Features of numerical resolution

Roch SMETS,

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

First attempt

Simple ODE for the function $y(t)$: $\mathrm{d}_t y = -\alpha y$
with $\alpha \in \mathbb{R}^+$ and $y(t=0) = y_0$.

Linear first order ODE has the simple analytical solution
$y(t) = y_0 \mathrm{e}^{-\alpha t}$.

Discret time steps $t^n = n\Delta t$

finite difference $\Delta t \mathrm{d}_t y(t^n) \sim y^{n+1} - y^n$, hence the scheme

$$y^{n+1} = y^n - \alpha y^n \Delta t \tag{1}$$

Explicit Euler scheme : y is approximated by a piecewise linear function

Definition : This solution is stable if $|y^n| < \infty$ for $n \to \infty$.
Definition : This solution is positive if $y^n > 0, \forall n$.

First attempt

- the criterion for **stability** is $\alpha \Delta t < 2$.
- the criterion for **positivity** is $\alpha \Delta t < 1$.

Remark : Using this scheme, accurate results require a very small time-step value. This strong constraint can be avoided by using an higher order scheme, a scheme with adjustment of the time step, or a scheme specially adapted to this ODE.

While apparently approximative, the foundation of this approach will be established in a more robust way in the next section when defining the order of a scheme.

Parabolic PDE

Heat equation : $\partial_t u = D\partial^2_{x^2} u$
with $D \in \mathbb{R}^+$.

Remark : The positivity of D is very important ; a negative diffusion coefficient would mean anti-diffusion. Hence, instead of the regularizing effect of the diffusion operator, anti-diffusion would increase in a dramatic way any gradients whatever its smallness.

Notation : $u_j^n = u(j\Delta x, n\Delta t)$.

$\Delta t \partial_t u(x_j, t^n) \sim u_j^{n+1} - u_j^n$
$\Delta x^2 \partial^2_{x^2} u(x_j, t^n) \sim u_{j+1}^n - 2u_j^n + u_{j-1}^n$

$$u_j^{n+1} = u_j^n + \eta(u_{j+1}^n - 2u_j^n + u_{j-1}^n) \tag{2}$$

New (dimensionless) unknown $\eta = D\Delta t/\Delta x^2$

- stability for $\eta < \frac{1}{2}$
- positivity for $\eta < \frac{1}{4}$

Parabolic PDE

Such a diffusion equation has the clear effect of "smoothing out" all kinds of structures. To say it more roughly, a diffusion operator on a mathematical function acts as "sun on butter".

When the stability criterion is satisfied, all small-scale structures will be wiped out, the stiffest the gradients, the faster their smoothing. In the other cases, when the stability criteron is not satisfied, spurious oscillations appear and grow with time.

Remark : The definition of η clearly shows that even if the Euler scheme is simple, the constraint on its stability is very expensive as the time step depends on the square of the grid size. Hence, a good spatial resolution will only be obtained at the expense of an expensive computation. Fortunately, some better schemes exist... but will not be detailed in this introductory course.

Linear advection equation

Most simple hyperbolic equation : $\partial_t u + A\partial_x u = 0$ with $A = \text{const}$.

With a centered finite difference

$$2\Delta x \partial_x u(x_j, t^n) \sim u_{j+1}^n - u_{j-1}^n \tag{3}$$

Notation : New parameter $\nu = \Delta t / \Delta x$

Explicit **centered Euler scheme** : unconditionally unstable !

$$w_j^{n+1} = w_j^n - \frac{1}{2} A\nu [w_{j+1}^n - w_{j-1}^n] \tag{4}$$

Notation : The numerical solution of a finite difference scheme is not a discretization of the exact solution u of the differential problem. We hence call w the exact solution of the discret problem (for a given scheme). That is $w_j^n \not\equiv u(x_j, t^n)$.

Taylor series for derivative approximation

Quite simple procedure, usually called "finite difference method".
$\rightarrow$ approximates the derivatives of the PDE by finite differences using a Taylor serie.

For a function $u(x)$ discretized on $x_j = j\Delta x$,

$$u_{j+1} = u_j + \Delta x u'_j + \frac{\Delta x^2}{2} u''_j + O(\Delta x^3) \tag{5}$$

$\rightarrow$ one straightforwardly obtains

$$u'_j = \frac{u_{j+1} - u_j}{\Delta x} + O(\Delta x) \tag{6}$$

The remainders coming from the Taylos serie in the O function.

Definition : The finite difference given by Eq. (7) is of order 1 because this is the power of Δx in the O function.

Taylor series for derivative approximation

Remark : The development at u_{j-1} around u_j would give

$$u'_j = \frac{u_j - u_{j-1}}{\Delta x} + O(\Delta x) \tag{7}$$

From Eq. (5), we could also obtain a finite difference expression of order 2, by cancelling the u''_j term. For that purpose, we would write the same expression as Eq. (5) at u_{j-1}, and by differenciation, we would then obtain

$$u'_j = \frac{u_{j+1} - u_{j-1}}{\Delta x} + O(\Delta x^2) \tag{8}$$

The procedure is to use Taylor serie(s) at appropriate location(s) and their combination in order to cancel-out low order terms.
$\rightarrow$ one can then "decide" which point is involved in the scheme.
As an illustration,

$$u'_{j+1} = \frac{u_{j-1} - 4u_j + 3u_{j+1}}{2\Delta x} + O(\Delta x^2) \tag{9}$$

Polynomial interpolation/extrapolation

Numerical solution of PDE : sampled values that approximate at a given location (on the set of grid points) the exact solution
$\rightarrow$ with these sampled values, one have access to a functional representation.

Definition : Any function created from samples is called a "reconstruction".

Definition : Any reconstruction passing through the sample points is called "interpolation" inside the domain of the samples.

Definition : Any reconstruction passing through the sample points is called "extrapolation" when it is used outside of the domain of the samples.

Polynomial for reconstruction function

There is a unique N^{th}-order polynomial passing through any set of $N+1$ samples.

Such polynomial can be :

- easy to reconstruct (e.g. the Lagrange form or the Newton form), but not that easy to manipulate (to derive for example).
- difficult to reconstruct (e.g. the Taylor series form) but then much simpler to use.

While not giving details on how to reconstruct the coefficients a_i in a Taylor serie, we remind its form.

$$p_N(x) = a_0 + a_1(x-b) + \cdots + a_N(x-b)^N \tag{10}$$

Discret problem

The linear advection equation can be approximated by Eq. (4).

Notation : We note $D(u) = 0$ the general differential problem and $\Xi(w) = 0$ the associated discret problem.

