = 0.3, w = 1$$

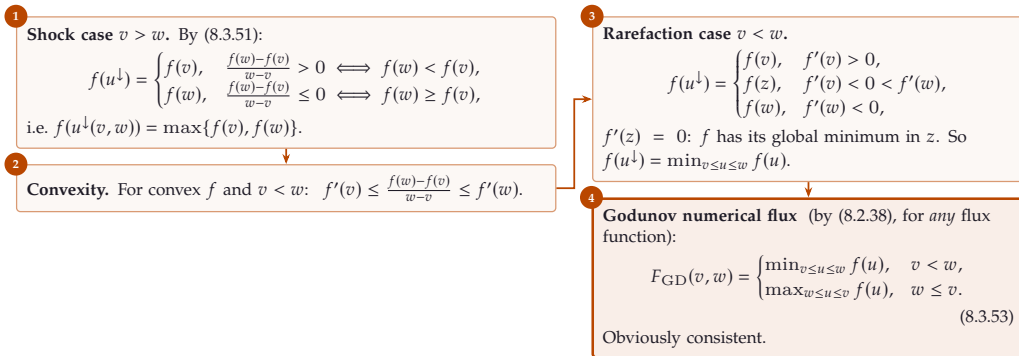


❺ transonic rarefaction

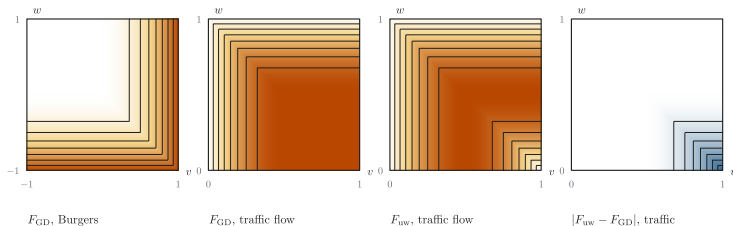
$$u^\downarrow = 0$$

$$v = -0.6, w = 0.8$$

(8.3.52) Formulas for the Godunov numerical flux

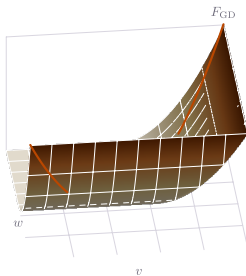


Godunov flux and Remark 8.3.54

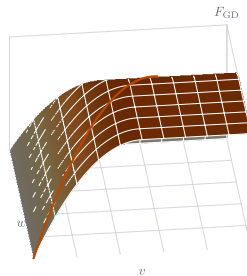


Remark 8.3.54. $F_{uw} = F_{GD}$ except for transonic rarefaction (right). There, e.g. convex f , $v < w$, $f'(v) < 0 < f'(w)$: $F_{uw}(v, w) = f(\psi(0))$, ψ an entropy-violating expansion shock. **Upwind treats transonic rarefaction as an expansion shock**: the explanation for Ex. 8.3.42.

The Godunov flux as a surface



Burgers

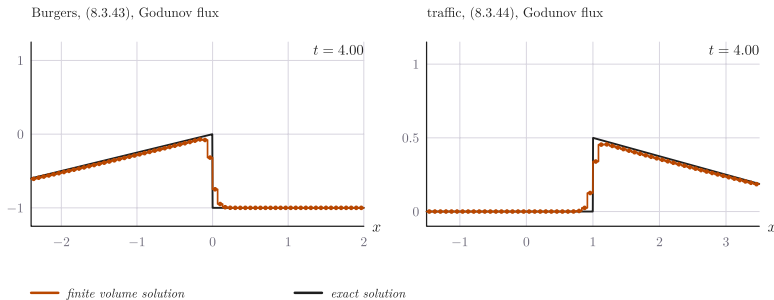


traffic flow

$F_{\text{GD}}(v, w)$ of (8.3.53) for Burgers (left) and traffic flow (right). Orange: the diagonal $F(u, u) = f(u)$ (consistency, Def. 8.3.18). Monotonicity (Lemma 8.3.64) is visible: the surface rises in v and falls in w .

Examples 8.3.55 / 8.3.56: Godunov flux

Same setting as Ex. 8.3.42 (transonic rarefactions); numerical flux (8.3.53):



Observation: the transonic rarefaction is captured by the discretization; only a small remnant of an expansion shock (a kink at the sonic point) is still observed.

8.3.4 Monotone schemes

Observations for Lax–Friedrichs (Ex. 8.3.37) and Godunov (Ex. 8.3.55):

$\min_x u_0 \leq u_N(x, t) \leq \max_x u_0$ and no new local extrema: **structure preservation**, cf. Thm. 8.2.41 and Sect. 8.2.7. Coincidence?

