Toward the ultimate conservative scheme : following the Quest

R. Abgrall
Mathématiques Appliquées de Bordeaux
Université Bordeaux I, 351 Cours de la Libération,
33 405 Talence Cedex, France
abgrall@math.u-bordeaux.fr

Keywords : Compressible flow solvers, Residual schemes, fluctuation splitting schemes, unstructured meshes, multidimensional up-winding.

AMS Subject Classification : 65C99, 65M60, 76N10

Proposed running head : A second order fluctuation splitting scheme for the Euler equations.

Rémi Abgrall
Mathématiques Appliquées de Bordeaux
351 Cours de la Libération
33 405 Talence Cedex
France
email : abgrall@math.u-bordeaux.fr
fax : +33 5 56 84 26 26

Send proof to Rémi Abgrall

Abstract

The aim of this paper is to develop a class of numerical schemes that work on triangular finite element type meshes, and devoted to the computation of steady transonic flows. The schemes are extensions of the PSI (positive stream-wise invariant) scheme of Struijs, and are built directly on the system of the Euler equation for Fluid Mechanics. They are a blending between a first order scheme and a second order scheme, the blending is realized from entropy considerations. It is formally second order accurate at steady state. Several numerical examples are shown to demonstrated the stability and the accuracy of these schemes.

1 Introduction

We are interested in the numerical approximation of the Euler equations of Fluid Mechanics in a domain Ω with boundary conditions,

$$\begin{array}{ll} \dfrac{\partial W}{\partial t} + \mathrm{div}\mathcal{F}(W) = 0 & t > 0 \text{ and } x \in \Omega \\ W(x,0) = W_0(x) & x \in \Omega \\ \text{Boundary conditions} & \text{on } \partial\Omega \end{array} \tag{1}$$

The flux $\mathcal{F} = (F, G)$ and the conserved variables are given by

$$W = (\rho, \rho u, \rho v, E)^T, \quad F(W) = (\rho u, \rho u^2 + p, \rho uv, u(E+p))^T$$
$$\text{and } G(W) = (\rho v, \rho uv, \rho v^2 + p, v(E+p))^T$$

where ρ is the density, u and v are the components of the velocity, ϵ the internal energy and $E = \rho\epsilon + \frac{1}{2}\rho(u^2+v^2)$ is the total energy. The system is closed by the equation of state relating the pressure p to the conserved variables,

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

The ratio of specific heats γ is kept constant, $\gamma = 1.4$ in the applications.

The system (1) has to be supplemented by the entropy inequality which translates the second law of thermodynamics,

$$\frac{\partial S}{\partial t} + \frac{\partial (uS)}{\partial x} + \frac{\partial (vS)}{\partial y} \leq 0 \text{ on } \Omega. \tag{2}$$

Here, the mathematical entropy is given by $S = -\rho h(s)$ [9], where s is the physical entropy

$$s = c_v \log\left(\frac{p}{\rho^\gamma}\right) + s_0 \tag{3}$$

and h is any real valued function such that

$$h' > 0 \text{ and } \frac{h''}{h'} < \gamma^{-1}.$$

In the practical examples, we take $h(x) = x$. If the flow is smooth, (3) is equivalent to

$$\frac{\partial s}{\partial t} + u\frac{\partial s}{\partial x} + v\frac{\partial s}{\partial y} = 0 (\geq 0) \tag{4}$$

and E. Tadmor has shown [14] that the solution (if it is bounded) enjoys the following minimum principle

$$s(x,t) \geq \min_{||y-x|| \leq t ||\vec{u}||_\infty} s(y, 0) \tag{5}$$

where $||x||$ is the Euclidean norm of x and $||\vec{u}||_\infty$ is the L^∞ norm of the velocity field.

In this paper, we are mainly interested in computing the steady solution of (1). This task has become a routine in many modern CFD codes. Many currently used schemes use the ideas developed in the 70's–80's for high resolution schemes by van Leer, Roe, Osher, Harten, Yee, Sweby, and many others. The list is enormous, and some of the most significant contributions has been collected in [10]. However, the quality of the solution is still questionable : some apparently simple problems such as computing the lift and drag of an airfoil is still a difficult task. One of the reasons is that the so–called high resolution schemes suffer a much too important entropy production. In fact, they have been devised on scalar 1D problems, then extended to multiD systems but their construction relies on "1D ideas". Another difficult problem is the sensitivity to the mesh : It is still difficult to construct 3D mesh with a very good quality and consequently, the quality of the solution itself maybe questionable in many cases. Hence, it is natural to try to develop methods that are as little as possible sensitive to the regularity of the mesh.

For these reasons, since several years, some researchers have tried to incorporate some ideas contained in the 1D high resolution schemes (up-winding) in a finite element like framework. Some of the major contributions has been done by P.L. Roe, H. Deconinck, D. Sidilkover and their coauthors. These fluctuation splitting schemes have first been developed for a scalar transport equation, then formally extended to the system (see [13, 6] for example) by incorporating as much physics as possible. These schemes share many common features with the SUPG scheme of Hughes or the streamline diffusion methods of Johnson, except for up-winding. These schemes are not constructed by deeply using any particular direction of the mesh. One of their advantage is that, at least for scalar equations, one can construct a fully second order accurate scheme on triangular meshes with a very compact stencil : the scheme uses only the neighboring nodes.

In our opinion, the maturity of these new schemes is still not sufficient : they may lack robustness, the formulation is not simple enough, etc. The aim of the present paper is to give some elements that might clarify the construction of second order upwind residual schemes.

We first give some generalities on fluctuation splitting schemes. In particular, we connect them to finite volume schemes and show why they offer more flexibility. We recall Roe–Struijs–Deconinck linearisation [5], give a simple condition that guarantees a Lax Wendroff like theorem and describe the design principle of our scheme. Then, we recall two important examples of the system N (Narrow) and the LDA (Low Diffusion Advection) system schemes introduced by van der Weide and Deconinck [15] after their scalar version. We show they are well defined for a symetrizable system. T. J. Barth [4, 2] has shown that for a linear symetrizable system the Nscheme is globally and locally dissipative. In a next section, we give a different interpretation of the PSI scheme, and we show how to extend it to (1). Last, numerical examples are given to illustrate the scheme.

2 The fluctuation splitting schemes

2.1 Generalities

Throughout the paper, we consider a two dimensional computational domain Ω that is triangulated. For the moment, we forget the boundary conditions. The set of triangles is $\{T_j\}_{j=1,\cdots,nt}$. The mesh points are $\{M_i\}_{i=1,\cdots,n_s}$. The vertices of a triangle T are M_{i_1}, M_{i_2}, M_{i_3}. When there is no ambiguity, they are denoted by their index in the list $\{M_i\}_{i=1,\cdots,n_s}$, namely i_1, i_2, i_3 or simply by $1, 2, 3$. To discretise (1), we consider the following residual scheme

$$|C_i|\frac{W_i^{n+1}-W_i^n}{\Delta t}+\sum_{T, M_i\in T}\Phi_i^T=0. \tag{6}$$

In this equation, W_i^n is an approximation of $W(M_i,t_n)$, $|C_i|$ is the area of the dual control volume (see Figure 1) and the residual Φ_i^T are function of W_i^n and its neighboring values. The residuals are supposed to fulfill the following condition :

$$\sum_{M_i\in T}\Phi_i^T=\int_T \text{div } \mathcal{F}^h dx \text{ for any triangle } T \tag{7}$$

where $\mathcal{F}^h$ is an approximation of $\mathcal{F}$. In [3], it is shown that under reasonable assumptions on the Φ_i^Ts (continuity, convergence of $\mathcal{F}^h$ towards $\mathcal{F}$, continuity of $\mathcal{F}^h$ on the edges of T) and the classical assumption of the Lax Wendroff theorem [11], the numerical solution converges to a weak solution of (1).

An example of such scheme is given by the following finite volume scheme (see Figure 2 for the notations) : the numerical flux on the edge $[M_i, M_j]$ is

$$\begin{aligned}F(W_i,W_j,\vec{n}_{ij}) &= \frac{1}{2}\left(\mathcal{F}.\vec{n}_{ij}^1(W_j)+\mathcal{F}.\vec{n}_{ij}^1(W_i)-Q(W_i,W_j,\vec{n}_{ij}^1).(W_i-W_j)\right)\\ &+ \frac{1}{2}\left(\mathcal{F}.\vec{n}_{ij}^2(W_j)+\mathcal{F}.\vec{n}_{ij}^2(W_i)-Q(W_i,W_j,\vec{n}_{ij}^2).(W_i-W_j)\right).\end{aligned} \tag{8}$$

In (8), we have used the notation $\mathcal{F}.\vec{n}$ for $n_x\, F(W)+n_y\, G(W)$ where n_x and n_y are the two components of $\vec{n}$. Since the boundary of C_i is closed, the scheme would be the same if we had set

$$\begin{aligned}\Phi_i^{T_1} &= \frac{1}{2}\left(\mathcal{F}.\vec{n}_{ij}^1(W_j)-\mathcal{F}.\vec{n}_{ij}^1(W_i)-Q(W_i,W_j,\vec{n}_{ij}^1).(W_i-W_j)\right)\\ &+ \frac{1}{2}\left(\mathcal{F}.\vec{n}_{ik}^1(W_k)-\mathcal{F}.\vec{n}_{ik}^1(W_i)-Q(W_i,W_k,\vec{n}_{ik}^1).(W_i-W_k)\right)\end{aligned} \tag{9}$$

and the same for T_2. In equation (9), the indices j and k denotes the indices of the two vertices of T_1 different of M_i. In the end, we get

$$\sum_{M_i\in T_1}\Phi_i^{T_1}=\frac{1}{2}\sum_{M_i\in T_1}\mathcal{F}(W_i).\vec{n}_i$$

where n_j is the inward normal of T_1 opposite to the node M_j. Here, the approximation $\mathcal{F}^h$ is the piecewise linear interpolant of the flux.

Other schemes that cast immediately into into this formulation are the finite element like schemes i.e. the SUPG schemes and the streamline diffusion method.

The main advantage of the residual formulation is that we are not anymore constraint by the geometry of the mesh, as for the finite volume schemes. Only relation (7) and the continuity of $\mathcal{F}^h$ through the edges are important, hence much more flexibility is possible.

2.2 The Roe Struijs Deconinck linearisation

The parameter vector is given by

$$Z = (\sqrt{\rho}, \sqrt{\rho}u, \sqrt{\rho}v, \sqrt{\rho}H)^T = (z_1, z_2, z_3, z_4)^T$$

where $H = \frac{E+p}{\rho}$ is the enthalpy. The remark is that $W = W(Z)$ and $\mathcal{F} = \mathcal{F}(Z)$ are quadratic in Z,

$$W(Z) = \frac{1}{2}D(Z)Z \ \ , \ \ \mathcal{F}(Z) = \frac{1}{2}\mathcal{R}(Z).Z \tag{10}$$

where $D(Z)$ and $\mathcal{R}(Z)$ are matrices that depend linearly on Z. The matrix $D(Z)$ is a triangular matrix and is invertible as soon as $z_1 \neq 0$. The matrices D and $\mathcal{R}$ have been chosen to be the Jacobian matrices of W (resp. $\mathcal{F}$) with respect to Z.

If Z is linearly interpolated on T by Z^h, we set

$$\mathcal{F}^h(x,y) = \mathcal{F}(Z^h)$$

and simple calculations shows that

$$\int_T \text{div } \mathcal{F}^h(x,y)dxdy = \int_T \bar{A}\frac{\partial W^h}{\partial x} + \bar{B}\frac{\partial W^h}{\partial y}dxdy$$

where

$$\bar{A} = \frac{\partial F}{\partial W}(W(\bar{Z})), \ \ \bar{B} = \frac{\partial G}{\partial W}(W(\bar{Z})) \quad \text{with } \bar{Z} = \frac{Z_1 + Z_2 + Z_3}{3}. \tag{11}$$

More explicitly, we have

$$(\bar{A}, \bar{B}) = \mathcal{R}(\bar{Z})D^{-1}(\bar{Z}).$$

Since the Jacobian matrices are functions of the velocity and the enthalpy only, the matrices $\bar{A}$ and $\bar{B}$ are functions of

$$\begin{aligned} \overline{u} &= \frac{\sqrt{\rho_1}u_1 + \sqrt{\rho_2}u_2 + \sqrt{\rho_3}u_3}{\sqrt{\rho_1} + \sqrt{\rho_2} + \sqrt{\rho_3}}, \\ \overline{v} &= \frac{\sqrt{\rho_1}v_1 + \sqrt{\rho_2}v_2 + \sqrt{\rho_3}v_3}{\sqrt{\rho_1} + \sqrt{\rho_2} + \sqrt{\rho_3}}, \\ \bar{H} &= \frac{\sqrt{\rho_1}H_1 + \sqrt{\rho_2}H_2 + \sqrt{\rho_3}H_3}{\sqrt{\rho_1} + \sqrt{\rho_2} + \sqrt{\rho_3}}. \end{aligned} \tag{12}$$

In the following we set

$$K_i = \frac{1}{2}\left(\bar{A}n_x^i + \bar{B}n_y^i\right)$$

where n_x^i, n_y^i are the component of the outward unit vector $\vec{n}_i$ of the side of T opposite to M_i.

One of the important properties of the linearisation is that the matrices K_i are easy to compute, and they are diagonalizable, with real eigenvalues.

The eigenvalues of K_i are $\lambda = u_n \pm c, u_n, u_n$. Hence, in order to show that the matrices K_i always have real eigenvalues, because of equation (11), it is enough to show that the averaged speed of sound defined by

$$\frac{\overline{c}^2}{\gamma - 1} = \overline{H} - \frac{1}{2}\left(\overline{u}^2 + \overline{v}^2\right) = \overline{h} + \delta,$$

where $\overline{h}$ stands for the average specific enthalpy with the same weight coefficients as in (12), is a *real* number. The rest δ is quadratic in speed and is given by :

$$\delta\left(\sqrt{\rho_1} + \sqrt{\rho_2} + \sqrt{\rho_3}\right)^2 = (u_1, u_2, u_3)P(u_1, u_2, u_3)^T + (v_1, v_2, v_3)P(v_1, v_2, v_3)^T$$

if $z_1 = \sqrt{\rho_1}$, etc, we get :

$$P = \begin{pmatrix} z_1(z_2 + z_3) & -z_1 z_2 & -z_1 z_3 \\ -z_1 z_2 & z_2(z_1 + z_3) & -z_3 z_1 \\ -z_1 z_3 & -z_3 z_2 & z_3(z_1 + z_2) \end{pmatrix}$$

This matrix is symmetric, and positive because its eigenvalues are :

$$\lambda_1 = 0, \quad \lambda_2 = \nu + \eta, \quad \lambda_3 = \nu - \eta$$

with

$$\nu = z_1 z_2 + z_1 z_3 + z_3 z_2$$

$$\eta^2 = z_1^2 z_2^2 + z_1^2 z_3^2 + z_2^2 z_3^2 - z_1^2 z_2 z_3 - z_1 z_2^2 z_3 - z_3^2 z_2 z_1$$

We can see that $\eta^2 = (z_1 z_2 + z_1 z_3 + z_2 z_3)^2 - 3 z_1 z_2 z_3 (z_1 + z_2 + z_3)$ is always positive. If we left z_2 and z_3 constant, η^2 becomes a second degree polynomial which discriminant $-3 z_3^2 z_2^2 (z_2 - z_3)^2$ is negative. The coefficient η^2 has thus a constant sign, positive. Moreover, $\eta^2 \leq z_1^2 z_2^2 + z_1^2 z_3^2 + z_2^2 z_3^2 \leq (z_1 z_2 + z_1 z_3 + z_2 z_3)^2$, so all the eigenvalues of P are positive. Thus, $\delta \geq 0$, and $\bar{c}^2 \geq 0$. Hence we get

Proposition 2.1. *If the densities ρ_i are positive, the Roe-Struijs-Deconinck's linearisation leads to diagonalizable with real eigenvalues matrices K_i.*

Note that the conservation relation (7) reads

$$\Phi^T = \sum_{M_i \in T} K_i Z_i. \tag{13}$$

where K_i is computed via the Jacobian of the fluxes with respect to Z.

3 Design principles

In this section, we present three design principles (up-winding, the linear preserving property and a monotonicity condition) and give some examples.

3.1 The up-winding property

Following Roe, Deconinck et al., the scheme is upwind if the following condition is true :

$$(\mathcal{U}) \text{ if all the eigenvalues of } K_i \text{ are negative, then } \Phi_i^T = 0.$$

3.2 Second order accuracy at steady state : the LP condition

At steady state, the scheme (6) satisfy

$$\text{for any node } M_i, \sum_{T, M_i \in T} \Phi_i^T = 0.$$

For any smooth function $\varphi \in C^1(\mathbb{R}^4)^4$, we have

$$\sum_{M_i} \varphi_i \cdot \left(\sum_{T, M_i \in T} \Phi_i^T \right) = 0.$$

Setting $\varphi_G^T = \frac{\varphi_1+\varphi_2+\varphi_3}{3}$ the value of the piecewise linear interpolant of φ at the centroid of T, we have, after having used (7)

$$\sum_T \varphi_G^T \int_T \text{div } \mathcal{F}^h dxdy + \sum_T \sum_{M_i \in T} \left(\varphi_i - \varphi_G^T\right) \cdot \Phi_i^T = 0. \tag{14}$$

To get second order of accuracy at steady state, the equivalent equation of the scheme must be of the form

$$\sum_{M_i} \left(\varphi_i - \varphi_G^T\right) \cdot \left(\sum_{T, M_i \in T} \Phi_i^T \right) = \mathcal{O}(h^2) \tag{15}$$

when the arguments of the residuals are replaced by a smooth solution of (1). In (15), h is the maximum diameter of the triangles T.

