Finite volume method for conservation laws I

Praveen. C
praveen@math.tifrbng.res.in

Tata Institute of Fundamental Research
Center for Applicable Mathematics
Bangalore 560065
http://math.tifrbng.res.in

February 11, 2013

Scalar conservation law

Given a smooth flux function $f : \mathbb{R} \to \mathbb{R}$, $f \in C^2(\mathbb{R})$ and an initial condition $u_0 \in L^\infty(\mathbb{R})$

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) = 0, \quad x \in \mathbb{R}, \quad t > 0 \tag{1}$$

$$u(x, 0) = u_0(x), \qquad x \in \mathbb{R} \tag{2}$$

We also define

$$a(u) = f'(u)$$

which is the slope of the characteristics.

Finite volume method

Divide space domain into finite volumes

$$\Omega = \bigcup_j (x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}), \quad h_j = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}, \quad x_j = \frac{1}{2}(x_{j-\frac{1}{2}} + x_{j+\frac{1}{2}})$$

Integrate conservation law over finite volume $(x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}})$ and time slab (t^n, t^{n+1})

$$\int_{t^n}^{t^{n+1}} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \left(\frac{\partial u}{\partial t} + \frac{\partial f}{\partial x} \right) \mathrm{d}x\mathrm{d}t = 0$$

Cell average value

$$u_j(t) = \frac{1}{h_j} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x,t)\mathrm{d}x$$

gives conservation law (exact)

$$(u_j^{n+1} - u_j^n)h_j + \int_{t^n}^{t^{n+1}} [f(x_{j+\frac{1}{2}}, t) - f(x_{j-\frac{1}{2}}, t)]\mathrm{d}t = 0$$

Finite volume method

Approximate time integral of flux using solution at t^n (explicit scheme)

$$\int_{t^n}^{t^{n+1}} f(x_{j+\frac{1}{2}}, t)\mathrm{d}t \approx f(x_{j+\frac{1}{2}}, t^n)\Delta t$$

leads to finite volume method

$$\frac{v_j^{n+1} - v_j^n}{\Delta t} + \frac{f_{j+\frac{1}{2}}^n - f_{j-\frac{1}{2}}^n}{h_j} = 0$$

Cell average values are the unknowns in the finite volume method.

$$v_j^n \approx u_j(t^n) = \frac{1}{h_j}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, t^n)\mathrm{d}x$$

We do not deal with point values. However for smooth solutions

$$u_j(t) - u(x_j, t) = \mathcal{O}\left(h_j^2\right)$$

Finite volume method

The finite volume solution is made of piecewise constant values

$$v(x,t) = v_j^n, \quad x \in (x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}), \quad t \in [t^n, t^{n+1})$$

At $x = x_{j+\frac{1}{2}}$ there are two values of v; it is not obvious how to approximate the flux $f(x_{j+\frac{1}{2}}, t^n)$. The simplest choice is

$$f(x_{j+\frac{1}{2}}, t^n) \approx \frac{1}{2}[f(v_j^n) + f(v_{j+1}^n)] \qquad \text{or} \qquad f(x_{j+\frac{1}{2}}, t^n) \approx f\left(\frac{v_j^n + v_{j+1}^n}{2}\right)$$

But this leads to a central difference type scheme, which does not respect the wave propagation property present in the hyperbolic problem. We will see how to construct good approximations to the flux in the form

$$f(x_{j+\frac{1}{2}}, t^n) \approx g(\ldots, v_j^n, v_{j+1}^n, \ldots) =: g_{j+\frac{1}{2}}^n$$

where g is called the numerical flux function. In the simplest case

$$g_{j+\frac{1}{2}} = g(v_j, v_{j+1}) \qquad \text{which leads to 3-point scheme}$$

Finite volume method

The finite volume method takes the form

$$\frac{v_j^{n+1} - v_j^n}{\Delta t} + \frac{g_{j+\frac{1}{2}}^n - g_{j-\frac{1}{2}}^n}{h_j} = 0$$

More accurate flux integral (Trapezoidal rule)

$$\int_{t^n}^{t^{n+1}} f(x_{j+\frac{1}{2}}, t)\mathrm{d}t \approx \frac{1}{2}[g_{j+\frac{1}{2}}^n + g_{j+\frac{1}{2}}^{n+1}]\Delta t$$

or, a mid-point integration

$$\int_{t^n}^{t^{n+1}} f(x_{j+\frac{1}{2}}, t)\mathrm{d}t \approx g(\ldots, v_j^{n+\frac{1}{2}}, v_{j+1}^{n+\frac{1}{2}}, \ldots)\Delta t, \quad v_j^{n+\frac{1}{2}} = \frac{1}{2}(v_j^n + v_j^{n+1})$$

These approximations lead to an implicit scheme which is second order accurate in time.

Remark: There are two approximations involved; time integral through some quadrature and the numerical flux function.

Method of lines approach

Integrate conservation law over one finite volume

$$\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \left(\frac{\partial u}{\partial t} + \frac{\partial f}{\partial x} \right) \mathrm{d}x = 0$$

Semi-discrete conservation law (exact)

$$h_j \frac{\mathrm{d}u_j}{\mathrm{d}t} + f(x_{j+\frac{1}{2}}, t) - f(x_{j-\frac{1}{2}}, t) = 0$$

Approximate flux with numerical flux function

$$h_j \frac{\mathrm{d}v_j}{\mathrm{d}t} + g_{j+\frac{1}{2}}(t) - g_{j-\frac{1}{2}}(t) = 0$$

$$g_{j+\frac{1}{2}}(t) = g(\ldots, v_j(t), v_{j+1}(t), \ldots)$$

System of ODE for the cell averages $(v_j(t))_j$; integrate in time using some ODE scheme like Runge-Kutta scheme. Explicit or implicit, high order accurate schemes can be constructed by this approach.