(8.3.57) Discrete comparison principle. (8.3.10) with the Lax–Friedrichs flux (8.3.36):

$$\frac{d\mu_j}{dt} = -\frac{1}{2h} \left(f(\mu_{j+1}) - f(\mu_{j-1}) - \max_{u \in [\mu_j, \mu_{j+1}]} |f'(u)|(\mu_{j+1} - \mu_j) + \max_{u \in [\mu_{j-1}, \mu_j]} |f'(u)|(\mu_j - \mu_{j-1}) \right). \quad (8.3.58)$$

Goal: u_N from (8.3.58) via (8.3.5) satisfies

$$\min_x u_N(x, 0) \leq u_N(x, t) \leq \max_x u_N(x, 0) \quad \forall x \in \mathbb{R}, \forall t \in [0, T]. \quad (8.3.59)$$

For exact solutions this follows from the comparison principle and the preservation of constants, the latter being straightforward here (8.3.15). $\Rightarrow$ establish a **comparison principle** for the semi-discrete solutions: $\vec{\eta}(0) \leq \vec{\mu}(0)$ componentwise $\Rightarrow \eta_j(t) \leq \mu_j(t) \forall j, t$.

Can η_j rise above μ_j ?

1

Touching. $\vec{\mu}, \vec{\eta}$ solve (8.3.58) with $\eta_k(t) \leq \mu_k(t) \forall k$ and $\xi := \eta_j(t) = \mu_j(t)$ for some j . Subtract:

$$\begin{aligned} \frac{d}{dt}(\mu_j - \eta_j) &= -\frac{1}{h}(F_{LF}(\xi, \mu_{j+1}) - F_{LF}(\xi, \eta_{j+1})) \\ &\quad - \frac{1}{h}(F_{LF}(\eta_{j-1}, \xi) - F_{LF}(\mu_{j-1}, \xi)). \end{aligned}$$

2

Sufficient for $\frac{d}{dt}(\mu_j - \eta_j) \geq 0$:

$$F(\xi, \mu_{j+1}) - F(\xi, \eta_{j+1}) \leq 0,$$

$$F(\eta_{j-1}, \xi) - F(\mu_{j-1}, \xi) \leq 0 \quad (8.3.60)$$

from $\mu_{j\pm 1} \geq \eta_{j\pm 1}$: true if F increases in its first and decreases in its second argument.

Definition 8.3.61 (Monotone numerical flux function)

$F = F(v, w)$ is *monotone*, if it is increasing in v ($\forall w$) and decreasing in w ($\forall v$).

Corollary 8.3.62 (Simple criterion)

A C^1 2-point flux is monotone iff $\frac{\partial F}{\partial v} \geq 0$ and $\frac{\partial F}{\partial w} \leq 0$ everywhere (8.3.63).

Lemma 8.3.64: LF and Godunov are monotone

Lemma 8.3.64 (Monotonicity of Lax–Friedrichs/Rusanov flux and Godunov flux)

For any continuously differentiable flux function f the associated Lax–Friedrichs/Rusanov flux (8.3.36) and Godunov flux (8.3.53) are monotone.

1

Local Lax–Friedrichs, criterion (8.3.63) with $M := \max_{\min\{v,w\} \leq u \leq \max\{v,w\}} |f'(u)|$ (treated as a constant locally; the genuine flux needs a case distinction):

$$\frac{\partial F_{\text{LF}}}{\partial v} = \frac{1}{2} f'(v) + \frac{1}{2} M \geq 0,$$

$$\frac{\partial F_{\text{LF}}}{\partial w} = \frac{1}{2} f'(w) - \frac{1}{2} M \leq 0,$$

since $|f'(v)|, |f'(w)| \leq M$.

2

Godunov, $v < w$: $F = \min_{[v,w]} f$. Raising v shrinks the interval: the minimum *increases*. Raising w enlarges it: the minimum *decreases*.



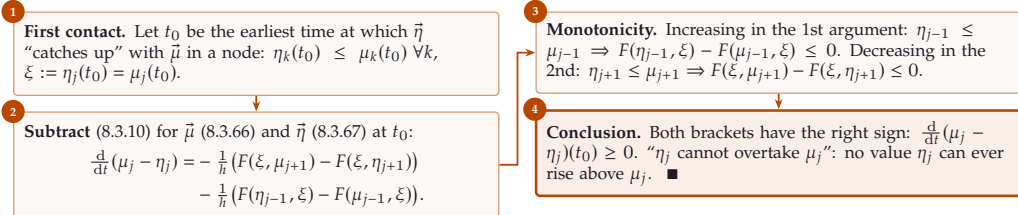
3

Godunov, $v \geq w$: $F = \max_{[w,v]} f$. Raising v enlarges the interval: the maximum *increases*. Raising w shrinks it: the maximum *decreases*. ■

Lemma 8.3.65: discrete comparison principle

Lemma 8.3.65 (Comparison principle for monotone semi-discrete conservative evolutions)

Let $F = F(v, w)$ be monotone and $\vec{\mu}(t), \vec{\eta}(t)$ solve (8.3.10). Then $\eta_k(0) \leq \mu_k(0) \forall k \in \mathbb{Z} \Rightarrow \eta_k(t) \leq \mu_k(t) \forall k \in \mathbb{Z}, \forall 0 \leq t \leq T$.

