Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Computational Physics 195 (2004) 474–507

JOURNAL OF COMPUTATIONAL PHYSICS

www.elsevier.com/locate/jcp

Construction of second-order accurate monotone and stable residual distribution schemes for steady problems

Rémi Abgrall *,[1], Mohamed Mezine

Mathématiques Appliquées de Bordeaux, Université Bordeaux I, 351 cours de la Libération, 33 405 Talence Cedex, France

Received 25 April 2003; received in revised form 9 September 2003; accepted 10 September 2003

Abstract

After having recalled the basic concepts of residual distribution (RD) schemes, we provide a systematic construction of distribution schemes able to handle general unstructured meshes, extending the work of Sidilkover. Then, by using the concept of simple waves, we show how to generalize this technique to symmetrizable linear systems. A stability analysis is provided. We formally extend this construction to the Euler equations. Several test cases are presented to validate our approach.

AMS: 65C99; 65M60; 76N10

Keywords: Compressible flow solvers; Residual schemes; Unstructured meshes; Multidimensional up-winding

1. Introduction

The numerical simulation of compressible flows is generally done via some generalization of the one-dimensional Lax–Wendroff scheme. It is well known that this scheme is stable in the energy norm, but does not have any stability property in the maximum norm. The simulation of flows with strong discontinuities can only be performed with schemes having properties in the maximum norm, because we want solutions without numerical oscillation. This goal can be reached by modifications of the Lax–Wendroff scheme, either by adding dissipative and fourth-order terms monitored by complex ad hoc, problem dependent, sensors, or via more automatic methods coming from the theory of scalar nonlinear schemes, see e.g. [14,15] and the numerous references therein.

In each case, the approximation relies strongly on the structure of the one-dimensional problem. If the flow is represented by the values of the conservative unknowns at the mesh points, a consequence is, for

* Corresponding author. Tel.: +33-5-40-00-60-68; fax: +33-5-40-00-26-26.
E-mail addresses: abgrall@math.u-bordeaux.fr (R. Abgrall), mezine@math.u-bordeaux.fr (M. Mezine).

[1] Institut Universitaire de France.

0021-9991/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2003.09.022

example, that the accuracy of the solution degrades when going to multidimensional problems mainly with irregular meshes. The analysis via Taylor expansion and the equivalent equation, that enables to study the formal accuracy of a scheme, is strictly valid only for one-dimensional problems. It can be extended to multidimensional problems only if the structure of the mesh allows special algebraic combinations resulting from the symmetries of the mesh. If the flow is represented by the averaged values of the conservative unknowns on control volumes, there exists techniques to develop high order accurate schemes, for example the ENO or WENO schemes [1,13,16,17,27]. In that case the analysis via the equivalent equation is different, but the main problem becomes a reconstruction problem, on variables that have to be chosen carefully [4]. The price to pay is a very large extension of the computational stencil. Even in that case, standard techniques use Riemann solvers, so the fluxes are still computed in a one-dimensional spirit, resulting in large errors as analyzed by van Leer [26]. Hence the overall quality of the scheme may be quite disappointing, in many cases.

In order to tackle these two problems – compactness of the stencil, effective accuracy of the solution – at least two classes of methods have emerged in recent years, the discontinuous Galerkin schemes (DG), and the residual distribution schemes (RD). Though different in spirit, they have a common core at least in their "unstabilized" versions, the residual property. This property that we discuss below in the framework of RD schemes, allows to show the formal accuracy of the scheme on a very general mesh. The DG schemes use a discontinuous polynomial representation of the unknowns that is a generalization of what is done in finite volume schemes. The solution is updated via evaluation of fluxes, and the stabilization mechanism is obtained by very similar techniques as in classical finite volume. The net effect of this is to loose the residual property.

On the contrary, the RD schemes use a pointwise representation of the solution, like in finite difference schemes. The unknowns are updated by evaluating the amount of residual sent to the vertices, and the stabilization mechanism can be similar to artificial viscosity, as in the SUPG-like finite element method [18,20], or inspired by the nonlinear techniques of the so-called high resolution schemes [15,19]. One can also take into account the genuinely multidimensional structure of the problem [23].

In this paper, we consider schemes of the RD class. Our goal is to propose a systematic construction of robust, high order and compact schemes on general meshes. The schemes are derived in such a way that *all* the decisions are made on elements only, the neighboring elements play no role in the way the residuals are sent to the element vertices. In that respect, the schemes we consider are the most compact possible. This is a very pleasant property for parallelism issues.

This type of schemes has received recently a lot of interest, one may quote the pioneering work of Roe, Sidilkover, Deconinck and their co-workers [11,22,23], and also more recently [2,3,10,25]. However, the solutions that are proposed in these contributions are not fully satisfactory, in particular, they may be not robust enough in some situations.

In this paper, we propose to reconsider the approach described in [22] for a very particular scheme, the first-order scalar N scheme. From it, one can construct Struijs' PSI scheme [11] that is second-order. These schemes have the special property that, for a triangular mesh, on each element, at least one residual vanishes, allowing very simple algebra. We show how to extend this approach to more general schemes where we do not need to assume a special structure of the underlying first-order scheme. This approach, via a stability analysis using simple waves, enables to extend the method to linear symmetrizable PDEs. Only steady problems are considered here, the unsteady case is discussed in [6].

The format of the paper is the following. We first recall the formalism of RD schemes, in particular we recall what is necessary to get second-order accuracy. Then we analyze the scalar problems and show how to construct second-order schemes. Going to system problems, we show some special properties of the Lax–Friedrichs scheme and the N scheme for symmetrizable systems. These stability properties enable to justify formally our construction. Several numerical tests, for linear and nonlinear PDEs demonstrate the qualities of our approach.

2. The residual distribution schemes

Let us consider the steady equation of fluid mechanics

$$\operatorname{div}\mathbf{F}(W) = 0, \quad x \in \Omega, \ t > 0, \tag{1}$$

supplemented by boundary conditions at the inflow, outflow and solid boundaries. The vector of conservative variables W is defined by

$$W = (\rho, \rho\mathbf{u}, E),$$

where ρ is the density, $\mathbf{u}$ is the velocity, and $E = \rho\epsilon + \frac{1}{2}\rho\mathbf{u}^2$ is the total energy, ϵ being the internal energy per unit volume. The flux is defined by

$$\mathbf{F} = \begin{pmatrix} \rho\mathbf{u} \\ \rho\mathbf{u}\otimes\mathbf{u} + p\mathbf{Id} \\ \mathbf{u}(E+p) \end{pmatrix}.$$

Lastly, the pressure is defined by the perfect gas equation of state

$$p = (\gamma - 1)\left(E - \frac{1}{2}\rho\mathbf{u}^2\right).$$

In the sequel, the ratio of specific heats γ is assumed to be constant.

In order to approximate (1), we consider a triangular mesh where the elements are denoted by $\{T_{jt}\}_{jt=1,n_t}$, the vertices are denoted by $\{M_{\mathrm{is}}\}_{\mathrm{is}=1,\mathrm{ns}}$. Strictly speaking, the vertices of T have to be indexed in the list $\{M_{\mathrm{is}}\}_{\mathrm{is}=1,\mathrm{ns}}$, namely M_{i_1}, M_{i_2}, M_{i_3}. When there is no ambiguity, we denote them by i_1, i_2, i_3 or more simply $1, 2, 3$. The vector $\vec{n}_i$ is the scaled inward vector normal to the boundary of T, opposite to the vertex i, i.e.

$$\vec{n}_i = 2|T|\nabla\Lambda_i,$$

where Λ_i is the barycentric coordinate at M_i.

The following iterative RD scheme approximates (1):

$$|C_i|\frac{W_i^{n+1} - W_i^n}{\Delta t} + \sum_{T, M_i \in T} \Phi_i^T = 0, \tag{2}$$

and we consider the limit, if it exists, of W_i^n when $n \to +\infty$. In (2), W_i^n is an approximation of W, solution of (1) at (M_i, t_n), $|C_i|$ is the area of the dual control volume associated to M_i, Δt is the (pseudo-) time step, and Φ_i^T stands for the residual sent by the element T to the vertex M_i. The residuals satisfy the following conservation relations:

$$\sum_{j=1,3} \Phi_{i_j}^T = \int_T \operatorname{div}\mathbf{F}^h \, \mathrm{d}x := \Phi^T. \tag{3}$$

In (3), $\mathbf{F}^h$ is a continuous interpolation of the flux $\mathbf{F}$ that converges in L^1_{loc} to $\mathbf{F}$. In [5], we show that under the classical assumptions of the Lax–Wendroff theorem, the limit solution of (2) are weak solutions of (1).

An example is given by finite volume type schemes on triangular meshes, another one by the SUPG scheme, see [2] for details.

Besides the conservation relation (3), several other requirements are needed: the scheme has to be stable and accurate. The stability is generally met by using a monotonicity preserving scheme. We briefly recall this concept for a scalar problem, we come back to it later in the system case where things are much less clear.

2.1. Monotonicity preserving schemes

In practice, all the known RD schemes can be written as

$$\Phi_i^T = \sum_{M_j \in T, M_i \neq M_i} c_{ij}^T (u_i - u_j). \tag{4}$$

For this scheme to be L^∞ stable, it is enough that

$$c_{ij}^T \geqslant 0 \quad \text{for all } i, j. \tag{5}$$

The stability is obtained thanks to a CFL-like condition [11]. This is the so-called monotonicity preserving condition.

2.2. Accuracy: the linear preserving (LP) condition

We briefly recall the analysis of [2]. It is shown that a *converged* RD scheme produces a formally second-order accurate solution of the *steady problem* (1) under the following three requirements:

1. The mesh is regular.
2. The approximation $\mathbf{F}^h$ is second-order accurate on smooth solutions.
3. For any smooth solution of (1), $\Phi_i^T(W) = \mathcal{O}(h^3)$ for any vertex M_i and any triangle T such that $M_i \in T$.

In most cases, the third condition is met by imposing that there exists a family of uniformly bounded coefficients (or matrices for system problems) β_i^T such that

$$\Phi_i^T = \beta_i^T \Phi^T.$$

This is the LP condition introduced in [11] which is satisfied by the SUPG scheme and the PSI scheme of Struijs [11] that we recall later.

It is known that it is not possible to have a linear scheme that is both monotonicity preserving and linearity preserving: this is Godunov's theorem [11]. The schemes that satisfy both requirements must be nonlinear. The construction of such schemes is the topic of the next section.

3. Discretization of scalar equations

Several constructions of monotonicity preserving scheme exist, so we start by indicating some motivations for revisiting the problem. Then we provide a general construction of LP schemes starting from a monotone first-order scheme, and then provide numerical examples.

3.1. Problem statement and relations to previous constructions

Defining $\langle x, y \rangle$ as the the dot-product of the vectors x and y, we consider the problem

$$\begin{aligned} &\langle \vec{\lambda}, \nabla u \rangle = 0, && x \in \Omega, \ t > 0, \\ &u = g && \text{on } \Gamma_-, \end{aligned} \tag{6}$$

where Γ_- is the inflow boundary of $\Gamma = \partial\Omega$. If the unknown u is piecewise linearly interpolated, the total residual Φ^T is given by

$$\Phi^T = \sum_{j=1}^{3} k_j u_j,$$

where

$$k_j = \frac{1}{2}\langle \vec{\lambda}, \vec{n_j} \rangle.$$

We notice that $\sum_{j=1}^{3} k_j = 0$. Here, and until the end of the paper, we have identified the vertices of T with the indices $j = 1, 2, 3$ because there is no ambiguity. Similarly, we drop the superscript T in $\Phi_i^T \equiv \Phi_i$.

We provide three examples of monotonicity preserving schemes. The first one is the Rusanov scheme,

$$\Phi_i = \frac{1}{3}\left(\Phi - \alpha \sum_{j \neq i} (u_i - u_j) \right) \tag{7}$$

with $\alpha \geqslant \max_i |k_i|$, so that $c_{ij} = \frac{1}{3}(\alpha - k_j) \geqslant 0$. We have clearly $\sum_j \Phi_j = \Phi$.

Another example is given by the N (narrow) scheme [11]. It can be written as

$$\Phi_i = k_i^+ (u_i - \widetilde{u}), \tag{8}$$

where $\widetilde{u}$ is obtained by recovering the conservation, i.e.

$$\widetilde{u} = \left(\sum_j k_j^- \right)^{-1} \left(\sum_j k_j^- u_j \right).$$

The scalar $n := (\sum_j k_j^-)^{-1}$ is always defined unless $\vec{\lambda} = 0$. [2] This scheme can be considered as a conservative method of characteristics. It is monotone under a CFL-like condition because

$$\Phi_i = \sum_j k_i^+ n k_j^- (u_i - u_j),$$

hence $c_{ij} = k_i^+ n k_j^- \geqslant 0$. A last example is provided by the classical upwind scheme. None of these scheme is linear preserving.

In order to get a monotonicity preserving LP scheme, several constructions exist, but all of them use the scalar N scheme as a base scheme. One may quote the PSI scheme of Struijs [9], Sidilkover's construction of the same scheme [22]. Another method is the hybridization technique of [2,24] which consists in blending a first-order monotone scheme (residual $\Phi_i^{(1)}$) and a second-order LP scheme (residual $\Phi_i^{(2)}$),

$$\Phi_i = \ell \Phi_i^{(1)} + (1 - \ell)\Phi_i^{(2)}.$$

One may also quote the B–scheme by Deconinck et al. [24]. It is not monotonicity preserving, even though it is not easy to produce an oscillatory counter–example. In both cases, the first-order scheme is the N scheme, and the second-order scheme is the LDA scheme. In [2], we show that a special choice of ℓ leads to the PSI scheme, but other choices are possible.

