Overview of classical schemes

Roch SMETS,

These slides and the pdf of the course are available on **ppf.public** :
`Courses/Cores/C5_Numerical_methods_and_simulation_codes/Smets/`

Conservative form

Any finite volume method uses the conservative form of the hyperbolic equation :

$$w_j^{n+1} = w_j^n + \nu[F_{j+1/2}^n - F_{j-1/2}^n] \tag{1}$$

with $\nu = \Delta t/\Delta x$.

- w_j^n : exact solution of $\Xi(w) = 0$ which approximates $D(u) = 0$.
$\rightarrow$ that is for a consistant scheme.

- F : numerical flux which approximates the physical flux f.
$\rightarrow$ is defined at the cell interface.

Naive approach

The simplest approach to set the value of the numerical flux $F^n_{j+1/2}$ could be the FTFS scheme with the flux

$$F^n_{j+1/2} = f(w^n_{j+1}) \tag{2}$$

With such a flux, small-scale oscillations grow even faster than for the the Euler explicit scheme.
$\rightarrow$ this scheme is useless.

A first alternative would be to choose a centered form for the numerical flux like

$$F^n_{j+1/2} = \frac{1}{2}[f(w^n_{j+1}) + f(w^n_j)] \tag{3}$$

which is also inoperant : even its fully implicit form (hence with a BTCS stencil) couple the two disadvantages of smearing out heavily any large scale structures but also let somme wiggles appear.

Lax-Wendroff schemes

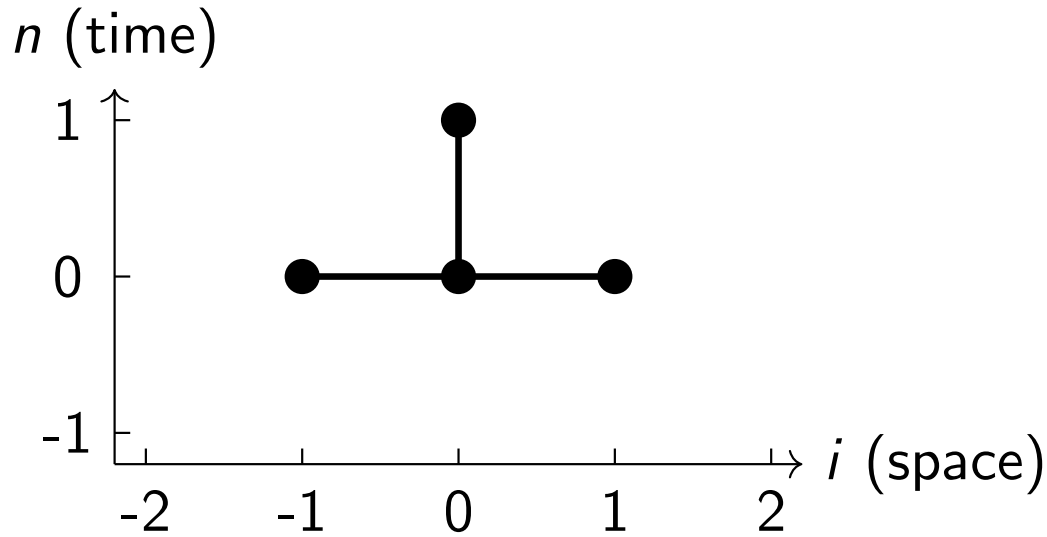
Remark : The main idea of all the Lax-Wendroff type schemes is to provide a correction of the flux given by Eq. (3) using the artificial viscosity form already suggested.

Notation : These types of schemes can be called "first generation" schemes.

This scheme by [lax, 1960] is based on a Taylor expansion in **time**, up to the third order. As a result, the Lax-Wendroff scheme is $O(\Delta x^2, \Delta t^2)$. In the linear case of a physical flux defined as $f(x,t) = Au(x,t)$ where A is a real constant, the numerical flux writes

$$F^n_{j+1/2} = \frac{1}{2}[f(w^n_{j+1}) + f(w^n_j)] - \frac{A^2\nu}{2}[w^n_{j+1} - w^n_j] \qquad (4)$$

Lax-Wendroff stencil



Remark : In the non-linear case, we need to find an appropriate form for $A = \partial_u f$. Different forms have been proposed, all constituting the wide class of Lax-Wendroff type scheme.

Lax-Wendroff is a "high-order" scheme

For initial conditions having strong gradients or even discontinuities, this scheme produces overshoots rising very quickly, even if they generally does not much grow with time.

Such a scheme is hence usefull for smooth initial conditions, providing that no stiff gradients appear later.

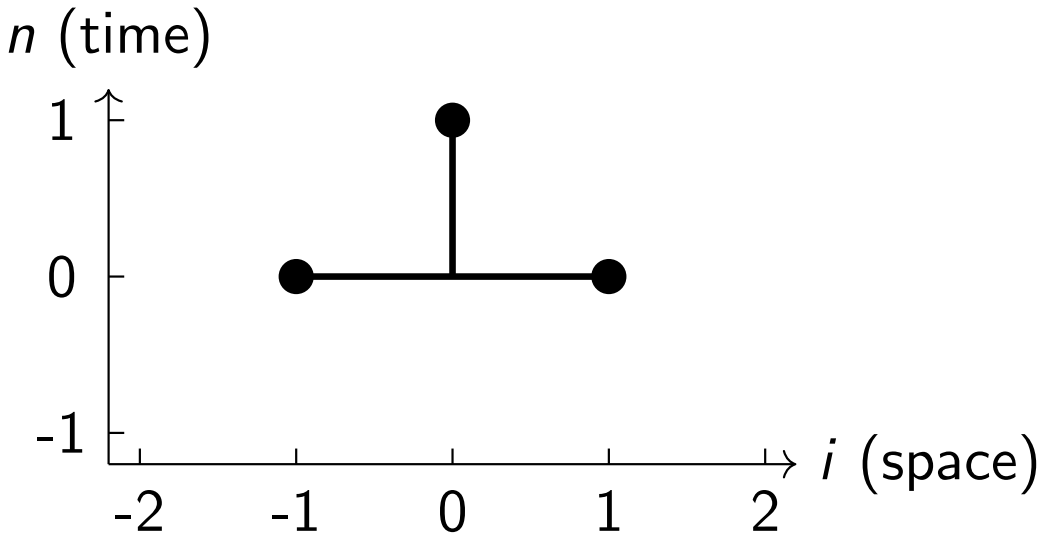
This Lax-Wendroff scheme can be generalized to the nonlinear cases. It is then necessary to replace the A value in Eq. (4) by a numerical approximation of a non-local $A^n_{j+1/2}$ value.