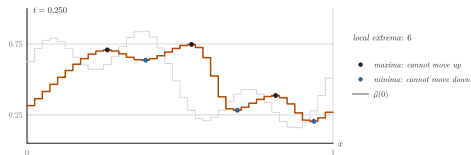


(8.3.68) No creation of discrete local extrema

Local maximum u_m : $\mu_j = u_m = \max_{k_l < l < k_r} \mu_l$,
 $\mu_{k_l}, \mu_{k_r} < u_m$ for some $k_l < j < k_r$.

Lemma 8.3.69 (Non-oscillatory monotone semi-discrete evolutions)

If $\vec{\mu}(t)$ solves (8.3.10) with a monotone F and $\vec{\mu}(0)$ has finitely many local extrema, then $\vec{\mu}(t)$ has no more local extrema than $\vec{\mu}(0)$.



Lax–Friedrichs flux, Burgers: the number of local extrema never grows.

Proof. i = index of a local maximum: $\mu_{i-1} \leq \mu_i \geq \mu_{i+1} \xRightarrow{\text{monotone}} F(\mu_i, \mu_{i+1}) \geq F(\mu_i, \mu_i) \geq F(\mu_{i-1}, \mu_i) \Rightarrow \frac{d}{dt} \mu_i = -\frac{1}{h} (F(\mu_i, \mu_{i+1}) - F(\mu_{i-1}, \mu_i)) \leq 0$: maxima subside, minima rise. ■

8.4 *Timestepping*

(8.4.1) Runge–Kutta single step timestepping

Method of lines: the ODE (8.3.9) lives on $\mathbb{R}^Z$, but single step methods apply formally. Explicit s -stage RK (Def. 6.1.40), Butcher coefficients a_{ij} (strictly lower triangular), b_l , for $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$:

$$\mathbf{k}_1 = \mathbf{f}(\mathbf{y}^{(k)}), \quad \mathbf{k}_i = \mathbf{f}\left(\mathbf{y}^{(k)} + \tau \sum_{j < i} a_{ij} \mathbf{k}_j\right), \quad \mathbf{y}^{(k+1)} := \mathbf{y}^{(k)} + \tau \sum_{l=1}^s b_l \mathbf{k}_l. \quad (8.4.2)$$

Abstract semi-discrete evolution $\frac{d\vec{\mu}}{dt} = \mathbf{L}_h(\vec{\mu}(t))$, $\vec{\mu}(0) = \vec{\mu}_0 \in \mathbb{R}^Z$ (8.4.3), with the (non-linear) finite difference operator, e.g. for 2-point fluxes

$$(\mathbf{L}_h \vec{\mu})_j := -\frac{1}{h} (F(\mu_j, \mu_{j+1}) - F(\mu_{j-1}, \mu_j)), \quad (8.4.4)$$

local: $(\mathbf{L}_h \vec{\mu})_j = L_j(\mu_{j-n_l}, \dots, \mu_{j+n_r})$ (8.4.5). Then explicit RK reads

$$\vec{\kappa}_1 = \mathbf{L}_h(\vec{\mu}^{(k)}), \quad \vec{\kappa}_i = \mathbf{L}_h\left(\vec{\mu}^{(k)} + \tau \sum_{j < i} a_{ij} \vec{\kappa}_j\right), \quad \vec{\mu}^{(k+1)} = \vec{\mu}^{(k)} + \tau \sum_{l=1}^s b_l \vec{\kappa}_l : \quad (8.4.6)$$

only evaluations of the (non-linear, non-smooth) $\mathbf{L}_h$, no non-linear systems of equations.

(8.4.7) Fully discrete evolutions, Ex. 8.4.8

Uniform space-time grid $x_j = hj$, $t_k = \tau k$: $\mu_j^{(k)} \approx u(x_j, t_k)$, $\vec{\mu}^{(k+1)} = \mathcal{H}_h(\vec{\mu}^{(k)})$ with the **fully discrete evolution operator** $\mathcal{H}_h : \mathbb{R}^{\mathbb{Z}} \rightarrow \mathbb{R}^{\mathbb{Z}}$. For the 2-point flux operator (8.4.4):

- explicit Euler (1-stage):

$$(\mathcal{H}_h \vec{\mu}^{(k)})_j = \mu_j^{(k)} - \frac{\tau}{h} (F(\mu_j^{(k)}, \mu_{j+1}^{(k)}) - F(\mu_{j-1}^{(k)}, \mu_j^{(k)})); \quad (8.4.9)$$