The solution provided by the blending technique seems interesting because it may be thought at first glance that any monotone first-order scheme $\{\Phi_i^{(1)}\}$ and any LP scheme $\{\Phi_i^{(2)}\}$ might be blended together, leading to a richer class of schemes. This is not true because the two schemes must have some compatibility relations otherwise the blended scheme might indeed reduce to the first-order one, i.e. $\ell \simeq 1$. An example where $\ell \simeq 1$ much too often is the blending of the N scheme and the Lax–Wendroff scheme

$$\Phi_i = \frac{\Phi}{3} - \frac{\Delta t}{2} k_i \Phi.$$

The problem here is that the N scheme is upwind, while the Lax–Wendroff is not, so that we might have $\Phi_i = 0$ for the N scheme and $\Phi_i \neq 0$ for the Lax–Wendroff one. Since ℓ is defined by a relation of the type

[2] Since for any i, $\Phi_i \to 0$ uniformly if $\vec{\lambda} \to 0$, there is no definition problem in the case $\vec{\lambda} = 0$.

$$\ell = \max_{i=1,3} \varphi(r_i),$$

where r_i is the ratio of the second-order residual versus the first-order one and φ a real valued function the graph of which satisfies some geometrical constraints similar to what happens in the TVD framework (see [2] for details), we have in general $\ell = 1$. In some cases, the LP condition is not so clear, for example, the blending of the Rusanov and the Lax–Wendroff scheme is monotonicity preserving by construction but its LP property is not clear.

This is why we have tried to develop another construction, partially inspired from Sidilkover [22], but in our opinion more powerful since non-triangular elements can be considered. The construction basically relies on the existence of a first-order monotone scheme, and provides a systematic way of defining an LP monotone scheme.

3.2. Construction

We start from a first-order monotonicity preserving scheme

$$\Phi_i = \sum_j c_{ij}(u_i - u_j), \quad c_{ij} \geqslant 0.$$

To simplify the notations, we present the technique on triangular elements, but the extension to more general elements is obvious though tedious.

As Sidilkover [22], we want to construct a residual Φ_i^* such that:

1. The scheme defined by Φ_i^* is monotonicity preserving.
2. The scheme is conservative, i.e. $\sum_j \Phi_j^* = \sum_j \Phi_j = \Phi$.
3. The scheme is LP, more precisely, we want $\beta_i := \Phi_i^*/\Phi$ to be bounded.

The first condition can be written

$$1 - \mu_i = \frac{\Phi_i^*}{\Phi_i} \geqslant 0,$$

i.e. $\mu_i \leqslant 1$. The second condition is written

$$\sum_j \mu_j \Phi_j = 0.$$

Before going further, let us examine the L^∞ stability condition. The stability condition of the first-order scheme is (we put back temporarily the superscript T)

$$\Delta t \max_{T \ni i} \frac{\sum_{j \in T} c_{ij}^T}{|T|} \leqslant 1,$$

i.e.

$$\Delta t \leqslant \Delta t^{(1)} := \min_i \left(\max_{T \ni i} \frac{\sum_{j \in T} c_{ij}^T}{|T|} \right)^{-1}. \tag{9}$$

Thus, the natural stability condition of the new scheme is

$$\Delta t \leqslant \Delta t^{(2)} := \min_i \left(\max_{T \ni i} \frac{(1 - \mu_i) \sum_{j \in T} c_{ij}^T}{|T|} \right)^{-1}. \tag{10}$$

The monotonicity constraint is $\mu_i \leqslant 1$ only: μ_i is allowed to be negative. It is clear that if μ_i is too negative, we might have $\Delta t^{(2)} \ll \Delta t^{(1)}$. If we want that the maximum time step of the first-order scheme is of the order of that of the new scheme, it is better that $|1-\mu_i|$ is not too large, that is μ_i is not too negative. For this reason, we ask

$$0 \leqslant \mu_i \leqslant 1. \tag{11}$$

Using the previous notations, the LP condition is

$$\beta_i = (1-\mu_i)\frac{\Phi_i^*}{\Phi} \quad \text{(bounded)}.$$

For that reason, we demand $\mu_i = 1$ as often as possible.

To summarize, the problem is: find $\{\mu_j\}_{j=1,N} \in [0,1]^N$ such that

$$\begin{aligned} &\textstyle\sum_{j=1}^{N} \mu_j \Phi_j = 0,\\ &\{\mu_j\}_{j=1,N} \in [0,1]^N,\\ &\mu_j = 1 \quad \text{(as often as possible)}. \end{aligned} \tag{12}$$

Assume now $N = 3$ for simplicity. We are looking for solutions where $\mu_j = 1$ (i.e. $\Phi_j^* = 0$) as often as possible, and for which the new scheme depends continuously on the parameters.

We may assume $\Phi_1 \neq 0$. The equation in (12) becomes

$$\mu_1 = \mu_2\left(-\frac{\Phi_2}{\Phi_1}\right) + \mu_3\left(-\frac{\Phi_3}{\Phi_1}\right).$$

The solutions we are seeking are those for which $\mu_1 = 1$ as often as possible, without violating the monotonicity condition.

(1) If $(-\Phi_2/\Phi_1)(-\Phi_3/\Phi_1) > 0$. The line defined by

$$1 = \mu_2\left(-\frac{\Phi_2}{\Phi_1}\right) + \mu_3\left(-\frac{\Phi_3}{\Phi_1}\right)$$

has either an empty intersection with $[0,1]^2$, or has two intersection points. For symmetry reasons, we assume $\mu_1 = \mu_2 = \mu$, hence

$$\mu = -\frac{\Phi_1}{\Phi_2 + \Phi_3}.$$

(a) If $\mu < 0$, then we take $\mu_2 = \mu_3 = \mu_1 = 0$ and then $\Phi_i^* = \Phi_i$, $i = 1, 3$.

(b) If $0 < \mu < 1$, then

$$\Phi_1^* = 0, \quad \Phi_2^* = \frac{\Phi_2}{\Phi_2 + \Phi_3}\Phi, \quad \Phi_3^* = \frac{\Phi_3}{\Phi_2 + \Phi_3}\Phi.$$

(c) If $\mu > 1$, then

$$\Phi_1^* = \Phi, \quad \Phi_2^* = 0, \quad \Phi_3^* = 0.$$

(2) If $(-\Phi_2/\Phi_1)(-\Phi_3/\Phi_1) < 0$, the line

$$1 = \mu_2\left(-\frac{\Phi_2}{\Phi_1}\right) + \mu_3\left(-\frac{\Phi_3}{\Phi_1}\right)$$

cuts the boundary $0 \leqslant \mu_1 \leqslant 1$ or $0 \leqslant \mu_2 \leqslant 1$ at most one point. The intersection points are

$$\mu_2 = 1, \ \ \mu_3 = -\frac{\Phi_2 + \Phi_1}{\Phi_3} \quad \text{and} \quad \mu_3 = 1, \ \ \mu_2 = -\frac{\Phi_3 + \Phi_1}{\Phi_2}.$$

(a) If $\mu_2 = 1, \mu_3 = -(\Phi_2 + \Phi_1)/\Phi_3 < 0$, we set

$$\Phi_1^* = \Phi_1 + \Phi_2, \quad \Phi_2^* = 0, \quad \Phi_3^* = \Phi_3.$$

(b) If $\mu_2 = 1, \mu_3 = -(\Phi_2 + \Phi_1)/\Phi_3 \in [0, 1]$, we set

$$\Phi_1^* = 0, \quad \Phi_2^* = 0, \quad \Phi_3^* = \Phi.$$

(c) If $\mu_3 = 1, \mu_2 = -(\Phi_3 + \Phi_1)/\Phi_2 \in [0, 1]$, we set

$$\Phi_1^* = 0, \quad \Phi_2^* = \Phi, \quad \Phi_3^* = 0.$$

(d) If $\mu_3 = 1, \mu_2 = -(\Phi_3 + \Phi_1)/\Phi_1 < 0$, we set

$$\Phi_1^* = \Phi_1 + \Phi_3, \quad \Phi_2^* = \Phi_2, \quad \Phi_3^* = 0.$$

In each case, one can easily check that the $\beta_i \in [0, 1]$ and depends continuously on the data.

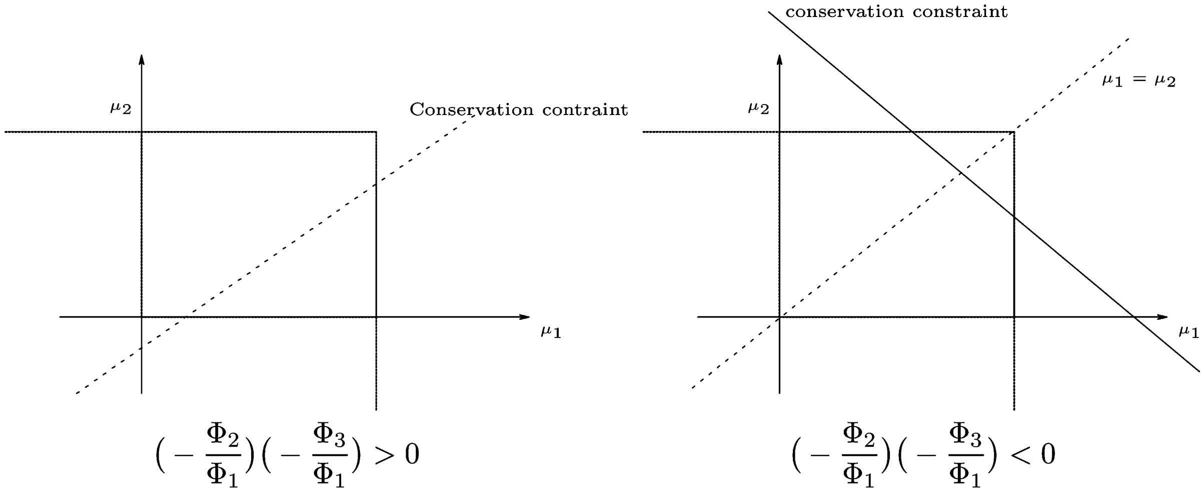
In this construction, the node $i = 1$ plays a special role, so the scheme is numbering dependent. This is why a better solution is to average the three solutions obtained by making the vertices $i = 1, \dots, 3$ play a pivot rôle. This scheme is now continuous and independent of the numbering of the mesh points.

A simpler solution is to make the construction on the pivot index i_0 the index for which $|\Phi_j|$ is maximum. This is the solution we use in all the numerical illustrations. The previous solution gives results that are indistinguishable from those obtained by this one (see Fig. 1).

3.3. Numerical illustrations

We apply the above construction to

$$\begin{aligned} &\frac{1}{2}\frac{\partial u^2}{\partial x} + \frac{\partial u}{\partial y} = 0, \qquad (x, y) \in [0, 1]^2, \\ &u(x, 0) = 1.5 - x, \quad u(0, y) = 1.5, \\ &u(1, y) = 0.5. \end{aligned} \tag{13}$$



$$\left(-\frac{\Phi_2}{\Phi_1}\right)\left(-\frac{\Phi_3}{\Phi_1}\right) > 0$$

$$\left(-\frac{\Phi_2}{\Phi_1}\right)\left(-\frac{\Phi_3}{\Phi_1}\right) < 0$$

Fig. 1. Some possible configurations in the solution of (12) for $N = 3$.

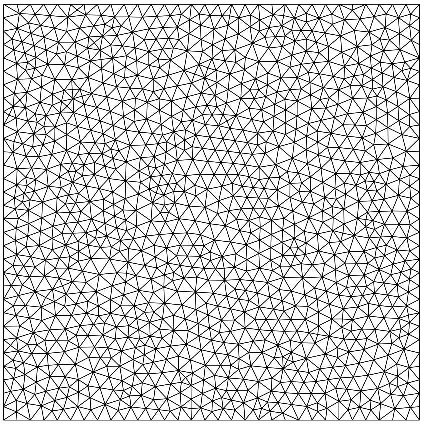


Fig. 2. Mesh for the problem (14).

The PDE (13) is nonlinear. The link between (13) to a linear PDE is obtained via a conservative linearisation [9], i.e. we determine for each triangle $\bar{u}$ such that

$$\int_T \left(\frac{1}{2} \frac{\partial u^2}{\partial x} + \frac{\partial u}{\partial y} \right) \mathrm{d}x \, \mathrm{d}y = \bar{u} \int_T \frac{\partial u}{\partial x} \mathrm{d}x \, \mathrm{d}y + \int_T \frac{\partial u}{\partial y} \mathrm{d}x \, \mathrm{d}y. \tag{14}$$

In the left-hand side of (14), u is piecewise linearly interpolated. The obvious solution is $\bar{u} = \frac{1}{3} \sum_{j=1,3} u_j$.

The total residual

$$\Phi = \bar{u} \int_T \frac{\partial u}{\partial x} \mathrm{d}x \, \mathrm{d}y + \int_T \frac{\partial u}{\partial y} \mathrm{d}x \, \mathrm{d}y$$

is distributed by mean of any of the three schemes described above, the Rusanov scheme, the N scheme and the first-order upwind scheme. The upwind scheme uses the classical one-dimensional Murman–Roe scheme adapted to (13), rewritten in the distribution framework as in [2]. For comparison purposes, we display the mesh on Fig. 2 and the solution obtained by a standard second-order ENO scheme on unstructured meshes.