Lax-Friedrichs scheme

The flux of the Lax-Friedrichs scheme [Lax, 1954] writes

$$F^n_{j+1/2} = \frac{1}{2}[f(w^n_{j+1}) + f(w^n_j)] - \frac{1}{2\nu}[w^n_{j+1} - w^n_j] \tag{5}$$

hence, reporting this expression in Eq. (3) gives the stencil



A corrected scheme by Richtmyer

The lax-Friedrichs scheme has the same kind of disadvantage of the FTFS scheme and is hence never used.

[Richtmyer, 1962] proposed a 2-steps form using a (staggered) predictor-corrector for w_j^n :

$$w_{j+1/2}^{\star} = \frac{w_{j+1}^n + w_j^n}{2} - \frac{1}{2}\nu[f(w_{j+1}^n) - f(w_j^n)] \quad (6)$$

$$w_j^{n+1} = w_j^n - \nu[f(w_{j+1/2}^{\star}) - f(w_{j-1/2}^{\star})] \quad (7)$$

A corrected scheme by Mac-Cormack

[Mac-Cormack, 1969] improved this scheme using a FS and BS centering in a non-staggered predictor-corrector to obtain :

$$w_j^\star = w_j^n - \nu[f(w_{j+1}^n) - f(w_j^n)] \quad (8)$$

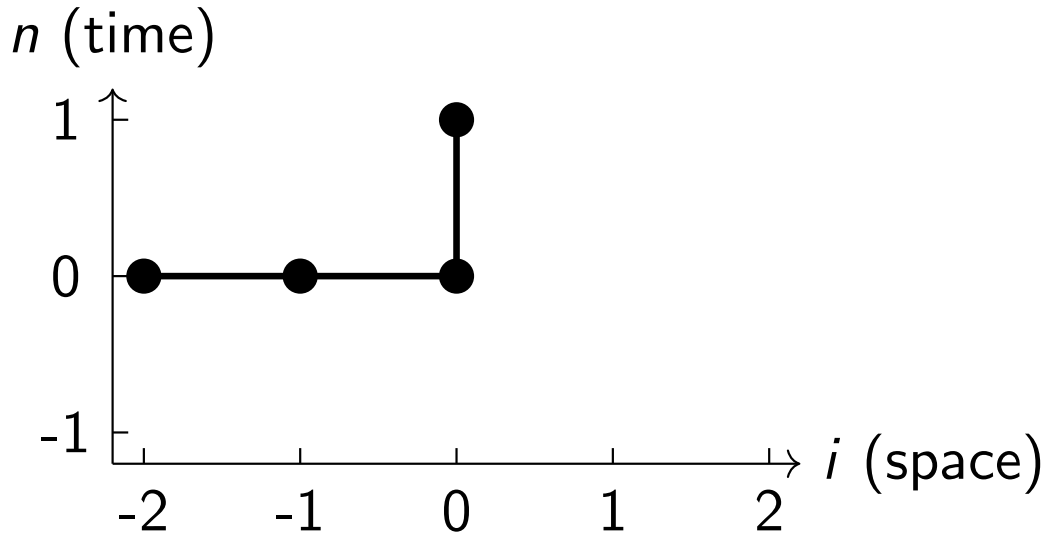
$$w_j^{n+1} = \frac{w_j^n + w_j^\star}{2} - \nu[f(w_j^\star) - f(w_{j-1}^\star)] \quad (9)$$

Beam-Warming scheme

The Beam-Warming scheme by [Beam, 1976] is also $O(\Delta x^2, \Delta t^2)$ with the flux

$$F^n_{j+1/2} = \frac{1}{2}[3f(w^n_j) - f(w^n_{j-1})] - \frac{A^2\nu}{2}[w^n_j - w^n_{j-1}] \qquad (10)$$

associated to the stencil



Fromm scheme

In the non-linear case, it is necessary to find a local evaluation of A.

It is clear that this scheme is decentered in the upwind direction.

It suffer from the same overshoot structures as the Lax-Wendroff scheme, but interestingly, these overshoots appear for the Beam-Warming scheme in the opposite direction.

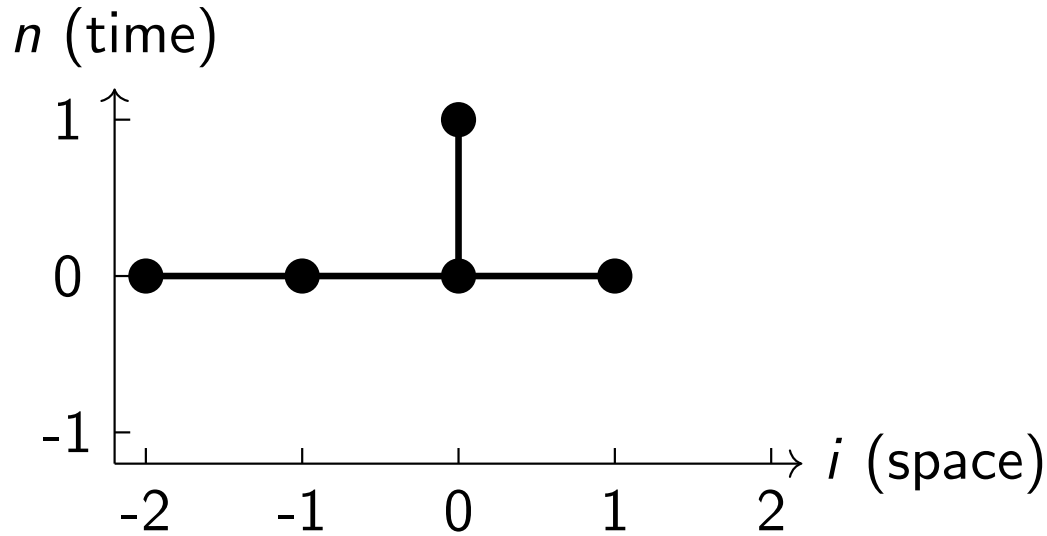
The Fromm scheme by [Fromm, 1968] is then an average of the Lax-Wendroff and Beam-Warming scheme : hence, it is also of order $O(\Delta x^2, \Delta t^2)$ with a flux given by

$$F^n_{j+1/2} = \frac{1}{2}[F^{\mathrm{Lax-Wendroff}}_{j+1/2} + F^{\mathrm{Beam-Warming}}_{j+1/2}] \qquad (11)$$

Fromm scheme

The results is very good for wave propagation, and also quite smooth for stiff gradients... so far the best scheme.

