Properties of numerical methods

The following criteria are crucial to the performance of a numerical algorithm:

1. Consistency — The discretization of a PDE should become exact as the mesh size tends to zero (truncation error should vanish)
2. Stability — Numerical errors which are generated during the solution of discretized equations should not be magnified
3. Convergence — The numerical solution should approach the exact solution of the PDE and converge to it as the mesh size tends to zero
4. Conservation — Underlying conservation laws should be respected at the discrete level (artificial sources/sinks are to be avoided)
5. Boundedness — Quantities like densities, temperatures, concentrations etc. should remain nonnegative and free of spurious wiggles

These properties must be verified for each (component of the) numerical scheme

Consistency

Relationship: discretized equation $\longleftrightarrow$ differential equation

Truncation errors should vanish as the mesh size and time step tend to zero

Example. Pure convection equation $\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial x} = 0$ discretized by

CDS in space, FE in time: $$\frac{u_i^{n+1} - u_i^n}{\Delta t} + v\frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} = \mathcal{O}[(\Delta t)^q, (\Delta x)^p]$$

Taylor series expansions: $$u_i^{n+1} = u_i^n + \Delta t \left(\frac{\partial u}{\partial t}\right)_i^n + \frac{(\Delta t)^2}{2}\left(\frac{\partial^2 u}{\partial t^2}\right)_i^n + \dots$$

$$u_{i\pm1}^n = u_i^n \pm \Delta x \left(\frac{\partial u}{\partial x}\right)_i^n + \frac{(\Delta x)^2}{2}\left(\frac{\partial^2 u}{\partial x^2}\right)_i^n \pm \frac{(\Delta x)^3}{6}\left(\frac{\partial^3 u}{\partial x^3}\right)_i^n + \dots$$

Hence, $\frac{u_i^{n+1} - u_i^n}{\Delta t} + v\frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} - \left(\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial x}\right)_i^n + \epsilon_\tau = 0$ where

$$\epsilon_\tau = -\frac{\Delta t}{2}\left(\frac{\partial^2 u}{\partial t^2}\right)_i^n - v\frac{(\Delta x)^2}{6}\left(\frac{\partial^3 u}{\partial x^3}\right)_i^n + \mathcal{O}[(\Delta t)^2, (\Delta x)^4]$$

residual of the difference scheme for the *exact* nodal values $u_j^m = u(j\Delta x, m\Delta t)$

Stability

Relationship: numerical solution of discretized equations $\longleftrightarrow$ exact solution of discretized equations

Definition 1 Numerical errors (roundoff due to final precision of computers) should not be allowed to grow unboundedly

Definition 2 The numerical solution itself should remain uniformly bounded

Stability analysis: can only be performed for a very limited range problems

Matrix method: $Au^{n+1} = Bu^n \quad \Rightarrow \quad u^{n+1} = Cu^n$, where $C = A^{-1}B$

is assumed to be a linear operator. In practice $u^n = \bar{u}^n + e^n$ so that

$u^{n+1} = Cu^n$ for the numerical solution u^n of the discretized equations

$\bar{u}^{n+1} = C\bar{u}^n$ for the exact solution $\bar{u}^n$ of the discretized equations

$e^{n+1} = Ce^n$ for the roundoff error e^n incurred in the solution process

Matrix method for stability analysis

In the linear case $u^{n+1} = Cu^n = \ldots = C^n u^0, \qquad e^{n+1} = Ce^n = \ldots = C^n e^0$

i.e., the error evolves in the same way as the solution and is bounded by

$$||e|| \leq ||C||^n ||e^0||, \qquad ||C|| \geq \rho(C) = \max_i |\lambda_i| \qquad \textit{spectral radius of } C$$

Unstable schemes: if $\rho(C) > 1$ then $||C|| \geq 1$ and the errors may grow

Example. Convection-diffusion equation $\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial x} = d\frac{\partial^2 u}{\partial x^2}$ discretized by