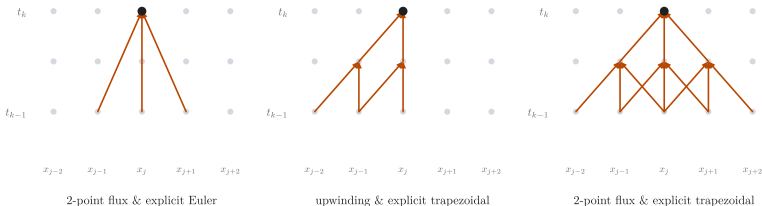
- explicit trapezoidal rule (Heun): $\vec{\kappa} = \vec{\mu}^{(k)} + \tau \mathbb{L}_h(\vec{\mu}^{(k)})$,

$$(\mathcal{H}_h \vec{\mu})_j = \mu_j^{(k)} - \frac{\tau}{2h} (F(\kappa_j, \kappa_{j+1}) - F(\kappa_{j-1}, \kappa_j) + F(\mu_j^{(k)}, \mu_{j+1}^{(k)}) - F(\mu_{j-1}^{(k)}, \mu_j^{(k)})); \quad (8.4.10)$$

- explicit midpoint rule: $\vec{\kappa} = \vec{\mu}^{(k)} + \frac{\tau}{2} \mathbb{L}_h(\vec{\mu}^{(k)})$, $(\mathcal{H}_h \vec{\mu})_j = \mu_j^{(k)} - \frac{\tau}{h} (F(\kappa_j, \kappa_{j+1}) - F(\kappa_{j-1}, \kappa_j))$ (8.4.11)

8.4.1 CFL-condition: *difference stencils*

Explicit timestepping in a method of lines often faces a mesh-dependent timestep constraint (Sects. 6.1.3, 6.2.5). Stencils visualize the flow of information in $\mathcal{H}_h$:



(8.4.13) Locality $(\mathcal{H}_h \vec{\mu})_j = H_j(\mu_{j-m_l}, \dots, \mu_{j+m_r})$ (8.4.14); **translation invariance** $H_j = H$ for $f = f(u)$;
 s -stage RK: $m_l \leq s n_l$, $m_r \leq s n_r$.

(8.4.15) Domains of dependence, CFL

Definition 8.4.16 (Numerical domain of dependence)

Explicit, translation-invariant, $(\mathcal{H}\vec{\mu})_j = H(\mu_{j-m}, \dots, \mu_{j+m})$ (8.4.17), uniform grid:

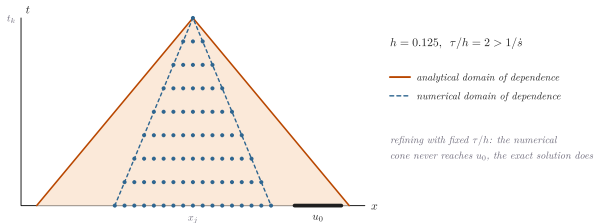
$$D_h^-(x_j, t_k) := \{(x_n, t_l) : j - m(k - l) \leq n \leq j + m(k - l), l \leq k\}.$$

Analytical (Thm. 8.2.43): $D^-(\bar{x}, \bar{t}) := \{\dot{s}_{\min}(\bar{t} - t) \leq \bar{x} - x \leq \dot{s}_{\max}(\bar{t} - t)\}$,
 $\dot{s}_{\min/\max} := \min/\max_{\xi} f'(\xi)$ (8.4.18), (8.4.19).

Definition 8.4.21 (Courant–Friedrichs–Lewy (CFL)-condition, cf. Rem. 6.2.56)

The convex hull of the numerical domain of dependence contains the maximal analytical one, $D^-(x_j, t_k) \subset \text{convex}(D_h^-(x_j, t_k))$. Sufficient: $\frac{\tau}{h} \leq \frac{m}{\dot{s}} \iff$ **timestep constraint** (8.4.22)

Why the CFL-condition is necessary



Sequence of grids with $\tau = \gamma h$, γ beyond (8.4.22): the point (x_j, t_k) is in every grid.

Initial data supported outside the numerical domain of dependence influence the exact solution at (x_j, t_k) , but never $\mu_j^{(k)} \Rightarrow$ possibly $\mu_j^{(k)} \not\rightarrow u(x_j, t_k)$ though $h, \tau \rightarrow 0$.

No convergence for fixed $\tau : h$ if the CFL-condition is violated. Implicit timestepping circumvents it (Rem. 6.1.101), at the price of non-linear systems: the moderate constraint (8.4.22) is usually accepted gladly.

8.4.2 Linear stability: linear advection

Is (8.4.22) also stipulated by stability (cf. $\tau \leq O(h^r)$ in Sects. 6.1.5.2, 6.2.5)? Only for $\frac{\partial u}{\partial t} + v \frac{\partial u}{\partial x} = 0$, $v > 0$ (8.1.10): semi-discretization yields a **linear, local, translation-invariant** evolution, a *stencil formula* of width m :

$$\frac{d\vec{\mu}}{dt}(t) = \mathbf{L}_h(\vec{\mu}(t)), \quad (\mathbf{L}_h \vec{\mu})_j = \sum_{l=-m}^m c_l \mu_{j+l}, \quad j \in \mathbb{Z}. \quad (8.4.24)$$

- *linear*: $\mathbf{L}_h : \mathbb{R}^{\mathbb{Z}} \rightarrow \mathbb{R}^{\mathbb{Z}}$ linear; *local*: $(\mathbf{L}_h \vec{\mu})_j$ depends on few μ_{j+l} ;
- *translation-invariant*: $\eta_j := \mu_{j+1} \Rightarrow (\mathbf{L}_h \vec{\eta})_j = (\mathbf{L}_h \vec{\mu})_{j+1}$.