Difference scheme

Some definitions:

$$\begin{aligned} \Delta t &= \text{time step} \\ \Delta &= \text{uniform spatial grid of size } \Delta x \\ \lambda &= \frac{\Delta t}{\Delta x} \end{aligned}$$

$(2k+1)$-point explicit difference scheme

$$v_j^{n+1} = H(v_{j-k}^n, \ldots, v_j^n, \ldots, v_{j+k}^n), \quad \forall n \geq 0, \quad j \in \mathbb{Z} \tag{3}$$

$H : \mathbb{R}^{2k+1} \to \mathbb{R}$ is a continuous function and

$$v_j^n \approx u_j^n = u(x_j, t^n), \quad x_j = j\Delta x, \quad t_n = n\Delta t$$

Define the operator H_Δ which maps a sequence $v = (v_j)_{j\in\mathbb{Z}}$ into the sequence $H_\Delta(v) = (H_\Delta(v)_j)_{j\in\mathbb{Z}}$

$$H_\Delta(v)_j = H(v_{j-k}, \ldots, v_j, \ldots, v_{j+k})$$

Then the difference scheme is

$$v^{n+1} = H_\Delta(v^n)$$

Definition

The difference scheme (3) can be put in the conservation form if there exists a continuous function $g : \mathbb{R}^{2k} \to \mathbb{R}$ such that

$$H(v_{-k}, \ldots, v_0, \ldots, v_k) = v_0 - \lambda[g(v_{-k+1}, \ldots, v_k) - g(v_{-k}, \ldots, v_{k-1})] \tag{4}$$

The function g is called the *numerical flux* function.

Hence (3) can be written as

$$v_j^{n+1} = v_j^n - \lambda[g(v_{j-k+1}^n, \ldots, v_{j+k}^n) - g(v_{j-k}^n, \ldots, v_{j+k-1}^n)] \tag{5}$$

If we set

$$g_{j+\frac{1}{2}}^n = g(v_{j-k+1}^n, \ldots, v_{j+k}^n)$$

then we get the difference scheme as

$$v_j^{n+1} = v_j^n - \lambda[g_{j+\frac{1}{2}}^n - g_{j-\frac{1}{2}}^n] \tag{6}$$

Proposition 1.1 (GR1)

The difference scheme (3) can be put in conservation form if and only if we have for any sequence $v = (v_j)_{j\in\mathbb{Z}} \in L^1(\mathbb{Z})$ such that $H_\Delta(v) \in L^1(\mathbb{Z})$

$$\sum_{j\in\mathbb{Z}} H(v_{j-k}, \ldots, v_{j+k}) = \sum_{j\in\mathbb{Z}} v_j \tag{7}$$

Remark: If the difference scheme (3) can be put in conservation form, the discrete solution operator H_Δ possesses the analogous property of the continuous solution operator

$$u_0 \in L^1(\mathbb{R}) \quad \Longrightarrow \quad \int_{\mathbb{R}} u(x,t)\mathrm{d}x = \int_{\mathbb{R}} u_0(x)\mathrm{d}x$$

The finite volume method satisfies conservation in **each** finite volume. It also satisfies a global conservation

$$\sum_j \frac{\mathrm{d}v_j}{\mathrm{d}t}\Delta x + \sum_{j=1}^{j=N} [g_{j+\frac{1}{2}}(t) - g_{j-\frac{1}{2}}(t)] = 0$$

Due to cancellation of fluxes, this gives the global balance equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_j v_j \Delta x = g_{\frac{1}{2}}(t) - g_{N+\frac{1}{2}}(t)$$

This is a discrete analogue of the conservation property of the PDE

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b u(x,t)\mathrm{d}x = f(u(a,t)) - f(u(b,t))$$

Remark: Continuous finite elements do not satisfy local conservation property, though they satisfy the global balance. Discontinuous finite elements, which are similar to finite volume method, do satisfy local conservation. This local conservation is important for correct shock capturing.

FDM for linear equation

$$u_t + a u_x = 0$$

- Upwind scheme

$$\frac{v_j^{n+1} - v_j^n}{\Delta t} + a^+ \frac{v_j^n - v_{j-1}^n}{\Delta x} + a^- \frac{v_{j+1}^n - v_j^n}{\Delta x} = 0$$

Numerical flux

$$g_{j+\frac{1}{2}} = \frac{1}{2} a (v_j + v_{j+1}) - \frac{1}{2} |a| (v_{j+1} - v_j)$$

- Lax-Friedrichs

$$\frac{v_j^{n+1} - \frac{v_{j-1}^n + v_{j+1}^n}{2}}{\Delta t} + a \frac{v_{j+1}^n - v_{j-1}^n}{2\Delta x} = 0$$

Numerical flux

$$g_{j+\frac{1}{2}} = \frac{1}{2} a (v_j + v_{j+1}) - \frac{1}{2} \frac{\Delta x}{\Delta t} (v_{j+1} - v_j) = \frac{1}{2} a (v_j + v_{j+1}) - \frac{1}{2\lambda} (v_{j+1} - v_j)$$

