Dipartimento di Ingegneria Meccanica e Aerospaziale

Lecture notes for the course

Numerical methods for compressible flows

Sergio Pirozzoli & Matteo Bernardini

Academic Year 2016/2017

Chapter 1

One-dimensional scalar conservation laws

1.1 Introduction

A scalar conservation law in a one-dimensional space reproduces the physical principle for which the rate of change of the integral of a given quantity over a fixed domain is equal to the net flux of that quantity through the boundaries; denoting with u the volumetric density of this conserved quantity, with φ its flux, and by considering the one-dimensional case, for which the control volume reduces to the interval $I = [a, b]$, the conservation principle can be expressed in mathematical form as

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b u \, \mathrm{d}x + \varphi[u(b)] - \varphi[u(a)] = 0, \qquad \forall \; a, b \tag{1.1}$$

where the total flux φ is given by the sum of a convective flux $f(u)$ and a diffusive flux

$$\varphi(u) = f(u) - \varepsilon \, \frac{\partial u}{\partial x}. \tag{1.2}$$

By applying the divergence theorem to equation (1.1), under the hypothesis that $u(x, t)$ is differentiable, one obtains the conservation equation in differential form

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = \varepsilon \, \frac{\partial^2 u}{\partial x^2}. \tag{1.3}$$

Equation (1.3) is a convection-diffusion equation, which (from the mathematical point of view) has a parabolic character (as the heat equation); it is relatively easy to show that the problem made from Eq (1.3) with the initial condition (IC)

$$u(x, 0) = u_0(x), \qquad -\infty < x < +\infty, \tag{1.4}$$

admits a unique solution, and such solution is continuous $\forall t > 0$, even though $u_0(x)$ is discontinuous.

Gasdynamic applications are mainly characterized by problems of pure convection, for which $\varepsilon \approx 0$, governed by equation of the type

$$\begin{cases} \dfrac{\partial u}{\partial t} + \dfrac{\partial f(u)}{\partial x} = 0 \\ u(x, 0) = u_0(x) \end{cases}, \qquad -\infty < x < +\infty, \quad t \geq 0. \tag{1.5}$$

The simplification step from equation (1.3) to (1.5) is not really straightforward, representing a classic problem of singular perturbation, where setting to zero the parameter ε leads to a change in the character of the differential equation, which becomes of hyperbolic nature. In such cases Eq. (1.5) loses its meaning. These generalized solutions are typically multiple, and it is needed to introduce additional constraints to re-establish the uniqueness of the solution. Nevertheless it is fundamental to remember that the hyperbolyc equation (1.5) derives from the more general (1.3), where a diffusive term is present. If we label as $u_\varepsilon(x,t)$ the solution of the differential problem (1.3) + (1.4) (unique and continuous for $\varepsilon \neq 0$), it is mandatory to guarantee that the solution $u(x,t)$ of the problem (1.5) satisfies

$$u(x,t) = \lim_{\varepsilon \to 0} u_\varepsilon(x,t). \tag{1.6}$$

Such solution is denoted as entropy solution.

1.2 Classic solutions - Characteristics Method

We call classic solutions of (1.5), solutions $u(x,t)$ everywhere continuous. Such solutions can be obtained by means of the characteristics method, described in the following. To that purpose it is useful to rewrite the conservation equation in quasi-linear form, by explicitly reporting the derivative of the convective flux with respect to u,

$$\frac{\partial u}{\partial t} + a(u)\,\frac{\partial u}{\partial x} = 0, \tag{1.7}$$

where $a(u) = f'(u)$ is denoted as propagation velocity of the signal or *speed of sound*.

A fundamental property of equation (1.7) is the existence of a particular class of curves, called *characteristics*, along which the solution is constant. Each particular characteristic Γ, satisfies by definition the condition

$$\left.\frac{\mathrm{d}x}{\mathrm{d}t}\right|_\Gamma = a[u(x,t)], \tag{1.8}$$

meaning that at each point along Γ, the slope of the tangent in the plane $x-t$ is equal to the local value of the speed of sound.

Computing the total differential of u along the generic characteristic Γ and using (1.8), (1.7), it is possible to obtain

$$\left.\frac{\mathrm{d}u}{\mathrm{d}t}\right|_\Gamma = \frac{\partial u}{\partial t} + \left.\frac{\mathrm{d}x}{\mathrm{d}t}\right|_\Gamma \cdot \frac{\partial u}{\partial x} = \frac{\partial u}{\partial t} + a(u)\,\frac{\partial u}{\partial x} \equiv 0, \tag{1.9}$$

which implies that u is constant along the Γ curve. It is interesting to observe that, since the speed of sound $a(u)$ is only a function of u, it is also constant along Γ. Therefore, the characteristics represent a family of straight lines, in general characterized by different slopes.

Such considerations provide a simple method to compute the solution $u(x,t)$ in a generic point $P(x,t)$ of the $x-t$ plane, given the initial conditions. To that purpose we consider figure 1.1.

Denoting as x_0 the intersection between the characteristic passing through P and the axis $t=0$, thanks to equation (1.9) we can write $u(x,t) = u_0(x_0)$. Since the slope of this line is $a(u)$, it will be $x_0 = x - a(u)\,t$, from which we obtain the condition

$$u(x,t) = u_0[x - a(u(x,t))t], \tag{1.10}$$

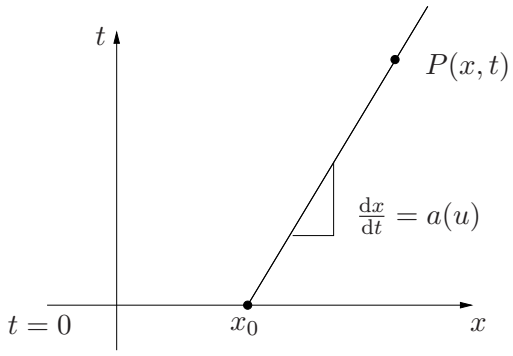


Figure 1.1: Geometrical properties of the characteristic curves.

This represents an algebraic equation that can be solved through iterative methods to find the solution u in P. By extending the same reasoning it is possible to find the solution to the whole $x - t$ plane.

1.3 Weak solutions and jump conditions

It is relatively easy to show that the procedure described in the previous section to find classic solutions by means of the characteristics method can lead to inconsistencies. Let consider for instance a case (see figure 1.2), for which the characteristics through two points $x_1 < x_2$ (at $t = 0$) converge at the same point x^* at the time t^*. At this point the solution provided by

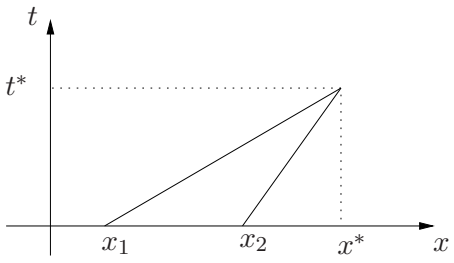


Figure 1.2: Intersection of characteristic lines.

the characteristic method is not longer univocally defined (single-valued), being at the same time $u(x^*, t^*) = u_0(x_1)$ and $u(x^*, t^*) = u_0(x_2)$, with $u_0(x_1)$ in general different from $u_0(x_2)$. Such condition occurs when

$$x^* = x_1 + a[u_0(x_1)]\, t^* = x_2 + a[u_0(x_2)]\, t^*, \tag{1.11}$$

implying

$$t^* = -\frac{x_2 - x_1}{a[u_0(x_2)] - a[u_0(x_1)]}. \tag{1.12}$$

The only possibility to avoid the intersection in a finite time $t^* > 0$ of the characteristics through the points x_1 and x_2 (with $x_1 < x_2$) is that $a[u_0(x_2)] \geq a[u_0(x_1)]$. A particular (trivial) case for which the intersection of characteristics never occurs is the linear advection

equation (hearinatfer LAE), for which $f(u) = c\,u$, with c constant. In this case $a(u) \equiv c$, and the characteristics lines are parallel.

Let be $a[u_0(x)]$ continuous and differentiable. Condition (1.12) can be written for two points at infinitesimal distance x and $x + \mathrm{d}x$, obtaining

$$t^*(x) = -\frac{1}{\frac{\mathrm{d}}{\mathrm{d}x}[a(u_0(x))]}. \tag{1.13}$$

The minimum t value for which the characteristics through the two points converge is obtained by considering the minim value of the critical times $t^*(x)$, for all points corresponding to the initial conditions, thus implying

$$T^* = \min_x t^*(x) = \frac{1}{\max_x \left\{-\frac{\mathrm{d}}{\mathrm{d}x}[a(u_0(x))]\right\}}. \tag{1.14}$$

As an example, we consider the Burgers equation, for which $f(u) = u^2/2$, with the initial condition $u_0(x) = -\sin(\pi x)$. From (1.14) we obtain

$$T^* = \frac{1}{\max_x \{\pi\, \cos(\pi x)\}} = \frac{1}{\pi}. \tag{1.15}$$

Starting from time T^*, defined by (1.14), it is not longer possible to build classic (single-valued) solutions of (1.5). There are some particular physical cases, characterized by hyperbolic equations, where multi-values solutions are plausible and can be accepted. However, in the cases of interest for gasdynamics, multi-values solutions cannot be accepted. It is then necessary to widen the class of the solutions for equation (1.5) including the possibility of having discontinuous solutions, provided they satisfy the conservation law written in integral form (1.1), here written for a pure convection equation (without diffusion)

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b u\ \mathrm{d}x = f[u(a)] - f[u(b)], \qquad \forall\ a, b, \tag{1.16}$$

where the continuity of $u(x,t)$ is not required, but only its integrability.

We now consider the case of a solution that satisfies equation (1.5) in the classic sense on both sides of a curve in the $x-t$ plane $x = x_s(t)$, across which u is discontinuous (see figure 1.3); let be also u_l and u_r the values assumed by u on the left and right side of the

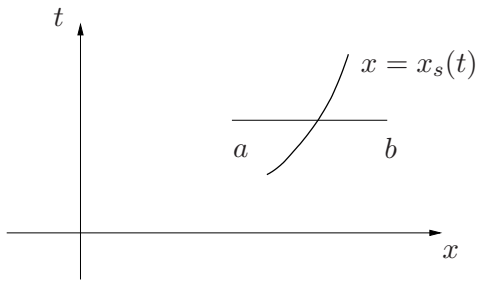


Figure 1.3: Jump relations across a curve in the $x-t$ plane.

discontinuity, respectively.

$$u_l(t) = \lim_{x \to x_s(t)^-} u(x,t), \tag{1.17}$$

$$u_r(t) = \lim_{x \to x_s(t)^+} u(x,t). \tag{1.18}$$

By choosing a and b in (1.16) in such a way that $a \leq x_s \leq b$ at time t, applying well-known theorems

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b u \, \mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}t}\int_a^{x_s(t)} u \, \mathrm{d}x + \frac{\mathrm{d}}{\mathrm{d}t}\int_{x_s(t)}^b u \, \mathrm{d}x = \tag{1.19}$$

$$= \int_a^{x_s(t)} \frac{\partial u}{\partial t} \, \mathrm{d}x + u_l \, \frac{\mathrm{d}x_s}{\mathrm{d}t} + \int_{x_s(t)}^b \frac{\partial u}{\partial t} \, \mathrm{d}x - u_r \, \frac{\mathrm{d}x_s}{\mathrm{d}t} = \tag{1.20}$$

$$= \int_a^{x_s(t)} \frac{\partial u}{\partial t} \, \mathrm{d}x + \int_{x_s(t)}^b \frac{\partial u}{\partial t} \, \mathrm{d}x + (u_l - u_r)\, s, \tag{1.21}$$

begin$s = \mathrm{d}x_s/\mathrm{d}t$ the propagation speed of the discontinuity at time t. Since equation (1.5) is satisfied on both sides of the discontinuity, we can replace $\partial u/\partial t$ with $-\partial f/\partial x$ and from (1.19) we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b u \, \mathrm{d}x = -f(u_l) + f[u(a)] - f[u(b)] + f(u_r) + (u_l - u_r)\, s. \tag{1.22}$$

By also using (1.16) we finally derive

$$s = \frac{f(u_r) - f(u_l)}{u_r - u_l} = \frac{[f]}{[u]}, \tag{1.23}$$

that allows to evaluate the local propagation speed of the discontinuity as a function of the jump (hereinafter denoted $[\bullet]$) of the conservative variable and values assumed by the flux function.

Equation (1.23) is called *jump relation* (or Rankine-Hugoniot) relation. In the case of the linear advection equation the jump relation assumes a particularly simple form,

$$s \equiv a, \tag{1.24}$$

implying that all the discontinuities propagates at the same velocity of the characteristics lines, equal to the speed of sound.

For the Burgers equation, (1.23) provides

$$s \equiv \frac{u_r^2/2 - u_l^2/2}{u_r - u_l} = \frac{1}{2}\,(u_l + u_r)\,, \tag{1.25}$$

that is equivalent to require that the speed of the discontinuity is the arithmetic mean of the two values assumed by the conservative variable at the left and right state.

Solutions of (1.5), that satisfy the conservation law in the classic sense everywhere, except in a finite number of discontinuity where the jump condition (1.23) holds are denoted as weak solutions. Such class of solutions is an extension of the classic one and it allows to avoid the problem of having multi-valued solution deriving from the application of the characteristics method. Indeed, it would be possible to demonstrate that, for a given initial condition, it is always possible to build a single-valued weak solution of the scalar conservation law.

Let consider for instance the following problem for the Burgers equation

$$\frac{\partial u}{\partial t} + u\,\frac{\partial u}{\partial x} = 0, \qquad u_0(x) = \begin{cases} 1 & x < 0 \\ 1 - x & 0 \le x \le 1 \\ 0 & x > 1 \end{cases}. \tag{1.26}$$

The classic solution of this problem (reported in figure 1.4a) is continuous and single-valued until time $T = 1$; starting from this time, there exists a region in the $x - t$ plane where the solution assumes three distinct values. For instance, three distinct characteristics Γ_1, Γ_2, Γ_3 cross the point $P(x^*, t^*)$, implying that in such a point the solution should assume three distinct values at the same time. However, we can easily verify that this problem also admits

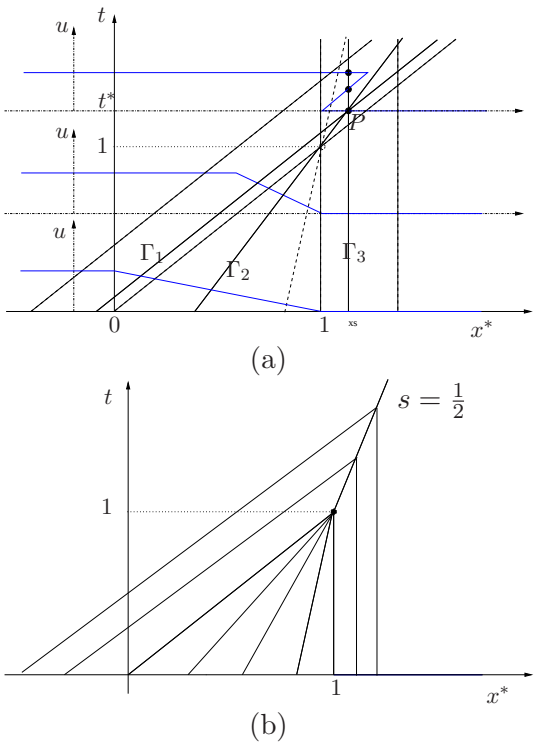


Figure 1.4: Solution of problem (1.26). (a) multi-valued solution computed with the characteristic method; (b) weak solution.

the following weak single-valued solution. (represented by figure 1.4b)

$$u(x,t) = \begin{cases} 1 & x < x_s(t) \\ 0 & x > x_s(t) \end{cases}, \tag{1.27}$$

where $x_s(t) = (1+t)/2$. This is an admissible solution because every point of the $x - t$ plane can be linked to a point at time $t = 0$ and the jump relation (1.23) is satisfied across the discontinuity, being

$$s = \frac{\mathrm{d}x_s}{\mathrm{d}t} = \frac{[f]}{[u]} = \frac{1/2 - 0}{1 - 0} = \frac{1}{2}. \tag{1.28}$$

1.4 Entropy solutions

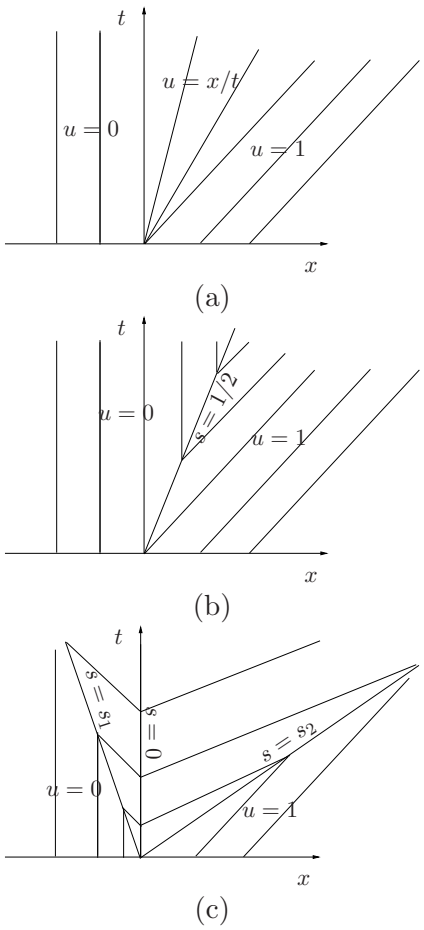


Figure 1.5: Solutions of the differential problem (1.29). (a) classic solution (1.30); (b) weak solution (1.31); (c) weak solution (1.32).

