Numerical schemes for systems of conservation laws in 1D

In this part, we are interested in the numerical approximation by finite volume methods of the solution of the 1D problem

$$\begin{cases} \partial_t u + \partial_x f(u) = 0, & x \in \mathbb{R}, \ t > 0, \\ u(x,0) = u_0(x), \end{cases}$$

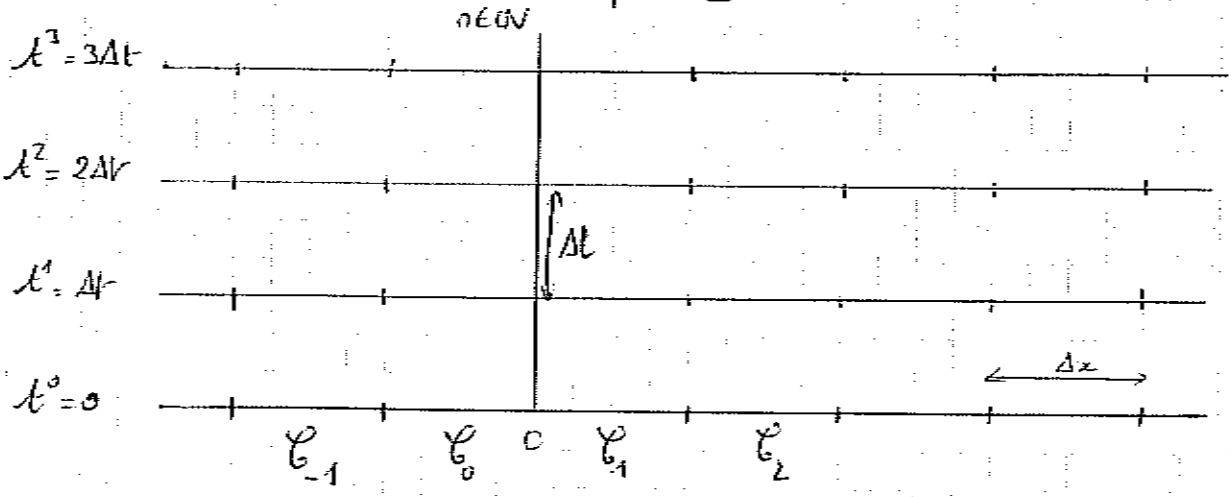
where $u = (u_1, \dots, u_p)^t \in \Omega \subset \mathbb{R}^p$ is the unknown and $f(u) \in \mathbb{R}^p$ is the flux function.

Let us first briefly recall the design principle of the finite volume approach.

We begin by introducing a time step Δt and a space step Δx that we assume to be constant for simplicity. We set $\lambda = \Delta t / \Delta x$ and define the mesh interfaces $x_{j+1/2} = j \Delta x$ for $j \in \mathbb{Z}$, and the intermediate times $t^n = n \Delta t$ for $n \in \mathbb{N}$. We thus have

$$\mathbb{R} = \bigcup_{j \in \mathbb{Z}} \mathcal{C}_j \quad \text{with} \quad \mathcal{C}_j = [x_{j-1/2}, x_{j+1/2}[$$

and $$\mathbb{R}^+ = \bigcup_{n \in \mathbb{N}} [t^n, t^{n+1}[$$



At each time t^n, we look for an approximation u_j^n of the mean value $\frac{1}{\Delta x} \int_{\mathcal{C}_j} u(x, t^n)\, dx$ of the exact solution at time t^n on the cell $\mathcal{C}_j$.

We assume that the initial condition u_0 is known, and set for instance for $n = 0$

$$u_j^0 = \frac{1}{\Delta x} \int_{C_j} u_0(x)\, dx, \quad \forall j \in \mathbb{Z}.$$

Being given the sequence $(u_j^n)_{j \in \mathbb{Z}}$ at time t^n, the main point is now to propose a definition of the sequence $(u_j^{n+1})_{j \in \mathbb{Z}}$ by a recurrence relation. To do so, we start from the governing equations to be solved, namely

$$\partial_t u + \partial_x f(u) = 0 \qquad (3)$$

and integrate them on the volume $C_j \times [t^n, t^{n+1}[= [x_{j-1/2}, x_{j+1/2}[\times [t^n, t^{n+1}[$

$$\int_{t^n}^{t^{n+1}} \int_{C_j} (\partial_t u + \partial_x f(u))\, dx\, dt = 0,$$

which gives

$$\int_{C_j} u(x, t^{n+1})\, dx - \int_{C_j} u(x, t^n)\, dx + \int_{t^n}^{t^{n+1}} f(u(x_{j+1/2}, t))\, dt - \int_{t^n}^{t^{n+1}} f(u(x_{j-1/2}, t))\, dt = 0.$$

Writing now

$$\int_{C_j} u(x, t^{n+1})\, dx \simeq \Delta x\, u_j^{n+1}, \quad j \in \mathbb{Z},$$

$$\int_{C_j} u(x, t^n)\, dx \simeq \Delta x\, u_j^n, \quad j \in \mathbb{Z},$$

we then deduce the following form for a natural finite volume approximation of the solution of (3):

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{\Delta x} \left(f_{j+1/2} - f_{j-1/2} \right)$$

where $f_{j+1/2} \simeq \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} f(u(x_{j+1/2}, t))\, dt$ is the so-called numerical flux and represents an approximation of the exact flux that passes through the interface $x_{j+1/2}$ in the time interval $[t^n, t^{n+1}[$.