FDM for linear equation

- Lax-Wendroff

$$v_j^{n+1} = v_j^n - \frac{1}{2}a\lambda(v_{j+1}^n - v_{j-1}^n) + \frac{1}{2}a^2\lambda^2(v_{j-1}^n - 2v_j^n + v_{j+1}^n)$$

  Numerical flux

$$g_{j+\frac{1}{2}} = \frac{1}{2}a(v_j + v_{j+1}) - \frac{1}{2}\lambda a^2(v_{j+1} - v_j)$$

All of these numerical fluxes have same structure

$$g_{j+\frac{1}{2}} = \frac{1}{2}a(v_j + v_{j+1}) - \frac{1}{2}q(v_{j+1} - v_j)$$

Non-linear conservation law

$$u_t + f_x = 0$$

- Lax-Friedrichs

$$\frac{v_j^{n+1} - \frac{v_{j-1}^n + v_{j+1}^n}{2}}{\Delta t} + \frac{f_{j+1}^n - f_{j-1}^n}{2\Delta x} = 0$$

Numerical flux

$$g_{j+\frac{1}{2}} = \frac{1}{2}(f_j + f_{j+1}) - \frac{1}{2}\frac{\Delta x}{\Delta t}(v_{j+1} - v_j) = \frac{1}{2}(f_j + f_{j+1}) - \frac{1}{2\lambda}(v_{j+1} - v_j)$$

- Lax-Wendroff

$$v_j^{n+1} = v_j^n - \frac{1}{2}\lambda(f_{j+1}^n - f_{j-1}^n) + \frac{1}{2}\lambda^2[a_{j+\frac{1}{2}}(f_{j+1}^n - f_j^n) - a_{j-\frac{1}{2}}(f_j^n - f_{j-1}^n)]$$

Numerical flux

$$g_{j+\frac{1}{2}} = \frac{1}{2}(f_j + f_{j+1}) - \frac{1}{2}\lambda a_{j+\frac{1}{2}}(f_{j+1} - f_j), \qquad a_{j+\frac{1}{2}} = f'\left(\frac{v_j + v_{j+1}}{2}\right)$$

Non-linear conservation law

- Ritchmeyer two-step Lax-Wendroff method

$$\begin{aligned} v_{j+\frac{1}{2}}^{n+\frac{1}{2}} &= \frac{1}{2}(v_j^n + v_{j+1}^n) - \frac{\lambda}{2}[f(v_{j+1}^n) - f(v_j^n)] \\ v_j^{n+1} &= v_j^n - \lambda[f(v_{j+\frac{1}{2}}^{n+\frac{1}{2}}) - f(v_{j-\frac{1}{2}}^{n+\frac{1}{2}})] \end{aligned}$$

  Avoids computation of flux Jacobian.

- MacCormack method

$$\begin{aligned} v_j^* &= v_j^n - \lambda[f(v_{j+1}^n) - f(v_j^n)] \\ v_j^{n+1} &= \frac{1}{2}(v_j^n + v_j^*) - \frac{\lambda}{2}[f(v_j^*) - f(v_{j-1}^*)] \end{aligned}$$

  Avoids computation of flux Jacobian.

- Upwind scheme for $u_t + uu_x = 0$: naive generalization

$$\frac{v_j^{n+1} - v_j^n}{\Delta t} + (v_j^n)^+ \frac{v_j^n - v_{j-1}^n}{\Delta x} + (v_j^n)^- \frac{v_{j+1}^n - v_j^n}{\Delta x} = 0$$

  Cannot be written as a finite volume scheme !!!

Example (Non-conservative scheme for Burgers equation)
Consider initial condition

$$u(x,0) = \begin{cases} 1 & x < 0 \\ 0 & x > 0 \end{cases}$$

Then naive upwind scheme

$$\frac{v_j^{n+1} - v_j^n}{\Delta t} + v_j^n \frac{v_j^n - v_{j-1}^n}{\Delta x} = 0$$

gives the solution

$$v_j^0 = \begin{cases} 1 & j \le 0 \\ 0 & j > 0 \end{cases} \qquad \Longrightarrow \qquad v_j^n = \begin{cases} 1 & j \le 0 \\ 0 & j > 0 \end{cases}$$

which is wrong solution. The correct solution has a shock moving with speed $s = \frac{1}{2}$.

Example: (Non-conservative scheme for Burgers equation)
Consider initial condition

$$u(x,0) = \begin{cases} 1.2 & x < 0 \\ 0.4 & x > 0 \end{cases}$$

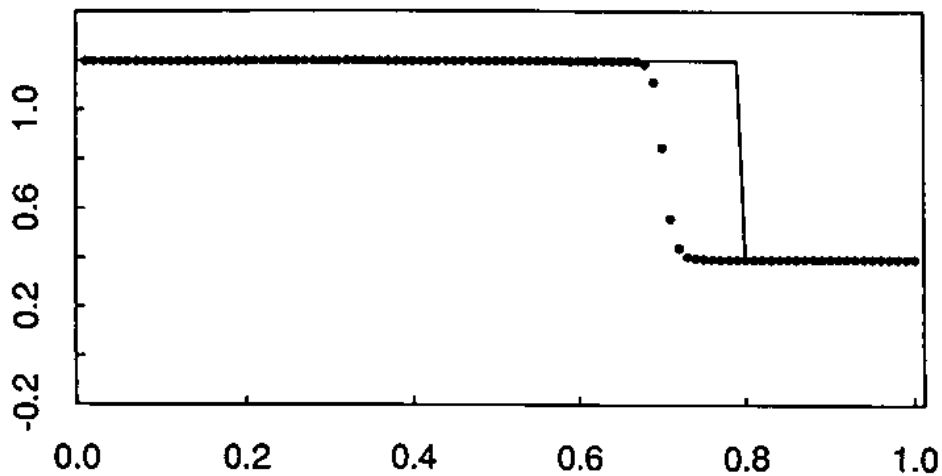


Figure 12.1. True and computed solutions to Burgers' equation using a nonconservative method.