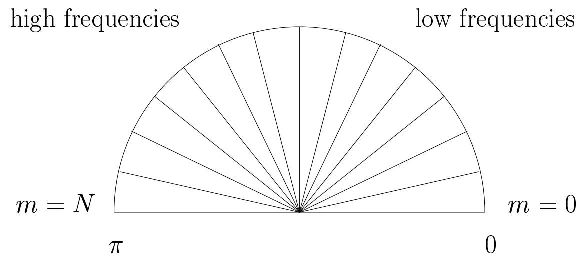
CDS in space, FE in time: $\frac{u_i^{n+1} - u_i^n}{\Delta t} + v\frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} = d\frac{u_{i-1}^n - 2u_i^n + u_{i+1}^n}{(\Delta x)^2}$

or $$u_i^{n+1} = u_i^n - \frac{\nu}{2}(u_{i+1}^n - u_{i-1}^n) + \delta(u_{i-1}^n - 2u_i^n + u_{i+1}^n)$$

where $\nu = v\frac{\Delta t}{\Delta x}$ is the *Courant number*, $\delta = d\frac{\Delta t}{(\Delta x)^2}$ is the *diffusion number*

Matrix method for stability analysis

$$C = \begin{bmatrix} \cdot & \cdot & \cdot & & & \\ & a & b & c & & \\ & & a & b & c & \\ & & & a & b & c \\ & & & & \cdot & \cdot & \cdot \end{bmatrix} \qquad a = \delta + \frac{\nu}{2}, \quad b = 1 - 2\delta, \quad c = \delta - \frac{\nu}{2}$$



Eigenvectors $\quad \varphi_j^{(m)} = \cos\theta_m j + i\sin\theta_m j = e^{i\theta_m j}, \quad \theta_m = m\frac{\pi}{N}, \quad i^2 = -1$

Eigenvalues $\quad a\varphi_{j+1}^{(m)} + b\varphi_j^{(m)} + c\varphi_{j-1}^{(m)} = \lambda_m \varphi_j^{(m)}$, $\quad$ divide by $\varphi_j^{(m)} = e^{i\theta_m j}$

$$\Rightarrow \quad \lambda_m = 1 + 2\delta(\cos\theta_m - 1) + i2\nu\sin\theta_m, \qquad \forall m = 1, \dots, N$$

Stability condition $\quad \boxed{|\lambda_m|^2 = [1 + 2\delta(\cos\theta_m - 1)]^2 + 4\nu^2\sin^2\theta_m \leq 1}$

pure convection: $\quad \delta = 0 \quad \Rightarrow \quad |\lambda_m| \geq 1$ $\quad$ unconditionally unstable :-(

pure diffusion: $\quad \nu = 0 \quad \Rightarrow \quad |\lambda_m| \leq 1$ if $\delta \leq \frac{1}{2}$ $\quad$ conditionally stable

Von Neumann's stability analysis

Objective: to investigate the propagation and amplification of numerical errors

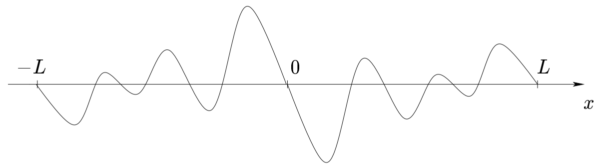
Assumptions: linear PDE, constant coefficients, periodic boundary conditions

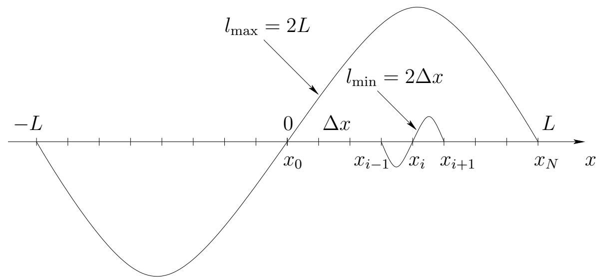
Continuous error representation

$e(x,t) = \sum_m a_m(t) e^{ik_m x}$ Fourier series

$i^2 = -1, \quad e^{ik_m x} = \cos k_m x + i \sin k_m x$

i.e. the error is a superposition of harmonics characterized by their wave number $k_m = \frac{2\pi}{l_m}$ (for wave length l_m) and amplitide $a_m(t)$





Discretization $\quad \Delta x = \frac{L}{N} \quad \Rightarrow \quad k_m = m\frac{\pi}{L} = \frac{m\pi}{N\Delta x}, \quad \theta_m = k_m \Delta x = m\frac{\pi}{N}$

