Explicit Runge-Kutta Residual Distribution schemes for Time Dependent Problems: second order case

Mario Ricchiuto, Remi Abgrall

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Explicit Runge-Kutta Residual Distribution

Mario Ricchiuto — Remi Abgrall

N° 6998

July 2009

Thème NUM

ISSN 0249-6399 ISRN INRIA/RR--6998--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
BORDEAUX - SUD OUEST

Explicit Runge-Kutta Residual Distribution

Mario Ricchiuto*, Remi Abgrall†

Thème NUM — Systèmes numériques
Équipes-Projets BACCHUS

Rapport de recherche n° 6998 — July 2009 — 55 pages

Abstract: In this paper we construct spatially consistent second order explicit discretizations for time dependent hyperbolic problems, starting from a given Residual Distribution (RD) discrete approximation of the steady operator. We explore the properties of the RD mass matrices necessary to achieve consistency in space, and finally show how to make use of second order mass lumping to obtain second order explicit schemes. The discussion is particularly relevant for schemes of the residual distribution type which we will use for all our numerical experiments. However, similar ideas can be used in the context of residual based finite volume discretizations.

Key-words: numerical analysis, second order schemes, hyperbolic problems, residual distribution, explicit schemes

* INRIA - Bordeaux Sud-Ouest
† INRIA - Bordeaux Sud-Ouest

Centre de recherche INRIA Bordeaux – Sud Ouest
Domaine Universitaire - 351, cours de la Libération 33405 Talence Cedex
Téléphone : +33 5 40 00 69 00

Explicit Runge-Kutta Residual Distribution schemes for Time Dependent Problems : second order case

Résumé : In this paper we construct spatially consistent second order explicit discretizations for time dependent hyperbolic problems, starting from a given Residual Distribution (RD) discrete approximation of the steady operator. We explore the properties of the RD mass matrices necessary to achieve consistency in space, and finally show how to make use of second order mass lumping to obtain second order explicit schemes. The discussion is particularly relevant for schemes of the residual distribution type which we will use for all our numerical experiments. However, similar ideas can be used in the context of residual based finite volume discretizations.

Mots-clés : numerical analysis, second order schemes, hyperbolic problems, residual distribution, explicit schemes

Contents

1 Introduction

The aim of this study is to understand, given a *residual based* discretization of steady limit of a hyperbolic conservation law, how to contruct fully explicit consistent discretizations for time dependent problems. As a case study we consider schemes of the residual distribution (RD) type [16, 3]. While well understood in the steady case, their formulation in the time dependent case has never been completely clarified, due to the lack of a rigorous formulation allowing a natural extension. In particular, the lack of sufficient constraints on the discretization has led in time to a number of different formulations all featuring different mass matrices [8, 20, 17, 18]. In this paper we show that all these formulations are equivalent, up to a dissipation operator. We also show the existence of entire families of additional consistent mass matrices.

Finally, we show how to combine all of the above discretization with high order (second order) mass-lumping to obtain fully explicit schemes. Note that in the case of nonlinear RD discretizations based on second order time integration in time, positivity preservation is obtained only under an explicit CFL-type condition [4, 16]. This fact, related to the properties of the underlying ODE

integrator [7], and the highly implicit nature of the schemes leads to poor efficiency. The explicit formulation propose here is one possible solution to this issue.

The work discussed here somehow generalizes the initial work of [32] where only central Lax-Wendroff type discretizations are considered. Moreover, in the paper we will show that the ideas presented here also apply to other classes of schemes, such as the ones proposed in [12].

The structure of the paper is as follows. We start by introducing the notations used throughout the discussion. Then, in section §3 we review different formulations of RD for time dependent problems. We discuss their relations, and show how other families of consistent formulations exist. We elaborate further on these ideas in §4 where we finally show how to obtain fully explicit formulations which still retain the same formal accuracy of the fully implicit formulations. In section §5 we give a summary of the different numerical schemes that we use in the numerical tests which are discussed in sections §6 and §7. We end the paper with some conclusive remarks and some thoughts for further developments.

2 Mathematical problem and notation

We seek approximations of solutions of the time dependent hyperbolic problem

$$r(u) = 0\,, \qquad r(u) = \partial_t u + \nabla \cdot \boldsymbol{\mathcal{F}}(u) \tag{1}$$

on some spatial domain Ω, and on some temporal domain $[0, t_f]$. We will mainly focus on the two-dimensional case $\Omega \in \mathbb{R}^2$, but the generalization to three spatial dimension is trivial.

We discretize Ω by an unstructured triangulation denoted by $\mathcal{T}_h$, with T denoting the generic element of the mesh, and h the mesh parameter (characteristic mesh size). When no confusion is generated we denote the nodes of T by $\{1, 2, 3\}$. In every element, we denote by $\vec{n}_j$ the inward pointing vector normal to the edge facing node j, scaled by the length of the edge. Denoting by φ_i the P^1 Lagrange basis function corresponding ot node $i \in \mathcal{T}_h$, we have

$$\nabla\varphi_i\big|_T = \frac{\vec{n}_i}{2|T|} \tag{2}$$

The P^1 approximation of u will be denoted by u_h, and it is given by

$$u_h = \sum_{i\in\mathcal{T}_h} u_i\varphi_i = \sum_{T\in\mathcal{T}_h}\sum_{j\in T} u_j\varphi_j\big|_T \tag{3}$$

The temporal domain is discretized by a set of non-overlapping time slabs $[t^n, t^{n+1}]$. We denote by $\Delta t = t^{n+1} - t^n$ the time step.

To simplify the presentation of the next sections we also introduce here the *element fluctuation* defined as

$$\phi(u_h) = \int_T \nabla \cdot \boldsymbol{\mathcal{F}}_h(u_h)\; dx\, dy = \oint_{\partial T} \boldsymbol{\mathcal{F}}_h(u_h)\cdot \hat{n}\; dl\;, \tag{4}$$

the *element residual*

$$\Phi(u_h) = \int_T r(u_h)\, dx\, dy = \int_T (\partial_t u_h + \nabla \cdot \boldsymbol{\mathcal{F}}_h(u_h))\, dx\, dy = \sum_{j\in T} \frac{|T|}{3}\frac{du_j}{dt} + \phi(u_h) \tag{5}$$

and the local Galerkin residual

$$\phi_i^{\mathrm{G}}(u_h) = \int_T \varphi_i \nabla \cdot \boldsymbol{\mathcal{F}}_h(u_h)\, dx\, dy \tag{6}$$

In the expressions above u_h represents the P^1 numerical approximation of the unknown, and $\boldsymbol{\mathcal{F}}_h(u_h)$ a discrete approximation of the flux. Note that all of the above quantities depend on time. Moreover, to simplify the notation, we do not introduce a super- or sub-script indicating the element T over which they are evaluated, this being always clear from the context.

3 Second order RD : the proliferation of mass matrices

Let us for the moment consider the particular case of (1) given by the linear constant advection problem

$$\partial_t u + \vec{a} \cdot \nabla u = 0 \tag{7}$$

We focus our attention on discrete counterparts of (7) that, on a slab $\mathcal{T}_h \times [t^n, t^{n+1}]$ can be written as

$$\sum_{T|i\in T} \left\{ \sum_{j\in T} m_{ij}^T \frac{du_j}{dt} + \beta_i \phi(u_h) \right\} = 0 \quad \forall\, i \in \mathcal{T}_h \tag{8}$$

Last definitions give a scheme that, requires the solution of a (generally) nonlinear system if the *mass matrix* m_{ij} is non-diagonal. Moreover, introducing the *nodal residuals*

$$\Phi_i(u_h) = \sum_{j\in T} m_{ij}^T \frac{du_j}{dt} + \beta_i \phi(u_h) \tag{9}$$

we also require that,

$$\sum_{j\in T} \Phi_j(u_h) = \Phi(u_h) \tag{10}$$

with $\Phi(u_h)$ given by (5).

The prototype (8) is meant to be a consistent generalization to the time dependent case of the *fluctuation splitting/residual distribution* discretization which approximates the steady limit of (7) as

$$\sum_{T|i\in T} \beta_i \phi(u_h) = 0 \quad \forall\, i \in \mathcal{T}_h \tag{11}$$

with

$$\sum_{j\in T} \beta_j = 1 \tag{12}$$

In order to distinguish the steady advective operator from the time dependent equation, we have chosen to keep a distinction between the fluctuation (4) and the residual (5), the latter representing the integral of the whole equation.

Historically, the first consistent approaches to obtain such a generalization were based on two different points of view. In the first approach [9, 8], one simply replaces the fluctuations in the discrete equations (11) with the full residual (5). This residual is then distributed exactly as in (11) (with the LDA scheme in the original reference [9, 8, 16]), leading to

$$0 = \sum_{T|i\in T} \beta_i \Phi(u_h) = \sum_{T|i\in T} \left(\sum_{j\in T} m_{ij}^{\mathrm{F1}} \frac{du_j}{dt} + \beta_i \phi(u_h) \right), \quad m_{ij}^{\mathrm{F1}} = \frac{|T|}{3}\beta_i \tag{13}$$

with δ_{ij} Kroenecker's delta, and F1 standing for Formulation 1.

A second approach [26, 20] uses an analogy with stabilized Galerkin finite element schemes in which the discrete equations (11) are obtained as

$$\sum_{T|i\in T} \beta_i \phi(u_h) = \sum_{T|i\in T} \phi_i^{\mathrm{G}}(u_h) + \sum_{T|i\in T} \delta\phi_i = \int_\Omega \varphi_i \vec{a}\cdot\nabla u_h \, dx\, dy + \sum_{T|i\in T} \int_T \delta_{\varphi_i} \vec{a}\cdot\nabla u_h \, dx\, dy$$

with the perturbation to the test function δ_{φ_i} depending on the distribution coefficients β_i^T. In particular, for constant advection and a P^1 variable approximation, one can assume δ_{φ_i} to be a constant, to find easily (in two space dimensions) $\delta_{\varphi_i}|_T = \beta_i - 1/3$. In the time dependent case this naturally leads to

$$\begin{aligned} 0 &= \int_\Omega \varphi_i \, r(u_h) \, dx\, dy + \sum_{T|i\in T} \int_T \delta_{\varphi_i} \, r(u_h) \, dx\, dy \\ &= \sum_{T|i\in T} \left(\sum_{j\in T} m_{ij}^{\mathrm{F2}} \frac{du_j}{dt} + \beta_i \phi(u_h) \right), \quad m_{ij}^{\mathrm{F2}} = \frac{|T|}{36} \left(3\,\delta_{ij} + 12\beta_i - 1 \right) \end{aligned} \tag{14}$$

with δ_{ij} Kroenecker's delta, and F2 standing for Formulation 2.

A different approach has instead been proposed in [18], where the authors use the idea that in every $T \in \mathcal{T}_h$ the combination of terms arising from the multiplication of the mass matrix with the nodal time derivatives should give back an integral of the time derivative of u_h over a dual sub-element $C_j \in T$. Consistency is guaranteed by the requirement $|C_j| = \beta_j |T|$[1]. In particular, in the paper the authors require *the node j to belong to C_j* (cf. figure 1). Conditions for second order of accuracy are shown to be

$$\sum_{i\in T} m_{ij} = \frac{|T|}{3}, \qquad \sum_{j\in T} m_{ij} = |T|\beta_i \tag{15}$$

In the reference, the authors ultimately arrive to the following formulation

$$0 = \sum_{T|i\in T} \int_{T_i} r(u_h) \, dx\, dy = \sum_{T|i\in T} \left(\sum_{j\in T} m_{ij}^{\mathrm{F3}} \frac{du_j}{dt} + \beta_i \phi(u_h) \right), \quad m_{ij}^{\mathrm{F3}} = \frac{|T|}{3}\beta_i \left(\delta_{ij} + 1 - \beta_j \right) \tag{16}$$

[1] which implicitly assumes $\beta_i \geq 0 \; \forall i$

with δ_{ij} Kroenecker's delta, and F3 standing for Formulation 3. This formulation is actually proposed mainly for schemes with a multidimensional upwind characher [16], and in particular for those schemes that, when the advection speed points toward an edge of element T, give $\beta_j \geq 0$ only if j belongs to this edge. In this case, only the rows of m_{ij}^{F3} relative to these nodes contain non-zero elements.

The idea of [18] can actually be used to derive still another member to the family of consistent mass matrices. It suffices to follow the exact same developments done in the reference, except that we allow the sub-triangle j *not to contain node* j *itself*. In particular, whenever $\beta_i \geq 0 \; \forall \, i$, we note that we can find a unique point, say $M \in T$, such that $\varphi_i(M) = \beta_i \; \forall \, i$. This means that the β_i coefficients represent the area coordinates of M (see figure 1). With the notation of figure 1, using the fact that $u_h(M) = \beta_1 \, u_1 + \beta_2 \, u_2 + \beta_3 \, u_3$ we find in the time dependent case :

$$0 = \sum_{T|i\in T} \int_{T_i} r(u_h) \, dx \, dy = \sum_{T|i\in T} \left(\sum_{j\in T} m_{ij}^{\mathrm{F4}} \frac{du_j}{dt} + \beta_i^T \, \phi(u_h) \right), \quad m_{ij}^{\mathrm{F4}} = \frac{|T|}{3} \beta_i \, (1 - \delta_{ij} + \beta_j) \tag{17}$$

with δ_{ij} Kroenecker's delta, and F4 standing for Formulation 4. Note that the difference with respect to the matrix proposed in [18] is that here $j \notin T_j$. One easily checks that conditions (15) are verified.