The choice of the numerical flux $f_{j+1/2}$ at each interface $x_{j+1/2}$ defines the numerical scheme under consideration. In the following, we propose to study several numerical schemes, focusing on the following form for the numerical flux:

$$\forall j: \quad f_{j+1/2} := f(u_j^n, u_{j+1}^n) \quad (6)$$

The function $f(\cdot,\cdot)$ is called the numerical flux function and this leads to the following update formula:

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{\Delta x}\left(f(u_j^n, u_{j+1}^n) - f(u_{j-1}^n, u_j^n)\right), \quad j \in \mathbb{Z}. \quad (7)$$

We observe that the definition of u_j^{n+1} only depends on the three values $u_{j-1}^n, u_j^n, u_{j+1}^n$. (7) is said to be a three-point explicit finite volume numerical scheme.

Definition

The scheme (7) is said to be consistent with (3) if $f(\cdot,\cdot)$ satisfies

$$f(u,u) = f(u) \quad \forall u \in \Omega.$$

Let us also briefly recall that we are basically interested in this course in entropy weak solutions of (1), that is weak solutions of (1) satisfying in addition the entropy condition

$$\partial_t S(u) + \partial_x G(u) \leq 0$$

in the distributional sense, where $(S(u), G(u))$ represents an entropy - entropy flux pair for (1) with $S: \Omega \to \mathbb{R}$ a convex function. We then introduce the following definition.

Definition

The scheme (7) is said to consistant with (3)-(9) if (8) holds true and if there exists a numerical entropy flux function $G(\cdot,\cdot)$,

consistant with $G(\cdot)$ in the sense that

$$G(u,u) = G(u) \quad \forall u \in \Omega$$

and such that

$$S(u_j^{n+1}) - S(u_j^n) - \frac{\Delta t}{\Delta x}\left(G_j(u_j^n, u_{j+1}^n) - G(u_{j-1}^n, u_j^n)\right) \leq 0 \quad \forall j \in \mathbb{Z}. \tag{9}$$

To conclude this introduction, we briefly recall the Lax-Wendroff theorem that asserts that when a consistent scheme of the form (7) converges in some sense to a function u, the limit is a weak solution of (3). In addition, if (9) is entropy consistent then the limit is an entropy weak solution of (3)-(9).

The Godunov method

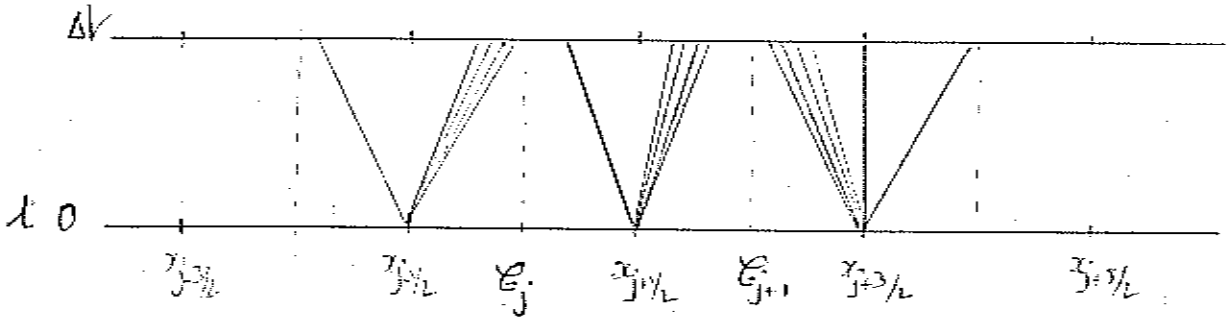
We begin with the celebrated Godunov's method. Let us define the piecewise constant approximate solution $x \to u_\Delta(x, t^n)$ at time t^n given by $u_\Delta(x, t^n) = u_j^n \quad \forall x \in C_j, j \in \mathbb{Z}, n \in \mathbb{N}$. The method is made of two steps.

Step 1 : Evolution in time

In this first step, one solves the Cauchy problem (1)-(9) exactly with initial data $u_0(x) = u_\Delta(x, t^n)$ and for times $t \in [0, \Delta t]$. Locally around each interface $x_{j+1/2}$, it is clear that this amounts to solve a Riemann problem. More precisely, one can even claim that under the so called CFL condition involving the characteristic speeds $\lambda_i, i = 1, \dots, p$ of (3)

$$\frac{\Delta t}{\Delta x} \max_u \{|\lambda_i(u)|, i = 1, \dots, p\} \leq 1/2, \tag{11}$$

for all the u under consideration, the solution of this Cauchy problem is known by glueing together the solutions of the Riemann problems set at each interface, see figure below.



More precisely, this solution is given by

$$\tilde{u}(x,t) = u_R\left(\frac{x-x_{j+1/2}}{t}; u_j^n, u_{j+1}^n\right) \text{ for all } (x,t) \in [x_j, x_{j+1}] \times [0, \Delta t],$$

where $x_j = \frac{1}{2}(x_{j-1/2} + x_{j+1/2})$ for all $j \in \mathbb{Z}$ and where $(x,t) \to u_R(x/t, u_L, u_R)$ denotes the self-similar solution of the Riemann problem (1)-(2) with initial data $u_0(x) = \begin{cases} u_L \text{ if } x < 0 \\ -u_R \text{ if } x > 0 \end{cases}$.