The class of weak solutions guarantees the possibility to build a single-valued solution for the scalar conservation law, which is not always possible in the classic sense. We now address the problem of the uniqueness of such solution. Let consider again the following problem for the Burgers equation

$$u_0(x) = \begin{cases} 0 & x < 0 \\ 1 & x > 0 \end{cases}, \tag{1.29}$$

It is imediate verify the the following solution

$$u(x,t) = \begin{cases} 0 & x \leq 0 \\ \dfrac{x}{t} & 0 \leq x \leq t \\ 1 & x \geq t \end{cases}, \tag{1.30}$$

reported in figure 1.5a is a classic (sinle-valued) solution of the problem under investigation.

However, it is possible to verify that also the following solution

$$u(x,t) = \begin{cases} 0 & x < t/2 \\ 1 & x > t/2 \end{cases}, \tag{1.31}$$

reported in figure 1.5b is a weak solution of the problem, being compatible with (1.29) and with the jump condition (1.23) ($s = 1/2$).

An other example of admissible solution for the problem under investigation is given by the following one parameter family solution, reported in figure 1.5c

$$u(x,t) = \begin{cases} 0 & x < s_1\, t \\ -\sigma & s_1\, t < x < s_2\, t \\ +\sigma & s_2\, t < x < s_3\, t \\ 1 & x > s_3\, t \end{cases}, \tag{1.32}$$

where $\sigma > 1$, $s_1 = -\sigma/2$, $s_2 = 0$, $s_3 = (1+\sigma)/2$.

The conclusion of this example is that, having extended the class of the possible solutions, including that with discontinuities, we now have the problem of guaranteeing the uniqueness of the solution. Infact, among all the (infinite) solutions of the problem (1.29) only one is admissible from the "physical" point of view, i.e. it satisfies the condition (1.6) and it is the limit for $\epsilon \downarrow 0$ of the family solutions (u_ϵ) of the scalar-diffusion equation (1.3). To select the unique correct solution (entropy solution) among all the possible weak solutions it is sufficient to satisfy the following condition

Oleinik entropy condition: The weak solution $u(x,t)$ is the entropy solution if, for each discontinuity, the following disequality holds

$$\frac{f(u) - f(u_l)}{u - u_l} \geq s \geq \frac{f(u) - f(u_r)}{u - u_r}, \quad \forall u : \min(u_l, u_r) \leq u \leq \max(u_l, u_r). \tag{1.33}$$

(1.33) expresses a simple geometric condition on the shape of the flux function in the interval between the values u_l and u_r (see figure 1.6) Since s represents the slope of the secant through u_l and u_r (see equation (1.23), the Oleinik condition implies that, if $u_l < u_r$ the graph of $f(u)$ must completely above the secant for all the values in the interval $[u_l, u_r]$. On the contrary, if $u_r < u_l$ the flux function must be completely below the secant. It is trivial to show that for flux functions without inflection points, convex ($f''(u) > 0 \ \forall u$, as for the Burgers equation), or concave ($f''(u) < 0 \ \forall u$), the Oleinik condition can be simplified and it is equivalent to require

$$f'(u_l) \geq s \geq f'(u_r), \tag{1.34}$$

as can be easily shown with geometrical considerations (see figure 1.7). It is important to notice the for the Burgers equation (and in general for equations with convex flux function) the entropy condition implies

$$u_l \geq u_r. \tag{1.35}$$

It is worth to notice that the condition (1.34) has a simple physical meaning, because it is equivalent to impose that the characteristics adjacent to the discontinuity converge inside the discontinuity itself (see figure 1.8a). If we take into account the interpretation of the characteristics, as lines along with the information propagates (being u = constant), the Oleinik condition implies that the information cannot emerge from a discontinuity, which

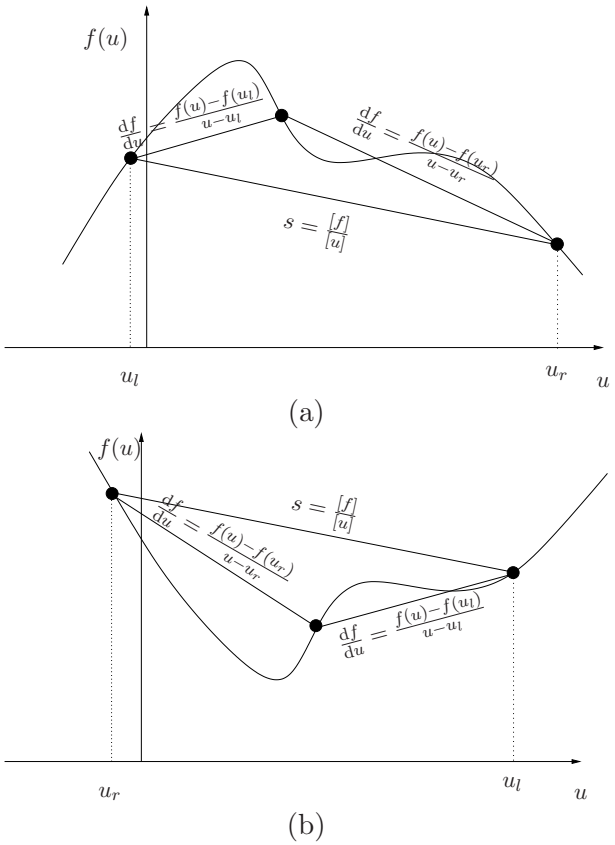


Figure 1.6: Representation of the entropy condition (1.33). (a) $u_l \leq u_r$; (b) $u_l \geq u_r$.

represents a sort of event horizon. This is consistent with the causality principle, because otherwise there would exist characteristics that cannot be linked to the initial condition (as can be verified in figure 1.8b.) Discontinuities that satisfy both the jump relation and the Oleinik entropy condition are called *shocks*.

Going back to problem (1.29), it easy verify that the only admissible solution is the classic (1.30), because both (1.31) and (1.32) do not satisfy (1.35) across all the discontinuities. In general it is possible to demonstrate that the problem (1.5) admits a unique (single-valued) weak solution for which all the discontinuities verify the Oleinik entropy condition and such solution coincides with the entropy solution defined by (1.6).

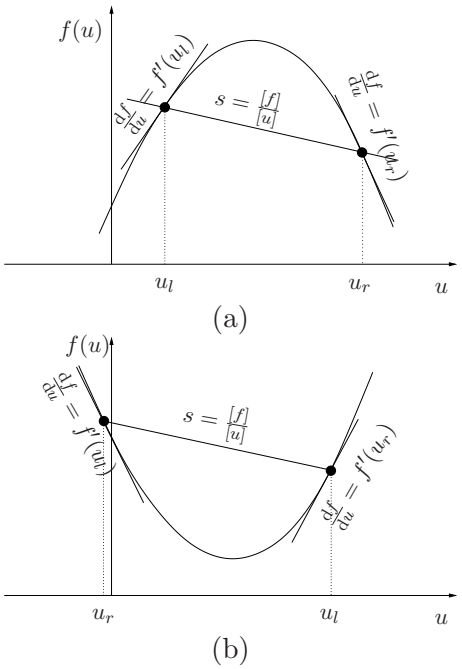


Figure 1.7: Entropy condition for a concave (a) and convex (b) flux function.

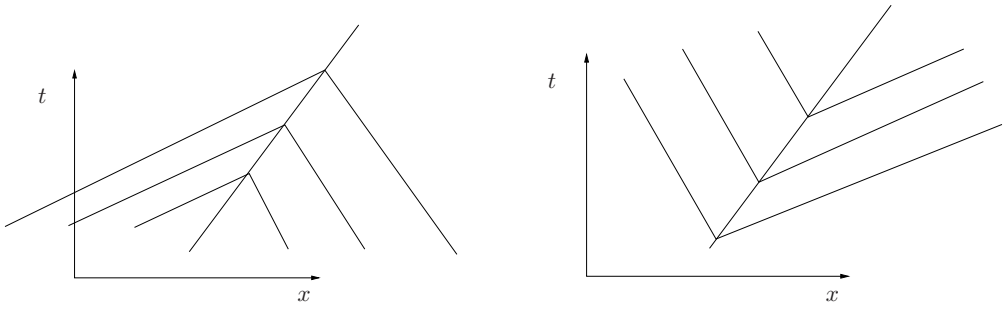


Figure 1.8: Representation of entropy and non-entropy solutions in the $x - t$ plane.

Chapter 2

Basic numerical methods for scalar hyperbolic equations

In this chapter we introduce fundamental numerical methods for the solution of pure convection, scalar, one-dimensional problems, whose mathematical properties have been highlighted in the previous section.

$$\begin{cases} \dfrac{\partial u}{\partial t} + \dfrac{\partial f(u)}{\partial x} = 0 \\ u(x,0) = u_0(x) \end{cases}, \qquad -\infty < x < +\infty, \quad t \geq 0. \tag{2.1}$$

To that purpose we start by considering the linear advection equation (LAE), obtained by assuming $f(u)0c\,u$, with c constant

$$\begin{cases} \dfrac{\partial u}{\partial t} + c\,\dfrac{\partial u}{\partial x} = 0 \\ u(x,0) = u_0(x) \end{cases}. \tag{2.2}$$

Equation (2.2) offers remarkable advantages for the analysis because it has a simple exact solution, given by

$$u(x,t) = u_0(x - c\,t). \tag{2.3}$$

To solve this equation from the numerical point of view, we introduce a discretization of the $(x-t)$ plane, with an equally spaced mesh in both the directions x and t. Let be h and k the mesh spacing and k the time step (see figure 2.1), assumed to be constant. We denote as *mesh nodes* the points with coordinates (x_j, t^n), with

$$x_j = j\,h, \qquad j = \ldots, -1, 0, 1, 2, \ldots \tag{2.4}$$

$$t^n = n\,k, \qquad n = 0, 1, 2, \ldots \tag{2.5}$$

For the sake of clarity it is also useful to introduce *intermediate nodes*, denoted by fractional indices

$$x_{j+1/2} = x_j + \frac{h}{2} = (j + \frac{1}{2})\,h. \tag{2.6}$$

Let be $v(x,t) \approx u(x,t)$ the approximate solution of equation (2.2), obtained with a generic numerical method. In the *finite difference* method (FD), we denote as U_j^n the discrete values of the approximate solution $v(x,t)$ at the mesh node x_j and time t^n

$$U_j^n = v(x_j, t^n). \tag{2.7}$$

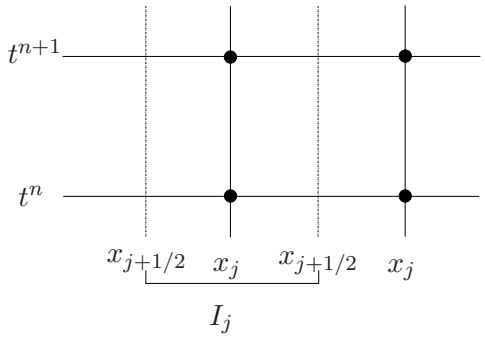


Figure 2.1: Nomenclature for the discretization of the $x-t$ plane.

In the *finite volume* method (FV) the unknowns represent the cell averages values of the approximate solution v. To that purpose we define as *computational cell* the interval (in 1D)

$$I_j = [x_{j_1/2}; x_{j+1/2}], \tag{2.8}$$

and the unknowns are then defined as

$$U_j^n = \frac{1}{h}\int_{x_{j-1/2}}^{x_{j+1/2}} v(x,t^n)\,\mathrm{d}x. \tag{2.9}$$

In FV the intermediate nodes are referred as *intercells*.

For most part of the analysis in the following we will not distinguish between FV and FD and we will use the same symbol U_j^n to denote both the discrete values of the approximate solution and its cell averages.

A simple and immediate technique to build discrete approximations of (2.2) is to replace the partial derivatives with finite difference formulas for the approximate solution v. Several different schemes can be generated by considering central or upwind (forward and backward) approximations and combining different discretizations for the time and spatial derivatives. A simple example is obtained by selecting a forward approximation for the time derivative (Forward-Time) at node (x_j, t^n)

$$\mathrm{FT}: \frac{\partial v}{\partial t}(x_j,t^n) \simeq \frac{v(x_j,t^n+k)-v(x_j,t^n)}{k} = \frac{U_j^{n+1}-U_j^n}{k}, \tag{2.10}$$

and a central one for the space derivative (Central-Space).

$$\mathrm{CS}: \frac{\partial v}{\partial x}(x_j,t^n) \simeq \frac{v(x_j+h,t^n)-v(x_j-h,t^n)}{2\,h} = \frac{U_{j+1}^n-U_{j-1}^n}{2h}. \tag{2.11}$$

By using these approximations in (2.2) the following finite difference scheme (known as FTCS) is obtained

$$U_j^{n+1} = U_j^n - \frac{ck}{2h}\,(U_{j+1}^n - U_{j-1}^n) = U_j^n - \frac{\sigma}{2}\,(U_{j+1}^n - U_{j-1}^n), \tag{2.12}$$

being $\sigma = ck/h$ the so-called *Courant number*. By considering a central approximation for the time derivative (Central-Time)

$$\mathrm{CT}: \frac{\partial v}{\partial t}(x_j,t^n) \simeq \frac{v(x_j,t^{n+k})-v(x_j,t^{n-k})}{2k} = \frac{U_j^{n+1}-U_j^{n-1}}{2k}, \tag{2.13}$$

and combining it with (2.11), we obtain the Leapfrog (or CTCS) scheme

$$U_j^{n+1} = U_j^{n-1} - \sigma\,(U_{j+1}^n - U_{j-1}^n). \tag{2.14}$$

It is also possible to consider backward and forward approximations for the spatial derivatives, (Backward-Space and Forward-Space, respectively)

$$\text{BS} : \frac{\partial v}{\partial x}(x_j, t^n) \simeq \frac{v(x_j, t^n) - v(x_{j-h}, t^n)}{h} = \frac{U_j^n - U_{j-1}^n}{h}, \tag{2.15}$$

$$\text{FS} : \frac{\partial v}{\partial x}(x_j, t^n) \simeq \frac{v(x_{j+h}, t^n) - v(x_j, t^n)}{h} = \frac{U_{j+1}^n - U_j^n}{h}. \tag{2.16}$$

Combining expressions (2.15) and (2.16) with (2.10) we obtain the following schemes

$$\text{FTBS(UW}^+\text{)} : U_j^{n+1} = U_j^n - \sigma\,(U_j^n - U_{j-1}^n) \tag{2.17}$$

$$\text{FTFS(UW}^-\text{)} : U_j^{n+1} = U_j^n - \sigma\,(U_{j+1}^n - U_j^n) \tag{2.18}$$

A simple modification of the FTCS scheme (denoted as Lax-Friedrichs scheme), needed for stability reasons (as we will see in the following), can be obtained replacing U_j^n in (2.12), with the arithmetic average of the adjacent nodes

$$\text{LF} : U_j^{n+1} = \frac{1}{2}(U_{j+1}^n + U_{j-1}^n) - \frac{\sigma}{2}(U_{j+1}^n - U_{j-1}^n). \tag{2.19}$$

2.1 Schemes based on the characteristics method

A greater level of complexity with respect to the schemes developed in the previous section is obtained by designing schemes based on the characteristics method, rather than on the direct discretization of the derivatives in equation (2.2). For the linear advection equation the characteristics are straight, parallel lines with slope c, along with the solution is constant. We can imagine to exploit this property to impose (see figure 2.2)

$$U_j^{n+1} = v(x^*, t^n), \tag{2.20}$$

x^* being the intersection between the characteristic through (x_j, t^{n+1}) with $t = t^n$. Since we

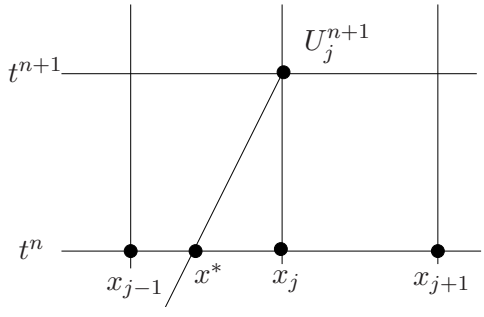


Figure 2.2: Discretization based on the characteristics method ($c > 0$).

know the slope of the characteristic we have

$$x^* = x_j - c\,k. \tag{2.21}$$

The problem is to compute the (interpolated) value of v in x^*, starting from the knowledge of the nodal values at time $t = t^n$. The simpler idea is to built an interpolation based on the adjacent nodes, x_j and x_{j-1}, which takes into account the direction of the propagation of the signal (in this case we are assuming $c > 0$). By considering the first order polynom

$$p(x) = a + b\,(x - x_j), \tag{2.22}$$

and imposing the interpolation conditions $p(x_j) = U_j^n$, $p(x_{j+1}) = U_{j+1}^n$, we get $a = U_j$, $b = (U_j^n - U_{j-1}^n)/h$. The resulting scheme is

$$U_j^{n+1} = p(x^*) = p(x_j - c\,k) = U_j^n + \frac{U_j^n - U_{j-1}^n}{h}\,(x_j - c\,k - x_j) = U_j^n - \sigma\,(U_j^n - U_{j-1}^n), \tag{2.23}$$

that coincides with the scheme FTBS reported in (2.17).