One way of ensuring this condition is, that for any smooth solution W of (1), to have $\Phi_i^T(W) = \mathcal{O}(h^3)$ because of (14). This is clear from (14) because (a) the number of vertices in a bounded domain is $\mathcal{O}(h^{-2})$ for a regular mesh, (b) Eqn (15) requires that $(\varphi_i - \varphi_G)\Phi_i^T = \mathcal{O}(h^4)$ which is true since $\varphi_i - \varphi_G = \mathcal{O}(h)$.

This condition is obtained for the SUPG and streamline diffusion schemes because the residual writes

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

with β_i^T uniformly bounded independently of the mesh. For a smooth solution of the *steady* version of (1), one has $\Phi^T = \int_T \text{div } \mathcal{F}^h(W)dx = \mathcal{O}(h^3)$. Indeed, we have

$$\begin{aligned}\int_T \text{div } \mathcal{F}^h(W)dx &= \int_T \text{div } \big(\mathcal{F}^h(W) - \mathcal{F}(W)\big)\, dx \\ &= \int_{\partial T} \big(\mathcal{F}^h(W) - \mathcal{F}(W)\big) \cdot \vec{n}dl\end{aligned}$$

Assuming that $\mathcal{F}^h$ is a *second order* approximation of $\mathcal{F}$, and since the length of ∂T is $\mathcal{O}(h)$, we get

$$\Phi^T = \int_T \text{div } \mathcal{F}^h(W)dx = \mathcal{O}(h^3).$$

This proof makes clear two facts :

1. The scheme we construct can be second accurate *only* for steady problems,
2. The approximation $\mathcal{F}^h$ of the flux must be second order accurate. This is true for the Roe–Deconinck–Struijs linearisation.

The condition $\Phi_T = \mathcal{O}(h^3)$ is not clear, and probably untrue, for finite volume schemes, because it necessitate geometrical cancellations that are not true in general (except for very regular meshes with geometrical invariance properties).

3.3 The monotonicity condition

This condition is very clear for a scalar equation, and quite intuitive but difficult to formalize for a system. The idea is to have a condition that avoids the creation of unphysical oscillations. For a scalar equation, this condition is met, up to a CFL like condition, if the residual sent at node M_i has the following structure

$$\Phi_i^T = \sum_{M_j \in T} c_{ij}(u_i - u_j)$$

with $c_{ij} \geq 0$ and uniformly bounded. These schemes are L^∞ stable under a CFL condition.

In the case of a system, this monotonicity condition is meaningless, but one still wants to avoid unphysical oscillations. Some well considered schemes admits unphysical solutions, such as the ENO schemes [1], but the constraints are such that one can reasonably expect that these oscillations are weak and vanish when the mesh size tends to zero ; and this is indeed the case in practice.

When a monotonicity condition exists, the scheme is L^∞ stable. In this paper, we replace a monotonicity condition by a *formal* approximation of the inequality (4). If (4) were numerically true, we would have a discrete version of (5) (under a CFL condition) and since s is concave, the scheme would be L^∞ stable.

4 Examples

We give two examples of upwind schemes : the system N scheme and the system LDA scheme of Deconinck and van der Weide [15].

4.1 The system N scheme

We set

$$\Phi_i^T = K_i^+ \left(\widetilde{W}_i - \widetilde{W}\right) \tag{16}$$

where $K_i = \bar{A}n_x^i + \bar{B}n_y^i$ and $\widetilde{W}_i = D(\bar{Z})Z_i$[1]. In order to recover the conservation relation (13), we must have

$$\left(\sum_{i=1,3} K_i^-\right)\widetilde{W} = \sum_{i=1,3} K_i^- \widetilde{W}_i \tag{17}$$

To define $\widetilde{W}$, one has *a priori* to invert the matrix

$$\sum_{i=1,3} K_i^-,$$

in some cases, this may be impossible. When it is possible, we denote by N the matrix

$$N = \left(\sum_{i=1,3} K_i^-\right)^{-1}. \tag{18}$$

However, we show in appendix B that, for the Euler equations, $K_i^+\widetilde{W}$ is always defined. More precisely, we show that for the scalar product $\langle\ .\ ,\ .\ \rangle$ defined by A_0 the Hessian of the mathematical entropy S evaluated at the same average state that is used to define $\bar{A}$ and $\bar{B}$, the space of state $\mathbb{R}^4$ can be written as

$$\mathbb{R}^4 = \mathbb{R}r_0 \bigoplus H$$

where

$$r_0 = \begin{pmatrix} 1 \\ \bar{u} \\ \bar{v} \\ \frac{1}{2}\left(\bar{u}^2 + \bar{v}^2\right) \end{pmatrix}.$$

The space H is the orthogonal complement of $r_0\mathbb{R}$ for $\langle\ .\ ,\ .\ \rangle$. The projector onto H is given by

$$\pi(W) = W - \frac{\langle W, v_0\rangle}{\langle r_0, v_0\rangle} r_0. \tag{19}$$

The vector v_0 is $\nabla_W s$ evaluated at the average state, it is the (common) left eigenvector of $\bar{A}$ and $\bar{B}$. We also denote by $\pi^\perp$ the projector

$$\pi^\perp(W) = \frac{\langle W, v_0\rangle}{\langle r_0, v_0\rangle} r_0. \tag{20}$$

[1]If K is a diagonalizable matrix with real eigenvalues ($K = L\Lambda R$) $\Lambda = \operatorname{diag}(\lambda)$, then $K^\pm = L\Lambda^\pm R$ where $\Lambda^\pm = \operatorname{diag}(\lambda^\pm)$

It gives the component of W along the entropy wave.The adjoint π^* of π is given by

$$\pi^*(W) = W - \frac{\langle W, r_0\rangle}{\langle r_0, v_0\rangle} v_0. \tag{21}$$

and satisfies $A_0\pi = \pi^* A_0$.

We can give a meaning to the inverse of $\sum_{i=1,3} K_i^-$, denoted by N. The proof is valid for a linearized symmetric system and is given in Appendix B.

The N scheme is linearly dissipative when the system is symetrizable. More precisely, if the linearisation is carried out in the entropy variables V, T. Barth [4, 2] has shown that,

Lemma 4.1. *If the matrices K_i are symmetric, one has*

$$\sum_{M_i\in T} \left\langle V_i, \Phi_i^T\right\rangle = \frac{1}{2}\sum_{M_i\in T} \langle V_i, K_i V_i\rangle + \mathcal{Q}_N(V_1, V_2, V_3) \tag{22}$$

where

$$\begin{aligned} 2\mathcal{Q}_N(V_1,V_2,V_3) &= -\left\langle \Phi^T N, \Phi^T\right\rangle \\ &+ \textstyle\sum_{M_i\in T} \left(\left\langle V_i, K_i^+ V_i\right\rangle - \left\langle K_i^+ V_i, N K_i^+ V_i\right\rangle\right) \\ &+ \textstyle\sum_{M_i\in T} \left(\left\langle V_i, -K_i^- V_i\right\rangle - \left\langle -K_i^+ V_i, N(-K_i^-) V_i\right\rangle\right). \end{aligned} \tag{23}$$

The quadratic form $\mathcal{Q}_N$ is positive : the N scheme is locally dissipative.

He shows that *each* of the three terms in (23) is positive. The proof of Lemma 4.1 is reproduced in Appendix A.

In the case of a linear problem, the sum of $\langle V_i, K_i V_i\rangle$ cancels, and we get a *global* energy stability result for the N-scheme. If we had a linearisation in the *entropy variable* $V = \nabla_W S$, we could interpret $\frac{1}{2}\sum_{M_i\in T} \langle V_i, K_i V_i\rangle$ as

$$\int_{\partial T} \langle V, K_n V\rangle^h \, d\sigma$$

where the "energy" $\langle V, K_n V\rangle$ is piecewise linearly interpolated by $\langle V, K_n V\rangle^h$. Hence, using exactly the same technique as in [3], if the conditions of Lax Wendroff theorem are true, then the limit of the numerical solutions satisfies an entropy inequality, namely

$$\frac{\partial S}{\partial t} + \text{ div } (uS) \leq 0.$$

4.2 The system LDA scheme

The straightforward extension of the scalar LDA scheme is given by

$$\Phi_i^T = -K_i^+ N \Phi^T \tag{24}$$

where the N matrix is given by (18). The conservation relation is obviously satisfied. This scheme is upwind and LP.

When the linearisation is carried out in the entropy variables, and if we set

$$V^+ = -N\left(\sum_{i=1,3} K_i^+ V_i\right) \text{ and } V^- = N\left(\sum_{i=1,3} K_i^- V_i\right),$$

we have

$$\sum_{M_i \in T} \left\langle V_i, \Phi_i^T \right\rangle = \frac{1}{2}\sum_{M_i \in T} \langle V_i, K_i V_i\rangle + \mathcal{Q}_{LDA}(V_1, V_2, V_3) \tag{25}$$

with

$$\begin{array}{rcl} 2\mathcal{Q}_{LDA}(V_1,V_2,V_3) & = & \sum_{i=1}^3 \left\langle V^+ - \widetilde{V}_i, K_i^+(V^+ - \widetilde{V}_i)\right\rangle \\ & + & \sum_{i=1}^3 \left\langle \widetilde{V}_i - V^-, K_i^+(\widetilde{V}_i - V^-)\right\rangle \\ & + & \sum_{M_i\in T}\left(\left\langle V_i, K_i^+ V_i\right\rangle + \left\langle K_i^+ V_i, N K_i^+ V_i\right\rangle\right) \\ & + & \sum_{M_i\in T}\left(\left\langle V_i, -K_i^- V_i\right\rangle + \left\langle -K_i^+ V_i, N(-K_i^-) V_i\right\rangle\right) \end{array} \tag{26}$$

Unfortunately, $\mathcal{Q}_{LDA}$ is not a positive quadratic form.

4.3 Additional properties of the LDA and N schemes

It is also possible to compare $\mathcal{Q}_N$ and $\mathcal{Q}_{LDA}$ for a symetrizable system when the linearisation is done via the entropy variables,

Lemma 4.2. *We have*

$$\mathcal{Q}_{LDA}(V_1, V_2, V_3) \le \mathcal{Q}_N(V_1, V_2, V_3)$$

This result states that the N scheme is more dissipative than the LDA scheme.

We have the following additional property

Lemma 4.3. *1. If Φ_i is the residual for the N scheme or the LDA scheme, we have*

$$\langle V, \pi\Phi_i(V_1, V_2, V_3)\rangle = \langle \pi^* V, \Phi_i(\pi^* V_1, \pi^* V_2, \pi^* V_3)\rangle.$$

2. The results of Lemmas 4.1, 4.2 are valid on H and $\mathbb{R}r_0$.

This result states that we can split the energy contribution of the residual into their contributions along the entropy wave and its orthogonal complement.

5 A LP positive scalar scheme

In this section, we consider a scalar equation

$$\frac{\partial u}{\partial t} + \left\langle \vec{\lambda}, \nabla u \right\rangle = 0. \tag{27}$$

Let us first describe in detail the N and LDA schemes on a triangle T. We set $k_j = \frac{1}{2}\left\langle \vec{\lambda}, \vec{n}_i \right\rangle$ Since $\sum_{j=1}^{3} k_j = 0$, there are two generic cases : either only one of the k_j is strictly positive, or two of them are strictly positive. The first case is called the "one target case", the second one the "two targets case". For the sake of simplicity, let us assume that $k_1 > 0$, and $k_2 > 0$ in the two targets case. We have

- One target case

$$\begin{aligned} \Phi_1^N &= \Phi \\ \Phi_2^N &= 0 \\ \Phi_3^N &= 0 \end{aligned}$$

  Since $-k_2 \geq 0$ and $-k_3 \geq 0$, the scheme is positive if

$$\Delta t \frac{\sum_{k_j \geq 0} k_j}{|C_i|} \leq 1.$$

- Two targets case

$$\begin{aligned} \Phi_1^N &= k_1(u_1 - u_3) \\ \Phi_2^N &= k_2(u_2 - u_3) \\ \Phi_3^N &= 0 \end{aligned}$$

  Since $k_1 \geq 0$ and $k_2 \geq 0$, the scheme is positive if

$$\Delta t \frac{\sum_{k_j \geq 0} k_j}{|C_i|} \leq 1$$

  .

The condition $\Delta t \frac{\sum_{k_j \geq 0} |k_j|}{|C_i|} \leq 1$ is a global positivity condition. Similarly, the LDA scheme reads

- One target case

$$\begin{aligned} \Phi_1^{LDA} &= \Phi = -k_2(u_1 - u_2) - k_3(u_1 - u_3) \\ \Phi_2^{LDA} &= 0 \\ \Phi_3^{LDA} &= 0 \end{aligned}$$

  Since $-k_2 \geq 0$ and $-k_3 \geq 0$, the scheme is positive.

- Two targets case

$$\begin{aligned} \Phi_1^{LDA} &= -\frac{k_1}{k_3}\Phi \\ \Phi_2^{LDA} &= -\frac{k_2}{k_3}\Phi \\ \Phi_3^{LDA} &= 0 \end{aligned}$$

In this section, we consider a scheme that is a blending between the N scheme and the LDA scheme. On any triangle T, the residual writes

$$\Phi_i = l\Phi_i^N + (1-l)\Phi_i^{LDA}.$$

We look for $l \in \mathbb{R}$ such that the scheme is positive and LP. The conservation constraints, as well as the upwind constraints are automatically satisfied. The problem is to compute l. For this, we follow Sidilkover's technique [12],

$$\Phi_i = l\Phi_i^N + (1-l)\Phi_i^{LDA} = \left(l + (1-l)r_i\right)\Phi_i^N. \tag{28}$$

where $r_i = \frac{\Phi_i^{LDA}}{\Phi_i^N}$.

One target case Any value of l works since $\Phi_i^N = \Phi_i^{LDA}$ for any i. We set $l = 1$ in that case.

Two targets case We have the relations (28) for $i = 1, 2$. Since the N scheme is positive (with a CFL constraint), the blended scheme may also be positive if

$$\begin{array}{l} l + (1-l)r_1 = l(1-r_1) + r_1 \geq 0 \\ l + (1-l)r_2 = l(1-r_2) + r_2 \geq 0. \end{array} \tag{29}$$

We can write

$$\Phi_1^{LDA} = \alpha\Phi = \alpha\left(\Phi_1^N + \Phi_2^N\right)$$

$$\Phi_2^{LDA} = \beta\Phi = \beta\left(\Phi_1^N + \Phi_2^N\right)$$

with $\alpha = -\frac{k_1}{k_3} \in [0,1]$ and $\beta = -\frac{k_2}{k_3} \in [0,1]$ and $\alpha + \beta = 1$. Setting $r = -\frac{\Phi_1^N}{\Phi_2^N}$, the inequalities (29) write

$$\begin{array}{l} l(\beta + \alpha r) + \alpha(1-r) \geq 0 \\ l(\alpha + \beta r) + \beta(1-r) \geq 0 \end{array}$$

A solution to this set of inequalities is

$$l = \begin{cases} 1 & \text{if } r \leq 0 \\ \max\left(\dfrac{\alpha(r-1)}{\beta + \alpha r}, \dfrac{\beta(1-r)}{\alpha + \beta r}\right) & \text{else} \end{cases} \tag{30}$$

The formulae (30) can be rewritten as

$$l = \begin{cases} 1 & \text{if } r \leq 0 \\ \dfrac{\beta(1-r)}{\alpha + \beta r} & \text{if } 0 \leq r \leq 1 \\ \dfrac{\alpha(r-1)}{\beta + \alpha r} & \text{if } 1 \leq r \end{cases}$$

so that an explicit calculation of the residuals $\Phi_i = l\Phi_i^N + (1-l)\Phi_i^{LDA}$ gives :

$$\Phi_1 = \begin{cases} \Phi_1^N & \text{if } r \leq 0 \\ \Phi & \text{if } 0 \leq r \leq 1 \\ 0 & \text{if } 1 \leq r \end{cases} \quad , \ \Phi_2 = \begin{cases} \Phi_2^N & \text{if } r \leq 0 \\ 0 & \text{if } 0 \leq r \leq 1 \\ \Phi & \text{if } 1 \leq r \end{cases}$$

which means that in the case $r \geq 0$, all the residuals are sent either to node 1 or 2 : this is nothing else than the PSI scheme.

It is also possible to rewrite the limiter l of (30) in different forms. Take $\xi \in [0,1[$ and define φ_ξ by

$$\varphi_\xi(x) = \begin{cases} \dfrac{r}{r-1} & \text{if } r \leq \xi \\ \dfrac{\xi}{\xi-1} & \text{else.} \end{cases} \tag{31}$$

We also define φ_1 as the limit of φ_ξ when $\xi \to 1$,

$$\varphi_1(x) = \begin{cases} \dfrac{r}{r-1} & \text{if } r < 1 \\ -\infty & \text{else.} \end{cases}$$

Then we can rewrite l of (30) as

$$\begin{aligned} l &= \min\left(1, \max\left(\varphi_\xi(r_1), \varphi_\xi(r_2)\right)\right) \\ &= \min\left(1, \max\left(\varphi_1(r_1), \varphi_1(r_2)\right)\right). \end{aligned} \tag{32}$$

This remark will be useful in the following.

