An explicit stable numerical scheme for the 1D transport equation

Yohan Penel

AN EXPLICIT STABLE NUMERICAL SCHEME FOR THE 1D TRANSPORT EQUATION

YOHAN PENEL*

Abstract. We derive in this paper a numerical scheme in order to calculate solutions of $1D$ transport equations. This $2nd$-order scheme is based on the method of characteristics and consists of two steps: The first step is about the approximation of the foot of the characteristic curve whereas the second one deals with the computation of the solution at this point. The main idea in our scheme is to combine two $2nd$-order interpolation schemes so as to preserve the maximum principle. The resulting method is unconditionaly stable and designed for classical solutions but turns out to remain valid when shocks occur.

Key words. method of characteristics, linear advection equation, numerical scheme, MOC2.

Subject classifications. 65M25, 65M06.

1. Introduction The method of characteristics has been used for 40 years as a theoretical tool to prove existence of smooth solutions to the linear transport equation. No matter what simple, this equation arises in several physical application as soon as the convection operator

$$\partial_t + \mathcal{U}\cdot\nabla$$

is involved, whether the velocity field $\mathcal{U}$ be a datum or an unknown. For instance, it appears when diagonalising hyperbolic conservation laws [11] or writing fluid mechanics equations under a nonconservative form [4], or else when tracking level sets of a smooth function [13].

Above all, this method provides an implicit solution to the initial value problem (IVP) in the bounded domain $\Omega\subset\mathbb{R}^d$, $d\in\{1,2,3\}$:

$$\begin{cases} \partial_t\mathcal{Y}(t,\boldsymbol{x}) + \mathcal{U}(t,\boldsymbol{x})\cdot\nabla\mathcal{Y}(t,\boldsymbol{x}) = f(t,\boldsymbol{x}), \\ \mathcal{Y}(0,\boldsymbol{x}) = \mathcal{Y}^0(\boldsymbol{x}), \end{cases} \tag{1.1}$$

for smooth enough velocity field $\mathcal{U}$ (with null normal component on the boundary) and for bounded source term f and initial datum $\mathcal{Y}^0$ [1]. The solution is said to be implicit because it requires to solve an ordinary differential equation (instead of a partial differential equation) to which there exists a unique solution:

$$\begin{cases} \dfrac{\mathrm{d}\mathcal{X}}{\mathrm{d}\tau} = \mathcal{U}\big(\tau,\mathcal{X}(\tau)\big), \\ \mathcal{X}(s) = \boldsymbol{x}_0. \end{cases} \tag{1.2}$$

More precisely, considering the initial condition $\boldsymbol{x}_0$ of ODE (1.2) as a parameter, one defines characteristic curves as the orbits of the solutions. Then the evolution of the corresponding solution of (1.1) along the characteristic curve is known:

$$\mathcal{Y}(t,\boldsymbol{x}) = \mathcal{Y}^0\big(\mathcal{X}(0;t,\boldsymbol{x})\big) + \int_0^t f\big(\tau,\mathcal{X}(\tau;t,\boldsymbol{x})\big)\,\mathrm{d}\tau. \tag{1.3}$$

*Commissariat à l'Énergie Atomique (CEA), DEN/DANS/DM2S/SFME/LETR, 91191 Gif-sur-Yvette, France, *et* Laboratoire d'Analyse, Géométrie et Applications (LAGA), Université Paris 13, 93430 Villetaneuse, France. Contact: yohan.penel@gmail.com.

In the framework of fluid mechanics, the field $\mathcal{X}(t;s,\boldsymbol{x}_0)$ represents the position at time t of a particule which was located in $\boldsymbol{x}_0$ at time s in a fluid driven at velocity $\mathcal{U}$. That is why the characteristic curves can be considered as trajectories of particles.

The method of characteristics[1] also turns out to be of great interest from a numerical point of view. Indeed, as the solution is known through (1.3), it suffices to locally approximate the characteristic curves to calculate the solution to the IVP in a time neighbourhood: given the semi-group property satisfied by $\mathcal{X}$

$$\mathcal{X}\big(t_1;t_2,\mathcal{X}(t_2;t_3,\boldsymbol{x}_0)\big)=\mathcal{X}(t_1;t_3,\boldsymbol{x}_0), \tag{1.4}$$

we have the identity

$$\mathcal{Y}(t+\Delta t,\boldsymbol{x})=\mathcal{Y}\big(t,\mathcal{X}(t;t+\Delta t,\boldsymbol{x})\big)+\int_t^{t+\Delta t} f\big(\tau,\mathcal{X}(\tau;t+\Delta t,\boldsymbol{x})\big)\,\mathrm{d}\tau. \tag{1.3'}$$

There exist two main strategies depending of the overall problem.

① If the sole transport equation (1.1) is concerned, it is possible to work with (1.3) or (1.3') formulations. However, the (1.3)-based method induces the propagation of numerical diffusion since one has to go back upstream to the origin of time $t=0$ while in (1.3'), there is only a local calculation between $t+\Delta t$ and t. Then, the integral is computed by means of a numerical integration formula (Euler, Gauss). The method consists in two steps: the **construction of the characteristic** to provide the foot $\mathcal{X}(t;t+\Delta t,\boldsymbol{x})$ of the curve passing through $\boldsymbol{x}$ at time $t+\Delta t$ (as well as other values required by the integration formula) and **the evaluation of the computed solution** at time t and position $\mathcal{X}(t;t+\Delta t,\boldsymbol{x})$. We remark that this algorithm only requires values of the solution of ODE (1.2) with $s=t+\Delta t$ over the interval $[t,t+\Delta t]$. See for instance [10] or the method we shall describe in § 2.

② If the convection operator is part of a more complex system (advection-diffusion, Euler, Navier-Stokes):

$$\partial_t\mathcal{Y}+\mathcal{U}\cdot\nabla\mathcal{Y}+\mathcal{F}(\mathcal{Y})=f,$$

where $\mathcal{F}$ is a differential operator w.r.t. $\mathcal{Y}$, then the equation is rewritten as:

$$\frac{\mathrm{d}\mathcal{Y}}{\mathrm{d}t}+\mathcal{F}(\mathcal{Y})=f.$$

The directional derivative along the characteristic $\dfrac{\mathrm{d}\mathcal{Y}}{\mathrm{d}t}$ corresponds to:

$$\left[\frac{\partial}{\partial\tau}\Big(\mathcal{Y}\big(\tau,\mathcal{X}(\tau;t,\boldsymbol{x})\big)\Big)\right]_{\tau=t}.$$

This term is generally discretized in time as:

$$\frac{\mathcal{Y}(t,x)-\mathcal{Y}\big(t-\Delta t,\mathcal{X}(t-\Delta t;t,\boldsymbol{x})\big)}{\Delta t}.$$

Then, the semi-discrete equation may be solved *via* a Finite Element method [1, 6, 7, 8, 16]. This combination is sometimes called Lagrange-Galerkin

[1] In a numerical framework, methods of characteristics are denoted by MOC.

method. In particular, an algorithm is designed in [1, 16] when the velocity field is approximated by a piecewise constant function. The foot of the characteristics $\mathcal{X}(t-\Delta t;t,\boldsymbol{x})$ is generally computed with a first order scheme [6]:

$$\mathcal{X}(t-\Delta t;t,\boldsymbol{x})=\boldsymbol{x}-\mathcal{U}(t-\Delta t,\boldsymbol{x})\Delta t,$$

then corrected in [7] to preserve the mass conservation, or with a second order scheme [8].