Ne now assume a central interpolation, based on nodes $j-1$, j, $j+1$. By considering the second order polynom

$$p(x) = a + b\,(x - x_j) + c\,(x - x_j)^2, \tag{2.24}$$

and imposing the interpolation conditions

$$\begin{aligned} p(x_{j-1}) &= U_{j-1}^n, \\ p(x_j) &= U_j^n, \\ p(x_{j+1}) &= U_{j+1}^n, \end{aligned}$$

we can derive $a = U_j^n$, $b = (U_{j+1}^n - U_{j-1}^n)/2h$, $c = (U_{j+1}^n - 2U_j^n + U_{j-1}^n)/2h^2$, and obtain the following (important) scheme (called Lax-Wendroff)

$$LW: U_j^{n+1} = U_j^n - \frac{\sigma}{2}(U_{j+1}^n - U_{j-1}^n) + \frac{\sigma^2}{2}(U_{j+1}^n - 2U_j^n + U_{j-1}^n). \tag{2.25}$$

It is worth to notice that equation (2.25) exactly reproduces the FTCS scheme, except for the presence of a further term, that can be interpreted as a central discretization of a second derivative. As we will see, this term is responsible for the stability of the scheme.

By considering a one-sided interpolation based on nodes $j-2$, $j-1$ e j, we easily obtain the Beam-Warming scheme

$$\mathrm{BW}^+: U_j^{n+1} = U_j^n - \frac{\sigma}{2}(3U_j^n - 4U_{j-1}^n + U_{j-2}) + \frac{\sigma^2}{2}(U_j^n - 2U_{j-1}^n + U_{j-2}). \tag{2.26}$$

The superscript + indicates that this scheme is specifically designed for problems characterized by a positive propagation speed ($c > 0$), for which the characteristics are represented in figure 2.2. In the case of $c < 0$, the most natural one-sided choice would be symmetrical with respect to the previous one and based on nodes j, $j+1$ e $j+2$. The resulting scheme would be

$$\mathrm{BW}^-: U_j^{n+1} = U_j^n + \frac{\sigma}{2}(3U_j^n - 4U_{j+1}^n + U_{j+2}) + \frac{\sigma^2}{2}(U_j^n - 2U_{j+1}^n + U_{j+2}). \tag{2.27}$$

Further generalizations are clearly possible and can be built by extending the set of nodes involved in the interpolation procedure for $v(x^*)$. For instance by considering $j-2$, $j-1$, j, $j+1$ we would get a scheme denoted as UW3 (third-order upwind scheme).

It is also possible to consider a linear combination of some of the previous schemes. For instance we can take the arithmetic average of the Lax-Wendroff and Beam-Warming$^+$ to obtain a new scheme, called Fromm. In the following table a survey of the schemes previously derived is reported. All the schemes considered until now are

	Finite difference scheme
FTCS	$U_j^{n+1} = U_j^n - \frac{\sigma}{2}\left(U_{j+1}^n - U_{j-1}^n\right)$
FTBS (UW$^+$)	$U_j^{n+1} = U_j^n - \sigma\left(U_j^n - U_{j-1}^n\right)$
FTFS (UW$^-$)	$U_j^{n+1} = U_j^n + \sigma\left(U_j^n - U_{j+1}^n\right)$
LF	$U_j^{n+1} = \frac{1}{2}\left(U_{j+1}^n + U_{j-1}^n\right) - \frac{\sigma}{2}\left(U_{j+1}^n - U_{j-1}^n\right)$
LW	$U_j^{n+1} = U_j^n - \frac{\sigma}{2}\left(U_{j+1}^n - U_{j-1}^n\right) + \frac{\sigma^2}{2}\left(U_{j+1}^n - 2U_j^n + U_{j-1}^n\right)$
BW$^+$	$U_j^{n+1} = U_j^n - \frac{\sigma}{2}\left(3U_j^n - 4U_{j-1}^n + U_{j-2}^n\right) + \frac{\sigma^2}{2}\left(U_j^n - 2U_{j-1}^n + U_{j-2}^n\right)$
BW$^-$	$U_j^{n+1} = U_j^n + \frac{\sigma}{2}\left(3U_j^n - 4U_{j+1}^n + U_{j+2}^n\right) + \frac{\sigma^2}{2}\left(U_j^n - 2U_{j+1}^n + U_{j+2}^n\right)$
FROMM$^+$	$U_j^{n+1} = U_j^n - \frac{\sigma}{4}\left(U_{j+1}^n + 3U_j^n - 5U_{j-1}^n + U_{j-2}^n\right)$ $+\frac{\sigma^2}{4}\left(U_{j+1}^n - U_j^n - U_{j-1}^n + U_{j-2}^n\right)$
FROMM$^-$	$U_j^{n+1} = U_j^n + \frac{\sigma}{4}\left(U_{j-1}^n + 3U_j^n - 5U_{j+1}^n + U_{j+2}^n\right)$ $+\frac{\sigma^2}{4}\left(U_{j-1}^n - U_j^n - U_{j+1}^n + U_{j+2}^n\right)$
UW3$^+$	$U_j^{n+1} = U_j^n - \frac{\sigma}{6}\left(2U_{j+1}^2 + 3U_j^n - 6U_{j-1}^n + U_{j-2}^n\right)$ $+\frac{\sigma^2}{2}\left(U_{j-1}^n - 2U_j^n + U_{j+1}^n\right) + \frac{\sigma^3}{6}\left(U_{j-2}^n - 3U_{j-1}^n + 3U_{j^n} - U_{j+1}^n\right)$
UW3$^-$	$U_j^{n+1} = U_j^n + \frac{\sigma}{6}\left(2U_{j-1}^2 + 3U_j^n - 6U_{j+1}^n + U_{j+2}^n\right)$ $+\frac{\sigma^2}{2}\left(U_{j-1}^n - 2U_j^n + U_{j+1}^n\right) - \frac{\sigma^3}{6}\left(U_{j+2}^n - 3U_{j+1}^n + 3U_{j^n} - U_{j-1}^n\right)$

Table 2.1: *fully discrete* finite difference schemes for the linear convection equation.

- (*fully discrete*)
- *explicit*, i.e. the solution at time $n+1$ can be directly computed on the basis of the solution at the previous time levels;
- *linear*, i.e. the right hand side (RHS) of the schemes is a linear combination of the nodal values U_j^n;
- *two levels* (except Leapfrog), i.e. only two time levels (n and $n+1$) are involved in the discretization.

For each scheme it is instructive to represent the set of nodes involved in the discretization process, denoted as computational stencil. Figure 2.3 shows those of the various schemes. It is important to highlight the fundamental difference between schemes with a symmetric stencil (defined as *central*) and those with an asymmetric one (*upwind schemes*). The latter are characterized by two different versions, denoted as + (asymmetry on the left) or − (asymmetry on the right).

In the following we will consider only explicit, two-levels schemes; such type of scheme can be symbolically written as

$$U^{n+1} = \mathrm{H}\,(U^n), \tag{2.28}$$

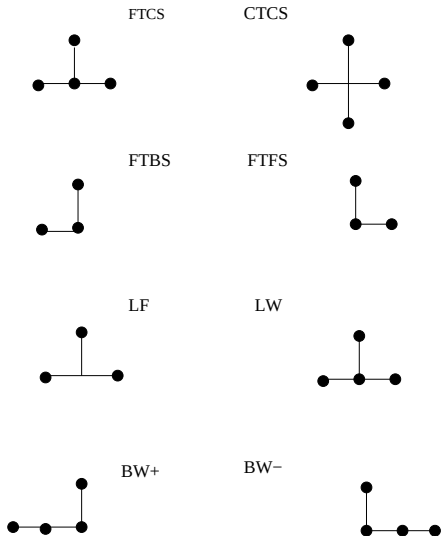


Figure 2.3: Stencil associated to various finite difference schemes.

where U^n is the vector of the nodal unknowns at time n and H the discrete operator (generally linear) for time advancement. For the schemes previously considered

$$U^{n+1} = \mathrm{H}\, U^n \quad \rightarrow \quad (\mathrm{H}\, U^n)_j = \sum_{l=-p}^{q} a_l\, U^n_{j+l} \tag{2.29}$$

where the coefficients a_l depend on the particular scheme under consideration. For instance, for Lax-Wendroff (see equation (2.25))

$$a_{-1} = \frac{\sigma}{2} + \frac{\sigma^2}{2}, \quad a_0 = 1 - \sigma^2, \quad a_{+1} = -\frac{\sigma}{2} + \frac{\sigma^2}{2}, \tag{2.30}$$

and the other coefficients are zero.

In the following sections we analyze the main properties of the various schemes derived, by considering the concepts if accuracy, stability and convergence.

2.2 Accuracy

The concepts of *consistency* (and *accuracy*) are related to the formal correspondence between the conservation equation and the finite difference scheme resulting from the discretization process, when the mesh spacing decreases ($h, k \downarrow 0$). Let consider the generic differential operator $\mathcal{L}$ (for the linear advection equation, $\mathcal{L} = \partial/\partial t + c\,\partial/\partial x$). The corresponding differential equation can be symbolically written as

$$\mathcal{L}\, u = 0, \tag{2.31}$$

where $u(x,t)$ is the exact solution of the PDE (partial differential equation). Discretizing the derivative operators in $\mathcal{L}$ (as shown in the previous section) we have a finite difference

scheme that involves only the nodal values of the approximate solution v. For instance, for the linear advection equation discretized with the FTCS scheme we have (see (2.12))

$$\frac{v(x,t+k)-v(x,t)}{k}+\frac{a}{2\,h}\left(v(x+h,t)-v(x-h,t)\right)=0, \tag{2.32}$$

that can be represented in compact form as

$$\mathcal{L}^*\,v=0 \tag{2.33}$$

having introduced the discretized differential operator $\mathcal{L}^*$. It is clear that $\mathcal{L}^*$ depends not only on $\mathcal{L}$ but also on the type of discretization applied. By definition, equations (2.31) and (2.33) are exactly satisfied by u and v, respectively. We define truncation error $\varepsilon_T(u)$ the residual obtained applying the discretized differential operator $\mathcal{L}^*$ to the exact solution u. Assuming that both the operators $\mathcal{L}$ and $\mathcal{L}^*$ are linear, and exploiting equation (2.31) we obtain

$$\varepsilon_T(u)=\mathcal{L}^*\,u=(\mathcal{L}^*-\mathcal{L})\,u. \tag{2.34}$$

We say that the scheme defined by (2.33) is consistent if

$$\lim_{h\downarrow 0,\,k/h=\lambda}\varepsilon_T=0, \tag{2.35}$$

being the limit considered at constant *mesh ratio* $\lambda\equiv k/h=O(1)$. Furthermore, the scheme is *formally accurate at order* r if the truncation error is proportional to the r-power of the mesh spacing, i.e.

$$\varepsilon_T(u)\overset{h\sim 0}{\sim}h^r. \tag{2.36}$$

By considering the residual obtained by plugging the approximate solution v in the differential equation, we have

$$\mathcal{L}\,v=(\mathcal{L}-\mathcal{L}^*)\,v=-(\mathcal{L}^*-\mathcal{L})\,v=-\varepsilon_T(v). \tag{2.37}$$

From inspection of (2.37) we can observe that the approximate solution v does not satisfy exactly the conservation equation; however it exactly satisfies a *modified equation*

$$\mathcal{L}\,v=-\varepsilon_T(v), \tag{2.38}$$

where the additional term at the right-hand-side is related to the truncation error of the numerical scheme. For instance, let consider the FTCS scheme applied to LAE. By expanding u in Taylor series around (x_j,t^n) we have

$$u(x_j,t^n+k)=u(x_j,t^n)+k\,u_t(x_j,t^n)+\frac{k^2}{2}u_{tt}(x_j,t^n)+O(k^3), \tag{2.39}$$

$$u(x_j\pm h,t^n)=u(x_j,t^n)\pm h\,u_x(x_j,t^n)+\frac{h^2}{2}u_{xx}(x_j,t^n)\pm\frac{h^3}{6}u_{xxx}(x_j,t^n)+O(h^4). \tag{2.40}$$

By replacing in (2.32), and exploiting the identity $u_t+a\,u_x=0$, we obtain the truncation error associated to the finite difference scheme

$$\varepsilon_T(u)=\mathcal{L}^*\,u=\ldots=\frac{k}{2}\,u_{tt}+O(h^2,k^2). \tag{2.41}$$

Differentiating with respect to x and t the linear advection equation

$$\begin{cases} u_{tt} + c\, u_{xt} = 0 \\ u_{tx} + c\, u_{xx} = 0 \end{cases} \quad \rightarrow \quad u_{tt} = c^2\, u_{xx}, \tag{2.42}$$

we can rewrite the truncation error as

$$\varepsilon_T(u) = \frac{\lambda h}{2}\, c^2\, u_{xx} + O(h^3) = \frac{h\sigma^2}{2\lambda}\, u_{xx} + O(h^2). \tag{2.43}$$

On the basis of the definitions (2.35) and (2.36), wen can state that the FTCS scheme is consistent and formally accurate at the first order and the modified equation associated to the scheme is

$$v_t + c\, v_x = -\frac{h\sigma^2}{2\lambda}\, v_{xx} + O(h^2). \tag{2.44}$$

It is important to notice the coefficient in front to the second derivative v_{xx} is negative, and thus leads to an anti-diffusive behavior of the numerical scheme. Indeed, as we will see in the next section, the FTCS scheme is useless in practice because it leads to unstable solutions. The first term in the expansion of the truncation error for the various schemes derived in

Schema	$\varepsilon_T(u)$
FTCS	$\frac{\sigma a}{2}\, h\, u_{xx}$
UW$^+$	$\frac{\sigma(\sigma-1)}{2\lambda}\, h\, u_{xx}$
LF	$\frac{-(1-\sigma^2)}{2\lambda}\, h\, u_{xx}$
LW	$\frac{(1-\sigma^2)a}{6}\, h^2\, u_{xxx}$
BW$^+$	$-\frac{(2-3\sigma+\sigma^2)a}{6}\, h^2\, u_{xxx}$
FROMM$^+$	$\frac{(1-3\sigma+2\sigma^2)a}{12}\, h^2\, u_{xxx}$
UW3$^+$	$-\frac{(2-\sigma-2\sigma^2+\sigma^3)a}{24} h^3\, u_{xxxx}$

Table 2.2: Truncation error for various finite difference schemes.

the previous sections is reported in table 2.2. Note that, for upwind schemes of type $-$, it is sufficient to replace $\sigma \rightarrow -\sigma$.