Step 2 : Projection

In this second step, we get back a piecewise constant approximate solution on each cell C_j at time t^{n+1} by averaging the solution $u(x, \Delta t)$:

$$u_j^{n+1} = \frac{1}{\Delta x} \int_{C_j} \tilde{u}(x, \Delta t)\, dx, \quad j \in \mathbb{Z}$$

Then, the process repeats.

In practice, this algorithm is considerably simplified by observing that the cell average (14) can be easily computed using the integral form of the governing equation (3), that is $\partial_t u + \partial_x f(u) = 0$. Indeed, since (13) is assumed to be an exact solution, integrating over $C_j \times [0, \Delta t]$ we know that

$$\int_{C_j} \tilde{u}(x, \Delta t)\, dx - \int_{C_j} \tilde{u}(x, 0)\, dx + \int_{t^n}^{t^{n+1}} f(\tilde{u}(x_{j+1/2}, t))\, dt - \int_{t^n}^{t^{n+1}} f(\tilde{u}(x_{j-1/2}, t))\, dt = 0$$

that is

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{\Delta x}\left(f_{j+1/2} - f_{j-1/2}\right), \quad j \in \mathbb{Z}$$

where the numerical flux $f_{j+1/2} = f(u_j^n, u_{j+1}^n)$ is given by

$$f_{j+1/2} = f(u_j^n, u_{j+1}^n) = \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} f(\tilde{u}(x_{j+1/2}, t))\,dt$$

$$= \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} f(u_R(0; u_j^n, u_{j+1}^n))\,dt$$

That is

$$f_{j+1/2} = f(u_j^n, u_{j+1}^n) = f(u_R(0; u_j^n, u_{j+1}^n)), \quad j \in \mathbb{Z}.$$

This shows that the Godunov method can be written in the form (7). It is also clear that the method is consistant with (3) since

$$f(u,u) = f(u_R(0; u, u)) = f(u).$$

As far as the entropy consistency is concerned, since (13) is assumed to satisfy the entropy inequality $\partial_t S(u) + \partial_x G(u) \le 0$, integrating over $C_j \times [0, \Delta t]$ leads to

$$\int_{C_j} S(\tilde{u}(x, \Delta t))\,dx - \int_{C_j} S(\tilde{u}(x,0))\,dx + \int_{t^n}^{t^{n+1}} G(\tilde{u}(x_{j+1/2}, t))\,dt - \int_{t^n}^{t^{n+1}} G(\tilde{u}(x_{j-1/2}, t))\,dt \le 0$$

That is

$$\frac{1}{\Delta x}\int_{C_j} S(\tilde{u}(x, \Delta t))\,dx - S(u_j^n) + \frac{\Delta t}{\Delta x}\left(G(u_j^n, u_{j+1}^n) - G(u_{j-1}^n, u_j^n)\right) \le 0 \qquad (16)$$

with $$G_{j+1/2} = G(u_j^n, u_{j+1}^n) = G(u_R(0, u_j^n, u_{j+1}^n)), \quad j \in \mathbb{Z} \qquad (17)$$

which is clearly consistent with G since $G(u,u) = G(u_R(0,u,u)) = G(u)$. But invoking the Jensen inequality and the convexity of $u \to S(u)$, we have

$$S\left(\frac{1}{\Delta x}\int_{C_j} \tilde{u}(x, \Delta t)\,dx\right) \le \frac{1}{\Delta x}\int_{C_j} S(\tilde{u}(x, \Delta t))\,dx$$

That is $$S(u_j^{n+1}) \le \frac{1}{\Delta x}\int_{C_j} S(\tilde{u}(x, \Delta t))\,dx.$$

Combining (16) and (17) proves that Godunov's method is consistant with (3).(9).

The main drawback of Godunov's method is that it requires at each time iteration and at each interface $x_{j+1/2}$, the resolution of exact Riemann problems. Although these Riemann problems can be solved in theory, in practice it necessitates the resolution of nonlinear equations which can be expensive. We note also that most of the structure of the exact Riemann solutions is not used in Godunov's method since an averaging process is involved. This discussion suggests that it is probably not worthwhile and relevant calculating the Riemann solutions exactly, instead of that defining approximate Riemann solutions could be less expensive and give equally good numerical results, provided that these approximate solutions are consistant in some "averaged" sense.

Approximate Riemann solvers

Our objective is to propose an approximation of $(x,t) \longrightarrow u_R(x/t, u_L, u_R)$. In this course, we will focus on <u>simple</u> approximate Riemann solvers of the following form

$$\tilde{u}_R(x/t, u_L, u_R) = \begin{cases} u_1 = u_L & \text{if } x/t < \tilde{\lambda}_1 \\ \vdots & \\ u_k & \text{if } \tilde{\lambda}_{k-1} < \frac{x}{t} < \tilde{\lambda}_k, \quad k = 2, \dots, l \\ \vdots & \\ u_{l+1} = u_R & \text{if } \frac{x}{t} > \tilde{\lambda}_l, \end{cases}$$

with u_k and $\tilde{\lambda}_k = \tilde{\lambda}_k(u_L, u_R)$, $k = 1, \dots, l$, $l \leq p$ to be defined.
Note that this approximate Riemann solver is self-similar.
Following the general theory of Harten-Lax and van Leer, a suitable approximate Riemann solver is subject to the following consistency properties.