Our method was first outlined in [15] in the framework of the modelling of bubbles, where the velocity field satisfies the elliptic equation:

$$\nabla\cdot\mathcal{U}(t,\boldsymbol{x})\propto\mathcal{Y}(t,\boldsymbol{x})-\frac{1}{|\Omega|}\int_\Omega\mathcal{Y}(t,\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}'.$$

We thus pay attention to the availability of values of $\mathcal{U}$ due to the fact that it depends on the solution $\mathcal{Y}$.

We shall describe in the sequel the derivation of our scheme (named MOC2), as well as qualitative properties (stability, consistancy, ...). We shall also present numerical results for the linear transport equation and the Burgers equation. This model is designed for dimension 1 and for smooth solutions to the transport equation. It is based on properties of the characteristic flow that we use in the approximation of the foot of the characteristic curve (2*nd*-order scheme) and on geometric considerations in the interpolation step so as to ensure the maximum principle.

2. Derivation of the scheme

2.1. Notations For sake of simplicity, we consider from now on the linear $1D$ advection equation (1.1) without source term (*i.e.* with $f=0$) over the domain $[0,1]$ and the time interval $[0,\mathscr{T}]$.

Then under smoothness assumptions on the velocity field, the solution is:

$$\mathcal{Y}(t,x)=Y^0\big(\mathcal{X}(0;t,x)\big)$$

according to (1.3) and satisfies the identity:

$$\mathcal{Y}(t+\Delta t,x)=\mathcal{Y}\big(t,\mathcal{X}(t;t+\Delta t,x)\big). \tag{2.1}$$

In order to derive the numerical method, we introduce the discretization parameters $N_t\in\mathbb{Z}_+$ and $N_x\in\mathbb{Z}_+$. Let Δt and Δx be the time step and the mesh given by:

$$\Delta t=\frac{\mathscr{T}}{n_t}\quad\text{and}\quad\Delta x=\frac{1}{n_x}.$$

We then define the uniform mesh:[2]

$$t^n=n\Delta t\quad\text{and}\quad x_i=(i-1)\Delta x,$$

$n\in\{0,\dots,N_t\}$ and $i\in\{1,\dots,N_x\}$, so that the unknowns are:

$$\mathcal{Y}_i^n=\mathcal{Y}(t^n,x_i).$$

[2]The method can be extended to variables meshes.

By virtue of (2.1), we compute:

$$\mathcal{Y}_i^{n+1} = \mathcal{Y}\big(t^n, \mathcal{X}(t^n; t^{n+1}, x_i)\big).$$

As we stated earlier, there are two main stages in MOCs, which correspond to the two next subsections:

- ❶ the computation of $\xi_i^n := \mathcal{X}(t^n; t^{n+1}, x_i)$ which is the foot of the characteristic curve going through x_i at time t^{n+1} (called the upstream point of x_i) – see Fig. 2.1. Remind that $\mathcal{X}$ is a solution to the nonlinear ODE (1.2);
- ❷ the calculation of $\mathcal{Y}(t^n, \xi_i^n)$ while we only know values of the numerical solution already computed, *i.e.* at nodes (x_j). But ξ_i^n is generally not a mesh node.

Theoretically, the first step can be interprated as a time procedure (it uses Δt as a parameter) while the second one mainly involves the mesh width Δx. This accounts for the unconditional stability of MOCs as a numerical scheme because Δt and Δx are used independantly from each other, even numerical approximations of the velocity make the analysis less straightforward.

2.2. Computation of the upstream point In the majority of MOCs, a 1*st*-order approach is used to build the characteristic curve which then reduces to a straight line – see $\bar{\xi}_i^n$ Fig. 2.1. When the velocity is constant,[3] this procedure is exact. But for variable velocity fields, it induces numerical errors. That is why we improve the accuracy with a second-order calculation for the upstream point.

We first recall some properties of the characteristic flow. Its derivatives (w.r.t. τ, s and x_0) satisfy the following equalities:

$$\partial_\tau \mathcal{X}(\tau; s, \boldsymbol{x}_0) = \mathcal{U}\big(\tau, \mathcal{X}(\tau; s, \boldsymbol{x}_0)\big), \tag{2.2a}$$

$$\partial_s \mathcal{X}(\tau; s, \boldsymbol{x}_0) = -\nabla_{\boldsymbol{x}_0} \mathcal{X}(\tau; s, \boldsymbol{x}_0) \mathcal{U}(s, \boldsymbol{x}_0), \tag{2.2b}$$

$$\det \nabla_{\boldsymbol{x}_0} \mathcal{X}(\tau; s, \boldsymbol{x}_0) = \exp \int_s^\tau \nabla \cdot \mathcal{U}\big(\sigma, \mathcal{X}(\sigma; s, \boldsymbol{x}_0)\big)\, \mathrm{d}\sigma. \tag{2.2c}$$

See for instance [2] for (2.2b) and [9, 12] for (2.2c).

As values at time t^{n+1} are not currently known, we use a Taylor expansion at (t^n, x_i):

$$\begin{aligned}\xi_i^n &= \mathcal{X}(t^n; t^{n+1}, x_i) \\ &= \mathcal{X}(t^n; t^n, x_i) + \Delta t \frac{\partial \mathcal{X}}{\partial s}(t^n; t^n, x_i) + \frac{\Delta t^2}{2} \frac{\partial^2 \mathcal{X}}{\partial s^2}(t^n; t^n, x_i) + \mathcal{O}\left(\Delta t^3\right).\end{aligned}$$

This formula requires to derive explicit expressions of derivatives of $\mathcal{X}$ w.r.t. s. Taking $\tau = s$ in (2.2b) and given that $\partial_{x_0} \mathcal{X}(s; s, x_0) = Id$, we have:

$$\partial_s \mathcal{X}(s; s, x_0) = -\mathcal{U}(s, x_0).$$

This deals with the Δt-term. For the Δt^2-term, we differentiate the last relation w.r.t. s so that we obtain:

$$\partial_{ss}^2 \mathcal{X}(s; s, x_0) = -\partial_t \mathcal{U}(s, x_0) - \partial_{\tau s}^2 \mathcal{X}(s; s, x_0).$$

[3] The constant case is useful insofar as the solution is explicitly known.