2.3 Numerical diffusion and dispersion

In order to predict the qualitative behavior of the numerical solutions obtained by different numerical schemes, it is instructive to analyze the evolution of the solution of the associated modified equations. To that purpose it is worth to notice that the dominant terms in all the various schemes presented in the previous sections (except for UW3$^{\pm}$), can be recast in the form $-\varepsilon_T(u) = \nu u_{xx} + \beta u_{xxx}$, and the corresponding modified equation results

$$v_t + c\, v_x = \nu\, v_{xx} + \beta\, v_{xxx}, \tag{2.45}$$

where the coefficients ν, β are functions of c, h, k. Let consider a solution for equation (2.45) in the form of a sinusoidal monochromatic wave, with wavenumber $w = 2\pi/\lambda$

$$v(x,t) = \hat{v}(t)\, e^{iwx}, \tag{2.46}$$

$\hat{v}(t)$ being the complex amplitude of the solution at time t. Be replacing in equation (2.45) we get

$$\frac{\mathrm{d}\hat{v}}{\mathrm{d}t} + i\,c\,w\,\hat{v} = -\nu\,w^2\hat{v} - \beta\,i\,w^3\hat{v} \quad \rightarrow \quad \hat{v}(t) = \hat{v}(0)\,e^{-\nu\,w^2\,t}\,e^{-i\,w(c+\beta\,w^2)\,t}, \tag{2.47}$$

and then

$$v(x,t) = \hat{v}_0\,e^{-\nu\,w^2\,t}\,e^{i\,w[x-(c+\beta\,w^2)\,t]}. \tag{2.48}$$

In the case of a pure linear advection equation we would have

$$v(x,t) = \hat{v}_0\,e^{i\,w(x-c\,t)}. \tag{2.49}$$

By comparing equation (2.47) with (2.49) it is possible to notice that, while for the exact solution of LAE $|\hat{v}(t)| = |\hat{v}(0)| = \text{cost.}$, the solution of a finite difference scheme is characterized by an exponential damping or amplification of $|\hat{v}(t)|$, associated with the coefficient ν. The effect of such coefficient is thus analogous to that of a term representing physical diffusion and the higher is the wavenumber w, the faster is the decay/amplification of the complex amplitude. We point out that, in the present case, this effect is a consequence of the discretization process and for this reason the coefficient ν is denoted as *numerical viscosity*.

The presence of the other coefficient $\beta \neq 0$ implies that the propagation speed associated with a particular wavenumber w is $\tilde{c} = c + \beta\,w^2$, and not identically equal to c. The effect of the third derivative is that of producing a dispersion of the harmonics of the solution due to the approximation introduced by the discretization process. Such effect is known as numerical dispersion and the coefficient β is called coefficient of *numerical dispersion*.

2.4 Linear stability: Von Neumann analysis

In order to apply a finite difference scheme it is necessary that a (small) error introduced in the solution at the initial time is not amplified exponentially by the computation. This requirements is usually denoted as *stability*.

Let consider a generic explicit, two-levels, linear, finite difference scheme

$$U^{n+1} = \mathrm{H}\,U^n, \quad U_j^{n+1} = \sum_{l=-p}^{q} a_l\,U_{j+l}^n. \tag{2.50}$$

We define a *local error* for the solution obtained by the scheme (related to both inaccuracy in the initial conditions or round-off errors) as the difference at node j and time level n between the value of the approximate solution and that of the solution of (2.50) obtained without numerical errors. [1]

$$\varepsilon_j^n = U_j^n - \overline{U}_j^n. \tag{2.51}$$

The finite difference scheme is stable if the error ε_j^n is not amplified in time. Since by definition $U^{n+1} = \mathrm{H}\,U^n$, $\overline{U}^{n+1} = \mathrm{H}\,\overline{U}^n$, and given the linearity of the scheme defined by H, the vector ε^n, containing the nodal values of the local error at time level n obeys the relation

$$\varepsilon^{n+1} = \mathrm{H}\,\varepsilon^n. \tag{2.52}$$

[1] Remark $\overline{U}_j^n \neq u(x_j, t^n)$, because it is not the solution of the differential equation but the solution of the finite difference scheme assuming the absence of numerical errors!

Imposing that the error must not increase in time is equivalent to impose that the approximate solution $v(x,t)$ must not grow.

In the classic linear stability analysis of finite difference schemes (introduced by Von Neumann) we consider the linear advection equation (2.2) as model equation, and we consider the evolution of an initial condition given by a monochromatic, sinusoidal wave, with wavelength λ (i.e. wavenumber $w = 2\pi/\lambda$) and complex amplitude $\hat{v}_0$

$$v_0(x) = \hat{v}_0\, e^{iwx}, \tag{2.53}$$

We also consider a uniform mesh, with spacing h and time step k. It is worth to remind that for the Nyquist theorem, the minimum wavelength that can be represented by the computational mesh (λ_{min}) is twice the spacing h, and the corresponding (maximum) wavenumber is $w_{max} = \pi/h$. The exact solution of the linear advection equation at time $T = n\,k$, with initial condition given by (2.53) is

$$u(x,T) = \hat{u}_n\, e^{iwx}, \quad \hat{u}_n = \hat{v}_0\, e^{-in\sigma\varphi}, \tag{2.54}$$

where $\varphi \equiv wh$ is the so called *reduced wavenumber*. The reduced wavenumber φ is an important quantity related to the number of mesh nodes used to discretize a single wavelength (points-per-wavelength, $PPW = \lambda/h = 2\pi/\varphi$), and then to the existence of a maximum wavenumber that is possible to resolve on a mesh with finite spacing, $|\varphi| \in (0, \pi]$ (see the example reported in figure 2.4, corresponding to PPW= 4, $\varphi = \pi/2$). Since the numerical solution satisfies the equation (2.50), and considering that

$$U_j^n = \hat{v}_n\, e^{iwx_j} = \hat{v}_n\, e^{iwjh} = \hat{v}_n\, e^{ij\varphi}, \tag{2.55}$$

we have

$$U_j^{n+1} = \hat{v}_{n+1}\, e^{ij\varphi} = \sum_{l=-p}^{q} a_l\, U_{j+l}^n = \sum_{l=-p}^{q} a_l\, \hat{v}_n\, e^{i(j+l)\varphi} \rightarrow \hat{v}_{n+1} = \hat{v}_n \sum_{l=-p}^{q} a_l\, e^{il\varphi}. \tag{2.56}$$

We can introduce the *amplification factor* of the finite difference scheme, defined as

$$g(\varphi) = \sum_{l=-p}^{q} a_l\, e^{il\varphi}, \tag{2.57}$$

and the evolution of the complex amplitude of the numerical solution can be written as

$$\hat{v}_{n+1} = g(\varphi)\, \hat{v}_n. \tag{2.58}$$

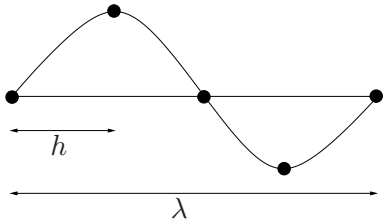


Figure 2.4: Discretization of a sinusoidal wave with 4 PPW.

Note that the amplification factor of a numerical scheme is a complex quantity, depending on the coefficients a_l of the scheme and on the reduced wavenumber. Since v is a real function, is easy to show that $g(-\varphi) = g^*(\varphi)$.

By considering the modulus of equation (2.58) we have

$$|\hat{v}_{n+1}| = |g(\varphi)|\,|\hat{v}_n|, \tag{2.59}$$

from which it is apparent that the numerical solution v (and the error) remain bounded for $t \uparrow \infty$, only if

$$|g(\varphi)| \leq 1 \qquad \forall \varphi \in (0, \pi]. \tag{2.60}$$

Relation (2.60) defines a stability region in the complex plane of the variable g, given by the internal region of the unit circle centred in the origin (see the next examples and figures).

FTCS scheme

Let consider the FTCS scheme, reported in (2.12). According to notation (2.50) we have $a_1 = \sigma/2$, $a_0 = 1$, $a_1 = -\sigma/2$. From (2.57), the amplification factor of the scheme results

$$g(\varphi) = \frac{\sigma}{2} e^{-i\varphi} + 1 - \frac{\sigma}{2} e^{i\varphi} = 1 - i\,\sigma\,\sin\varphi. \tag{2.61}$$

When φ varies, (2.61) defines a segment of a straight line that is external to the stability region, as shown in figure 2.5. Therefore, the FCTS scheme, as previously mentioned, is *unconditionally unstable* and useless from practical applications.

FTBS scheme

The FTBS scheme (reported in (2.17)) is defined by $a_{-1} = \sigma$, $a_0 = 1 - \sigma$. The amplification factor is

$$g(\varphi) = (1 - \sigma) + \sigma\, e^{-i\,\varphi}. \tag{2.62}$$

When φ varies, (2.62) describes a circle in the complex plane centred in $(1 - \sigma, 0)$ with radius $|\sigma|$. Depending on the value of σ we can have three distinct situations, represented in figure 2.6. From the analysis it follows that the FTBS scheme is stable when $0 \leq \sigma \leq 1$

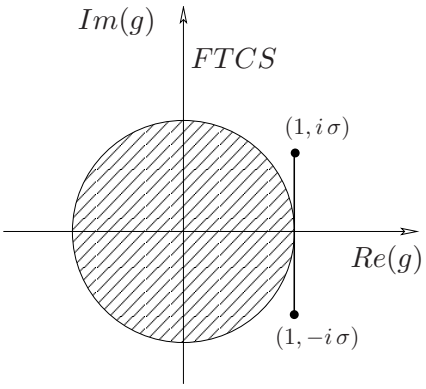


Figure 2.5: Stability diagram for the FTCS scheme.

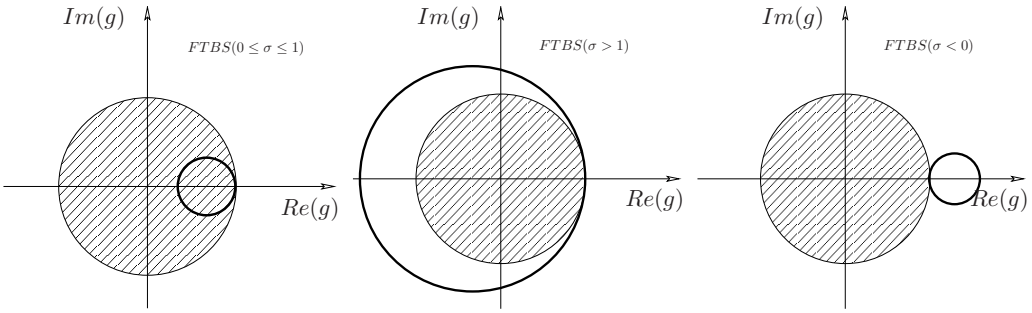


Figure 2.6: Stability diagram for the UW^+ scheme (FTBS).

(*conditionally stable*).

The following table reports the stability conditions obtained with the Von Neumann analysis for the various finitie difference schemes It is important to observe that the central

Schema	Interv. di stabilità
FTCS	*incond. instabile*
UW^+	$0 \leq \sigma \leq 1$
UW^-	$-1 \leq \sigma \leq 0$
LF	$-1 \leq \sigma \leq 1$
LW	$-1 \leq \sigma \leq 1$
BW^+	$0 \leq \sigma \leq 2$
BW^-	$-2 \leq \sigma \leq 0$
$FROMM^+$	$0 \leq \sigma \leq 1$
$FROMM^-$	$-1 \leq \sigma \leq 0$
$UW3^+$	$0 \leq \sigma \leq 1$
$UW3^-$	$-1 \leq \sigma \leq 0$

Table 2.3

schemes (LF, LW), for which according to notation (2.50) $p = q$, the stability is not influenced by the sign of σ, i.e. by the sign of the speed of sound c We can then imagine to extend the use of central schemes also to problems for which the propagation speed is not known at priori, as for non linear conservation problems. On the contrary, the stability of upwind schemes crucially depends on the sign of c. In particular, left-sided upwind schemes (type '+') are suited for waves propagating in the positive x direction, whereas right-sided upwind schemes (type '-') are suited for waves propagating in the negative x direction.

2.5 CFL condition

The CFL condition (or Courant-Friedrichs-Lewy principle) is a condition needed for the stability of a numerical algorithm, based on the concept of *domain of dependence* The physical domain of dependence represents the set of points that have an influence the value of the solution in a given point $P(x, t)$. In a similar way, the numerical domain of dependence

associated to a finite difference scheme represents the set of mesh nodes that are involved in the computation of the value of the solution at node x_j at time t^n. The CFL stability condition requires that the physical domain of dependence must be entirely included in the numerical domain of dependence. If such conditions wan not satisfied, the numerical solution would not converge to the physical solution, because a change of the conditions in a point of the physical domain of dependence would not have any effect on the value of the numerical solution computed by the finite difference scheme.

For better clarity, let consider the linear advection equation and figure 2.7. The physical domain of dependence associated to point (x_j, t^n) is given by the points of the characteristic line passing through that point. The trace of the domain of dependence at the time level $t^{n-m} < t^n$ thus consists of the point

$$D_f(x_j; t^n, t_m) = \{x_j - c\,m\,k\}, \tag{2.63}$$

denoted by a square symbol in figure 2.7. If we consider a linear, two-levels, explicit discretization of the linear advection equation

$$U_j^{n+1} = \sum_{l=-p}^{q} a_l\, U_{j+l}^n, \tag{2.64}$$

the numerical domain of dependence at time t^{n-m} is given by the set of nodes

$$D_n(x_j; t^n, t_m) = \{x_{j-mp}, \ldots, x_j, \ldots, x_{j+mq}\}, \tag{2.65}$$

that is represented with circles in figure 2.7 (where we assume $p = q = 1$). The CFL condition can be expressed by imposing

$$D_f(x_j; t^n, t_m) \in D_n(x_j; t^n, t_m), \tag{2.66}$$

that in the present case implies

$$-p\,m\,h \leq -c\,m\,k \leq q\,m\,h \qquad \Longrightarrow \qquad -q \leq \sigma \leq p. \tag{2.67}$$

Note that the CFL is a necessary condition for stability and it is satisfied for all the schemes reported in table 2.3. However, this is not a sufficient condition to guarantee the stability of the algorithms, and in general it is important to rely on the Von Neumann stability analysis.

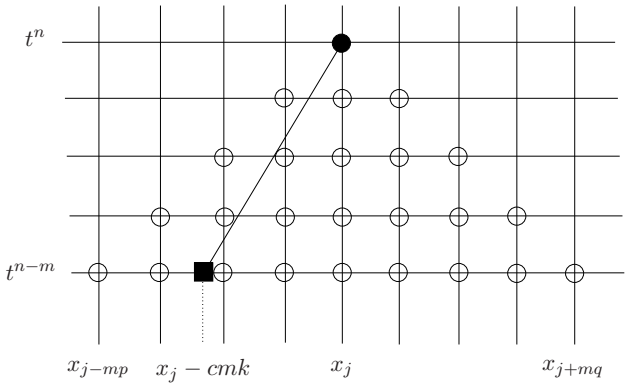


Figure 2.7: CFL condition

2.6 Dispersion and dissipation in the Fourier space

The analysis we carried out to determine the linear stability properties of a numerical scheme also allows an interesting interpretation of the numerical dispersion and dissipation in the Fourier space. In particular, from equation (2.54) it is possible to derive the amplification factor associated to the exact solution of the linear advection equation

$$g^*(\varphi) = \frac{\hat{u}(t+k)}{\hat{u}(t)} = 1 \cdot e^{-i\,c\,w\,k} = 1 \cdot e^{-i\,\sigma\,\varphi}. \tag{2.68}$$

This relation indicates that the exact solution proceeds in time from one step to the successive in such a way that the modulus of the complex amplitude $\hat{u}$ remains constant (being $|g^*(\varphi)| = 1$), whereas its phase changes by a factor $(-\sigma\,\varphi)$. For the discrete solution we have

$$g(\varphi) = \frac{\hat{v}^{n+1}}{\hat{v}^n} = g(\varphi) = |g(\varphi)|\, e^{-i\,\Phi} = |g(\varphi)| \cdot e^{-i\,\tilde{c}\,w\,k} \tag{2.69}$$

where $\Phi = -\arg(\varphi)$, and $\tilde{c}$, comparing (2.69) with (2.68), can be interpreted as a discrete phase velocity. Since in general $|g^*(\varphi)| \neq 1$, we observe that contrary to the exact solution, the numerical one evolves in such a way that every time step the harmonics are damped or amplified, depending on the value of $|g(\varphi)|$. Therefore it is useful to define the dissipation error in the Fourier space associated to a numerical scheme as

$$\varepsilon_d(\varphi) \equiv \frac{|g|}{|g^*|} = |g(\varphi)|, \tag{2.70}$$

to quantify the rate of dissipation (or amplification) associated to an harmonic with reduced wavenumber φ. In particular, if the stability condition is satisfied, we will have $\varepsilon_d(\varphi) \leq 1\,\forall\varphi$.