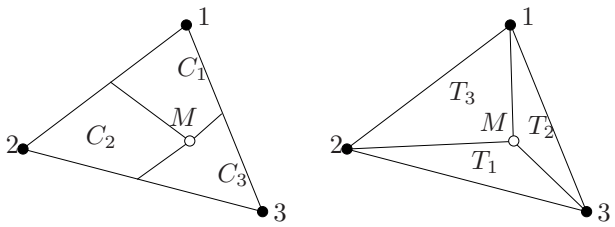


Figure 1: Left. Formulation 3 : dual areas C_j, $|C_j| = \beta_j |T|$. Right. Formulation 4 : area coordinates of the distribution point M ; $|T|_j = \beta_j |T|$

All of the above construction can be still generalized by noting that, given the distribution coefficients β_i^T, the only constranits available are given by (15). These constraints actually correspond to the conservation requirement (10), and to the requirement that, whenever the differential operator

$$r(u_h) = \partial_t u_h + \vec{a} \cdot \nabla u_h$$

is (as in the steady case) locally constant, say $r(u_h)\big|_T = r^T$, then

$$\Phi_i = \beta_i^T \, |T| \, r^T$$

These properties are always verified if we can find a Petrov-Galerkin test function ω_i such that on every T

$$\Phi_i(u_h) = \int_T \omega_i \, r(u_h) \, dx \, dy \tag{18}$$

with

$$\sum_{j\in T}\omega_j = 1\ , \quad \frac{1}{|T|}\int_T \omega_i\ dx\,dy = \beta_i^T \tag{19}$$

As we will show later, the formal second order of accuracy is guaranteed as long as ω_i is locally bounded. Note also that

$$m_{ij}^T = \int_T \omega_i\ \varphi_j\ dx\,dy \tag{20}$$

The number of functions that verify these constraints is infinite. For example, to obtain the formulation F1 (cf. equation (13)) one can choose on each T

$$\omega_i^{\rm F1}\Big|_T = \beta_i^T \chi_T$$

having denoted by χ_T the characteristic function

$$\chi_T(x,y) = \left\{ \begin{array}{ll} 1 & \text{if } (x,y)\in T \\ 0 & \text{if } (x,y)\notin T \end{array}\right.$$

Conversely, formulation F2 (cf. equation (14)) is obtained for

$$\omega_i^{\rm F2} = \varphi_i + \sum_{T|i\in T} \delta_{\varphi_i}\chi_T$$

Formulations F3 and F4 are instead obtained by taking for example (cf. equations (16) and (17))

$$\omega_i^{\rm F3/F4}\Big|_T = \chi_{T_i}$$

Moreover, for any given test function ω_i verifying all the consistency, conservation, and accuracy constraints, we can easily come up with a modified function, say $\widetilde{\omega}_i$ with all the desirable properties. For example, if we can find three bounded functions, say f_1, f_2, and f_3 such that

$$\sum_{j=1}^3 f_j = C_f$$

with C_f a constant, we can modify ω_i as follows

$$\overline{\omega}_i = \omega_i + K(f_i - \overline{f}_i)\ , \quad \overline{f}_i = \frac{1}{|T|}\int_T f_i\ dx\,dy$$

with K an arbitrary parameter ! Note that this term does not affect consistency or conservation, due to the fact that (using for the nodes of T the local renumbering $\{i,j,k\}\to\{1,2,3\}$)

$$\sum_{j=1}^3 \left(f_j - \overline{f}_j\right) = 0\ , \quad \int_T \left(f_j - \overline{f}_j\right) dx\,dy = 0$$

nor it does pollute the accuracy of the discretization, as long as the each f_i is bounded. Moreover, in the P^1 case ∇u_h is constant per element, so that

$$\int_T \left(f_j - \overline{f}_j\right) \vec{a} \cdot \nabla u_h \, dx \, dy = 0$$

so that the extra term only affects the form of the mass matrix. This leads clearly to quite a large number of consistent mass matrices, and extra constraints are needed to make sure one does the right thing. For example stability is not at all taken into account in the analysis performed so far.

The last observation leads to very interesting consequences if we take $f_i = \varphi_i$. In this case we have

$$\overline{\varphi}_i = \frac{1}{|T|} \int_T \varphi_i = \frac{1}{3}$$

So any mass matrix can be modified as

$$\overline{m}_{ij}^T = m_{ij}^T + K \int_T (\varphi_i - \overline{\varphi}_i)\varphi_j \, dx \, dy = \int_T \omega_i \, \varphi_j \, dx \, dy + K \int_T (\varphi_i - \overline{\varphi}_i)\varphi_j \, dx \, dy$$

that leads to the semi-discrete scheme

$$\sum_{T|i\in T} \left(\sum_{j\in T} \left(m_{ij}^T + K\delta m_{ij}\right) \frac{du_j}{dt} + \beta_i \phi(u_h) \right) = 0 \, , \quad \delta m_{ij} = \frac{|T|}{36} \left(3\delta_{ij} - 1\right) \tag{21}$$

with δ_{ij} Kroenecker's delta. As already noted in [29], the matrix δm_{ij} is symmetric, and defines a dissipation operator, that is

$$v^T \left[\delta m_{ij}\right] v \geq 0 \, , \quad \forall \, v \in \mathbb{R}^3$$

In the last reference, this term has been used to provide further stabilization to a nonlinear second-order variant of a Lax-Friedrich's scheme. The interesting observation is that if we take $K = 3$ and apply the modification to the Galerkin scheme we obtain :

$$\overline{m}_{ij} = \overbrace{\frac{|T|}{12}(\delta_{ij} + 1)}^{\text{Galerkin}} + \overbrace{\frac{|T|}{12}(3\delta_{ij} - 1)}^{3\delta m_{ij}} = \frac{|T|}{3}\delta_{ij}$$

Which is just another way to show that mass lumping for the Galerkin scheme does not reduce the accuracy in the P^1 case but it does introduce a degree of dissipation.

It is also worth noting that comparing (13) and (14) with (21), one immediately sees that

$$m_{ij}^{\mathrm{F2}} = m_{ij}^{\mathrm{F1}} + \delta m_{ij} \tag{22}$$

The third and first formulations are linked by a very similar relation :

$$m_{ij}^{\mathrm{F3}} = m_{ij}^{\mathrm{F1}} + \widetilde{\delta m}_{ij} \, , \qquad \widetilde{\delta m}_{ij} = \frac{|T|}{3} \left(\beta_i \delta_{ij} - \beta_i \beta_j^T\right) \tag{23}$$

Provided that $\beta_i \geq 0 \; \forall \, i$, then the symmetric matrix $\widetilde{\delta m_{ij}}$ also defines a dissipation operator. In particular, $\forall \, v \in \mathbb{R}^3$ we have

$$v^T \left[\widetilde{\delta m_{ij}}\right] v = \frac{|T|}{3}\beta_1^T\beta_2^T(v_1-v_2)^2 + \frac{|T|}{3}\beta_1^T\beta_3^T(v_1-v_3)^2 + \frac{|T|}{3}\beta_3^T\beta_2^T(v_3-v_2)^2 \geq 0$$

A similar relation holds for the last formulation, only this time we have

$$m_{ij}^{\mathrm{F1}} = m_{ij}^{\mathrm{F4}} + \widetilde{\delta m_{ij}} \tag{24}$$

This last relation allows finally to show that *all the formulations are equivalent up to a dissipation term* :

$$\begin{aligned} m_{ij}^{\mathrm{F1}} &= m_{ij}^{\mathrm{F4}} + \widetilde{\delta m_{ij}} \\ m_{ij}^{\mathrm{F2}} &= m_{ij}^{\mathrm{F1}} + \delta m_{ij} = m_{ij}^{\mathrm{F4}} + \widetilde{\delta m_{ij}} + \delta m_{ij} \\ m_{ij}^{\mathrm{F3}} &= m_{ij}^{\mathrm{F1}} + \widetilde{\delta m_{ij}} = m_{ij}^{\mathrm{F4}} + 2\widetilde{\delta m_{ij}} \end{aligned} \tag{25}$$

Remark 3.1. *The last relations show that, for a fixed approximation of the advection operator, the formulation F4 is the least dissipative of all.*

Remark 3.2. *In one space dimension, if the spatial discretization is given by the classical 1d upwind scheme, the formulations F1, F3, and F4 become identical. The formulation F2, instead, reduces to the 1D SUPG scheme obtained by defining the* τ *SUPG parameter as [25, 34, 22]*

$$\tau = \frac{\Delta \, x}{2|a|}$$

The objective of the following sections is to show how to make use of the properties discussed above to contruct fully explicit residual based schemes.

4 Mass lumping and bubble stabilization

In this section we make use of the elements already discussed to propose a fully explicit variant of the residual distribution schemes described in the previous sections. The start with we make the assumption that, however complex the definition of the β_i coefficients and of the mass matrix, there exists a uniformly bounded and locally differentiable function γ_i, such that we can rewrite the discretization (8) as

$$\int_\Omega \varphi_i \left(\partial_t u_h + \vec{a}\cdot\nabla u_h\right) dx\, dy + \sum_{T|i\in T} \int_T \gamma_i \left(\partial_t u_h + \vec{a}\cdot\nabla u_h\right) dx\, dy = 0 \tag{26}$$

where γ_i plays the role of a "stabilizing" bubble function, and satisfies

$$\begin{cases} \sum\limits_{j\in T} \gamma_j = 0 \\ \\ \int\limits_T (\varphi_i + \gamma_i)\partial_t u_h \; dx\, dy = \sum\limits_{j\in T} m_{ij}\dfrac{du_j}{dt} \\ \\ \dfrac{1}{|T|}\int\limits_T (\varphi_i+\gamma_i) dx\, dy = \beta_i \end{cases} \tag{27}$$

The last three relations guarantee the satisfaction of the conservation property, and of the consistency with the (given) spatial discretization, so that ultimately (cf. equation (9))

$$\Phi_i(u_h) = \sum_{j\in T} m_{ij}\frac{du_j}{dt} + \beta_i\phi(u_h) = \int_T \varphi_i\, r(u_h)\, dx\, dy + \int_T \gamma_i\, r(u_h)\, dx\, dy \quad (28)$$

The role of the bubble is of course that of providing stabilization to the otherwise unstable Galerkin scheme. *Note that it is not necesary to actually show particular forms of such function which, as we shall see in the following, is just an artifact allowing to analyze the accuracy of the schemes proposed in the paper.* Nevertheless, whenever we can exhibit the existence of a Petrov-Galerkin test function ω_i such that (18) holds, we can simply set

$$\gamma_i|_T = \omega_i|_T - \varphi_i|_T \quad (29)$$

For example, for the formulations seen in the previous sections we have :

$$\begin{aligned} \gamma_i^{\mathrm{F1}}\big|_T &= \beta_i - \varphi_i|_T \ ; \\ \gamma_i^{\mathrm{F2}}\big|_T &= \delta_{\varphi_i} = \beta_i - \frac{1}{3}\,; \\ \gamma_i^{\mathrm{F3/F4}}\Big|_T &= \chi_{T_i} - \varphi_i|_T \ . \end{aligned} \quad (30)$$

4.1 Construction of explicit schemes

The starting point of the contruction is to choose an explicit time-stepping scheme. We will focus here on Runge-Kutta (RK) schemes, however other possibilities exist, and will be studied in the future. Let us denote by $\delta u^k = u^k - u^n$ the increment of the k-th step of a given explicit RK scheme. Similarly, let f^k be the k-th step evolution operator so that for the problem

$$\frac{du}{dt} + f(u) = 0$$

we can rewrite each RK step as

$$\frac{\delta u^k}{\Delta t} + f^k = 0$$

In particular, in the following we will denote by r^k the quantity

$$r^k = \frac{\delta u^k}{\Delta t} + f^k \quad (31)$$

For example for the classical TVD RK2 scheme we have

$$\begin{cases} r^1 = \dfrac{\delta u^1}{\Delta t} + f^1 = 0\,, \qquad f^1 = \vec{a}\cdot\nabla u^n \\ \\ r^2 = \dfrac{\delta u^2}{\Delta t} + f^2 = 0\,, \qquad f^2 = \dfrac{1}{2}\vec{a}\cdot\nabla u^n + \dfrac{1}{2}\vec{a}\cdot\nabla u^1 \end{cases} \quad (32)$$

Similarly, the TVD RK3 scheme gives

$$\left\{\begin{array}{ll} r^1 = \dfrac{\delta u^1}{\Delta t} + f^1 = 0\,, & f^1 = \vec{a}\cdot\nabla u^n \\ r^2 = \dfrac{\delta u^2}{\Delta t} + f^2 = 0\,, & f^2 = \dfrac{1}{4}\vec{a}\cdot\nabla u^n + \dfrac{1}{4}\vec{a}\cdot\nabla u^1 \\ r^3 = \dfrac{\delta u^3}{\Delta t} + f^3 = 0\,, & f^3 = \dfrac{1}{6}\vec{a}\cdot\nabla u^n + \dfrac{1}{6}\vec{a}\cdot\nabla u^1 + \dfrac{2}{3}\vec{a}\cdot\nabla u^2 \end{array}\right. \tag{33}$$

With this notation we can write the k-th step of the RK time integrator as

$$r^k = 0$$

Its Galerkin discretization writes

$$\int_\Omega \varphi_i r^k(u_h)\,dx\,dy = \int_\Omega \varphi_i \frac{\delta u_h^k}{\Delta t} dx\,dy + \int_\Omega \varphi_i f^k(u_h)\,dx\,dy = 0$$

The next step is to add the contribution of the bubble. The standard approach would be to write this contribution as

$$\sum_{T|i\in T}\int_T \gamma_i\, r^k(u_h)\,dx\,dy = \sum_{T|i\in T}\int_T \gamma_i \left(\frac{\delta u_h^k}{\Delta t} + f^k(u_h)\right) dx\,dy\,,$$

however, even when lumping the Galerkin component of the mass matrix, this would lead to a scheme with a non-diagonal mass matrix, still requiring the solution of an *a-priori* nonlinear system at each RK step. What we propose is to replace *only in the bubble contribution* the k-th step residual $r^k(u_h)$ by a modified residual $\overline{r}^k(u_h)$, which makes use of a different approximation of the time derivative. In practice, we will look for $\overline{r}^k(u_h)$s differing from $r^k(u_h)$ only in the definition of the time increment, that is

$$\overline{r}^k(u_h) = \overline{\frac{\delta u^k}{\Delta t}} + f^k \tag{34}$$

The constraints on $\overline{r}^k(u_h)$ guaranteeing that the overall accuracy of the discretization is not deteriorated will be given in the next section. For the moment

we observe that when adding this contribution we obtain :

$$
\begin{aligned}
0 &= \int_\Omega \varphi_i \, r^k(u_h) \, dx\, dy + \sum_{T|i\in T} \int_T \gamma_i \, \overline{r}^k(u_h) \, dx\, dy \\
&= \int_\Omega \varphi_i \frac{\delta u_h^k}{\Delta t} dx\, dy + \int_\Omega \varphi_i \, f^k(u_h) \, dx\, dy + \sum_{T|i\in T} \int_T \gamma_i \left(\overline{\frac{\delta u_h^k}{\Delta t}} + f^k(u_h) \right) dx\, dy \\
&= \int_\Omega \varphi_i \frac{\delta u_h^k}{\Delta t} dx\, dy + \int_\Omega \varphi_i \left(\overline{\frac{\delta u_h^k}{\Delta t}} + f^k(u_h) \right) dx\, dy + \sum_{T|i\in T} \int_T \gamma_i \left(\overline{\frac{\delta u_h^k}{\Delta t}} + f^k(u_h) \right) dx\, dy - \int_\Omega \varphi_i \overline{\frac{\delta u_h^k}{\Delta t}} dx\, dy \\
&= \int_\Omega \varphi_i \frac{\delta u_h^k}{\Delta t} + \sum_{T|i\in T} \int_T (\varphi_i + \gamma_i) \left(\overline{\frac{\delta u_h^k}{\Delta t}} + f^k(u_h) \right) dx\, dy - \int_\Omega \varphi_i \overline{\frac{\delta u_h^k}{\Delta t}} dx\, dy \\
&= \int_\Omega \varphi_i \frac{\delta u_h^k}{\Delta t} + \sum_{T|i\in T} \Phi_i^{\mathrm{RK}(k)} - \int_\Omega \varphi_i \overline{\frac{\delta u_h^k}{\Delta t}} dx\, dy
\end{aligned}
\tag{35}
$$