$\rightarrow$ D is a **differential form**,
$\rightarrow$ Ξ is a **finite difference scheme**,
$\rightarrow$ u is the excat solution of the differential problem
$\rightarrow$ w is the exact solution of the discrete problem

For the scheme given by Eq. (4), the Ξ operator acting on the $u(x,t)$ function is then

$$\Xi(u) = \frac{u_j^{n+1} - u_j^n}{\Delta t} + a\frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} \tag{11}$$

with $u_j^n = u(x_j, t^n) = u(j\Delta x, n\Delta t)$.

Truncation error

All the terms in the right-hand-side of Eq. (11) can be written using a Taylor expansion around u_j^n. Using a Taylor-Lagrange formulation of the serie expansion, we have

$$u_j^{n+1} = u_j^n + \Delta t \partial_t u(x_j, t^n) + \frac{\Delta t^2}{2} \partial_{t^2}^2 u(x_j, t^\star) \tag{12}$$

with $t^\star \in [t^n, t^{n+1}]$. This expansion is said to be at first order because the first order term is the last one that is explicited ($t^\star$ is not equivocally defined). In the same spirit, for the $u_{j\pm1}^n$ terms at second order,

$$u_{j\pm1}^n = u_j^n \pm \Delta x \partial_x u(x_j, t^n) + \frac{\Delta x^2}{2} \partial_{x^2}^2 u(x_j, t^n) \pm \frac{\Delta x^3}{6} \partial_{x^3}^3 u(x_\dagger, t^n) \tag{13}$$

with $x_\dagger \in [x_j, x_{j\pm1}]$.

Truncation error

From Eq. (11), (12) and (13), we have

$$\begin{aligned}\Xi[u(x_j,t^n)] &= \partial_t u(x_j,t^n) + \frac{\Delta t}{2}\partial^2_{t^2} u(x_j,t^\star) \\ &\quad + a\left[\partial_x u(x_j,t^n) + \frac{\Delta x^2}{3}\partial^3_{x^3} u(x_\dagger,t^n)\right] \\ &= [\partial_t u|^n_j + a\partial_x u|^n_j] \\ &\quad + \frac{\Delta t}{2}\partial^2_{t^2} u(x_j,t^\star) + \frac{a\Delta x^2}{3}\partial^3_{x^3} u(x_\dagger,t^n)\end{aligned}$$

The first bracket in the right-hand-side is null as it satisfes the discrete problem $D(u) = 0$. This equation then writes

$$\Xi(u) = D(u) + T(x,t) \tag{14}$$

Definition : In Eq. (14), $T(x,t)$ is the **truncation error**

Truncation error

Because we have $D(u) = 0$, the truncation error writes

$$T(x,t) = \frac{\Delta t}{2}\partial^2_{t^2} u(x_j, t^\star) + \frac{a\Delta x^2}{3}\partial^3_{x^3} u(x_\dagger, t^n) \tag{15}$$

If we now consider that $u(x,t)$ is C^2 for t and C^3 for x, then the derivative in Eq. (15) can be bounded.

$\rightarrow$ we note M_{tt} the upper bound for $|\partial^2_{t^2} u(x,t)|$
$\rightarrow$ we note M_{xxx} the upper bound for $|\partial^3_{x^3} u(x,t)|$

$$T(x,t) \leq \frac{1}{2}M_{tt}\Delta t + \frac{a}{3}M_{xxx}\Delta x^2 \tag{16}$$

In the expression of $T(x,t)$, Δt and Δx are as small as we want

$\rightarrow$ for a given problem and a given scheme, we can (in principle) make $T(x,t)$ as small as we want.

Scheme order

Definition : Given a numerical scheme, if the truncation error writes $T(x,t) = A\Delta t^N + B\Delta x^M$ with $(A,B) \in \mathbb{R}^2$, the scheme is said to be of order N in time and M in space.

Remark : In Eq. (16), the terms that are explicited are the leading one because Δt and Δx being small, higher power of these terms are even smaller.

Note : The order of a scheme is of course an integer, larger or equal to 1.

Modified equation

We solve the discrete problem $\Xi(u) = 0$
$\rightarrow$ because of Eq. (14), we then have

$$D(u) = -T(x,t) \tag{17}$$

If we only keep the leading term in the truncation error (the first one in Eq. (15)), u being a solution of $D(u) = 0$, we have $\partial^2_{t^2} u = A^2 \partial^2_{x^2} u$ so the truncation error writes

$$T(x,t) = \frac{A^2 \Delta t}{2} \partial^2_{x^2} u(x,t) \tag{18}$$

As a result, the scheme $\Xi(w) = 0$ does not provide a solution of the initial differential problem $D(u) = 0$, but rather gives an approximate solution of the modified equation $D(u) = -T(x,t)$.

Definition : The modified equation of a numerical scheme is the differential equation that takes into account the (lowest order) terms of the truncation error.

Numerical diffusion

For the linear advection problem, the modified equation is then

$$\partial_t u + A\partial_x u = -\frac{A^2 \Delta t}{2} \partial^2_{x^2} u(x,t) \tag{19}$$

Remark : The right-hand side of Eq. (19) is a diffusion term with a diffusion coefficient $-A^2 \frac{\Delta t}{2}$
$\rightarrow$ this term being negative, it gives rise to anti-diffusion
$\rightarrow$ any fluctuations are growing with time
$\rightarrow$ this scheme is unstable as we already saw

Definition : A scheme is said to be diffusive/anti-diffusive if the modified equation contains a diffusive/anti-diffusive term.

Better schemes generally have a diffusion term.
$\rightarrow$ this coefficient is generally a function of Δt, Δx and A
$\rightarrow$ while generally small, this is a diffusive term
$\rightarrow$ it is important to find a scheme that minimizes or counter-balance its growth.

Numerical dispersion

Definition : When the modified equation of a numerical scheme is of the form

$$\partial_t w + A\partial_x w = \mu \partial^3_{x^3} w \tag{20}$$

the scheme is said to be dispersive.

The Fourier transform (both in space and time) of Eq. (20)

$$\widetilde{w}(k,\omega) = \int_{\mathbb{R}} w(x,t) \mathrm{e}^{\imath(kx-\omega t)} \mathrm{d}x\mathrm{d}t \tag{21}$$

has a solution satisfying $\omega = Ak + \mu k^3$.

The structures of w propagate at a phase speed $\omega/k = A + \mu k^2$
$\rightarrow$ all the Fourier modes propagate at a speed depending on k
$\rightarrow$ this is dispersion : the form of a wave packet will not be preserved with time

Remark : In most cases, the modified equation of a given scheme contains both diffusion and dispersion.

Stability, consistency and convergence

Definition : We call Ω_j the set of j-indices so that $x_j \in \mathcal{D}$. We call $\mathcal{D}$ the definition domain (in space) of the function $u(x,t)$.

There are three fundamental properties associated to the numerical resolution of a PDE.

Definition : A numerical scheme is stable if $|w_j^n| < \infty$, $\forall j \in \mathcal{D}_j$ and $t \to \infty$.