The stencil diagramm is



Approximate Riemann solvers

Solution of the Riemann problem is **challenging** in the vectorial case.

The exact solution of the Riemann problem is a 4 steps procedure :

- calculate the diagonal matrix $\mathbf{\Lambda}$ from $\mathbf{A}$, the eigen values λ_i and the transition matrix $\mathbf{Q}$
- calculate the components of $\mathbf{v}_L = \mathbf{Q}^{-1} \,.\, \mathbf{u}_L$ and $\mathbf{v}_R = \mathbf{Q}^{-1} \,.\, \mathbf{u}_R$
- compare x/t to the λ_i's and deduce the v_i (hence the $\Delta\mathbf{v}_i$'s) using the characteristics
- then get the solution $\Delta\mathbf{u}$ from the values of $\Delta\mathbf{v}$.

Notation : We note $\mathbf{u}_0 \equiv \mathbf{u}(x = 0, t) = u(x/t = 0)$.

Property : The solution $\mathbf{u}(x, t)$ of the Riemann problem is self-similar, meaning that it only depends on the ratio x/t.

Approximate Riemann solvers

Hence, $\mathbf{u}$ is constant along any lines passing through $(x = 0, t = 0)$.

The solution of $\mathbf{u}(x, t > 0)$ is then "connected" to $\mathbf{u}_0$ along the $x = 0$ line.

Solving the Riemann problem is then restricted to studying the evolution of $\mathbf{u}$ along $x = 0$. As a consequence, the positive λ_i's will play a different role as the negative ones.

It is clear that the exact solution of the Riemann problem is very expensive as these solutions request to calculate $\mathbf{\Lambda}$, $\mathbf{Q}^{-1}$ and some of their products.

Then, it, can hardly be achieved in a numerical code at the edges of each cells, and at each time steps.

For that reason, many approximate Riemann solvers have been proposed and few of these are reported below.

Why do we need $\mathbf{A}.\mathbf{u}_0$?

Notation :

$$\lambda_i^- = \min(0, \lambda_i) \qquad \lambda_i^+ = \max(0, \lambda_i) \tag{12}$$

The main challenge in solving a conservative equation is to calculate a numerical flux $\mathbf{F}_{i+1/2}$ as a numerical approximation of the physical flux. As the numerical flux is calculated at the interface between 2 adjacent cells, we need to perform some interpolation.

Remembering that

$$\mathbf{f}(\mathbf{u}_2) \sim \mathbf{f}(\mathbf{u}_1) + \partial_{\mathbf{u}}\mathbf{f}.(\mathbf{u}_2 - \mathbf{u}_1) \equiv \mathbf{f}(\mathbf{u}_1) + \mathbf{A}.(\mathbf{u}_2 - \mathbf{u}_1) \tag{13}$$

the previous remark makes clear that most of the work lies in finding a good approximation of $\mathbf{A}.\mathbf{u}_0$ (in the last term).

Why do we need $\mathbf{A}.\mathbf{u}_0$?

We can verify that along $x = 0$,

$$\mathbf{A} \,.\, \mathbf{u}_0 = \mathbf{A} \,.\, \mathbf{u}_L + \sum_{i=1}^{3} \mathbf{r}_i \lambda_i^- \Delta v_i = \mathbf{A} \,.\, \mathbf{u}_R - \sum_{i=1}^{3} \mathbf{r}_i \lambda_i^+ \Delta v_i \qquad (14)$$

so by averaging, we have

$$\mathbf{A} \,.\, \mathbf{u}_0 = \frac{1}{2} \mathbf{A} \,.\, (\mathbf{u}_R + \mathbf{u}_L) - \frac{1}{2} \sum_{i=1}^{3} \mathbf{r}_i |\lambda_i| \Delta v_i \qquad (15)$$

$\rightarrow$ it is clear that we hence need to calculate λ_i, $\mathbf{r}_i$ and $\Delta \mathbf{v}_i$.

Reformulation of $\mathbf{A}.\mathbf{u}_0$

We define $\mathbf{\Lambda}^+$ and $\mathbf{\Lambda}^-$ the diagonal matrix with diagonal elements λ_i^+ and λ_i^-, respectively and $|\mathbf{\Lambda}|$ the diagonal matrix with diagonal elements $|\lambda_i|$. We then have

$$\mathbf{\Lambda} = \mathbf{\Lambda}^+ + \mathbf{\Lambda}^- \qquad\qquad |\mathbf{\Lambda}| = \mathbf{\Lambda}^+ - \mathbf{\Lambda}^- \tag{16}$$

We straightforwardly define

$$|\mathbf{A}| = \mathbf{Q} \,.\, |\mathbf{\Lambda}| \,.\, \mathbf{Q}^{-1} \tag{17}$$

and $\mathbf{A}.\mathbf{u}_0$ then writes

$$\mathbf{A} \,.\, \mathbf{u}_0 = \frac{1}{2}\mathbf{A} \,.\, (\mathbf{u}_R + \mathbf{u}_L) - \frac{1}{2}|\mathbf{A}| \,.\, (\mathbf{u}_R - \mathbf{u}_L) \tag{18}$$

Approximate Riemann solvers

The first term is quite straightforward because $\mathbf{A}.\mathbf{u}_R = \mathbf{F}(w_R) \equiv \mathbf{F}_R$ and $\mathbf{A}.\mathbf{u}_L = \mathbf{F}(w_L) \equiv \mathbf{F}_L$.

The second one is less obvious because $\mathbf{A}$ being non-linear, that is local, one needs to define "where" to calculate $|\mathbf{A}|$.

The approximate Riemann solvers are then the methods to approximate $|A|$.

The interested reader could check the very nice and brief review of the Godunov-type methods by [Sweby, 1999].