Another choice of limiter, that does not give back the scalar PSI scheme is obtained by applying the formula (32) where φ_ξ is replaced by ψ given by

$$\psi(x) = \frac{|x|}{|x|+1}, \tag{33}$$

namely

$$l = \min\left(1, \max\left(\psi(r_1), \psi(r_2)\right)\right). \tag{34}$$

We note that the limiters (32) (31) or (34) satisfies

$$\lim_{\phi_2^N \to 0} l = 1$$

which ensure the continuity of the limiter function.

6 A LP stable scheme for (1)

6.1 Comments on the system N scheme

All the numerical experiments that have been conducted with the system N scheme indicate it is a very stable, robust and monotone scheme. By saying it is monotone, we mean that whatever strong the discontinuities are, there are no pre- or post- discontinuity oscillations. In particular, it is a numerical fact that the physical entropy s follows the following semi-discrete relation

$$|C_i|\left(\frac{ds}{dt}\right)_i + \sum_{T,M_i\in T}\sum_{M_j\in T} c_{ij}^T(s_i-s_j)\geq 0 \tag{35}$$

for some *positive* numbers c_{ij}.

Since r_0 is a common eigenvector of $\bar{A}$ and $\bar{B}$, we have

$$\left\langle v_i, \pi^\perp\Phi_i^{N,T}\right\rangle = \sum_{M_j\in T} c_{ij}^T\left\langle v_i, \pi^\perp(W_i-W_j)\right\rangle. \tag{36}$$

where the c_{ij}^T are the coefficients for the scalar N scheme where $\lambda=\overline{\vec{u}}$, i.e.

$$k_j = \frac{1}{2}\left\langle \overline{\vec{u}}, \vec{n}_j\right\rangle,$$

$$c_{ij}^T = \frac{k_i^+k_j^-}{\sum_{j=1,3}k_j^-}.$$

Equation (36) states that the system N scheme is *positive* on the (linearized) entropy wave.

Throughout the paper, we assume that a similar relation does exists on the shear and acoustic waves modes ; more precisely, we assume that a relationship of the type

$$\left\langle v_i, \pi\Phi_i^{N,T}\right\rangle \leq \sum_{M_j\in T} c_{ij}^T\left\langle v_i, \pi(W_i-W_j)\right\rangle \tag{37}$$

holds even if we have been unable to prove it. The first inequality (36) states a monotone behavior of the projection of Φ_i^T on the entropy wave. The second inequality (37) states the same for the projection of Φ_i^T on the acoustic and shear modes.

6.2 Construction of an entropy stable LP scheme

In the following analysis, we set

$$v_i = \nabla_W s(W_i)$$

and we consider the semi–discrete scheme

$$|C_i|\left(\frac{dW}{dt}\right)_i + \sum_{T,M_i\in T}\Phi_i^T = 0$$

where

$$\Phi_i^T = \ell\Phi_i^{N,T} + (\mathbf{Id}-\ell)\Phi_i^{LDA,T}. \tag{38}$$

Here ℓ is a matrix which structure has to be defined in order to get a monotone entropy stable scheme.

We first left multiply $\left(\frac{dW}{dt}\right)_i$ by v_i, and we get

$$|C_i| \left(\frac{ds}{dt}\right)_i + \sum_{T, M_i \in T} \left\langle v_i, \Phi_i^T \right\rangle = 0$$

The idea is to construct ℓ in such a way that

$$\frac{\left\langle v_i, \Phi_i^T \right\rangle}{\left\langle v_i, \Phi_i^{N,T} \right\rangle} \geq 0. \tag{39}$$

If it is possible, by combining this inequality to (36-37), provided that $\frac{\left\langle v_i, \Phi_i^T \right\rangle}{\left\langle v_i, \Phi_i^{N,T} \right\rangle}$ is bounded, we recover formally a bound on the solution, under a CFL like condition. Now we show how it is possible to construct such a matrix ℓ.

To begin with, recall the decomposition of the state space, $\mathbb{R}^4 = \mathbb{R}r_0 \bigoplus H$ given, for any state $W \in \mathbb{R}^4$ by

$$W = \pi(W) + \pi^\perp(W), \quad \pi^\perp(W) = l(W) r_0 = \frac{\langle W, v_0 \rangle}{\langle r_0, v_0 \rangle} r_0$$
$$\text{and } y = \pi(W) = W - \frac{\langle W, v_0 \rangle}{\langle r_0, v_0 \rangle} r_0.$$

The vectors r_0 and v_0 are evaluated for an averaged set of Jacobian matrices $\bar{A}$ and $\bar{B}$. From a physical point of view, $l(W)$ is the component of W on the entropy wave r_0, while $\pi(W)$ is the sum of the acoustic and shear waves.

The key remark is to notice that the N scheme and the LDA scheme have a simple expression for this decomposition because r_0 is a common eigenvector of $\bar{A}$ and $\bar{B}$. More precisely, we have

$$\begin{aligned}
\Phi_i^N &= \sum_{j=1; j\neq i}^{3} K_i^+ N K_j^- \pi^\perp \left(\tilde{W}_i - \tilde{W}_j\right) + \sum_{j=1; j\neq i}^{3} K_i^+ N K_j^- \pi \left(\tilde{W}_i - \tilde{W}_j\right) \\
&= \left(\frac{\sum\limits_{j=1; j\neq i}^{3} k_i^+ k_j^- l(W_i - W_j)}{\sum_{j=1,3} k_j^-} \right) r_0 + \sum_{j=1; j\neq i}^{3} K_i^+ N K_j^- \pi \left(\tilde{W}_i - \tilde{W}_j\right) \\
\Phi_i^{LDA} &= -\sum_{j=1; j\neq i}^{3} K_i^+ N K_j \pi^\perp \left(\tilde{W}_i - \tilde{W}_j\right) - \sum_{j=1; j\neq i}^{3} K_i^+ N K_j \pi \left(\tilde{W}_i - \tilde{W}_j\right) \\
&= \left(\frac{\sum\limits_{j=1; j\neq i}^{3} k_i^+ k_j l(W_i - W_j)}{\sum\limits_{j=1; j\neq i}^{3} k_i^+} \right) r_0 - \sum_{j=1; j\neq i}^{3} K_i^+ N K_j \pi \left(\tilde{W}_i - \tilde{W}_j\right)
\end{aligned}$$

For this reason, we set

$$\ell = l_1 \pi^\perp + l_2 \pi. \tag{40}$$

In other words, the matrix ℓ has two components. One acts only on the components on the entropy wave, and the other component only plays on the shear/acoustic waves. Then, thanks to Lemma 4.3, we evaluate the entropy production within a single triangle,

$$\begin{aligned}
\langle v_i, \Phi_i \rangle &= \left\langle v_i, \ell \Phi_i^N \right\rangle + \left\langle v_i, (Id - \ell)\Phi_i^{LDA} \right\rangle \\
&= \left(l_1 \left\langle v_i, \pi^\perp(\Phi_i^N) \right\rangle + (1-l_1) \left\langle v_i, \pi^\perp(\Phi_i^{LDA}) \right\rangle\right) \\
&\quad + \left(l_2 \left\langle v_i, \pi(\Phi_i^N) \right\rangle + (1-l_2) \left\langle v_i, \pi(\Phi_i^{LDA}) \right\rangle\right) \\
&= \left(l_1 + (1-l_1) \frac{\left\langle v_i, \pi^\perp(\Phi_i^{LDA}) \right\rangle}{\left\langle v_i, \pi^\perp(\Phi_i^N) \right\rangle} \right) \left\langle v_i, \pi^\perp(\Phi_i^N) \right\rangle \\
&\quad + \left(l_2 + (1-l_2) \frac{\left\langle v_i, \pi(\Phi_i^{LDA}) \right\rangle}{\left\langle v_i, \pi(\Phi_i^N) \right\rangle} \right) \left\langle v_i, \pi(\Phi_i^N) \right\rangle
\end{aligned}$$

We define l_1 and l_2 by the two conditions

- For $i = 1, 2, 3$,
$$l_1 + (1-l_1) \frac{\left\langle v_i, \pi^\perp(\Phi_i^{LDA}) \right\rangle}{\left\langle v_i, \pi^\perp(\Phi_i^N) \right\rangle} \geq 0,$$
- For $i = 1, 2, 3$,
$$l_2 + (1-l_2) \frac{\left\langle v_i, \pi(\Phi_i^{LDA}) \right\rangle}{\left\langle v_i, \pi(\Phi_i^N) \right\rangle} \geq 0.$$

Following the developments of section 5, we set

$$\begin{aligned}
l_1 &= \min(1, \max\left(\varphi_\xi(r_1'), \varphi_\xi(r_2'), \varphi_\xi(r_3')\right)) \\
l_2 &= \min(1, \max\left(\varphi_\xi(r_1), \varphi_\xi(r_2), \varphi_\xi(r_3)\right))
\end{aligned} \tag{41}$$

where $r_i = \frac{\left\langle v_i, \pi(\Phi_i^{LDA}) \right\rangle}{\left\langle v_i, \pi(\Phi_i^N) \right\rangle}$ and $r_i' = \frac{\left\langle v_i, \pi^\perp(\Phi_i^{LDA}) \right\rangle}{\left\langle v_i, \pi^\perp(\Phi_i^N) \right\rangle}$ for $i = 1, 2, 3$,$\xi \in [0, 1[$ and φ_ξ is defined by (31). Contrarily to the scalar case where the value of ξ, was unimportant, it is not clear whether different values of ξ furnishes the same value of l_2.

An other solution is given by

$$\begin{aligned}
l_1 &= \min(1, \max\left(\psi(r_1'), \psi(r_2'), \psi(r_3')\right)) \\
l_2 &= \min(1, \max\left(\psi(r_1), \psi(r_2), \psi(r_3)\right))
\end{aligned} \tag{42}$$

where ψ is defined by (33).

Remark 6.1. *1. In the continuous case, we have $ds = \langle v, dW \rangle$. Since v belongs to $\mathbb{R}r_0$, v is orthogonal to the acoustic and shear modes of the Euler equations. In the discrete case however, the situation is a bit more complex : there is no reason why $\langle v_i, \pi^\perp \Phi_i^N \rangle$ should be zero. When the flow is smooth, $\langle v_i, \pi^\perp \Phi_i^N \rangle$ is likely to be small, but certainly not when a discontinuity exists.*

2. We have developed another version of the scheme (refereed as V-scheme later in this remark), where we replace v by $V = \nabla_W(\rho s)$. At the continuous level, one has $\left\langle V,\ div\ \vec{F}(W) \right\rangle = \rho \left\langle v,\ div\ \vec{F}(W) \right\rangle + s div\ (\rho \vec{u})$, and then

$$\left\langle V, \pi(\ div\ \vec{F}(W)) \right\rangle = \rho \left\langle v,\ div\ \vec{F}(W) \right\rangle$$

and

$$\left\langle V, \pi^\perp(\ div\ \vec{F}(W)) \right\rangle = s div\ (\rho \vec{u}).$$

This suggests that the same numerical technique, where v_i is replaced by V_i, would be a very similar scheme on $\mathbb{R}r_0$, but would act yield a limitation of the mass flow on H. We have implemented this scheme. Its results are indistinguishable from those obtained by the scheme (v-scheme) developed in this section. However, we have preferred the v-scheme because the interpretation of the V-scheme is not clear. The v-scheme uses an approximation of the transport of the physical entropy where a minimum principle exists, while the V-scheme uses an approximation of a conservation operator,

$$\frac{\partial(\rho s)}{\partial t} + \ div\ (\rho s \vec{u})$$

for which no minimum or maximum principle exists.

7 Boundary conditions

To set the boundary conditions, we utilize the fact that a finite volume scheme is a residual distributive scheme, according to equation (9). The inflow and outflow boundary conditions are prescribed via the modified Steger and Warming [8] flux splitting,

$$\mathcal{F}(W_i, W_\infty, \vec{n}) = A(W_i)^+.\vec{n}W_i + A(W_i)^-.\vec{n}W_\infty,$$

and the wall condition are simply obtained by setting $\langle \vec{u}, \vec{n} \rangle = 0$, i.e.

$$\mathcal{F}_{\text{wall}}(W_i, \vec{n}) = \begin{pmatrix} 0 \\ pn_x \\ pn_y \\ 0 \end{pmatrix}.$$

8 Numerical examples

We have run this scheme on many examples ; each of them are steady computations. We report the most significant ones, and give comparison with the N scheme, the LDA scheme, a first order finite volume scheme (Roe scheme), and a Muscl scheme (Roe scheme with Muscl extrapolation on the primitive variables with van Leer limiter). The finite volume schemes use the dual cell control volume formulation [8]. For a given problem and variable, the same ISO-lines have been used to draw the pictures whatever the scheme. In all the numerical computations, we have taken the matrix ℓ defined by (42–33).

8.1 A shock tube problem

The initial condition consists in two parallel uniform flows which conditions are listed above :

- Top : Mach number $= 2.4$, $\rho = \gamma$, $p = 1$,
- Bottom : Mach number $= 4$, $\rho = \frac{\gamma}{2}$, $p = 0.25$

The conditions on the left boundary are identical to the initial conditions. The flow is everywhere supersonic ; no exit boundary condition is needed. The steady solution consists of a shock wave, a contact discontinuity and a fan. On any line orthogonal to the initial velocity vector, the solution looks like a 1D Riemann problem. On Figure 3, we display the mesh, on Figures 4, 5, 6, the density, the pressure and the Mach number.

From these results, it appears clearly that :

- the resolution of the shock between the MUSCL scheme and the blended scheme is in favor of the new scheme,
- the resolution of the slip line between the MUSCL scheme and the blended scheme is without any doubt in favor of the blended scheme,
- the fan is better represented by the blended scheme because the plateau, where the solution is constant between the discontinuities, starts earlier for the blended scheme than for MUSCL scheme.

The display of the results for the N scheme illustrates the mono tic behavior of this scheme, where as the LDA scheme is clearly not monotone.

8.2 A transonic test case

We take one of the test case of the Game workshop held at India Rocquencourt in 1987, [7]. It is the Naca 0012 case, the Mach number at infinity is $M_\infty = 0.85$, $\alpha = 1^o$. The solution has two shocks one on the top of the airfoil, and another weaker on its bottom. A slip line comes out of the trailing edge. This case is interesting because the more dissipative the scheme, the more symmetric the solution. We present the results on two different meshes. Both are unstructured, one is very irregular 7-(a), the other is much more regular 7-(b). Our purpose is to illustrate the effects of the numerical dissipation and the mesh. Table 1 gives the minimum and maximum of the Mach number, the pressure coefficient $c_p = \frac{p-p_\infty}{\frac{1}{2}\rho_\infty u_\infty^2}$ and the entropy deviation $\Sigma = \frac{s-s_\infty}{s_\infty}$ for the irregular mesh Figures 8, 9 and 10 and the Mach number for the regular one Figure 11. The main facts are

1. Behavior of the entropy : By looking at Table 1 and Figure 8, we see that the entropy has a more physical behavior for the blended scheme than the Muscl scheme. There is no numerical artifact, see Figure 8-a, lower and upper shocks. The slip line seems less diffused. We should have $\Sigma \geq 0$, numerically this is violated by both schemes but by an order of magnitude less by the blended scheme, see Table 1.

2. The shocks are better resolved because the isolines goes into the shock much more for the blended scheme than the Muscl one, see Figure 9 and 10.

3. When comparing the results of Figure 10, one can see a strange behavior of the Mach number contours. This is due to the large entropy layer created by the Muscl scheme and the strong dependency of the scheme on the mesh. Compare to Figure 11 where the same schemes have been used on a much more regular mesh. There is still a larger entropy layer for the Muscl scheme, but its influence on the Mach number is much weaker.

8.3 Engine inlet

This another test case of the Gamm workshop [7]. The condition are set so that the Mach number at infinity is $M_\infty = 2$ and the Mach number at the exit of the engine inlet is $M = 0.27$. We display on Figure 12 the Mach number in the whole computational domain. The solution has a lambda shock at the entrance of the inlet. This lambda shock produces a slip line coming out of the triple point. By comparing the solutions of Figures 12-(c) and 12-(d), it is clear that a slip line is much better resolved by the blended scheme than the Muscl one, even though the mesh is quite coarse, see figure 13.

Next, on Figures 14, 15 and 16, we see that the location of the secondary shock depend on the scheme. The shock is hardly visible for Roe's. By increasing strength, we have the N scheme first, then Muscl and last the blended scheme. Moreover, the stronger, the further from the inlet entrance. This is quite consistent with the conclusion of the Gamm workshop : the less diffusive, the further from the entrance the shock is. We have run this case with the LDA scheme, the results are very oscillatory but comfort this conclusion. In table 2, one can notice that even if $\Sigma \geq 0$ theoretically, this is not the case for the new scheme. However, the deviation form $\Sigma \geq 0$ is more that an order of magnitude smaller from the blended than the Muscl scheme.