Definition

The simple approximate Riemann solver (19) is said to be consistant with the integral form of (3) if and only if for Δx and Δt such that

$$\max_{1\le k\le \ell} |\tilde\lambda_k(u_L,u_R)|\,\frac{\Delta t}{\Delta x} \le 1/2,$$

we have

$$\frac{1}{\Delta x}\int_{-\frac{\Delta x}{2}}^{\frac{\Delta x}{2}} \tilde u_R(x/\Delta t; u_L,u_R)\,dx = \frac{u_L+u_R}{2} - \frac{\Delta t}{\Delta x}\big(f(u_R)-f(u_L)\big)\,.$$

(and if $\tilde u_R(x/t;u,u)=u \quad \forall (x,t)$)

Where does this equality come from and what does it mean? Actually, starting from $\partial_t u + \partial_x f(u)=0$ and integrating over $[-\frac{\Delta x}{2},\frac{\Delta x}{2}]\times[0,\Delta t]$ gives for the exact Riemann solution $u_R(\cdot,u_L,u_R)$

$$\frac{1}{\Delta x}\int_{-\frac{\Delta x}{2}}^{\frac{\Delta x}{2}} u_R(x/\Delta t,u_L,u_R)\,dx = \frac{1}{\Delta x}\int_{-\frac{\Delta x}{2}}^{\frac{\Delta x}{2}} u_0(x)\,dx - \frac{\Delta t}{\Delta x}\big(f(u_R)-f(u_L)\big)$$

$$= \frac{u_L+u_R}{2} - \frac{\Delta t}{\Delta x}\big(f(u_R)-f(u_L)\big)$$

since u_0 is given by $u_0(x)=\begin{cases} u_L & \text{if } x<0\\ u_R & \text{if } x>0\end{cases}$.

Then, (21) says that the exact and approximate Riemann solutions may be different "in the details", but must coincide "in average" over a cell $[-\frac{\Delta x}{2},\frac{\Delta x}{2}]$ with length Δx and centered on the position of the initial discontinuity.

Remark

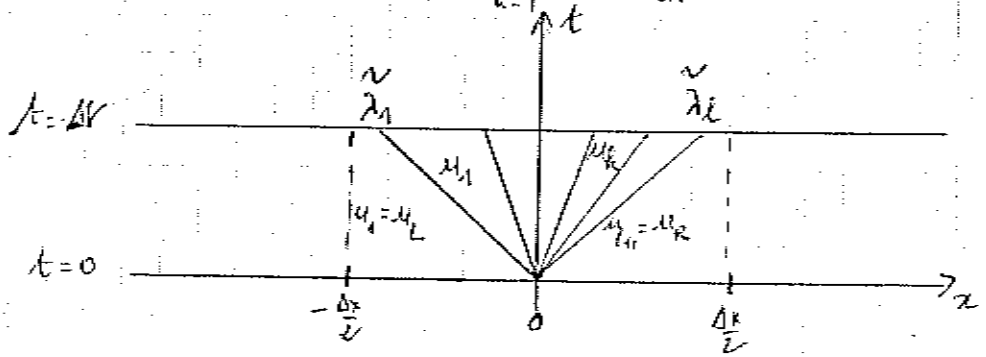
Since we clearly have (see figure below)

$$\int_{-\frac{\Delta x}{2}}^{\frac{\Delta x}{2}} \tilde u_R(x/\Delta t;u_L,u_R)\,dx = \Big(\frac{\Delta x}{2}+\tilde\lambda_1\Delta t\Big)u_L + (\tilde\lambda_2-\tilde\lambda_1)\Delta t\,u_2 + \cdots + (\tilde\lambda_\ell-\tilde\lambda_{\ell-1})\Delta t\,u_\ell + \Big(\frac{\Delta x}{2}-\tilde\lambda_\ell\Delta t\Big)u_R$$

$$= \frac{\Delta x}{2}(u_L+u_R) - \Delta t\sum_{k=1}^{\ell}\tilde\lambda_k(u_{k+1}-u_k),$$

the consistency relation (21) also writes

$$f(u_R)-f(u_L) = \sum_{k=1}^{\ell}\tilde\lambda_k(u_{k+1}-u_k)$$



Similarly, we introduce the notion of consistency with the entropy inequality $\partial_t S(u) + \partial_x G(u) \leq 0$.

integral form of the

Definition

The simple approximate Riemann solver (19) is said to be consistant with the integral form of the entropy inequality (3) if and only if for Δx and Δt satisfying (20) we have

$$\frac{1}{\Delta x}\int_{-\frac{\Delta x}{2}}^{\frac{\Delta x}{2}} S(\tilde{u}_R(\tfrac{x}{\Delta t}; u_L, u_R))\,dx \leq \frac{S(u_L)+S(u_R)}{2} - \frac{\Delta t}{\Delta x}\left(G(u_R) - G(u_L)\right),$$

or equivalently

$$G(u_R) - G(u_L) \leq \sum_{k=1}^{\ell} \tilde{\lambda}_k^- \left(S(u_{k+1}) - S(u_k)\right).$$

The associated Godunov-type method

The idea is to follow the same approach as in the original Godunov method but replacing the exact Riemann solution $u_R(\cdot; u_L, u_R)$ with the approximate one $\tilde{u}_R(\cdot; u_L, u_R)$. Then we get:

Step 1 Approximate evolution in time

One solves approximately the Cauchy problem (1)-(3) with initial condition $u_0(x) = u_j^n$ for times $t \in [0, \Delta t]$ with Δt such that

$$\frac{\Delta t}{\Delta x} \max_{k=1,\dots,\ell} |\lambda_k(u_j^n, u_{j+1}^n)| < \frac{1}{2} \qquad (23)$$

and using the approximate Riemann solution. We get

$$\tilde{u}(x,t) = \tilde{u}_R\left(\frac{x - x_{j+1/2}}{t}; u_j^n, u_{j+1}^n\right) \text{ for all } (x,t) \in [x_j, x_{j+1}] \times [0, \Delta t].$$

Step 2 Projection

This step is unchanged and we set

$$u_j^{n+1} = \frac{1}{\Delta x}\int_{I_j} \tilde{u}(x, \Delta t)\,dx, \quad j \in \mathbb{Z},$$

and the process repeats.