Godunov schemes

In the special case of scalar linear advection equation, each cell-average values (separated by the discontinuities) are "simply" advected (Lagrangian stage).

$\rightarrow$ But in an Eulerian perspective, the cell averaged advection equation is needed.

Let's consider the one-dimensional scalar conservation la

$$\partial_t u + \partial_x f = 0 \tag{19}$$

Following the procedure in the course, one can demonstrate that the flux function at the origin writes

$$f(u(0,t)) = \begin{cases} \min\limits_{u_L \leq u \leq u_R} f(u) & \text{if } u_L < u_R \\ \max\limits_{u_L \geq u \geq u_R} f(u) & \text{if } u_L > u_R \end{cases} \tag{20}$$

Godunov schemes

Applying this formula for the flux function in the cell between j and $j+1$, one obtains the **Godunov's first-order upwind flux** by [Godunov, 1959]

$$\overline{F}^{\text{Godunov}}_{j+1/2}(w_j, w_{j+1}) = \begin{cases} \min\limits_{w_j^n \leq u \leq w_{j+1}^n} f(u) & \text{if } w_j < w_{j+1} \\ \max\limits_{w_j^n \geq u \geq w_{j+1}^n} f(u) & \text{if } w_j > w_{j+1} \end{cases} \tag{21}$$

Notation : The overline on the numerical flux F is intended to outline that such a flux can be read as an "average" using the two points given as parameters.

Godunov schemes

It is then clear that the trick is to calculate the min or max of the physical flux function in a given cell.

$\rightarrow$ This min/max is either at an endpoint of the interval for a monotonic physical flux function, or where the derivative of the flux function is null (sonic point). The Godunov numerical flux function is then in a simpler form

$$\overline{F}^{\text{Godunov}}_{j+1/2}(w_j, w_{j+1}) = \begin{cases} \min[f(w_j^n), f(w_{j+1}^n), f(u^\star)] & \text{if } w_j < w_{j+1} \\ \max[f(w_j^n), f(w_{j+1}^n), f(u^\star)] & \text{if } w_j > w_{j+1} \end{cases} \tag{22}$$

where $u^\star$ refers to any and all sonic points between w_j and w_{j+1}.

Remark : With any analytical form of the flux, the derivative $\partial_u f$ can be calculated, and for a gentle enough form of this flux, the $u^\star$ value at which this derivative is null can be analyticaly obtained.

Godunov schemes

Example : For the Burger equation, the physical flux writes $f(u) = \frac{1}{2}u^2$, so $u^\star = 0 = f(u^\star)$. Obviously, there is a $u^\star$ in a range $[w_j, w_{j+1}]$ if and only if $w_j.w_{j+1} < 0$.

It si obvious that in the linear case of a physical flux given by $f(u) = au$, the Godunov first-order upwind scheme gives the classical FTBS upwind scheme. Of course, some more complicated form of this flux can be used in the class of the Godunov schemes.

Definition : The **entropy condition** for a numerical scheme is equivalent to the second law of thermodynamics. This condition ensures that the entropy of a system numerically integrated through time is necessarily increasing or stay constant.

The Godunov first-order upwind scheme satisfies the entropy condition of the Euler's equations and preserves the positivity of the variables. But this scheme costs a lot as it needs to solve the Riemann problem at each interface and for each time steps.

Godunov theorem

Theorem : (Godunov's theorem). Linear numerical schemes for solving partial differential equations (PDEs), having the property of not generating new extrema (monotone scheme), can be at most first-order accurate.

• In the case of shocks or discontinuities, any scheme with an order of accuracy larger than one will fail close to these shocks or discontinuities.

• In the case of smooth regions where complex coupling of mode can be at play, a first-order accuracy scheme will have bad performances .

$\rightarrow$ This is for this reason that most of efforts have later been dedicated to find non-linear schemes as they are the only ones ensuring both monotonicity preserving property and high order.

The need of Flux hybridation

A proper way to manage the consequences of the Godunov theorem can also be to use a smart symbiosis of a first-order scheme with a larger order one.

For that purpose, "smart" means that we should find a way to weight the associated fluxes in a sort of average.

Roe-first order upwind method

The main idea of this scheme by [Roe, 1981] is to find a linear approximation of the Jacobian matrix $\mathbf{A}$ in the quasi-linear form of the advection equation.

For a scalar conservative equation, the Taylor expansions of the flux function around $\mathbf{u}_R$ and $\mathbf{u}_L$ are

$$\begin{aligned} \mathbf{f}(\mathbf{u}) &\sim \mathbf{A}(\mathbf{u}_L).(\mathbf{u}-\mathbf{u}_L)+\mathbf{f}(\mathbf{u}_L) && (23)\\ &\sim \mathbf{A}(\mathbf{u}_R).(\mathbf{u}-\mathbf{u}_R)+\mathbf{f}(\mathbf{u}_R) && (24)\end{aligned}$$

We hence would like to define a $\mathbf{A}_{RL}$ matrix such that

$$\mathbf{f}(\mathbf{u}_R)-\mathbf{f}(\mathbf{u}_L)=\mathbf{A}_{RL}.(\mathbf{u}_R-\mathbf{u}_L) \quad (25)$$

The solutions satisfying these N equations is not unique for the $N \times N$ matrix $\mathbf{A}_{RL}$ (with its N^2 free parameters).

Roe-first order upwind method

Roe proposed the simplified form

$$\mathbf{A}_{RL} = \mathbf{A}(\mathbf{u}_{RL}) \tag{26}$$

where the unknown $\mathbf{u}_{RL}$ lies between $\mathbf{u}_R$ and $\mathbf{u}_L$. The matrix $\mathbf{A}$ being known, solving Eq. (26) gives the $\mathbf{u}_{RL}$ components. For the one-dimensional Euler equations, the $\mathbf{A}_{RL}$ matrix is called the **Roe-average Jacobian matrix**.

Once the Roe-average conservative variables of the problem $\mathbf{u}_{RL}$ have been defined with Eq. (26), the Roe-average wave speed λ_i can be computed as well as the wave speed Δv_i.