Here θ_m is the phase angle, m is the number of waves fitted into the interval $(-L, L)$ and Δx determines the highest frequency resolvable on the mesh

Von Neumann's stability analysis

Representation of numerical error (trigonometric interpolation)

$$e_j^0 = \sum_m a_m^0 e^{i\theta_m j} \quad \Rightarrow \quad e_j^n = \sum_m a_m^n e^{i\theta_m j}, \quad \text{where} \quad a_m^n = a_n^0 \lambda_m^n$$

Due to linearity, the error satisfies the discretized equation and so does each harmonic. Hence, it suffices to check stability for $e_j^n = a_m^n e^{i\theta_m j}, \quad \forall m$

Amplification factor

$$G_m = \frac{a_m^{n+1}}{a_m^n} = \lambda_m$$

the enhancement of the $m-$th harmonic during one time step

Stability condition

$$|G_m| \leq 1, \quad \forall m$$

guarantees that the error component $e_j^n = (G_m)^n e_j^0$ remains bounded

Remark. The accuracy of approximation can be assessed by analyzing phase errors i.e. the actual speed of harmonics as compared to the exact speed

Example: pure convection equation in 1D

1. Let $\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial x} = 0$ be discretized by CDS in space, FE in time

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + v\frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} = 0 \qquad \text{and substitute} \quad e_j^n = a^n e^{i\theta j}$$

The resulting difference equation for the error can be written as

$$(a^{n+1} - a^n)e^{i\theta j} + \frac{\nu}{2}a^n(e^{i\theta(j+1)} - e^{i\theta(j-1)}) = 0, \qquad \nu = v\frac{\Delta t}{\Delta x}$$

Divide by $e^{i\theta j}$ $\Rightarrow$ $a^{n+1} = a^n - \frac{\nu}{2}a^n(e^{i\theta} - e^{-i\theta})$ and note that

$$e^{i\theta} - e^{-i\theta} = \cos\theta + i\sin\theta - \cos\theta + i\sin\theta = 2i\sin\theta$$

Amplification factor $\boxed{G = \frac{a^{n+1}}{a^n} = 1 - i\nu\sin\theta}$ is responsible for stability

$|G|^2 = 1 + \nu^2\sin^2\theta \geq 1$ $\Rightarrow$ the scheme is unconditionally unstable :-(

Example: pure convection equation in 1D

2. Let $\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial x} = 0$ be discretized by BDS in space, FE in time

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + v\frac{u_j^n - u_{j-1}^n}{\Delta x} = 0 \qquad \text{and substitute} \quad e_j^n = a^n e^{i\theta j}$$

which yields $\quad (a^{n+1} - a^n)e^{i\theta j} + \nu a^n(e^{i\theta j} - e^{i\theta(j-1)}) = 0, \qquad \nu = v\frac{\Delta t}{\Delta x}$

$$G = 1 - \nu + \nu e^{-i\theta} = 1 - \nu + \nu(\cos\theta - i\sin\theta) = 1 - 2\nu\sin^2\frac{\theta}{2} - i\nu\sin\theta$$

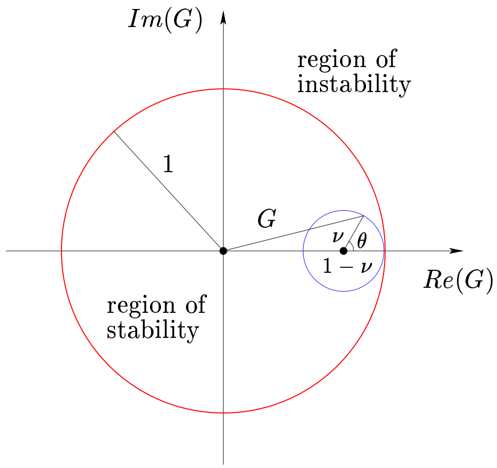
$$Re(G) = 1 - \nu + \nu\cos\theta, \quad Im(G) = -\nu\sin\theta$$

Stability restriction $|G|^2 \leq 1$ means that G must lie within the unit circle in the complex plane.