The scheme constructed from the Rusanov scheme (resp. the N scheme, the one-dimensional upwind scheme) is denoted by L-Rusanov (resp. PSI and L-upwind). The solutions are plotted on Figs. 3 and 4. We plot cross-sections in the discontinuity in Fig. 5. On Fig. 6, we have displayed cross-section plots in the fan.

We see that the quality of the results is *always* better for the second-order distribution schemes than for the second-order ENO scheme. This is a consequence of the LP property. Among them, there is a hierarchy: the PSI scheme is the best, the results for the upwind scheme are almost identical, followed by those for L-Rusanov. For the latter scheme, the results seem a bit wiggly in the fan. This is not a stability problem, since the results are converged. We believe that since the Rusanov scheme is very dissipative, the limitation mechanism that we propose is probably too over-compressive. The way to remedy to this drawback is not known.

4. Going to linear hyperbolic systems

The main difficulty to step from scalar to symmetrizable systems of PDEs of the type

$$A \frac{\partial U}{\partial x} + B \frac{\partial U}{\partial y} = 0 \tag{15}$$

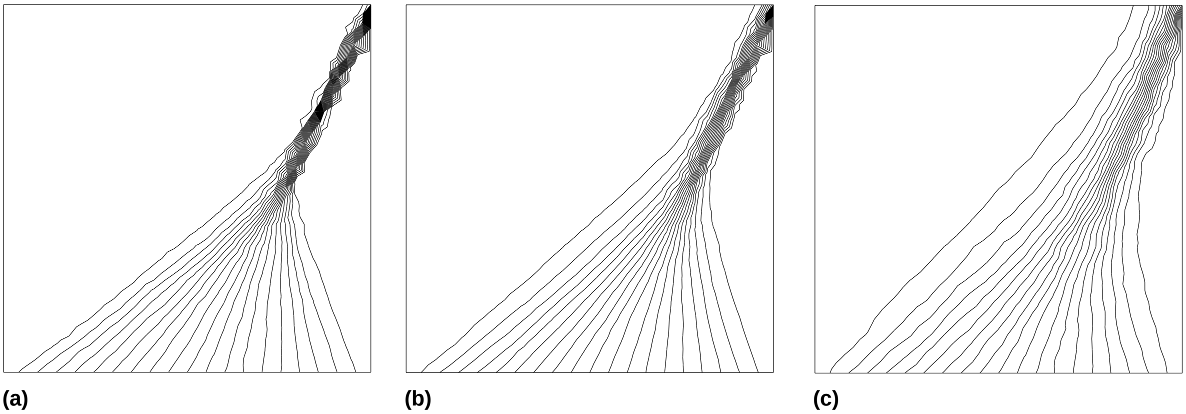


Fig. 3. (a) N scheme, (b) upwind scheme, (c) Rusanov scheme.

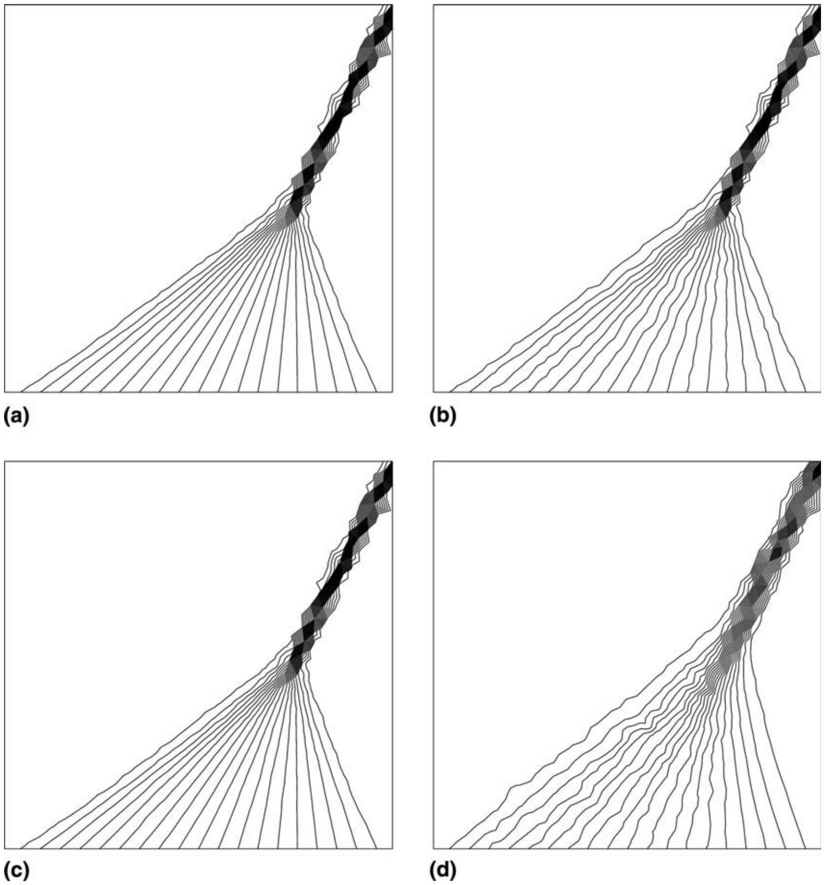


Fig. 4. (a) PSI, (b) L-Rusanov, (c) L-upwind, (d) second-order ENO scheme.

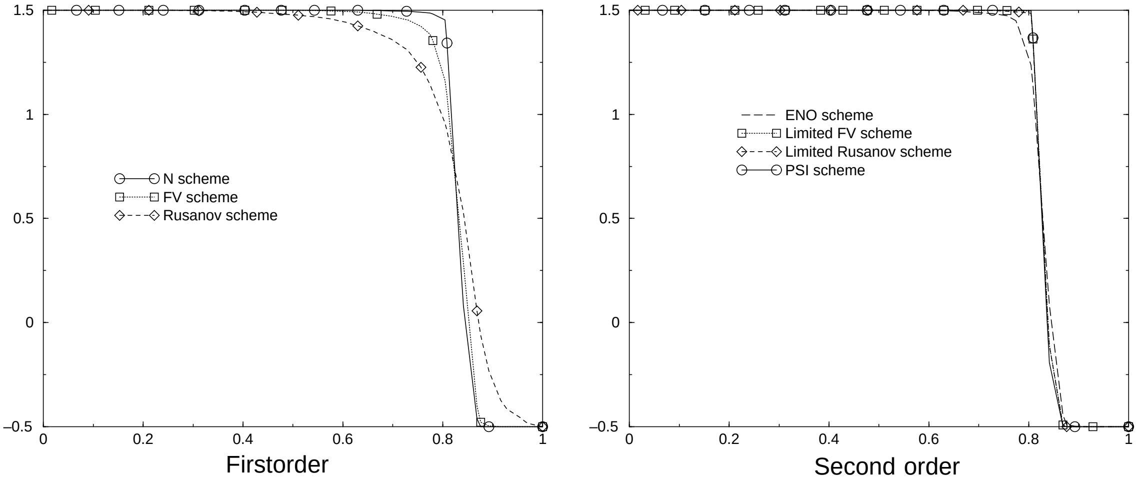


Fig. 5. Cross-sections in the shock for the problem (14).

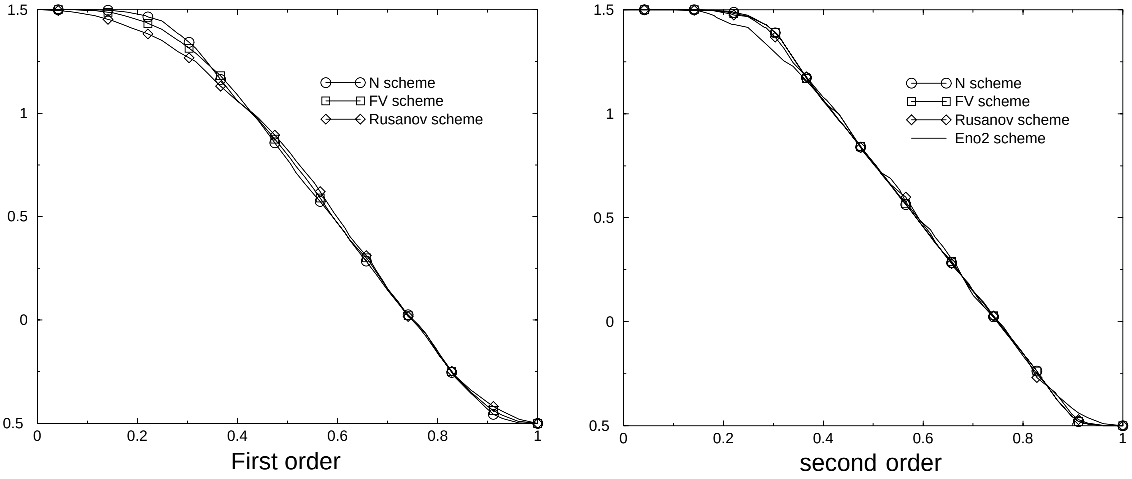


Fig. 6. Cross-section in the fan for the problem (14).

supplemented by boundary conditions, is that, in general, the two Jacobian matrices A and B do not commute. The consequence is that there exists no basis of common eigenvectors to the matrices A and B. However, the analysis of the Cauchy problem

$$\frac{\partial U}{\partial t}+A\frac{\partial U}{\partial x}+B\frac{\partial U}{\partial y}=0 \quad \text{(supplemented by boundary and initial conditions)} \tag{16}$$

shows that the solution is piecewise smooth, without high frequency oscillations at the discontinuities when the initial and boundary conditions are piecewise smooth and the Jacobian matrices are symmetric.

Unfortunately, systems (15), (16) are well posed only in L^2 or in Hilbert spaces which use in the analysis of a numerical scheme seems very complex [7,8]. It seems legitimate, and all the numerical experiments support this, to seek for a stability criterion that has a L^∞ flavor in order to control the oscillations of an approximating scheme for (15).

In the following, we first recall some particular distribution schemes. We show some of their properties, in particular when applied to simple waves. These properties justify, in our opinion, the heuristic arguments we use to construct stable LP schemes. In order to simplify the text, we assume in the following that A and B are symmetric. The discussion can easily be generalized to symmetrizable systems by changing the canonical dot-product to the one associated to the symmetrization variables.

4.1. Some RD schemes for (15)

If one interpolates U linearly in each triangle, the total residual Φ associated to (15) is

$$\Phi = \int_T \left(A\frac{\partial U}{\partial x} + B\frac{\partial U}{\partial y} \right) \mathrm{d}x\,\mathrm{d}y = \sum_{j=1,3} K_j U_j, \tag{17}$$

where U_j denotes the conservative variables at the vertices of T and the matrices K_j are

$$K_j = n_x^j A + n_y^j B,$$

where n_x^j and n_y^j represent the components of $\vec{n}_j$. We denote by K_n the matrix $K_n = n_x A + n_y B$ where $\vec{n} = (n_x, n_y)$.

Thanks to these notations, the Rusanov scheme is

$$\Phi_i = \frac{1}{3}\left(\Phi - \alpha \sum_j (U_i - U_j) \right) \tag{18}$$

with $\alpha \geqslant \max_j \|K_j\|$. Another example is provided by the upwind finite volume scheme formulated as a RD scheme.

The system N scheme [25] can be written as

$$\Phi_i = K_i^+ (U_i - \widetilde{U}), \tag{19}$$

where $\widetilde{U}$ is computed to recover the conservation property

$$\left(\sum_j K_j^- \right) \widetilde{U} = \sum_j K_j^- U_j.$$

The matrix $\sum_j K_j^-$ is invertible whenever A and B have no common eigenvectors [2]. When there exist common eigenvectors, $\sum_j K_j^-$ may be no longer invertible but the matrices NK_j^- have a meaning and in each case, the scheme (19) is well defined [2].

These three examples satisfy the conservation relation $\sum_j \Phi_j = K_j U_j$.

Lastly, we notice that the scheme (2) can be rewritten as

$$U_i^{n+1} = \sum_{T, M_j \in T} \frac{|T|}{|C_i|} \widetilde{U}_i^{n+1} \tag{20}$$

with

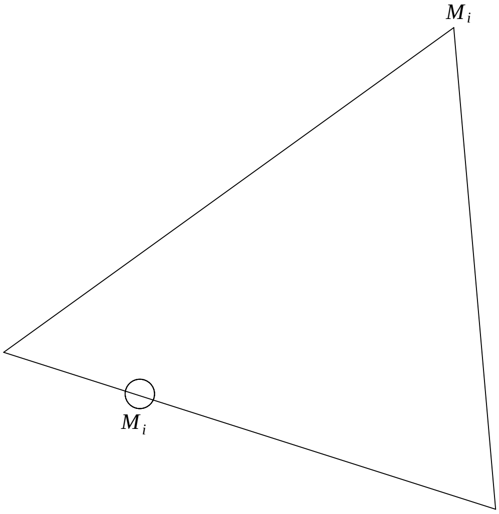


Fig. 7. Definition of M_i' in (A.4).

$$\widetilde{U}_i^{n+1} = U_i^n - \frac{\Delta t}{|T|} \Phi_i^T. \tag{21}$$

This enables to localize the analysis on each of the triangles of the mesh.