Finally, the computation of the conservative variables $\mathbf{u}$ needs also to compute the transition matrix $\mathbf{Q}$. Remembering that $\mathbf{r}_i$ is the i^{th} column of $\mathbf{Q}$, the numerical flux of this method then writes

$$\mathbf{F}^n_{j+1/2} = \frac{1}{2}[\mathbf{f}(w^n_{j+1}) + \mathbf{f}(w^n_j)] - \frac{1}{2}\sum_{k=1}^{N} \mathbf{r}_k |\lambda_k| \Delta v_k \tag{27}$$

Upwind schemes

This scheme is written in a way to illustrate the idea of flux vector splitting. In a pedagogical perspective, we treat the linear scalar advection equation.

• In the case $A \in \mathbb{R}^+$, the numerical flux for a BS centering is simply

$$F^n_{j+1/2} = f(w^n_j) = Aw^n_j \tag{28}$$

and we already saw that it is of order $O(\Delta t, \Delta x)$, consistent and stable with the CFL condition $A\Delta t \leq \Delta x$.

• In the case $A \in \mathbb{R}^-$, such an upwind scheme which is also called **donor cell** has the numerical flux

$$F^n_{j+1/2} = f(w^n_{j+1}) = Aw^n_{j+1} \tag{29}$$

Upwind schemes

For a more general case of a velocity A which can have both signs during the time evolution of the flow, the upwind scheme should then write

$$\frac{w_j^{n+1} - w_j^n}{\Delta t} + \left(\frac{A + |A|}{2}\right) \frac{w_j^n - w_{j-1}^n}{\Delta x} + \left(\frac{A - |A|}{2}\right) \frac{w_{j+1}^n - w_j^n}{\Delta x} = 0 \tag{30}$$

meaning that with the two new unknowns

$$A^+ \equiv \frac{A + |A|}{2} \qquad A^- \equiv \frac{A - |A|}{2} \tag{31}$$

we can define two fluxes

$$F^+_{j+1/2} = A^+ w_j^n \qquad F^-_{j+1/2} = A^- w_{j+1} \tag{32}$$

Upwind schemes

The upwind scheme writes

$$\frac{w_j^{n+1} - w_j^n}{\Delta t} + \frac{F^+_{j+1/2} - F^+_{j-1/2}}{\Delta x} + \frac{F^-_{j+1/2} - F^-_{j-1/2}}{\Delta x} = 0 \qquad (33)$$

which means that we could use the conservative form given by Eq. (1) with the numerical flux defined as

$$F^n_{j+1/2} = F^+_{j+1/2} + F^-_{j+1/2} = A^+ w_j + A^- w_{j+1} \qquad (34)$$

Remark : The concept of "Flux vector splitting" now clearly appears ; the idea is to split the numerical flux in two parts, F^+ associated to positive flux, that is $\partial_u F^+ \geq 0$ and F^- associated to negative flux, that is $\partial_u F^- \leq 0$

Upwind schemes

Finally, in artificial viscosity form, this flux writes

$$F^n_{j+1/2} = \frac{1}{2}[f(w^n_j) + f(w^n_{j+1})] - \frac{|A|}{2}(w^n_{j+1} - w^n_j) \qquad (35)$$

This form can only be applied to scalar equations. In the more general case, the game is to find, depending on the kind of waves, the most important(s) one(s) and retain the absoluite value of their velocity for $|A|$.

Notation : The form of the flux given by Eq. (35) is clearly the same as the one given by Eq. (27) for a scalar equation, and will then be later called the Roe approximate Riemann solver.
The two forthcoming subsection deal with the cases where A should be replaced by a Jacobian matrix,

Rusanov scheme

One remember that the Godunov flux has a general form qiven by Eq. (21).

$\rightarrow$ Numerical flux being $F^n_{j+1/2} = \overline{F}^{\text{Godunov}}(w^n_j, w^n_{j+1})$:

- the solution of the RP is $w^\star_{i+1/2}(x/t) = \text{RP}[w^n_j, w^n_{j+1}]$
- the flux is $F^n_{j+1/2} = f[w^\star_{j+1/2}(0)] = F^\star(w^n_j, w^n_{j+1})$.

The Rusanov scheme is then a way to write this averaged numerical flux as

$$F^\star(w^n_j, w^n_{j+1}) = \frac{1}{2}[f(w^n_j) + f(w^n_{j+1})] - \frac{1}{2}A(w^n_j, w^n_{j+1})(w^n_{j+1} - w^n_j) \tag{36}$$

with de definition of $A(w^n_j, w^n_{j+1})$ as

$$A(w^n_j, w^n_{j+1}) = \max(|f'(w^n_j)|, |f'(w^n_{j+1})|) \tag{37}$$

HLL scheme

This scheme is part of the familly of the two-waves schemes. Thes two waves have a wave speed given by $S_L = \min_{w_L,w_R} \min_i \lambda_i(w)$ and $S_R = \max_{w_L,w_R} \max_i \lambda_i(w)$. These two terms can also write

$$S_L = \min(w_L, w_R) - \max(A_L, A_R) \quad S_R = \max(w_L, w_R) + \max(A_L, A_R) \tag{38}$$

Hence, the numerical flux is

$$\begin{aligned} F^\star(w_j^n, w_{j+1}^n) &= f(w_j^n) \text{ for } S_L > 0 \\ F^\star(w_j^n, w_{j+1}^n) &= \frac{S_R f(w_j^n) - S_L f(w_{j+1}^n) + S_L S_R (w_j^n w_{j+1}^n - w_j^n)}{S_R - S_L} \text{ for } S_L \\ F^\star(w_j^n, w_{j+1}^n) &= f(w_{j+1}^n) \text{ for } S_R < 0 \end{aligned}$$

Flux limiter schemes : the big picture

The Fromm scheme is defined as an average of the Lax-Wendroff and the Beam-Warming schemes.

$\rightarrow$ doing so, we tried to get the best from each scheme.
$\rightarrow$ but the average between these two fluxes is not case-sensitive

• Hence, might not always be the best linear combination one could hope.

The idea of flux limiter methods is to use a weighted average of two fluxes, one dedicated to correctly treat regular regions and another one dedicated to shocks and/or discontinuities.
$\rightarrow$ The associated weights in this average for a new time step should then adjust, depending on the gradients of the numerical solution obtained at the current time step.
the flux could be defined as

$$F^n_{j+1/2} = F^{(1)}_{j+1/2} + \phi^n_{j+1/2}(F^{(2)}_{j+1/2} - F^{(1)}_{j+1/2}) \qquad (43)$$

Flux limiter schemes : the big picture

$F^{(1)}_{j+1/2}$ and $F^{(2)}_{j+1/2}$ are the conservative numerical fluxes of two different methods.