The last relations are obtained by first adding and substracting the Galerkin integral of the approximate time increment $\overline{\delta u^k}/\Delta t$, then using the properties of the bubble function γ_i (cf. equations (9) and (30)), and finally introducing the fully discrete split residuals :

$$
\Phi_i^{\mathrm{RK}(k)} = \int_T (\varphi_i + \gamma_i) \overline{r}^k(u_h) \, dx\, dy = \sum_{j\in T} m_{ij} \overline{\frac{\delta u_j^k}{\Delta t}} + \beta_i \phi^{\mathrm{RK}(k)}(u_h) \tag{36}
$$

with

$$
\phi^{\mathrm{RK}(k)} = \int_T f^k(u_h) \, dx\, dy \quad \text{and} \quad \sum_{j\in T} \Phi_j^{\mathrm{RK}(k)} = \int_T \overline{r}^h(u_h) \, dx\, dy = \Phi^{\mathrm{RK}(k)} \tag{37}
$$

At this point two possibilities exist, leading to two distinct classes of methods.

Selectively Lumped (SL) schemes. If in the last line of (35) only mass-matrix corresponding to the first Galerkin integral is lumped we obtain the following explicit formulation :

$$
|S_i| \frac{\delta u_i^k}{\Delta t} = - \sum_{T|\in T} \left(\Phi_i^{\mathrm{RK}(k)} - \int_T \varphi_i \overline{\frac{\delta u_h^k}{\Delta t}} dx\, dy \right) \tag{38}
$$

In this case, the effect of the lumping simply leads to the following modification of the mass matrix. If by m_{ij}^{G} we denote the Galerkin mass matrix, we have for the selectively lumped schemes :

$$
m_{ij}^{\mathrm{SL}} = m_{ij}^T - m_{ij}^{\mathrm{G}}, \quad m_{ij}^{\mathrm{G}} = \frac{|T|}{12}(\delta_{ij} + 1) \tag{39}
$$

Globally Lumped (GL) schemes. If we lump all the Galerkin integrals we obtain the following explicit formulation :

$$
|S_i| \frac{\delta u_i^k - \overline{\delta u_i^k}}{\Delta t} = - \sum_{T|\in T} \Phi_i^{\mathrm{RK}(k)} \tag{40}
$$

In this case, there is no modification at all on the residual distribution formulation, however, the lumping modifies the explicit iterations that now depend on the definition of $\overline{\delta u^k}$.

All that remains to do is to properly define $\overline{r}^k$ and $\overline{\delta u^k}$, such that we can still keep the desired accuracy.

4.2 Accuracy and time-stepping

We want to derive a sufficient condition on the $\overline{r}^k$ guaranteeing that the accuracy of the Runge-Kutta Galerkin approximation is not lost when adding the bubble contribution. To do this we use a truncation error analysis, following the approach of [28]. All the details of the analysis are given in two appendices at the end of the paper. The general idea of the proof is, given a classical solution w, and a smooth function $\psi \in C_0^1(\Omega)$, to verify under which conditions the truncation error

$$\begin{aligned}\mathcal{E}_n = \Bigg| & \sum_{i\in\mathcal{T}_h} \psi_i \int_\Omega \varphi_i \left(\frac{\delta w_h^{n+1}}{\Delta t} + f^{n+1}(w_h) \right) dx\, dy + \\ & \sum_{i\in\mathcal{T}_h} \psi_i \sum_{T|i\in T} \int_T \gamma_i \left(\frac{\delta \overline{w_h^{n+1}}}{\Delta t} + f^{n+1}(w_h) \right) dx\, dy \Bigg|\end{aligned} \tag{41}$$

is of an order $\mathcal{O}(h^p)$. The analysis reported in appendix 1 and 2 is done for the general case $p \geq 2$, even though the paper focuses only on the case $p = 2$. Note that in the definition of the error we have used the notation introduced in the previous section. This means that $f^{n+1}(w_h)$ represents the discrete evolution operator of the last RK step, which atually makes use of flux values at *known* time-steps (cf. equations (32) and (33)). The analysis makes use of the following two hypotheses.

Hypothesis 4.1 (RK truncation error)**.** *Given a smooth classical solution w such that $\partial_t w + \nabla \cdot \boldsymbol{\mathcal{F}}(w) = 0$, a p-th order RK scheme verifies the truncation error estimate*

$$r^{n+1}(w) = \frac{\delta w^{n+1}}{\Delta t} + f^{n+1}(w) = C_{RK} \Delta t^p$$

Hypothesis 4.2 (Approximate semi-discrete residual estimate)**.** *Given a smooth classical solution w such that $\partial_t w + \nabla \cdot \boldsymbol{\mathcal{F}}(w) = 0$, the approximate semidiscrete residual $\overline{r}$ verifies the estimate*

$$\overline{r}^{n+1}(w) = \frac{\overline{\delta w^{n+1}}}{\Delta t} + f^{n+1}(w) = \overline{C}_{RK} \Delta t^l$$

for some $l \leq p$.

The main result is summarized by the following proposition.

Proposition 4.3 (Accuracy and time-stepping)**.** *Given a p-th order spatial approximation and a p-th order RK scheme verifying hypothesis 4.1, the truncation error* (41) *verifies an estimate of the type*

$$\mathcal{E}_n \leq C\, h^p$$

provided that

1. the bubble γ_i is uniformly bounded
2. *the approximate semi-discrete residual verifies hypothesis 4.2 with*

$$l \geq p-1$$

In particular, in the second order case of interest here, it is enough to provide definitions of the approximate time increments yielding a first order semi-discrete operator.

Remark 4.4 (Accuracy, time-stepping, and distribution coefficients)**.** *As seen in section §3, for all the known consistent formulations of RD we can provide define the bubble fuction as $\gamma_i = \omega_i - \psi_i$ is always bounded. For the formulations recalled in section §3, ω_i, and hence γ_i, is bounded whenever the distribution coefficients β_i^T are.*

To end the construction we give particular definitions of $\overline{r}^k$ that satisfy hypothesis 4.2 (see appendix 2) :

RK2 scheme

$$\begin{aligned} \overline{\delta u^1} = 0 \Rightarrow\ \overline{r}^1 &= \nabla\cdot\boldsymbol{\mathcal{F}}(u^n) \\ \overline{\delta u^2} = \overline{\delta u^{n+1}} = u^1 - u^n \Rightarrow\ \overline{r}^2 &= \frac{u^1-u^n}{\Delta t} + \frac{\nabla\cdot\boldsymbol{\mathcal{F}}(u^n)+\nabla\cdot\boldsymbol{\mathcal{F}}(u^1)}{2} \end{aligned} \tag{42}$$

When combining this definition with the updates (38), and (40), we obtain for the SL schemes

$$\begin{cases} |S_i|\dfrac{u_i^1-u_i^n}{\Delta t} & = -\displaystyle\sum_{T|i\in T}\beta_i\phi(u_h^n) \\ |S_i|\dfrac{u_i^{n+1}-u_i^n}{\Delta t} & = -\displaystyle\sum_{T|i\in T}\left(\Phi_i^{\mathrm{RK2(2)}} - \sum_{j\in T} m_{ij}^{\mathrm{G}}\frac{u_j^1-u_j^n}{\Delta t}\right) \end{cases} \tag{43}$$

with m_{ij}^{G} as in (39), and with

$$\Phi_i^{\mathrm{RK2(2)}} = \sum_{j\in T} m_{ij}\frac{u_j^1-u_j^n}{\Delta t} + \frac{1}{2}\beta_i\left(\phi(u_h^n)+\phi(u_h^1)\right)$$

The update for the GL schemes is somewhat simpler and given by

$$\begin{cases} |S_i|\dfrac{u_i^1-u_i^n}{\Delta t} & = -\displaystyle\sum_{T|i\in T}\beta_i\phi(t^n) \\ |S_i|\dfrac{u_i^{n+1}-u_i^1}{\Delta t} & = -\displaystyle\sum_{T|i\in T}\Phi_i^{\mathrm{RK2(2)}} \end{cases} \tag{44}$$

RK3 scheme

$$\begin{aligned} \delta\overline{u}^1 = 0 \Rightarrow\ \overline{r}^1 &= \nabla\cdot\boldsymbol{\mathcal{F}}(u^n) \\ \delta\overline{u}^2 = \frac{u^1-u^n}{2} \Rightarrow\ \overline{r}^2 &= \frac{u^1-u^n}{2\Delta t} + \frac{\nabla\cdot\boldsymbol{\mathcal{F}}(u^n)+\nabla\cdot\boldsymbol{\mathcal{F}}(u^1)}{2} \\ \delta\overline{u}^3 = \delta\overline{u}^{n+1} = 2(u^2-u^n) \Rightarrow\ \overline{r}^3 &= \frac{2(u^2-u^n)}{\Delta t} + \frac{\nabla\cdot\boldsymbol{\mathcal{F}}(u^n)+\nabla\cdot\boldsymbol{\mathcal{F}}(u^1)+4\nabla\cdot\boldsymbol{\mathcal{F}}(u^2)}{6} \end{aligned} \tag{45}$$

Note that in this case the coefficients involved in the definition take into account the fact that u^1 and u^2 are initial guesses for the solution at times $t^n + \Delta t$ and $t^n + \Delta t/2$, respectively. When combining this definition with the updates (38), and (40), we obtain for the selectively lumped schemes

$$\left\{\begin{array}{ll} |S_i|\dfrac{u_i^1-u_i^n}{\Delta t} = & -\displaystyle\sum_{T|i\in T}\beta_i\phi(u_h^n) \\ |S_i|\dfrac{u_i^2-u_i^n}{\Delta t} = & -\displaystyle\sum_{T|i\in T}\left(\Phi_i^{\mathrm{RK3(2)}}-\sum_{j\in T}m_{ij}^{\mathrm{G}}\frac{u_j^1-u_j^n}{2\Delta t}\right) \\ |S_i|\dfrac{u_i^{n+1}-u_i^n}{\Delta t} = & -\displaystyle\sum_{T|i\in T}\left(\Phi_i^{\mathrm{RK3(3)}}-\sum_{j\in T}m_{ij}^{\mathrm{G}}\,2\frac{u_j^2-u_j^n}{\Delta t}\right) \end{array}\right. \tag{46}$$

with m_{ij}^{G} as in (39), and with

$$\Phi_i^{\mathrm{RK3(2)}} = \sum_{j\in T} m_{ij}\frac{u_j^1-u_j^n}{2\Delta t} + \frac{1}{4}\beta_i\left(\phi(u_h^n)+\phi(u_h^1)\right)$$

and

$$\Phi_i^{\mathrm{RK3(3)}} = \sum_{j\in T} m_{ij}\,2\frac{u_j^2-u_j^n}{\Delta t} + \beta_i\left(\frac{1}{6}\phi(u_h^n)+\frac{1}{6}\phi(u_h^1)+\frac{2}{3}\phi(u_h^2)\right)$$

As before, the update for the globally lumped schemes is somewhat simpler and given by

$$\left\{\begin{array}{ll} |S_i|\dfrac{u_i^1-u_i^n}{\Delta t} & = -\displaystyle\sum_{T|i\in T}\beta_i\phi(u_h^n) \\ \dfrac{|S_i|}{\Delta t}\left(u_i^2-\dfrac{u_i^1+u_i^n}{2}\right) & = -\displaystyle\sum_{T|i\in T}\Phi_i^{\mathrm{RK3(2)}} \\ \dfrac{2|S_i|}{\Delta t}\left(\dfrac{u_i^{n+1}+u_i^n}{2}-u_i^2\right) & = -\displaystyle\sum_{T|i\in T}\Phi_i^{\mathrm{RK3(3)}} \end{array}\right. \tag{47}$$

Remark 4.5 (Fluctuations/signals)**.** *Both formulations, the one based on selective lumping and the one based on global lumping, allow to see the RD component of the discretization as an error between two different approximations of the unknown at certain time levels. When using the formulation F1 of the RD discretization (cf. section §3, equation* (13)*), the second step of the RK2 scheme with selective lumping can be recast as*

$$|S_i|\frac{u_i^{n+1}-u_i^n}{\Delta t} - \int_\Omega \varphi_i\frac{u_h^1-u_h^n}{\Delta t} = -\sum_{T|i\in T}\beta_i\Phi^{RK2(2)} \tag{48}$$

where

$$\Phi^{RK2(2)} = \int_T\left(\frac{u_h^1-u_h^n}{\Delta t}+\frac{1}{2}\nabla\cdot\boldsymbol{\mathcal{F}}_h(u_h^n)+\frac{1}{2}\nabla\cdot\boldsymbol{\mathcal{F}}_h(u_h^1)\right)dx\,dy$$