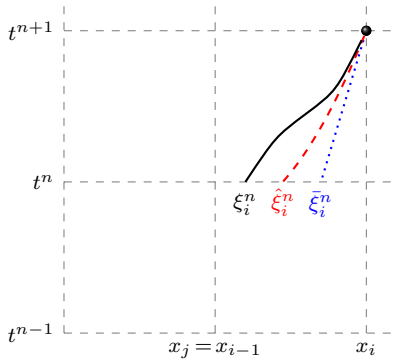


FIG. 2.1. *1st-order and 2nd-order upstream points*

As we suppose enough regularity to apply the Schwarz lemma, the derivative $\partial^2_{\tau s}\mathcal{X}$ is equal to $\partial^2_{s\tau}\mathcal{X}$ and can be computed by differentiate (2.2a) w.r.t. s, which reads:

$$\partial^2_{\tau s}\mathcal{X}(\tau;s,x_0) = \partial_x\mathcal{U}\big(\tau,\mathcal{X}(\tau;s,x_0)\big)\partial_s\mathcal{X}(\tau;s,x_0).$$

This leads to the semi-discrete formulation:

$$\xi_i^n = x_i - \mathcal{U}(t^n,x_i)\Delta t + \frac{\Delta t^2}{2}\left[\mathcal{U}(t^n,x_i)\partial_x\mathcal{U}(t^n,x_i) - \partial_t\mathcal{U}(t^n,x_i)\right] + \mathcal{O}\left(\Delta t^3\right). \tag{2.3}$$

Note that this relation holds in higher dimensions:

$$\xi_i^n = \boldsymbol{x}_i - \mathcal{U}(t^n,\boldsymbol{x}_i)\Delta t + \frac{\Delta t^2}{2}\left[(\mathcal{U}\cdot\nabla)\mathcal{U}(t^n,\boldsymbol{x}_i) - \partial_t\mathcal{U}(t^n,\boldsymbol{x}_i)\right] + \mathcal{O}(\Delta t^3).$$

In the sequel, $\hat{\xi}_i^n$, Y_i^n and U_i^n denote the numerical approximations of ξ_i^n, $\mathcal{Y}(t^n,x_i)$ and $\mathcal{U}(t^n,x_i)$.

Eq. (2.3) raises the issue of computing spatial derivatives of $\mathcal{U}$. Indeed, if $\mathcal{U}$ is a given function independent from $\mathcal{Y}$, then such derivatives can be approximated by finite-difference (FD) formulae (for instance $2nd$-order centered FD in space, $1st$-order upwind FD in time):[4]

$$\xi_i^n \approx x_i - \frac{\Delta t}{2}\left[3U_i^n - U_i^{n-1}\right] + \frac{\Delta t^2}{2}\left[\mathcal{U}_i^n\frac{U_{i+1}^n - U_{i-1}^n}{2\Delta x}\right].$$

However, if $\mathcal{U}$ is a solution to another PDE, this equation can be used to deal with differential terms appearing in (2.3). For instance, if $\mathcal{U}$ is a solution to the Burgers equation [3], the second order term can be rewritten:

$$\Delta t^2\mathcal{U}(t^n,x_i)\partial_x\mathcal{U}(t^n,x_i).$$

We could also reformulate as $-\Delta t^2\partial_t\mathcal{U}(t^n,x_i)$ but this would require to store two time levels of the unknown. That is why we prefer keeping only spatial derivatives.

[4]The term $\mathcal{U}\partial_x\mathcal{U}$ can also be considered as the exact derivative $\partial_x(\mathcal{U}^2/2)$.

2.3. Interpolation step Once the upstream point computed, we find out the interval $[x_j,x_{j+1})$ from which the characteristic curve issues at time t^n. Let θ_{ij}^n be the position of $\hat{\xi}_i^n$ in this interval, *i.e.*:

$$\theta_{ij}^n = \frac{x_{j+1}-\hat{\xi}_i^n}{\Delta x}.$$

By definition, $\theta_{ij}^n \in (0,1]$ and $\theta_{ij}^n = 1$ for $\hat{\xi}_i^n = x_j$.

As $\hat{\xi}_i^n \in [x_j,x_{j+1}]$, a natural choice to compute $Y_i^{n+1} \approx \mathcal{Y}(t^n,\xi_i^n)$ is to use a linear interpolation involving Y_j^n and Y_{j+1}^n. But as we remark a progressive loss of accuracy, we get interested in second-order interpolation schemes.

One may either take (x_{j-1},x_j,x_{j+1}) or (x_j,x_{j+1},x_{j+2}) into account. The Lagrange interpolation polynomials associated to these two sets of points, when expressed in the θ variable, read:

$$\mathbf{Y}_l(\theta) = -\frac{\theta(1-\theta)}{2}Y_{j-1}^n + \theta(2-\theta)Y_j^n + \frac{(1-\theta)(2-\theta)}{2}Y_{j+1}^n \tag{2.4a}$$

$$= \frac{\theta^2}{2}\left(Y_{j-1}^n - 2Y_j^n + Y_{j+1}^n\right) - \frac{\theta}{2}\left(Y_{j-1}^n - 4Y_j^n + 3Y_{j+1}^n\right) + Y_{j+1}^n. \tag{2.4b}$$

and

$$\mathbf{Y}_r(\theta) = \frac{\theta(1+\theta)}{2}Y_j^n + \left(1-\theta^2\right)Y_{j+1}^n - \frac{\theta(1-\theta)}{2}Y_{j+2}^n \tag{2.5a}$$

$$= \frac{\theta^2}{2}\left(Y_{j+2}^n - 2Y_{j+1}^n + Y_j^n\right) - \frac{\theta}{2}\left(Y_{j+2}^n - Y_j^n\right) + Y_{j+1}^n. \tag{2.5b}$$

We have noticed that formulae (2.4a) and (2.5a) induce round-off errors. Although useful in the theoretical study, we do not use them numerically. We rather use (2.4b) and (2.5b).

These two possibilities thus differ from the stencil and the sign of the weights: In both cases (2.4a) and (2.5a), Y_j^n and Y_{j+1}^n have positive coefficients unlike the third value (Y_{j-1}^n or Y_{j+2}^n) which has a negative weight. This proves that each scheme (of 2nd-order) is not monotonicity-preserving (as stated by Godunov [11, Th. 16.1]). We see on Fig. 2.2(a) that the linear combination in $\mathbf{Y}_r$ may provide a negative value. Likewise, the other scheme $\mathbf{Y}_l$ may not suit in other configurations (Fig. 2.2(b)).

These considerations can be easily interprated from a geometrical point of view. In the (x,Y) plane, we set $\mathbf{X}_k = (x_k, Y_k^n)$, $k \in \{j-1,j,j+1,j+2\}$ and $\mathbf{X}_p^\theta = \left(\hat{\xi}_i^n, \mathbf{Y}_p(\theta_{ij}^n)\right)$, $p \in \{l,r\}$. Hence:

$$\mathbf{X}_l^\theta = -\frac{\theta_{ij}^n(1-\theta_{ij}^n)}{2}\mathbf{X}_{j-1} + \theta_{ij}^n(2-\theta_{ij}^n)\mathbf{X}_j + \frac{(1-\theta_{ij}^n)(2-\theta_{ij}^n)}{2}\mathbf{X}_{j+1},$$

$$\mathbf{X}_r^\theta = \frac{\theta_{ij}^n(1+\theta_{ij}^n)}{2}\mathbf{X}_j + \left(1-\theta_{ij}^n\right)\left(1+\theta_{ij}^n\right)\mathbf{X}_{j+1} - \frac{\theta_{ij}^n(1-\theta_{ij}^n)}{2}\mathbf{X}_{j+2}.$$

Notice that even in the degenerate case where the three points are aligned, these formulae have a sense and enable to compute a relevant value.