Last, the lambda shock reflects on the wall of the engine. The reflection is much clearer for the blended scheme than the muscl one. It can be seen that the shock wave has also more reflections, but this is difficult to notice on Figure 12.

8.4 Flow over a cylinder

This is another Gamm test case. The Mach number at infinity is $M_\infty = 0.38$. The solution should be symmetric with respect to the $x-$ and $y-$ axis. The Mach number is always less than 1, but its maximum is very close to 1. Since the flow is sub-sonic, the entropy deviation should be 0. In fact, the respect of the symmetry properties, and the departure from $\Sigma = 0$ are good criterion to compare the solutions. Once more the mesh is completely unstructured, see Figure 17.

We display the solution given by the Muscl scheme, the blended scheme and the LDA scheme. As expected the LDA scheme gives the best results, see Figures 18-(c) and 19-(c). But the blended scheme is not that far away, see 18-(b) and 19-(b), and compare the max/min values on Table 3. It also gives much better results than the Muscl scheme, as reported on 18-(a) and 19-(a), and Table 3. The entropy production is five times more important than the other schemes. The entropy layer is also much thicker.

8.5 Comments on the iterative convergence

As van der Weide system PSI scheme [15], the iterative convergence of the scheme we present in this paper are very poor, basically the L^2 residual on the density stagnates at about 10^{-2}–10^{-3}, where the residual is compared to the first iteration.

There is a way to improve this behavior by modifying the arguments in equations (32) or (34). To simplify the text, take the example of the function ψ, equation (33). The arguments are essentially

$$\psi(r_i) = \frac{|\left\langle v_i, \Phi_i^{LDA}\right\rangle|}{|\langle v_i, \Phi_i^{LDA}\rangle| + |\langle v_i, \Phi_i^N\rangle|}$$

or

$$\psi(r_i) = \frac{|\left\langle v_i, \pi^\perp\Phi_i^{LDA}\right\rangle|}{|\langle v_i, \pi^\perp\Phi_i^{LDA}\rangle| + |\langle v_i, \pi^\perp\Phi_i^N\rangle|}.$$

To prevent zero divide, we have considered instead

$$\psi_\epsilon(r_i) = \frac{|\left\langle v_i, \Phi_i^{LDA}\right\rangle|}{|\langle v_i, \Phi_i^{LDA}\rangle| + |\langle v_i, \Phi_i^N\rangle| + \epsilon}$$

and

$$\psi_\epsilon(r_i) = \frac{|\left\langle v_i, \pi^\perp\Phi_i^{LDA}\right\rangle|}{|\langle v_i, \pi^\perp\Phi_i^{LDA}\rangle| + |\langle v_i, \pi^\perp\Phi_i^N\rangle| + \epsilon}.$$

The parameter ϵ should be $O(h^3)$ for it to be negligible compared to the residual of the N and LDA schemes. In all the calculations, we have set ϵ in the range $[10^{-5}, 10^{-6}]$. If ϵ is too small, the iterative convergence is erratic. If $\epsilon \to 0$ the scheme resembles the LDA scheme. In our experiments, we have noticed that the results are quite unsensitive to ϵ : they are non-oscillatory.

9 Concluding remarks

In this paper, we have presented the construction of an upwind residual scheme that is formally second order accurate at steady state. It is a blending between the system N scheme and the Low Diffusion Advection schemes formally extended to system by van der Weide and Deconinck. The present construction relies on the analysis of the entropy production of the scheme within a single element. This scheme is robust and much less diffusive that a state of the art Muscl scheme on unstructured mesh. The stencil of the scheme is also more compact, so a parallel implementation is much easier.

The limiter we consider here has only two degrees of freedom. Future work will consist of constructing limiters with more parameter, in the hope to decrease the numerical diffusion. Extensions to unsteady flows will also be considered.

Acknowledgements

During this study, I have been helped by B. Nkonga (Université Bordeaux I) : I owe him (at least) the squeleton of the code that have been used for the numerical simulations and all the finite volumes subroutines. The infancy of this study was done with with K. Mer, CEA CESTA and she has coded the N scheme. I have been encouraged by my colleagues P. Charrier and B. Dubroca. Numerous discussions with H. Deconinck, P.L. Roe, T. Barth and K. Sermeus have also been very helpful. The referees are also acknowledged for their very accurate comments which led to drastic improvements of the paper.

References

[1] R. Abgrall. On Essentially Non-Oscillatory Schemes on Unstructured Meshes : Analysis and Implementation. *J. Comp. Phys.*, 114(1):45–54, September 1994.

[2] R. Abgrall and T.J. Barth. Linearisation via the entropy variables : application to residual distributive schemes. , Rapport de Recherches n^0 00-15, Mathématiques Appliquees de Bordeaux, 2000. Submitted.

[3] R. Abgrall and K. Mer. Un théorème de type Lax–Wendroff pour les schémas distributifs. Technical Report 98010, Mathématiques Appliquées de Bordeaux, Mars 1998.

[4] T.J. Barth. Some working notes on the N scheme. Private communication, 1997.

[5] H. Deconinck, P.L. Roe, and R. Struijs. A multidimensional generalization of Roe's difference splitter for the Euler equations. *Computer and Fluids*, 22:215–222, 1993.

[6] H. Deconinck, R. Struijs, G. Bourgeois, and P.L. Roe. Compact advection schemes on unstructured meshes. VKI Lecture Series 1993–04, Computational Fluid Dynamics, 1993.

[7] A. Dervieux, B. van Leer, J. Périaux, and A. Rizzi, editors. *Numerical Simulation of Compressible Euler Flows*, volume 26 of *Note on Numerical Fluid mechanics*. Vieweg, 1989.

[8] L. Fezoui and B. Stoufflet. A class of implicit upwind schemes for Euler simulations with unstructured meshes. *J. Comput. Phys.*, 84(1):174–206, 1989.

[9] A. Harten. On the symmetric form of conservation laws with entropy. *J. Comp. Phys.*, 49:151–164, 1983.

[10] M.Y. Hussaini, B. van Leer, and J. Van Rosendal, editors. *Upwind and High Resolution Schemes*. Springer Verlag, 1997.

[11] P. Lax and B. Wendroff. Systems of conservation laws. *Comm. Pure Appl. Math.*, 13:381–394, 1960.

[12] P.L. Roe and D. Sidilkover. Optimum positive linear schemes for advection in two and three dimensions. *Siam J. Numer. Anal.*, 29(6):1542–1588, 1992.

[13] R. Struijs, H. Deconinck, and P. L. Roe. Fluctuation Splitting Schemes for the 2D Euler equations. *VKI LS 1991-01, Computational Fluid Dynamics*, 1991.

[14] E. Tadmor. The numerical viscosity of entropy stable schemes for systems of conservation laws. *Math. Comp.*, 49:91–103, 1987.

[15] E. van der Weide and H. Deconinck. Positive matrix distribution schemes for hyperbolic systems. In *Computational Fluid Dynamics '96*, pages 747–753. Wiley, 1996.

[16] O. Axelsson. *Iterative Solution Methods*. Cambridge University Press, Cambridge, England, 1996.

A Proof of Lemma 4.1

The proof of the energy inequality was made by T. Barth and is presented in [2]. It is included in the present paper for completeness, with permission from the original author.

For ease of exposition, we will show the development in two space dimensions but the generalization to $\mathbb{R}^d$ will be clear. The analysis is done for a symetrized linear hyperbolic system : we assume that the matrices K_i are symmetric, the state variables are V_i.

Setting $V = (V_1, V_2, V_3)^T$, the N scheme can be rewritten as

$$L_T V_T = \begin{pmatrix} \Phi_1 \\ \Phi_2 \\ \Phi_3 \end{pmatrix} = \left[\begin{bmatrix} K_1^+ & & \\ & K_2^+ & \\ & & K_3^+ \end{bmatrix} + \begin{bmatrix} K_1^+ \\ K_2^+ \\ K_3^+ \end{bmatrix} [N] \begin{bmatrix} K_1^- \\ K_2^- \\ K_3^- \end{bmatrix}^T \right] \begin{pmatrix} V_1 \\ V_2 \\ V_3 \end{pmatrix} \tag{43}$$

with $K^\pm$ symmetric and $[N]$ a block diagonal matrix $[N] \equiv \mathrm{diag}(N, N, N)$.

The study of the quadratic form

$$\langle V_1, \Phi_1 \rangle + \langle V_2, \Phi_2 \rangle + \langle V_3, \Phi_3 \rangle$$

amounts to studying the symmetric matrix $(L + L^T)/2$. The study of $\mathcal{Q}_N$,

$$\mathcal{Q}_N(V_1, V_2, V_3) = \langle V_1, \Phi_1 \rangle + \langle V_2, \Phi_2 \rangle + \langle V_3, \Phi_3 \rangle - \frac{1}{2}\left(\sum_{i=1}^{3} \langle V_i, K_i V_i \rangle\right)$$

is this equivalent to the study of

$$\frac{1}{2}\left(L + L^T\right) - \frac{1}{2}\begin{bmatrix} K_1 & & \\ & K_2 & \\ & & K_3 \end{bmatrix}.$$

The symmetric part of L is given by

$$L_T = \begin{bmatrix} K_1^+ & & \\ & K_2^+ & \\ & & K_3^+ \end{bmatrix} + \frac{1}{2}\begin{bmatrix} K_1^+ \\ K_2^+ \\ K_3^+ \end{bmatrix} [N] \begin{bmatrix} K_1^- \\ K_2^- \\ K_3^- \end{bmatrix}^T + \frac{1}{2}\begin{bmatrix} K_1^- \\ K_2^- \\ K_3^- \end{bmatrix} [N] \begin{bmatrix} K_1^+ \\ K_2^+ \\ K_3^+ \end{bmatrix}^T . \quad (44)$$

Examining rows of L_T or L_T, observe that the row sum is nonzero. However, we can add the following block diagonal matrix to the element matrix L

$$-\frac{1}{2}\begin{bmatrix} K_1 & & \\ & K_2 & \\ & & K_3 \end{bmatrix} \quad (45)$$

so that rows and columns of the L_T sum to zero. These additional terms have no impact on the constant coefficient discretization of the Cauchy problem. These terms all vanish identically when summed for all elements sharing a mesh vertex since the geometry surrounding the vertex is closed. Hence forward, we will include these terms in our definition of L_T and T_T yielding

$$L_T = \frac{1}{2}\begin{bmatrix} |K|_1 & & \\ & |K|_2 & \\ & & |K|_3 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} K_1^+ \\ K_2^+ \\ K_3^+ \end{bmatrix} [N] \begin{bmatrix} K_1^- \\ K_2^- \\ K_3^- \end{bmatrix}^T + \frac{1}{2}\begin{bmatrix} K_1^- \\ K_2^- \\ K_3^- \end{bmatrix} [N] \begin{bmatrix} K_1^+ \\ K_2^+ \\ K_3^+ \end{bmatrix}^T . \quad (46)$$

Next, rewrite off-diagonal term such as

$$K_i^+ N K_j^- + K_i^- N K_j^+$$

in the following form

$$K_i^+ N K_j^- + K_i^- N K_j^+ = K_i N K_j - K_i^+ N K_j^+ - K_i^- N K_j^- \ .$$

Consequently, L_T can be rewritten as

$$\begin{aligned} L_T &= \frac{1}{2}\begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} [N] \begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix}^T \\ &+ \frac{1}{2}\left[\begin{bmatrix} K_1^+ & & \\ & K_2^+ & \\ & & K_3^+ \end{bmatrix} - \begin{bmatrix} K_1^+ \\ K_2^+ \\ K_3^+ \end{bmatrix} [N] \begin{bmatrix} K_1^+ \\ K_2^+ \\ K_3^+ \end{bmatrix}^T\right] \\ &+ \frac{1}{2}\left[\begin{bmatrix} -K_1^- & & \\ & -K_2^- & \\ & & -K_3^- \end{bmatrix} - \begin{bmatrix} -K_1^- \\ -K_2^- \\ -K_3^- \end{bmatrix} [N] \begin{bmatrix} -K_1^- \\ -K_2^- \\ -K_3^- \end{bmatrix}^T\right]. \end{aligned} \tag{47}$$

Note that the first term appearing on the right hand side of Eqn. (47) gives rise to a quadratic form with positive energy so our only concern is the remaining terms on the right hand side on this equation. Before proving positive semi-definiteness of (47), we first review a simple result concerning the spectra of non-commuting matrices.

Lemma A.1. *The nonzero parts of the spectrum of AB and BA are identical for all matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times m}$.*

Proof. Omitted, see for example Axelsson [16] p. 69. □

Next we prove positive semi-definiteness of a specialized matrix in product form.

Lemma A.2 (Golub). *The matrix*

$$L = \begin{bmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{bmatrix} - \begin{bmatrix} A \\ B \\ C \end{bmatrix} N \begin{bmatrix} A \\ B \\ C \end{bmatrix}^T, \quad N = [A + B + C]^{-1}$$

is positive semidefinite for all $A, B, C \in \mathbb{R}^{n\times n}$ symmetric positive definite.

Proof. Let

$$Z = \begin{bmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{bmatrix}$$

and congruence transform L

$$Z^{-1/2} L Z^{-1/2} = \begin{bmatrix} I_n & & \\ & I_n & \\ & & I_n \end{bmatrix} - \begin{bmatrix} A^{1/2} \\ B^{1/2} \\ C^{1/2} \end{bmatrix} N \begin{bmatrix} A^{1/2} \\ B^{1/2} \\ C^{1/2} \end{bmatrix}^T = I_{3n} - P.$$

Next use Lemma A.1 concerning the spectra of non-square matrix products. In the present case Lemma A.1 this implies that

$$
\begin{aligned}
\text{Eigenvalues}\left(\begin{bmatrix} A^{1/2} \\ B^{1/2} \\ C^{1/2} \end{bmatrix} N \begin{bmatrix} A^{1/2} \\ B^{1/2} \\ C^{1/2} \end{bmatrix}^T\right) &= \text{Eigenvalues}\Big(N^{1/2}(A+B+C)N^{1/2}\Big) + 2n \text{ zeros} \\
&= \text{Eigenvalues}\Big(N(A+B+C)\Big) + 2n \text{ zeros} \\
&= \text{Eigenvalues}(I_n) + 2n \text{ zeros} \qquad (48)
\end{aligned}
$$

and consequently

$$I_{3n} - P$$

is positive semidefinite. From this result it follows immediately that

$$L = Z^{1/2}(I_{3n} - P)Z^{1/2}$$

is also positive semidefinite. □

The extension to $A, B, C \geq 0$ and $(A+B+C) > 0$ follows by considering the perturbed matrices $A_\epsilon = A + \epsilon I$, $B_\epsilon = B + \epsilon I$, and $C_\epsilon = C + \epsilon I$ and letting $\epsilon \downarrow 0$.

Returning to the system N-scheme, we now can prove the

Lemma A.3 (Lemma 4.1). *If the matrices K_i are symmetric, one has*

$$\sum_{M_i \in T} \left\langle V_i, \Phi_i^T \right\rangle = \frac{1}{2} \sum_{M_i \in T} \langle V_i, K_i V_i \rangle + \mathcal{Q}_N(V_1, V_2, V_3) \qquad (49)$$

where

$$
\begin{aligned}
2\mathcal{Q}_N(V_1, V_2, V_3) &= -\left\langle \Phi^T N, \Phi^T \right\rangle \\
&+ \textstyle\sum_{M_i \in T} \left(\left\langle V_i, K_i^+ V_i \right\rangle - \left\langle K_i^+ V_i, N K_i^+ V_i \right\rangle \right) \qquad (50) \\
&+ \textstyle\sum_{M_i \in T} \left(\left\langle V_i, -K_i^- V_i \right\rangle - \left\langle -K_i^+ V_i, N(-K_i^-) V_i \right\rangle \right).
\end{aligned}
$$

The quadratic form $\mathcal{Q}_N$ is positive : the N scheme is locally dissipative.

Proof. Since $N = [K_1^+ + K_2^+ + K_3^+]^{-1} = [-K_1^- - K_2^- - K_3^-]^{-1}$, the result follows immediately after application of the Golub lemma to (47). □

B The N and LDA schemes are well defined

We show that for a linearized symetrizable system, the system N and LDA schemes are well defined. We carry out the proof only for the N scheme, the extension to the LDA one is obvious.

We consider a linearisation of the system

$$W_t + AW_x + BW_y = 0$$

by mean of some parameter vector. The Jacobian are evaluated at some average state, for example that given by the parameter vector Z. It could be any other linearisation provided the symetrisation property of the system are kept.

Formally, the residual within a triangle T of the system N scheme writes

$$\Phi_i^N = \sum_{M_j \in T} C_{ij}(\widetilde{W}_i - \widetilde{W}_j)$$

where the matrices C_{ij} are

$$C_{ij} = K_i^+ \left(\sum_{l=1,3} K_l^- \right)^{-1} K_j^- \tag{51}$$

Here $K_l = An_x^l + Bn_y^l$ where (n_x^l, n_y^l) are the component of a vector $\vec{n}_l$. The vectors $\vec{n}_l$, $l = 1, 2, 3$ satisfies $\sum_{l=1}^3 \vec{n}_l = 0$. In the sequel, we set $N = (\sum_{l=1,3} K_l^-)^{-1}$. The question is to see whether the matrices C_{ij} are well defined.