Definition : A numerical scheme is consistent if $D \to \Xi$ for $\Delta x \to 0$ and $\Delta t \to 0$.

Definition : A numerical scheme converges if $|u_j^n - w_j^n| \to 0$, $\forall j \in \mathcal{D}_j$ and $t \to \infty$.

Stability, consistency and convergence

The wide range of schemes (only some of them are discussed in this introduction) are all consistent.
$\rightarrow$ it is so because they were defined with the constraint that the associated truncation error are at least of order 1 in both space and time.

The stability is the minimal condition that one can expect from a numerical scheme.
$\rightarrow$ Obviously, if its solution w_j^n is diverging, it is also diverging from the exact solution u of the differential problem $D(u) = 0$.

Definition : The **error** of a scheme is defined at each grid point (and each time step) as $e_j^n = |u_j^n - w_j^n|$.

A good scheme is hence a scheme that converges.

Stability, consistency and convergence

A convergent scheme is a scheme for which the error goes to zero, meaning that the discrete solution w_j^n is a good approximation of the exact solution $u(x_j, t^n)$.

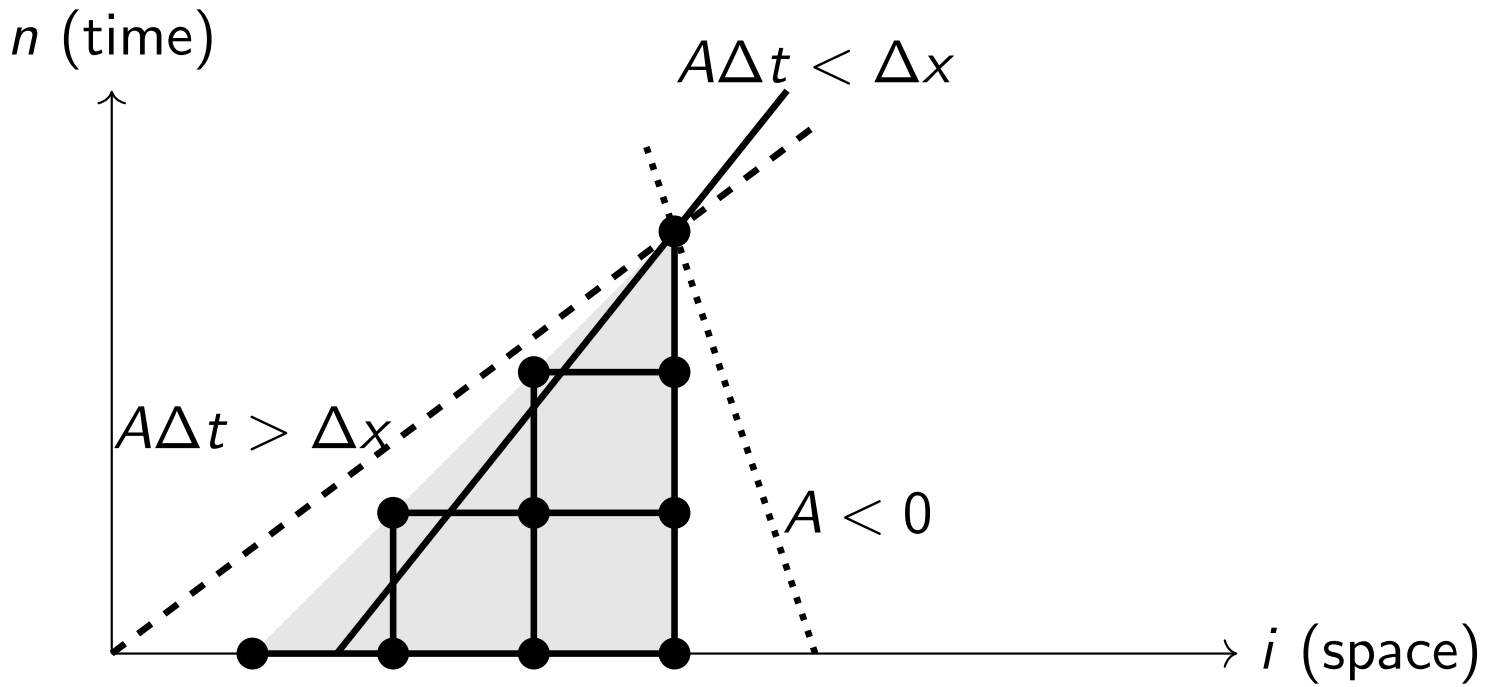
The way we generally characterize the error of a scheme will be given in the next subsection of this chapter. While the convergence of a scheme is not always easy to demonstrate, the Lax equivalence theorem is generally at the heart of demonstrating the convergence of a scheme.

Theorem : (Lax equivalence theorem). Given a well-posed initial value problem and a finite-difference approximation, that satisfies the consistency condition, stability is the necessary and sufficient condition for convergence.

CFL condition

Because of the finite speed of waves, hyperbolic PDEs have a finite physical domain of dependence.

For the simple advection problem with the initial condition
$u(x, t = 0) = u_0(x)$
$\rightarrow$ the solution is $u(x, t) = u_0(x - At)$
$\rightarrow u(x, t)$ is const. along a caracteristic, which equation is $d_t x = A$



CFL condition

Consider a non-centered scheme of the form

$$w_j^{n+1} = \gamma w_{j-1}^n + \beta w_j^n \tag{22}$$

for $(\gamma, \beta) \in \mathbb{R}^2$, meaning that w_j^{n+1} depends on w_{j-1}^n and w_j^n.
$\rightarrow$ in the same way, this set of points (w_{j-1}^n, w_j^n) depends on the three points $(w_{j-2}^{n-1}, w_{j-1}^{n-1}, w_j^{n-1})$.

Going up to $n = 0$ we then define the numerical domain of dependance, represented in gray in the previous Fig.

Because the solution is in this case $u(x, t) = u_0(x - At)$, it is clear that the stability of a scheme cannot be guaranted if the characteristic does not lie in the numerical domain of dependence.

This is the CFL condition, from the paper by [Courant, 1928].

CFL condition

In the Fig., the two caracteristics in dashed line (for a larger A value) and dotted line (for negative A value) do not satisfy the CFL condition :

- for the dashed line, Δx is too small or Δt too large, so that $a\Delta t > \Delta x$
- for the dotted line, A is negative. To satisfy the CFL condition, we should better have a scheme of the form $w_j^{n+1} = \gamma' w_{j+1}^n + \beta' w_j^n$

Property : To satisfy the CFL condition, the full numerical domain of dependence must contain the physical domain of dependence

Remark : The CFL condition is a necessary, but insufficient condition, for the stability of a scheme

Property : The CFL condition is not $A\Delta t < \Delta x$ in the general case. For example, for a scheme with the five points at $j-2$, $j-1$, j, $j+1$ and $j+2$, the CFL condition is $A\Delta t < 2\Delta x$

CFL condition

Definition : Using the notation already introduced $\nu = \dfrac{\Delta t}{\Delta x}$, the **CFL number** is $A\nu$.