At this stage, it is natural to wonder if such a Godunov-type scheme admits an equivalent formulation of the form (7) with a given numerical flux $f_{i+1/2}$. Note that since the exact Riemann solution has been replaced with an approximate one, the calculations carried out in the context of the Godunov method and leading to (11) are not valid anymore here. However, we can write

$$u_j^{n+1} = \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} \tilde{u}(x,\Delta t)\,dx =$$

$$= \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_j} \tilde{u}_R\left(\frac{x-x_j}{\Delta t}; u_{j-1}^n, u_j^n\right)dx + \frac{1}{\Delta x}\int_{x_j}^{x_{j+1/2}} \tilde{u}_R\left(\frac{x-x_{j+1/2}}{\Delta t}; u_j^n, u_{j+1}^n\right)dx$$

$$= \frac{1}{\Delta x}\int_0^{\frac{\Delta x}{2}} \tilde{u}_R\left(\frac{x}{\Delta t}; u_{j-1}^n, u_j^n\right)dx + \frac{1}{\Delta x}\int_{-\frac{\Delta x}{2}}^{0} \tilde{u}_R\left(\frac{x}{\Delta t}; u_j^n, u_{j+1}^n\right)dx$$

$$= \frac{1}{2}\left(u^+(u_{j-1}^n, u_j^n) + u^-(u_j^n, u_{j+1}^n)\right)$$

where we have set

$$\begin{cases} u^-(u_L,u_R) = \dfrac{2}{\Delta x}\displaystyle\int_{-\frac{\Delta x}{2}}^{0} \tilde{u}_R\left(\frac{x}{\Delta t}, u_L, u_R\right)dx, \\ u^+(u_L,u_R) = \dfrac{2}{\Delta x}\displaystyle\int_{0}^{\frac{\Delta x}{2}} \tilde{u}_R\left(\frac{x}{\Delta t}; u_L, u_R\right)dx. \end{cases}$$

Using the form (19) of the approximate Riemann solver, we can easily check that

$$u^-(u_L,u_R) = u_L - \frac{2\Delta t}{\Delta x}\sum_{k=1}^{l} \tilde{\lambda}_k^- (u_{k+1} - u_k) \qquad x^+ = \max(x,0),\ x^- = \min(x,0)$$

$$u^+(u_L,u_R) = u_R - \frac{2\Delta t}{\Delta x}\sum_{k=1}^{l} \tilde{\lambda}_k^+ (u_{k+1} - u_k)$$

so that, with clear notations

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{\Delta x}\left(\sum_{k=1}^{l} \tilde{\lambda}_{k,j+1/2}^- (u_{k,j+1}^{j+1/2} - u_{k,j}^{j+1/2}) + \sum_{k=1}^{l} \tilde{\lambda}_{k,j-1/2}^+ (u_{k,j}^{j-1/2} - u_{k,j-1}^{j-1/2})\right).$$

Let us assume that the approximate Riemann solver under consideration is consistent with the integral form of (3).

Using the property $\begin{cases} x = x^+ + x^- \\ |x| = x^+ - x^- \end{cases}$, we can write

$$\sum_{k=1}^{\ell} \tilde{\lambda}^-_{k\,j+1/2} \left(u^{j+1/2}_{k+1} - u^{j+1/2}_k\right) = \frac{1}{2} \sum_{k=1}^{\ell} \left(\tilde{\lambda}_{k\,j+1/2} - |\tilde{\lambda}_{k\,j+1/2}|\right)\left(u^{j+1/2}_{k+1} - u^{j+1/2}_k\right)$$

$$= \frac{1}{2}\left(f(u_{j+1}) - f(u_j)\right) - \frac{1}{2} \sum_{k=1}^{\ell} |\tilde{\lambda}_{k\,j+1/2}| \left(u^{j+1/2}_{k+1} - u^{j+1/2}_k\right).$$

Similarly we get

$$\sum_{k=1}^{\ell} \tilde{\lambda}^+_{k\,j-1/2} \left(u^{j-1/2}_{k+1} - u^{j-1/2}_k\right) = \frac{1}{2}\left(f(u_j) - f(u_{j-1})\right) + \frac{1}{2} \sum_{k=1}^{\ell} |\tilde{\lambda}_{k\,j-1/2}| \left(u^{j-1/2}_{k+1} - u^{j-1/2}_k\right)$$

So that we eventually have

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{\Delta x}\left(f_{j+1/2} - f_{j-1/2}\right), \quad j \in \mathbb{Z}$$

with

$$f_{j+1/2} = f(u_j^n, u_{j+1}^n) = \frac{1}{2}(f(u_j) + f(u_{j+1})) - \frac{1}{2} \sum_{k=1}^{\ell} |\tilde{\lambda}_{k\,j+1/2}| \left(u^{j+1/2}_{k+1} - u^{j+1/2}_k\right), \quad j \in \mathbb{Z}.$$

The proposed Godunov-type method can then be written in the form (7) and is clearly consistant with (3) (keep in mind that we have assumed that the approximate Riemann solver is consistent in the integral sense to get this result).