Clearly equation (48) *expresses the error between two local approximations of the time variation of the unknown as a function of signals proportional to elemental errors represented by the residual* $\Phi^{RK2(2)}$*. This is even more apparent in the case of the globally lumped scheme which reads, in RK2 case :*

$$|S_i|\frac{u_i^{n+1}-u_i^1}{\Delta t} = -\sum_{T|i\in T}\beta_i\Phi^{RK2(2)} \tag{49}$$

The RK3 version of the last equation is obtained immediately from equation (47)*. In this case the RD wheighted average on the left expresses the between the two different approximations of the unknown at time* t^{n+1}*. The same remarks applies of course to the case of the RK3 schemes. In some way the explicit formulations proposed here lead us back to the original ideas of P.L.Roe [31] in which the nodal error is proportional to the signals sent by surrounding elements.*

Remark 4.6 (Relations with explicit predictor-corrector)**.** *The explicit formulation proposed here is also related to the explicit predictor/multi-corrector formulation of the SUPG scheme used for example in [22, 23, 24] (see also [37, 25, 33]). In the simplest setting, in this formulation on replaces an implicit time integrator by a finite number of explicit steps. In the case of the Crank-Nicholson time integrator for example the idea is to rewrite the SUPG scheme as*

$$|S_i|\frac{u_i^1-u_i^n}{\Delta t} = -\int_\Omega \varphi_i\vec{a}\cdot\nabla u_h^n\,dx\,dy + \sum_{T|i\in T}\int_T \vec{a}\cdot\nabla\varphi_i\,\tau\,\vec{a}\cdot\nabla u_h^n\,dx\,dy$$

$$|S_i|\frac{u_i^k-u_i^n}{\Delta t} = -\int_\Omega \varphi_i\vec{a}\cdot\nabla\frac{u_h^{k-1}+u_h^n}{2}\,dx\,dy + \sum_{T|i\in T}\int_T \vec{a}\cdot\nabla\varphi_i\,\tau\left(\frac{u_h^{k-1}-u_h^n}{\Delta t}+\vec{a}\cdot\nabla\frac{u_h^{k-1}+u_h^n}{2}\right)dx\,dy$$

where $k\geq 2$ *and* $u_i^{n+1}=u_i^{k_{\max}}$*. The second relation can immediately be recast as*

$$|S_i|\frac{u_i^k-u_i^n}{\Delta t} = -\sum_{T|i\in T}\left(\sum_{T|i\in T}\Phi_i^{SUPG(k)} - \int_T \varphi_i\frac{u_h^{k-1}-u_h^n}{\Delta t}dx\,dy\right)$$

with

$$\Phi_i^{SUPG(k)} = \int_T (\varphi_i+\vec{a}\cdot\nabla\varphi_i\,\tau)\left(\frac{u_h^{k-1}-u_h^n}{\Delta t}+\vec{a}\cdot\nabla\frac{u_h^{k-1}+u_h^n}{2}\right)dx\,dy$$

and

$$\sum_{j\in T}\Phi_j^{SUPG(k)} = \Phi^k = \int_T\left(\frac{u_h^{k-1}-u_h^n}{\Delta t}+\vec{a}\cdot\nabla\frac{u_h^{k-1}+u_h^n}{2}\right)dx\,dy$$

This is basically the selectively lumped formulation of the SUPG scheme in RD form. In particular, when using only one correction step we end up exactly with the RK2 scheme (43)*.*

Remark 4.7 (Explicit Residual Based Finite Volume formulation)**.** *The approach presented here finds application also in the case of Finite Volume discretizations where the stabilization operator is proportional to some local approximation of the residual, rather than to local variations of the solution. Such*

schemes have been proposed for example in [12, 10] and, in a different spirit, in [11]. The schemes of [12, 10] in their basic formulation can be rewritten as

$$|C_i|\frac{du_i}{dt} + \oint_{\partial C_i} \mathcal{H}_C \cdot \hat{n}\, dl - \frac{1}{2}\sum_j h_j \Psi_{ij} \Phi_{ij} = 0 \tag{50}$$

where $\mathcal{H}_C$ is a centered finite volume numerical flux, while the last term represent stabilization terms. These terms are function of a local residual Φ_{ij} computed on the staggered cell C_{ij} (cf. figure 2) and defined as [12, 10]

$$\Phi_{ij} = \int_{C_{ij}} \left(\frac{du_h}{dt} + \nabla \cdot \boldsymbol{\mathcal{F}}_h(u_h) \right) dx\, dy$$

where now u_h is a polynomial approximation of the unknown recontructed starting from cell averages. *We refer to [12, 10] for further details, and in particular for the definition of the local mesh size h_j, and of the Ψ_{ij} parameter in* (50)*. The important point is that the residual Φ_{ij} has to include the time derivative of the numerical unknown to attain consistency in the spatial discretization. In [12, 10] the authors use the same discrete operator to approximate both the time derivative of u_i in* (50)*, and in Φ_{ij}. this naturally leads to the appearance of a mass matrix rendering the scheme implicit* in space. *The approach proposed here allows to overcome this limitation allowing the contruction of an explicit RK schemes in which the time derivative in Φ_{ij} is approximated by time increments using known values of the discrete solution. In the RK2 case for example the scheme would read :*

$$|C_i|\frac{u_i^1 - u_i^n}{\Delta t} + \oint_{\partial C_i} \mathcal{H}_C^n \cdot \hat{n}\, dl - \frac{1}{2}\sum_j h_j \Psi_{ij} \int_{C_{ij}} \nabla \cdot \boldsymbol{\mathcal{F}}_h^n\, dx\, dy = 0$$

$$|C_i|\frac{u_i^{n+1} - u_i^n}{\Delta t} + \oint_{\partial C_i} \frac{\mathcal{H}_C^n + \mathcal{H}_C^1}{2} \cdot \hat{n}\, dl - \frac{1}{2}\sum_j h_j \Psi_{ij} \int_{C_{ij}} \left(\frac{u_h^1 - u_h^n}{\Delta t} + \frac{\nabla \cdot \boldsymbol{\mathcal{F}}_h^n + \nabla \cdot \boldsymbol{\mathcal{F}}_h^1}{2} \right) dx\, dy = 0$$

with $\mathcal{H}_C^1 = \mathcal{H}_C(u_h^1)$ and $\boldsymbol{\mathcal{F}}_h^1 = \boldsymbol{\mathcal{F}}_h(u_h^1)$.

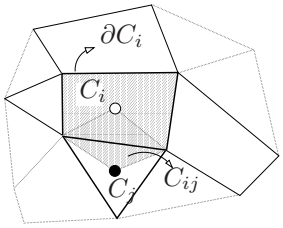


Figure 2: Residual Based Finite Volume. Cells C_i and C_j, and staggered cell C_{ij}.

5 Schemes used in the numerical experiments

This section is devoted to the description of the schemes actually used in the numerical tests discussed in the following paragraphs, and of some details relative to their implementation. We will discuss the results obtained with four well known schemes : the LDA scheme, the blended LDAN scheme or B scheme, the Streamline Upwind scheme, or SU scheme for short, and a centered Blended scheme, Bc for short, constructed starting from the limited stabilized Lax-friedrich's scheme of [2]. The results obtained with other RD schemes are very similar in nature. An importat remark is that *so far we still have not worked on the adaptation of nonlinear RD discretizations to the construction proposed in tha paper.* This means that we limited ourselves to code the schemes as they are presented in the literature. Improvements will be made in the future concerning strict preservation of positivity. Even so, as we will see, the numerical results are excellent, and confirm our theoretical analysis.

5.1 LDA scheme

We test our construction on the well known second order linear multidimensional upwind LDA scheme, defined by the distribution coefficients [16] :

$$\beta_i^{\mathrm{LDA}} = k_i^+ \left(\sum_{j \in T} k_j^+ \right)^{-1} \tag{51}$$

where, using the notation of equation (2), we define $\forall T \in \mathcal{T}_h$

$$k_i = \frac{1}{2} \frac{\partial \boldsymbol{\mathcal{F}}(\overline{u})}{\partial u} \cdot \vec{n}_i \tag{52}$$

with $\overline{u}$ the arithmetic average of the values of u_h in the nodes of T. Note that in the case of a system of conservation laws, the k_is are matrices, and their sign in (51) is computed in the standard matrix sense, via eigenvalue decomposition. For more details on the definition and properties of the LDA scheme, the reader is referred to [16, 1].

In the scalar case we will compare the results obtained when using the different formulations recalled in section §3. In particular, the scheme has been coded exactly as decribed in equations (43) and (46), for the selectively lumped scheme, and in equations (44) and (47) for the global lumped scheme. In both cases, we replace the quantities $\Phi_i^{\mathrm{RK2}(k)}$ and $\Phi_i^{\mathrm{RK3}(k)}$ by (see equations (37), (42), and (45) for the notation)

$$\Phi_i^{\mathrm{LDA}(k)} = \sum_{j \in T} m_{ij}^{\mathrm{LDA}} \overline{\delta u_j^k} + \Delta t \, \beta_i^{LDA} \phi^{\mathrm{RK}(k)} \tag{53}$$

In particular, the form of the mass matrix m_{ij}^{LDA} will depend of the formulation chosen (cf. section §3). To shorten the text we will lump together the acronyms when referring to a scheme. For example, we shall speak of the LDA-F1-SL-RK2 when referring to the scheme obtained using the LDA distribution coefficients, the mass matrix of the formulation 1, selective lumping, and the RK2 scheme in time. Similarly for all the other combinations.

5.2 Blended LDA-N scheme

As suggested by its name, the Blended LDA-N scheme, or B scheme for short, is a blending between the LDA scheme of section §5.1 with the first order positive multidimensional upwind N scheme defined by the spatial splitting [16, 1, 14]

$$\phi_i^{\rm N} = k_i^+ \left(u_i - u_{\rm in}\right) , \quad u_{\rm in} = \left(\sum_{j\in T} k_j^+\right)^{-1} \left(-\phi(u_h) + \sum_{j\in T} k_j^+ u_j\right)$$

In particular, following [4], we set for the B scheme

$$\Phi_i^{{\rm B}(k)} = (1 - l(u_h))\Phi_i^{{\rm LDA}(k)} + l(u_h)\Phi^{{\rm N}(k)} \tag{54}$$

with $\Phi_i^{{\rm LDA}(k)}$ given by (53) and with

$$\Phi_i^{{\rm N}(k)} = \frac{|T|}{3}\overline{\delta u_i^k} + \Delta t\, \phi_i^{{\rm N}(k)}$$

having denoted by $\phi_i^{{\rm N}(k)}$ the spatial contribution of the N scheme corresponding to the k-the RK step. Expression (54) is used in (43) , (46), (44) and (47) to replace $\Phi_i^{{\rm RK2}(k)}$ and $\Phi_i^{{\rm RK3}(k)}$.

Concerning the blending paramenter $l(u_h)$ we have used the standard definition of Deconinck *et al.* [17, 14] (cf. also equation (37)) :

$$l(u_h) = \frac{|\Phi^{{\rm RK}(k)}|}{\sum\limits_{j\in T} |\Phi_j^{{\rm N}(k)}|}$$

for systems of equations, the blending procedure has been performed on residuals projected in characteristic directions, as explained in [5, 4].

As a last remark, we note that only when using global lumping for $l(u_h) = 1$ does the B scheme defined by (54) reduce to the N scheme with RK time integration. In the selective lumping case, for $l(u_h) = 1$ we get (cf. equation (38))

$$|S_i|\frac{\delta u_i^k}{\Delta t} + \sum_{T|i\in T} \phi_i^{{\rm N}(k)} = |S_i|\frac{\overline{\delta u_i^k}}{\Delta t} - \int\limits_\Omega \varphi_i \frac{\overline{\delta u_h^k}}{\Delta t}\, dx\, dy$$

where the left hand side corresponds to the k-th RK step of the N scheme, while the right hand side contains some kind of anti-diffusive correction (cf. section §3.).

5.3 The SU scheme

To test the behavior of our formulation with different type of discretizations, we also consider centered schemes. The first is referred to in the RD literature either as SUPG scheme or as LW scheme. It is defined by the distribution coefficients

$$\beta_i^{\rm SU} = \frac{1}{3} + k_i\,\tau \tag{55}$$

Independently on the definition of the scaling parameter τ, the second term on the last definition introduces some Streamline Upwinding in the distribution [16], which is why we refer to this scheme as to the SU scheme. In our computations we have taken

$$\tau = \left(\sum_{j \in T} |k_j| \right)^{-1} \tag{56}$$

For the Euler equations, last expression is meant in the usual matrix sense.

Finally, we replace the quantities $\Phi_i^{\mathrm{RK2}(k)}$ and $\Phi_i^{\mathrm{RK3}(k)}$ by (see equations (37), (42), and (45) for the notation)

$$\Phi_i^{\mathrm{SU}(k)} = \sum_{j \in T} m_{ij}^{\mathrm{SU}} \overline{\delta u_j^k} + \Delta t \; \beta_i^{SU} \phi^{\mathrm{RK}(k)} \tag{57}$$

As for the LDA scheme, also for the SU scheme the form of the mass matrix m_{ij}^{SU} depends of the formulation chosen (cf. section §3).

5.4 Central blended scheme

In [2, 29] the authors introduce a centered discretization based on a nonlinear variant of a Lax-Friedrich's scheme. This limited stabilized Lax-Friedrich's scheme, or LLFs scheme, is obtained starting from the positive first order Lax-Friedrich's (LF) splitting

$$\phi_i^{\mathrm{LF}} = \frac{1}{3} \left(\phi(u_h) + \alpha_{\mathrm{LF}} \sum_{j \in T} (u_i - u_j) \right) \tag{58}$$

where α_{LF} is the Lax-Friedrich's dissipation coefficient which we set to

$$\alpha_{\mathrm{LF}} = \frac{1}{2} a_T \, h_T \,, \quad a_T = \max_{j] \in T} \left\| \frac{\partial \boldsymbol{\mathcal{F}}(u_j)}{\partial u} \right\|$$

in the scalar case, while for the Euler equations we have set

$$\alpha_{\mathrm{LF}} = \frac{1}{2} \max_{j] \in T} (\|\vec{u}_j\| + a_j) \, h_T$$

with $\vec{u}$ the flow speed, a the speed of sound, and h_T a reference length for element T.