Definition : In Eq. (61), the parameter $\phi^n_{j+1/2}$ controlling the weight of the linear combination is called a **flux limiter**.

Remark : In Eq. (61), the numerical flux is then given by the first term $F^{(1)}_{j+1/2}$, but then "limited" by the second term.

In order to save computation, the spatial index of the flux limiter is oftenly bump, up or down, by one half.

or $a > 0$,

$$F^n_{j+1/2} = F^{(1)}_{j+1/2} + \phi^n_j (F^{(2)}_{j+1/2} - F^{(1)}_{j+1/2}) \tag{44}$$

for $a < 0$,

$$F^n_{j+1/2} = F^{(1)}_{j+1/2} + \phi^n_{j+1} (F^{(2)}_{j+1/2} - F^{(1)}_{j+1/2}) \tag{45}$$

Flux limiter schemes : the big picture

The whole game is to find an analytical expression for the ϕ function, as well as the appropriate unknown on which it applies.

$\rightarrow$ ϕ should be a function depending on the numerical solution w_j^n and more precisely on its derivative, approximated by finite differences.

As a general idea, we need to evaluate how close we are from a shock or from a wave. This means that we should certainly calculate a 1-grid cell difference, but also evaluate how it change from one cell to the directly adjacent one.

$\rightarrow$ we introduce r_j^+ and r_j^-, defined as

$$r_j^+ = \frac{w_j^n - w_{j-1}^n}{w_{j+1}^n - w_j^n} \qquad r_j^- = \frac{w_{j+1}^n - w_j^n}{w_j^n - w_{j-1}^n} \tag{46}$$

Remark : It is clear that $r_j^+ = 1/r_j^-$.

Flux limiter schemes : the big picture

An alternative approach is to consider a shock indicator based on the numerical flux difference $F^{(2)}_{j+1/2} - F^{(1)}_{j+1/2}$.

- In smooth regions, these numerical fluxes should be almost equal.
- In shock regions, such a difference should then be "large".

Hence, the r parameters defined below could also be defined as

$$r_j^+ = \frac{F^{(2)}_{j-1/2} - F^{(1)}_{j-1/2}}{F^{(2)}_{j+1/2} - F^{(1)}_{j+1/2}} \qquad r_j^- = \frac{F^{(2)}_{j+1/2} - F^{(1)}_{j+1/2}}{F^{(2)}_{j-1/2} - F^{(1)}_{j-1/2}} \tag{47}$$

Remember the way numerical fluxes writes in term of artificial viscosity,

$$F^{(1)}_{j+1/2} = \frac{1}{2}[f(w^n_{j+1}) + f(w^n_j)] - \frac{1}{2}\epsilon^{(1)}_{j+1/2}(w^n_{j+1} - w^n_j) \tag{48}$$

$$F^{(2)}_{j+1/2} = \frac{1}{2}[f(w^n_{j+1}) + f(w^n_j)] - \frac{1}{2}\epsilon^{(2)}_{j+1/2}(w^n_{j+1} - w^n_j) \tag{49}$$

$$\tag{50}$$

Flux limiter schemes : the big picture

The flux difference writes

$$F^{(2)}_{j+1/2} - F^{(1)}_{j+1/2} = \frac{1}{2}[\epsilon^{(2)}_{j+1/2} - \epsilon^{(1)}_{j+1/2}](w^n_{j+1} - w^n_j) \qquad (51)$$

With such expression, we can write the ratios of flux difference, and study their properties. It appears that for $\epsilon^{(1)}_{j+1/2} > \epsilon^{(2)}_{j+1/2}$

$r^{\pm}_j \geq 0$ if the w^n_j's are monotonically increasing or decreasing,
$r^{\pm}_j \leq 0$ if the w^n_j's have a local maximum or a minimum.

- There is a link between the sign of $r^{\pm}_j$ and sonic point.

$\rightarrow$ we investigate various form of the $\phi(r)$ function and will try to emphasize the constraints on this function.

Van Leer flux limiter

The flux-limited method by [Vanleer1974] for the linear advection equation with $a > 0$ is

$$w_j^{n+1} = w_j^n - \nu[F_{j+1/2}^n - F_{j-1/2}^n] \tag{52}$$

with the flux

$$F_{j+1/2}^n = \frac{1+\eta_j^n}{2} F_{j+1/2}^{\text{Lax-Wendroff}} + \frac{1-\eta_j^n}{2} F_{j+1/2}^{\text{Beam-Warming}} \tag{53}$$

The new unknown η_j^n s defined as

$$\eta_j^n = \frac{|r_j^+| - 1}{|r_j^+| + 1} \tag{54}$$

and

$$r_j^+ = \frac{w_j^n - w_{j-1}^n}{w_{j+1}^n - w_j^n} \tag{55}$$

Sweby flux limiter

The Roe-first order upwind scheme and the Lax-Wendroff scheme have complementary properties :

$\rightarrow$ the first one is doing well near jump discontinuities
$\rightarrow$ the last one does well in smooth region.

For the Sweby flux-limited scheme by [Sweby, 1984],

$$F^n_{j+1/2} = F^{\text{Roe}}_{j+1/2} + \phi^n_j [F^{\text{Lax-Wendroff}}_{j+1/2} - F^{\text{Roe}}_{j+1/2}] \qquad (56)$$

There exist conditions on the Sweby's flux-limited function, as the one that satisfies the proper "upwinding" of the scheme, depending on the sign of A.