Let us now focus on the validity of an entropy inequality similar to (11). For that, we assume that the approximate Riemann solver under consideration is consistent with the integral form of the entropy inequality (9), that is (22)' is valid:

$$G(u_R) - G(u_L) \leq \sum_{k=1}^{\ell} \tilde{\lambda}_k \left(S(u_{k+1}) - S(u_k)\right).$$

Let us first recall that

$$u_j^{n+1} = \frac{1}{2}\left(u^+(u_{j-1}^n, u_j^n) + u^-(u_j^n, u_{j+1}^n)\right)$$

so that by a convexity argument we have

$$S(u_j^{n+1}) \le \frac{1}{2}\left(S(u^+(u_{j-1}^n, u_j^n)) + S(u^-(u_j^n, u_{j+1}^n)) \right).$$

The next calculations now follow the same lines as before. Namely, we first note that using the form (19) of the approximate Riemann solver we can check that (using the Jensen inequality)

$$S(u^-(u_L,u_R)) \le \frac{2}{\Delta x}\int_0^{\frac{\Delta x}{2}} \tilde{u}_{RS}\left(\frac{x}{\Delta t}, u_L, u_R\right)dx = S(u_L) - \frac{2\Delta t}{\Delta x}\sum_{k=1}^{\ell} \tilde{\lambda}_k^- \left(S(u_{k+1}) - S(u_k)\right),$$

$$S(u^+(u_L,u_R)) \le \frac{2}{\Delta x}\int_{\frac{\Delta x}{2}}^{\Delta x} \tilde{u}_{RS}\left(\frac{x}{\Delta t}, u_L, u_R\right)dx = S(u_R) - \frac{2\Delta t}{\Delta x}\sum_{k=1}^{\ell} \tilde{\lambda}_k^+ \left(S(u_{k+1}) - S(u_k)\right),$$

so that with clear notations

$$S(u_j^{n+1}) \le S(u_j^n) - \frac{\Delta t}{\Delta x}\left\{ \sum_{k=1}^{\ell} \tilde{\lambda}^-_{k\,j+1/2}\left(S(u_{k+1}^{j+1/2}) - S(u_k^{j+1/2})\right) + \sum_{k=1}^{\ell} \tilde{\lambda}^+_{k\,j-1/2}\left(S(u_{k+1}^{j-1/2}) - S(u_k^{j-1/2})\right) \right\}$$

But by consistency with the entropy inequality we have

$$\sum_{k=1}^{\ell} \tilde{\lambda}^-_{k\,j+1/2}\left(S(u_{k+1}^{j+1/2}) - S(u_k^{j+1/2})\right) = \frac{1}{2}\sum_{k=1}^{\ell}\left(\tilde{\lambda}_{k\,j+1/2} - |\tilde{\lambda}_{k\,j+1/2}|\right)\left(S(u_{k+1}^{j+1/2}) - S(u_k^{j+1/2})\right)$$

$$\ge \frac{1}{2}\left(G(u_{j+1}) - G(u_j)\right) - \frac{1}{2}\sum_{k=1}^{\ell} |\tilde{\lambda}_{k\,j+1/2}|\left(S(u_{k+1}^{j+1/2}) - S(u_k^{j+1/2})\right)$$

and

$$\sum_{k=1}^{\ell} \tilde{\lambda}^+_{k\,j+1/2}\left(S(u_{k+1}^{j-1/2}) - S(u_k^{j+1/2})\right) \ge$$

$$\frac{1}{2}\left(G(u_j) - G(u_{j-1})\right) + \frac{1}{2}\sum_{k=1}^{\ell} |\tilde{\lambda}_{k\,j+1/2}|\left(S(u_{k+1}^{j+1/2}) - S(u_k^{j+1/2})\right).$$

This means that we eventually have

$$S(u_j^{n+1}) \le S(u_j^n) + \frac{\Delta t}{\Delta x}\left(G(u_j^n, u_{j+1}^n) - G(u_{j-1}^n, u_j^n)\right), \quad j \in \mathbb{Z}$$

with

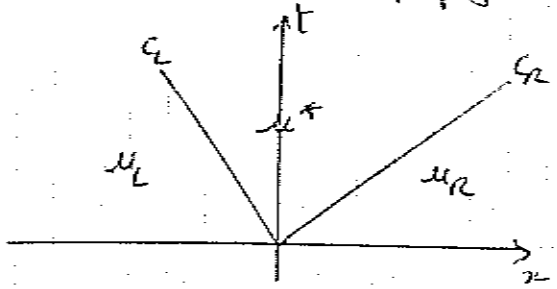
$$G_{j+1/2} = G(u_j^n, u_{j+1}^n) = \frac{1}{2}\left(G(u_j^n) + G(u_{j+1}^n)\right) - \frac{1}{2}\sum_{k=1}^{\ell} |\tilde{\lambda}_{k\,j+1/2}|\left(S(u_{k+1}^{j+1/2}) - S(u_k^{j+1/2})\right), \quad j \in \mathbb{Z}.$$

The proposed Godunov method is then consistent with (9) provided that the underlying approximate Riemann solver is consistent with the integral form of (9).