The LF scheme is only first order. To obtain a formally second order nonlinear splitting we proceed as follows. First we define the LF-RK splitting

$$\Phi_i^{\mathrm{LF}(k)} = \frac{|T|}{3} \overline{\delta u_i^k} + \Delta t \, \phi_i^{\mathrm{LF}(k)}$$

having denoted by $\phi_i^{\mathrm{LF}(k)}$ the k-the RK step of the spatial operator (58). Next, we compute bounded distribution coefficients by applying a sign preserving nonlinear mapping. Several ways of doing this exist, and we refer to [6, 2] for a discussion. Here, we set (cf. equation (37)) :

$$\beta_i^{\mathrm{LLF}} = \frac{\max \left(0 \,, \; \Phi_i^{\mathrm{LF}(k)} \Phi^{\mathrm{RK}(k)} \right)}{\sum\limits_{j \in T} \max \left(0 \,, \; \Phi_j^{\mathrm{LF}(k)} \Phi^{\mathrm{RK}(k)} \right)} \tag{59}$$

The limited LF scheme is then defined by

$$\Phi_i^{\mathrm{LLF}(k)} = \beta_i^{\mathrm{LLF}} \Phi^{\mathrm{RK}(k)}$$

As shown in previous work [2, 29, 27], the limiter (60) not taking into account the directional propagation of the information typical of hyperbolic problems, the LLF scheme shows mild spurious modes that eventually reduce its accuracy to first order. This is cured as in the above references by adding an upwind bias inspired by the SU scheme :

$$\beta_i^{\mathrm{LLFs}} = \beta_i^{\mathrm{LLF}} + \delta(u_h) k_i \tau \tag{60}$$

with τ as in (56). We refer the reader to [2, 29, 27] for more details on the theoretical background leading to this choice. We limit ouselves to recall that $\delta(u_h)$ is a smoothness sensor such that $\delta(u_h) = 1$ in smooth areas, while $\delta = \mathcal{O}(h_T)$ in presence of discontinuities. In our computations we have set in the scalar case [2, 29, 27]

$$\delta(u_h) = \min\left(1, \frac{\Delta t\, h_T^2\, a_T |u|_T}{|\Phi^{\mathrm{RK}(k)}|}\right) \tag{61}$$

where $|u|_T$ is the maximum of the absolute value of the solution over the element. For the Euler equations, the extension is done following [2] : the limiter (60) is evaluated on residual projected on local characteristic directions, while the $|\Phi^{\mathrm{RK}(k)}|$ in (61) is replaced by the *scalar* entropy component of $\Phi^{\mathrm{RK}(k)}$. This is computed as

$$\varphi_s = l_0 \cdot \Phi^{\mathrm{RK}(k)}$$

where l_0 is the left eigenvector of the flux Jacobian corresponding to the entropy wave. For the Euler equations $\delta(u_h)$ is then the scalar quantity (see [2] for more)

$$\delta(u_h) = \min\left(1, \frac{\Delta t\, h_T^2}{|\varphi_s|}\right) \tag{62}$$

Normally, we would set

$$\Phi_i^{\mathrm{LLFs}(k)} = \beta_i^{\mathrm{LLFs}} \Phi^{\mathrm{RK}(k)} \tag{63}$$

and replace $\Phi_i^{\mathrm{RK2}(k)}$ and $\Phi_i^{\mathrm{RK3}(k)}$ in (43) , (46), (44) and (47) by (63). However, we found that much better results are obtained, at negligible extra cost, by using the central blended scheme, or Bc scheme for short, defined by

$$\Phi_i^{\mathrm{Bc}(k)} = \beta_i^{\mathrm{Bc}} \Phi^{\mathrm{RK}(k)}, \quad \beta_i^{\mathrm{Bc}} = \delta(u_h)\beta_i^{\mathrm{SU}} + (1 - \delta(u_h))\beta_i^{\mathrm{LLF}} \tag{64}$$

From definitions (55) and (60) we immediately see that

$$\Phi_i^{\mathrm{Bc}(k)} = \Phi_i^{\mathrm{LLFs}(k)} + \delta(u_h)\frac{1}{3}\Phi^{\mathrm{RK}(k)}$$

Compared to the LLFs scheme, the only extra cost to evaluate (64) is the addition of the term $\delta(u_h)\Phi^{\mathrm{RK}(k)}/3$. Being $\delta(u_h)$ always a scalar, this cost is clearly negligible.

5.5 Computation of the time step

All the numerical results presented in the folllowing section have been obtained by computing the timestep as (cf. section §5.3) :

$$\Delta t = \min_{i \in \mathcal{T}_h} \frac{|S_i|}{\sum\limits_{T|i \in T} \alpha_{\text{LF}}} \tag{65}$$

For all the nonlinear problems considered, α_{LF} is evaluated using solution values at the last known time step.

A fourier analysis on structured triangulations is under way to have a better estimate of the time step stability limit for the linear schemes.

6 Scalar results

The scalar tests we present have two objectives : verify the accuracy of our explicit formulation for different forms of the mass matrix, and for schemes of different nature (multidimensional upwind, and centered) ; test the non-oscillatory nature of the results obtained with the nonlinear schemes, when no modifications are introduced to take into account the additional terms introduced by RK formulation.

Unless stated, all the numerical tests, including the Euler tests, have been performed on unstructured triangulations with the topology shown on figure 3.

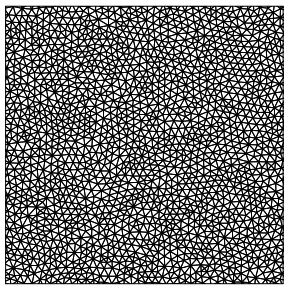


Figure 3: Typical topology of the meshes used in the numerical tests

6.1 Advection of a smooth profile : grid convergence

The first test involves the simple scalar equation

$$\partial_t u + \partial_x u = 0$$

solved on the rectangular domain $[0, 2] \times [0, 1]$. The initial solution is set to

$$u_0 = \begin{cases} \cos^2(2\pi r) & \text{if} \quad r \leq 0.25 \\ 0 & \text{otherwise} \end{cases}$$

with $r^2 = (x - 0.5)^2 + (y - 0.5)^2$. We solve the problem up to time $t = 1$ on a series of 5 meshes with the topology shown on figure 3. The coarsest mesh

has a reference element size $h \approx 1/20$ (10 points in the smooth cosinusoidal profile). The other meshes are obtained via 4 steps of conformal refinement. We use this test to study the accuracy of the different schemes discussed in the paper. The accuracy is monitored by the convergence of the L^1 norm of the error with respect to the exact solution. The behaviour of the L^∞ and L^2 norms is qualitatively and quantitatively very similar.

The first exercice is to verify that indeed our RK formulation leads to second order discretizations, independently on the starting form of the (consistent) mass matrix. We perform the test for all the mass matrix formulations for the LDA scheme, which is the most popular multidimensional upwind RD scheme.

The results are summarized in figures 4 and 5, where we report the grid convergence history and the rate of convergence history, respectively. The first remark we can make is that our explicit formulation does lead to a second order discretization. This is clear especially from the rates of convergence observed. What is more interesting is that the RK2 schemes all yield the same accuracy, while the RK3 scheme with global lumping seem to actually be less and less accurate as the mesh is refined. We believe this might be the consequence of a (mild) linear stability problem. We are currently performing a Fourier analysis on structured grids to better understand this behaviour. There are minor differences between the different mass matrix forms which, in our opinion, do not justify the use of the more complex formulations F3 and F4 (cf. section §3), especially in view of the extension to systems.

We repeat the same exercise with the SU scheme, only this time we only test the mass matrix formulations F1 and F2 (cf. section §3). The results are shown on figure 6. The same remarks made for the LDA scheme apply also to the SU distribution : second order of accuracy is obtained already with the RK2 scheme, independently of the mass matrix and lumping choices ; the RK3 scheme with global lumping suffers from a drop in the convergence rate, which might be caused by the presence of a linear instability.

We now come to the nonlinear schemes. We firt test the B scheme, using either formulation F1, or formulation F2 for the LDA mass matrix. The results are displayed on figure 7. The asymptotic rate of convergence obtained is about 1.75-1.8, independently on the formulation. Clearly, when using global lumping, the drop in convergence speed of the LDA affacts the B scheme as well. Lastly, on figure 8 we report the results obtained with the Bc scheme. Once more, we observe asymptotic convergence rates ranging from 1.7 to 1.9, with the exception of the RK3 scheme in conjunction with global lumping.

We believe these tests confirm our theoretical contruction. In particular the fact that with the RK2 scheme one already obtains a second order discretization. Moreover, the fact that different forms of the mass matrix lead to very similar accuracy properties leads us to the conclusion that the choice of the form of the mass matrix should be done on the basis of stability (or positivity eventually) considerations. This is the objective of our current investigations.

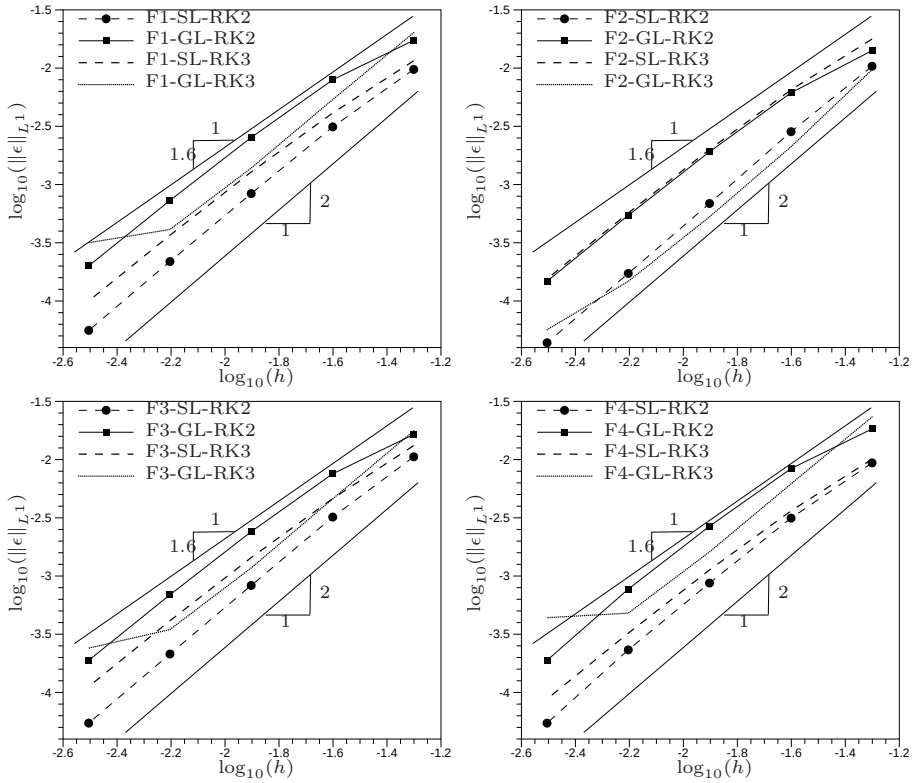


Figure 4: Scalar advection : grid convergence for the LDA scheme. Top-left : formulation F1. Top-right : formulation F2. Bottom-left : formulation F3. Bottom-right : formulation F4.

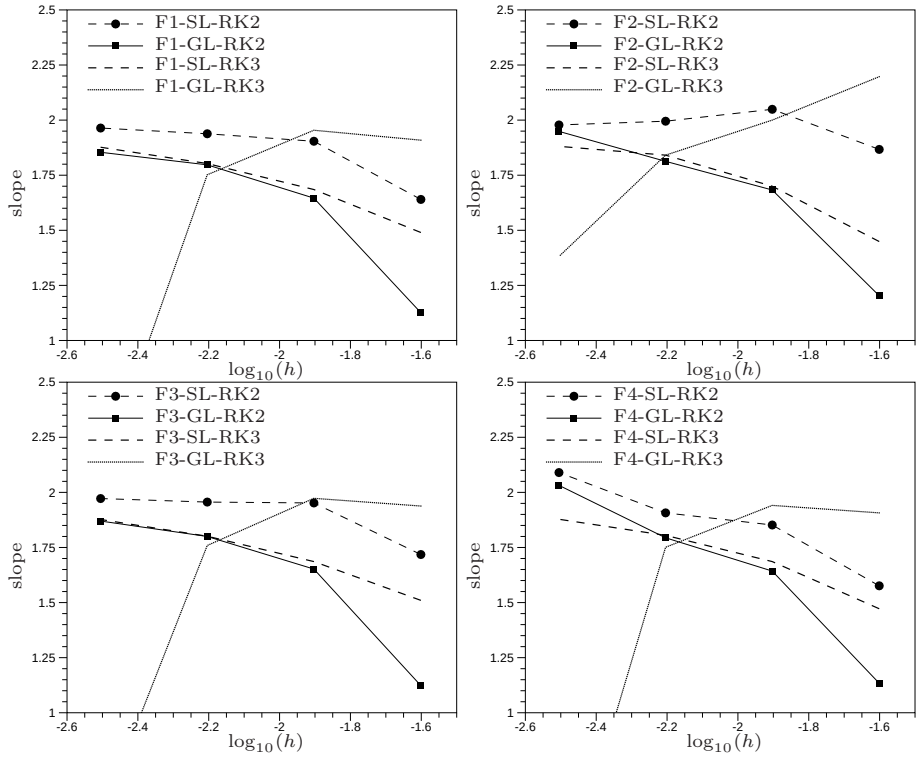


Figure 5: Scalar advection : convergence rates for the LDA scheme. Top-left : formulation F1. Top-right : formulation F2. Bottom-left : formulation F3. Bottom-right : formulation F4.