This also proves that the relation satisfied by y-coordinates (values of Y_i^n) is also satisfied by x-coordinates. Thus, $\mathbf{X}_l^\theta$ is a barycenter of the points $\mathbf{X}_{j-1}$, $\mathbf{X}_j$ and $\mathbf{X}_{j+1}$. We represent on Fig. 2.3 the areas in which $\mathbf{X}_l^\theta$ and $\mathbf{X}_r^\theta$ may be located. More precisely, $\mathbf{X}_l^\theta$ lies outside the triangle $\mathbf{X}_{j-1}\mathbf{X}_j\mathbf{X}_{j+1}$ (in the half-plane with boundary $(\mathbf{X}_j\mathbf{X}_{j+1})$ which does not contain $\mathbf{X}_{j-1}$).

The idea is thus to combine these two formulae depending on the failure of the maximum principle in one case or another. In order to ensure a global maximum principle, we would like to impose it locally:

$$\mathbf{Y}_l(\theta_{ij}^n), \mathbf{Y}_r(\theta_{ij}^n) \in [\min(Y_j^n, Y_{j+1}^n), \max(Y_j^n, Y_{j+1}^n)]. \tag{2.6}$$

These two criteria cannot be satisfied everywhere for both schemes as explained before and they even fail together at the same time. The point is that one at least is admissible in the critical cases, namely close to the areas where $Y_i^n = \max_j Y_j^n$ and where $Y_i^n = \min_j Y_j^n$.

One way to check that $\mathbf{Y}_l(\theta_{ij}^n)$ satisfies (2.6) consists in determining the position of the extremum of the $2nd$-order polynomial function $\mathbf{Y}_l$: If $\mathbf{Y}_l$ reaches its extremum in θ_l with $\theta_l \in (0,1)$ (which corresponds to (Y_j^n, Y_{j+1}^n) in x-variable), then the maximum principle may fail. We calculate:

$$\theta_l = \frac{Y_{j-1}^n - 4Y_j^n + 3Y_{j+1}^n}{2(Y_{j-1}^n - 2Y_j^n + Y_{j+1}^n)} = \frac{1}{2} + \frac{Y_{j+1}^n - Y_j^n}{Y_{j-1}^n - 2Y_j^n + Y_{j+1}^n} = \vartheta_l(Y_{j-1}^n).$$

Given Y_j^n and Y_{j+1}^n, the admissible values for Y_{j-1}^n so that $\vartheta_l(Y_{j-1}^n) \notin (0,1)$ are such that $Y_{j+1}^n - Y_j^n$, $Y_{j+1}^n - Y_{j-1}^n$ and $Y_{j-1}^n - 4Y_j^n + 3Y_{j+1}^n$ have the same sign (due to the variations of ϑ_l), which reduces to:

$$\left(Y_{j+1}^n - Y_{j-1}^n\right)\left(Y_{j-1}^n - 4Y_j^n + 3Y_{j+1}^n\right) \geq 0. \tag{2.7a}$$

We similarly get the following constraint for the right scheme $\mathbf{Y}_r$:

$$\left(Y_{j+2}^n - Y_j^n\right)\left(-3Y_j^n + 4Y_{j+1}^n - Y_{j+2}^n\right) \geq 0. \tag{2.7b}$$

We shall make a few comments about these strong sufficient conditions. First of all, they are independent from θ_{ij}^n, which decouples the two steps of the procedure. Secondly, they define an admissible domain for Y_{j-1}^n and Y_{j+2}^n (see FIG. 2.4). However, if $Y_j^n = Y_{j+1}^n$, we expect to have either $Y_{j-1}^n = Y_j^n$ or $Y_{j+2}^n = Y_{j+1}^n$ so as to prevent irrelevant values. But we remark that dividing Eqs. (2.7) by $4\Delta x$, they are equivalent in the limit as $\Delta x \to 0$ to $(\partial_x Y)^2 \geq 0$, which is trivially true. This shows that when refining the mesh width, the two conditions tend to be satisfied (for smooth solutions).

Nevertheless, situations may occur where $\mathbf{Y}_p(\theta_{ij}^n) \in [Y_j^n, Y_{j+1}^n]$ although $\theta_p \in (0,1)$. Indeed, the FIG. 2.5 shows that the dashed blue part of the parabola provides values in the good interval. That is why (2.7) seem to be too restrictive. To go further, we are lead to introduce the equations:

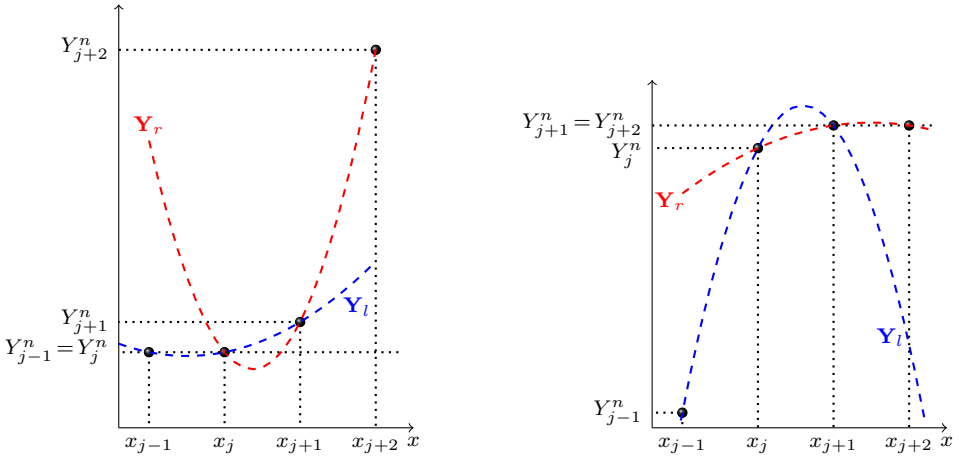
$$\mathbf{Y}_p(\theta) = Y_k^n,\ p \in \{l,r\},\ k \in \{j, j+1\}.$$

There are trivial solutions for $k = j$ (which is $\theta = 1$) and $k = j+1$ ($\theta = 0$). In the non-degenerate case, the other solutions are:

$$\kappa_{l,j} = \frac{2(Y_{j+1}^n - Y_j^n)}{Y_{j-1}^n - 2Y_j^n + Y_{j+1}^n}, \qquad \kappa_{r,j} = \frac{2(Y_{j+1}^n - Y_j^n)}{Y_j^n - 2Y_{j+1}^n + Y_{j+2}^n},$$
$$\kappa_{l,j+1} = \frac{Y_{j-1}^n - 4Y_j^n + 3Y_{j+1}^n}{Y_{j-1}^n - 2Y_j^n + Y_{j+1}^n}, \qquad \kappa_{r,j+1} = \frac{Y_{j+2}^n - Y_j^n}{Y_j^n - 2Y_{j+1}^n + Y_{j+2}^n}.$$