This leads to the *CFL condition* $\boxed{0 \leq \nu \leq 1}$

$v > 0$ upwind scheme, stable for $\Delta t \leq \frac{\Delta x}{v}$

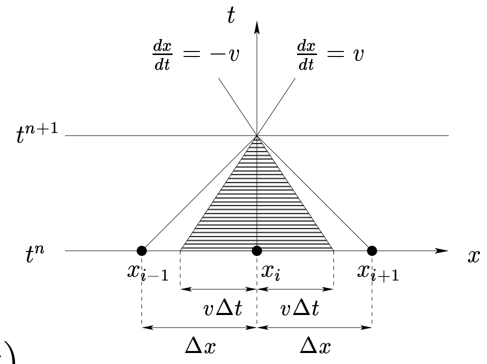
$v < 0$ downwind scheme, unconditionally unstable



Example: pure convection equation in 1D

The numerical domain of dependence should contain the analytical one:

- if $\nu > 1$, then the data at some grid point may affect the true solution but not the numerical one
- on the other hand, for $\nu < 1$ some grid points influence the solution although they should not
- for accuracy reasons it is desirable to have $\nu \approx 1$; some schemes are exact for $\nu = 1$ (unit CFL property)



3. Let $\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial x} = 0$ be discretized by CDS in space, BE in time

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + v\frac{u_{j+1}^{n+1} - u_{j-1}^{n+1}}{2\Delta x} = 0 \qquad \text{and substitute} \quad e_j^n = a^n e^{i\theta j}$$

$$(a^{n+1} - a^n)e^{i\theta j} + \tfrac{\nu}{2}a^{n+1}(e^{i\theta(j+1)} - e^{i\theta(j-1)}) = 0 \qquad \Rightarrow \qquad G = \tfrac{1}{1+i\nu\sin\theta}$$

It follows that $\quad |G|^2 = G \cdot \bar{G} = \frac{1}{1+\nu^2\sin^2\theta} \leq 1 \quad$ unconditional stability

Spectral analysis of numerical errors

Consider $\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial x} = d\frac{\partial^2 u}{\partial x^2}$ convection-diffusion equation

Exact solution: $u(x,t) = e^{ik(x-vt)-k^2 dt} = a(t)e^{ikx}, \quad a(t) = e^{-ikvt-k^2 dt}$

a wave with exponentially decaying amplitude traveling at constant speed

Amplification factor $G_{\text{ex}} = \frac{a(t^{n+1})}{a(t^n)} = \frac{e^{-(ikv+k^2 d)(n+1)\Delta t}}{e^{-(ikv+k^2 d)n\Delta t}} = e^{-(\delta+i\omega)}$

$$|G_{\text{ex}}| = e^{-\delta}, \qquad \delta = k^2 d\Delta t, \qquad \omega = kv\Delta t = -\arg(G_{\text{ex}})$$

Amplitude error $\epsilon_\delta = \frac{|G_{\text{num}}|}{|G_{\text{ex}}|} = |G_{\text{num}}|e^{\delta}$ numerical damping

Phase error $\epsilon_\omega = \frac{\arg(G_{\text{num}})}{\arg(G_{\text{ex}})} = \frac{\arg(G_{\text{num}})}{-\omega} = \frac{\tilde{v}}{v}$ numerical dispersion

where $\tilde{v} = \frac{\arg(G_{\text{num}})}{-k\Delta t}$ is the numerical propagation speed

Harmonics travel too fast if $\epsilon_\omega > 1$ (leading phase error)
and too slow in the case $\epsilon_\omega < 1$ (lagging phase error)

Convergence

Relationship: numerical solution of discretized equations $\longleftrightarrow$ exact solution of the differential equation

Definition: A numerical scheme is said to be convergent if it produces the exact solution of the underlying PDE in the limit $h \to 0$ and $\Delta t \to 0$

Lax equivalence theorem: **stability + consistency = convergence**

- For practical purposes, convergence can be investigated numerically by comparing the results computed on a series of successively refined grids
- The rate of convergence is governed by the leading truncation error of the discretization scheme and can also be estimated numerically:

$$u = u_h + e(u)h^p + \ldots = u_{2h} + e(u)(2h)^p + \ldots = u_{4h} + e(u)(4h)^p + \ldots$$

$$\begin{aligned} u_{2h} - u_h &\approx e(u)h^p(1-2^p) \\ u_{4h} - u_{2h} &\approx e(u)h^p(1-2^p)2^p \end{aligned} \quad \Rightarrow \quad p \approx \frac{\log\left(\frac{u_{4h}-u_{2h}}{u_{2h}-u_h}\right)}{\log 2}$$