Since the system is symetrizable, there exists a symmetric positive definite (s.d.p.) matrix A_0 such that

$$\widetilde{A} = AA_0^{-1}\ ,\quad \widetilde{B} = BA_0^{-1}.$$

are symmetric. Here, A_0 is the Hessian of the mathematical entropy S evaluated at the average state. We set $\widetilde{K}_l = \widetilde{A}n_x^l + \widetilde{B}n_y^l = K_l A_0^{-1}$. Since A_0 is s.d.p., we can left and right multiply by $A_0^{1/2}$ to see that

$$\widetilde{K}_l^+ = K_l^+ A_0^{-1} \quad , \quad \widetilde{K}_l^- = K_l^- A_0^{-1}$$

$$|\widetilde{K}_l| = |K| A_0^{-1}.$$

We also have $\sum_l K_l = 0$, $\sum_l \widetilde{K}_l = 0$.

Since $\widetilde{A}$ and $\widetilde{B}$ are symmetric, the $\widetilde{K}_l$, $\widetilde{K}_l^{\pm}$, $|\widetilde{K}_l|$ are also symmetric. Hence

$$\sum_l \widetilde{K}_l^+ \geq 0$$

$$\sum_l \widetilde{K}_l^- \leq 0$$

$$\sum_l |\widetilde{K}_l| \geq 0$$

This shows that $\sum_l |K_l|$, $\sum_l K_l^+$, $-\sum_l K_l^-$ have positive eigenvalues ; this can be seen still by left and right multiplying by $A_0^{1/2}$ and $A_0^{-1/2}$.

If one matrix K_l^+ has a system of strictly positive eigenvalues, $\sum_l K_l^+$ has only strictly positive eigenvalues. Thus $\sum_l K_l^+$ is invertible.

Assume now there exists $x \in \mathbb{R}^4$ such that $\sum_l K_l^+ x = 0$. By setting $y = A_0 x$, we have $\sum_l \widetilde{K}_l^+ y = 0$. Thus,

$$0 = \left\langle \sum K_l^+ x, x \right\rangle = \sum_l \left\langle \widetilde{K}_l^+ y, y \right\rangle .$$

If there exists l such that $\left\langle \widetilde{K}_l^+ y, y \right\rangle > 0$, y cannot be in the kernel of $\sum \widetilde{K}_i'^+$ unless $y = 0$ and then $x = 0$.

Then we have to assume that $\widetilde{K}_l^+ y = 0$ for $l = 1, 2, 3$. Since $\sum \widetilde{K}_l^+ = -\sum \widetilde{K}_l^-$, the same arguments as above, applied to $\widetilde{K}_l^-$ shows that $\widetilde{K}_l^- y = 0$ for $l = 1, 2, 3$. Thus $K_l x = 0$, with $x \neq 0$. Coming back to the definition of K_l, since any two among $\{\vec{n}_1, \vec{n}_2, \vec{n}_3\}$ are linearly independent, we have $Ax = Bx = 0$: the matrices A and B have a common eigenvector, associated to the eigenvalue 0.

In the case of the Euler equations, the eigenvectors of A are

$$R_1^+ = \begin{pmatrix} 1 \\ u+a \\ v \\ H+ua \end{pmatrix} \quad R_1^- = \begin{pmatrix} 1 \\ u-a \\ v \\ H-ua \end{pmatrix} \quad R_1^0 = \begin{pmatrix} 1 \\ u \\ v \\ \dfrac{u^2+v^2}{2} \end{pmatrix} \quad R_1^t = \begin{pmatrix} 0 \\ 0 \\ 1 \\ v \end{pmatrix}$$

with the eigenvalues $u+a$, $u-a$, u et u. Those of B are

$$R_2^+ = \begin{pmatrix} 1 \\ u \\ v+a \\ H+ua \end{pmatrix} \quad R_2^- = \begin{pmatrix} 1 \\ u \\ v-a \\ H-ua \end{pmatrix} \quad R_2^0 = \begin{pmatrix} 1 \\ u \\ v \\ \dfrac{u^2+v^2}{2} \end{pmatrix} \quad R_2^t = \begin{pmatrix} 0 \\ -1 \\ 0 \\ u \end{pmatrix}$$

with the eigenvalues $v+a$, $v-a$, v, v. We see that $R_1^0 = R_2^0 = r_0$

The only solution to the problem is $u = v = 0$ and $x = \lambda r_0 = \lambda r_0$ (stagnation point) otherwise $a = 0$ which corresponds to vacuum.

What remains to do is to show that one can give a meaning to C_{il} even in that case. More precisely, we show there exists a decomposition of $\mathbb{R}^4$ such that $\mathbb{R}^4 = \mathbb{R}r_0 \bigoplus H$ where H contains all the eigenvectors of A and B that are different of r_0.

Lemma B.1. *If A and B are two matrices with one common eigenvector r_0, and if there exists a s.p.d. matrix A_0 that symmetrises A and B there exists a vector space H that can be explicitly computed such that the other eigenvectors of A and B belong to H and $\mathbb{R}^n = (\mathbb{R}r_0) \bigoplus H$.*

Proof. The matrices AA_0^{-1} and BA_0^{-1} are symmetric, so are $A_0^{1/2} A A_0^{-1/2}$ and $A_0^{1/2} B A_0^{-1/2}$ (left and right multiply by $A_0^{1/2}$). They are also congruent to A and B respectively.

Let $\{r_k\}_{k=0,n-1}$ and $\{r_k'\}_{k=0,n-1}$ complete systems of eigenvectors of A and B respectively. We assume $r_0 = r_0'$. Then $\{A_0^{1/2} r_k\}_{k=0,n-1}$ and $\{A_0^{1/2} r_k'\}_{k=0,n-1}$ are complete systems of eigenvectors of symmetric matrices : they are orthogonal. Hence

$$\mathbb{R}^n = \mathbb{R}A_0^{1/2} r_0 \bigoplus (\mathbb{R}A_0^{1/2} r_0)^\perp .$$

Clearly , $A_0^{1/2} r_k^{'} \in (\mathbb{R}A_0^{1/2} r_0)^\perp$ et $A_0^{1/2} r_k \in (\mathbb{R}A_0^{1/2} r_0)^\perp$ for $k > 0$. By defining

$$H = A_0^{-1/2} (\mathbb{R}A_0^{1/2} r_0)^\perp$$

we have the expected decomposition. □

$$x = l(x) r_0 + x^\perp, x^\perp \in H$$

where $l(x) = \dfrac{\langle A_0 x, x\rangle}{\langle A_0 r_0, r_0\rangle}$, we have

$$M_{ij} x = \left(\frac{\langle \vec{u}, \vec{n_i}\rangle^+ \langle \vec{u}, \vec{n_j}\rangle^-}{\sum_{i=1,3} \langle \vec{u}, \vec{n_i}\rangle^+} \right) l(x) r_0 + M_{ij} x^\perp.$$

When $\vec{u} \to 0$, the first term tends to 0 because

$$\left| \frac{\langle \vec{u}, \vec{n_i}\rangle^+ \langle \vec{u}, \vec{n_j}\rangle^-}{\sum_{i=1,3} \langle \vec{u}, \vec{n_i}\rangle^+} \right| \leq |\langle \vec{u}, \vec{n_j}\rangle^-| \longrightarrow 0$$

and $M_{ij} x^\perp$ converges to a finite limit because none of the eigenvalues of the restriction of N to the space H vanishes.

Remark B.2. *Since $\pi^* A_0 = A_0 \pi$, we can see that $A_0 r_0 = v_0$. Hence H is the kernel of $\pi^\perp$.*

C Proof of Lemmas 4.2 and 4.3

In this section, we assume that the linearisation is done via the entropy variable V and we use the notations of the previous section.

Throughout this section, we can assume that the N, at least for the symmetrisable system of the Euler equations. We set $M = -N$ and

$$\begin{aligned} V^+ &= M \left(\textstyle\sum_{i=1}^3 \widetilde{K_i}^+ V_i) \right) \\ V^- &= N \left(\textstyle\sum_{i=1,3} \widetilde{K}_i^- V_i \right) \end{aligned}$$

so that we have $\Phi^T = M^{-1}(V^+ - V^-)$.

Proof of Lemma 4.2. A direct calculation shows (with obvious notations)

$$\mathcal{Q}_{LDA}(V_1, V_2, V_3) - \mathcal{Q}_N(V_1, V_2, V_3) = \sum_{i=1}^3 \left\langle V_i, \Phi_i^{LDA} - \Phi_i^N \right\rangle.$$

Since $\Phi_i^{LDA} - \Phi_i^N = \widetilde{K}_i^+ (V^+ - V_i)$ and since

$$\sum_{i=1}^3 K_i^+ V^+ = \sum_{i=1}^3 K_i^+ V_i,$$

we get

$$\begin{aligned}\mathcal{Q}_{LDA}(V_1,V_2,V_3)-\mathcal{Q}_N(V_1,V_2,V_3) &= \sum_{i=1}^3 \left\langle V_i, \widetilde{K}_i^+(V^- - V_i)\right\rangle \\ &= -\sum_{i=1}^3 \left\langle V^- - V_i, \widetilde{K}_i^+(V^- - V_i)\right\rangle \\ &\leq 0\end{aligned}$$

□

Proof of Lemma 4.3. The result is true if the first part of the lemma is shown. The N scheme writes (in symmetric variables)

$$\Phi_i = \sum_{j=1}^3 \widetilde{K}_i^+ \widetilde{N} \widetilde{K}_j^- (V_i - V_j)$$

Since $\widetilde{K_i^+} = K_i^+ A_0^{-1}$, $\widetilde{K_j^-} = K_j^- A_0^{-1}$ and $\widetilde{N} = A_0 N$, we have

$$\Phi_i = \sum_{j=1}^3 K_i^+ N K_j^- A_0^{-1}(V_i - V_j).$$

Now, π commutes with K_i^+, K_j^- and N, and since $\pi^2 = \pi$ and $\pi A_0 = A_0 \pi^*$, we get the result for the N scheme. The proof is identical for the LDA scheme. The second part of the lemma is a consequence of its first part. □

D The blended scheme is LP at convergence

In order to show that the blended scheme is LP at convergence, we have to check the condition (15) when the arguments in the residuals are replaced by the exact smooth solution. Here,

$$\sum_T \sum_{M_i \in T} (\varphi_i - \varphi_G^T).\Phi_i^T = \sum_T \sum_{M_i \in T} (\varphi_i - \varphi_G^T).\left(\Phi_i^{LDA} + \ell(\Phi_i^N - \Phi_i^{LDA})\right)$$

The LDA scheme is LP because the matrices $-K_i^+ N K_j$ are uniformly bounded is the solution is itself bounded Since ℓ is also bounded, it is enough to check if

$$\sum_T \sum_{M_i \neq M_j} (\varphi_i - \varphi_G^T).\ell\Phi_i^N = \mathcal{O}(h^2).$$

In fact ,

$$\Phi_i^N = \sum_{j\neq i} K_i^+ N K_j^- (\widetilde{W}_i - \widetilde{W}_j) = \sum_{j \neq i} K_i^+ N K_j^- \left\langle \nabla \widetilde{\pi_h(W)}, \vec{M_i M_j}\right\rangle$$

where $\pi_h(W)$ is the piecewise linear interpolant of W.[2] The matrices NK_j^- are bounded, the matrices K_i^+ are $\mathcal{O}(h)$, and $\vec{M_iM_j}$ is also $\mathcal{O}(h)$ if the mesh is regular. Hence $\Phi_i^N = \mathcal{O}(h^2)$

[2] $\vec{M_iM_j}$ is the vector starting at point M_i and ending at point M_j.

on any triangle. Moreover, $\varphi_i - \varphi_G^T = \mathcal{O}(h)$, so it is enough to check if $\ell = \mathcal{O}(h)$. Since W is the exact solution and $\pi_h(W)$ the piecewise linear interpolant of W, we have

$$\langle v_i, l(\Phi_i^{LDA} \rangle = \mathcal{O}(h^3) \text{ and } \langle v_i, l(\pi\Phi_i^{LDA} \rangle = \mathcal{O}(h^3).$$

As we have seen just above,

$$\langle v_i, l(\Phi_i^{N} \rangle = \mathcal{O}(h^2) \text{ and } \langle v_i, l(\pi\Phi_i^{N} \rangle = \mathcal{O}(h^2).$$

and thus $\ell = \mathcal{O}(h)$.

However, there is an important difference between what is done on the entropy wave and its orthogonal. On the entropy wave, the scheme can be naturally associated to a finite element method with discontinuous test functions, as the SUPG method. This is much less clear for what we do on the orthogonal complement of the entropy wave.

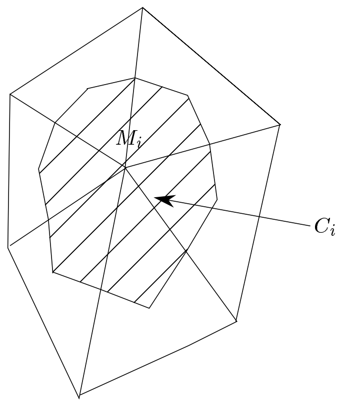


Figure 1: The dual cell is obtained by joining the midpoints of the edges starting from M_i and the centroids of the triangles containing M_i as a vertex.

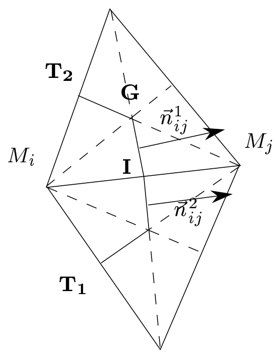


Figure 2: Geometrical elements for the finite volume scheme. The normal $\vec{n}^1_{ij}$ is orthogonal to $[G, I]$ with the same length. Same for $\vec{n}^2_{ij}$.

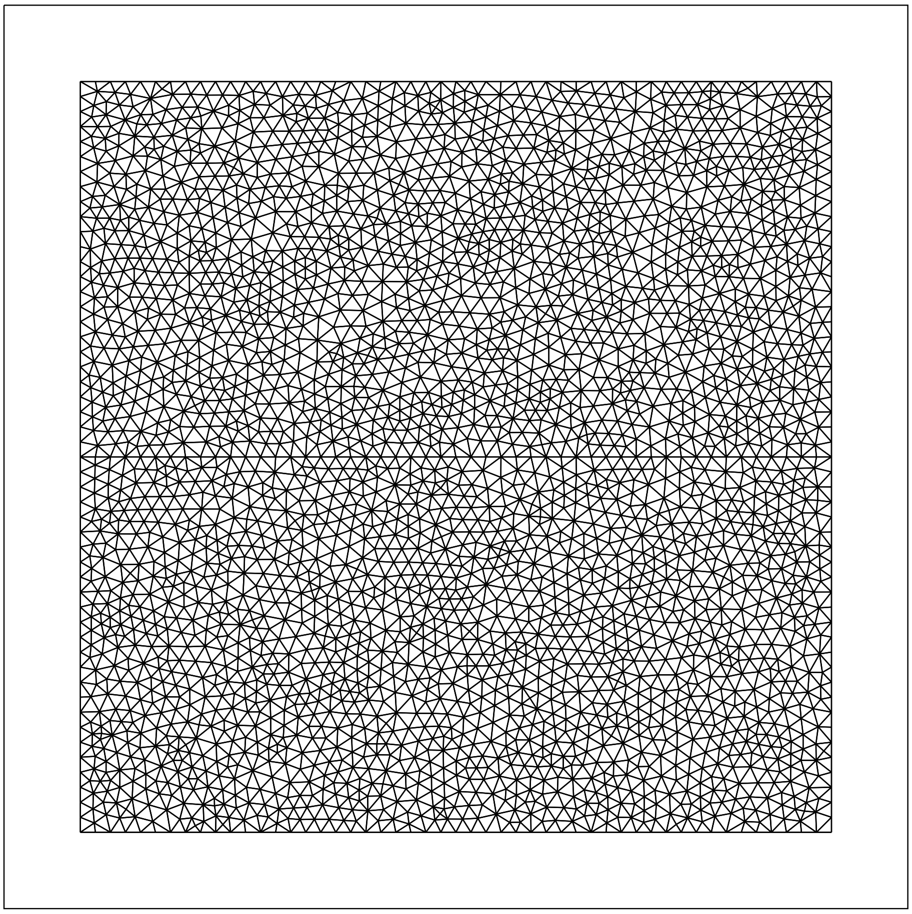


Figure 3: Mesh for the shock tube problem

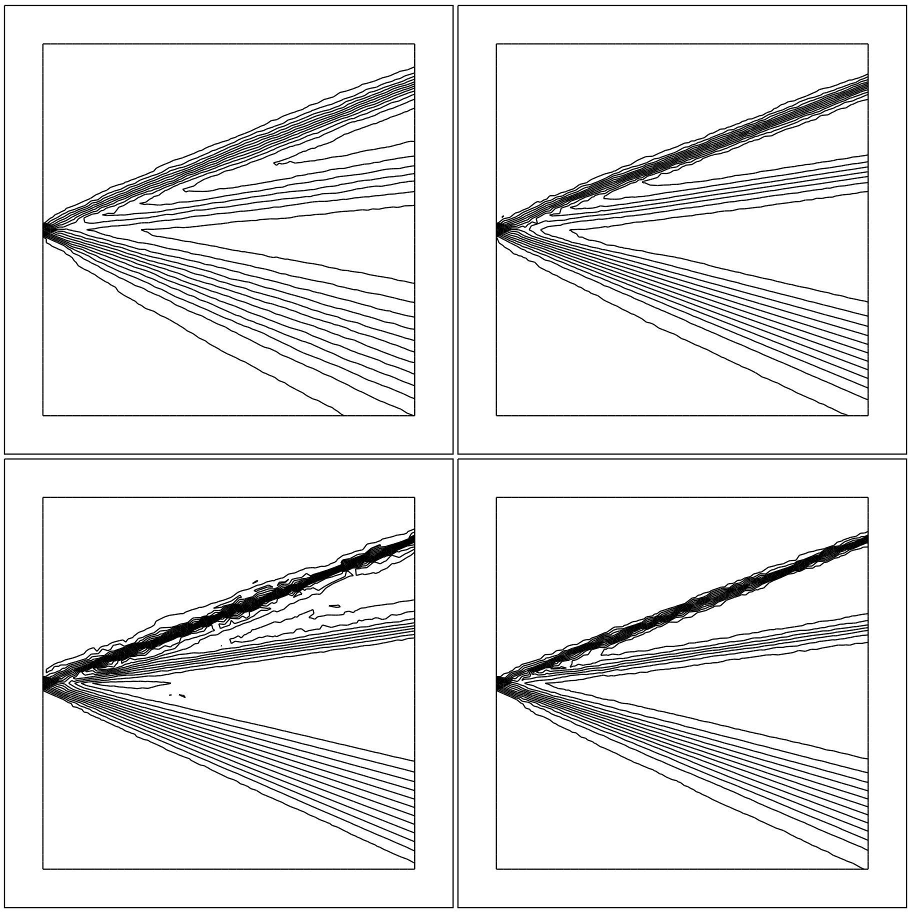


Figure 4: Density isolines for the shock tube problem. Top : N scheme (left), Muscl scheme (right), Bottom : LDA scheme (left), Blended scheme (right). N scheme : $0.7 \leq \rho \leq 1.4$, Muscl : $0.69 \leq \rho \leq 1.40$, LDA : $0.61 \leq \rho \leq 1.42$, Blended : $0.70 \leq \rho \leq 1.40$