We have the relation $\kappa_{p,j+1} - \kappa_{p,j} = 1$, $p \in \{l,r\}$. We also recover Eqs. (2.7) through the equivalence:

$$\theta_p \notin (0,1) \Longleftrightarrow \kappa_{p,j} \notin (0,1) \text{ and } \kappa_{p,j+1} \notin (0,1).$$



(a) Convex monotone-increasing case (b) Nonconvex monotone-increasing case

FIG. 2.2. *Configurations of nodes and location of the 2nd-order interpolated points*

But according to FIG. 2.5, $\mathbf{Y}_l(\theta^n_{ij}) \in [Y_j, Y_{j+1}]$ if $\theta^n_{ij} \leq \kappa_{l,j}$. More generally, $\mathbf{Y}_p$ is admissible iff:

$$\theta^n_{ij} \leq \kappa_{p,j} \text{ or } \theta^n_{ij} \geq \kappa_{p,j+1}. \tag{2.8}$$

The point is now to choose one scheme or the other when (2.8) is satisfied for both $p=l$ and $p=r$. Naturally, when only one of the two schemes is admissible at one node, it has to be applied for the interpolation step. If both are admissible, there are several possible strategies:

- ❶ Similarly to the spirit of the antidiffusive scheme designed by Després and Lagoutière [5], we can use downwind data ($\mathbf{Y}_r$ if $U^n_j > 0$, $\mathbf{Y}_l$ if $U^n_j < 0$) as soon as it is possible, and upwind ones otherwise.
- ❷ We can also take into account the configurations that lead us to combine two schemes (see FIG. 2.2 and [14]). If the solution at time t^n is convex and monotone-increasing or nonconvex and monotone-decreasing, we would like to use $\mathbf{Y}_l$.

Strategy ❶ depends on the sign of the local velocity while ❷ is based on the local properties of the solution itself. However, it requires to define discrete monotonicty and convexity concepts.

3. Numerical analysis We present in this part the numerical analysis of our scheme. We first start paying attention to the coupling between the 2*nd*-order upstream point construction and the left scheme $\mathbf{Y}_l$ (we leave aside the second scheme for now). For sake of simplicity, the analysis is done under the hypothesis that $\mathcal{U}$ is constant in time and space (we shall note u its value), which enables to compare with conservative methods since the equation now reads:

$$\partial_t \mathcal{Y} + \partial_x(u\mathcal{Y}) = 0. \tag{3.1}$$

Then $\xi^n_i = \hat{\xi}^n_i = x_i - u\Delta t \in [x_j, x_{j+1})$ for j such that:

$$j = i + \left\lfloor \frac{-u\Delta t}{\Delta x} \right\rfloor = i + \lfloor -\lambda \rfloor, \tag{3.2}$$

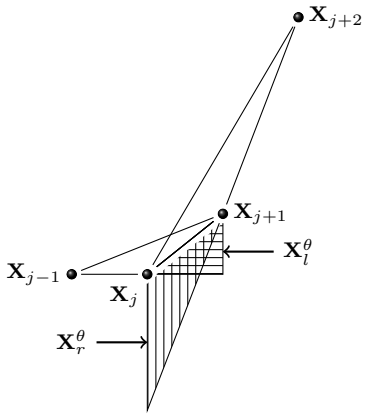


FIG. 2.3. *Location of barycenters* $\left(\xi, \mathbf{Y}_p(\theta)\right)$, $p \in \{l,r\}$

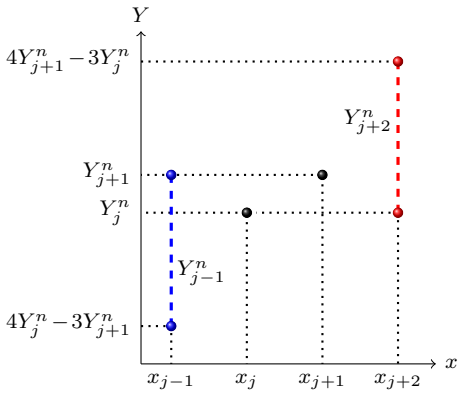


FIG. 2.4. *Authorized values for* Y_{j-1}^n *and* Y_{j+2}^n

with the usual Courant number notation $\lambda = u\Delta t/\Delta x$. Then:

$$\theta_{ij}^n = \theta = j - i + 1 + \lambda = 1 + \lambda + \lfloor -\lambda \rfloor.$$

In particular, for $u > 0$ such that $\lambda < 1$, our scheme is equivalent to the Beam-Warming scheme. Likewise, for $u < 0$ and $|\lambda| < 1$, it corresponds to the Lax-Wendroff scheme.

No matter what λ, as previously stated, this scheme is conservative but not monotonicity-preserving when (2.8) is not satisfied for $p = l$, which induces oscillations in numerical simulations. The method is stable without restriction on λ because even when $\Delta x \to 0$, θ is bounded[5]. We now show the consistency of the scheme. Most MOCs are proved to be of order $\mathcal{O}\left(\Delta t + \Delta x + \dfrac{\Delta x^2}{\Delta t}\right)$ [16], which is a conditionnal consistency. But we show that our scheme is consistent whatever Δx and Δt.

[5]It can be shown with a Von Neumann analysis.

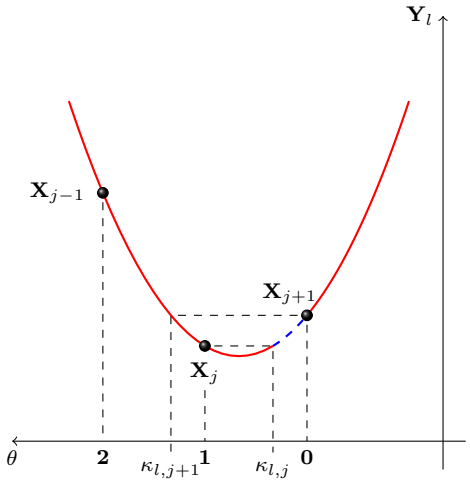


FIG. 2.5. *Admissibility of the interpolation scheme* $\mathbf{Y}_l$

For $\mathcal{Y}$ solution to (3.1), we set:

$$\mathcal{E}_i^n(\Delta t,\Delta x) := \frac{1}{\Delta t}\Big[\mathcal{Y}(t^{n+1},x_i) - \frac{\theta^2}{2}\big[\mathcal{Y}(t^n,x_{j-1}) - 2\mathcal{Y}(t^n,x_j) + \mathcal{Y}(t^n,x_{j+1})\big] + \frac{\theta}{2}\big[\mathcal{Y}(t^n,x_{j-1}) - 4\mathcal{Y}(t^n,x_j) + 3\mathcal{Y}(t^n,x_{j+1})\big] - \mathcal{Y}(t^n,x_{j+1})\Big].$$