Examples of approximate Riemann Solvers

The HLL approximate Riemann Solver

One of the most simple and popular approximate Riemann solver consists in considering only two discontinuities propagating with speeds c_L and c_R:



so that only one intermediate state u^* has to be calculated. For that we use the consistency with the integral form of (3), namely (2.1) or equivalently (2.1)', that writes here

$$f(u_R) - f(u_L) = c_L (u^* - u_L) + c_R (u_R - u^*) .$$

Then we have

$$u^* = \frac{c_R u_R - c_L u_L}{c_R - c_L} - \frac{1}{c_R - c_L} \left(f(u_R) - f(u_L) \right) .$$

Note that we have implicitly assumed $c_R - c_L > 0$.

We shall not give the details here, but it can be proved that this approximate Riemann solver is also consistent with the entropy inequality (9) provided that the so-called "subcharacteristic condition"

$$c_L \leq \lambda_k(u) \leq c_R , \quad k = 1, \dots, p$$

is valid. It means that the eigenvalues of the system to be solved lie between the eigenvalues of the approximate Riemann solver. In other words, the information propagates faster in the approximate Riemann solver.

A local optimization of c_L is

$$c_L = \inf_{u = u_L, u_R} \inf_k \lambda_k(u) \quad \text{and} \quad c_R = \sup_{u = u_L, u_R} \sup_k \lambda_k(u) .$$

In practice, the corresponding scheme rather diffusive, but less than the next one.

The Lax-Friedrichs approximate Riemann solver

It consists in choosing $c_L = -c_R \equiv -c$, with

$$|\lambda_k(u)| \leq c \quad , \quad k = 1, \dots, p .$$

Again, stability properties can be proved in this context.

The Rusanov approximate Riemann solver

Here we simply take

$$c = \sup_{u = u_L, u_R} \sup_k |\lambda_k(u)| .$$

This works well in practice, except the excessive numerical diffusion on the waves associated with intermediate eigenvalues.

The Roe approximate Riemann solver

The idea of the Roe approximate solver is to exactly solve an linearized problem, associated with a so-called Roe linearization $A(u_L, u_R)$. More precisely, the Roe approximate Riemann solver is defined by the exact Riemann solution of the linear system

$$\begin{cases} \partial_t u + A(u_L, u_R)\, \partial_x u = 0 \\ u(x,0) = u_0(x) = \begin{cases} u_L & \text{if } x < 0 , \\ u_R & \text{if } x > 0 . \end{cases} \end{cases}$$

The matrix $A(u_L, u_R)$ is called a Roe-type linearization if the mapping $(u_L, u_R) \to A(u_L, u_R)$ from $\Omega \times \Omega$ into $\mathbb{R}^{p \times p}$ satisfies the following properties.

(i) $f(u_R) - f(u_L) = A(u_L, u_R)(u_R - u_L)$ (conservativity)

(ii) $A(u_L, u_R)$ is $\mathbb{R}$-diagonalizable (hyperbolicity)

(iii) $A(u, u) = \nabla_u f(u)$. (consistency)

In practice, (i) is the most difficult property to satisfy.

Let us now prove the existence of a Roe-type linearization when the system under consideration admits a strictly convex entropy.

Theorem (Harten-Lax, 1983)
Assume that (1) admits a strictly convex entropy S. Then there exists at least one Roe-type linearization.

Proof
The Roe-type linearization will be of the form

$$A(u_L, u_R) = R(u_L, u_R)\, P(u_L, u_R)$$

where R and P are symmetric matrices, P being in addition positive-definite. Then, $A = P^{-1/2} P^{1/2} R P^{1/2} P^{1/2}$ is similar to the matrix $P^{1/2} R P^{1/2}$ which is clearly symmetric, and thus $\mathbb{R}$-diagonalizable. Then, A is $\mathbb{R}$-diagonalizable as well, and property (ii) is immediately satisfied. Properties (i) and (iii) will follow from the definitions of R and P.
S being a strictly convex function, the change of variables $v(u) = \nabla S(u)$ is admissible and we can consider the function

$$g(v) = f(u(v))$$

Let us prove that $g'(v) = \nabla_u f(u(v)) \cdot u'(v)$ is symmetric. Actually, since $u'(v) = v'(u)^{-1} = \nabla^2 S(u)^{-1}$, it amounts to prove that the matrix $\nabla_u f(u(v)) \cdot \nabla^2 S(u)^{-1} = \nabla^2 S(u)^{-1} \left(\nabla^2 S(u)\, \nabla_u f(u)\right) \nabla^2 S(u)^{-1}$ is symmetric. Which clearly holds true since the similar matrix $\nabla^2 S(u)\, \nabla_u f(u)$ is symmetric by the strict convexity property of S.
For any given states u_L, u_R, let us set $v_L = \nabla S(u_L)$, $v_R = \nabla S(u_R)$ and write

$$g(v_R) - g(v_L) = \int_0^1 \frac{d}{d\theta}\, g(\theta v_R + (1-\theta) v_L)\, d\theta$$

$$= \left\{ \int_0^1 g'(\theta v_R + (1-\theta) v_L)\, d\theta \right\} \cdot (v_R - v_L)$$

We propose to set

$$R(u_L, u_R) = \int_0^1 g'(\theta v_R + (1-\theta) v_L)\, d\theta$$

R is then symmetric, and satisfies

$$g(v_R) - g(v_L) = R(u_L, u_R)\,(v_R - v_L)$$

On the other hand

$$v_R - v_L = \int_0^1 \frac{d}{d\theta} v(\theta u_R + (1-\theta) u_L)\, d\theta$$

$$= \left\{ \int_0^1 v'(\theta u_R + (1-\theta) u_L)\, d\theta \right\} \cdot (u_R - u_L).$$

We then naturally propose to set

$$P(u_L, u_R) = \int_0^1 v'(\theta u_R + (1-\theta) u_L)\, d\theta$$

so that, since $v'(u) = \nabla S(u)$, P is clearly symmetric and positive definite, while we clearly have

$$A(u_L, u_R)(u_R - u_L) = R(u_L, u_R) P(u_L, u_R)(u_R - u_L)$$

$$= R(u_L, u_R)(v_R - v_L)$$

$$= g(v_R) - g(v_L)$$

$$= f(u_R) - f(u_L)$$

Then property (i) is satisfied.