4.2. Stability analysis by simple waves

We call simple wave a solution of the type

$$U(x) = \mathbf{C} + \beta \langle \vec{n}, \mathbf{x} \rangle \mathbf{r},$$

where $\mathbf{C}$ is a constant vector, $\mathbf{x}$ is any point, $\mathbf{r}$ is a normalized eigenvector of the matrix $K_{\vec{n}}$ and $\beta \in \mathbb{R}$ is arbitrary. The function U is linear on T, its nodal values are still denoted by U_j. Lastly, we still denote $\varphi_j = \langle U_j, \mathbf{r} \rangle$, i.e. $U_j = \varphi_j \mathbf{r} + \mathbf{C}$.

We notice that

$$A \frac{\partial U}{\partial x} + B \frac{\partial U}{\partial y} = K_{\vec{n}} \mathbf{r} = \lambda(\mathbf{r}) \mathbf{r},$$

where $\lambda(\mathbf{r})$ is the eigenvalue associated with $\mathbf{r}$, hence Φ is proportional to $\mathbf{r}$.

Our aim is to justify the experimental fact that the Rusanov, N and finite volume schemes are monotonicity preserving, i.e., there is no creation of numerical oscillation.

4.2.1. Wave decomposition

The first remark is that, locally in a triangle, U^n is the sum of simple waves. In fact, in T,

$$U(\mathbf{x}) = \sum_j U_j \Lambda_j(\mathbf{x})$$

with

$$\Lambda_j(\mathbf{x}) = \frac{\langle \vec{n}_j, \mathbf{x} \rangle}{2|T|} + C_j.$$

Calling $\{\mathbf{r}_\xi\}$ an orthogonal basis of eigenvectors of $K_{\vec{n}}$, we may write

$$U(\mathbf{x}) = \frac{1}{2|T|}\sum_{j=1,3}\sum_{\xi}\langle U_j, \mathbf{r}_\xi\rangle\langle \vec{n}_j, \mathbf{x}\rangle \mathbf{r}_\xi + \mathbf{C}. \tag{22}$$

This shows that a piecewise linear U is a sum of simple waves.

The decomposition is not unique. Depending on the scheme, we might need adapted wave decomposition to prove our claim, but the central idea is contained in (22).

4.2.2. Results of the stability analysis

Within a triangle T, the piecewise linear interpolation of U_j can be decomposed as a sum of simple waves

$$U(\mathbf{x}) = \sum_{\sigma:\ \text{wave}} \varphi_\sigma(\mathbf{x})\mathbf{r}_\sigma, \tag{23}$$

where φ_σ is of the form

$$\varphi_\sigma(\mathbf{x}) = \alpha_\sigma\langle \vec{n}_\sigma, \mathbf{x}\rangle + C_\sigma$$

with $\alpha_\sigma \in \mathbb{R}$, $\vec{n}_\sigma$ is a unitary vector and $\mathbf{r}_\sigma$ is an eigenvector of $\mathbf{K}_{\vec{n}_\sigma}$.

The three schemes considered here are linear, so the residuals sent to node M_j is the sum of the residual sent to this node by $U_\sigma(\mathbf{x}) = \varphi_\sigma(\mathbf{x})\mathbf{r}_\sigma$, namely

$$\Phi_i = \sum_{\sigma:\text{wave}} \Phi(U_\sigma(\mathbf{x}))_i$$

with some abuse of language.

The analysis carried out in Appendix A for the Rusanov and the N schemes can be extended without difficulty to the finite volume scheme. We show that for any simple wave U_σ, the updated quantities $\widetilde{U}_i$ defined by

$$\widetilde{U}_i = U_\sigma(M_i) - \frac{\Delta t}{3|T|}\Phi_i(U_\sigma)$$

satisfy

$$\|\widetilde{U}_i\| \leqslant \max_{M_j\in T}\|U_\sigma(M_j)\| \tag{24}$$

under a CFL-type condition.

Gathering the properties and relations (20), (21), (23) and (24), we say that the scheme is stable and monotone. We are not able to exhibit any norm for which a relation like

$$\|U_i^n\| \leqslant C(U^0)$$

would be true in general under a CFL-like condition. From [7,8], the classical L^p norms are not suitable, maybe the less standard $L_{p,\alpha}$ norms of [8] for which the Cauchy problem is well posed. But we have not found a way to use them in a practical way.

From an experimental point of view, the conditions (20), (21), (23) and (24) seem sufficient; this is why we call this stability.

4.3. Construction of LP schemes for systems

We present now a method which, starting from a stable monotone scheme, enables the construction of monotone second-order schemes at steady state. These LP schemes have the residual

$$\Phi_i = \mathbf{B}_i \Phi,$$

where $\mathbf{B}_i$ is a matrix. They are oscillation free. We show that they also satisfy the stability requirement described in Section 4.2.2.

4.3.1. Construction

The idea of the construction is the following. Starting from a monotone scheme, it is possible to decompose the solution as a sum of simple waves indexed by σ. The residual can then be splitted into a sum of residuals, Φ_i^σ, each of them acting on a single simple wave. These sub-residuals can be written as a positive weighted sum of difference,

$$\Phi_i^\sigma = \left(\sum_j c_{ij}^\sigma \left(\varphi_j^\sigma - \varphi_i^\sigma\right)\right)\mathbf{r}_\sigma. \tag{25}$$

Here, the states U_j are described by mean of a sum of simple waves

$$U_j = \sum_{\sigma:\text{ wave}} \varphi_j^\sigma \mathbf{r}_\sigma.$$

This is a geometrical property of the scheme: it states that simple waves evolve in a non-oscillatory manner. This non-oscillatory behavior of the scheme should be independent of the way we choose to describe it.

Thus, the idea is to choose an orthonormal basis $\{\mathbf{t}_l\}_l$, to notice that

$$\sum_j \langle \Phi_j, \mathbf{t}_l\rangle = \langle \Phi, \mathbf{t}_l\rangle$$

and to interpret the coefficients $\langle \Phi_j, \mathbf{t}_l\rangle$ as scalar residuals to which we apply the limitation technique of Section 3. We construct residuals $\varphi_{l,i}^*$ such that

$$\begin{aligned} &\textstyle\sum_i \varphi_{l,i}^* = \langle \Phi, \mathbf{t}_l\rangle, \\ &\varphi_{l,i}^* = \beta_i^l \langle \Phi, \mathbf{t}_l\rangle, \end{aligned} \tag{26}$$

with $\beta_i^l \in [0,1]$. The residual

$$\Phi_i^* = \sum_l \varphi_{l,i}^* \mathbf{t}_l \tag{27}$$

can be rewritten as

$$\Phi_i^* = \mathbf{B}_i \Phi, \tag{28}$$

where the matrix $\mathbf{B}_i$ is uniformly bounded by construction. In the next section, we show the scheme preserves the monotonicity.

4.3.2. Analysis

Consider an orthonormal basis $(\mathbf{t}_l)_l$. We construct the limited scheme by

$$\Phi_i^* = \mathbf{B}_i \Phi,$$

that is

$$\Phi_i^* = \sum_\sigma \mathbf{B}_i \Phi(\varphi_\sigma(x)\mathbf{r}_\sigma).$$

The matrix $\mathbf{B}_i$ is constructed by

$$\langle \Phi_i^*, \mathbf{t}_l \rangle = \beta_i^l \langle \Phi, \mathbf{t}_l \rangle$$

with $\beta_i^l \in [0,1]$.

We set $\lambda = \Delta t / 3|T|$. We have

$$\|U_i^n - \lambda \Phi_i\| = \sum_l \left(\langle U_i^n, \mathbf{t}_l \rangle - \lambda \beta_i^l \langle \Phi, \mathbf{t}_l \rangle \right)^2. \tag{29}$$

In Appendix B, we show that if $\beta_i^l \in [0,1]$, relation (29) implies, for simple waves, the following inequality:

$$\|U_i^n - \lambda \Phi_i\| \leqslant \max_{M_j \in T} \|\varphi_\sigma(M_j)\|. \tag{30}$$

4.4. Example of the Cauchy–Riemann equations

We consider the Riemann problem

$$\frac{\partial U}{\partial t} + A\frac{\partial U}{\partial x} + B\frac{\partial U}{\partial y} = 0, \quad t > 0, \tag{31}$$

supplemented by Riemann data per quadrant. The matrices A and B are

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

We test the scheme of (26)–(28) on the following Riemann data $U = (u, v)$:

$$u = \begin{cases} 1 & \text{if } x > 0 \text{ and } y > 0, \\ -1 & \text{if } x < 0 \text{ and } y > 0, \\ -1 & \text{if } x > 0 \text{ and } y < 0, \\ 1 & \text{if } x < 0 \text{ and } y < 0, \end{cases} \quad \text{and} \quad v = \begin{cases} 1 & \text{if } x > 0 \text{ and } y > 0, \\ -1 & \text{if } x < 0 \text{ and } y > 0, \\ -1 & \text{if } x > 0 \text{ and } y < 0, \\ 2 & \text{if } x < 0 \text{ and } y < 0. \end{cases} \tag{32}$$

The solution is self similar, $U(x,y,t) = \widetilde{U}(x/t, y/t)$. The function $\widetilde{U}$ satisfies

$$-\xi\frac{\partial U}{\partial \xi} - v\frac{\partial U}{\partial v} + A\frac{\partial U}{\partial \xi} + B\frac{\partial U}{\partial v} = 0 \tag{33}$$

with the boundary conditions at infinity given by the Riemann data (31), (32) at time $t = 1$. The problem is solved by a time marching technique. The computational domain is $[-2,2] \times [-2,2]$. The solution of (33) corresponds to the solution of the Riemann problem (31), (32) at time $t = 1$. The PDE (33) is elliptic in $\xi^2 + v^2 \leqslant 1$ and hyperbolic in $\xi^2 + v^2 \geqslant 1$: the boundary conditions can be easily computed by solving one dimensional Riemann problems.

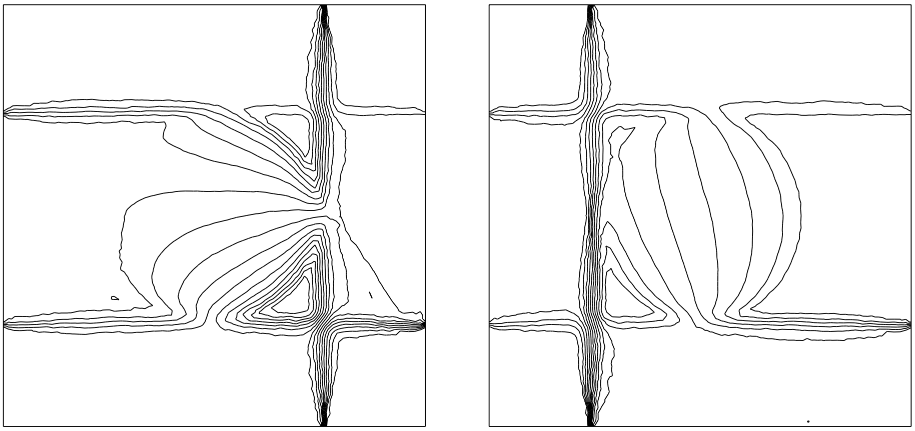


Fig. 8. Solutions of the Riemann problem for the Cauchy–Riemann equations. N scheme, right: u, left: v.

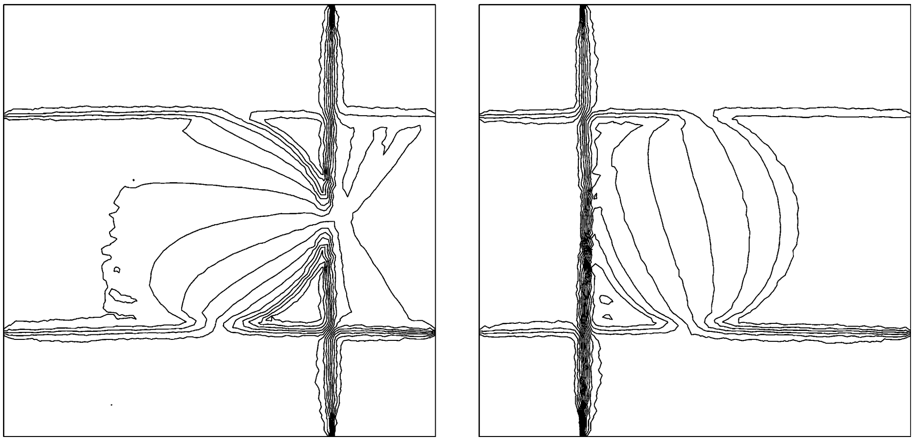


Fig. 9. Solutions of the Riemann problem for the Cauchy–Riemann equations. Limited N scheme, $\theta = 0^\circ$, right: u, left: v.

In the method, the choice of the orthogonal basis is free. In $\mathbb{R}^2$, they can be indexed by the angle θ. We have chosen the eigenvectors of $\cos\theta A + \sin\theta B$. The results are presented for the limited N scheme that reduces to the PSI one for scalar problems. Of course, they will depend on the choice of θ: two different angles give two different schemes. What we want to check numerically is that, first, the non-oscillatory behavior of the results is independent of θ, and second, that their global quality is the same whatever θ.

From Figs. 8–11, we see that the limited solutions are much more accurate than the first-order scheme. The quality of the solution does not depend on the angle, as conjectured, even if the angle is solution dependent as on Fig. 11.