Shock location is wrong !!!

The first basic property that the finite volume scheme is required to satisfy is consistency with the conservation law. Observe that

$$\frac{v_j^{n+1} - v_j^n}{\Delta t} \approx \frac{\partial v}{\partial t}(x_j, t_n)$$

Thus we require that

$$\frac{1}{\Delta x}[g_{j+\frac{1}{2}}^n - g_{j-\frac{1}{2}}^n] \approx \frac{\partial f}{\partial x}(x_j, t_n)$$

Definition

The difference scheme (3) is said to be consistent with equation (1) if

$$g(v, \ldots, v) = f(v), \qquad \forall v \in \mathbb{R} \tag{8}$$

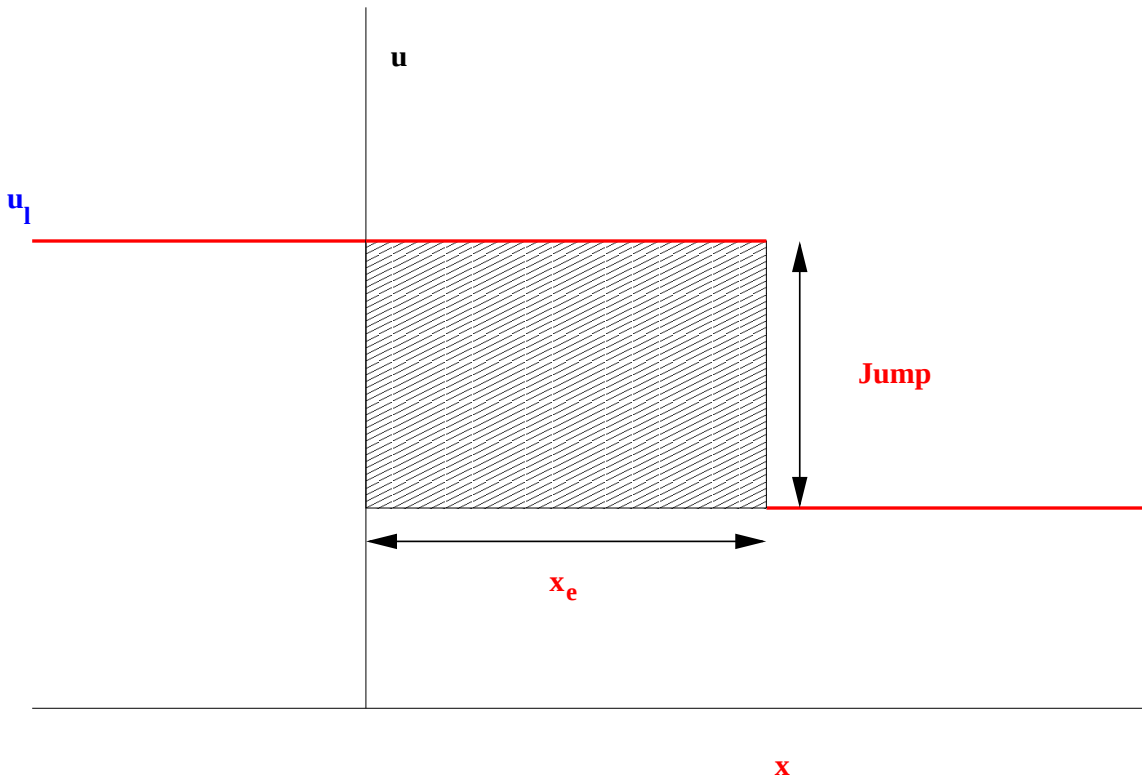
upto an additive constant. Such a numerical flux function g is said to be *consistent*.

Consistency + conservation = correct shocks

$$v_j^{n+1} = v_j^n - \lambda(g^n_{j+\frac{1}{2}} - g^n_{j-\frac{1}{2}}), \qquad g(u,\ldots,u) = f(u)$$

Let us consider a step function as an initial condition

$$u_0(x) = \begin{cases} u_l, & x \le 0 \\ u_r, & x > 0 \end{cases}$$



Consistency + conservation = correct shocks

After time t the exact shock location is $x_e(t)$, which is given by

$$x_e(t) = \frac{\text{shaded area}}{\text{jump in } u} = \frac{\int_{-\infty}^{\infty}[u(x,t) - u_0(x)]}{u_l - u_r}$$

Note that for finite time t, $u(\cdot, t) - u_0$ has compact support so that the integral makes sense. We now apply the same definition to the numerical solution, except that now the integral is replaced by a summation

$$x_{num}(n+1) - x_{num}(n) = \frac{\Delta x \sum_{j=-\infty}^{\infty}(v_j^{n+1} - v_j^n)}{u_l - u_r}$$

Using finite volume scheme we get

$$x_{num}(n+1) - x_{num}(n) = -\frac{\Delta t \sum_{j=-\infty}^{\infty}(g_{j+\frac{1}{2}}^n - g_{j-\frac{1}{2}}^n)}{u_l - u_r}$$