We apply the integral Taylor expansion formula around (t^n,x_i):

$$\mathcal{Y}(t^n,x_k) = \mathcal{Y}(t^n,x_i) + (k-i)\Delta x\,\partial_x\mathcal{Y}(t^n,x_i) + \frac{(k-i)^2\Delta x^2}{2}\partial^2_{xx}\mathcal{Y}(t^n,x_i) + \int_{x_i}^{x_k}\frac{(x_k-z)^2}{2}\mathcal{Y}^{(3)}(t^n,z)\,\mathrm{d}z,$$

which leads to :

$$\begin{aligned}\Delta t\,\mathcal{E}_i^n(\Delta t,\Delta x) = {}& \Delta t\,\mathcal{R}_i^n + \Delta t\,\partial_t\mathcal{Y}(t^n,x_i) + \frac{\Delta t^2}{2}\partial^2_{tt}\mathcal{Y}(t^n,x_i) + \mathcal{O}(\Delta t^3)\\ & + \partial_x\mathcal{Y}(t^n,x_i)\Delta x\left[\frac{-\theta^2}{2}\big((j-i-1) - 2(j-i) + (j-i+1)\big)\right.\\ & \left. + \frac{\theta}{2}\big((j-i-1) - 4(j-i) + 3(j-i+1)\big) - (j-i+1)\right]\\ & + \partial^2_{xx}\mathcal{Y}(t^n,x_i)\frac{\Delta x^2}{2}\left[\frac{-\theta^2}{2}\big((j-i-1)^2 - 2(j-i)^2 + (j-i+1)^2\big)\right.\\ & \left. + \frac{\theta}{2}\big((j-i-1)^2 - 4(j-i)^2 + 3(j-i+1)^2\big) - (j-i+1)^2\right],\end{aligned}$$

where :

$$\mathcal{R}_i^n = -\frac{(1-\theta)(2-\theta)}{2\Delta t}\bar{\mathcal{R}}^n_{i,j+1} - \frac{\theta(2-\theta)}{\Delta t}\bar{\mathcal{R}}^n_{i,j} + \frac{\theta(1-\theta)}{2\Delta t}\bar{\mathcal{R}}^n_{i,j-1},$$

and:

$$\bar{\mathcal{R}}_{i,k}^n = \int_{x_i}^{x_k} \frac{(x_k - z)^2}{2} \mathcal{Y}^{(3)}(t^n, z)\, \mathrm{d}z.$$

Thus:

$$\mathcal{E}_i^n(\Delta t, \Delta x) = \Big[\underbrace{\partial_t \mathcal{Y} + u\partial_x \mathcal{Y}}_{=0}\Big](t^n, x_i) + \frac{\Delta t}{2}\Big[\underbrace{\partial_{tt}^2 \mathcal{Y} - u^2 \partial_{xx}^2 \mathcal{Y}}_{=0}\Big](t^n, x_i) + \mathcal{R}_i^n + \mathcal{O}(\Delta t^2).$$

Indeed, since $\mathcal{Y}$ is a solution to (3.1) with constant velocity, $\mathcal{Y}$ also satisfies the $1D$ wave equation $\partial_{tt}^2 \mathcal{Y} - u^2 \partial_{xx}^2 \mathcal{Y} = 0$.

Given the equality $x_k - x_i = (k-i)\Delta x$, we derive a bound for each integral term, assuming $Y^{(3)}$ bounded:

$$\big|\bar{\mathcal{R}}_{i,k}^n\big| \leq \Big\|\mathcal{Y}^{(3)}\Big\|_\infty \frac{(k-i)^2 \Delta x^2}{2}.$$

Then, taking (3.2) into account, this leads to:

$$|\mathcal{R}_i^n| \leq \frac{\Delta x^3}{12\Delta t}\Big\|\mathcal{Y}^{(3)}\Big\|_\infty \Big[2(1-\theta)\big|\lfloor -\lambda \rfloor + 1\big|^3 + 4\theta\big|\lfloor -\lambda \rfloor\big|^3 + \theta(1-\theta)\big|\lfloor -\lambda \rfloor - 1\big|^3\Big], \quad (3.3)$$

with $\theta = 1 + \lambda + \lfloor -\lambda \rfloor$. We have decided to bound $2-\theta$ (by 4) and to keep the terms θ and $1-\theta$ due to their asymptotic behaviour:

$$\theta \underset{\lambda \to 0_+}{\sim} \lambda \to 0 \text{ and } 1 - \theta \underset{\lambda \to 0_-}{\sim} -\lambda \to 0.$$

To show the consistency, we have to prove that the right hand side in (3.3) considered as a function of λ tends to 0 as Δx and Δt tend to 0.

Assume first that the two parameters satisfy $\Delta t = \Delta x^\alpha$ for some $\alpha > 0$. We distinguish three cases:

① $0 < \alpha < 1$: as $\lambda = u\Delta x^{\alpha - 1} \to \infty$, we bound θ and $1-\theta$ by 1. Then, each term in (3.3) is equivalent (up to a multiplicative constant) to:

$$\frac{\Delta x^3}{\Delta t}\big|\lfloor -\lambda \rfloor\big|^3 = \Delta x^{3-\alpha}\big|\lfloor -u\Delta x^{\alpha-1} \rfloor\big|^3 \sim |u|^3 \Delta x^{2\alpha} \to 0.$$

② $\alpha = 1$: $\lambda = u$ is a constant, which implies that the whole term between brackets is a constant. The factor Δx^2 ensures the convergence.

③ $\alpha > 1$: for $\lambda = u\Delta x^{\alpha-1}$ small enough, each term in the right hand side of (3.3) is either 0 ($\lfloor -\lambda \rfloor + 1 = 0$ for $u > 0$, $\lfloor -\lambda \rfloor = 0$ for $u < 0$) or bounded by $\theta \Delta x^{3-\alpha}$ ($u > 0$) or $(1-\theta)\Delta x^{3-\alpha}$ ($u < 0$). As:

$$\frac{\theta}{\Delta x^{\alpha-3}} \underset{\lambda \to 0_+}{\sim} u\Delta x^2 \text{ and } \frac{1-\theta}{\Delta x^{\alpha-3}} \underset{\lambda \to 0_-}{\sim} |u|\Delta x^2,$$

we deduce that the right hand side tends to 0 in all cases when Δt depends on Δx.

Hence, the global order is:

- $\mathcal{E}_i^n(\Delta t, \Delta t^{1/\alpha}) = \mathcal{O}(\Delta t^2)$, $\alpha < 1$;
- $\mathcal{E}_i^n(\Delta x^\alpha, \Delta x) = \mathcal{O}(\Delta x^2)$, $\alpha \geq 1$.