We also highlight that in general $\tilde{c} \neq c$, implying that the numerical phase velocity is different from the exact one and it is a function of the reduced wave number. We can introduce the dispersion error (in Fourier space) associated to a specific numerical scheme as

$$\varepsilon_\varphi(\varphi) = \frac{\tilde{c}}{c} = \frac{\Phi}{\sigma\,\varphi}. \tag{2.71}$$

For a sinusoidal wave with reduced wavenumber φ the discrete propagation velocity will be higher than the exact one when $\varepsilon_\varphi > 1$. Vice-versa, if $\varepsilon_\varphi < 1$ the propagation velocity will be smaller.

In table 2.4 we have reported the amplification factors of various schemes and in the next figures the corresponding plots for the dissipation and dispersion errors for different values of the Courant number. The results provide a further confirmation of the stability characteristics of the schemes obtained on the basis of the analysis of the truncation error. In particular we note that UW^+ and LF are stable in the interval $0 \leq \sigma \leq 1$ and show significant deviations with respect to the unitary value. For LW we have $\varepsilon_\varphi < 1$ for most part of the reduced wavenumber interval, independently of the value assumed by σ; this behavior suggests that the numerical solutions computed with the LW scheme typically presents a phase lag with respect to the exact solution. On the contrary, the BW^+ scheme does not present a unique behavior in terms of dispersion, and it produces solutions with a phase lead when $0 \leq \sigma \leq 1$ (for which $\varepsilon_\varphi > 1$) and with a phase lead when $1 \leq \sigma \leq 2$ (being $\varepsilon_\varphi < 1$). A similar behavior is observed for the FROMM^+ scheme, for which the change of the sign of the

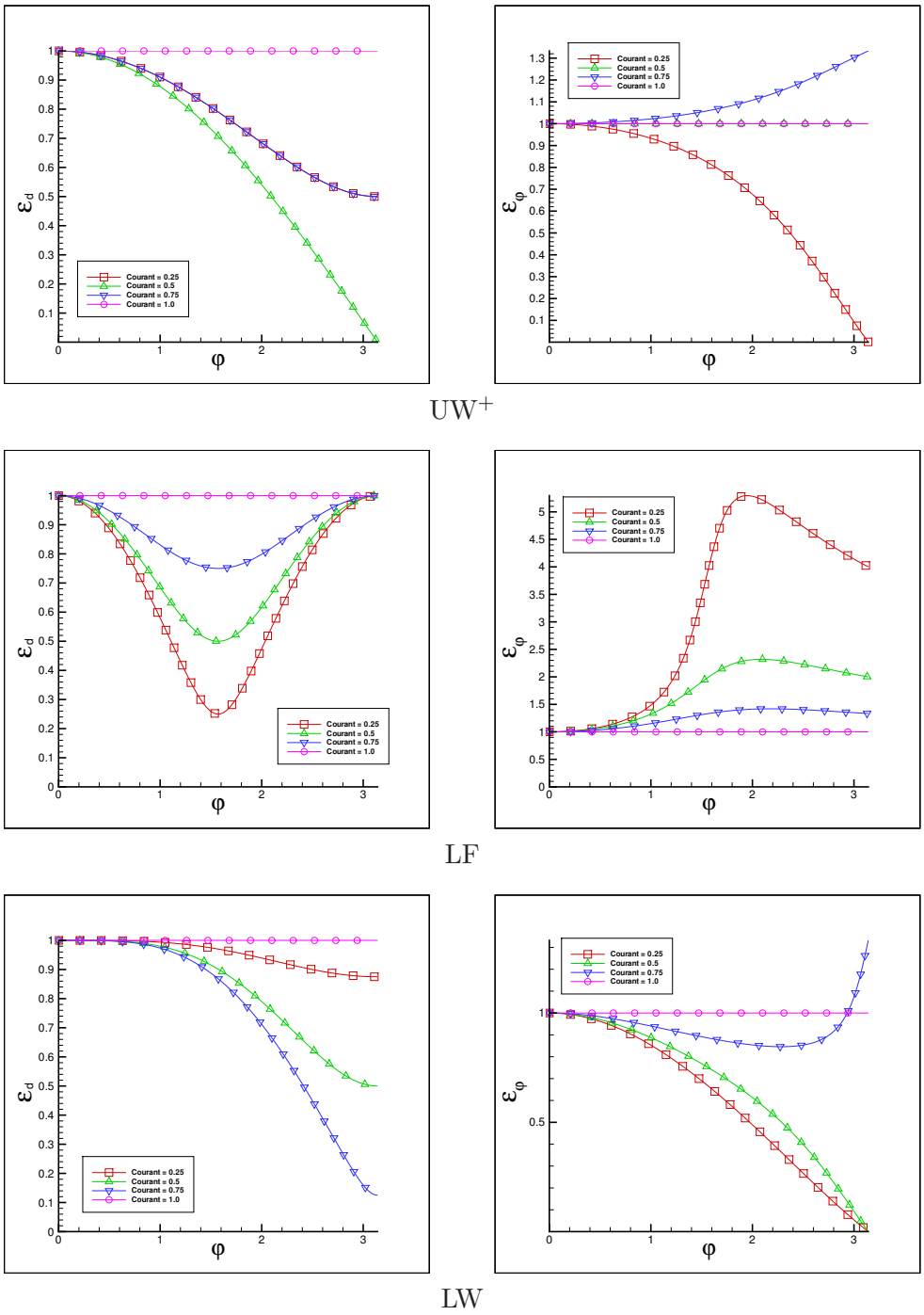
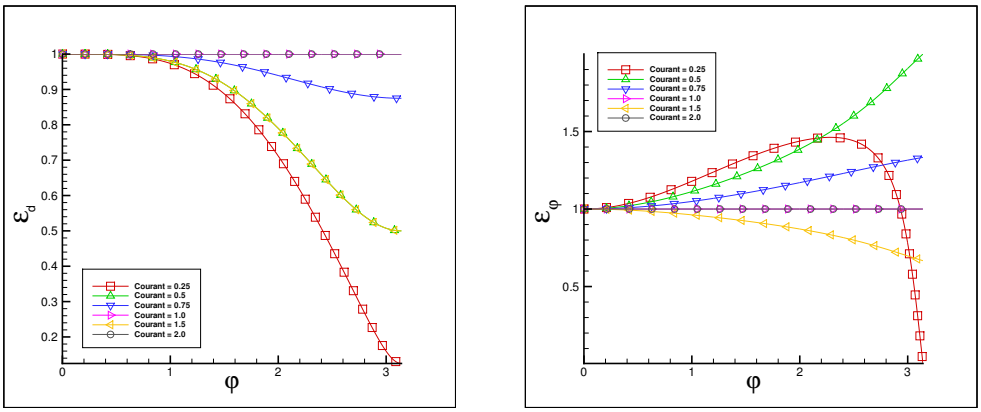
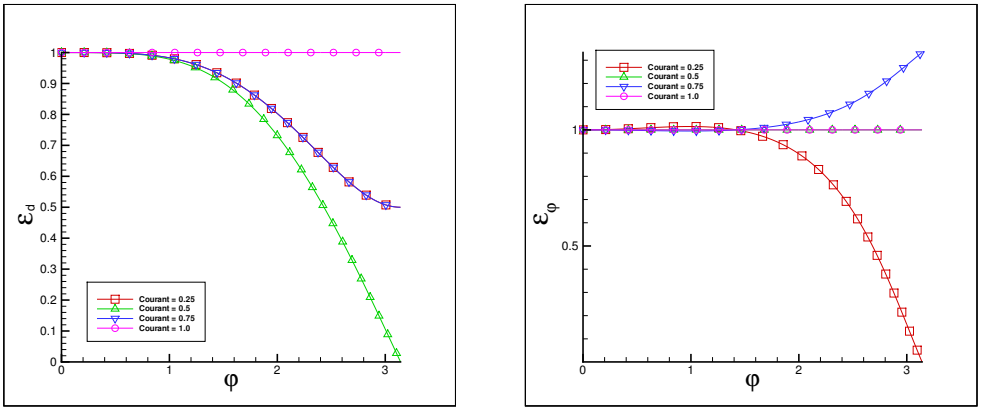


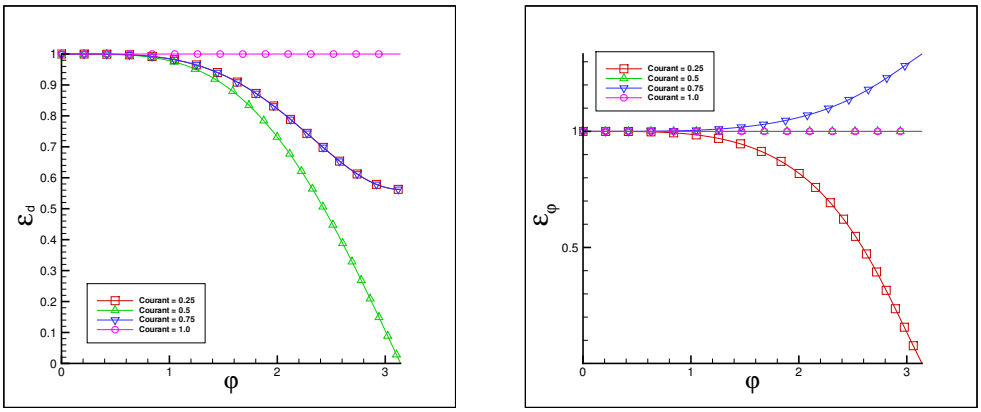
Figure 2.8: Dispersion and dissipation error for various finite difference scheme.



BW$^+$



Fromm$^+$



UW3$^+$

Figure 2.8: Dispersion and dissipation error for various finite difference scheme.

Schema	$g(\sigma,\varphi)$
FTCS	$1-i\sigma\sin(\varphi)$
UW$^+$	$1-\sigma+\sigma e^{-i\varphi}$
LF	$\frac{1}{2}(\sigma+1)e^{-i\varphi}-\frac{1}{2}(\sigma-1)e^{i\varphi}$
LW	$1-\sigma^2+\frac{\sigma}{2}(\sigma+1)e^{-i\varphi}+\frac{\sigma}{2}(\sigma-1)e^{i\varphi}$
BW$^+$	$1+\frac{\sigma}{2}(\sigma-3)+\frac{\sigma}{2}(\sigma-1)e^{-2i\varphi}-\sigma(\sigma-2)e^{-i\varphi}$
FROMM$^+$	$1-\frac{\sigma}{4}(\sigma+3)+\frac{\sigma}{4}(\sigma-1)e^{-2i\varphi}-\frac{\sigma}{4}(\sigma-5)e^{-i\varphi}+\frac{\sigma}{4}(\sigma-1)e^{i\varphi}$
UW3$^+$	$\frac{1}{2}(1-\sigma^2)(2-\sigma)+\frac{\sigma}{6}(\sigma^2-1)e^{-2i\varphi}+\frac{\sigma}{2}(1+\sigma)(2-\sigma)e^{-i\varphi}$ $+\frac{\sigma}{6}(\sigma-1)(2-\sigma)e^{i\varphi}$

Table 2.4: Amplification factor associated to various finite difference scheme.

dispersion error occurs at $\sigma = 0.5$. It is interesting to note that the error phase associated to the FROMM scheme is remarkably lower than the error associated to both LW and BW. It would possible to show that the analysis reported herein and in section 2.3 are closely related to each other. in particular the truncation error of a specific numerical scheme is linked to the shape of $\varepsilon_d(\varphi)$ and $\varepsilon_\varphi(\varphi)$ for $\varphi \simeq 0$. Effectively, the analysis based on the truncation error is based on Taylor series expansion and it only provides asymptotic information in the limit $h \downarrow 0$, i.e. valid only for $\varphi \downarrow 0$. The analysis of the error in the Fourier space characterizes the properties of a numerical scheme in a more efficient and complete way, being also valid for finite values of φ.

2.7 Convergence

Convergence is the property of a numerical scheme for which when the mesh spacing goes to zero, the numerical solution tends towards the solution of the differential equation. Note that this concept is different from both consistency, where the differential equation is compared with the finite difference scheme, and stability, which pertains to the behavior of the numerical solution. To quantify convergence we start by the definition of a *global error*, that includes both the discretization errors and the numerical errors deriving from the solution of the finite difference scheme, at node x_j and time t^n

$$e_j^n = U_j^n - u(x_j, t^n), \tag{2.72}$$

and we introduce the global error norm at time level t^n as

$$||e^n|| = \begin{cases} \max\limits_{j=1,N} |e_j^n| & Norma-\infty \\ \sqrt{\dfrac{1}{N}\displaystyle\sum_{j=1}^{N}(e_j^n)^2} & Norma-2 \\ \dfrac{1}{N}\displaystyle\sum_{j=1}^{N}|e_j^n| & Norma-1 \end{cases}. \tag{2.73}$$

A finite difference scheme is convergent if, for a given fixed time $T = nk$, results

$$\lim_{h\downarrow 0, k/h=\lambda} ||e^n|| = 0, \tag{2.74}$$

begin the limit intended at constant mesh ratio λ Note that in (2.74) we do not specify the type of norm considered because the convergence can in general be dependent on the specific norm applied to defined the global error. In particular we highlight that the norm-∞ requires the punctual convergence of the solution, a condition that is difficult (almost impossible) to obtain in the presence of discontinuous solutions. The convergence (or not) of a numerical method to the exact solution (if available) can be 'measured' by means of numerical experiments by considering different meshes. To that purpose it is convenient to represent the results in log-log plot, by reporting the error norm as a function of the mesh spacing h. Generally, for $h \downarrow 0$, the global error decades as a power of the spacing h

$$||e^n|| \sim c\, h^p \qquad \Longrightarrow \qquad \log ||e^n|| \sim p\, \log h, \tag{2.75}$$

and to determine p, called global order of accuracy (or convergence speed) of the numerical scheme it is sufficient to measure the asymptotic slope of error curves in the plot. A typical result of a convergence analysis (in this case for the LW scheme) is reported in figure 2.9.

Lax equivalence theorem

A bridge between the concepts introduced in the previous section, i.e. consistency, stability, convergence, is provided by the following theorem

Lax equivalence theorem A linear numerical scheme, which is consistent and stable, converges (in all the norms) to the exact solution of the differential equation. Furthermore, the global order of accuracy coincides with the formal order of accuracy.

Note that the theorem can be applied only to solutions characterized by the proper regularity conditions. In the case of discontinuous solutions it is not even possible to expand it in Taylor series and demonstrate the consistency of the numerical scheme. In that case we typically observe a lack of convergence of the solution, or a global order of accuracy which is lower than the formal one. The utility of the Lax equivalence theorem for numerical gasdynamics is then rather limited, because the equations have a nonlinear character and the solution can be discontinuous.

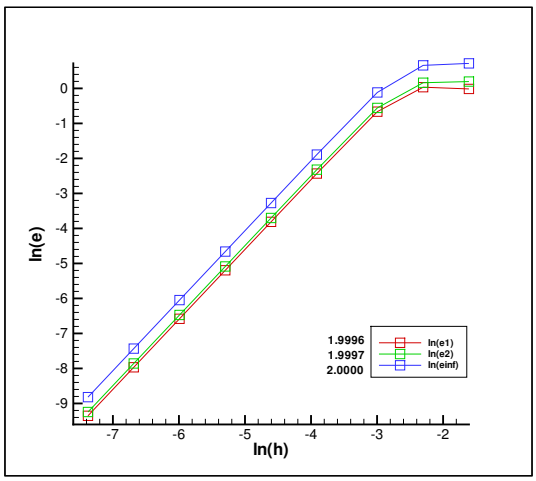


Figure 2.9: Typical convergence curve for a second order scheme (Lax-Wendroff).