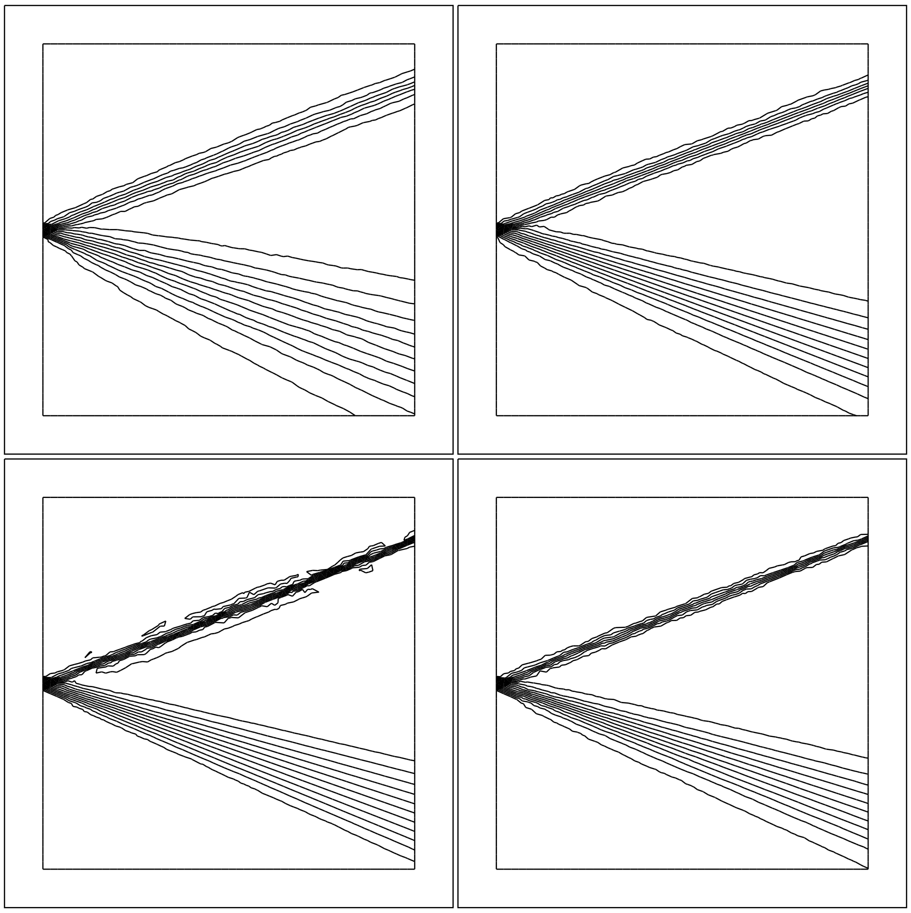


Figure 5: Pressure isolines for the shock tube problem. Top : N scheme (left), Muscl scheme (right), Bottom : LDA scheme (left), Blended scheme (right). N scheme : $0.25 \leq p \leq 1.$, Muscl : $0.24 \leq p \leq 1.$, LDA : $0.20 \leq p \leq 1.02$, Blended : $0.25 \leq p \leq 1$.

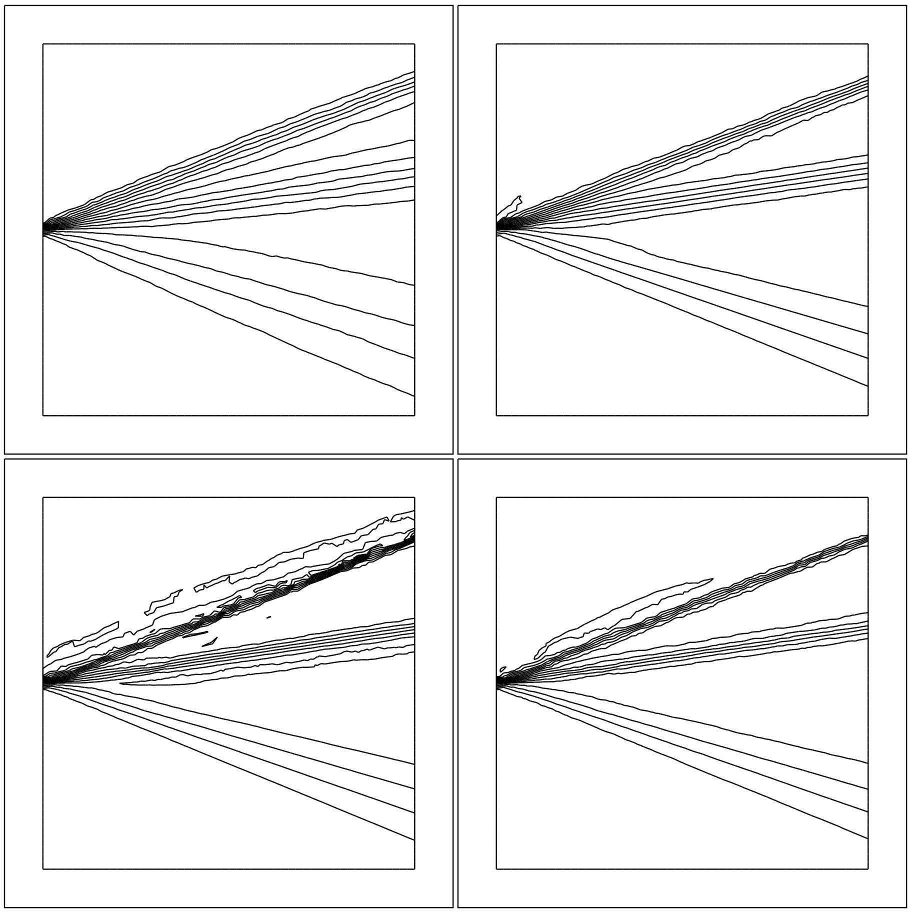
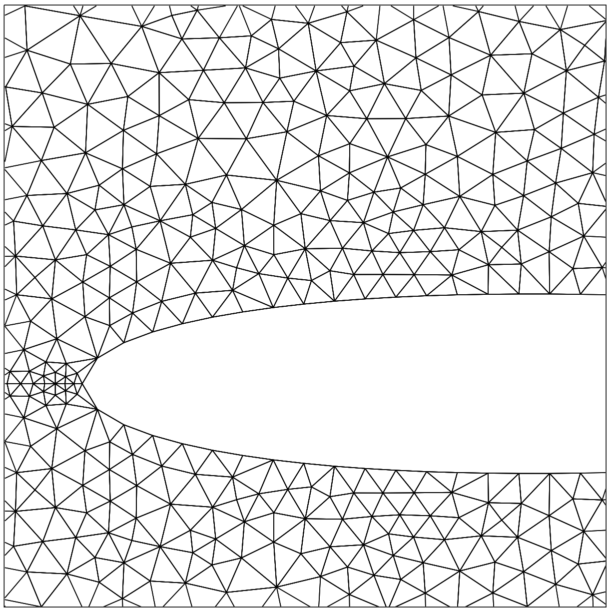
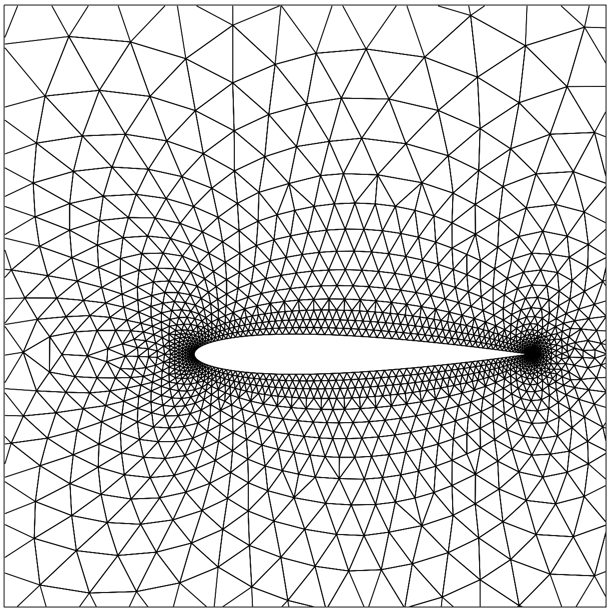


Figure 6: Mach number isolines for the shock tube problem. Top : N scheme (left), Muscl scheme (right), Bottom : LDA scheme (left), Blended scheme (right). N scheme : $2.4 \leq M \leq 4.$, Muscl : $2.40 \leq M \leq 4.10$, LDA : $2.39 \leq M \leq 4.27$, Blended : $2.40 \leq M \leq 4.00$

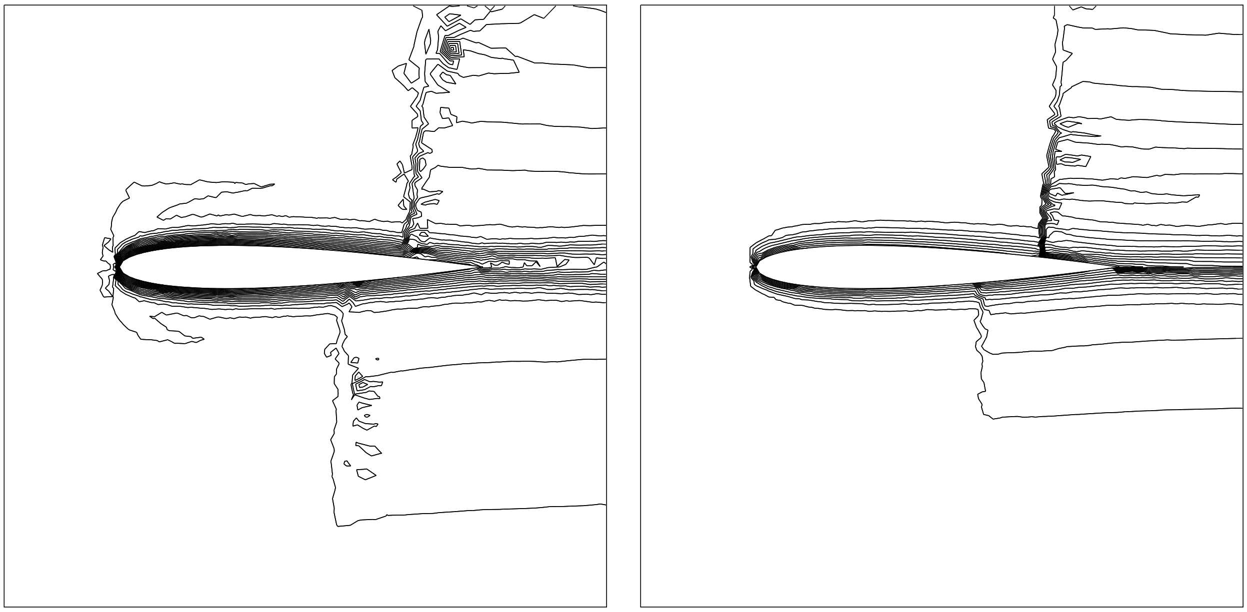


(a)



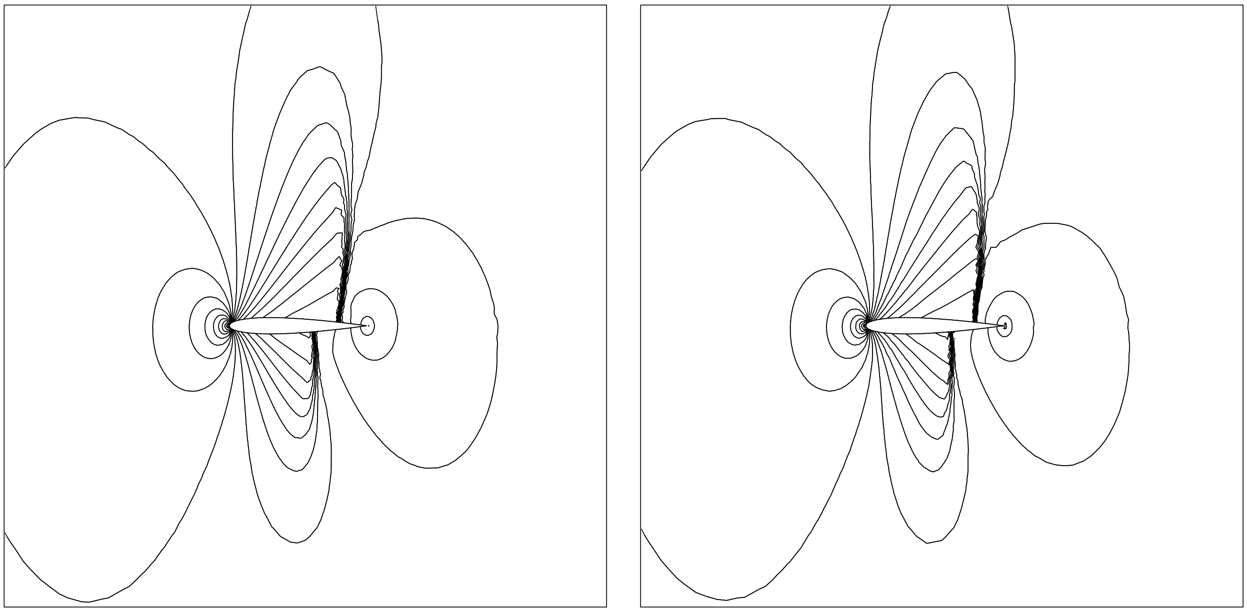
(b)

Figure 7: Meshes for the Naca 0012 problem



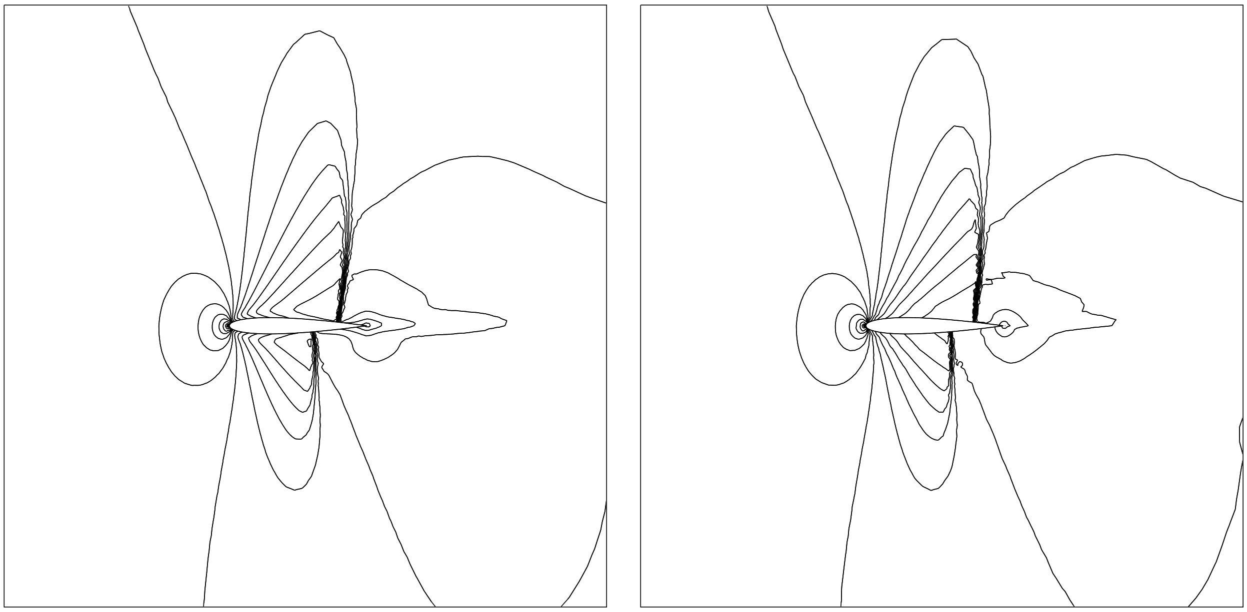
(a) (b)

Figure 8: Deviation of physical entropy ($\Sigma = \frac{s-s_\infty}{s_\infty}$) for the MUSCL scheme (a) and the blended scheme (b).