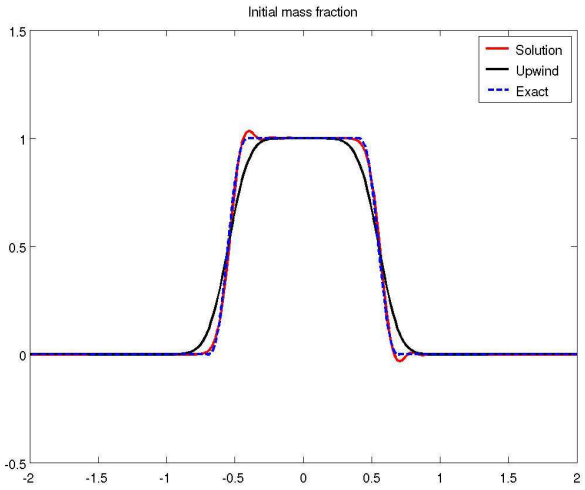


FIG. 4.1. *MOC using only* $\mathbf{Y}_l$ *interpolation scheme*

For the general case $\Delta t, \Delta x \to 0$, the same arguments hold, depending on whether Δt or Δx tends first to 0. For instance, we rewrite the first term in (3.3) either:

$$\Delta t^2(1-\theta)\left|\frac{\lfloor -\lambda \rfloor + 1}{\lambda}\right|^3 = \mathcal{O}(\Delta t^2),$$

in the limit as $\Delta x \to 0$ ($\lambda \to \infty$), or:

$$\Delta x^2(1-\theta)\frac{\left|\lfloor -\lambda \rfloor + 1\right|^3}{|\lambda|} = \mathcal{O}(\Delta x^2),$$

in the limit as $\Delta t \to 0$ ($\lambda \to 0$).

We have thus constructed an explicit linear stable and consistent scheme but which turns out to be dissipative (see FIG. 4.1). By combining the two interpolation schemes ($\mathbf{Y}_l$ and $\mathbf{Y}_r$), the maximum principle is now ensured (by construction) even if we loose linearity (through the choice between the two schemes at each node) and conservativity. Its major property is that it is **unconditionaly stable**, which enables to choose Δt and Δx independently from each other.

We should underline that it has been designed for **smooth** functions modelling bubbles (*i.e.* equal to 1 or to 0 on large intervals with smooth transitions between 1 and 0). That is why we do not present any discontinuous simulation in the next part.

4. Simulations

In order to highlight the numerical properties of the MOC2 scheme, we present in this section some numerical simulations. The first case shows the advantage of combining two 2*nd*-order schemes for the linear transport equation with constant velocity ($u = 0.5$). We consider the initial condition over the domain $[-2,2]$:

$$\mathcal{Y}_1^0(x) = \begin{cases} 1 & \text{if } x \in [-0.4, 0.4], \\ 0 & \text{if } x \in [-2, -0.7] \cup [0.7, 2], \end{cases}$$

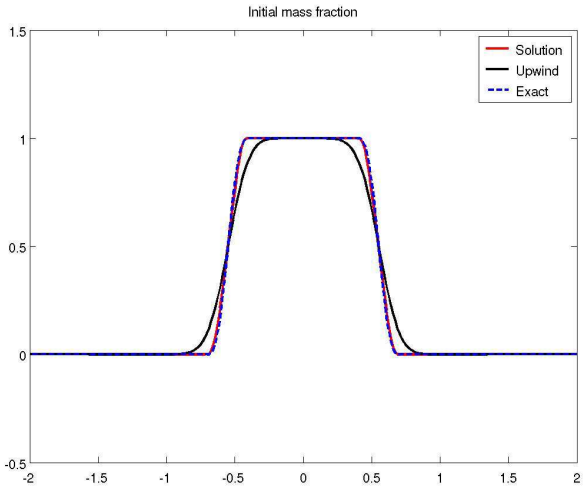


FIG. 4.2. *MOC2 scheme for a bubble-kind solution*

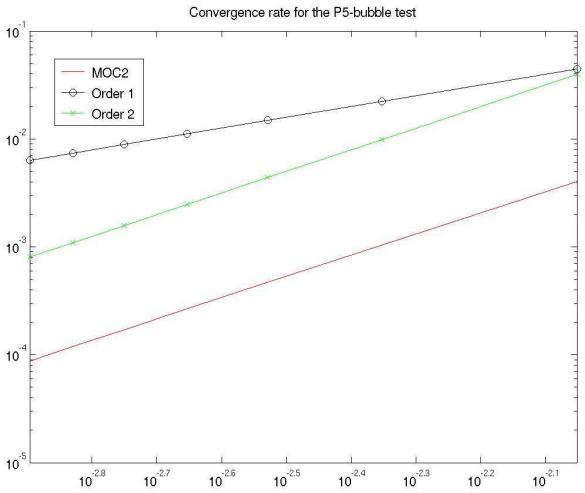
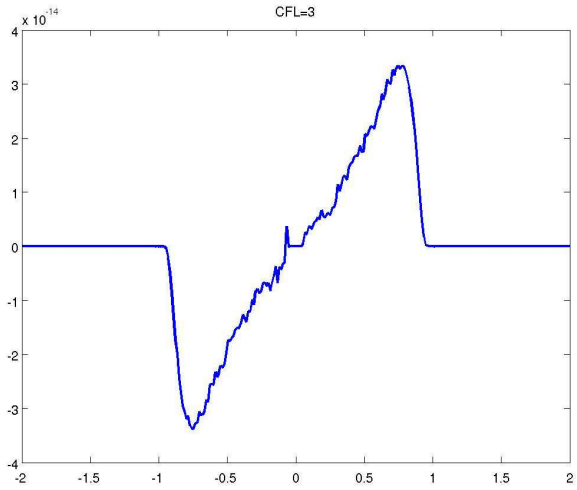
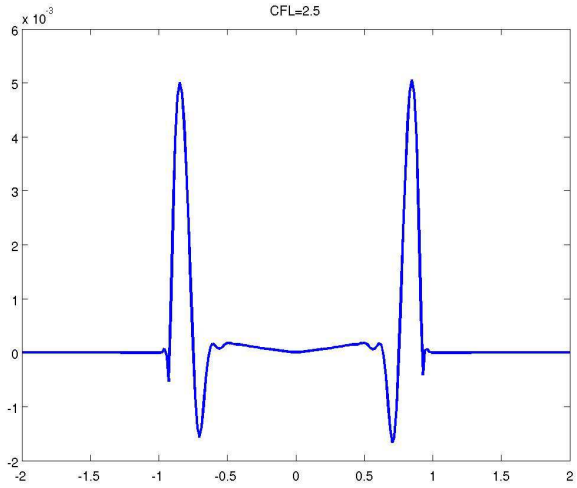


FIG. 4.3. *Convergence of the MOC2 scheme*

with a polynomial regularization (of $5th$ degree). We impose periodic boundary conditions so that we must recover the initial condition at time $\mathscr{T}=8$. In this case, $\Delta t=8/300$ and $\Delta x=4/250$ (Δt is chosen sufficiently small so that the upwind scheme is stable and can be used for comparisons). We see on FIG. 4.1 that when interpolating exclusively by means of $\mathbf{Y}_l$, numerical dissipation occurs whereas the combination of $\mathbf{Y}_l$ and $\mathbf{Y}_r$ enforces the maximum principle: Indeed, the solution provided by MOC2 matches with the exact solution (FIG. 4.2). Under the CFL condition, the upwind scheme converges too but is much more diffusive.