5. Case of the Euler equations

First, as in [9], the nonlinear problem is replaced by a linearized one in each triangle. It is well known that the state vector and the Euler fluxes are quadratic in

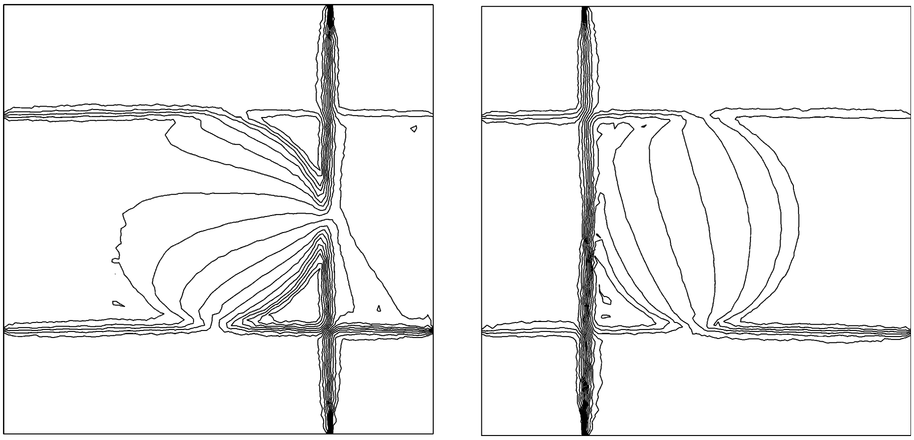


Fig. 10. Solutions of the Riemann problem for the Cauchy–Riemann equations. Limited N scheme, $\theta = 45^\circ$, right: u, left: v.

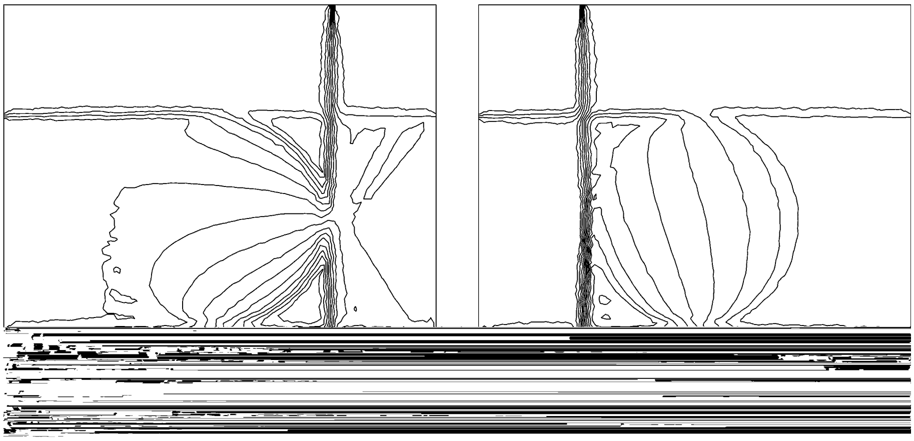


Fig. 11. Solutions of the Riemann problem for the Cauchy–Riemann equations. Limited N scheme $\theta = \arctan(u/v)$, right: u, left: v.

$$Z = \sqrt{\rho}(1, \mathbf{u}, H)^T,$$

where H is the enthalpy of the fluid. For example, we have $W = \frac{1}{2}D(Z)Z$ where $D(Z)$ is linear in Z. For this to be true, one needs that the ratio of specific heats γ be constant. This is what we assume.

Thanks to the linearisation, the problem reduces to

$$\frac{\partial W}{\partial t} + \bar{A}\frac{\partial Z}{\partial x} + \bar{B}\frac{\partial Z}{\partial y} = 0,$$

where the Jacobians $\bar{A}$ and $\bar{B}$ are functions of the average of Z on T, see [9]. Once this is done, we consider the N scheme and the limited scheme as described in Section 4.3. For example the N scheme is

$$|C_i|\frac{W_i^{n+1}-W_i^n}{\Delta t} + \sum_{T,M_i\in T}\Phi_i^T = 0,$$
$$\Phi_i^T = \sum_{M_j\in T} K_i^+ N K_j^-\left(\widetilde{W}_i^n - \widetilde{W}_j^n\right).$$

Here, $\widetilde{W} = \frac{1}{2}D(\bar{Z})Z$, and K_i is evaluated at the state $\bar{W} = \frac{1}{2}D(Z)\bar{Z}$.

In the simulations, we consider two types of boundary conditions: wall and inflow/outflow conditions. They are approximated as in [2]. More precisely, the inflow/outflow conditions are obtained using Steger–Warming flux modified as in [12], and written in fluctuation form. The wall condition is imposed weakly. Here, as in [12], we have chosen to impose the flux

$$\mathscr{F} = \begin{pmatrix} 0 \\ pn_x \\ pn_y \\ 0 \end{pmatrix}$$

written in fluctuation form. Other solutions are possible, such as the ones described in Paillère's thesis [21].

The limited N scheme has been tested against numerous test cases in subsonic, transonic and supersonic situations. We only present the most significant examples. The NACA 0012 is very classical and well documented. The sphere problem is fully subsonic, so one can check the amount of numerical dissipation. The bow shock problem enables to check the robustness of the scheme. Lastly the scramjet problem enables to check the behavior of the schemes on complex waves, and their interactions.

In each problem, the initial condition is a uniform flow given by the conditions at infinity. The scheme is implicit, the implicit phase is provided by an approximate linearisation of the first-order Roe solver. Our aim is not to have a maximum efficiency, but to show the accuracy of the new scheme, as well as the robustness of the method.

5.1. Flow around a NACA 0012, $\mathscr{M}_\infty = 0.85$, $\theta = 1^\circ$

The mesh is plotted in Fig. 12. We plot the velocity for the N scheme and the limited N scheme on Fig. 13. This shows the improvement of the slip line at the trailing edge of the airfoil. We also present the pressure isolines in Fig. 14. The pressure coefficient along the airfoil is provided in Fig. 15. Fig. 16 presents the entropy deviation; there is a clear improvement. In Fig. 15, we compare the pressure coefficient obtained by the scheme of [2] and the present one. We see that the new scheme provides oscillation free solutions, whereas the blending of the N and LDA scheme [3] of [2] was providing slight oscillations. The comparison of the entropy deviation between the new scheme and the blended scheme of [2] also shows a clear improvement.

5.2. Hypersonic flow, $\mathscr{M}_\infty = 8$

It was not possible to run this test case with the scheme of [2], because very quickly negative pressure problems were occurring. A zoom of the mesh is given in Fig. 17. The Mach number for the N scheme and the limited N scheme are given in Fig. 18. We also give cross-sections of the density (Fig. 20), of the Mach number (Fig. 19) and the entropy deviation (Fig. 21) along the symmetry axis. Our results are clearly oscillation free. The boundary conditions are better taken into account with the limited N scheme.

[3] The LDA scheme is defined by $\Phi_i = -NK_i^+\Phi$.

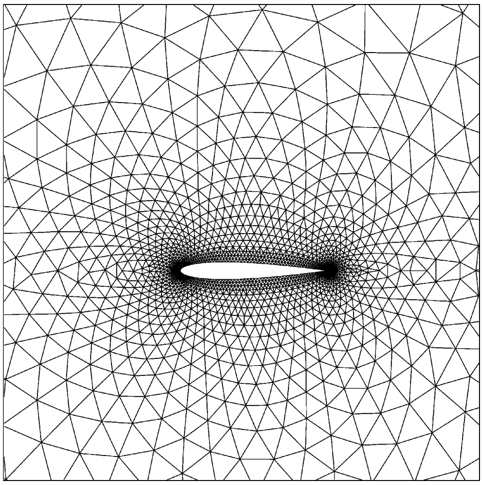


Fig. 12. Zoom of the mesh for the NACA 0012 problem.

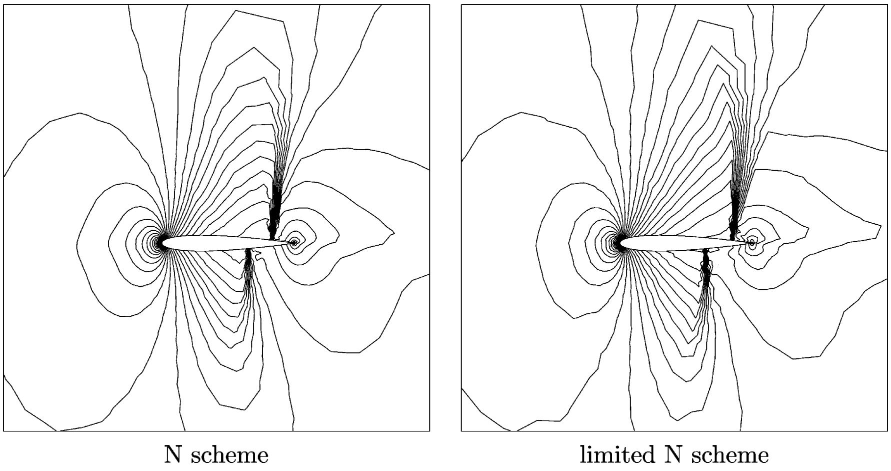


Fig. 13. Velocity isolines.

5.3. Subsonic cylinder, $\mathcal{M}_\infty = 0.35$

We have run this subsonic test case with the N scheme, the limited N scheme, the LDA scheme and the blended N/LDA scheme of [6]. We display the Mach number in Fig. 22 and the entropy deviation in Fig. 23. In both case, we have plotted the same isolines. The mesh is similar as in Fig. 17.

As expected the best results are obtained with the LDA scheme. This is particularly clear from the entropy deviation and Mach number isolines. The Mach isolines are symmetric with respect to the axis orthogonal to the velocity at infinity, as it should be. As expected, the worst results are obtained for the N scheme. The results for the limited N and blended scheme are of similar quality. The entropy deviation is better for the blended LDA/N scheme, but the symmetry of the Mach number isolines is better respected with the limited N scheme.

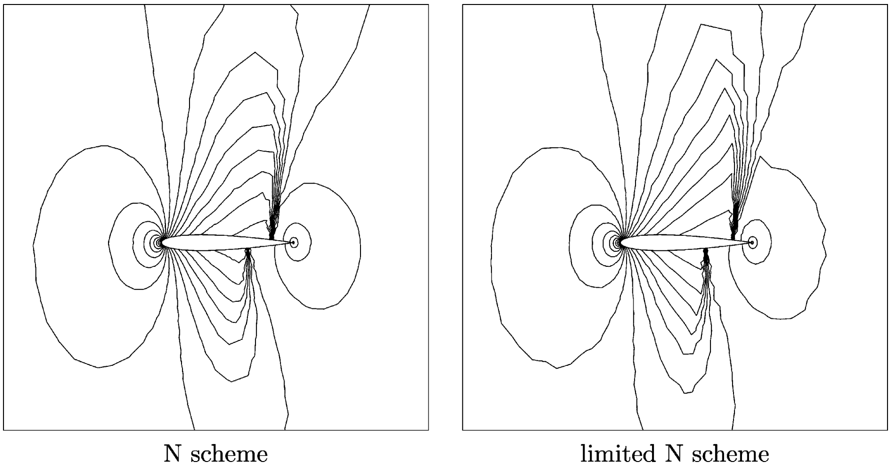


Fig. 14. Pressure isolines.

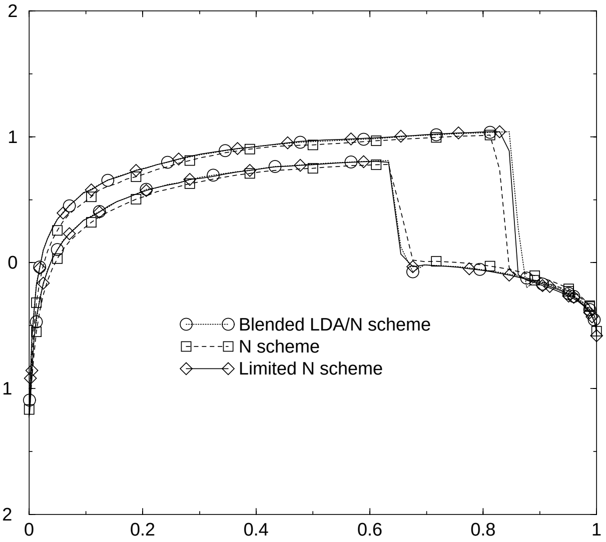


Fig. 15. Plot of cp along the airfoil.

5.4. Scramjet, $\mathcal{M}_\infty = 3.5$

We have run the N scheme, the LDA scheme, the blended scheme of [6], Deconinck et al. B scheme [10] and the limited N scheme on a scramjet-like case. Our implementation of the B scheme is the following. We first compute the N and LDA residuals denoted, respectively, Φ_i^{N} and Φ_i^{LDA} for $i = 1, \ldots, 3$. Then we consider the right and left eigenvectors associated to the flow direction (the result is rather independent of

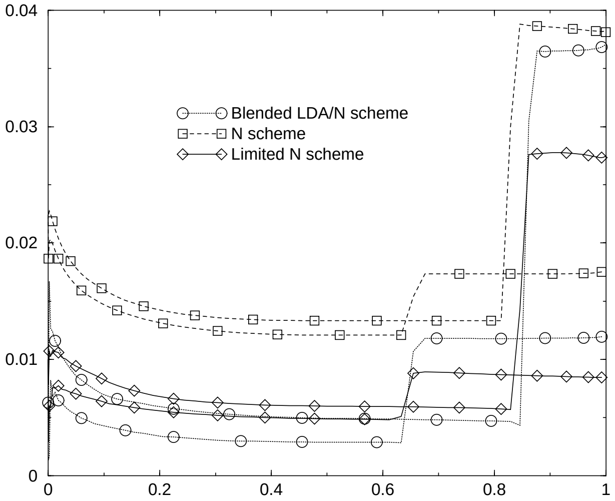


Fig. 16. Entropy deviations along the airfoil.