Example 8.4.25 (Upwind difference operator for linear advection). Godunov flux (= upwind flux here; Lax–Friedrichs gives the same $\mathbf{L}_h$): $(\mathbf{L}_h \vec{\mu})_j = -\frac{v}{h}(\mu_j - \mu_{j-1})$ (8.4.26), i.e. $c_0 = -\frac{v}{h}$, $c_{-1} = \frac{v}{h}$.

The symbol of a difference operator

Diagonalize, with a new twist: L_h acts on $\mathbb{R}^{\mathbb{Z}}$. Trial “eigenvectors” in $\mathbb{C}^{\mathbb{Z}}$, **complex waves**

$$(\vec{\psi}^\xi)_j := \exp(i\xi j), \quad j \in \mathbb{Z}, \quad -\pi < \xi \leq \pi : \quad (8.4.27)$$

infinitely many, matching $\dim \mathbb{C}^{\mathbb{Z}} = \infty$.

1

Apply (8.4.24), $\exp(x + y) = \exp x \exp y$:

$$(L_h \vec{\psi}^\xi)_j = \sum_l c_l e^{i\xi(j+l)} = \left(\sum_{l=-m}^m c_l e^{i\xi l} \right) e^{i\xi j}.$$

2

Spectrum and **symbol** $\hat{c}_h$ of L_h :

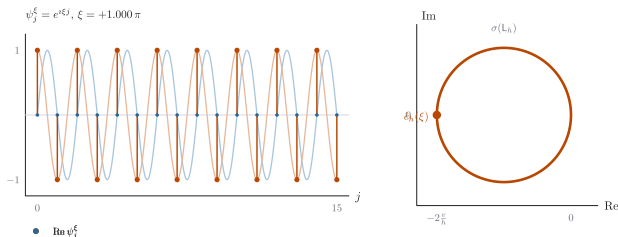
$$\sigma(L_h) = \left\{ \hat{c}_h(\xi) := \sum_{l=-m}^m c_l e^{i\xi l} : -\pi < \xi \leq \pi \right\} \quad (8.4.29)$$

Remark 8.4.28 (Diagonalization in $\mathbb{C}$). Purely real matrices may have complex eigenvalues: purely real difference operators have a complex spectrum.

Remark 8.4.30. L_h is an infinite banded circulant matrix; Fourier harmonics diagonalize circulant matrices [33], the complex waves generalize them to $\mathbb{R}^{\mathbb{Z}}$.

Ex. 8.4.31, (8.4.32): diagonalization

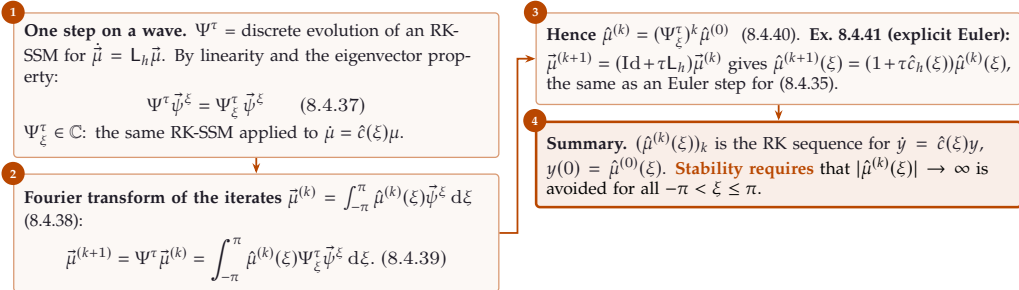
Example 8.4.31. Upwind operator (8.4.26), $c_0 = -\frac{v}{h}$, $c_{-1} = \frac{v}{h}$: $\sigma(L_h) = \left\{ \frac{v}{h} (e^{-i\xi} - 1) : -\pi < \xi \leq \pi \right\}$, a circle of radius $\frac{v}{h}$ around $-\frac{v}{h}$. Each complex wave is an eigenvector; its eigenvalue sits on the circle:



(8.4.32) $\vec{\mu}(0) = \vec{\psi}^\xi \Rightarrow \vec{\mu}(t) = e^{\hat{c}_h(\xi)t} \vec{\psi}^\xi$ (8.4.33); with $\mu_j(t) = \int_{-\pi}^{\pi} \hat{\mu}(t, \xi) e^{i\xi j} d\xi$ (8.4.34):

$\frac{\partial \hat{\mu}}{\partial t}(t, \xi) = \hat{c}_h(\xi) \hat{\mu}(t, \xi)$ (8.4.35), decoupled scalar ODEs. **Rem. 8.4.36:** $\vec{\mu} \leftrightarrow \hat{\mu}$ is the Fourier series, an isometry $\sum_j |\mu_j|^2 = 2\pi \int_{-\pi}^{\pi} |\hat{\mu}|^2 d\xi$.