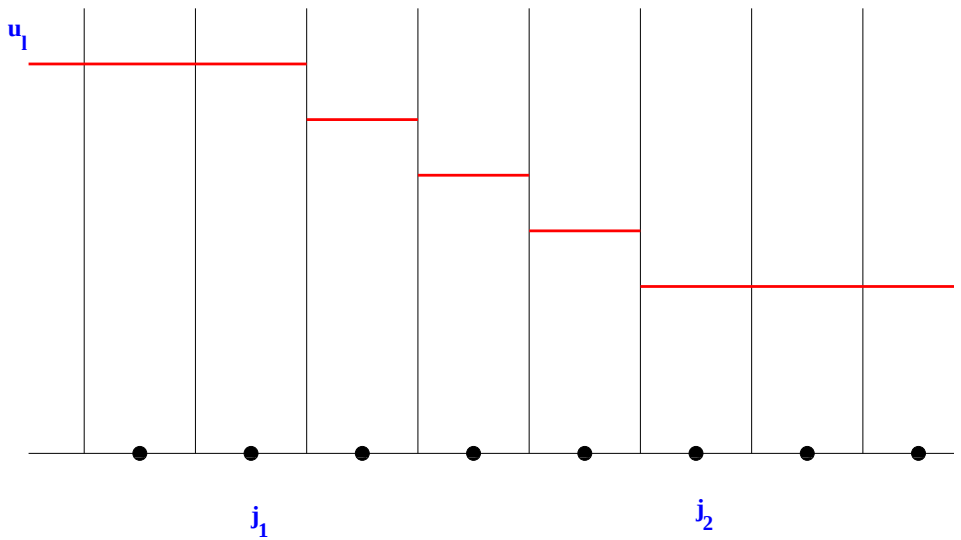
Assume that the scheme is TVD (this is not strictly necessary). Then for j very large or very small, we have by flux consistency

$$g_{j+\frac{1}{2}} = g_{j-\frac{1}{2}}$$

Consistency + conservation = correct shocks

Then

$$x_{num}(n+1) - x_{num}(n) = -\frac{\Delta t \sum_{j=j_1}^{j_2} (g^n_{j+\frac{1}{2}} - g^n_{j-\frac{1}{2}})}{u_l - u_r}$$



Consistency + conservation = correct shocks

From figure we see that there is a cancellation of the fluxes and the previous equation reduces to

$$x_{num}(n+1) - x_{num}(n) = -\frac{\Delta t(-g^n_{j_1-\frac{1}{2}} + g^n_{j_2+\frac{1}{2}})}{u_l - u_r}$$

But by consistency of the flux function

$$g^n_{j_1-\frac{1}{2}} = f(u_l), \qquad g^n_{j_2+\frac{1}{2}} = f(u_r)$$

Hence we get

$$\begin{aligned} x_{num}(n+1) - x_{num}(n) &= -\frac{\Delta t(-f_l + f_r)}{u_l - u_r} \\ &= \Delta t \left(\frac{f_l - f_r}{u_l - u_r} \right) \\ &= \Delta t \cdot S \end{aligned}$$

where S is the exact shock speed. We thus see that the numerical shock moves with the correct speed and hence has the correct shock location. The

Consistency + conservation = correct shocks

constraint of TVD can be easily removed. Even if there are wiggles in the solution we can still find a j_1, j_2 such that solution is constant for large $|j|$ and the proof will still hold. Also the solution need not be of a step function and can have any shape except that there should be only one discontinuity. The remarkable fact about this result is that ALL conservative schemes will give correct shock location; CONSERVATION is the only condition required.

In order to analyze the convergence of the solution v_j^n of the difference scheme (3) we introduce the *piecewise constant function* v_Δ defined a.e. in $\mathbb{R} \times (0, \infty)$ by

$$v_\Delta(x,t) = v_j^n, \quad x_{j-\frac{1}{2}} < x < x_{j+\frac{1}{2}}, \quad t_n \le t < t_{n+1}$$

where $x_{j+\frac{1}{2}} = (j + \frac{1}{2})\Delta x = \frac{1}{2}(x_j + x_{j+1})$. Then we study the convergence of v_Δ towards the weak solution of (1). We can define the initial condition as

$$v_j^0 = \frac{1}{\Delta x} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u_0(x)\mathrm{d}x \tag{9}$$

Lax-Wendroff Theorem 1.1 (GR1)

Consider any conservative and consistent scheme with v^0 given by (9). Assume that there exists a sequence $\Delta_k x$ which tends to zero such that $\lambda = \frac{\Delta_k t}{\Delta_k x}$ is kept constant. Assume that

1. $\|v_{\Delta_k}\|_{L^\infty(\mathbb{R}\times(0,\infty))} \le C$
2. v_{Δ_k} converges in $L^1_{loc}(\mathbb{R} \times (0, \infty))$ and a.e. to a function u

Then u is a weak solution of (1)-(2).

1. For non-conservative schemes there are counter examples where the scheme converges to some function which is not a weak solution; the RH conditions are not satisfied.
2. Shortcomings of the theorem
   1. Not a convergence theorem
   2. We dont know whether the entropy condition is satisfied.
   3. No result on the speed of convergence.

Definition

The order of accuracy of difference scheme (3) is the largest number $p \geq 1$ such that we have for any smooth solution u of (1) and for $\lambda = \frac{\Delta t}{\Delta x}$ fixed

$$\tau(x,t) := u(x,t+\Delta t) - H(u(x-k\Delta x,t),\ldots,u(x+k\Delta x,t)) = \mathcal{O}\left(\Delta t^{p+1}\right)$$

as $\Delta t \to 0$. The left hand side is called the *local truncation error*.