Because of the structure and properties of the advection equation, the value $u(x_j, t^{n+1})$ is inherited from its "first" ancestors, that is the ones defined at t^n. If all the ancestors are in a given range of values, then $u(x_j, t^{n+1})$ also lie in this range. This results in an important property that is to be satisfied by a scheme

Property : The **upwind range condition** states that

$$\min_{x_{j-1} \leq x \leq x_j} u(x, t^n) \leq u(x_j, t^{n+1}) \leq \max_{x_{j-1} \leq x \leq x_j} u(x, t^n) \text{ for } 0 \leq \eta A_j^{n+1} \leq 1$$

$$\min_{x_j \leq x \leq x_{j+1}} u(x, t^n) \leq u(x_j, t^{n+1}) \leq \max_{x_j \leq x \leq x_{j+1}} u(x, t^n) \text{ for } -1 \leq \eta A_j^{n+1} \leq 0$$

Stability analysis

If we ignore the effects of the boundary conditions, we can investigate the stability of a scheme by Fourier analysis
$\rightarrow$ for that purpose, we use the Fourier modes

$$w_j^n = (\lambda)^n \mathrm{e}^{\imath(kj\Delta x)} \tag{23}$$

We did not demonstrate that this form is a solution
$\rightarrow$ but this form is convenient, essentially because $w_j^{n+1} = \lambda w_j^n$
$\rightarrow$ a stability criterion is clearly $|\lambda| < 1$

For the scheme given by Eq. (4), we straightforwardly obtain

$$\lambda(k) = 1 - \imath A\nu \sin k\Delta x \tag{24}$$

then, $|\lambda| > 1$, for all mesh ratio and almost all modes
$\rightarrow$ this scheme is always unstable
$\rightarrow$ the numerical solution w_j^n will unboundly grow with time

Stability analysis

We saw that the scheme given by Eq. (4) is definitively very bad for the linear advection equation.
$\rightarrow$ one introduce a new one, with much better performance while very simple : the upwind scheme

We use two points in order to discretize the spatial derivative
$\rightarrow$ to satisfy the CFL condition, these points depend on the sign of A

$$w_j^{n+1} = \begin{cases} w_j^n - A\nu(w_{j+1}^n - w_j^n) & \text{for } A < 0 \\ w_j^n - A\nu(w_j^n - w_{j-1}^n) & \text{for } A > 0 \end{cases} \tag{25}$$

Remark : This scheme was proposed in the paper [Courant, 1952] and can also be called the CIR method from the names of its authors

Stability analysis

The upwind scheme can be written in a single line using the quantities $\frac{1}{2}(|A| - A)$ and $\frac{1}{2}(|A| + A)$.

The name of this scheme is straightforward, as it uses the points in the direction from which the wind (the wave speed) is blowing.

On can then easily makes the stability analysis of the upwind scheme and then obtain for the amplification coefficient.

$$\lambda(k) = 1 - 2A\nu(1 - A\nu)(1 - \cos k\Delta x) \tag{26}$$

which gives

$$|\lambda|^2 = 1 - 4A\nu(1 - A\nu)\sin^2 \frac{1}{2}k\Delta x \tag{27}$$

It then follows that $|\lambda(k)| \leq 1$ for all k, provided that the CFL condition $|A\nu| < 1$ is satisfied.

The upwind scheme is then stable when the CFL condition is satisfied.

Stability analysis

Remark : While the upwind scheme is of order 1, that is smaller than the order of the explicit centered Euler scheme, it is a stable scheme
$\rightarrow$ the order of a scheme is not a guarantee of its stability

In Eq. (23), space and time indices are not treated in the same way
$\rightarrow$ we could have used a Fourier serie in time as we did in space

But the form that is used with λ is enough for stability analysis.

Obviously, $\lambda \in \mathbb{C}$, so that its modulus quantify how instable is the scheme, and its phase quantify its dispersion.

Grid deposition

Finite difference. With the finite difference approach, each quantity involved in the equations to be solved (that is the components of u and the components of the associated flux f) are **discretized** on a structured grid, generally uniform.
$\rightarrow$ with such an approach, we have

$$w_j^n \sim u(j\Delta x, n\Delta t) \tag{28}$$

that is, using words : w_j^n is a discret set of values on a structured uniform grid, which approximates the exact solution (which is a continuous function) sampled on this grid.

Definition : The scheme is said to be implicit if w_j^{n+1} depends at least from another w value also defined at time step $n+1$.

Grid deposition

Finite volume. For this grid deposition, we also consider a spatial discretization on $N+1$ nodes uniformally distributed with a mesh $\Delta x = L/N$
$\rightarrow$ in the same way, the time is discretized with a time-step Δt

For the finite volume approach, the conservative laws are **integrated** on a elementary mesh $[x_{j-1/2}, x_{j+1/2}]$
$\rightarrow$ the integration of the linear advection equation then gives

$$\partial_t \int_{x_{j-1/2}}^{x_{j+1/2}} u \, \mathrm{d}x + \int_{x_{j-1/2}}^{x_{j+1/2}} \partial_x f \, \mathrm{d}x = 0 \tag{29}$$

As a consequence, such approach is dealing with the average value $\overline{u}_j^n$ of the continuous function $u(x,t)$ calculated on the mesh $[x_{j-1/2}, x_{j+1/2}]$ defined as

$$\overline{u}_j^n = \int_{x_{j-1/2}}^{x_{j+1/2}} u(x, n\Delta t) \, \mathrm{d}x \tag{30}$$

Grid deposition

With such a grid deposition, w_j^n is hence an approximation of $\overline{u}_j^n$.

The physical flux has also to be approximated.
$\rightarrow$ because of the structure of Eq. (29) and the spatial derivative in the second integral, this equation simply writes

$$\partial_t w_j^n + \frac{F_{j+1/2}^n - F_{j-1/2}^n}{\Delta x} = 0 \tag{31}$$

F_j^n is the **numerical flux**.
$\rightarrow$ it approximates the **physical flux** $f[u(j\Delta x, n\Delta t)]$

The power of this approach is that the flux are then defined on the interface of a cell which is an important property of the associated conservative law.

Property : In the finite volume approach, two adjacent cells share the same flux : the outgoing flux of a given cell is then exactly the one ingoing in the next adjacent cell.

Grid deposition

Remark : It is then clear that a finite volume scheme is consistent if $\lim_{\Delta t \to 0\ ,\ \Delta x \to 0} F_j^n = f(j\Delta x, n\Delta t)$.

The power of the finite volume approach is also that, because of the integration in a cell, the spatial derivative of the physical flux has disappeared.