Runge–Kutta timestepping in Fourier domain



Theorem 8.4.42 (Stability function of explicit Runge–Kutta methods)

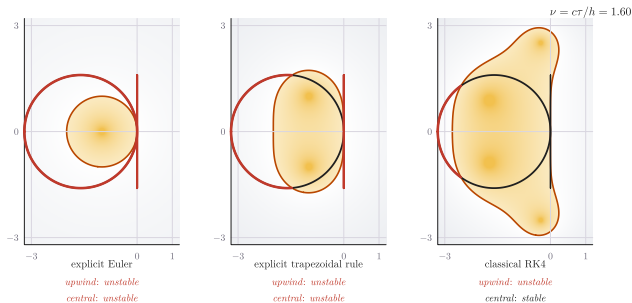
One step of size τ of an explicit s -stage RK-SSM with Butcher scheme $\begin{smallmatrix} c & \mathfrak{A} \\ & \mathbf{b}^T \end{smallmatrix}$ for $\dot{y} = \lambda y$, $\lambda \in \mathbb{C}$, multiplies by

$$\Psi_{\lambda}^{\tau} = \underbrace{1 + z\mathbf{b}^T(\mathbf{I} - z\mathfrak{A})^{-1}\mathbf{1}}_{\text{stability function } S(z)} = \det(\mathbf{I} - z\mathfrak{A} + z\mathbf{1}\mathbf{b}^T), \quad z := \lambda\tau, \quad \mathbf{1} = (1, \dots, 1)^T.$$

Ex. 8.4.43. Euler: $S(z) = 1 + z$; trapezoidal: $1 + z + \frac{1}{2}z^2$; RK4: $1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \frac{1}{24}z^4$. So $\hat{\mu}^{(k+1)}(\xi) = S(\tau\hat{c}(\xi))\hat{\mu}^{(k)}(\xi)$, and

Stability of RK-timestepping of a linear semi-discrete evolution $\iff \max_{-\pi < \xi \leq \pi} |S(\tau\hat{c}(\xi))| \leq 1$ (**von Neumann stability analysis**).

Remark 8.4.44: stability domains



Sweeping $\nu = c\tau/h$: the loci $\tau\hat{c}(\xi)$ of the upwind flux (circle) and the central flux (segment); red where they leave the domain.

Stability domain [33, Def. 12.1.49]:

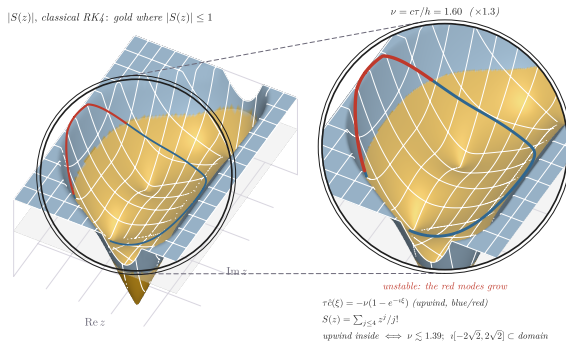
$$\{z \in \mathbb{C} : |S(z)| \leq 1\}.$$

For explicit RK-SSM S is a polynomial: stability domains are *bounded*.

Necessary stability condition in $\mathbb{R}^Z$

$\{\tau\hat{c}(\xi), -\pi < \xi \leq \pi\} \subset \text{stability domain of the RK-SSM}.$

Stability functions over the complex plane



Left: $|S(z)|$ (clipped at 1.7) for Euler, the explicit trapezoidal rule and RK4; gold where $|S| \leq 1$. The lens magnifies the domain: the upwind locus $\tau\hat{c}(\xi)$ lies on the surface, blue while $|S| \leq 1$, red where a Fourier mode is amplified, as ν grows.

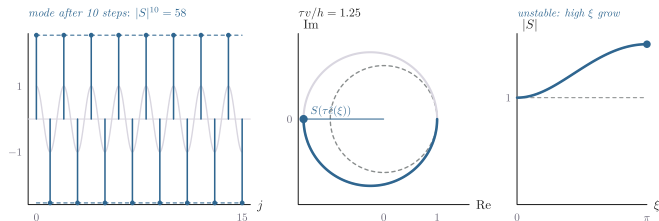
Example 8.4.45: stability and CFL condition

1

Upwind & explicit Euler. $\hat{c}_h(\xi) = \frac{v}{h}(e^{-i\xi} - 1)$, $S(z) = 1 + z$:
 $S(\tau\hat{c}_h(\xi)) = 1 - \frac{\tau v}{h} + \frac{\tau v}{h} e^{-i\xi}$, a circle of radius $\frac{\tau v}{h}$ around $1 - \frac{\tau v}{h}$.