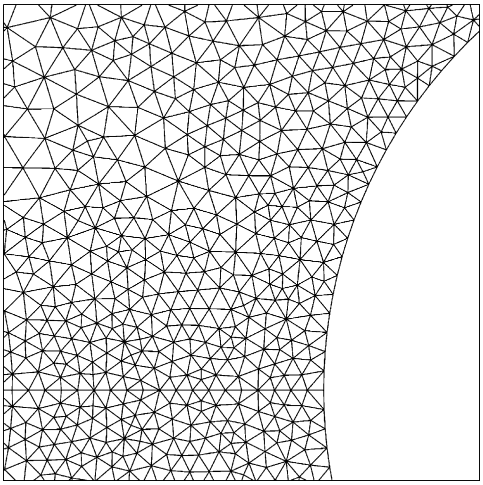


Fig. 17. Mesh for the blunt body problem.

the choice, and qualitatively independent of this choice). We denote them by $\mathbf{r}_l$ and ℓ_l, $l = 1,4$. Then we compute $\langle \ell_l, \Phi_i^{\mathrm{N}} \rangle$ and $\langle \ell_l, \Phi_i^{\mathrm{LDA}} \rangle$. Next we introduce the blending parameters μ_l, $l = 1,\ldots,4$,

$$\mu_l = \frac{|\sum_{i=1}^{3} \langle \ell_l, \Phi_i^N \rangle|}{\sum_{i=1}^{3} |\langle \ell_l, \Phi_i^N \rangle| + \epsilon}, \quad \epsilon = 10^{-10}.$$

Then the B scheme can be written as

$$\Phi_i^B = \sum_{l=1}^{4} \big(\langle \ell_l, \Phi_i^{LDA} \rangle + \mu_l \big(\langle \ell_l, \Phi_i^N - \Phi_i^{LDA} \rangle \big) \big) \mathbf{r}_l.$$

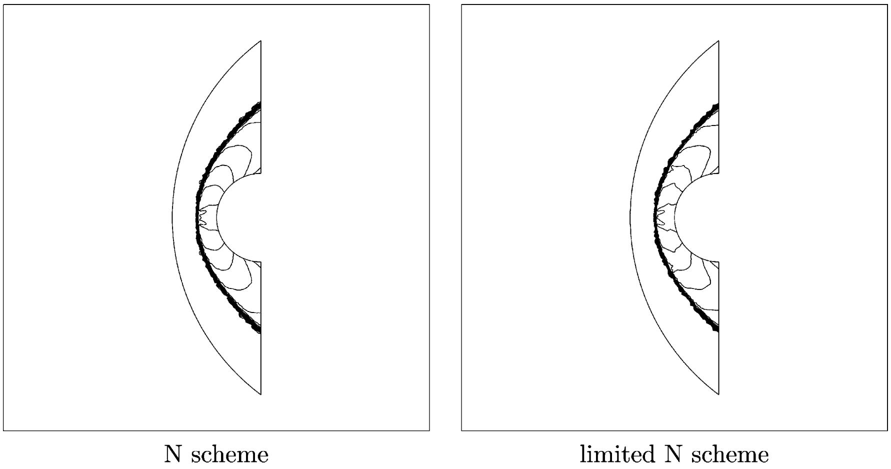


Fig. 18. Iso–Mach lines, limited N scheme.

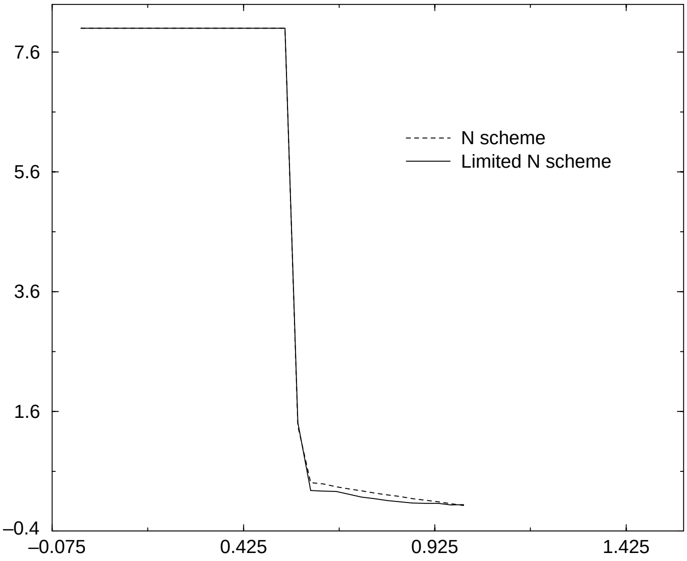


Fig. 19. Cross-section of the Mach number.

The geometry has been found on http://www.inria.fr/, the Mach number at infinity is set to 3.5. A zoom of the mesh is given in Fig. 24.

This case is interesting because it provides a good example of the difference between the schemes. The Mach number is displayed in Fig. 25. We have used the same isolines.

The most dissipative scheme is once more the N scheme. The least dissipative is the LDA scheme, but it is very oscillatory. The three other schemes are monotone. The blended scheme of [6] is the most dissipative among the three (this is clear from the structure of the reflected shocks). The B scheme has a better behavior

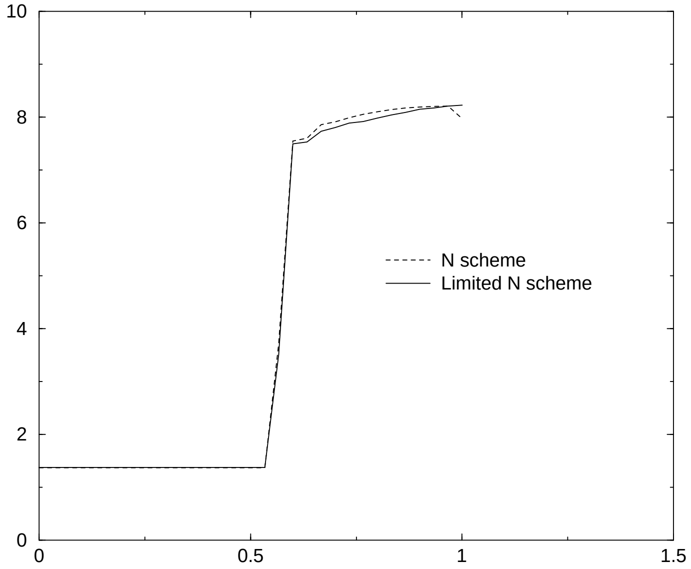


Fig. 20. Cross-section of the density.

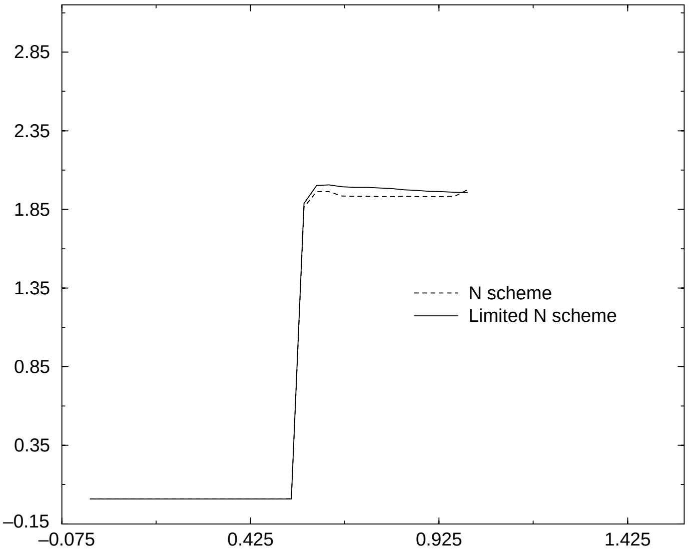


Fig. 21. Cross-section of the entropy deviation.

with respect to this criterion, but the best is the limited N scheme. This is confirmed by the Mach number distribution on the line $y = 0.3$ in the throat, see Fig. 26.

This shows that the choice of the blending parameters is important (the B scheme has a richer structure than the scheme of [6]), but the construction presented in this paper seems the most efficient.

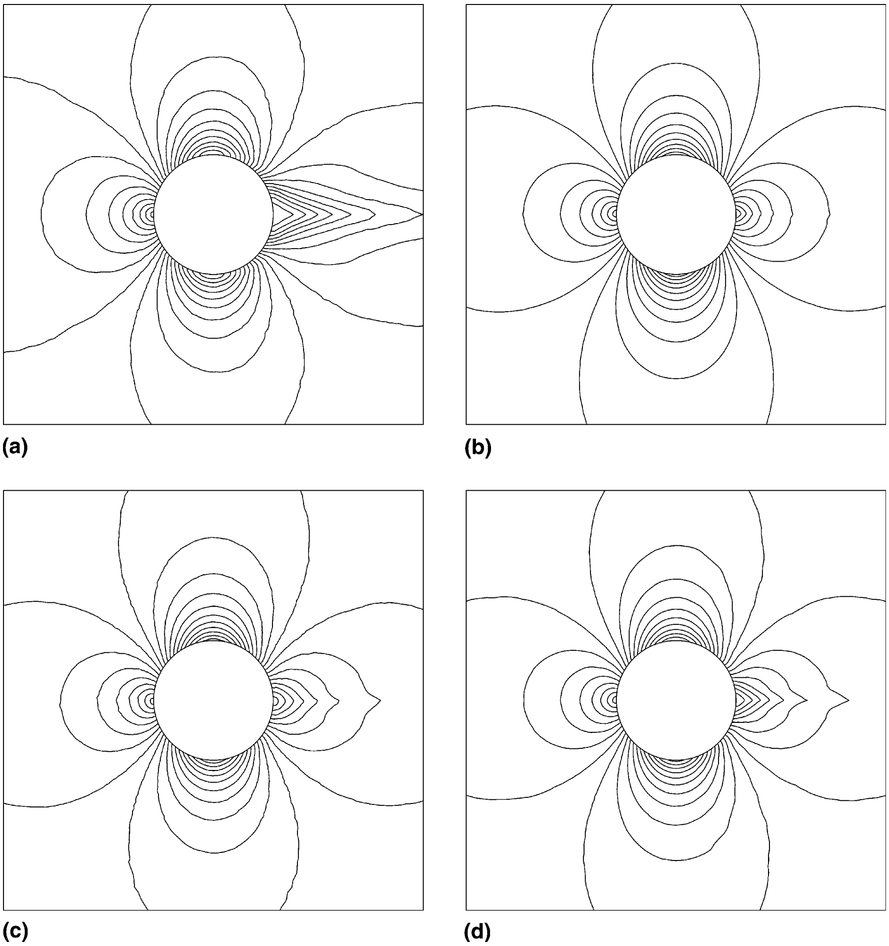


Fig. 22. Mach number isolines: (a) N scheme (min = 0, max = 0.756), (b) LDA scheme (min = 0.001, max = 0.84), (c) limited N scheme (min = 0.002, max = 0.83), (d) blended LDA/N scheme (min = 0.002, max = 0.82).

6. Conclusions

We have presented and analyzed a stable and monotone method for the computation of compressible flows. The schemes we develop are formally second-order accurate on regular unstructured meshes. The capabilities of the schemes are presented on several subsonic, transonic and supersonic flows. The results are good. Compared with any scheme constructed by blending two schemes, as in [2] or [10], the computational complexity is reduced by a half, since only one first-order scheme has to be evaluated.

We have used very crude boundary conditions in this paper, in particular the wall boundary conditions could be improved along the lines of Paillère's thesis, where it becomes easy to incorporate the limitation technique developed in the paper. This is done in [6].