Definition : We call **weak formulation** the way to rewrite a PDE in a way that no derivative of the solution appears.

While surprising, PDE can have solution which derivative is not necessarily defined. The finite volume formalism is hence well-suited to handle the weak solution of a PDE when discontinuities are initially imposed or are rising with time integration.

Definition : We call **weak solution** of a PDE the function that satisfies the integral form of the original PDE.

Grid deposition

Remark : The derivative of the weak solution of a PDE is not necessarily defined !
One should remember that in the first chapter, the derivative form of the Euler equations was obtained from the integral form. The integral form was "simply" a balance in a small (but finite) control volume. The derivative form was obtained in the limit of a control volume size going to zero. While it mathematically makes sense, it is physically a nonsense because as already discussed, the control volume has to be large enough so that one can define a macroscopic average value.

Remark : Some schemes are obtained with a grid deposition of finite-difference type. Nevertheless, some of them can also be obtained with a different grid deposition, namely using finite volume. It is then important to keep in mind that in these cases (see for example the classical linear Lax-Wendroff scheme), the discret values w_j^n are not defined in the same way.

Time integration

Whatever finite difference or finite volume approach, w_j^n is an approximate value of the continuous function $u(x,t)$ at $x_j = j\Delta x$ and $t^n = n\Delta t$.

The temporal derivative is generally obtained using an explicite Euler scheme of the form

$$\partial_t w_j^n = \frac{w_j^{n+1} - w_j^n}{\Delta t} \tag{32}$$

The time step Δt has to be defined in order to satisfy the CFL condition (for an explicit scheme).

For that issue, we generally need to calculate the spectral radius of the Jacobian matrix $\partial_u f$, that is the largest of the absolute values of its eigen values.
$\rightarrow$ in codes where the time step is const., it is important to be sure that this condition will be fulfilled during its time integration.

Time integration

The conservative form of any finite volume scheme is then generally of the form

$$w_j^{n+1} = w_j^n + \frac{\Delta t}{\Delta x}[F_{j+1/2}^n - F_{j-1/2}] \tag{33}$$

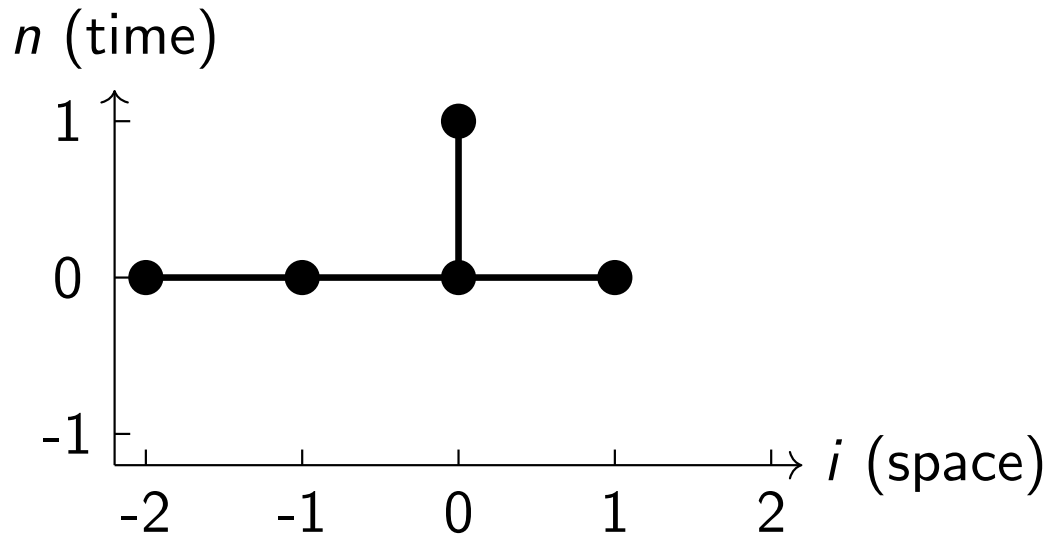
meaning that when the form of the numerical flux is the only one provided, it has to be used in Eq. (33).

In the other cases, the expression of w_j^{n+1} is provided, as well as all the numerical flux needed.

Stencils diagrams

A finite volume scheme generally gives the form of the numerical flux $F^n_{j+1/2}$ as a function (linear or not) of the quantities w^n_j, w^n_{j+1}.... Hence, in Eq. (33), it means that the value of w^{n+1}_j depends on a given set of values of w^n.

This set can then be graphically represented in a stencil diagram as the one given in Fig. below,



Stencils diagrams

In this example, w_j^{n+1} depends on w_{j-2}^n, w_{j-1}^n, w_j^n and w_{j+1}^n. The **width** of this (asymmetric) stencil is 4.

Definition : The **width of a stencil** is the number of points involved at time t^n to calculate w_j^{n+1}.

There is a terminology associated to these stencils that simplifies the naming of several classes. For the spatial discretization,

- **Forward Space** means that the derivative at grid point j uses a set of w_{j+k} points with $k \in \mathbb{N}^+$.
- **Backward Space** means that the derivative at grid point j uses a set of w_{j-k} poionts with $k \in \mathbb{N}^+$.
- **Centered Space** means that the derivative at grid point j uses a difference on indices symetric around j (for example $j+1$ and $j-1$). A centered scheme then uses an odd number of points.

The same terminology is used for the time discretization.

Stencils diagrams

Example : The finite difference scheme given by Eq. (4) is FTCS (Forward Time Centered Space).

Example : The finite difference scheme given by Eq. (25) is FTBS.

A last word on stencils : it appears that in some schemes, the stencil is not constant, but depends on the numerical solution w_j^n.

Definition : An **adaptative sencil** is a stencil wich width and structure depends on the local values of the numerical solution.

Ghost cells

The width of a stencil is important for the appropriate treatment of the boundary conditions.

In the example of the previous Fig., for any points on the left border of the domain, the scheme needs two points that will be out of the domain.
$\rightarrow$ In the same way, on the right border, the scheme will need one point out of the domain.

While these points are needed, but out of the domain, it means that we then need some extra cells, out of the domain, in order to have these discret values of w.

Definition : The cell points out of the domain but needed to apply a stencil are called **ghost cells**.

Ghost cells

For **periodic simulations**, the values of thes w_j^ns in the ghost cells are simply a copy of the appropriate values of the w_j^ns that are inside the domain.

For **non-periodic simulations**, we then need to set some dedicated boundary condition, and then "translate" these physical conditions to deduce the values of the w_j^n in the ghosts cells.

The most classical boundary conditions for a plasma are **perfectly conducting wall** and **dielectric wall**, where one can deduce the values of the electric field, magnetic field, charge density...

Remark : The treatment of boundary conditions can be very touchy... meaning that in many cases, this task is more an art than a science !