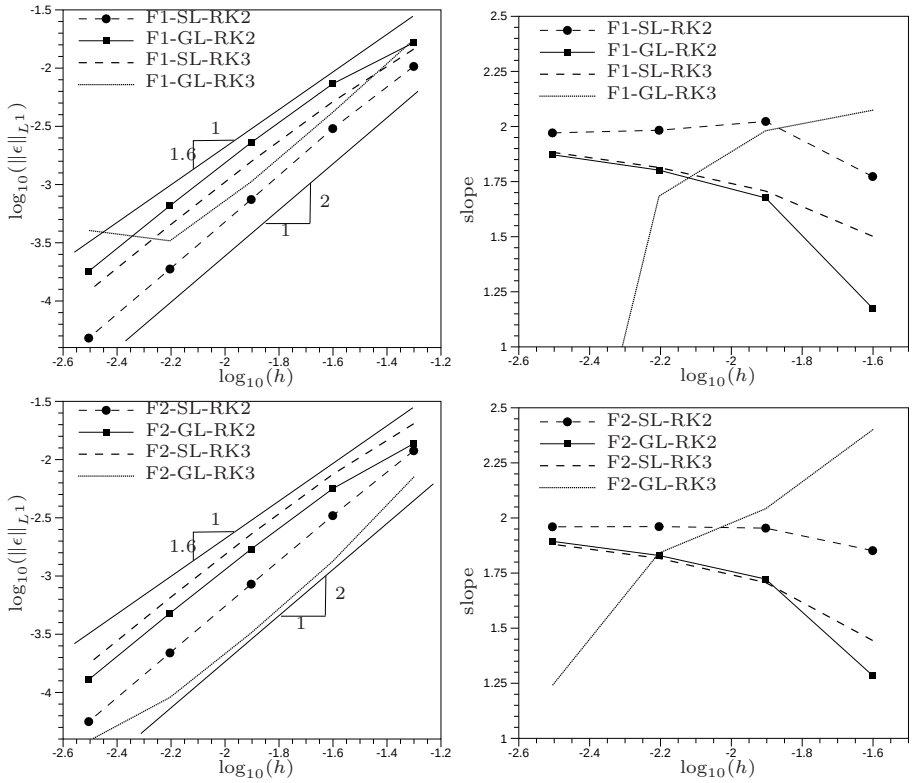


Figure 6: Scalar advection : grid convergence for the SU scheme. Top row : formulation F1. Bottom row : formulation F2. Left column : convergence history. Right column : convergence rates.

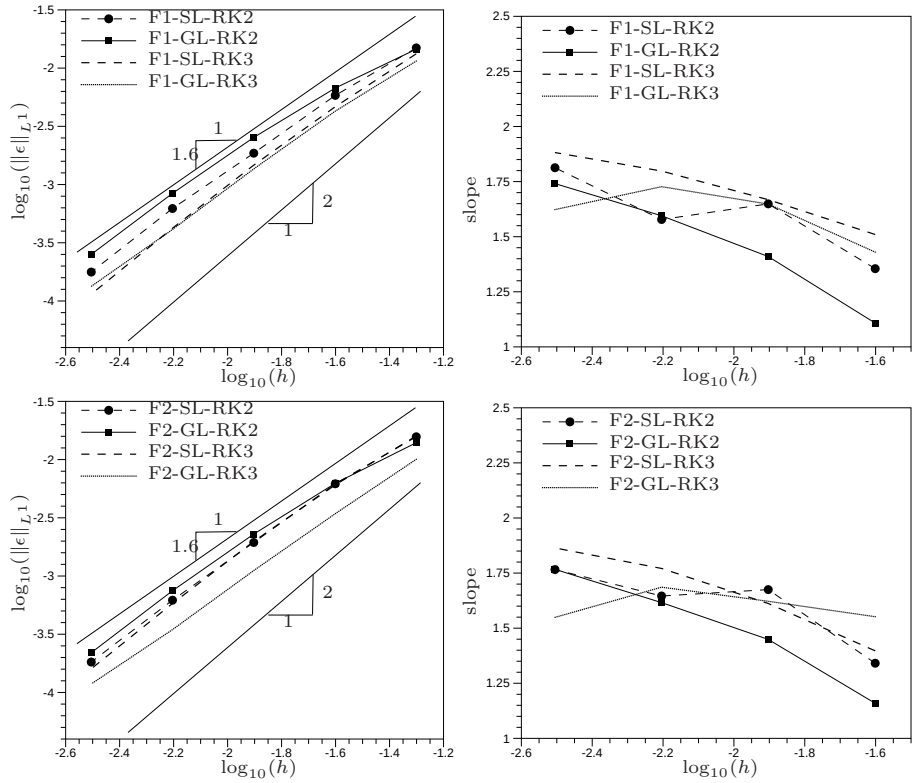


Figure 7: Scalar advection : grid convergence for the B scheme. Top row : formulation F1. Bottom row : formulation F2. Left column : convergence history. Right column : convergence rates.

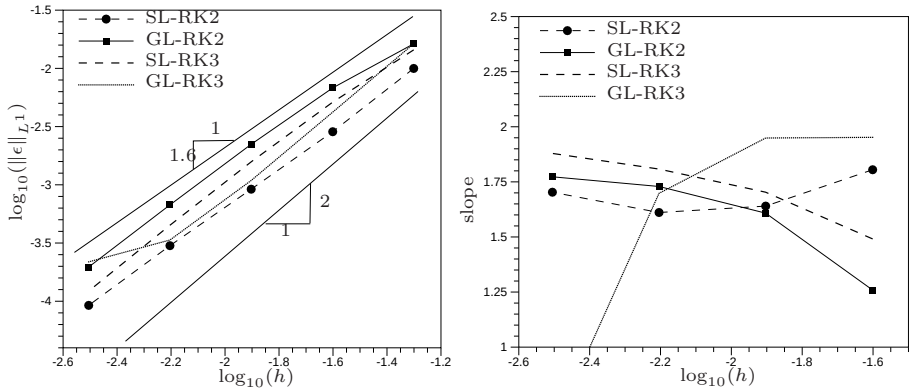


Figure 8: Scalar advection : grid convergence for the Bc scheme. Left column : convergence history. Right column : convergence rates.

6.2 Discontinuous solutions : 2d Burger's equation

We consider now the nonlinear 2d Burger's equation

$$\partial_t u + \partial_x \left(\frac{u^2}{2} \right) + \partial_y \left(\frac{u^2}{2} \right) = 0$$

We solve the problem on the square $[-1, 1]^2$ with the discontinuous initial solution

$$u_0 = \begin{cases} 1 & \text{if } \; x \in [-0.6, -0.1] \times [-0.35, 0.15] \\ 0 & \text{otherwise} \end{cases}$$

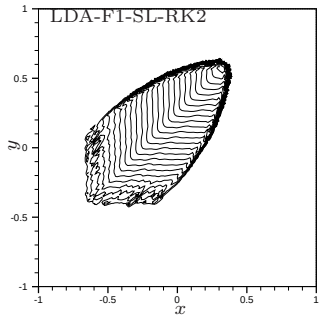
The problem is solved up to the final time $t = 1$ on an unstructured triangulation with the topology shown on figure 3, and reference size $h \approx 1/80$. We compare the results of all the schemes considered. Only the simplest mass matrix form F1 (cf. section §3) has been used.

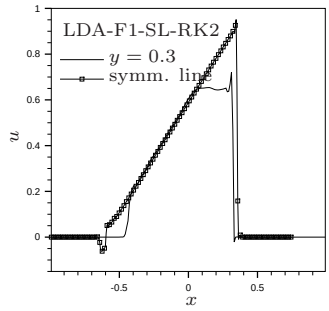
We first consider the multidimensional upwind LDA and B schemes. The results for different RK schemes and lumping strategy are shown on figures from 9 to 16. Concerning the LDA scheme, as one would expect, the solution exhibits oscillations near the discontinuities. These oscillations are much more pronounced when using selective lumping. More interesting are however the results of the B scheme, shown on figures 13 to 16. From all the contour plots we can see that the solution is smoother (the kinks close to the shock are less pronounced) when compared to the LDA scheme. When using selective lumping oscillations still appear close to the discontinuity. This, as observed at the end of section §5.2, is a consequence of the non-positive coefficients introduced by the Galerkin integral present when lumping selectively. A mixed formulation, in which these terms are also multiplied by the blending coefficient, might be used to cure the problem, but this is beyond the scopes of this paper and left for future work. The results obtained with global lumping show the expected monotone resolution of the discontinuities. Probably, the small negative undershoots can also be avoided by properly redefining the blending. Again, this is

beyon the scopes of this paper and left for the future.

The results obtained with the centered distributions are displayed on figures from 17 to 24. The qualitative bahavior of these schemes is similar to the one of the multidimensional upwind discretizations. The linear SU scheme gives oscillations near the discontinuities. Milder oscillations are obtained when using global lumping. Concerning the Bc scheme, the results show smoother contours (the kinks close to the shock are less pronounced), and less oscillations. However, only the results obtained using global lumping show a monotone resolution of the moving shock.

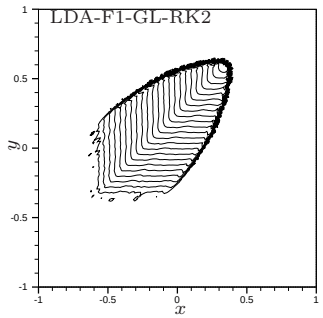
We judge the results obtained on this nonlinear problem very encouraging : even without modifying the basic RD distribution, the nonlinear second order explicit RK-RD schemes can yield monotone solutions. This is further confirmed by the Euler results discussed hereafter.

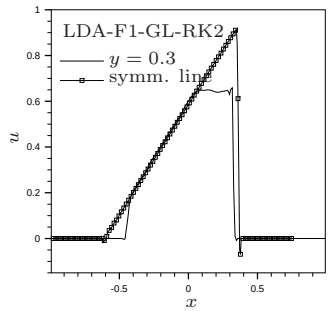




$u_{\min}$	$u_{\max}$
-0.128	1.006

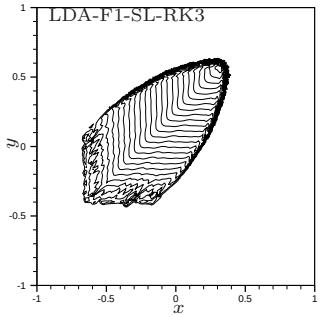
Figure 9: 2d Burger's equation : LDA-F1-SL-RK2 scheme. Left : contours at time $t = 1$. Middle : solution along the line $y = 0.3$ and along the symmetry line. Right : minimum and maximum values of the solution.

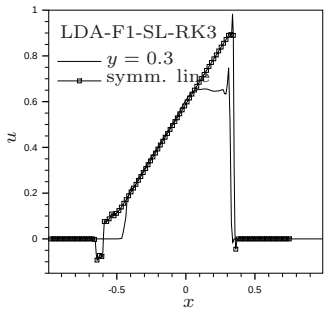




$u_{\min}$	$u_{\max}$
-0.144	0.951

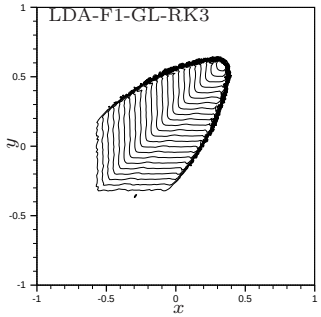
Figure 10: 2d Burger's equation : LDA-F1-GL-RK2 scheme. Left : contours at time $t = 1$. Middle : solution along the line $y = 0.3$ and along the symmetry line. Right : minimum and maximum values of the solution.

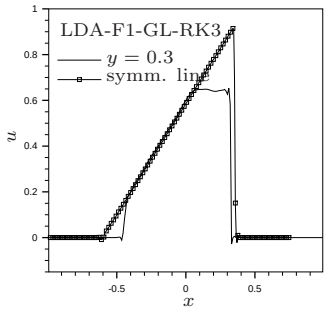




$u_{\min}$	$u_{\max}$
-0.144	1.081

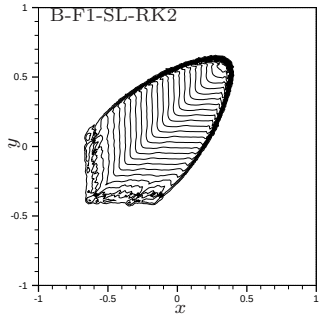
Figure 11: 2d Burger's equation : LDA-F1-SL-RK3 scheme. Left : contours at time $t = 1$. Middle : solution along the line $y = 0.3$ and along the symmetry line. Right : minimum and maximum values of the solution.

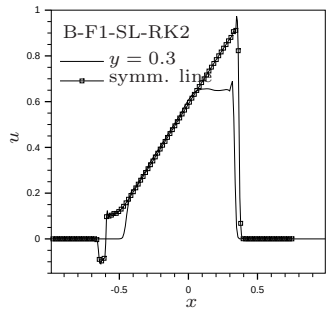




$u_{\min}$	$u_{\max}$
-0.243	0.952

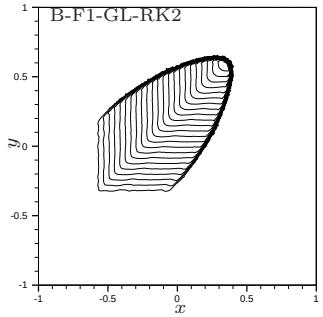
Figure 12: 2d Burger's equation : LDA-F1-GL-RK3 scheme. Left : contours at time $t = 1$. Middle : solution along the line $y = 0.3$ and along the symmetry line. Right : minimum and maximum values of the solution.

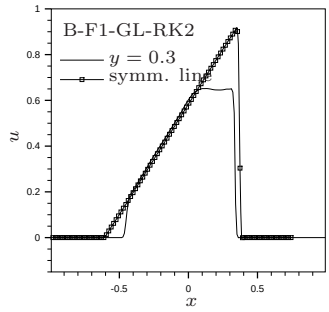




$u_{\min}$	$u_{\max}$
-0.134	1.018

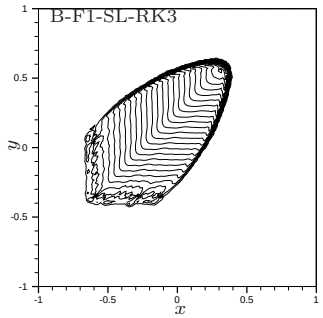
Figure 13: 2d Burger's equation : B-F1-SL-RK2 scheme. Left : contours at time $t = 1$. Middle : solution along the line $y = 0.3$ and along the symmetry line. Right : minimum and maximum values of the solution.