(a) $\Delta t = 0.08, \Delta x = \frac{4}{3} \times 10^{-2}, \lambda = 3$



(b) $\Delta t = 0.08, \Delta x = 0.016, \lambda = 2.5$

FIG. 4.4. *Error plot for different* λ

We emphasize the order 2 of the method on FIG. 4.3 for $\Delta t \propto \Delta x$ by representing the numerical error $\|Y_{MOC2} - \mathcal{Y}\|/\|\mathcal{Y}\|$ with respect to Δx. The underlying initial condition is still Func. $\mathcal{Y}_1^0$ described above.

As for the discretization parameters, we know that the MOC2 scheme allows to choose Δt and Δx independently from each other. For $\lambda \in \mathbb{Z}$, the scheme is exact in this constant case, unlike when $\lambda \notin \mathbb{Z}$. It can be infered from the comparison between FIG. 4.4(b) (where $\lambda = 2.5$ and the error is of order 10^{-3}) and FIG. 4.4(a) (where $\lambda = 3$ and the error is about 10^{-14}). The initial condition associated to this test is the $\mathscr{C}^\infty$

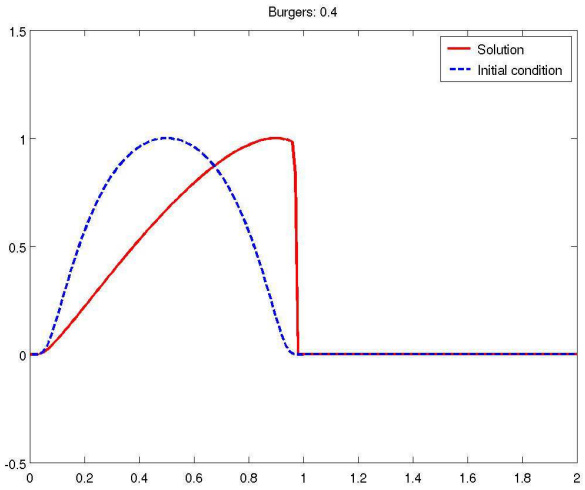


FIG. 4.5. *Solution of the Burgers equation at time* 0 *and* 0.4

function with compact support:

$$\mathcal{Y}_2^0(x) = \exp\frac{-x^2}{1-x^2}\mathbf{1}_{(-1,1)}(x).$$

We end the numerical section with a nonlinear application, namely the $1D$ Burgers equation:

$$\partial_t \mathcal{U} + \mathcal{U}\partial_x \mathcal{U} = 0.$$

As explained in § 2.2, the upstream point is localized by means of a $2nd$-order projection formula involving the term $\mathcal{U}\partial_x\mathcal{U}(t^n,x_i)$ instead of $-\partial_t\mathcal{U}(t^n,x_i)$ whose discretization would require two time levels of data. We endow the equation with the initial condition $\mathcal{Y}_2^0$. It is proven that there exists a finite time when characteristics cross and a shock forms [11]. We see on FIG. 4.5 the initial condition together with the numerical solution given by MOC2 after the shock formation. Not only do these results prove that our scheme is able to cope with nonlinear problems, but they also go to show that it may capture shocks, although it has been designed for smooth contexts.

5. Conclusion We have designed in this paper a new numerical method of characteristics which is $2nd$-order accurate so as to simulate smooth solutions to convection problems such as the linear transport equation or the Burgers equation. The main point is the combination of two $2nd$-order interpolation formulae to ensure the maximum principle by construction. This procedure results in a nonlinear stable scheme which has proven to give expected results relative to the order of accuracy in various cases and without CFL-kind condition. Moreover, numerical resultats tend to show that the scheme is able to capture shocks.

However, the strategy of choosing one interpolation formula or the other has to be refined. In addition, this scheme may be extended to nonhomogeneous meshes in $1D$ before being adapted to higher dimensions. We thus have to pay attention to the two steps composing our method. On $2D$ and $3D$ cartesian meshes, both seem to be

applicable. For general meshes, several procedures of construction of characteristic curves had been published [8, 16]. In our case, we have to bear in mind that it is necessary to discretize derivatives of the velocity field. As for the interpolation step, the same idea may adapt, based on geometrical arguments, as soon as there exists a structure of neighbouring elements. For instance, on conformal triangular or quadrangular meshes, we can think of a 6-point interpolation formula involving the 3 or 4 neighbours.

REFERENCES

[1] C. Bardos, M. Bercovier and O. Pironneau, *The vortex method with finite elements*, Math. Comp., **36** (1981), no. 153, 119–136.

[2] F. Boyer and P. Fabrie, "Éléments d'analyse pour l'étude de quelques modèles d'écoulements de fluides visqueux incompressibles", Springer-Verlag, Berlin, 2006.

[3] J. Burgers, *A mathematical model illustrating the theory of turbulence*, Adv. Appl. Mech., **1** (1948), 171–199.

[4] S. Dellacherie, *On a diphasic low Mach number system*, ESAIM: M2AN, **39** (2005), no. 3, 487–514.

[5] B. Després and F. Lagoutière, *Contact discontinuity capturing schemes for linear advection and compressible gas dynamics*, J. Sci. Comput., **16** (2001), no. 4, 479–524.

[6] J. Douglas Jr. and T. Russell, *Numerical methods for convection-dominated diffusion problems based on combining the method of characteristics with finite element or finite difference procedures*, SIAM J. Numer. Anal., **19** (1982), no. 5, 871–885.

[7] J. Douglas Jr., C.-S. Huang and F. Pereira, *The modified method of characteristics with adjusted advection*, Numer. Math., **83** (1999), no. 3, 353–369.

[8] G. Fourestey, "Simulation numérique et contrôle optimal d'interactions fluide incompressible / structure par une méthode de Lagrange-Galerkin d'ordre 2", Ph.D Thesis, École Nationale des Ponts et Chaussées, 2002.

[9] E. Godlewski and P.-A. Raviart, "Numerical Approximation of Hyperbolic Systems of Conservation Laws," Springer-Verlag, New-York, 1996.

[10] F. Holly and A. Preissmann, *Accurate calculation of transport in two dimensions*, J. Hydr. Div., **103** (1977), no. 11, 1259–1277.

[11] R. LeVeque, "Numerical Methods for Conservation Laws," Birkhäuser-Verlag, Basel, 1992.

[12] J. Marsden and A. Chorin, "A Mathematical Introduction to Fluid Mechanics," Springer-Verlag, New-York, 1979.

[13] S. Osher and J. Sethian, *Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations*, J. Comput. Phys., **79** (1988), 12–49.

[14] Y. Penel, "Étude théorique et numérique de la déformation d'une interface séparant deux fluides non-miscibles à bas nombre de Mach," Ph.D Thesis, Univ. Paris 13, (To appear).

[15] Y. Penel, S. Dellacherie and O. Lafitte, *Global solutions to the 1D Abstract Bubble Vibration model*, (Subm.).

[16] O. Pironneau, *On the transport-diffusion algorithm and its applications to the Navier-Stokes equations*, Numer. Math., **38** (1982), no. 3, 309–332.