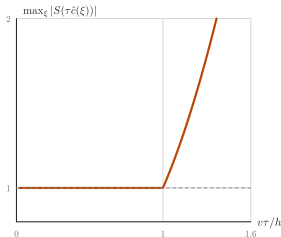
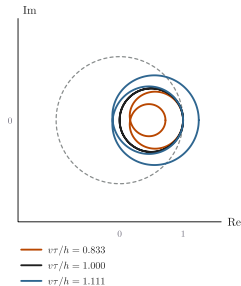
2

Inside the unit circle iff $|1 - \frac{\tau v}{h}| + \frac{\tau v}{h} \leq 1$, i.e. $|S(\tau\hat{c}_h(\xi))| \leq 1 \forall \xi \iff \frac{\tau v}{h} v \leq 1$: the **CFL-condition** of Def. 8.4.21.



Sweeping the frequency ξ from 0 to π : at $\tau v/h = 0.8$ every mode decays; at 1.25 the high frequencies leave the unit disk, and the sawtooth mode near $\xi = \pi$ grows by $|S|^{10} \approx 52$ in ten steps.

Upwind & explicit trapezoidal rule



$S(z) = 1 + z + \frac{1}{2}z^2$: again

$$|S(\tau\hat{c}(\xi))| \leq 1 \quad \forall \xi \iff \frac{\tau}{h}v \leq 1.$$

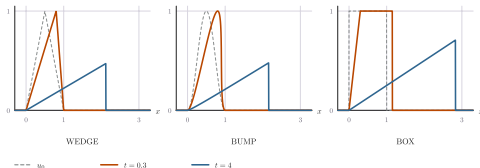
Tighter than the mere CFL-condition (8.4.22): the trapezoidal rule has 2 stages, so the stencil has width 2 in the upwind direction (Fig. (8.4.12), middle).

Stability induced timestep constraint. For explicit RK-SSM on (8.4.24), $\max_{\xi} |S(\tau\hat{c}(\xi))| \leq 1$ implies a timestep constraint.

8.4.3 Convergence: Experiment 8.4.47

Burgers (8.1.46), data supported in $[0, 1]$:

- BUMP: $u_0 = 1 - \cos^2(\pi x)$,
- WEDGE: $u_0 = 1 - 2|x - \frac{1}{2}|$,
- BOX: $u_0 = 1$ on $[0, 1]$;



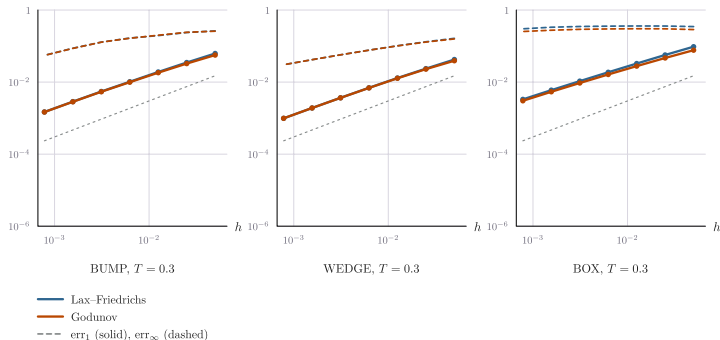
$\dot{s} = 1$; LF (8.3.36), Godunov (8.3.53), RK4,
 $\tau : h = 1, h = \frac{1}{20}, \dots, \frac{1}{1280}$.

$$\text{err}_1(h) := \max_{k>0} h \sum_j |\mu_j^{(k)} - u(x_j, t_k)|, \quad (8.4.48)$$

$$\text{err}_\infty(h) := \max_{k>0} \max_{j \in \mathbb{Z}} |\mu_j^{(k)} - u(x_j, t_k)|, \quad (8.4.49)$$

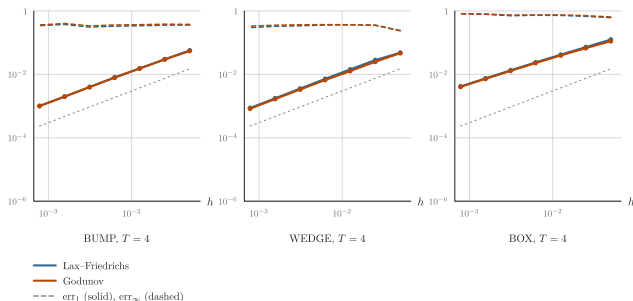
the natural norms by Thms. 8.2.41, 8.2.43 (cf. the energy norm for elliptic BVPs). Here against exact cell averages (Hopf–Lax), $t = \frac{T}{4}, \dots, T$.

Convergence before shocks: $T = 0.3$



BUMP is still smooth at $T = 0.3$ (characteristics intersect only at $t = 1/\pi$), but already very steep.

Convergence after shock formation: $T = 4$



Observations (either flux): (near) **first order** algebraic convergence in err_1 ; algebraic convergence in err_∞ before shocks form; **no convergence** in err_∞ after shock formation. Best we get: $O(h)$.