Conservation

Physical principles should apply at the discrete level: if mass, momentum and energy are conserved, they can only be distributed improperly

Integral form of a generic conservation law

$$\underbrace{\frac{\partial}{\partial t}\int_V u\,dV}_{accumulation} + \underbrace{\int_S \mathbf{f}\cdot\mathbf{n}\,dS}_{influx} = \underbrace{\int_V q\,dV}_{source/sink}, \qquad \underbrace{\mathbf{f} = \mathbf{v}u - d\nabla u}_{flux\ function}$$

Caution: nonconservative discretizations may produce reasonably looking results which are totally wrong (e.g. shocks moving with a wrong speed)

- even nonconservative schemes can be consistent and stable
- correct solutions are recovered in the limit of very fine grids

Problem: it is usually unclear whether or not the mesh is sufficiently fine

Discrete conservation

1. Any **finite volume scheme** is conservative by construction both locally (for every single control volume) and globally (for the whole domain)

2. A **finite difference scheme** is conservative if it can be written in the form

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + \frac{f_{i+1/2} - f_{i-1/2}}{\Delta x} = q$$

which is equivalent to a vertex-centered finite volume discretization

3. Any **finite element scheme** is conservative, at least globally

$$\sum_{i=1}^{N} \varphi_i \equiv 1, \qquad \int_V \varphi_i \left[\frac{\partial u_h}{\partial t} + \nabla \cdot f_h - q_h \right] dV = 0, \quad i = 1, \ldots, N$$

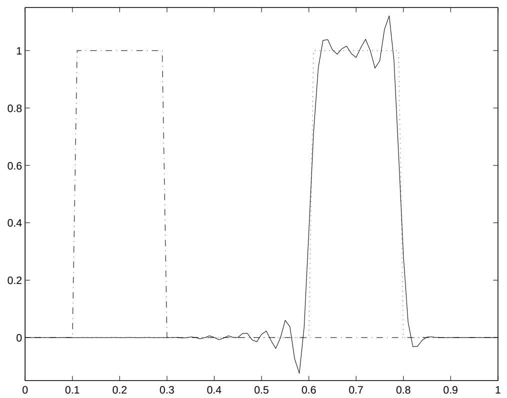
Summation over i yields a discrete counterpart of the integral conservation law

$$\int_V \left[\tfrac{\partial u_h}{\partial t} + \nabla \cdot f_h - q_h \right] dV = 0 \quad \Rightarrow \quad \tfrac{\partial}{\partial t} \int_V u_h \, dV + \int_S \mathbf{f}_h \cdot \mathbf{n} \, dS = \int_V q_h \, dV$$

Boundedness

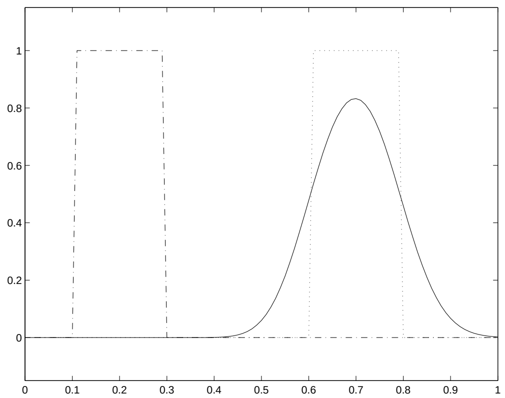
Convection-dominated / hyperbolic PDEs $\qquad Pe \gg 1, \quad Re \gg 1$

- spurious undershoots and overshoots occur in the vicinity of steep gradients
- quantities like densities, temperatures and concentrations become negative
- the method may become unstable or converge to a wrong weak solution



low-order $\longrightarrow$

$\longleftarrow$ high-order



Idea: make sure that important properties of the exact solution (monotonicity, positivity, nonincreasing total variation) are inherited by the numerical one