A consistent difference scheme whose numerical flux function g is a C^2 function is atleast first order accurate. More precisely we have the following proposition.

Proposition 1.2 (GR1)

Let (3) be a conservative, consistent difference scheme for (1). Assume that H is a C^2 function. Then for any solution u of (1) which is smooth enough and λ kept fixed, the truncation error has the following expression

$$\tau(x,t) = -\Delta t^2 \frac{\partial}{\partial x}\left[\beta(u,\lambda)\frac{\partial u}{\partial x}(x,t)\right] + \mathcal{O}\left(\Delta t^3\right) \tag{10}$$

where

$$\beta(u,\lambda) = \frac{1}{2\lambda^2}\sum_{j=-k}^{k} j^2 \frac{\partial H}{\partial v_j}(u,\ldots,u) - \frac{1}{2}a^2(u) \tag{11}$$

Proof: If we set

$$\bar{v} = (v_{-k},\ldots,v_{k-1}) \in \mathbb{R}^{2k} \qquad T\bar{v} = (v_{-k+1},\ldots,v_k) \in \mathbb{R}^{2k}$$

we have

$$H(v_{-k},\ldots,v_k) = v_0 - \lambda[g(T\bar{v}) - g(\bar{v})] \tag{12}$$

Differentiating (12) we obtain

$$\frac{\partial H}{\partial v_j} = \delta_j^0 - \lambda \left[\frac{\partial g}{\partial v_{j-1}}(T\bar{v}) - \frac{\partial g}{\partial v_j}(\bar{v}) \right], \quad -k \leq j \leq k$$

with the convention that

$$\frac{\partial g}{\partial v_{-k-1}} = 0 = \frac{\partial g}{\partial v_k}$$

Then

$$\begin{aligned} \sum_{j=-k}^{k} j \frac{\partial H}{\partial v_j}(u, \ldots, u) &= -\lambda \sum_{j=-k}^{k} j \left[\frac{\partial g}{\partial v_{j-1}}(u, \ldots, u) - \frac{\partial g}{\partial v_j}(u, \ldots, u) \right] \\ &= -\lambda \sum_{j=-k}^{k} \frac{\partial g}{\partial v_j}(u, \ldots, u) \end{aligned}$$

Differentiation of the consistency condition (8) gives

$$\sum_{j=-k}^{k} \frac{\partial g}{\partial v_j}(u, \ldots, u) = a(u)$$

we find

$$\sum_{j=-k}^{k} j \frac{\partial H}{\partial v_j}(u, \ldots, u) = -\lambda a(u)$$

Next differentiating (12) twice gives

$$\frac{\partial^2 H}{\partial v_i \partial v_j} = -\lambda \left[\frac{\partial^2 g}{\partial v_{i-1} \partial v_{j-1}}(T\bar{v}) - \frac{\partial^2 g}{\partial v_i \partial v_j}(\bar{v}) \right]$$

and

$$\sum_{i,j=-k}^{+k} (i-j)^2 \frac{\partial^2 H}{\partial v_i \partial v_j}(u, \ldots, u) =$$

$$-\lambda \sum_{i,j=-k}^{+k} (i-j)^2 \left[\frac{\partial^2 g}{\partial v_{i-1} \partial v_{j-1}}(u, \ldots, u) - \frac{\partial^2 g}{\partial v_i \partial v_j}(u, \ldots, u) \right] = 0$$

so that by symmetry of $(\partial^2 H / \partial v_i \partial v_j(u, \ldots, u))_{i,j}$

$$\sum_{i,j=-k}^{+k} (ij - j^2) \frac{\partial^2 H}{\partial v_i \partial v_j}(u, \ldots, u) = -\frac{1}{2} \sum_{i,j=-k}^{+k} (i-j)^2 \frac{\partial^2 H}{\partial v_i \partial v_j}(u, \ldots, u) = 0$$

Next we do Taylor expansion of H around the point $u, \ldots, u$ with the notation $u = u(x,t)$ and $u_j = u(x + j\Delta x, t)$

$$\begin{aligned} H(u_{-k}, \ldots, u_k) &= u + \sum_{j=-k}^{+k} \frac{\partial H}{\partial v_j}(u, \ldots, u)(u_j - u) \\ &\quad + \frac{1}{2} \sum_{i,j=-k}^{+k} \frac{\partial^2 H}{\partial v_i \partial v_j}(u, \ldots, u)(u_i - u)(u_j - u) + \mathcal{O}\left(\Delta x^3\right) \end{aligned}$$

Now

$$u_j - u = j\Delta x \frac{\partial u}{\partial x} + \frac{1}{2}(j\Delta x)^2 \frac{\partial^2 u}{\partial x^2} + \mathcal{O}\left(\Delta x^3\right)$$

and

$$(u_i - u)(u_j - u) = ij\Delta x^2 \left(\frac{\partial u}{\partial x}\right)^2 + \mathcal{O}\left(\Delta x^3\right)$$

Hence

$$H(u_{-k},\dots,u_k) = u + \Delta x \sum_{j=-k}^{+k} j \frac{\partial H}{\partial v_j}(u,\dots,u)\frac{\partial u}{\partial x} + \frac{\Delta x^2}{2}\sum_{j=-k}^{+k} j^2 \frac{\partial H}{\partial v_j}(u,\dots,u)\frac{\partial^2 u}{\partial x^2} + \frac{\Delta x^2}{2}\sum_{i,j=-k}^{+k} ij \frac{\partial^2 H}{\partial v_i \partial v_j}(u,\dots,u)\left(\frac{\partial u}{\partial x}\right)^2 + \mathcal{O}\left(\Delta x^3\right)$$