Notation : In the following developments, ϕ^n_j has to be read as the function $\phi(r^+_j)$ on the unknown r^n_j given by Eq. (55)

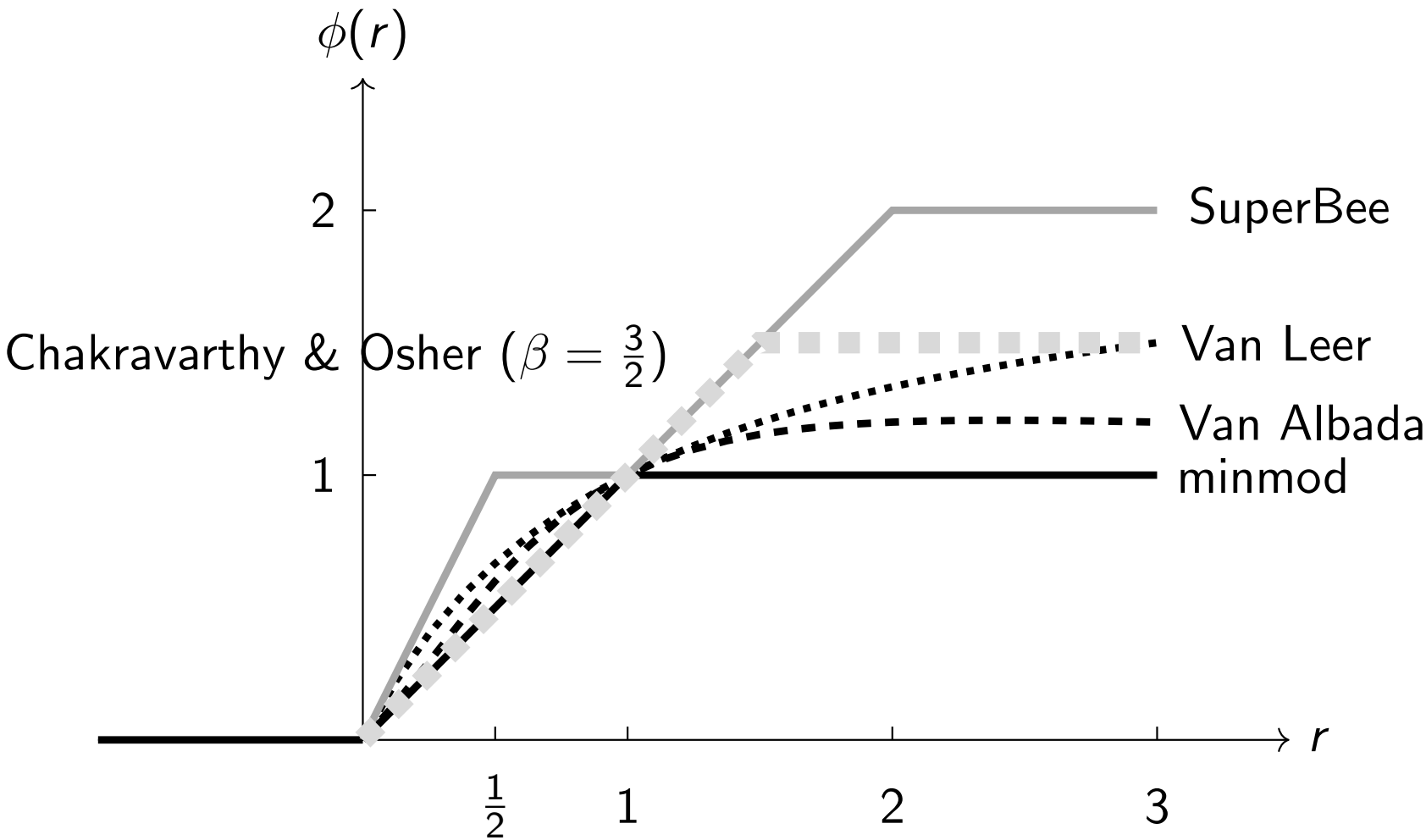
Various flux limiters

There exist several flux limiters among which,

- **minmod** : $\phi(r) = \max[0, \min(1, r)]$
- **Chakravarthy & Osher** : $\phi(r) = \max[0, \min(\beta, r)]$ with $1 \leq \beta \leq 2$
- **Van Leer** : $\phi(r) = \frac{r+|r|}{1+r}$
- **Van Albada** : $\phi(r) = \max\left[0, \frac{r+r^2}{1+r^2}\right]$
- **SuperBee** : $\phi(r) = \max[0, \min(1, 2r), \min(2, r)]$

These flux limiters are displayed below. It can be shown that all these flux limiters should lie between the minmod and the Superbee limiters.

comparison between flux limiters



Limits & properties of these flux limiters

Property : The minmod limiter is the most robust flux limiter, but then also the most dissipative.

Property : The Superbee limiter is the less robust flux limiter, but allows the stieffest shock fronts.

Remark : Obviously, $\phi(r) = 0$ gives the Roe-first order upwind scheme and $\phi(r) = 1$ gives the Lax-Wendroff scheme.

TVD property & ENO

[Harten, 1983] introduced the concept of **Total Variation Diminishing**.

$\rightarrow$ Very important in order to prevent the growth of any small-scale oscillations, generally close to stiff gradients.

For any function $u(x,t)$ of class C^1, its total variation is defined as

$$\mathrm{TV}[u,\tau] = \int_{\mathbb{R}} |\partial_x u(x,\tau)| \, \mathrm{d}x \tag{57}$$

Definition : The Total Variation Diminishing (**TVD**) property means that for a given function $u(x,t)$, $\mathrm{TV}[u,\tau]$ is decreasing with time τ.

Hence, if $\mathrm{TV}[u,t_2] \leq \mathrm{TV}[u,t_1]$ for $t_2 > t_1$, we say that $u(x,t)$ verifies the TVD property.

TVD property & ENO

The TVD property is then a way to prevent the birth of spurious oscillation, meaning that any solution verifying the TVD property should behave smoothly.

The total variation can also be defined in the same way for a discrete serie ; for the approximate solution w_j^n the total variation writes

$$\mathrm{TV}[w^n] = \sum_{j=0}^{N-1} |w_{j+1}^n - w_j^n| \tag{58}$$

so a discret solution will verify the TVD property if

$$\mathrm{TV}[w^{n+1}] \leq \mathrm{TV}[w^n] \tag{59}$$

Property : A numerical scheme is TVD if the numerical solution verifies the TVD property given by Eq. (59)

TVD property & ENO

As a consequence, any local enhancement of a gradient of the solution w_j^n during its time evolution will be balanced by a (larger) decrease of such a gradient in a different location.

For any TVD method, spurious oscillations close to discontinuities can neither birth nor grow.

Property : Any flux limited method verifies the TVD property.

Definition : The acronym **ENO** means **Essentially Non Oscillatory**. A ENO scheme is a scheme which guarantees that no spurious oscillation will birth and grow close to stiff gradients of the solution w_j^n.