Design of nonoscillatory methods

Monotone methods $\quad \frac{\partial u}{\partial t} + \frac{\partial f}{\partial x} = 0 \quad \rightarrow \quad u_i^{n+1} = H(u^n; i) \quad$ such that

$$\frac{\partial H(u^n; i)}{\partial u_j^n} \geq 0, \quad \forall i, j \qquad \text{Then} \quad v_i^n \geq u_i^n, \quad \forall i \quad \Rightarrow \quad v_i^{n+1} \geq u_i^{n+1}, \quad \forall i$$

Example. Let $\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial x} = 0$ be discretized by UDS in space, FE in time

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + v\frac{u_i^n - u_{i-1}^n}{\Delta x} = 0, \qquad H(u^n; i) = u_i^n - v\frac{\Delta t}{\Delta x}(u_i^n - u_{i-1}^n)$$

Derivatives $\quad \frac{\partial H(u^n;i)}{\partial u_i^n} = 1 - \nu, \quad \frac{\partial H(u^n;i)}{\partial u_{i-1}^n} = \nu, \quad$ where $\quad \nu = v\frac{\Delta t}{\Delta x}$

$\Rightarrow$ monotone under the CFL condition $\nu \leq 1$ (cf. stability analysis)

Lax-Wendroff theorem: *If a monotone consistent and conservative method converges, then it converges to a physically acceptable weak solution*

Design of nonoscillatory methods

Godunov's theorem: *Monotone method are at most first-order accurate*

Monotonicity-preserving methods (monotone if linear)

$$u_i^0 \geq u_{i+1}^0, \quad \forall i \qquad \Rightarrow \qquad u_i^n \geq u_{i+1}^n, \quad \forall i, \ \forall n$$

If the initial data u^0 is monotone, then so is the solution u^n at all times

It is known that the *total variation* defined as $TV(u) = \int_{-\infty}^{\infty} \left|\frac{\partial u}{\partial x}\right| dx$ is a nonincreasing function of time for any physically admissible weak solution

Total variation diminishing methods (monotone if linear)

$$TV(u^{n+1}) \leq TV(u^n), \quad \text{where} \quad TV(u^n) = \sum_i |u_i^n - u_{i-1}^n|$$

Classification *monotone* $\Rightarrow$ *TVD* $\Rightarrow$ *monotonicity-preserving*

Total variation diminishing methods

Harten's theorem: *An explicit finite difference scheme of the form*

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

is total variation diminishing (TVD) provided that the coefficients satisfy

$$c_{i-1/2} \geq 0, \qquad c_{i+1/2} \geq 0, \qquad c_{i-1/2} + c_{i+1/2} \leq 1$$

Semi-discrete problem $\frac{du_i}{dt} + \frac{f_{i+1/2} - f_{i-1/2}}{\Delta x} = 0$ conservation form

Idea: switch between high- and low-order flux approximations depending on the local smoothness of the solution so as to enforce Harten's conditions:

$$\boxed{f_{i+1/2} = f_{i+1/2}^L + \Phi_{i+1/2}[f_{i+1/2}^H - f_{i+1/2}^L]}$$

where $0 \leq \Phi_{i+1/2} \leq 2$ is a solution-dependent correction factor (flux limiter)

TVD discretization of convective terms

Example. Pure convection equation $\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial x} = 0, \quad v > 0, \quad f = vu$

Linear flux approximations

$f^L_{i+1/2} = vu_i$ *upwind difference*

$f^H_{i+1/2} = v\frac{u_{i+1}+u_i}{2}$ *central difference*

Smoothness indicator

$$r_i = \frac{u_i - u_{i-1}}{u_{i+1} - u_i}$$

Nonlinear TVD flux $f_{i+1/2} = vu_i + \frac{v}{2}\Phi(r_i)(u_{i+1} - u_i)$

Harten's coefficients $c_{i-1/2} = \frac{v}{2\Delta x}\left[2 + \frac{\Phi(r_i)}{r_i} - \Phi(r_{i-1})\right], \quad c_{i+1/2} = 0$

Standard flux limiters: $\Phi(r) = \frac{r+|r|}{1+|r|}$ Van Leer

$\Phi(r) = \max\{0, \min\{1, r\}\}$ minmod

$\Phi(r) = \max\{0, \min\left\{\frac{1+r}{2}, 2, 2r\right\}\}$ MC

$\Phi(r) = \max\{0, \min\{1, 2r\}, \min\{2, r\}\}$ superbee

1D stencil