The main problem is in the nonlinear convergence history. The scalar results are converged: the L^2 residual of our results is below 10^{-7}. Going to systems, the situation is much less clear. We have been able to

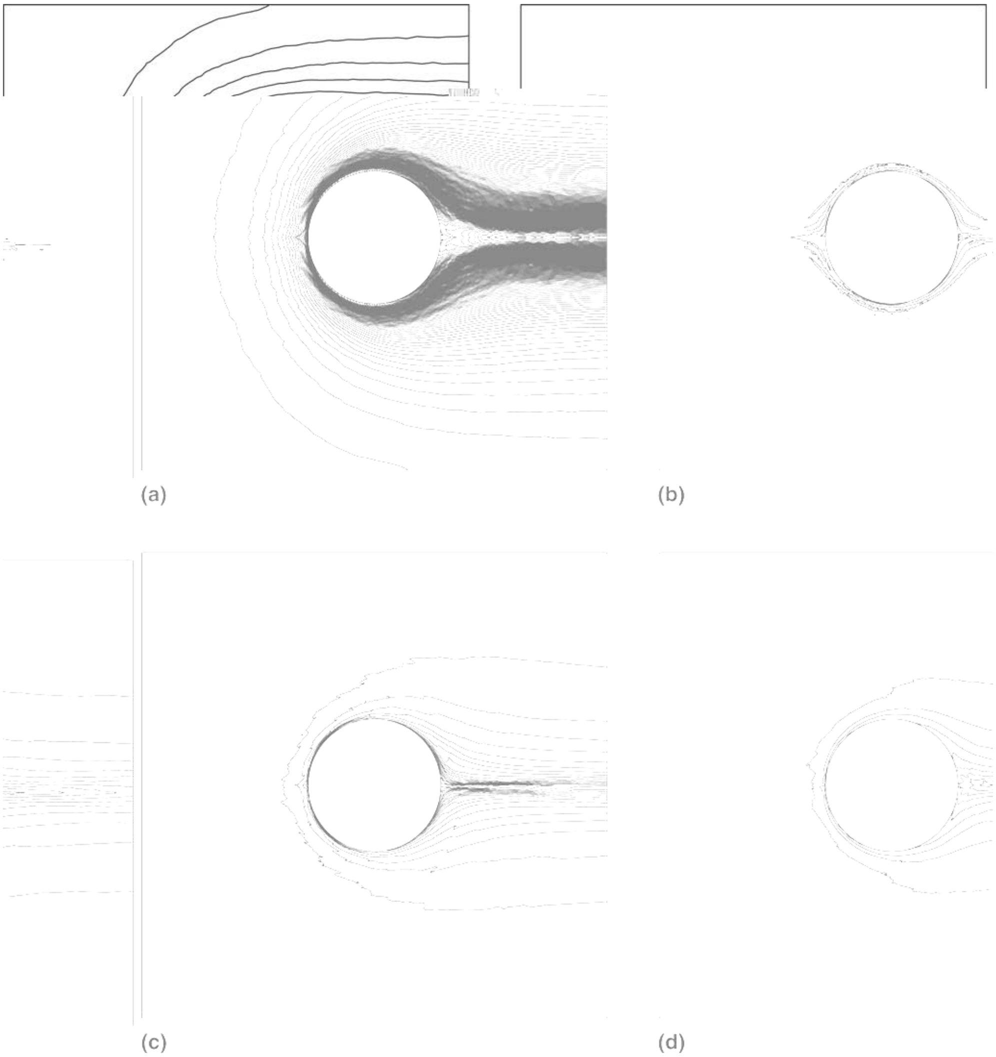


Fig. 23. Entropy deviation isolines: (a) N scheme (min = 0, max = 0.007), (b) LDA scheme (min = −0.0002, max = 0.0002), (c) limited N scheme (min = 0, max = 0.0014), (d) blended LDA/N scheme (min = 0, max = 0.0007).

drop the L^2 and L^∞ residual for the Cauchy–Riemann problem below 10^{-7}, but the convergence history depends very much on the mesh, and on the angle θ in Section 4.4. In particular the convergence history is much smoother when θ is the same throughout the mesh. In some other tests, we were not able to drop the residual below 10^{-3}. In the Euler case, except for very coarse meshes, we were never able to drop the residual below 10^{-3}. However, in each case, most of the mesh points are converged, only very few of them have an erratic convergence history. The reasons of this behavior are not understood and will be investigated elsewhere.

Lastly, an extension of these methods to unsteady flow fields is presented in [6].

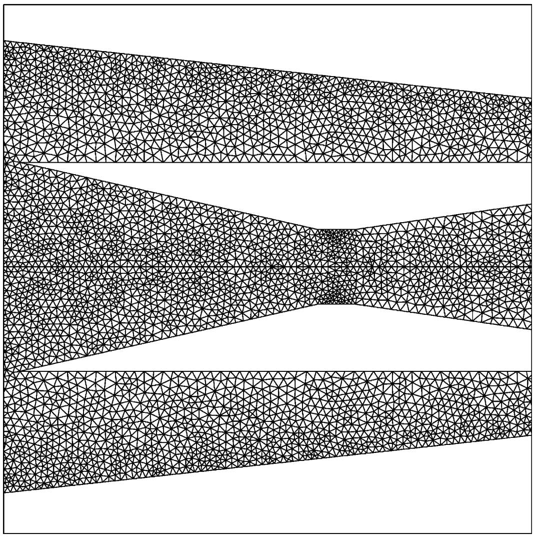


Fig. 24. Zoom of the mesh for the scramjet case.

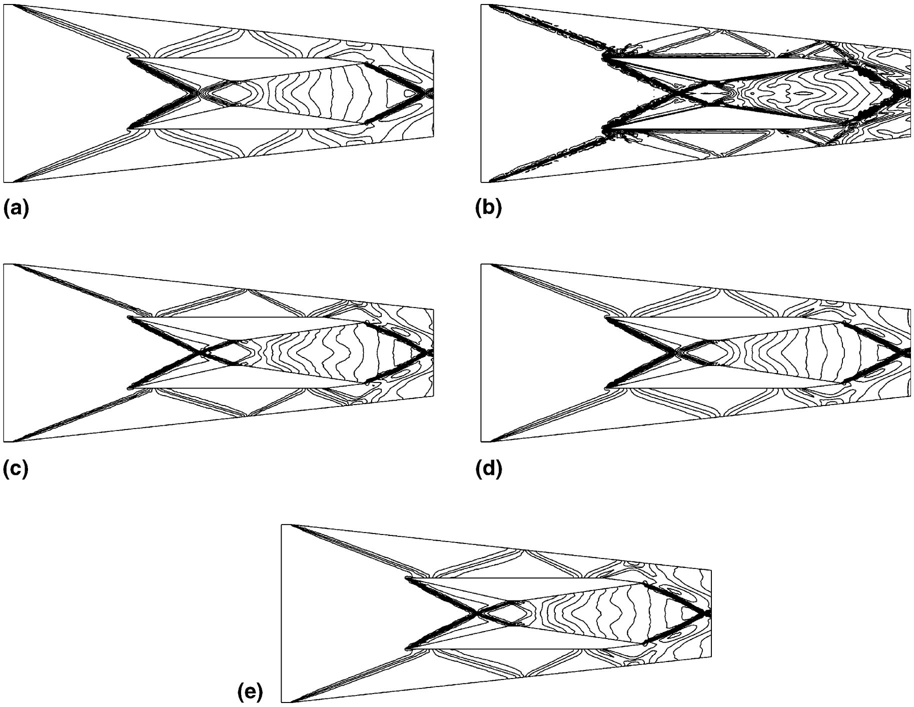


Fig. 25. Mach number isolines: (a) N scheme (min = 1.97, max = 3.6), (b) LDA scheme (min = 1.44, max = 6.47), (c) limited N scheme (min = 1.85, max = 3.6), (d) blended LDA/N scheme (min = 1.87, max = 3.6), (e) B scheme (min = 1.88, max = 3.6).

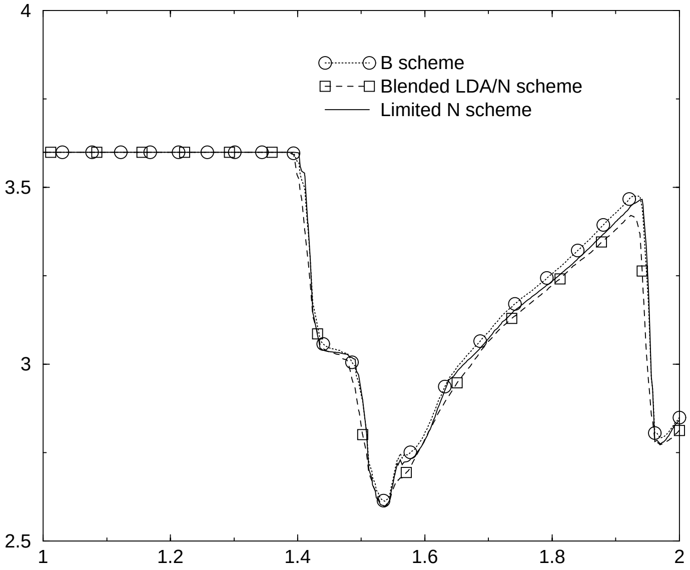


Fig. 26. Mach number distribution along $y = 0.3$ in the throat.

Acknowledgements

This work has been performed under a CEA grant through the "Laboratoire correspondant du Cea" LRC-03. The first author also acknowledge the hospitality of RIACS, NASA Ames Research Center, where part of this work has been performed.

Appendix A. Stability analysis for first-order schemes

A.1. Case of the Rusanov scheme

In this case, we have

$$\Phi_i = \frac{1}{3}\left(\Phi - \alpha\sum_j (U_i - U_j)\right)$$

with U_j proportional to $\mathbf{r}$. Hence, Φ_i is proportional to $\mathbf{r}$. This shows that

$$\Phi_i = \langle \Phi_i, \mathbf{r}\rangle \mathbf{r}$$

with

$$\langle \Phi_i, \mathbf{r}\rangle = \sum_j \frac{1}{3}\langle (K_j - \alpha\mathbf{Id})\mathbf{r}, \mathbf{r}\rangle(\varphi_i - \varphi_j) := \sum_j c_{ij}^R(\varphi_i - \varphi_j)$$

and, by definition of α,

$$c_{ij}^R := \frac{1}{3}\langle (K_j - \alpha \mathbf{Id})\mathbf{r}, \mathbf{r}\rangle \geqslant 0.$$

We consider then the iterative scheme

$$\widetilde{U}_j^{n+1} = U_j^n - \frac{\Delta t}{|C_j|}\sum_{T, M_j \in T} \Phi_j.$$

For a simple wave it reduces to

$$\widetilde{U}_j^{n+1} = \left(\varphi_i - \frac{\Delta t}{|C_i|}\sum_{T, M_i \in T}\sum_{M_j \in T} c_{ij}^R(\varphi_i - \varphi_j)\right)\mathbf{r}. \tag{A.1}$$

This shows that the simple wave evolves proportionally to $\mathbf{r}$, and $||\widetilde{U}_i^{n+1}|| \leqslant \max_{M_j \in T}||U_j^n||$ under a CFL-type condition, and a single time step. Then, thanks to (20),

$$\|U_i^{n+1}\| \leqslant \max_{M_j \text{ neighbor of } M_i}\|U_j^n\|.$$

This does *not* show that if U^n is globally a simple wave, so will be U^{n+1}. But it shows that the profile of the solution remains monotone.

We see that in general, $\widetilde{U}_j^{n+1}$ is a linear combination of terms like (A.1) that are averaged according to (20). This is a strong indication that there is no creation of spurious oscillations.

A.2. Case of the system N scheme

The system N scheme can be written as

$$\Phi_i = \mathbf{K}_i^+ \sum_j \mathbf{N}\mathbf{K}_j^-(U_i - U_j) \tag{A.2}$$

with $\mathbf{N} = (\sum_j \mathbf{K}_j^-)^{-1}$. We start by reducing the problem to the case where $\mathbf{N} = -\mathrm{Id}$ using a (local) change of variable.

We write

$$\begin{aligned}\mathbf{K}_i^+ \sum_j \mathbf{N}\mathbf{K}_j^-(U_i - U_j) &= \mathbf{K}_i^+\mathbf{M}\sum_j \mathbf{K}_j^-(U_j - U_i)\\ &= \mathbf{M}^{1/2}(\mathbf{M}^{-1/2}\mathbf{K}_i^+\mathbf{M}^{1/2})\sum_j(\mathbf{M}^{1/2}\mathbf{K}_j^-\mathbf{M}^{1/2})(\mathbf{M}^{-1/2}U_j - \mathbf{M}^{-1/2}U_i)\end{aligned}$$

with $\mathbf{M} = -\mathbf{N} > 0$.

We make the change of variable $\widetilde{\mathbf{K}}_j = \mathbf{M}^{1/2}K_j\mathbf{M}^{1/2}$, $\widetilde{\mathscr{K}}_j = \mathbf{M}^{-1/2}K_j\mathbf{M}^{1/2}$, and $\mathbf{V} = \mathbf{M}^{-1/2}U$. We get

$$\widetilde{\mathbf{K}}_j^\pm = \mathbf{M}^{1/2}K_j^\pm\mathbf{M}^{1/2} \quad \text{and} \quad \widetilde{\mathscr{K}}_j^\pm = \mathbf{M}^{-1/2}K_j^\pm\mathbf{M}^{1/2},$$

because the matrices are symmetric. Hence, the residual becomes

$$\mathbf{\Phi}_i = \mathbf{M}^{1/2}\sum_j \widetilde{\mathscr{K}}_i^+\widetilde{\mathbf{K}}_j^-(\mathbf{V}_j - \mathbf{V}_i),$$

and the scheme (A.2) becomes

$$\mathbf{V}_i^{n+1} = \mathbf{V}_i^n - \frac{\Delta t}{|T|}\widetilde{\mathscr{K}}_i^+\left(\sum_j \widetilde{\mathbf{K}}_j^-(\mathbf{V}_j - \mathbf{V}_i)\right) = \mathbf{V}_i^n - \frac{\Delta t}{|T|}\widetilde{\mathscr{K}}_i^+\left[\left(\sum_j \widetilde{\mathbf{K}}_j^-\mathbf{V}_j\right) + \mathbf{V}_i\right] \tag{A.3}$$

for which $\mathbf{N} = -\text{Id}$ and the matrices $\widetilde{\mathbf{K}}_i, \widetilde{\mathscr{K}}_i$ are symmetric. This is why, in the following, we assume that $\mathbf{N} = -\text{Id}$. From now on, we write with the form (A.3) of the scheme, even though, strictly speaking, it is not any longer the N scheme. As we have seen before, it is important to write the vector

$$\sum_j \widetilde{\mathbf{K}}_j^-\mathbf{V}_j + \mathbf{V}_i$$

as a sum of simple waves. The matrix $\widetilde{\mathscr{K}}_i$ plays the role of a combination of projectors.