Last two terms become

$$\sum_{j=-k}^{+k} j^2 \frac{\partial H}{\partial v_j}(u,\dots,u)\frac{\partial^2 u}{\partial x^2} + \sum_{i,j=-k}^{+k} ij \frac{\partial^2 H}{\partial v_i \partial v_j}(u,\dots,u)\left(\frac{\partial u}{\partial x}\right)^2$$

$$= \frac{\partial}{\partial x}\sum_{j=-k}^{+k} j^2 \frac{\partial H}{\partial v_j}(u,\dots,u)\frac{\partial u}{\partial x} + \cancel{\sum_{i,j=-k}^{+k} (ij - j^2)\frac{\partial^2 H}{\partial v_i \partial v_j}(u,\dots,u)\left(\frac{\partial u}{\partial x}\right)^2}$$

$$= \frac{\partial}{\partial x}\sum_{j=-k}^{+k} j^2 \frac{\partial H}{\partial v_j}(u,\dots,u)\frac{\partial u}{\partial x}$$

Hence

$$H(u_{-k},\dots,u_k) = u - \Delta t\, a(u)\frac{\partial u}{\partial x} + \frac{1}{2}\Delta x^2 \frac{\partial}{\partial x}\sum_{j=-k}^{+k} j^2 \frac{\partial H}{\partial v_j}(u,\dots,u)\frac{\partial u}{\partial x} + \mathcal{O}\left(\Delta x^3\right)$$

Expand $u(x, t+\Delta t)$ using Taylors formula

$$u(x,t+\Delta t) = u + \Delta t\frac{\partial u}{\partial t} + \frac{\Delta t^2}{2}\frac{\partial^2 u}{\partial t^2} + \mathcal{O}\left(\Delta t^3\right)$$

For smooth solution, conservation law can be written as

$$\frac{\partial u}{\partial t} = -\frac{\partial f}{\partial x} = -a(u)\frac{\partial u}{\partial x}$$

and

$$\frac{\partial^2 u}{\partial t^2} = -\frac{\partial}{\partial t}\frac{\partial f}{\partial x} = -\frac{\partial}{\partial x}\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x}a(u)\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}a^2(u)\frac{\partial u}{\partial x}$$

Combining everything we get

$$u(x,t+\Delta t) - H(u(x-k\Delta x,t),\dots,u(x+k\Delta x,t)) = \Delta t\cancel{\left(\frac{\partial u}{\partial x} + a(u)\frac{\partial u}{\partial x}\right)}$$

$$+ \Delta t^2 \frac{\partial}{\partial x}\left\{\frac{1}{2}[a^2(u) - \frac{1}{\lambda^2}\sum_{j=-k}^{+k} j^2 \frac{\partial H}{\partial v_j}(u,\dots,u)]\frac{\partial u}{\partial x}\right\} + \mathcal{O}\left(\Delta t^3\right)$$

□

Modified PDE

Assume that difference scheme (6) is first order accurate in which case $\beta(u, \lambda) \neq 0$. Consider the following second order differential equation

$$\frac{\partial v}{\partial t} + \frac{\partial}{\partial x} f(v) - \lambda \Delta x \frac{\partial}{\partial x} \beta(u, \lambda) \frac{\partial v}{\partial x} = 0 \tag{13}$$

Then

$$v(x, t + \Delta t) - H(v(x - k\Delta x, t), \ldots, v(x + k\Delta x, t)) = \mathcal{O}\left(\Delta t^3\right)$$

Hence the scheme can be viewed as a second order accurate approximation of (13), which is called the *modified partial differential equation*. This equation can be used for a heuristic stability analysis based on the nature of the extra terms in this equation.

L^2 stability

When the flux is a linear function of u, $f(u) = au$ and the difference scheme is also linear, we can perform L^2-stability by means of Fourier transforms. Hence, let us consider the linear advection equation

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0 \tag{14}$$

and suppose that the difference scheme (3) is linear, i.e., of the form

$$v_j^{n+1} = \sum_{l=-k}^{+k} c_l v_{j+l}^n \tag{15}$$

for some coefficients $c_l = c_l(a, \lambda)$. Define the l^2-norm of a sequence $v = (v_j)_{j\in\mathbb{Z}}$ as

$$\|v\|_{L^2(\Delta)} = \left(\Delta x \sum_{j\in\mathbb{Z}} v_j^2\right)^{\frac{1}{2}} \tag{16}$$

L^2 stability

Then the difference scheme is L^2-stable if $\exists C > 0$ independent of Δt such that

$$\left\|v^n\right\|_{L^2(\Delta)} \leq C \left\|v^0\right\|_{L^2(\Delta)}, \qquad \forall n \geq 0 \tag{17}$$

For convenience, we extend the scheme (15) to the whole real line by setting

$$v^{n+1}(x) = \sum_{l=-k}^{k} c_l v^n(x + l\Delta x) \tag{18}$$

so that the stability condition (17) becomes

$$\left\|v^n\right\|_{L^2(\mathbb{R})} \leq C \left\|v^0\right\|_{L^2(\mathbb{R})} \tag{19}$$

By using the Fourier transform

$$\widehat{\phi}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}x\xi} \phi(x) \mathrm{d}x$$

L^2 stability

equation (18) can be written as

$$\widehat{v}^{n+1}(\xi) = h(\xi)\widehat{v}^{n}(\xi) \qquad \text{where} \qquad h(\xi) = \sum_{l=-k}^{+k} c_l \mathrm{e}^{\mathrm{i}l\xi\Delta x} \tag{20}$$

Von Neumann condition

The linear difference scheme (15) is L^2-stable if and only if its amplification factor h defined by (20) satisfies

$$|h(\xi)| \leq 1, \qquad \forall \xi \in \mathbb{R} \tag{21}$$

This is known as *von Neumann condition*.