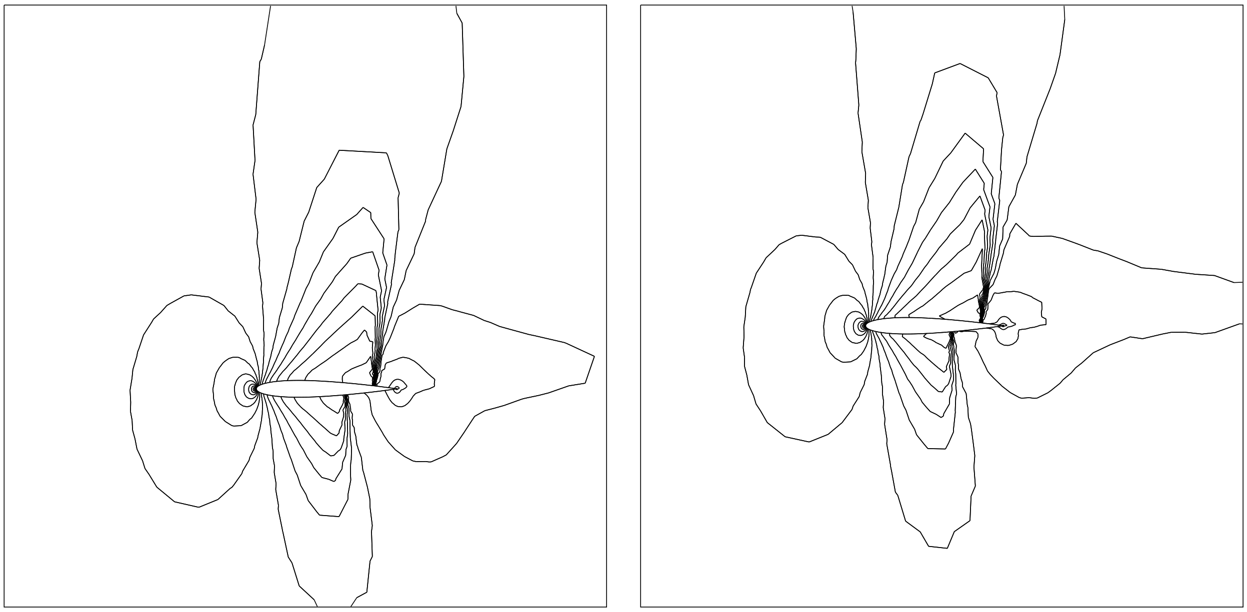
(a) (b)

Figure 9: Pressure coefficient for the MUSCL scheme (a) and the blended scheme (b).



(a) (b)

Figure 10: Mach number contours for the MUSCL scheme (a) and the blended scheme (b).



(a) (b)

Figure 11: Mach number contours on the regular mesh of Figure 7-b for the MUSCL scheme (a) and the blended scheme (b).

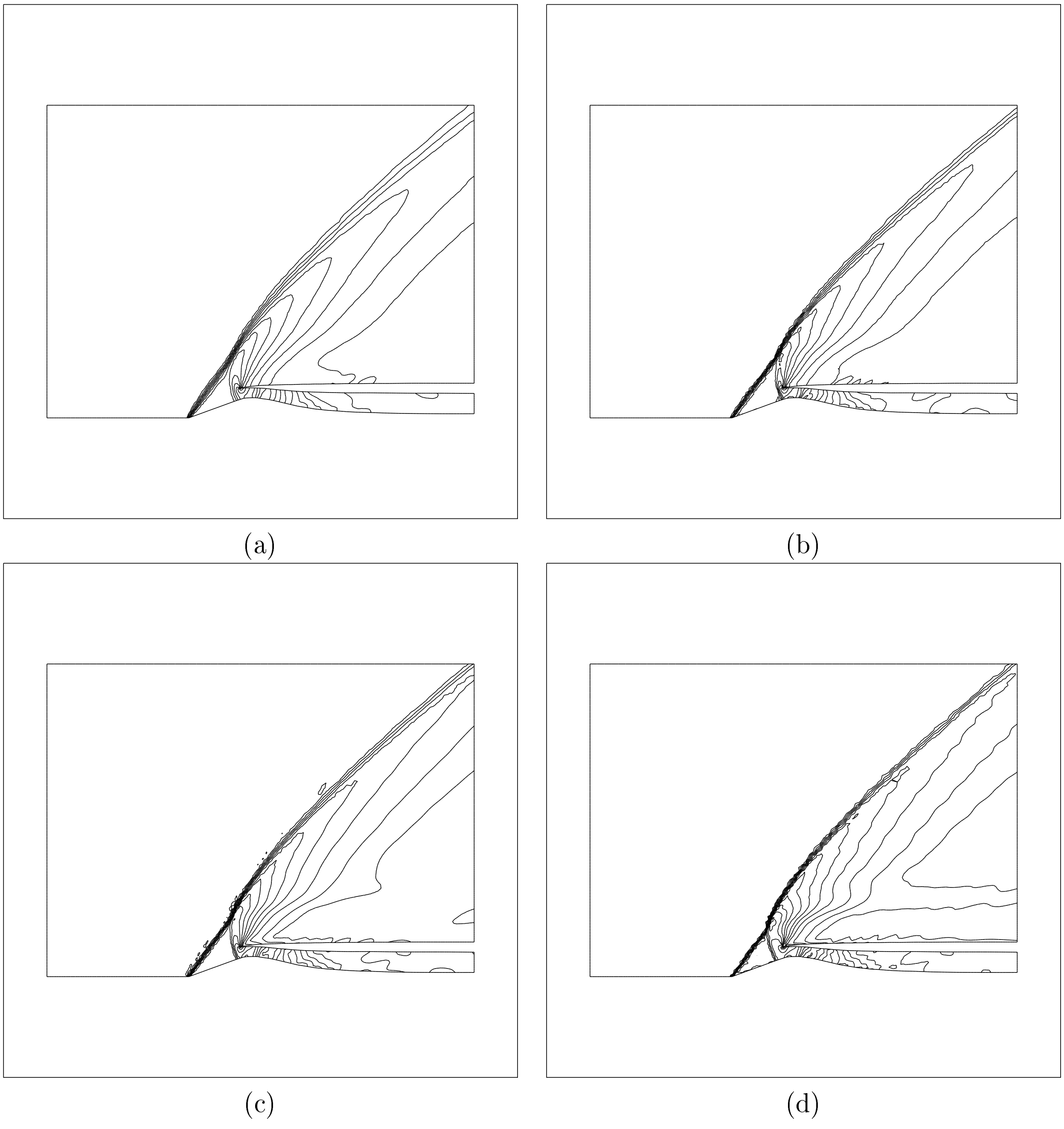
Irregular mesh, Figure 7-a

Scheme	Mach (Min,Max)	c_p (Min,Max)	Σ (Min,Max)
Muscl	0.053, 1.34	-1.00, 1.05	-0.005, 0.054
Blended	0.022, 1.39	-1.00, 1.08	-0.0005, 0.052

Regular mesh, Figure 7-b

Scheme	Mach (Min,Max)
Muscl	0.047, 1.42
Blended	0.05, 1.43

Table 1: Naca 0012 problem on meshes 7-a,b. Min and max of the Mach number, the pressure coefficient and the entropy deviation.



(a) (b) (c) (d)

Figure 12: Mach number contours for the 1st order Roe scheme (a), the system N scheme (b), the MUSCL scheme (c) and the blended scheme (d).

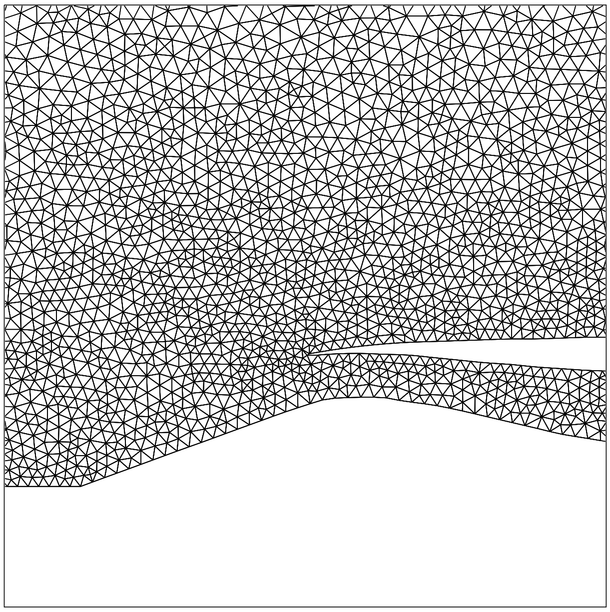


Figure 13: Zoom of the mesh near the inlet entrance

Scheme	Mach (Min,Max)	Pressure (Min,Max)	Σ (Min,Max)
Roe	0.50, 2.21	0.57, 5.76	0.0, 0.14
N	0.53, 2.25	0.58, 6.06	0.0, 0.10
Muscl	0.48, 2.26	0.56, 5.90	-0.01, 0.16
Blended	0.38, 2.27	0.56, 6.75	-0.0008, 0.17

Table 2: Engine inlet problem, Min and max of the Mach number, pressure and entropy deviation Σ for the first order roe scheme, the N scheme, the Muscl scheme and the blended scheme

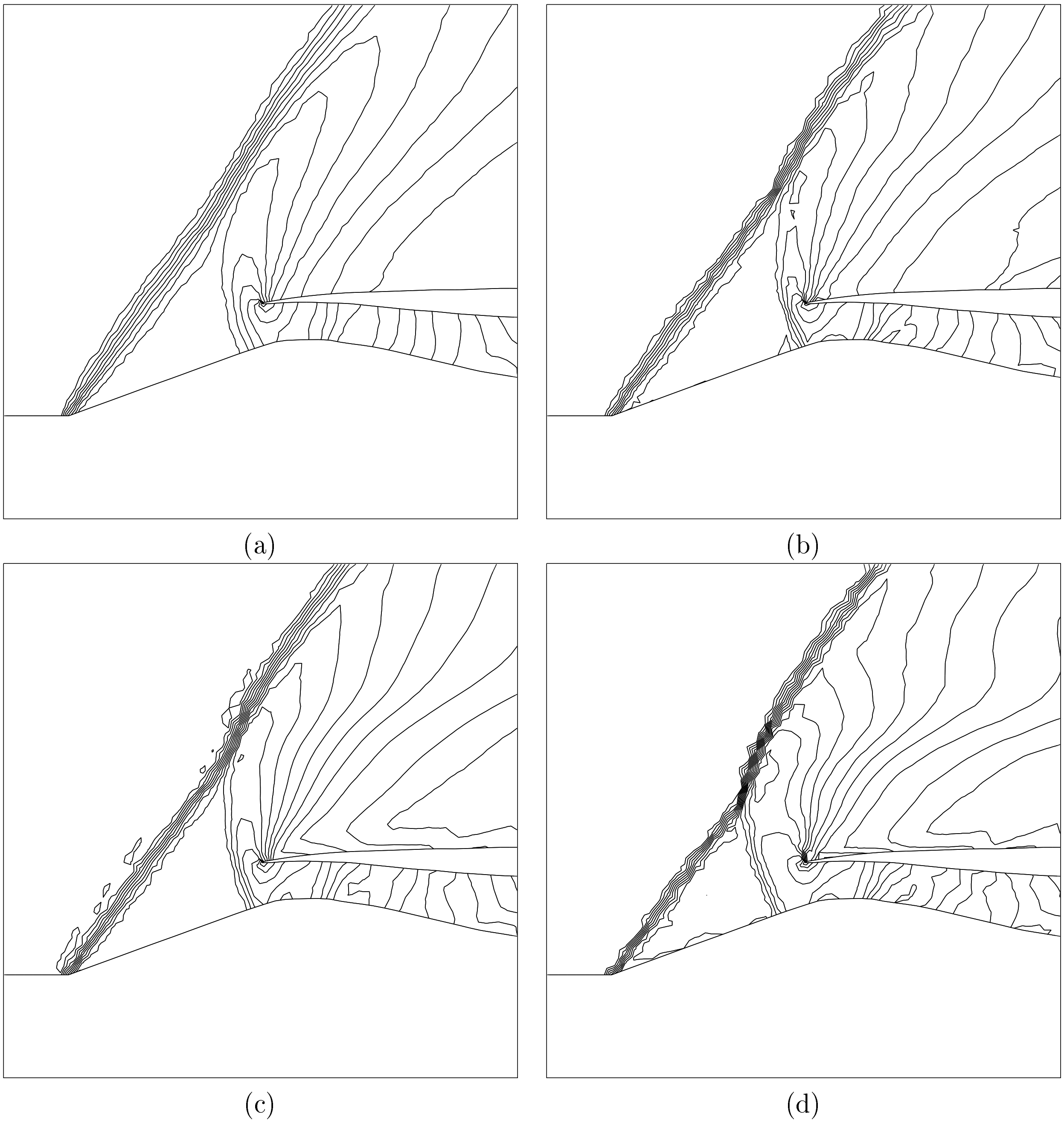


Figure 14: Zoom of the Mach number contours for the 1st order Roe scheme (a), the system N scheme (b), the MUSCL scheme (c) and the blended scheme (d).

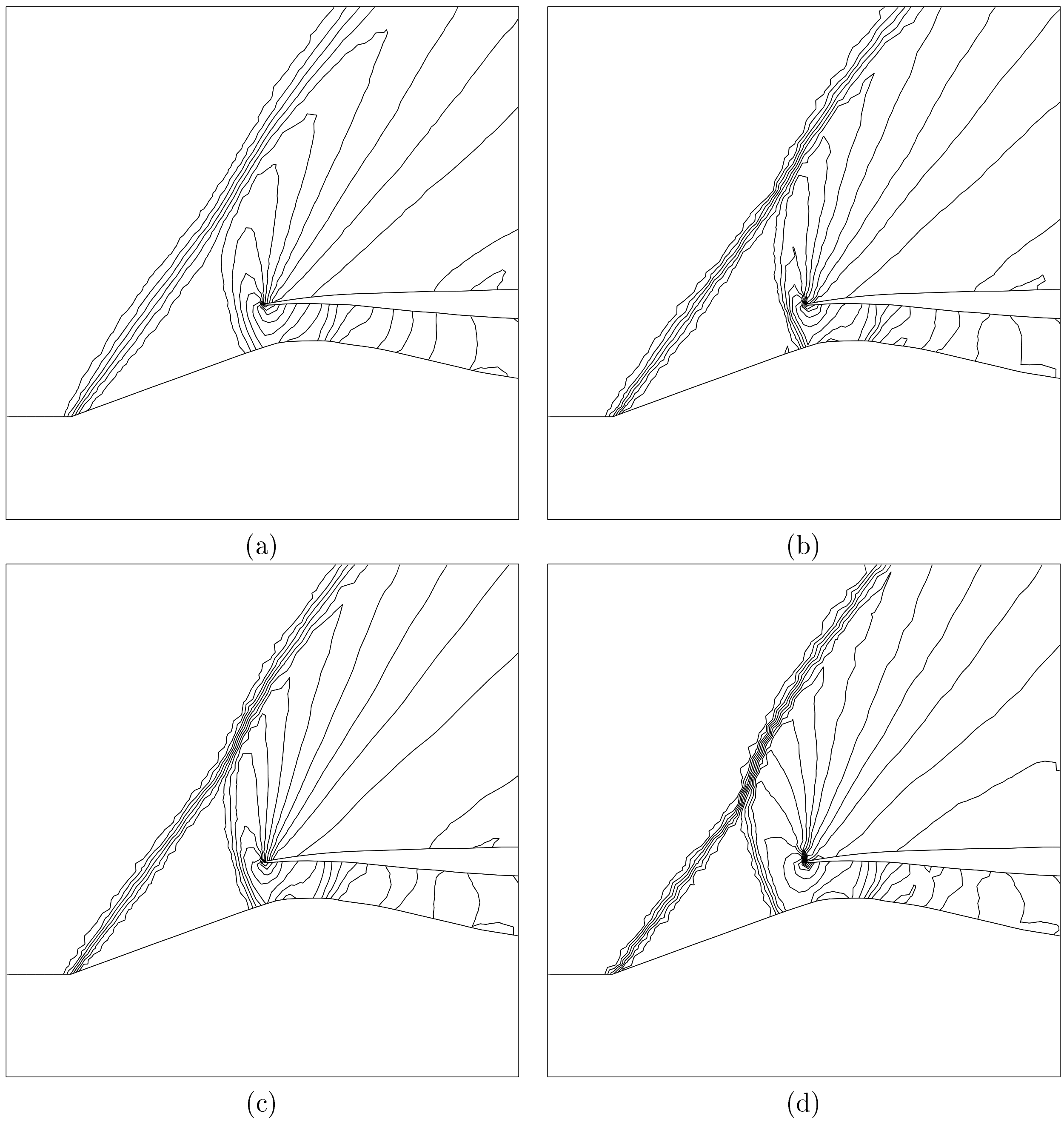


Figure 15: Zoom of the pressure contours for the 1st order Roe scheme (a), the system N scheme (b), the MUSCL scheme (c) and the blended scheme (d).