Artificial viscosity

As already saw, the first derivative (at second order) can be approximated as

$$\partial_x f|_{x_j} = \frac{1}{2\Delta x}[f(x_{i+1}) - f(x_{j-1})] + O(\Delta x^2) \qquad (34)$$

Such an approximation gives awful results in a finite-difference scheme
$\rightarrow$ the reason is that in Eq. (34), the derivative at odd indices j will only depends on fluxes calculated at even indices (and vice versa).

This lead to a separation between the odd- and even-indices.

Such a phenomenon is usually called "odd-even $2\Delta x$-wave oscillations".
We saw in Eq. (19) that the FTCS scheme exhibits an anti-diffusive term that make it unstable.

Artificial viscosity

As an alternative, Eq. (9) gives the derivative at x_{j+1}.

$$\partial_x f|_{x_{j-1}} = \frac{1}{2\Delta x}[-f(x_{i+1}) + 4f(x_j) - 3f(x_{j-1})] + O(\Delta x^2) \quad (35)$$

that we call the "second order forward difference".
$\rightarrow$ this derivative can be written in a different way as

$$\partial_x f|_{x_{j-1}} = \frac{f(x_j) - f(x_{j-1})}{\Delta x} - \frac{\Delta x}{2}\frac{f(x_{i+1}) - 2f(x_j) + f(x_{j-1})}{\Delta x^2} + O(\Delta x^2) \quad (36)$$

In Eq. (36), the first term approximates a first order derivative (forward difference) and the second term approximates a second order derivative (centered difference).

Such a second order derivative appears in the Navier-Stokes equation, and while the difference with the Euler equation is a viscous term, the second term in Eq. (36) is called artificial viscosity.

Artificial viscosity

Remark : It has to be clear that the artificial viscosisity generally arises from the way a scheme is build, and has absolutely nothing to do with the physical viscosity.

Definition : Any second, fourth, sixth... and other even-order differences in a modified equation are called **artificial viscosity**.

Definition : Any third, fifth... and other odd-order differences in a modified equation are called **artificial dispersion**.

From a general point of view, a viscous-like term in an advection equation add a term on the right hand side of the conservation law as

$$\partial_t u + \partial_x f = \partial_x[\epsilon(u)\partial_x u] \tag{37}$$

with $\epsilon \geq 0$. This last condition is essential. With $\epsilon \leq 0$, such a term is associated to **anti-diffusion**, and drives fluctutions growth at an unacceptable level deadly fast.

Artificial viscosity

A conservative FTCS discretization of Eq. (37) writes

$$\frac{w_j^{n+1} - w_j^n}{\Delta t} + \frac{1}{2\Delta x}[f(w_{j+1}^n) - f(w_{j-1}^n)] = \frac{1}{\Delta x}\left[\epsilon_{j+1/2}^n \frac{w_{j+1}^n - w_j^n}{\Delta x} - \epsilon_{j-1/2}^n \frac{w_j^n - w_{j-1}^n}{\Delta x}\right]$$

Rearanging these terms, the numerical flux (which is used in the conservative form of the advection equation) can be written as the sum of two terms

$$F_{j+1/2}^n = \frac{1}{2}[f(w_{j+1}^n) + f(w_j^n)] - \frac{\epsilon_{j+1/2}^n}{2}(w_{j+1}^n - w_j^n) \qquad (38)$$

Artificial viscosity

- The first term is widely encountered in finite volume method ; the numerical flux at $x_{j+1/2}$ is simply an average of the physical flux at x_j and x_{j+1}.
- The second term clearly contains a first order derivative of w. While in Eq. (31) the numerical flux appears in a first order derivative, this second term is clearly a viscous term associated to a diffusion operator.

The artificial viscosity then appears as a flux correction.

Remark : Eq. (36) can be written as Eq. (38) with $\epsilon_{j+1/2} = -\Delta x/2$.

Artificial viscosity

While the viscosity ϵ_j^n is strictly numerical, some schemes can use this formalism. Hence, the game is to find the proper way to define the value of the artificial viscosity. This one has to be

- as small as possible in smooth region where no smoothing is needed
- just large enough in shocks region in order to prevent a too large steepening of the wave front that could turn the scheme unstable

Then, ϵ_j^n should be a function of differences involving w_j^n's values. This will be discussed later on in the techniques of flux limiters.

The Riemann problem

$\rightarrow$ Solve the advection equation with the initial condition

$$u(x, t = 0) = u_0(x) = \begin{cases} u_L & \text{for } x < 0 \\ u_R & \text{for } x > 0 \end{cases} \tag{39}$$

The Riemann problem has an exact analytical solution for the Euler equations, for any scalar conservation laws, as well as for any linear system of equations.

The system is **self-similar** (or **self-preserving**).

Definition : A PDE depending on x and t coordinates is **self-similar** if its solution only depends on the ratio x/t rather than on x and t separately.

A consequence is that the solution is constant along any lines $x = \eta t$ (with $\eta \in \mathbb{R}$) passing through the origin in the (x, t) plane.

The Riemann problem is hence the simplest test-case for numerical approximation of discontinuities.

The Riemann problem

Property : By self-similarity, numerical methods using Riemann solvers require **only the flux along** $x = 0$.

Even in the cases where the flux function $f(u)$ as well as the initial conditions $u_0(x)$ are C^1, discontinuities can arise meaning that the solution $u(x, t)$ of the problem could eventually not be C^1.
$\rightarrow$ such **weak solutions** describe shocks or discontinuities.

The differential form of the conservation equation is not mathematically adapted, to treat shocks and discontinuities.
$\rightarrow$ integral form : integration over small space-control volume.
$\rightarrow$ closely related to the finite volume formulation.

Theorem : (Lax-Wendroff theorem). If a conservative numerical scheme for a hyperbolic system of conservation laws converges, then it converges towards a **weak solution**.

The jump conditions

Let $u(x,t)$ be the weak solution that is discontinuous along a (regular) curve Γ in the (x,t) plan.