Chapter 3

Nonlinear scalar conservation laws; conservative schemes

In this chapter we present techniques for the numerical solution of the scalar conservation law (1.5). Compared with the case of the linear advection equation, discussed in the previous chapter, we can identify a list of potential problems for a finite difference scheme, as the follows:

1. appearance of discontinuities also in the presence of smooth and continuous initial conditions;
2. different speed of sound ($a(u) = f'(u)$) (also the sign) at each computational node
3. oscillations can lead to non-linear instabilities and to non physical solutions (e.g. $p < 0$, $\rho < 0$ in the case of Euler equations)
4. possible convergence to weak solutions that do not satisfy the entropy condition
5. possible convergence to solutions that do not satisfy the jump conditions

We start by discussing the last of the points mentioned above. To that purpose it is convenient to consider the following Riemann problem for the Burgers equation (here reported in quasi-linear form)

$$u_t + u\,u_x = 0, \qquad u_0(x) = \begin{cases} 1 & x < 0 \\ 0 & x > 0 \end{cases}, \qquad -\infty < x < +\infty,\ t \geq 0 \tag{3.1}$$

Let consider the discretization

$$U_j^{n+1} = U_j^n - \lambda\, U_j^n\,(U_j^n - U_{j-1}^n) \tag{3.2}$$

It is easy to verify that the scheme (3.2) is consistent with (3.1), and accurate to the first order. However, the solution obtained with this method (easy to show) is

$$U_j^n = U_j^0 \qquad \forall\, j, n; \lambda, \tag{3.3}$$

which consists of a steady discontinuity separating two uniform states. However, such discontinuity does not satisfy the jump relations (1.23), according to which the propagation speed

should be $s = 1/2$. On the contrary, we can verify that the following finite difference scheme (obtained by a discretization of the conservative form of the Burgers equation)

$$U_j^{n+1} = U_j^n - \lambda/2\,[(U_j^n)^2 - (U_{j-1}^n)^2)] \tag{3.4}$$

generates a solution that is not steady and (in the limit $h \to 0$) converges to the weak solution of the problem under investigation.

3.1 Conservative schemes

We say that a numerical scheme is *conservative* if there exists a function (h, named numerical flux function)

$$\hat{f}_{j+\frac{1}{2}}^n = h(U_{j-p+1}^n, \ldots, U_{j+q}^n) \tag{3.5}$$

such that it is possible to write the finite difference scheme in the form (called conservative)

$$U_j^{n+1} = U_j^n - \lambda\,(\hat{f}_{j+\frac{1}{2}}^n - \hat{f}_{j-\frac{1}{2}}^n) \tag{3.6}$$

where $\hat{f}_{j+\frac{1}{2}}^n$ is denoted as *numerical flux* at the intermediate node $x_{j+1/2}$. Note that the stencil associated to the scheme (3.6) includes the nodes with indices $j-p, \ldots, j+q$ and that in general the function h depends on both the flux function $f(u)$, and the type of discretization. It would be possible to demonstrate that, in order the approximation (3.6) to be consistent, it is sufficient to satisfy the property

$$h(u, u, \ldots, u) = f(u). \tag{3.7}$$

Note that the finite difference scheme given by (3.6)+(3.5) is a *nonlinear*, explicit, two-levels scheme, where the nonlinearity is contained in the flux function $f(u)$ and/or in the form of the numerical flux function.

3.2 Integral conservation properties. Lax-Wendroff theorem

By considering the conservative approximation (3.6) and summing over all nodes with indices between M and N, we get

$$h \sum_{j=M}^{N} U_j^{n+1} - h \sum_{j=M}^{N} U_j^n = k \sum_{j=M}^{N} \left(\hat{f}_{j+\frac{1}{2}}^n - \hat{f}_{j-\frac{1}{2}}^n\right). \tag{3.8}$$

By expanding the summation on the right hand side we have

$$\begin{aligned}\sum_{j=M}^{N} \left(\hat{f}_{j+\frac{1}{2}}^n - \hat{f}_{j-\frac{1}{2}}^n\right) &= \hat{f}_{M+1/2}^n - \hat{f}_{M-1/2}^n + \hat{f}_{M+3/2}^n - \hat{f}_{M+1/2}^n + \hat{f}_{M+5/2}^n - \hat{f}_{M+3/2}^n + \ldots \\ &+ \ldots \hat{f}_{N-5/2}^n - \hat{f}_{N-3/2}^n \hat{f}_{N-3/2}^n - \hat{f}_{N-1/2}^n \hat{f}_{N-1/2}^n - \hat{f}_{N+1/2}^n,\end{aligned}$$

where all the contributions cancel two by two (telescopic property), except for the boundary values, and (3.8) simplifies to

$$h \sum_{j=M}^{N} \frac{U_j^{n+1} - U_j^n}{k} = \hat{f}_{N+1/2}^n - \hat{f}_{M-1/2}^n. \tag{3.9}$$

By setting $x_{M+1/2} = a$, $x_{N+1/2} = b$, and observing that $(U_j^{n+1} - U_j^n)/k \approx \partial u/\partial t$, we have

$$h \sum_{j=M}^{N} \frac{U_j^{n+1} - U_j^n}{k} \approx \frac{\mathrm{d}}{\mathrm{d}t} \int_a^b u \, \mathrm{d}x. \tag{3.10}$$

Furthermore, since $\hat{f}^n_{M-1/2} \approx f(a)$, $\hat{f}^n_{N+1/2} \approx f(b)$ thanks to the consistency of the numerical scheme, we can conclude that the equation (3.9) reproduces, in the discrete sense, the integral form of the scalar conservation equation (1.1). Since the Rankine-Hugoniot jump relations have been obtained by using the integral form of the equation (see Section 1.3), we can expect that numerical schemes written in conservative form are able to detect possible discontinuities of the numerical solution in the proper way. In this regard there exists an exact mathematical properties of conservative schemes:

Lax-Wendroff theorem: let consider a conservative and consistent numerical scheme. If the scheme converges at least in norm-1, then it converges to a weak solution of the conservation law.

The Lax-Wendroff theorem provides a strong theoretical support to the use of conservative schemes, because it guarantees the onset of incorrect behaviors as that shown with the example (3.2). However, it is important to notice that this is not a convergence theorem. The convergence of the theorem is an hypothesis and must be verified in some way. Moreover, the Lax-Wendroff theorem does not specify that the solution converges to the weak solution that is also the entropy one.

3.3 Conservative formulation for schemes based on the characteristics method

Let consider again the schemed developed for the linear advection equation in chapter 2. It is straightforward to verify that all these schemes can be rewritten in the conservative form (3.6), by properly defining the numerical flux. In table 3.1 we report the numerical fluxes associated to all the schemes considered until now, both central and upwind.

Schema	$\hat{f}^n_{j+1/2}$
FTCS	$\frac{c}{2}\left(U_j^n + U_{j+1}^n\right)$
UW^+	cU_j^n
UW^-	cU_{j+1}^n
LF	$\frac{c}{2}\left(U_j^n + U_{j+1}^n\right) - 1/(2\lambda)\,\delta U^n_{j+1/2}$
LW	$\frac{c}{2}\left(U_j^n + U_{j+1}^n\right) - 1/2\,\lambda c^2\,\delta U^n_{j+1/2}$
BW^+	$\frac{c}{2}\left(3U_j^n - U_{j-1}^n\right) - 1/2\,\lambda c^2\,\delta U^n_{j-1/2}$
BW^-	$\frac{c}{2}\left(3U_{j+1}^n - U_{j+2}^n\right) - 1/2\,\lambda c^2\,\delta U^n_{j+3/2}$

Table 3.1: Numerical flux associated to various discretizations of the linear advection equation.

The extension of the numerical fluxes to the scalar conservation equation in its general form (1.5) can be done through the formal substitution

$$cU_j^n \rightsquigarrow f_j^n, \quad c \rightsquigarrow a_{j+1/2},$$

where $f_j^n = f(U_j^n)$, and the speed of sound at the intermediate state $j+1/2$ is defined as follows (see figure 3.1) [1]

$$a_{j+1/2}^n = \begin{cases} \delta f_{j+1/2}^n / \delta U_{j+1/2}^n & \delta U_{j+1/2}^n \neq 0 \\ f'(U_j^n) & \delta U_{j+1/2}^n = 0 \end{cases} . \tag{3.11}$$

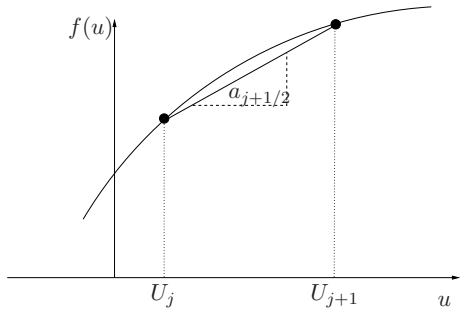


Figure 3.1: Definition of the speed of sound at the intermediate state $a_{j+1/2}$.

Schema	$\hat{f}_{j+1/2}^n$
FTCS	$\frac{1}{2}\left(f_j^n + f_{j+1}^n\right)$
UW$^+$	f_j^n
UW$^-$	f_{j+1}^n
LF	$1/2\left(f_j^n + f_{j+1}^n\right) - 1/(2\lambda)\,\delta U_{j+1/2}^n$
LW	$1/2\left(f_j^n + f_{j+1}^n\right) - 1/2\,\lambda a_{j+1/2}^2\,\delta U_{j+1/2}^n$
BW$^+$	$1/2\left(3f_j^n - f_{j-1}^n\right) - 1/2\,\lambda a_{j-1/2}^2\,\delta U_{j-1/2}^n$
BW$^-$	$1/2\left(3f_{j+1}^n - f_{j+2}^n\right) - 1/2\,\lambda a_{j+3/2}^2\,\delta U_{j+3/2}^n$

Table 3.2: Numerical flux associated to different discretizations of the nonlinear scalar conservation equation (1.5).

[1] hereinafter we use the notation $\delta z_{j+1/2} = z_{j+1} - z_j$ to denoted the variation of a generic discrete variable z across the intermediate node $j+1/2$.

3.4 Artificial viscosity

Let consider the convection-diffusion equation (1.3) and the application of the FTCS discretization. We obtain the following finite difference scheme

$$U_j^{n+1} = U_j^n - \frac{\lambda}{2}\left(f_{j+1}^n - f_{j-1}^n\right) - \frac{\varepsilon h}{k^2}\left(U_{j-1}^n - 2U_j^n + U_{j+1}^n\right), \tag{3.12}$$

that, with simple algebraic manipulations, can be rewritten in the conservative form (3.6), provided the numerical flux is defined as follows

$$\hat{f}_{j+1/2}^n = \frac{1}{2}\left(f_j^n + f_{j+1}^n\right) - \frac{\varepsilon}{h}\,\delta U_{j+1/2}^n, \tag{3.13}$$

which is the sum of the contribution of the FCTS scheme applied to (1.5) (as reported in table 3.2), plus a contribution related to the physical diffusion (ε here represents a coefficient of physical diffusion).

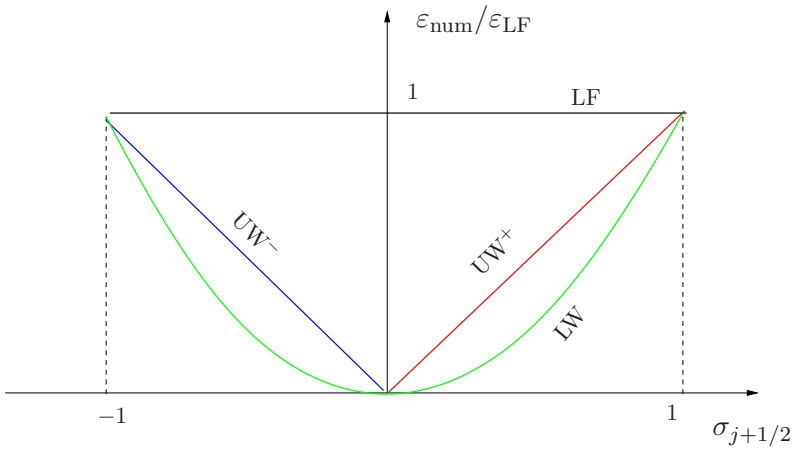


Figure 3.2: Numerical diffusion coefficients (normalized for the value corresponding to the LF scheme).

By comparing the expressions of the numerical fluxes for the non-viscous equation (1.5) associated to the various schemes reported in table 3.2 (at least for schemes with a three nodes stencil $p = q = 1$) we notice that all can be recast in a form which analogous to (3.13) (named *numerical viscosity form*, provided that the coefficient of numerical viscosity ε is properly defined, as shown in table 3.3. Note that in this case ε is not associated to diffusive mechanisms of physical nature/origin, but it comes out as a result of the numerical discretization. For this reason we will denote this coefficient as *numerical diffusion* or *artificial viscosity* and it will be denoted as ε_{num}.

Table 3.3 shows that the artificial viscosity of a finite difference scheme typically depends by the grid spacing and by the local propagation speed. It is instructive to represent the coefficients of numerical diffusion in graphic form by normalizing with respect to that associated to the Lax-Friedrichs scheme (as shown in figure 3.2). Note that the result can be expressed

Schema	ε_{num}
FTCS	0
UW$^+$	$a_{j+1/2}h/2$
UW$^-$	$-a_{j+1/2}h/2$
LF	$h/(2\lambda)$
LW	$\lambda h a_{j+1/2}^2/2$
Roe	$\lvert a_{j+1/2}\rvert h/2$
LFFS-UW	$\alpha h/2$

Table 3.3: Numerical diffusion coefficients for fully discrete with a three nodes stencil ($p = q = 1$).

i terms of the *local Courant number* $\sigma_{j+1/2} = h\, a_{j+1/2}$, as for instance

$$\varepsilon_{UW^\pm}/\varepsilon_{LF} = \pm\frac{a_{j+1/2}h}{2}\cdot\frac{2\lambda}{h} = \pm\sigma_{j+1/2} \tag{3.14}$$

$$\varepsilon_{LW}/\varepsilon_{LF} = \pm\frac{\lambda a_{j+1/2}^2 h}{2}\cdot\frac{2\lambda}{h} = \sigma_{j+1/2}^2. \tag{3.15}$$

The capability of the numerical flux of a finite difference scheme to reproduce (in the discrete sense) mechanisms of numerical diffusion that had been removed in the passage from the convection diffusion equation (1.3) to the pure convection equation is of extreme importance from the practical point of view. We note in particular that for all the schemes considered (except the case for which $a_{j+1/2} = 0$), $\varepsilon_{\text{num}} \overset{h\to 0}{\to} 0$. This fact implies that the numerical viscosity of a finite difference scheme plays the role of the (physical) viscosity that appears in the definition of the entropy solution of a conservation equation (see the discussion reported in section 1.4). The presence of a (positive) contribution of artificial viscosity in a numerical scheme is a suitable precondition to hope in the convergence towards the correct entropy solution of the conservation equation.

The exam of the coefficient of numerical diffusion also allows to draw several important conclusions on its qualitative behavior. First, we can expect that the coefficient ε_{num} must be positive to have a stable numerical scheme. This is indeed essential to avoid an exponential divergence of the solution (recall section 2.6). Looking at table 3.3, we can infer that upwind schemes of type + can be stable only when $a_{j+1/2} \geq 0$, i.e. for waves propagating from left to right), whereas the opposite holds for upwind schemes of type −.

Secondly, some of the schemes (specifically UW±, LW) present a cancellation of the coefficient of numerical diffusion when $a_{j+1/2} = 0$. Such a situation typically occurs in the presence of *sonic transitions*, i.e. when $a(u)$ changes sign from U_j to U_{j+1} (see figure 3.3). In this case we can expect the onset of problems to the correct convergence towards the entropy solution, because of the absence of the dissipative mechanisms incorporated in (1.3).

3.5 Linear stability

The Von Neumann stability analysis cannot be straightforwardly extended to the analysis of schemes considered in this chapter, because without the initial assumption of linearity we

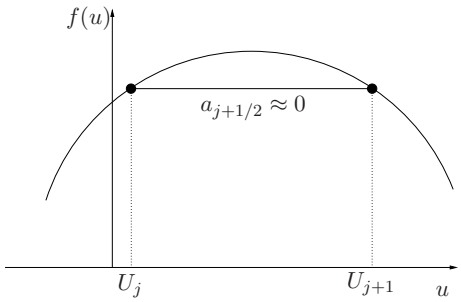


Figure 3.3: Representation of sonic transition in the u-$f(u)$ plane.

cannot consider the autonomous evolution of the various harmonics. However, we can expect that, in order to have a stable nonlinear scheme of the type (3.6)+(3.5) the Von Neumann stability condition must be satisfied at each node (locally). For instance, the linear stability of scheme UW$^+$ for the linear advection equation requires (see section 2.4) that $0 \leq \sigma \leq 1$. By extrapolating this results, we have a stability condition $0 \leq \sigma_{j+1/2} \leq 1$. Furthermore, this condition must be satisfied for each intermediate node, i.e. $\forall j$. The validity of this extrapolation is (almost) always verified in practice. In particular, at least in the case of *smooth* solutions, a numerical scheme is stable when the Von Neumann condition is verified at each node. On the other hand, in the case of discontinuous solutions there can be the occurrence of nonlinear instability conditions that lead to the divergence of the numerical solution, even though the conditions of linear stability are satisfied.