Remark : While their definitions are different, the TVD and ENO properties are not that far.

Remark : The ENO (and WENO) schemes are based on a "reconstruction-evolution" method.

Flux corrected schemes

The idea behind this class of method is not that far from the flux limited methods.

In conservative form, the numerical solution depends on the numerical fluxes

$$w_j^{n+1} = w_j^n - \nu[F_{j+1/2}^n - F_{j-1/2}^n] \tag{60}$$

but then, the numerical flux is (generally) defined as

$$F_{j+1/2}^n = F_{j+1/2}^{(1)} + F_{j+1/2}^{(C)} \tag{61}$$

in which $F_{j+1/2}^{(1)}$ is a flux that is "corrected" by $F_{j+1/2}^{(C)}$.

This corrected flux is defined, in several (but not all) cases as

$$F_{j+1/2}^{(C)} = \mathrm{d}_{j+1/2}^n(F_{j+1/2}^{(1)}, F_{j+1/2}^{(2)}) \tag{62}$$

where $\mathrm{d}_{j+1/2}^n$ is a function depending on the two fluxes $F_{j+1/2}^{(1)}$ and $F_{j+1/2}^{(2)}$.

Flux corrected schemes

Up to now, this class of method really looks like a reformulation of flux limited methods...

Definition : In flux limited methods, $\phi^n_{j+1/2}(r)$ generally depends on **ratios** of solutions or fluxes.

Definition : In flux corrected methods $\mathrm{d}^n_{j+1/2}(f^1, f^2)$ generally depends on **differences** between solutions or fluxes.

As a consequence, the function $\phi^n_{j+1/2}(r)$ in flux limiter depends on a single parameter, while $\mathrm{d}^n_{j+1/2}(F^1, F^2)$ depends on two (or eventually more) parameters.

The "Flux Corrected Transport" (FCT) method

This scheme proposed by [Boris, 1973] is a blend between a first order upwind method and the Lax-Wendroff method.

We firstly introduce the **modified Boris-Book first-order upwind method** which flux si given by

$$\nu F_{j+1/2}^{\mathrm{Boris-Book}} = \nu F_{j+1/2}^{\mathrm{Lax-Wendroff}} - \frac{1}{8}(w_{j+1}^n - w_j^n) \qquad (63)$$

so that, by developping the Lax-Wendroff flux, this flux can be written in artificial viscosity form as

$$\nu F_{j+1/2}^{\mathrm{Boris-Book}} = \frac{1}{2}\nu[f(w_{j+1}^n) - f(w_j^n)] - \frac{1}{2}\left[\left(\nu A_{j+1/2}^n\right)^2 + \frac{1}{4}\right](w_{j+1}^n - w_j^n) \qquad (64)$$

The "Flux Corrected Transport" (FCT) method

We can apply the flux limited method to $F_{j+1/2}^{\text{Boris}-\text{Book}}$ and use its difference with $F_{j+1/2}^{\text{Lax}-\text{Wendroff}}$ in the limitor

$$F_{j+1/2}^{n} = F_{j+1/2}^{\text{Boris}-\text{Book}} + \phi_{j+1/2}^{n}(F_{j+1/2}^{\text{Lax}-\text{Wendroff}} - F_{j+1/2}^{\text{Boris}-\text{Book}}) \quad (65)$$

and then work on the $\phi_{j+1/2}^{n}$ function.

Considering that shocks are the cause of spurious oscillations and extremas, we then would like to have $F_{j+1/2}^{n} \sim F_{j+1/2}^{\text{Boris}-\text{Book}}$ near extrema, and $F_{j+1/2}^{n} \sim F_{j+1/2}^{\text{Lax}-\text{Wendroff}}$ elsewhere.

In term of corrective flux,

$$F_{j+1/2}^{(C)} = \begin{cases} 0 & \text{near extrema} \\ F_{j+1/2}^{\text{Lax}-\text{Wendroff}} - F_{j+1/2}^{\text{Boris}-\text{Book}} & \text{elsewhere} \end{cases} \quad (66)$$

The "Flux Corrected Transport" (FCT) method

This flux can write in term of flux limiters

$$\phi_{j+1/2}^n = \begin{cases} 0 & \text{near extrema} \\ 1 & \text{elsewhere} \end{cases} \tag{67}$$

Boris & Book also impose conditions on $F_{j+1/2}^n$, to stay between $F_{j+1/2}^{\text{Boris}-\text{Book}}$ and $F_{j+1/2}^{\text{Lax}-\text{Wendroff}}$.

$\rightarrow$ this is equivalent to $0 \leq \phi_{j+1/2}^n \leq 1$.

- A simple way to satisfy these 2 conditions is to take

$$\nu F_{j+1/2}^{(C)} = \text{minmod}\left[w_j^n - w_{j-1}^n, \frac{1}{8}(w_{j+1}^n - w_j^n), w_{j+2}^n - w_{j+1}^n \right] \tag{68}$$

The "Flux Corrected Transport" (FCT) method

This flux limiter also writes

$$\phi_{j+1/2}^n = \text{minmod}\left(8r_j^+, 1, \frac{8}{r_j^+}\right) \tag{69}$$

with

$$r_j^+ = \frac{w_j^n - w_{j-1}^n}{w_{j+1}^n - w_j^n} \tag{70}$$

Definition : The minmod (minimum modulus) function equals the argument with the least absolute value if all the arguments have the same sign, and equal zero otherwise.

This is the **One-step Boris-Book flux-corrected method**.

The two-steps FCT method

The most useful method is a two-setps variant of this one, that is a predictor

$$w_j^\star = w_j^n - \eta[F_{j+1/2}^{\text{Lax}-\text{Wendroff}} - F_{j+1/2}^{\text{Boris}-\text{Book}}] \tag{71}$$

followed by a corrector

$$w_j^{n+1} = w_j^\star - \eta[F_{j+1/2}^{(C)} - F_{j-1/2}^{(C)}] \tag{72}$$

with

$$\eta F_{j+1/2}^{(C)} = \text{minmod}(w_j^\star - w_{j-1}^\star, \frac{1}{8}[w_{j+1}^\star - w_j^\star], w_{j+2}^\star - w_{j+1}^\star) \tag{73}$$

This scheme is pretty good, both at shocks/discontinuities and regular solutions, provided that the CFL number is not too large (that is for low Mach number).