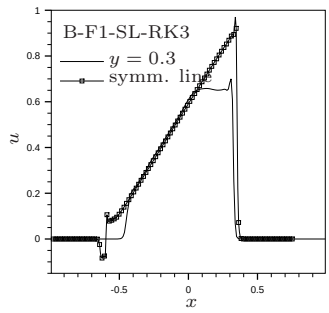




$u_{\min}$	$u_{\max}$
-0.004	0.920

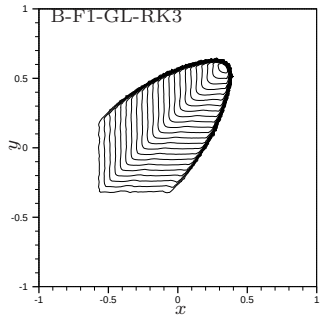
Figure 14: 2d Burger's equation : B-F1-GL-RK2 scheme. Left : contours at time $t = 1$. Middle : solution along the line $y = 0.3$ and along the symmetry line. Right : minimum and maximum values of the solution.

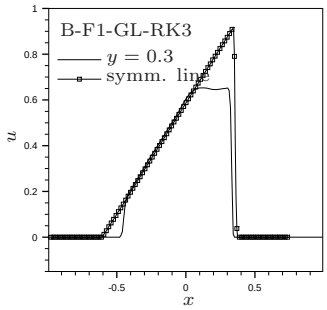




$u_{\min}$	$u_{\max}$
-0.121	1.002

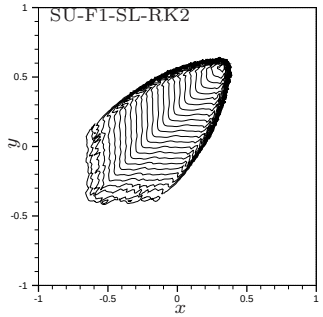
Figure 15: 2d Burger's equation : B-F1-SL-RK3 scheme. Left : contours at time $t = 1$. Middle : solution along the line $y = 0.3$ and along the symmetry line. Right : minimum and maximum values of the solution.

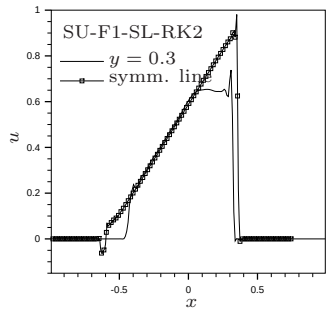




$u_{\min}$	$u_{\max}$
-0.027	0.913

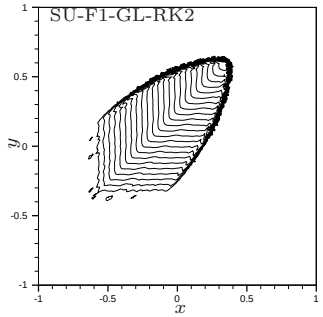
Figure 16: 2d Burger's equation : B-F1-GL-RK3 scheme. Left : contours at time $t = 1$. Middle : solution along the line $y = 0.3$ and along the symmetry line. Right : minimum and maximum values of the solution.

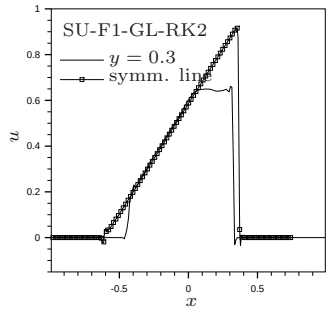




$u_{\min}$	$u_{\max}$
-0.103	1.064

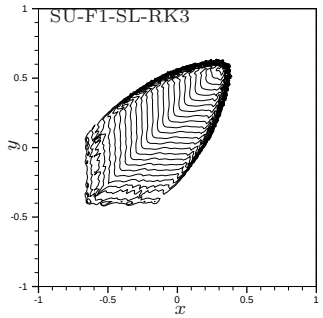
Figure 17: 2d Burger's equation : SU-F1-SL-RK2 scheme. Left : contours at time $t = 1$. Middle : solution along the line $y = 0.3$ and along the symmetry line. Right : minimum and maximum values of the solution.

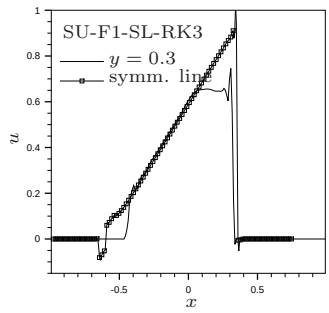




$u_{\min}$	$u_{\max}$
-0.146	0.969

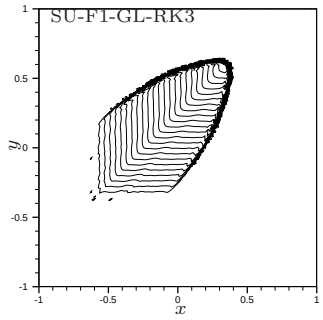
Figure 18: 2d Burger's equation : SU-F1-GL-RK2 scheme. Left : contours at time $t = 1$. Middle : solution along the line $y = 0.3$ and along the symmetry line. Right : minimum and maximum values of the solution.

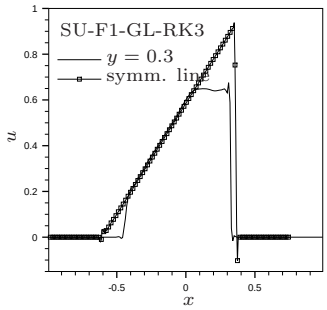




$u_{\min}$	$u_{\max}$
-0.110	1.113

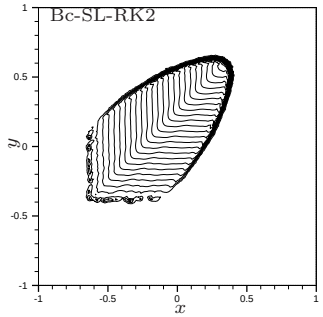
Figure 19: 2d Burger's equation : SU-F1-SL-RK3 scheme. Left : contours at time $t = 1$. Middle : solution along the line $y = 0.3$ and along the symmetry line. Right : minimum and maximum values of the solution.

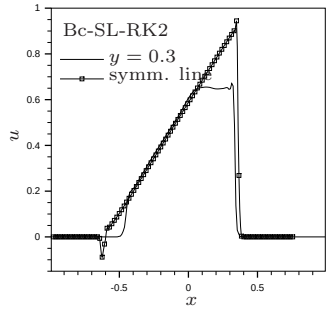




$u_{\min}$	$u_{\max}$
-0.211	0.985

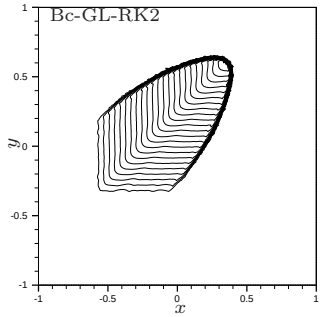
Figure 20: 2d Burger's equation : SU-F1-GL-RK3 scheme. Left : contours at time $t = 1$. Middle : solution along the line $y = 0.3$ and along the symmetry line. Right : minimum and maximum values of the solution.

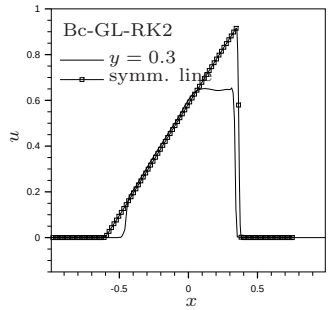




$u_{\min}$	$u_{\max}$
-0.131	0.973

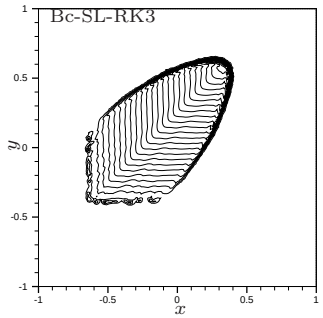
Figure 21: 2d Burger's equation : Bc-SL-RK2 scheme. Left : contours at time $t = 1$. Middle : solution along the line $y = 0.3$ and along the symmetry line. Right : minimum and maximum values of the solution.

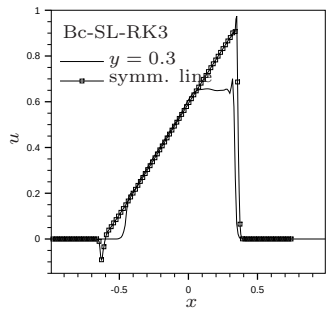




$u_{\min}$	$u_{\max}$
-0.014	0.919

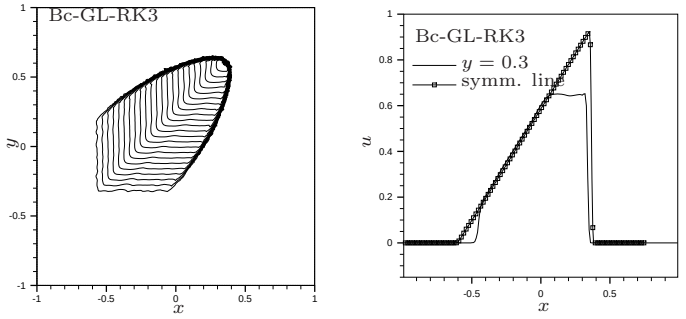
Figure 22: 2d Burger's equation : Bc-GL-RK2 scheme. Left : contours at time $t = 1$. Middle : solution along the line $y = 0.3$ and along the symmetry line. Right : minimum and maximum values of the solution.





$u_{\min}$	$u_{\max}$
-0.135	1.017

Figure 23: 2d Burger's equation : Bc-SL-RK3 scheme. Left : contours at time $t = 1$. Middle : solution along the line $y = 0.3$ and along the symmetry line. Right : minimum and maximum values of the solution.



$u_{\min}$	$u_{\max}$
-0.013	0.929

Figure 24: 2d Burger's equation : Bc-GL-RK3 scheme. Left : contours at time $t = 1$. Middle : solution along the line $y = 0.3$ and along the symmetry line. Right : minimum and maximum values of the solution.

7 Euler equations

As already said in section §5, the extension of the schemes to the system of Euler equations is performed formally. As in the scalar case, the nonlinear schemes are not modified to take into account the additional terms coming from the explicit RK formulation and to improve their behavior close to discontinuities. For simplicity, only the schemes based on the simple mass matrix formulation F1 (cf. section §3) are tested. The objective of the tests is to asses the accuracy of the discretizations, and the behavior of the nonlinear schemes in presence of a strong moving planar shock, and for more complex flow structures involving several contact lines and interactions between shocks and expansions.

7.1 Advection of a vortex : grid convergence

The accuracy of the schemes is measured on the advection of a constant density vortex. The test has been initially proposed in [19], to which we refer for all the details concerning its implementation. The solution involves the advection of a vortex with a constant density profile, and a smooth pressure variation of which the analytical form is known [19]. We solve the problem on a set of 5 unstructured grids with the topology shown on figure 3. The coarsest grid as a reference size of $h \approx 1/20$. The other meshes are obtained by means of 4 successive steps of conformal refinement. We measure the accuracy by means of the L^2 norm of the relative pressure error

$$\epsilon_p = \frac{p - p^{\text{exact}}}{p_\infty}$$

see [19] for the definition of p_∞ and of $-p^{\text{exact}}$. The behaviour of the L^∞ and L^1 norms is qualitatively and quantitatively very similar. The results are displayed on figures from 25 to 28 in terms of error convergence history, and rate of convergence history. The results are qualitatively very similar to the ones discussed in section §6.1. With the exception of the first refining step, we do obtain roughly second order of convergence with all the schemes except the RK3 ones when using global lumping. These schemes, exactly as in the scalar case, show a more or less evident decrease in accuracy, as the mesh is refined.

The poor convergence rate at the first refinement step might be explained by the coarseness of the first meshes : the starting mesh only has 10 points through the vortex core, the second one 20 points. The drop in convergence rates of the RK3-GL schemes might be a consequence of a linear stability problem. This is under investigation.

The main difference between the distribution strategies is that the B scheme gives a slighly smaller asymptotic accuracy of about 1.7, while all the others attain convergence rates closer to 2. While this is expected for the linear schemes, we believe the face that the Bc scheme shows a better convergence is due to the definition of the entropy smoothness sensor $\delta(u_h)$ proposed in [2] and used for the blending, which really is turned on only very close to discontinuities. Once more, the improvement of the nonlinear schemes is a topic for future work. Nevertheless, the results obtained confirm once more our theoretical analysis.

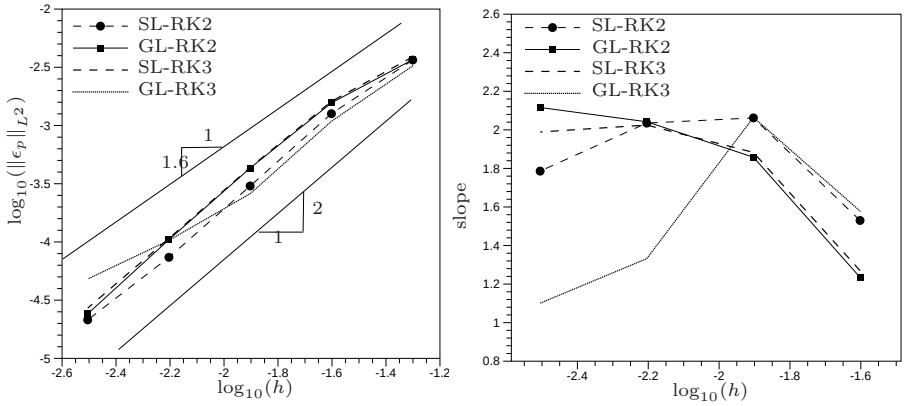


Figure 25: Vortex advection : grid convergence for the LDA scheme with F1. Left column : convergence history. Right column : convergence rates.

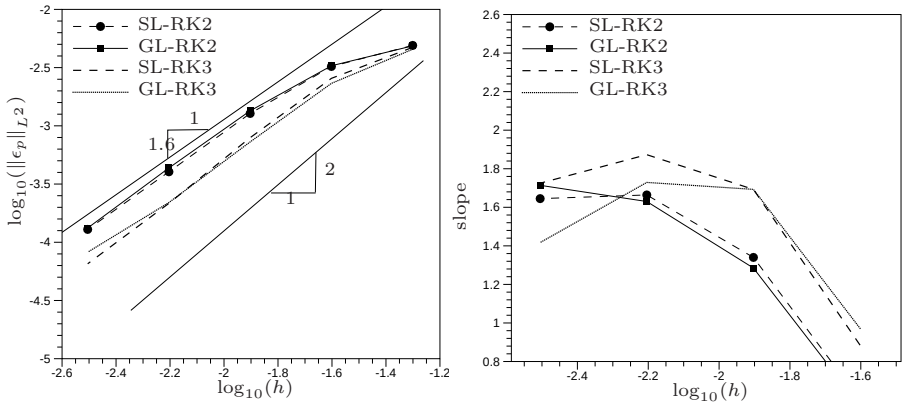


Figure 26: Vortex advection : grid convergence for the B scheme with F1. Left column : convergence history. Right column : convergence rates.