3.6 Central vs upwind schemes

On the basis of the considerations made in the previous section, and recalling the findings concerning the stability of the various schemes for the linear advection equation, it results natural to separate the numerical schemes in two categories

1. *central schemes*, whose stability does not depend on the sign of the local speed of sound

2. *upwind schemes*, whose stability depends on the sign of the local speed of sound.

FTCS, LW and LW belong to the first category, whereas UW$^\pm$, BW$^\pm$, FROMM$^\pm$, UW3$^\pm$ are in the second one.

Concerning the schemes of the second category, their extension to the case of nonlinear equations, for which $a(u)$ can change its sign , requires some caution. The simplest idea to realize such an extension consists in verifying the sign of the speed of sound associated to the generic intermediate state $j + 1/2$, defining the relative numerical flux applying an upwind scheme of type '+' when $a_{j+1/2} \geq 0$, and vice versa, of type '-' if $a_{j+1/2} < 0$. The extension of the first-order upwind scheme to the nonlinear case leads to the following numerical flux [2]

$$\hat{f}^n_{j+1/2} = \begin{cases} f^n_j & a^n_{j+1/2} \geq 0 \\ f^n_{j+1} & a^n_{j+1/2} < 0 \end{cases} , \tag{3.16}$$

[2]the sign function is defined as $\mathrm{sgn}(x) = x/|x|$.

that with simple algebraic manipulations cab be recast in form of numerical viscosity

$$\hat{f}^n_{j+1/2} = \frac{1}{2}\left(1+\mathrm{sgn}(a^n_{j+1/2})\right) f^n_j + \frac{1}{2}\left(1-\mathrm{sgn}(a^n_{j+1/2})\right) f^n_{j+1} = \frac{1}{2}\left(f^n_j + f^n_{j+1}\right) - \frac{|a^n_{j+1/2}|}{2}\,\delta U^n_{j+1/2}. \tag{3.17}$$

The finite difference scheme defined by the numerical flux reported in (3.17) is known as *Roe scheme*, and it represent the prototype of a first order upwind scheme applied to a nonlinear conservation equation. The associated coefficient of numerical viscosity results $\varepsilon_{\mathrm{ROE}} = |a_{j+1/2}|h/2$, whose graph is essentially the union of the stable branches of the coefficients associated to UW$^{\pm}$ (see figure 3.2). As for other schemes, $\varepsilon_{\mathrm{ROE}}$ is zero in correspondence of sonic transitions, and as previously discussed, the Roe scheme can present some problems to convergence to the entropy solution. There exist more general approaches to that of Roe to build generalized upwind schemes, described in the next two sections.

3.7 Flux splitting

The basic idea of the *flux splitting* procedure consists in decomposing (at the continuous level) the flux function $f(u)$ in two parts, one associated with waves exclusively propagating from left to right ($f^+(u)$) and an other one associated with waves from right to left ($f^-(u)$).

Recalling the definition of local speed of sound given in section 1.2, this is equivalent to require that

$$f(u) = f^+(u) + f^-(u), \quad a^+(u) = \frac{\mathrm{d}f^+}{\mathrm{d}u} \geq 0, \quad a^-(u) = \frac{\mathrm{d}f^-}{\mathrm{d}u} < 0, \quad \forall u. \tag{3.18}$$

Denoted as $\hat{f}^+_{j+1/2}$ associated to an upwind discretization of type '+' applied to $f^+(u)$, and denoted as $\hat{f}^-_{j+1/2}$ the numerical flux obtained by an upwind '-' discretization applied to $f^-(u)$, the global numerical flux is obtained by recombining the two contributions, obtaining

$$\hat{f}_{j+1/2} = \hat{f}^+_{j+1/2} + \hat{f}^-_{j+1/2}. \tag{3.19}$$

For instance, the application of the fluxes for UW$^{\pm}$ leads to

$$\hat{f}_{j+1/2} = f^+_j + f^-_{j+1}, \tag{3.20}$$

where, as usual, $f^{\pm}_j = f^{\pm}(U_j)$.

There exist various type of flux splitting that satisfy the condition (3.18). We present two typical examples in the following.

Steger & Warming flux splitting

This splitting (also known as *physical flux splitting*), is based on the following decomposition of the flux function

$$f^+(u) = \begin{cases} f(u) & f'(u) \geq 0 \\ 0 & f'(u) < 0 \end{cases}, \quad f^-(u) = \begin{cases} 0 & f'(u) \geq 0 \\ f(u) & f'(u) < 0 \end{cases}, \tag{3.21}$$

that can be recast as

$$f^{\pm}(u) = \frac{1}{2}\left(1 \pm \mathrm{sgn}(f'(u))\right) \cdot f(u). \tag{3.22}$$

It is straightforward to verify that the decomposition 3.21 satisfy the requirements (3.18). The Steger & Warming flux splitting leads to numerical schemes that are very accurate, bit it is affected by the presence of a discontinuity in the first derivative in correspondence of the points where $f'(u)$ changes its sign. For this reason in practical applications it is preferable to use an alternative, more robust flux splitting.

Lax & Friedrichs flux splitting

The Lax & Friedrichs flux splitting is based on the following decomposition of the flux function

$$f^{\pm}(u) = \frac{1}{2}\left(f(u) \pm \alpha u\right), \quad \alpha = \max_u |f'(u)|, \tag{3.23}$$

where the parameter α represents the maximum among the possible values of the speed of sound at a given time. It is straightforward to verify that (3.23) automatically satisfies the condition (3.18).

By applying the flux splitting (3.23) to (3.20) we obtain a scheme, known as generalized Lax-Friedrichs (LFFS-UW), whose numerical flux is given by

$$\hat{f}^n_{j+1/2} = \frac{1}{2}\left(f^n_j + f^n_{j+1}\right) - \frac{\alpha}{2}\,\delta U^n_{j+1/2}. \tag{3.24}$$

Similarly to the Lax-Friedrichs scheme, this method presents an artificial viscosity that does not depend on the local speed of sound $a_{j+1/2}$, and that is never zero. Consequently we can expect that this scheme is not affected by problems related to the lack of convergence to the entropy solution. Moreover, is it easily to show that $\varepsilon_{\mathrm{LFFS-UW}}/\varepsilon_{\mathrm{LF}} = \sigma^*$, where

$$\sigma^* = \lambda\alpha, \tag{3.25}$$

is the so-called *global Courant number*, equal to the maximum absolute vale among the local Courant numbers. Note also that, for the linear stability of the scheme must be $|\sigma_{j+1/2}| \leq 1, \forall j$, that is certainly satisfied when $\sigma^* \leq 1$. This implies that $\varepsilon_{\mathrm{LFFS-UW}}/\varepsilon_{\mathrm{LF}} \leq 1$, and the scheme LFFS-UW is thus less dissipative than the LW scheme in its original form.

3.8 The Godunov's method

The flux splitting approach is commonly used to develop upwind schemes in the context of the finite difference method. The development of schemes based on the finite volume method (FV) follows completely different routes, although they lead to results similar to those we have seen until now. The strategy originally developed by Godunov is based on the discretization of the conservation equation written in the integral form.... By integrating equation (1.5) over a rectangular region of the $x-t$ plane defined by the Cartesian product $\left[x_{j-1/2}, x_{j+1/2}\right] \times [t^n, t^{n+1}]$, (as shown in figure 3.4), we obtain

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

$$= \int_{x_{j-1/2}}^{x_{j+1/2}} \left(v(x,t^{n+1}) - v(x,t^n)\right)\,\mathrm{d}x + \int_{t^n}^{t^{n+1}} \left(f(v(x_{j+1/2},t)) - f(v(x_{j-1/2},t))\right)\,\mathrm{d}t, \tag{3.26}$$

where the divergence theorem has been applied to partially integrate the first integrand with respect to t and the second to x. Recalling that in the finite volume method the unknowns represent average cell values of the approximate solutions, as defined in equation (2.9), and introducing the numerical flux defined as

$$\hat{f}^n_{j+1/2} = \int_{t^n}^{t^{n+1}} f(v(x_{j+1/2}, t))\,\mathrm{d}t, \tag{3.27}$$

equation (3.26) leads to

$$U^{n+1}_j = U^n_j - \lambda\left(\hat{f}^n_{j+1/2} - \hat{f}^n_{j+1/2}\right), \tag{3.28}$$

which coincides with the generic conservative form of a finite difference scheme (as from equation (3.6)). Note that the evaluation of the numerical flux (3.27) requires the computation of the integral of $f(v)$ along the boundary (in this case the right side highlighted in red in figure) of the control cell. In principle this computation requires in turn the knowledge of the values of the approximate solution $v(x_{j+1/2}, t)$ at the intercell $x_{j+1/2}$, for $t \in [t^n, t^{n+1}]$.

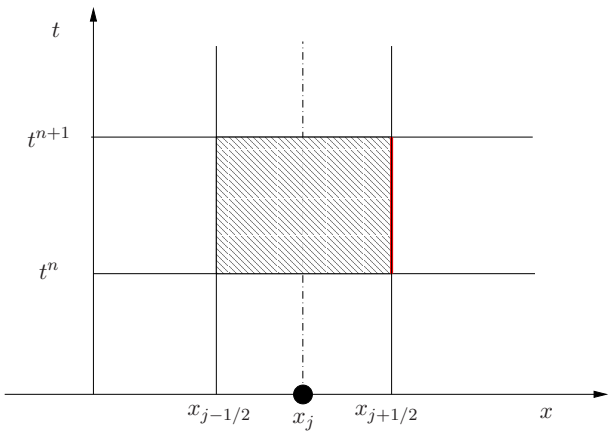


Figure 3.4: Numerical cell in the $x-t$ plane for the integration of equation (1.5).

In the Godunov approach the evaluation of the numerical flux at the intercell is done by solving a series of Riemann problems. Let assume that, cell by cell, the approximate solution (of which we know only the cell-average values U^n_j, $\forall j$),) have the following functional representation (at time t^n)

$$v(x, t^n) = U^n_j, \quad x \in [x_{j-1/2}, x_{j+1/2}], \tag{3.29}$$

which is apparently consistent with the condition (2.9). The *reconstruction*[3] expressed by (3.29) is equivalent to suppose that the approximate solution at time t^n is constant in every cell. Consequently, if we consider the intercell $x_{j+1/2}$, we will have an associated Riemann problem where the 'left' state is given by $u_l = U^n_j$, and the 'right' by $u_r = U^n_{j+1}$. The computation of the punctual value of v at the intercell in the time interval $[t^n, t^{n+1}]$requires

[3]with this term we will denote the process that allows to determine the punctual values of a function, starting from the cell-averages values. Note that this procedure is different from the classic interpolation procedure.

the solution of the Riemann problem. The analysis is remarkably simplified from the *self-similarity* property of the solutions of the Riemann problem for equation (1.5), according to which in a suitable interval around the point (in this case $x_{j+1/2}$) where the discontinuity has its origin the following property holds

$$v(x,t) = V\left(\frac{x - x_{j+1/2}}{t - t^n}\right), \tag{3.30}$$

which is equivalent to say that the solution is constant along rays in the plane $x-t$ that start from the origin of the discontinuity (see figure 3.5).

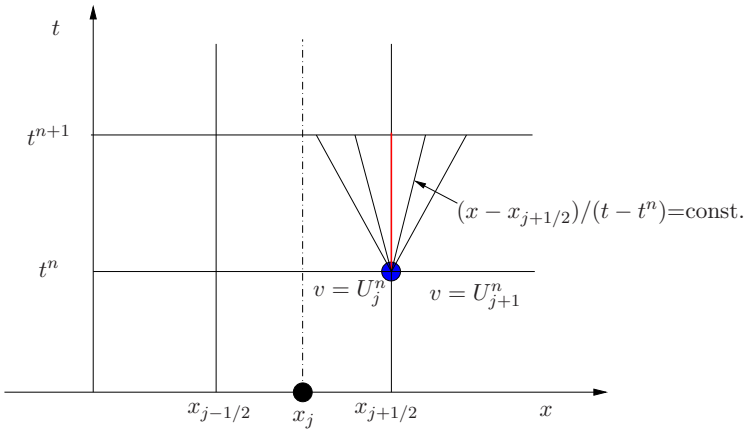


Figure 3.5: Solution of the Riemann problem at the intercell $x_{j+1/2}$ at time t^n.

This property implies that $v(x_{j+1/2}, t) \equiv v^n_{j+1/2}$ =cost., and the integral (3.27) drastically simplifies, leading to

$$\hat{f}^n_{j+1/2} = f(v^n_{j+1/2}). \tag{3.31}$$

The problem of determining the numerical flux reduces to find a numerical approximation for the 'average state' ($v^n_{j+1/2}$) between U^n_j and U^n_{j+1}.

It is worth to notice that the scheme defined by the numerical flux (3.31), named *Godunov scheme* presents a formal order of accuracy equal to one, independently on the computation of $v^n_{j+1/2}$. Indeed, while no approximations have been made to obtain the conservative form (3.28), the assumption (3.29) leads to an error of order $O(h)$ in the evaluation of the values of v.

To the purpose of evaluating $v^n_{j+1/2}$ it is possible to adopt two approaches:

1. exact solution of the Riemann problem;

2. approximate solution of the Riemann problem.

The first approach can be pursued only when the exact solution of the problem is known, as for instance for our model scalar conservation law. However, such a solution is not available for the three dimensional Euler equations, where it is possible to proceed through an iterative process that is too expensive for practical applications. As a consequence approach two if frequently applied, by using approximate estimates of the solution of the Riemann problem that are sufficiently accurate to preserve the accuracy of the scheme.

Exact Riemann solver

Let consider equation (1.5) with initial conditions [4]

$$u(x,0) = \begin{cases} u_l & x < 0 \\ u_r & x > 0 \end{cases}, \tag{3.32}$$

where, for the Godunov scheme, $u_l = U_j^n$, $u_r = U_{j+1}^n$. We want to evaluate the solution of the Riemann problem along the straight line $x = 0$ ($v_{j+1/2}$ according to our notation).

We consider a convex flux function, i.e. $f''(u) \geq 0$ (as for the Burgers equation). Two fundamental cases can occur

A. Expansion, if $u_l < u_r$;

B. Shock, if $u_l > u_r$.

In the first case the characteristics originating from the left and right states tend to diverge, being $f'(u_l) < f'(u_r)$. Vice versa, in the second case the characteristics coming from the left state propagate faster than that from the right state and a shock forms (recall section 1.3). It is possible to distinguish between various sub-cases, as shown in the following through the solution plots in the $x - t$ and $u - f$ planes.

A-1. Subsonic expansion

In this case $f'(u_l) \leq f'(u_r) \leq 0$ and the solution of the Riemann problem (highlighted by a red circle in figure 3.6b) is $v_{j+1/2} = u_r$.

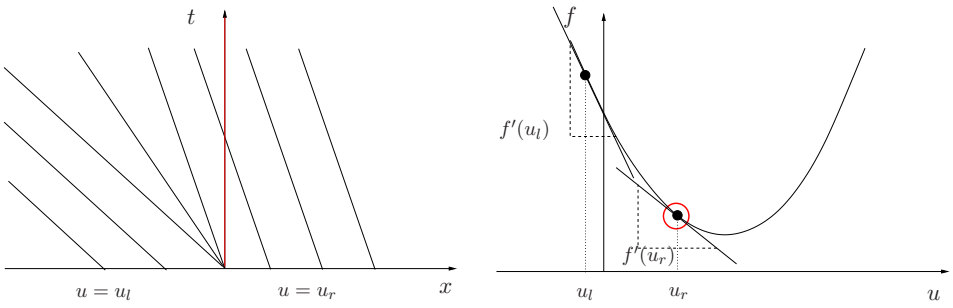


Figure 3.6: Solution of the Riemann problem for case A-1.