It remains to check (ii):

$$A(u, u) = R(u, u)\, P(u, u)$$

$$= g'(v)\, v'(u)$$

$$= f'(u),$$

which concludes the proof. ∎

Remark

It is easily checked that $P(u_L, u_R) = P(u_R, u_L)$ and $R(u_L, u_R) = R(u_R, u_L)$, so that we can reverse the roles of u_L and u_R ($A(u_L, u_R) = A(u_R, u_L)$) in the

Roe-type linearization proposed in the proof above.

Now that the Roe approximate Riemann solver can be defined, let us check that it gives rise to a simple approximate Riemann solver that is consistent with the integral form of (1).

First of all, we already saw that solving (33) gives rise to a simple approximate Riemann solver of the form (19), where $\tilde{\lambda}_k$, $k=1,\dots,\ell=p$, coincide with the eigenvalues of $A(u_L,u_R)$, and where the intermediate states u_k are given by

$$u_k = \sum_{\ell=1}^{k-1} (u_R, l_\ell)\, r_\ell + \sum_{\ell=k}^{p} (u_L, l_\ell)\, r_\ell$$

where r_ℓ and l_ℓ represent the right and left eigenvectors of the eigenvalues $\tilde{\lambda}_\ell$.

Then, remind that

$$\begin{aligned} u_{k+1} - u_k &= (u_k, l_k)\, r_k - (u_L, l_k)\, r_k \\ &= \{ (u_R, l_k) - (u_L, l_k) \}\, r_k \end{aligned}$$

so that

$$A(u_L,u_R)\,(u_{k+1}-u_k) = \tilde{\lambda}_k\,(u_{k+1}-u_k)\,.$$

Then we successively obtain

$$\begin{aligned} f(u_R) - f(u_L) &= A(u_L,u_R)\,(u_R - u_L) && \text{(by (1))} \\ &= \sum_{k=1}^{p} A(u_L,u_R)\,(u_{k+1}-u_k) \\ &= \sum_{k=1}^{p} \tilde{\lambda}_k\,(u_{k+1}-u_k) \end{aligned}$$

which is nothing but the consistency property (21)' with the integral form of (1).

Then, the Roe approximate Riemann solver, with a Roe-type linearization $A(u_L,u_R)$ satisfying (i), (ii), (iii) above, gives rise to

an approximate Riemann solver which is consistent with the integral form of (1). The numerical flux function associated with the corresponding Godunov-type method is then, here again, given by (26), that is

$$f(u_L,u_R) = \frac{1}{2}\left(f(u_L)+f(u_R)\right) - \frac{1}{2}\sum_{k=1}^{p} |\tilde{\lambda}_k|\,(u_{k+1}-u_k)\,.$$

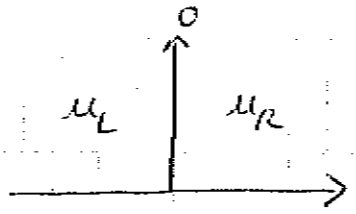
A well-know drawback of this Roe's scheme is that it is not consistant with the entropy inequality (9), that is the consistency relation (22)', that is

$$G(u_R) - G(u_L) \le \sum_{k=1}^{p} \tilde{\lambda}_k \left(S(u_{k+1}) - S(u_k)\right)$$

is not satisfied in general. To be convinced, it suffices to consider u_L and u_R such that $f(u_L) = f(u_R)$ and $G(u_L) \neq G(u_R)$. Then, $A(u_L,u_R)(u_R-u_L) = 0$ which means that 0 is an eigenvalue of $A(u_L,u_R)$ and that (u_R-u_L) is a corresponding eigenvector. Then

$$u_{k+1}-u_k = \left\{(u_R,l_k) - (u_L,l_k)\right\} r_k$$

$$= (u_R-u_L,\, l_k)\, r_k$$



equals 0 except across the discontinuity $\tilde{\lambda}_k = 0$, so that we immediately get in that case $\sum_{k=1}^{p} \tilde{\lambda}_k \left(S(u_{k+1}) - S(u_k)\right) = 0$. Then, the consistency relation (22)' would write

$$G(u_R) - G(u_L) \le 0\,.$$

This cannot be true since we can reverse the role of u_L and u_R in our argument just above.

This means that the Roe scheme may resolve nonphysical solutions. Various entropy corrections have been proposed, for instance by Harten and Hyman in 1983, Roe in 1983, Roe and Pike in 1983, Huynh in 1991, ...

Another drawback of Roe's scheme is that it may generate in some particular test cases nonphysical intermediate states in the approximate Riemann solution, for instance with negative internal energies or densities in the frame of the Euler equations. A recent way to overcome this difficulty, as well as the one associated with the entropy inequality consistency, is given by the so-called Relaxation approach, that we now address. Note however that besides these drawbacks, the Roe scheme is again widely used in the industry.

The Relaxation approach