Proof: Use Parseval identity

$$\int_{\mathbb{R}} |v(x)|^2 \mathrm{d}x = \int_{\mathbb{R}} |\hat{v}(\xi)|^2 \mathrm{d}\xi$$

Linear 3-point scheme

$$v_j^{n+1} = c_{-1} v_{j-1}^n + c_0 v_j^n + c_1 v_{j+1}^n$$

This scheme is conservative iff

$$\sum_j v_j^{n+1} = c_{-1} \sum_j v_{j-1}^n + c_0 \sum_j v_j^n + c_1 \sum_j v_{j+1}^n = (c_{-1} + c_0 + c_1) \sum_j v_j^n$$

i.e. if

$$c_{-1} + c_0 + c_1 = 1$$

In that case

$$v_j^{n+1} = c_{-1} v_{j-1}^n + (1 - c_{-1} - c_1) v_j^n + c_1 v_{j+1}^n$$

and the numerical flux is given by

$$g(u, v) = \frac{1}{\lambda}(c_{-1} u - c_1 v)$$

and consistency of g implies

$$g(u, u) = \frac{1}{\lambda}(c_{-1} - c_1) u = a u \quad \Longrightarrow \quad c_{-1} - c_1 = a\lambda$$

Linear 3-point scheme

Define $q = c_{-1} + c_1 = 1 - c_0$, we get

$$v_j^{n+1} = v_j^n - \frac{a\lambda}{2}(v_{j+1}^n - v_{j-1}^n) + \frac{q}{2}(v_{j-1}^n - 2v_j^n + v_{j+1}^n) \tag{22}$$

Hence we obtain that linear 3-point schemes which are conservative and consistent form a one parameter family of schemes. We will interpret the parameter $q = q(a, \lambda)$ as the coefficient of numerical viscosity and we call

$$\nu = a\lambda = \frac{a\Delta t}{\Delta x}$$

as the Courant number or CFL number.

Proposition

A linear 3-point consistent, conservative difference scheme (22) is L^2-stable iff the coefficient q satisfies

$$\nu^2 \leq q \leq 1$$

Linear 3-point scheme

Proof: The amplification factor is given by

$$\begin{aligned} h(\xi) &= 1 - \frac{\nu}{2}(\mathrm{e}^{\mathrm{i}\xi\Delta x} - \mathrm{e}^{-\mathrm{i}\xi\Delta x}) + \frac{q}{2}(\mathrm{e}^{-\mathrm{i}\xi\Delta x} - 2 + \mathrm{e}^{\mathrm{i}\xi\Delta x}) \\ &= 1 - q[1 - \cos(\xi\Delta x)] + \mathrm{i}\nu \sin(\xi\Delta x) \end{aligned}$$

Setting $y = \cos^2(\frac{1}{2}\xi\Delta x)$, we obtain

$$|h(\xi)|^2 = (1 - 2qy)^2 + 4\nu^2 y(1 - y) =: m(y)$$

Since $0 \le y \le 1$, the second term is always positive. We first make $(1 - 2qy)^2 < 1$ which requires $0 < q \le 1$. Since $m(0) = 1$, we should have $m'(0) \le 0$.

$$m'(y) = 4(\nu^2 - q) + 8y(q^2 - \nu^2), \qquad m'(0) = 4(\nu^2 - q)$$

Hence

$$m'(0) \le 0 \quad \Rightarrow \quad \nu^2 \le q$$

Linear 3-point scheme

From these two conditions, we get $0 < \nu^2 \leq q \leq 1$. We now show that this is sufficient.

$$\begin{aligned} m(y) &= (1-2qy)^2 + 4\nu^2 y(1-y) \\ &\leq (1-2qy)^2 + 4qy(1-y) \\ &= 1 - 4y^2 q(1-q) \\ &\leq 1 \end{aligned}$$

Linear 3-point scheme

Second order accuracy: From local truncation error

$$\beta(u,\lambda) = \frac{q}{2\lambda^2} - \frac{a^2}{2}$$

For second order accuracy

$$\beta(u,\lambda) = 0 \qquad \Longrightarrow \qquad q = \lambda^2 a^2 = \nu^2$$

There is only one 3-point, consistent, conservative, second order scheme and it is the Lax-Wendroff scheme. This is L^2 stable if the CFL condition

$$|a|\lambda \leq 1 \qquad \Longrightarrow \qquad \Delta t \leq \frac{\Delta x}{|a|}$$

is satisfied.

Remark: β determines the numerical viscosity. A large value of β (hence q) means there is more numerical viscosity.

Linear 3-point scheme

Conservation form: Numerical flux

$$g_{j+\frac{1}{2}} = \frac{1}{2}a(v_j + v_{j+1}) - \frac{q}{2\lambda}(v_{j+1} - v_j)$$

q	$a^2\lambda^2$	$\lvert a\rvert\lambda$	1
Scheme	Lax-Wendroff	Upwind	Lax-Friedrichs

- All of these schemes are stable since they satisfy CFL condition $|a|\lambda \leq 1$
- Lax-Wendroff scheme has least numerical dissipation
- Lax-Friedrichs scheme has the highest numerical dissipation