A-2. Supersonic expansion

In this case $0 \leq f'(u_l) \leq f'(u_r)$, and the solution of the Riemann problem is $v_{j+1/2} = u_l$.

A-3. Transonic expansion

In this case $f'(u_l) \leq 0 \leq f'(u_r)$. The solution of the Riemann problem can be obtained by noticing that the straight line $x = 0$ coincides with the characteristic of the expansion

[4] obviously the choice of the origin of the coordinate system does not influence the solution of the problem, therefore we consider a problem centered in the origin of the plane$x - t$.

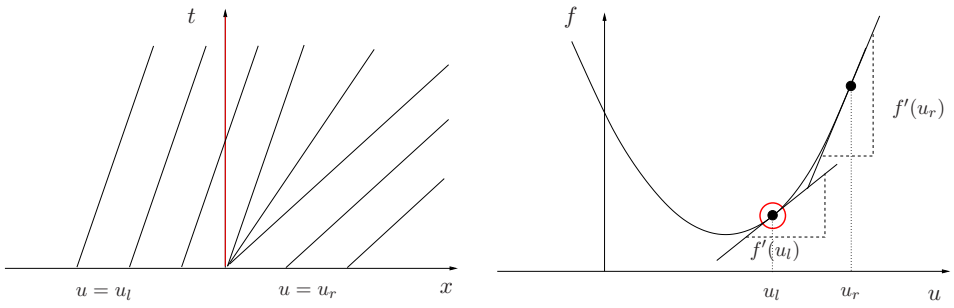


Figure 3.7: Solution of the Riemann problem for case A-2.

fan of equation $\mathrm{d}x/\mathrm{d}t = 0$. By definition of characteristic curve (equation 1.8), it results $\mathrm{d}x/\mathrm{d}t = f'(v_{j+1/2}) = 0$, from which it is possible to get $v_{j+1/2}$. If $f(u)$ is convex, $f'(v_{j+1/2})$ can be zero (at most) for a value of u (denoted as u^*), and the solution of the Riemann problem is $v_{j+1/2} = u^*$. For instance, for Burgers ($f(u) = u^2/2$), $u^* = 0$.

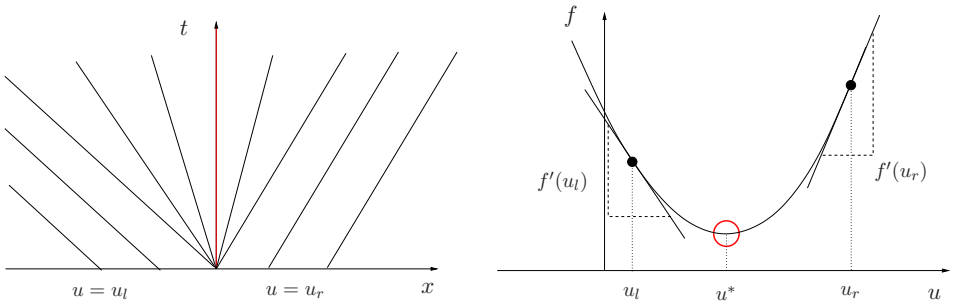


Figure 3.8: Solution of the Riemann problem for case A-3.

B-1. Subsonic shock

In this case $f'(u_r) \leq f'(u_l) \leq 0$, and the solution of the Riemann problem is $v_{j+1/2} = u_r$.

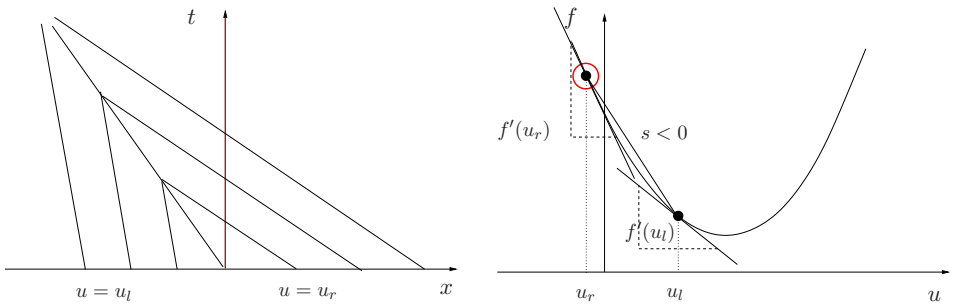


Figure 3.9: Solution of the Riemann problem for case B-1.

B-2. Supersonic shock

In this case $0 \leq f'(u_r) \leq f'(u_l)$, and the solution of the Riemann problem is $v_{j+1/2} = u_l$.

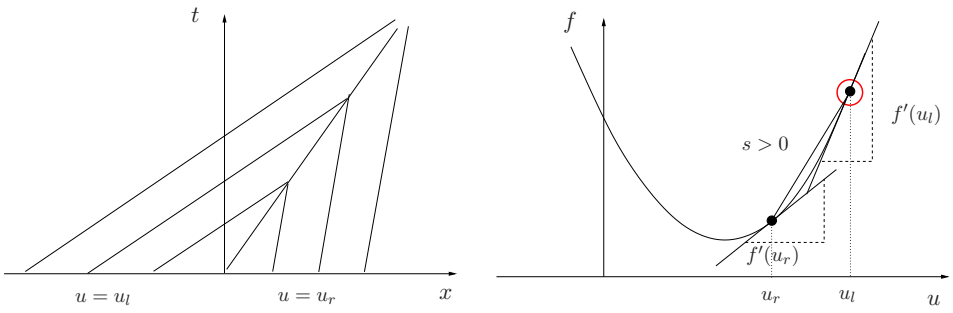


Figure 3.10: Solution of the Riemann problem for case B-2.

B-3. Transonic shock with $s > 0$

In this case $f'(u_r) \leq 0 \leq f'(u_l)$, and the shock linking the 'left' and 'right' states propagates at speed $s > 0$. The solution of the Riemann problem is $v_{j+1/2} = u_l$.

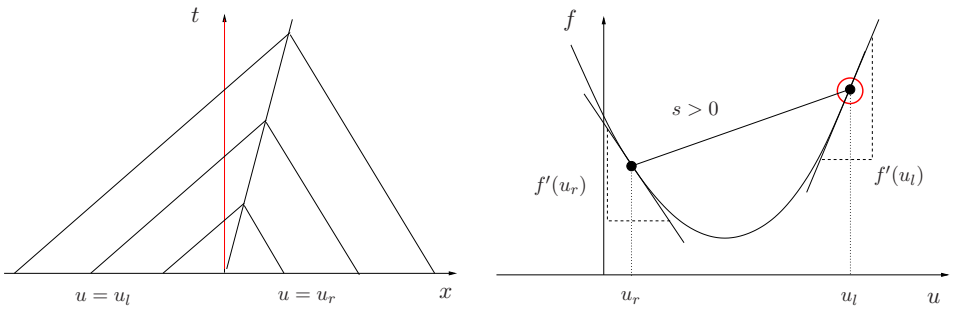


Figure 3.11: Solution of the Riemann problem for case B-3.

B-4. Transonic shock with $s < 0$

In this case $f'(u_r) \leq 0 \leq f'(u_l)$, and the shock linking the 'left' and 'right' states propagates at speed $s < 0$. The solution of the Riemann problem is $v_{j+1/2} = u_r$.

By collecting the results for all sub-cases it is possible to write the following expression for the numerical flux (recall that $\hat{f}_{j+1/2} = f(v_{j+1/2})$)

$$\hat{f}_{j+1/2} = \begin{cases} \min_{u_l \leq u \leq u_r} f(u) & u_l \leq u_r \\ \max_{u_r \leq u \leq u_l} f(u) & u_l \geq u_r \end{cases} . \tag{3.33}$$

Roe's Riemann solver

Roe has developed an approximate Riemann solver based on the exact solution of the problem deriving from the linearization of the conservation equation (1.5). Let consider the quasi-linear form (1.7) of the conservation equation and replace the local speed of sound $a(u)$ with

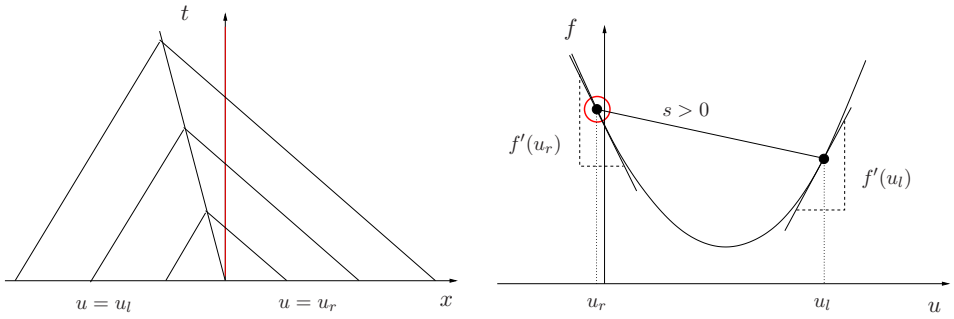


Figure 3.12: Solution of the Riemann problem for case B-4.

its value at the intermediate state $j + 1/2$ (in correspondence of the Riemann problem). We have the following linearised equation

$$\frac{\partial u}{\partial t} + a_{j+1/2} \frac{\partial u}{\partial x} = 0, \tag{3.34}$$

for which it is possible to exactly solve the Riemann problem with initial conditions (3.32). Indeed, the linear convection equation (3.34) is characterized by characteristics lines that are parallel (with slope $\mathrm{d}x/\mathrm{d}t = a_{j+1/2}$)) that propagate from left to right if $a_{j+1/2} \geq 0$, and vice versa, from right to left when $a_{j+1/2} \leq 0$. Therefore (see figure 3.13) the solution of the Riemann for the intermediate state results

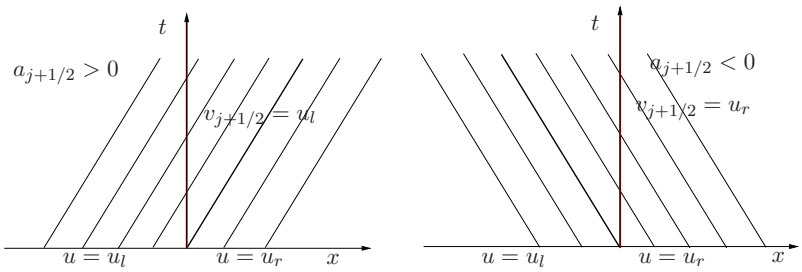


Figure 3.13: Diagram of characteristics for the linearised Riemann problem (3.34)+(3.32).

$$v_{j+1/2} = \begin{cases} u_l & a_{j+1/2} \geq 0 \\ u_r & a_{j+1/2} < 0, \end{cases} . \tag{3.35}$$

and the corresponding numerical flux is

$$\hat{f}_{j+1/2} = f(v_{j+1/2}) = \begin{cases} f(u_l) & a_{j+1/2} \geq 0 \\ f(u_r) & a_{j+1/2} < 0 \end{cases} . \tag{3.36}$$

Note that the numerical flux obtained with the Roe linearization applied to the Godunov scheme is formally identical to the Roe scheme derived in the context of the finite difference scheme (see equation (3.16)). It is also easy to verify that equation (3.36) can be equivalently written as

$$\hat{f}_{j+1/2} = \begin{cases} \min\left(f(u_l), f(u_r)\right) & u_l \leq u_r \\ \max\left(f(u_l), f(u_r)\right) & u_l \geq u_r \end{cases} , \tag{3.37}$$

that, compared with (3.37), shows that the exact solution of the Riemann problem is recovered in all cases, except for the transonic expansion (A-3). In this case, the prediction (3.37) provides $\hat{f}_{j+1/2} = \min(f(u_l), f(u_r))$, while the exact solution is $\hat{f}_{j+1/2} = f(u^*)$.

It is possible to show that schemes built on the basis of this approach (Godunov type) converge to the entropy solution of the conservation equation, provided that the Riemann solver gives solutions consistent with the entropy condition. The approximate Roe Riemann solver does not satisfy this requirement.

An example of the problems associated to the use of this approximate solver is given by the following problem for the Burgers equation

$$u(x,0) = \begin{cases} -1 & x < 0 \\ 1 & x > 0 \end{cases} . \tag{3.38}$$

From (3.38) we derive the following discrete initial conditions

$$U_j^0 = \begin{cases} -1 & j \leq 0 \\ 1 & j > 0 \end{cases} . \tag{3.39}$$

We see that, by applying the numerical flux of the Roe scheme (3.36), we obtain $\hat{f}_{j+1/2} = 1/2$, $\forall j$. This implies that the solution of the scheme (3.28) at step 1 is identical to the initial solution, i.e. $U_j^1 = U_j^0$, $\forall j$. By reducing the mesh spacing the same behavior occurs and the solution thus converges to the following discontinuous solution

$$u(x,t) = \begin{cases} -1 & x < 0 \\ 1 & x > 0 \end{cases} , \tag{3.40}$$

that satisfies the Rankine-Hugoniot jump relations ($0 = s = 1/2(-1+1) = 0$) However, such solution does not satisfy the entropy condition, as can be seen+ from inspection of the characteristics field (see figure 3.14a). In fact, the entropy solution of the problem is

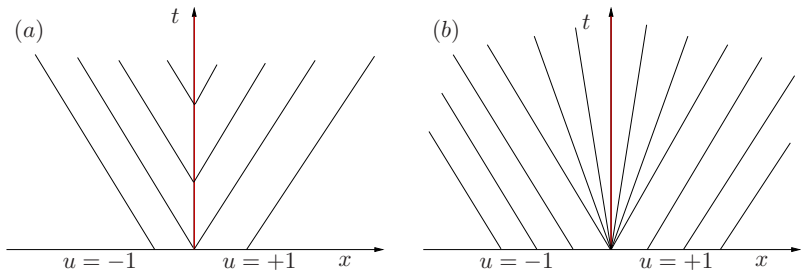


Figure 3.14: Diagram of characteristics for the Burger equation with initial conditions (3.38). (a) solution obtained with Roe scheme; (b) entropy solution.

$$u(x,t) = \begin{cases} -1 & x < -t \\ 1 & x > t \\ x/t & -t < x < t \end{cases} . , \tag{3.41}$$

ans it consists of an expansion fan centered in the origin (as in figure 3.14b).

In the latter example we have verified what we anticipated in section 3.4, i.e. the possible lack of convergence to the correct entropy solution for those schemes characterized by a zero

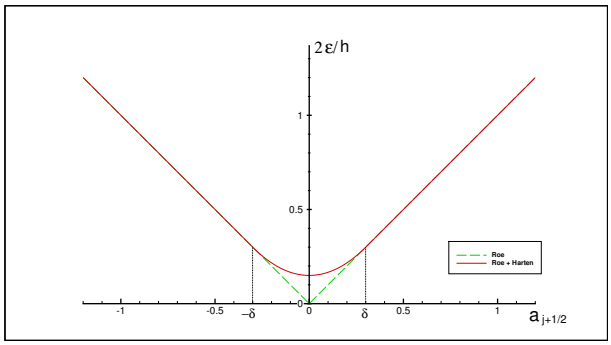


Figure 3.15: Numerical diffusion coefficient for the Roe scheme with and without Harten entropy fix.

value of the coefficient of numerical viscosity $a_{j+1/2}$ The same problem would also happen for the Lax-Wendroff scheme. A simple way to overcome the problem consists in modifying the expression of the numerical flux to avoid a zero value of the coefficient of numerical viscosity. In particular we can modify the numerical flux of the Roe scheme (written in the form of numerical viscosity (3.17)) as follows

$$\hat{f}^n_{j+1/2} = \frac{1}{2}\left(f^n_j + f^n_{j+1}\right) - \frac{\psi(a^n_{j+1/2})}{2}\,\delta U^n_{j+1/2}, \tag{3.42}$$

where the function ψ, called *entropy fix*, can be defined for instance as (Harten)

$$\psi(a_{j+1/2}) = \begin{cases} \frac{1}{2}\left(\delta + a^2_{j+1/2}/\delta\right) & |a_{j+1/2}| \le \delta \\ |a_{j+1/2}| & |a_{j+1/2}| > \delta \end{cases}, \tag{3.43}$$

where $\delta > 0$ is known as *entropy parameter*. Such correction implies that the coefficient of artificial viscosity for $\sigma_{j+1/2} = 0$ is $\varepsilon = \delta h/4 \neq 0$ (see figure 3.15). As a consequence, as can be seen from numerical experiments, the problem to the lack of convergence to the entropy solution is solved.