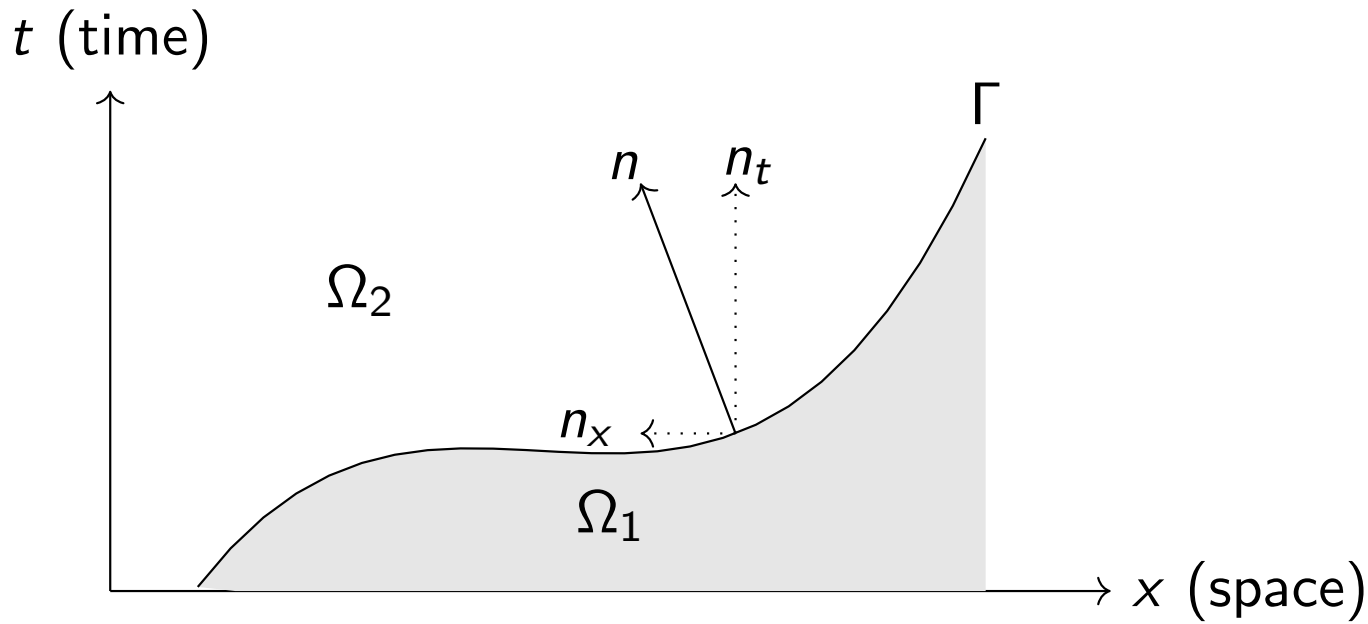
We call Ω_L and Ω_R the two domains separated by Γ where u_L and u_R (the restriction of u to Ω_L and Ω_R, respectively) are regular solutions of the problem.

Let n be the normal to Γ, oriented from Ω_R to Ω_L.

The jump conditions between $u(x,t)$ and its associated flux function $f(u)$ across the Γ curve writes

$$(u_R - u_L)n_t + [f(u_R) - f(u_L)]n_x = 0 \tag{40}$$

The jump conditions



Not a nice formulation :

- define the normal vectors is not that easy
- the physical dimension of its coordinates are not the same

The jump conditions

A way to solve this problem is to write the advection equation in integral form,

$$\int_{x_1}^{x_2} [u(x,t_2) - u(x,t_1)]\mathrm{d}x + \int_{t_1}^{t_2} [f(x_2,t) - f(x_1,t)]\mathrm{d}t = 0 \quad (41)$$

For a schock traveling at speed s, we choose $(x_1,\ t_1)$ and $(x_2,\ t_2)$ along the shock :

$$x_2 - x_1 = s(t_2 - t_1) \quad (42)$$

We introduce the quantities $\Delta t = t_2 - t_1$ and $\Delta x = x_2 - x_1$.

For small enough Δt and Δx,
$\int_{x_1}^{x_2} u(x,t_2)\mathrm{d}x = u_L \Delta x$ and $\int_{t_1}^{t_2} f(x_2,t)\mathrm{d}t = f_R \Delta t$.
One can do the same at t_1 and x_1.

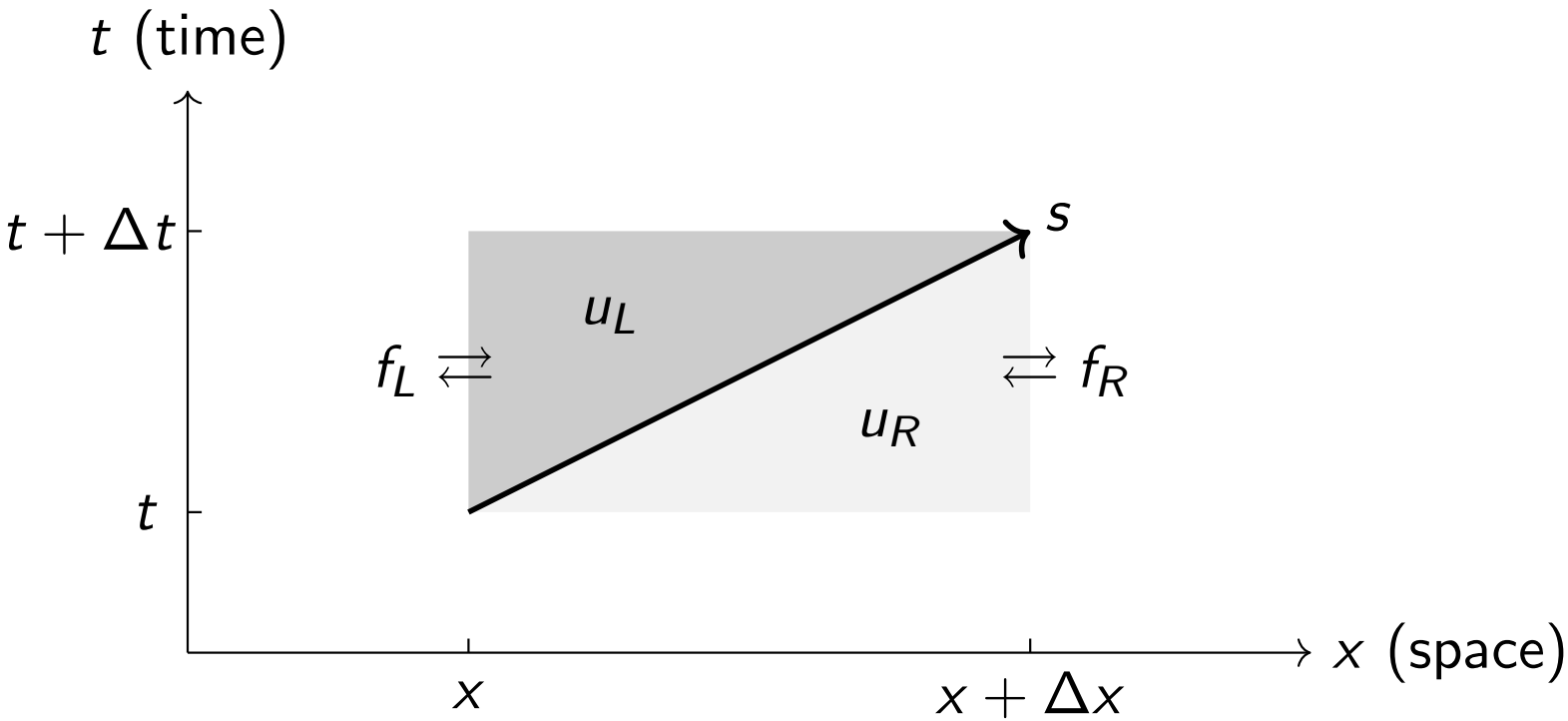
Notation : For the sake of simplification, we note $f_q = f(u_q)$ where q can stand for a spatial index or a side index in the context of a discontinuity/shock.