Smooth entropy solution u ; **consistency error** of a numerical flux $F = F(v, w)$:

$$(\vec{\tau}_F(t))_j = F(u(x_j, t), u(x_{j+1}, t)) - f(u(x_{j+\frac{1}{2}}, t));$$

$\max_j (\vec{\tau}_F(t))_j = O(h^q) \leftrightarrow$ consistent of order q (8.4.50) **Rule of thumb**: the order of consistency of the numerical flux limits the algebraic order of convergence.

1 **Example 8.4.51 (upwind).** $f \in C^1$, $u_0 \geq 0$, $f'(u) \geq 0$ for $u \geq 0$ (no transonic rarefactions): $F_{\text{uw}} = F_{\text{GD}}$ and $F_{\text{uw}}(u(x_j), u(x_{j+1})) = f(u(x_j))$.

2 **Taylor of f and u :**

$$(\vec{\tau}_{F_{\text{uw}}})_j = f(u(x_j)) - f(u(x_{j+\frac{1}{2}}))$$
$$= -f'(u) \frac{\partial u}{\partial x} \Big|_{x_{j+\frac{1}{2}}} \frac{h}{2} + O(h^2) :$$

only **first order** consistent.

Ex. 8.4.52: consistency of the LF flux

- 1 **Central part.** Taylor expansion around $x_{j+\frac{1}{2}}$:
$$\begin{aligned} & \frac{1}{2} (f(u_j) + f(u_{j+1})) - f(u_{j+\frac{1}{2}}) \\ &= \frac{1}{2} f'(u_{j+\frac{1}{2}}) (u_j + u_{j+1} - 2u_{j+\frac{1}{2}}) + O(h^2) \\ &= \frac{1}{2} f' \frac{\partial u}{\partial x} \left(-\frac{h}{2} + \frac{h}{2} \right) + O(h^2) = O(h^2) \end{aligned}$$

($u_j := u(x_j, t)$): the central flux is *second order* consistent (8.4.53), but useless alone (Sect. 8.3.3.1).
- 2 **Diffusive part.** $u(x_{j+1}) - u(x_j) = \frac{\partial u}{\partial x}(x_{j+\frac{1}{2}})h + O(h^2)$,
multiplied by $\frac{1}{2} \max |f'|$: an $O(h)$ term.
- 3 **Hence** $F_{\text{LF}}(u(x_j), u(x_{j+1})) - f(u(x_{j+\frac{1}{2}})) = O(h)$: only first order; the error is dominated by the diffusive flux, needed for stability.

Order barrier for monotone numerical fluxes

Monotone numerical fluxes (Def. 8.3.61) are at most first order consistent.

1. Conduct a von Neumann stability analysis for the linear evolution

$$\dot{\mu}_j = \frac{\mu_{j+1} - 2\mu_j + \mu_{j-1}}{h^2}, \quad h > 0, \quad (8.4.56)$$

when explicit Euler timestepping / the explicit trapezoidal rule is used for discretization in time.

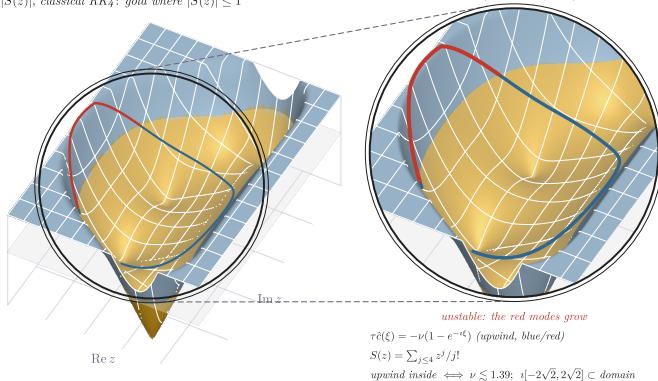
Hint. Symbol $\hat{c}_h(\xi) = \frac{2}{h^2}(\cos \xi - 1) \in [-\frac{4}{h^2}, 0]$; compare with the real intervals contained in the stability domains of Rem. 8.4.44.

- Naive finite differences can converge to the **wrong** solution (Ex. 8.3.1).
- **Conservation form** (8.3.9) with a consistent numerical flux: discrete conservation and the correct discrete shock speed.
- Central fluxes oscillate; Lax–Friedrichs adds diffusion; upwind can produce **expansion shocks**; Godunov solves local Riemann problems.
- **Monotone** fluxes: comparison principle, no new extrema.
- **CFL-condition**: the numerical domain of dependence must contain the analytical one.
- Von Neumann analysis:
 $\max_{\xi} |S(\tau \hat{c}(\xi))| \leq 1$; for upwind & Euler exactly $\tau v/h \leq 1$.
- Convergence at best $O(h)$ in L^1 ; monotone fluxes are at most first order consistent.

Next lecture. Higher order with slope limiting (MUSCL), systems, and finite elements for the Stokes equations.

$|S(z)|$, classical RK4: gold where $|S(z)| \leq 1$

$$\nu = c\tau/h = 1.60 \quad (\times 1.3)$$



Thank you.

Johannes Herter