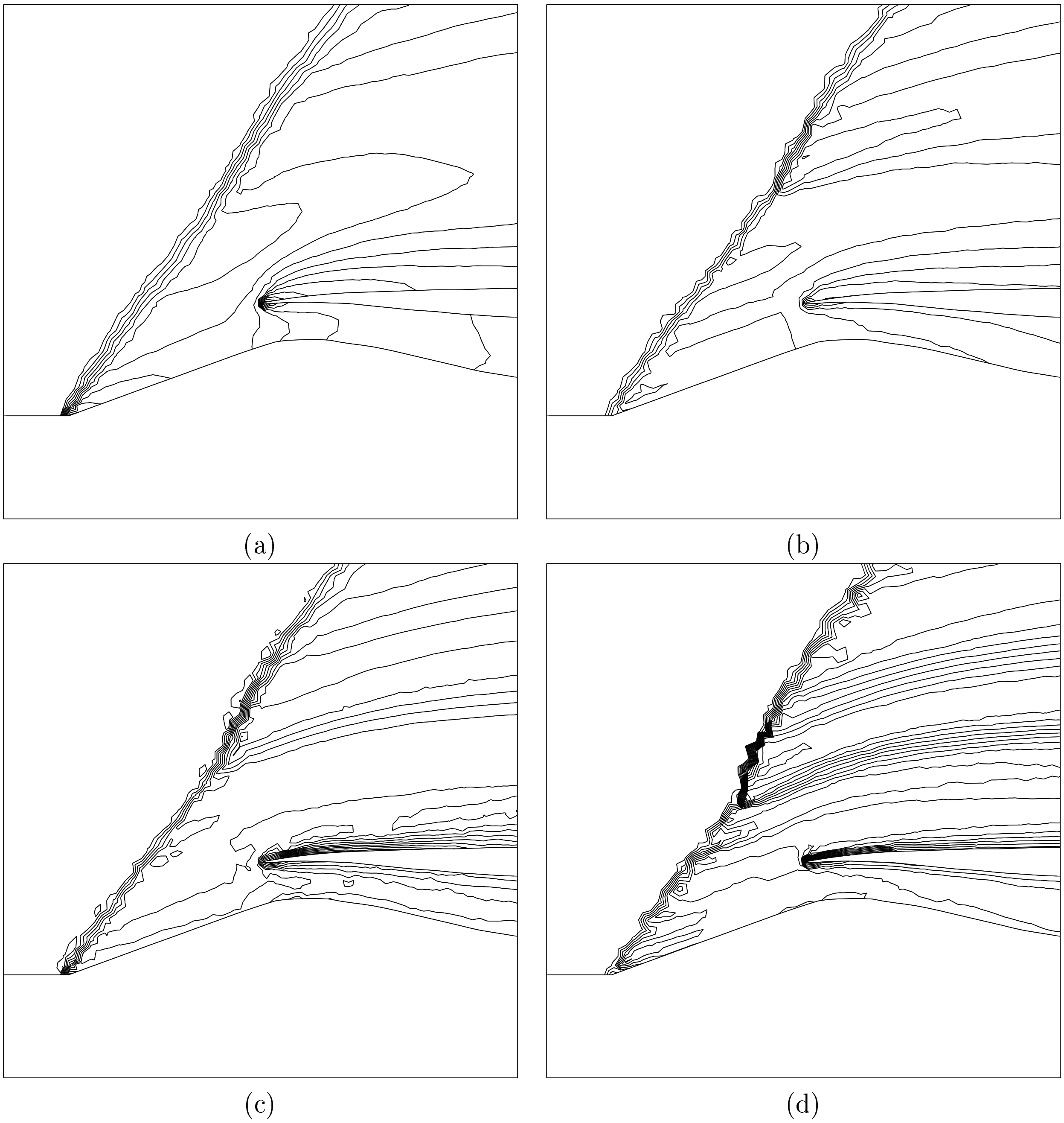


Figure 16: Zoom of the entropy deviation contours for the 1st order Roe scheme (a), the system N scheme (b), the MUSCL scheme (c) and the blended scheme (d).

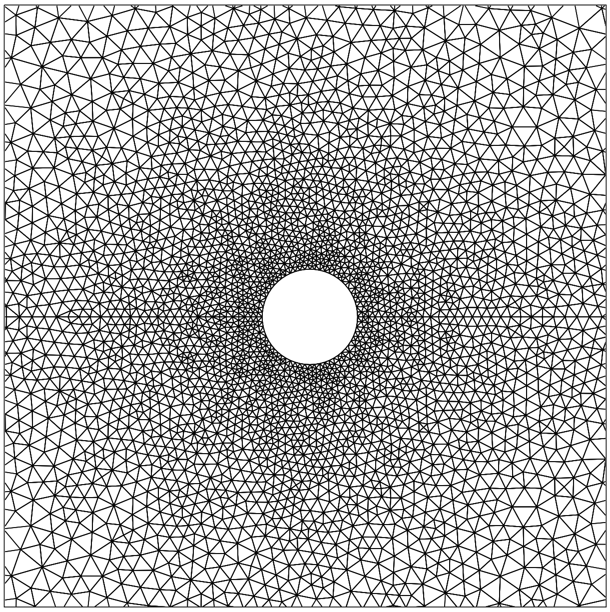
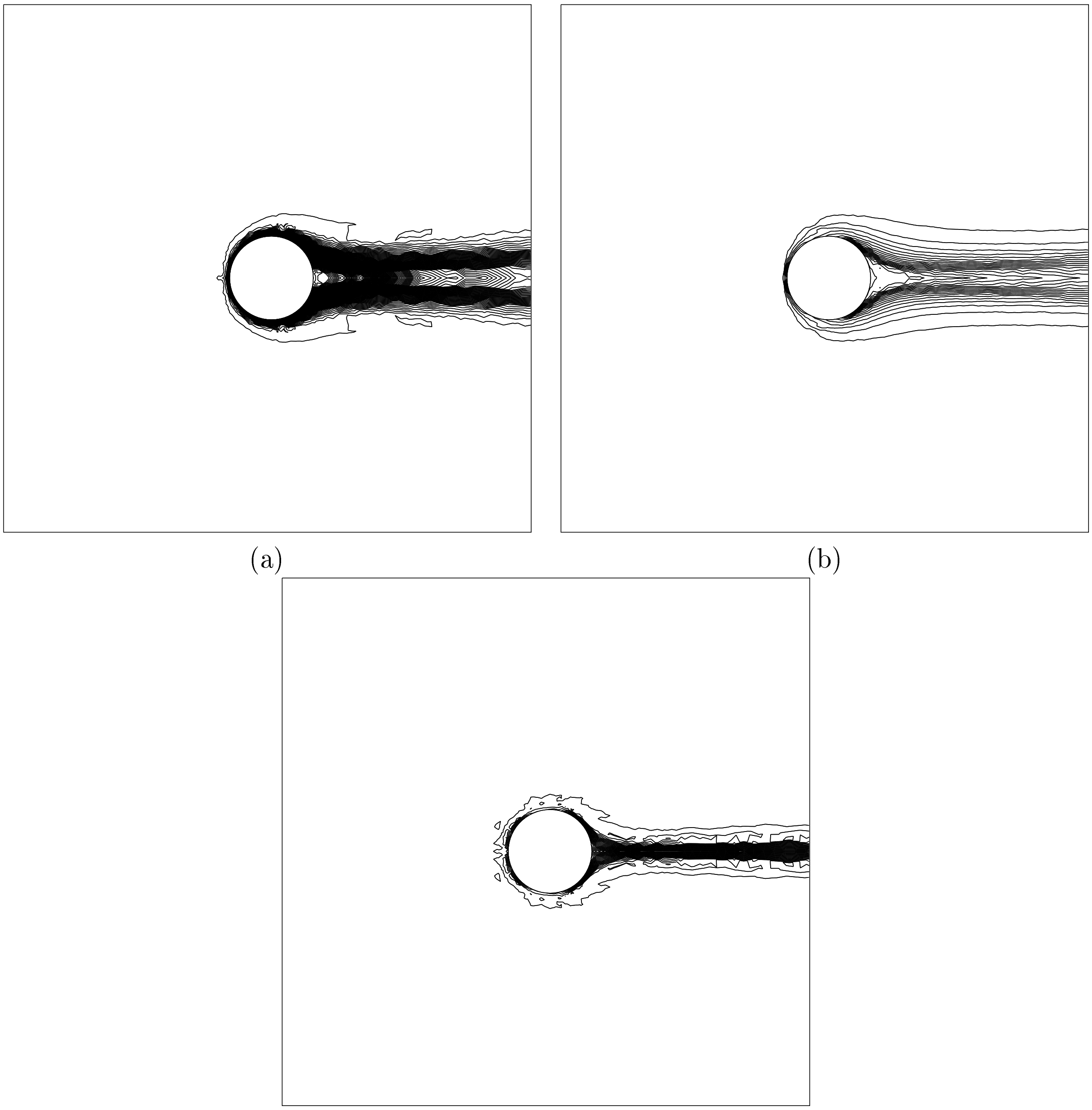


Figure 17: Mesh for the cylinder problem

Scheme	Mach (Min,Max)	Pressure (Min,Max)	Σ (Min,Max)
Muscl	0.0001, 0.82	0.67, 1.10	-0.0004 0.048
Blended	0.0001, 0.89	0.64, 1.11	0.0, 0.009
LDA	0.0, 0.94	0.62, 1.10	-0.001, 0.010

Table 3: Flow over a cylinder problem, Min and max of the mach number, pressure and entropy deviation Σ for the Muscl scheme, the blended scheme and the LDA scheme.



(a) (b)

(c)

Figure 18: Entropy deviation contours for the Muscl scheme (a), the blended scheme (b) and the LDA scheme (c).

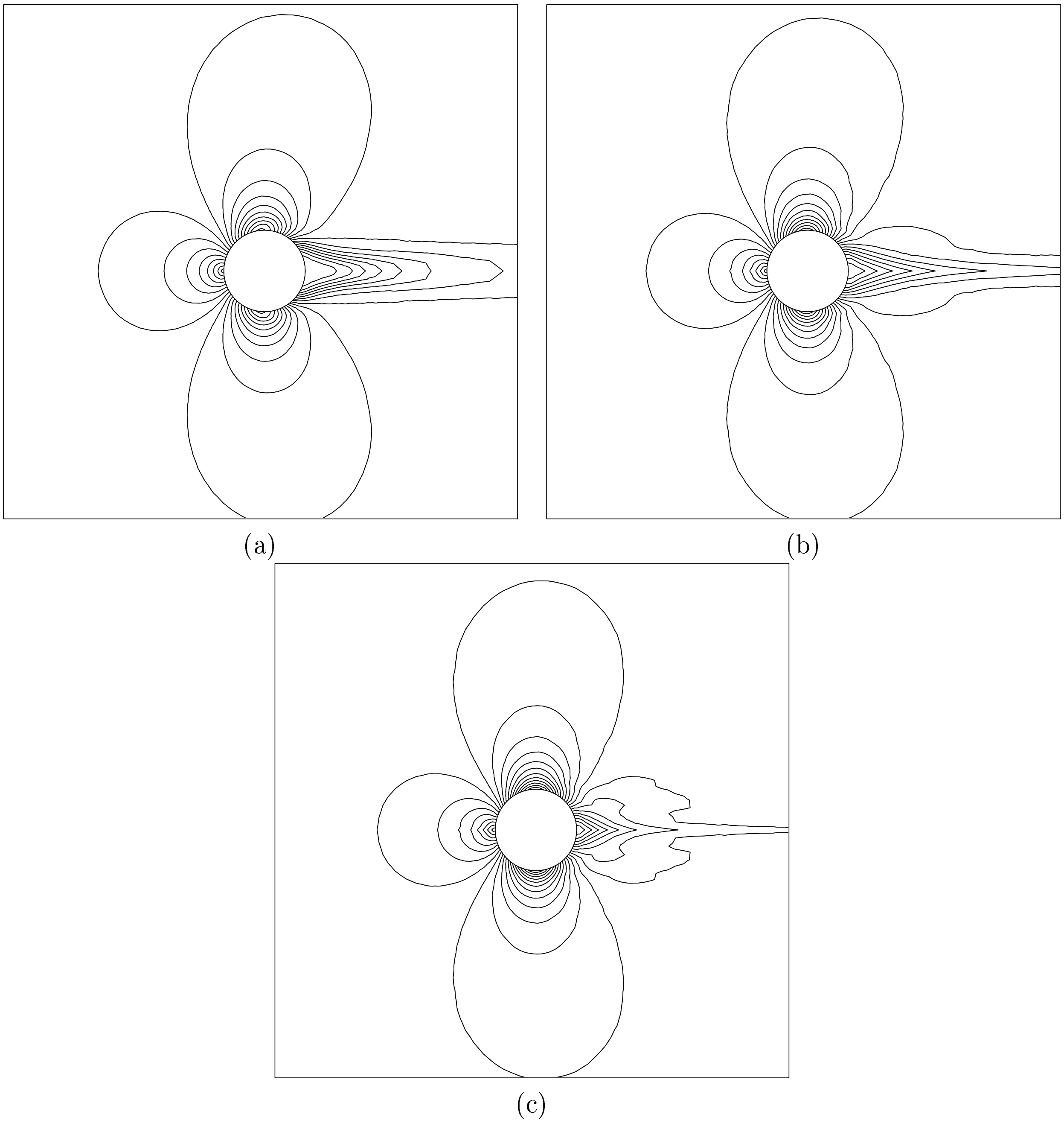


Figure 19: Mach number contours for the Muscl scheme (a), the blended scheme (b) and the LDA scheme (c).

Figure legends

Figure 1 : The dual cell is obtained by joining the midpoints of the edges starting from M_i and the centroids of the triangles containing M_i as a vertex.

Figure 2 : Geometrical elements for the finite volume scheme. The normal $\vec{n}_{ij}^1$ is orthogonal to $[G, I]$ with the same length. Same for $\vec{n}_{ij}^2$.

Figure 3 : Mesh for the shock tube problem

Figure 4 : Density isolines for the shock tube problem. Top : N scheme (left), Muscl scheme (right), Bottom : LDA scheme (left), Blended scheme (right). N scheme : $0.7 \leq \rho \leq 1.4$, Muscl : $0.69 \leq \rho \leq 1.40$, LDA : $0.61 \leq \rho \leq 1.42$, Blended : $0.70 \leq \rho \leq 1.40$

Figure 5 : Pressure isolines for the shock tube problem. Top : N scheme (left), Muscl scheme (right), Bottom : LDA scheme (left), Blended scheme (right). N scheme : $0.25 \leq p \leq 1.$, Muscl : $0.24 \leq p \leq 1.$, LDA : $0.2 \leq p \leq 1.04$, Blended : $0.25 \leq p \leq 1.$

Figure 6 : Mach number isolines for the shock tube problem. Top : N scheme (left), Muscl scheme (right), Bottom : LDA scheme (left), Blended scheme (right). N scheme : $2.4 \leq M \leq 4.$, Muscl : $2.4 \leq M \leq 4.07$, LDA : $2.39 \leq M \leq 4.27$, Blended : $2.4 \leq M \leq 4.0$

Figure 7 : Meshes for the Naca 0012 problem.

Figure 8 : Deviation of physical entropy ($\Sigma = \frac{s - s_\infty}{s_\infty}$) for the MUSCL scheme (a) and the blended scheme (b).

Figure 9 : Pressure coefficient for the MUSCL scheme (a) and the blended scheme (b).

Figure 10 : Mach number contours for the MUSCL scheme (a) and the blended scheme (b).

Figure 11 : Mach number contours on the regular mesh of Figure 7-b for the MUSCL scheme (a) and the blended scheme (b).

Figure 12 : Mach number contours for the 1st order Roe scheme (a), the system N scheme (b), the MUSCL scheme (c) and the blended scheme (d).

Figure 13 : Zoom of the mesh near the inlet entrance.

Figure 14 : Zoom of the Mach number contours for the 1st order Roe scheme (a), the system N scheme (b), the MUSCL scheme (c) and the blended scheme (d).

Figure 15 : Zoom of the pressure contours for the 1st order Roe scheme (a), the system N scheme (b), the MUSCL scheme (c) and the blended scheme (d).

Figure 16 : Zoom of the entropy deviation contours for the 1st order Roe scheme (a), the system N scheme (b), the MUSCL scheme (c) and the blended scheme (d).

Figure 17 : Mesh for the cylinder problem

Figure 18 : Entropy deviation contours for the Muscl scheme (a), the blended scheme (b)and the LDA scheme (c).

Figure 19 : Mach number contours for the Muscl scheme (a), the blended scheme (b)and the LDA scheme (c).