We consider $\mathbf{V}$, the linear interpolation of $\mathbf{V}_j$ on T. We see that

$$\mathbf{V}(x) = \mathbf{V}_0 + \sum_{j=1}^{3} \frac{\langle \vec{n}_j, \mathbf{x}\rangle}{2|T|}\ \mathbf{V}_j,$$

where $\vec{n}_j$ is the inward normal unit opposite to the node j of T and $\mathbf{V}_0$ is a constant vector.

The second step is to rewrite this equality with the right (resp. left) eigenvectors $\mathbf{r}_l^j$ (resp. $\underline{\ell}_l^j$) of $\mathbf{K}_j$ (with $\langle \underline{\ell}_l^k, \mathbf{r}_l^j\rangle = \delta_j^k$),

$$\mathbf{V}(x) = \mathbf{V}_0 + \sum_{j=1}^{3}\sum_{l=1}^{4} \frac{\langle \vec{n}_j, \mathbf{x}\rangle}{2|T|}\langle \underline{\ell}_l^j, \mathbf{V}_j\rangle \mathbf{r}_l^j.$$

The last step is to introduce a point M_j' on the side opposite to the vertex j, see Fig. 7, and to rewrite $\mathbf{V}(x)$ as

$$\mathbf{V}(x) = \mathbf{V}_0' + \sum_{j=1}^{3}\sum_{l=1}^{4} \frac{\langle \vec{n}_j, \overrightarrow{M_j'\mathbf{x}}\rangle}{2|T|}\langle \underline{\ell}_l^j, \mathbf{V}_j\rangle \mathbf{r}_l^j := \mathbf{V}_0 + \sum_{j=1}^{3}\sum_{l=1}^{4}\langle \underline{\ell}_l^j, \mathbf{V}_j\rangle \varphi_l^j(\mathbf{x})\mathbf{r}_l^j, \tag{A.4}$$

i.e., as a sum of simple waves.

We notice, thanks to the definition of the points M_l', $l = 1, \ldots, 3$,

$$\varphi_j^l(M_i)\mathbf{r}_l^j = 0, \quad \varphi_j^l(M_k)\mathbf{r}_l^j = 0, \quad \varphi_j^l(M_j)\mathbf{r}_l^j \neq 0.$$

Since the N scheme is linear, it is enough to evaluate the N scheme on each of the simple waves. Let us consider the wave that vanishes for $j = 2$ and 3. It is proportional to $\mathbf{r}_l^1$:

$$\mathbf{V}_i = \alpha\varphi_l^1(M_i)\mathbf{r}_l^1.$$

We may assume that $\alpha = 1$. We have

$$\mathbf{V}_i^{n+1} = \varphi_l^1(M_i)\mathbf{r}_l^1 - \frac{\Delta t}{|T|}\widetilde{\mathscr{K}}_i^+(\varphi_l^1(M_i)\mathbf{r}_l^1 + \widetilde{\mathbf{V}})$$

with

$$\widetilde{\mathbf{V}} = \varphi_l^1(M_1)\ \mathbf{K}_1^-\mathbf{r}_l^1 = \lambda_l^-\varphi_l^1(M_1)\mathbf{r}_1^j.$$

Hence,

$$\varphi_l^1(M_i)\mathbf{r}_l^1 + \widetilde{\mathbf{V}} = (\varphi_l^1(M_i) + \lambda_l^-\varphi_l^1(M_1))\mathbf{r}_l^j,$$

where λ_l is the eigenvalue of $\mathbf{K}_1$ associated with $\mathbf{r}_l^1$, and then

$$\mathbf{V}_i^{n+1} = \varphi_l^1(M_i)\mathbf{r}_l^1 - \frac{\Delta t}{|T|}\left(\varphi_l^1(M_i) + \lambda_l^- \varphi_l^1(M_1)\right)\widetilde{\mathscr{K}}_i^+\mathbf{r}_l^1.$$

Let us introduce $\boldsymbol{\Upsilon}_i^p$ the left eigenvector of $\widetilde{\mathscr{K}}_i$ and μ_i^p the corresponding eigenvalue. We have

$$\boldsymbol{\Upsilon}_i^p(\mathbf{V}_i^{n+1}) = \left\{\varphi_l^1(M_i) - \frac{\Delta t}{|T|}\left[(\mu_i^p)^+\left(\varphi_l^1(M_i) + \lambda_l^- \varphi_l^1(M_1)\right)\right]\right\}\boldsymbol{\Upsilon}_i^p(\mathbf{r}_l^1),$$

that is

$$\boldsymbol{\Upsilon}_i^p(\mathbf{V}_i^{n+1}) = \left\{\left(1 - \frac{\Delta t}{|T|}(\mu_i^p)^+\right)\varphi_l^1(M_i) - \frac{\Delta t}{|T|}\ (\mu_i^p)^+\lambda_l^- \varphi_l^1(M_1)\right\}\boldsymbol{\Upsilon}_i^p(\mathbf{r}_l^1).$$

Since $\lambda_l^- \leqslant 0$ and $(\mu_i^p)^+ \geqslant 0$, *this shows the stability of the scheme under a CFL-like condition,*

$$\forall i \text{ and } p, \quad \frac{\Delta t}{|T|}(\mu_i^p)^+ \leqslant 1. \tag{A.5}$$

Appendix B. Stability analysis for the limited scheme

Using the notations of Section 4.3.2, we show here that if $\beta_i^l \in [0,1]$, we have for simple waves,

$$\|U_i^n - \lambda\Phi_i\| \leqslant \max_{M_j\in T}\|\varphi_\sigma(M_j)\|.$$

For a simple wave, we have

$$(\langle \widetilde{U}_i^{n+1}, \mathbf{t}_l\rangle)^2 = (\varphi_\sigma\langle r_\sigma, \mathbf{t}_l\rangle - \beta_i^l\lambda\langle\Phi_i, \mathbf{t}_l\rangle)^2,$$

which is convex in β_i^l, so

$$(\varphi_\sigma\langle r_\sigma, \mathbf{t}_l\rangle - \beta_i^l\lambda\langle\Phi_i, \mathbf{t}_l\rangle)^2 \leqslant \max((\varphi_\sigma\langle r_\sigma, \mathbf{t}_l\rangle)^2, (\varphi_\sigma\langle r_\sigma, \mathbf{t}_l\rangle - \lambda\langle\Phi_i, \mathbf{t}_l\rangle)^2).$$

Hence, we can rewrite

$$\sum_l(\langle \widetilde{U}_i^{n+1}, \mathbf{t}_l\rangle)^2 = \|\mathbf{P}U_i^n\|^2 + \|\mathbf{Q}(U_i^n - \lambda\Phi_i)\|^2,$$

where $\mathbf{P}$ (resp. $\mathbf{Q}$) is the orthogonal projector onto the subspace generated by the vectors of $\{\mathbf{t}_l\}$ for which

$$\max((\varphi_\sigma\langle r_\sigma, \mathbf{t}_l\rangle)^2, (\varphi_\sigma\langle r_\sigma, \mathbf{t}_l\rangle - \lambda\langle\Phi_i, \mathbf{t}_l\rangle)^2) = (\varphi_\sigma\langle r_\sigma, \mathbf{t}_l\rangle)^2$$

(resp. $\max((\varphi_\sigma\langle r_\sigma, \mathbf{t}_l\rangle)^2, (\varphi_\sigma\langle r_\sigma, \mathbf{t}_l\rangle - \lambda\langle\Phi_i, \mathbf{t}_l\rangle)^2) = (\varphi_\sigma\langle r_\sigma, \mathbf{t}_l\rangle - \lambda\langle\Phi_i, \mathbf{t}_l\rangle)^2$).

We consider the examples of the Rusanov scheme and the system N scheme for which we have shown that (30) is true.

B.1. Case of the Rusanov scheme

For a simple wave, we have shown that$\widetilde{U}_i^{n+1}$ is proportional to $\mathbf{r}$, and

$$\|\widetilde{U}_i^{n+1}\| \leqslant \max_{M_j\in T}\|\varphi_\sigma(M_j)\|$$

so $\mathbf{Q}(U_i^n - \lambda\Phi_i)$ is proportional to $\mathbf{Q}(\mathbf{r})$ and we have

$$\begin{aligned}\|\mathbf{P}U_i^n\|^2 + \|\mathbf{Q}(U_i^n - \lambda\Phi_i)\|^2 &\leqslant |\varphi_\sigma(M_i)|^2\|\mathbf{Pr}\|^2 + \max_{M_j\in T}\|\varphi_\sigma(M_j)\|^2\|\mathbf{Qr}\|^2 \\ &\leqslant \max_{M_j\in T}\|\varphi_\sigma(M_j)\|^2(\|\mathbf{Pr}\|^2 + \|\mathbf{Qr}\|^2) = \max_{M_j\in T}\|\varphi_\sigma(M_j)\|^2,\end{aligned}$$

because $\mathbf{P}$ and $\mathbf{Q}$ are two orthogonal projectors. This ends the proof in the case of the Rusanov scheme.

B.2. Case of the system N scheme

For the sake of simplicity, we only consider the 2 × 2 case of the Cauchy–Riemann system. The proof is similar in the general case using the arguments of Appendix A.2.

The Cauchy–Riemann system reads

$$\frac{\partial U}{\partial t} + \begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}\frac{\partial U}{\partial x} + \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\frac{\partial U}{\partial y} = 0.$$

Considering a direction $\vec{n}_i$, the eigenvalues of $K_i = K_{\vec{n}_i}$ are $\pm\|\vec{n}_i\|$, and the normalized orthogonal eigenvectors are denoted by $\mathbf{r}_i^\pm$. An easy calculation shows that $\|K_i\| = \|\vec{n}_i\|\mathbf{Id}$, hence

$$\sum_{i=1}^{3} K_i^- = -\frac{2}{\sum_{j=1}^{3}\|\vec{n}_i\|}\mathbf{Id} := \alpha\mathbf{Id}.$$

Using this remark, we see that for a general simple wave $U_i = \varphi_\sigma(M_i)\mathbf{r}$, we have

$$\widetilde{U} = N\left(\sum_{j=1}^{3} K_j^- U_j\right) = \alpha\left(\sum_{j=1}^{3}\|\vec{n}_j\|\varphi_\sigma(M_j)\langle\mathbf{r},\mathbf{r}_i^-\rangle^2\right)\mathbf{r},$$

and then, with $c_{ij} = \alpha\|\vec{n}_j\|\|\vec{n}_i\|\langle\mathbf{r},\mathbf{r}_i^-\rangle^2 \geqslant 0$,

$$\widetilde{U}_i^{n+1} = U_i^n - \lambda\sum_{j=1}^{3} c_{ij}\big(\varphi_\sigma(M_i) - \varphi_\sigma(M_j)\big)\langle\mathbf{r}_i^+,\mathbf{r}\rangle\mathbf{r}_i^+$$

because (setting $\varphi_\sigma(M_j) \equiv 1$), $\alpha(\sum_{j=1}^{3}\|\vec{n}_j\|\langle\mathbf{r},\mathbf{r}_i^-\rangle^2) = 1$.

Another way of stating this result is

$$\begin{aligned}\langle\widetilde{U}_i^{n+1},\mathbf{r}_i^+\rangle &= \left\{\varphi_\sigma(M_i) - \lambda\left(\sum_{j=1}^{3} c_{ij}\big(\varphi_\sigma(M_i) - \varphi_\sigma(M_j)\big)\right)\right\}\langle\mathbf{r},\mathbf{r}_i^+\rangle := \mathscr{A}\ \langle\mathbf{r},\mathbf{r}_i^+\rangle \\ \langle\widetilde{U}_i^{n+1},\mathbf{r}_i^-\rangle &= \varphi_\sigma(M_i)\langle\mathbf{r},\mathbf{r}_i^-\rangle.\end{aligned}$$

We have

$$\begin{aligned}\|\mathbf{Q}(\widetilde{U}_i^{n+1})\|^2 &= \|\mathbf{Q}(\mathscr{A}\langle\mathbf{r},\mathbf{r}_i^+\rangle\mathbf{r}_i^+)\|^2 + \|\mathbf{Q}(\varphi_\sigma(M_i)\langle\mathbf{r},\mathbf{r}_i^-\rangle\mathbf{r}_i^-)\|^2 \\ &\leqslant \max_{M_j\in T}\|\varphi_\sigma(M_j)\|^2\{\|\mathbf{Q}(\langle\mathbf{r},\mathbf{r}_i^+\rangle\mathbf{r}_i^+)\|^2 + \|\mathbf{Q}([\langle\mathbf{r},\mathbf{r}_i^-\rangle]\mathbf{r}_i^-)\|^2\} \\ &= \max_{M_j\in T}\|\varphi_\sigma(M_j)\|^2\|\mathbf{Q}(\mathbf{r})\|^2\end{aligned}$$

because **Q** is an orthogonal projector. Then, since **P** and **Q** are orthogonal projectors, we have

[25] E. van der Weide, H. Deconinck, Positive matrix distribution schemes for hyperbolic systems, in: Désidéri, C. Hirch, P.Le Tallec, M. Pandolfi, J. Périaux (Eds.), Computational fluid dynamics '96, Wiley, New York, 1996, pp. 747–753.

[26] B. van Leer, Progress in multidimensional upwinding, Tech. Rep. 92-43, ICASE, 1992.

[27] P. Vankeirsbilk, H. Deconinck, Solution of the compressible euler equations with high order eno-schemes on general unstructured meshes, in: C. Hirsch, P.Le Périaux, W. Kordulla (Eds.), Computational fluid dynamics '92, Elsevier, Amsterdam, 1992, pp. 843–850.