The jump conditions

$$\Delta x(u_L - u_R) + \Delta t(f_R - f_L) = 0 \tag{43}$$

Shock speed $s = \Delta x/\Delta t$: **Rankine-Hugoniot** jump condition

$$s(u_R - u_L) = f_R - f_L \tag{44}$$



Exact solution of the Riemann problem

The Riemann problem in vectorial form writes

$$\partial_t \mathbf{u} + \mathbf{A} \,.\, \partial_x \mathbf{u} = 0 \tag{45}$$

$$\mathbf{u}(x, t = 0) = \mathbf{u}_0(x) = \begin{cases} \mathbf{u}_L & \text{for } x < 0 \\ \mathbf{u}_R & \text{for } x > 0 \end{cases} \tag{46}$$

for a constant $N \times N$ diagonalizable matrix $\mathbf{A}$. Remember

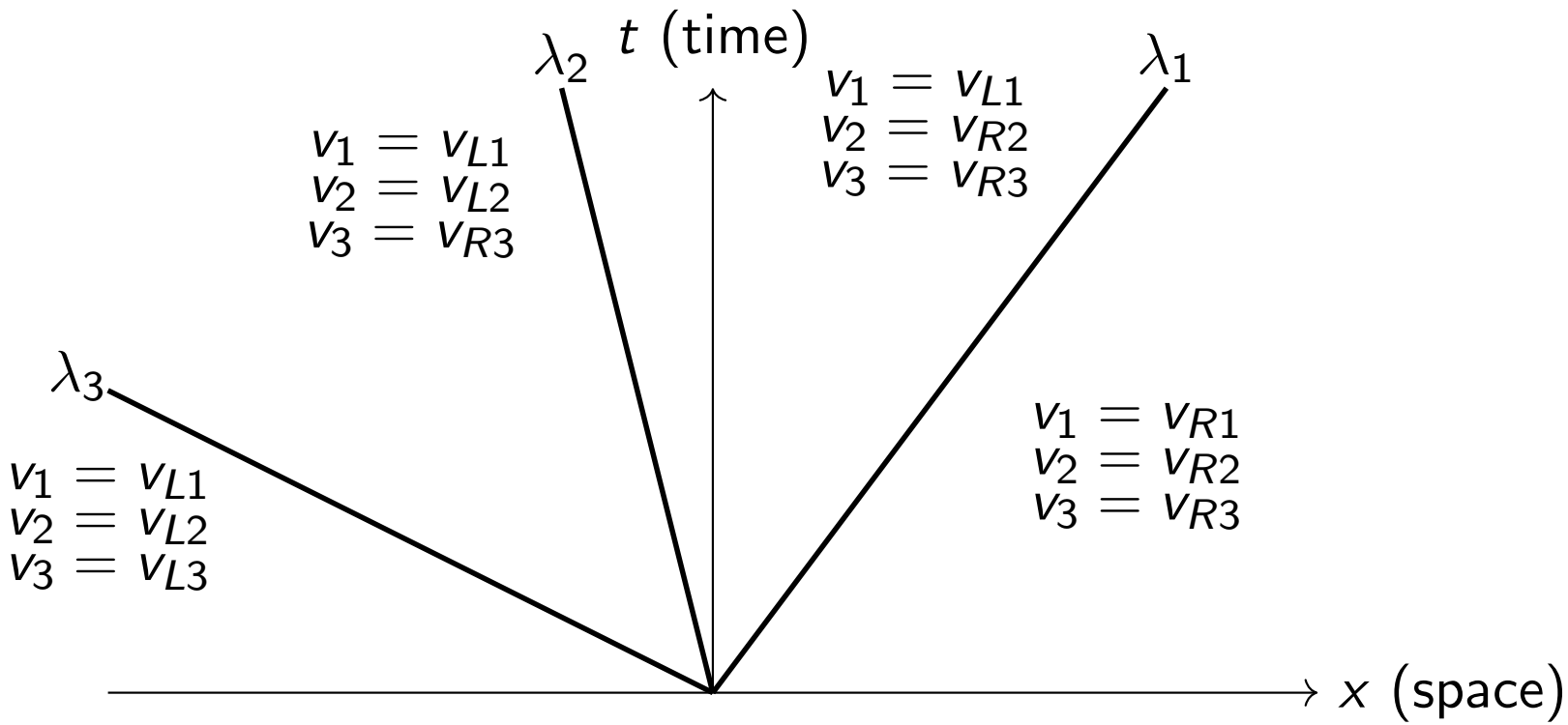
$$\mathbf{A} = \mathbf{Q} \,.\, \mathbf{\Lambda} \,.\, \mathbf{Q}^{-1} \text{ and } \mathbf{v} = \mathbf{Q}^{-1} \,.\, \mathbf{u} \tag{47}$$

The advection equation for $\mathbf{v}$ is the same as the one for $\mathbf{u}$, provided that $\mathbf{A}$ is replaced by $\mathbf{\Lambda}$, $\mathbf{u}_L$ by $\mathbf{v}_L = \mathbf{Q}^{-1} \,.\, \mathbf{u}_L$ and $\mathbf{u}_R$ by $\mathbf{v}_R = \mathbf{Q}^{-1} \,.\, \mathbf{u}_R$.

$\rightarrow$ the problem for $\mathbf{v}$ is still a Riemann problem, but with a diagonal Jacobian.

Exact solution of the Riemann problem

The matrix $\mathbf{\Lambda}$ is diagonal with three eigen values λ_i in the case $N = 3$ (which is the case for the one-dimensional Euler equations). As an example, we report the 3 caracteristics $\mathrm{d}_t x = \lambda_i$



Exact solution of the Riemann problem

Let $\Delta v_i = v_{Ri} - v_{Li}$ be the jump in the i^{th} variable defined as

$$\Delta \mathbf{v} = \mathbf{v}_R - \mathbf{v}_L = \mathbf{Q}^{-1} \,.\, \Delta \mathbf{u} \tag{48}$$

so we have

$$\Delta \mathbf{u} = \mathbf{Q} \,.\, \Delta \mathbf{v} \tag{49}$$

The i component of $\Delta \mathbf{u}$ is the i component of $\Delta \mathbf{v}$ multiplied by the i^{th} column of $\mathbf{Q}$, that is the i^{th} eigen vector.

We then get $\mathbf{u}(x/t)$, that is the solution of the Riemann problem.