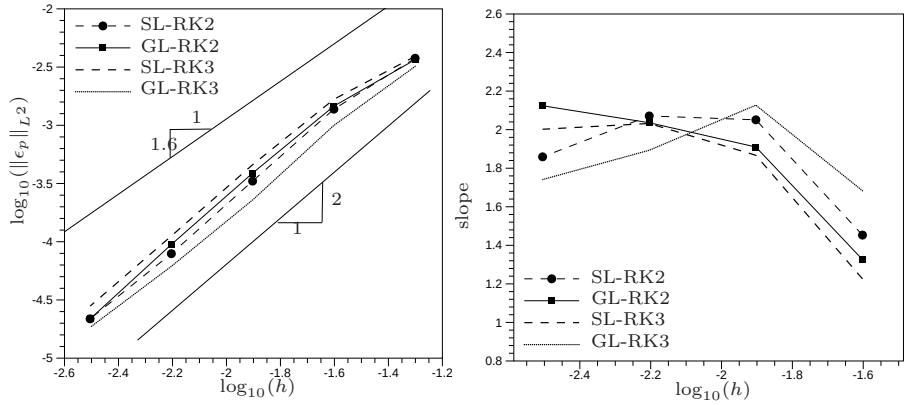


Figure 27: Vortex advection : grid convergence for the SU scheme with F1. Left column : convergence history. Right column : convergence rates.

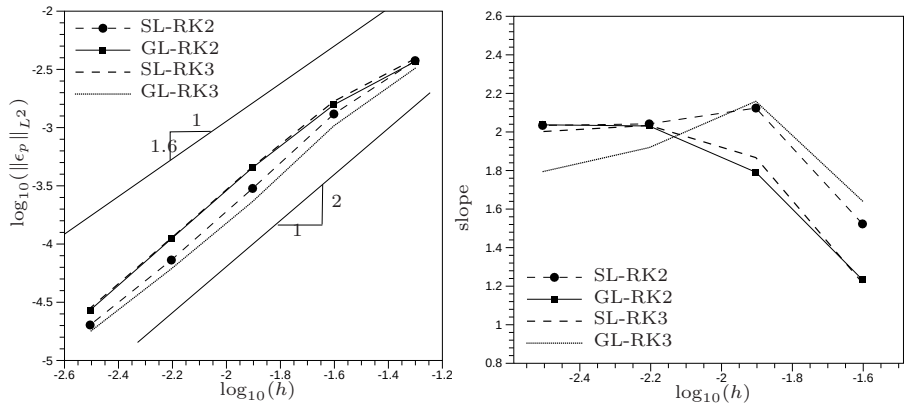


Figure 28: Vortex advection : grid convergence for the Bc scheme. Left column : convergence history. Right column : convergence rates.

7.2 Double Mach reflection

In this section we check the behavior of the nonlinear schemes in presence of a strong moving planar shock. The test case is that of a reflection of a Mach 10 oblique shock over a ramp proposed by Woodward and Colella in [38], to which we refer for details concerning the implementation. The computations have been run on an untructured triangulation with the topology shown on figure 3 and reference mesh size $h \approx 1/100$.

We display on figures 29 and 30 the density contours obtained with the B and Bc scheme, respectively. The first remark we can make is that even in presence of a strong moving shock both nonlinear discretizations yield quite smooth and non oscillatory results. To confirm this we report on figures 31 and 32 the density and entropy distributions across the shock and on the wall, respectively. From figure 31 we see that the shock is resolved very sharply. Only a small overshoot in its vicinity is observed in almost all the solutions. An

exception to this is the B scheme with RK3 time stepping and global lumping. In this case, as it can be also seen from the contour plot on figure 29 (bottom-right), we obtain some strange behavior in correspondence of the compression region where the bent incoming shock changes curvature. This seems to be a feature propagating from the upper boundary, where the exact shock movement is strongly imposed. This affects both the shock profile, as seen on figure 31, and the structures on the lower wall, as seen on figure 32. So far we have not been able to explain this behavior.

Apart from the above remarks, the solutions obtained are very satisfactory. The minumim and maximum values of the density, reported on table 1 also show that the minumim of the density is always very close to 1.4 (its analytical value), again with the exception of the B-GL-RK3 scheme.

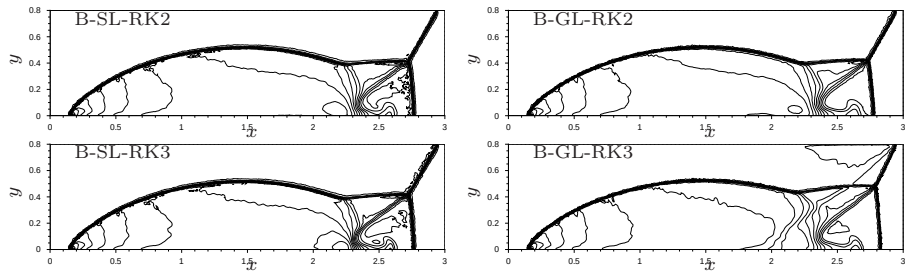


Figure 29: Double Mach reflection. Density contours for the B scheme. 30 equally spaced contours from 1 to 24. Top row : RK2. Bottom row : RK3. Left column : selective lumping. Right column : global lumping.

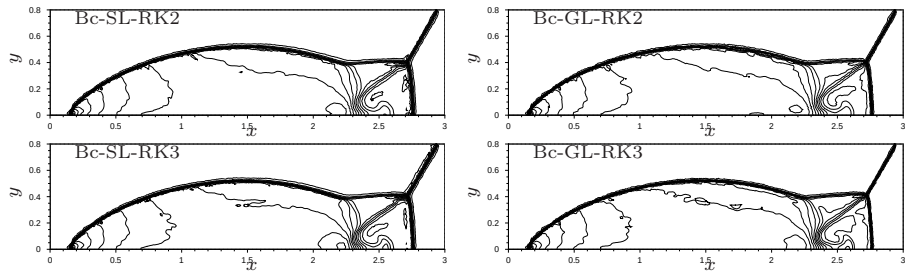


Figure 30: Double Mach reflection. Density contours for the Bc scheme. 30 equally spaced contours from 1 to 24. Top row : RK2. Bottom row : RK3. Left column : selective lumping. Right column : global lumping.

7.3 Mach 3 wind tunnel with a step

This final test is also taken from [38] and involves the formation and evolution of a moving shock in a Mach 3 wind tunnel with a step. We refer to [38] for details concerning the implementation of the test case. The mesh used for the computations is the same used in [13, 15]. A close up view in vicinity of the corner of the step is displayed on figure 33. The reference mesh size far from

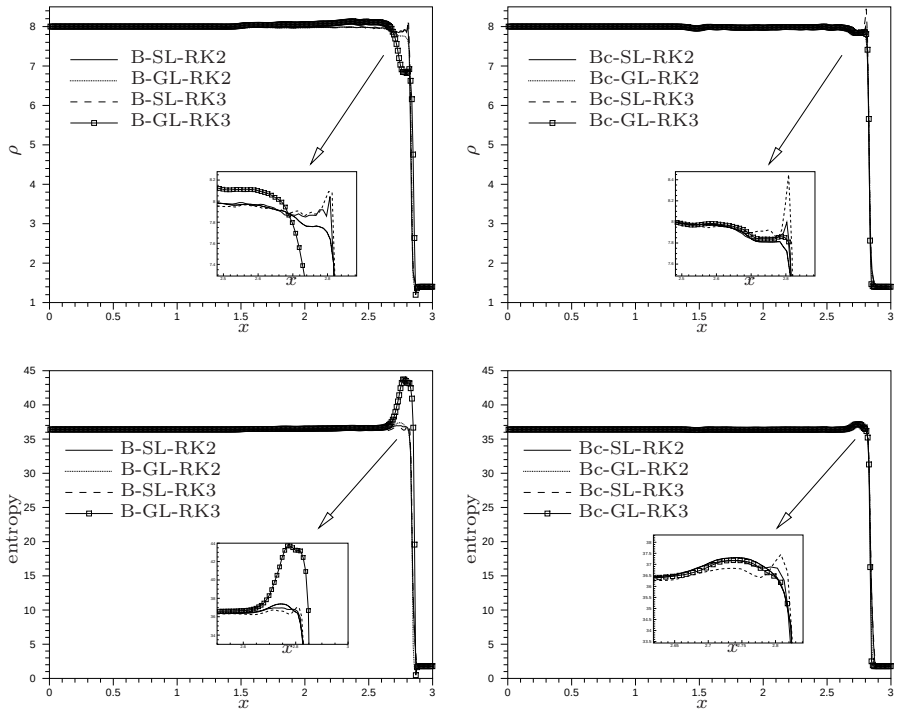


Figure 31: Double Mach reflection. Solution across the oblique shock ($y = 0.6$). Top row : density. Bottom row : entropy. Left column : B scheme. Right column : Bc scheme.

	ρ_{min}	$\rho_{\max}$
B-SL-RK2	1.40	22.30
B-SL-RK3	1.29	22.30
B-GL-RK2	1.37	22.30
B-GL-RK3	0.77	22.20

	ρ_{min}	$\rho_{\max}$
Bc-SL-RK2	1.398	24.12
Bc-SL-RK3	1.4	24.07
Bc-GL-RK2	1.397	24.00
Bc-GL-RK3	1.396	23.98

Table 1: Double Mach reflection : minimum and maximum values of the density. Left : B scheme. Right : Bc scheme

the corner is $h \approx 1/80$. The mesh is refined at the corner to attain a mimium size of $h \approx 1/1000$. No particular numerical treatment has been used near the corner to handle the supersonic expansion taking place during the transient.

The solutions obtained at times $t = 0.5$, $t = 1.5$, and $t = 4.0$ with the B and Bc schemes are shown on figures from 34 to 41. All the figures show a monotone and sharp resolution of the shocks, and of the contact lines. The non-oscillatory character of the results is confirmed by the line plots of the solution along the upper wall of the step (line $y = 0.2$ containing the corner singularity). We never obtained negative densities. Note that this is a test case where the

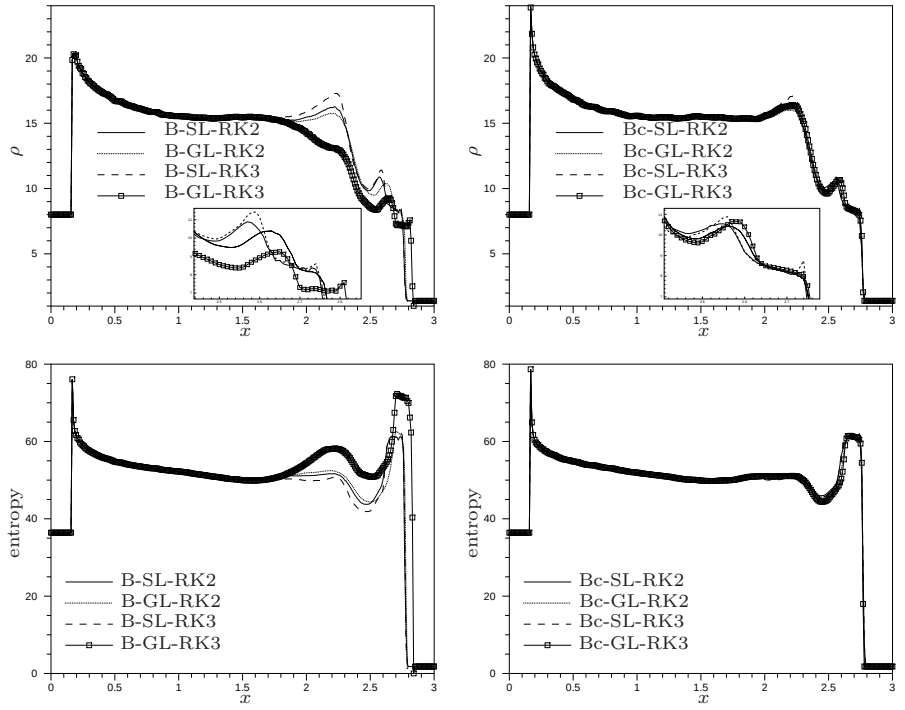


Figure 32: Double Mach reflection. Solution along the wall ($y = 0$). Top row : density. Bottom row : entropy. Left column : B scheme. Right column : Bc scheme.

explicit formulation does give an advantage with respect to the implicit schemes based on Crank-Nicholson time integration [4, 30]. Even if implicit in time, the positivity of the schemes proposed in the last references is still guaranteed by an explicit type time step restriction which, in presence of mesh refinement, renders the implicit formulation extremely time consuming.

As a last remark, we note that the Bc scheme with selective lumping yields a much better resolution of the flow, as seen for example from the kinks of the initial shock (top-left on figures 38 and 39), and from the resolution of the contact emanating from the interaction of the corner expansion with the reflected shock (middle-left on figures 38 and 39).

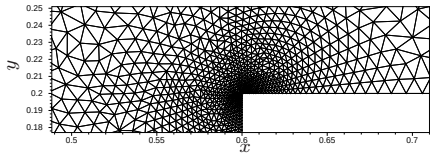
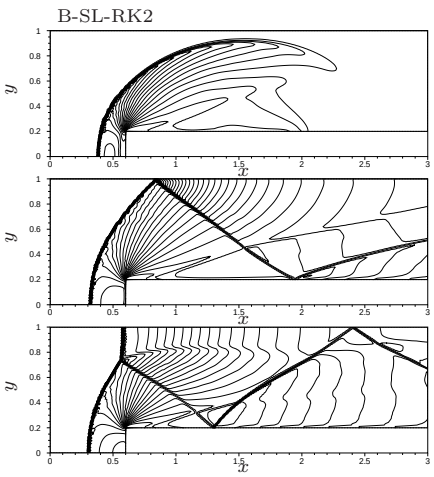
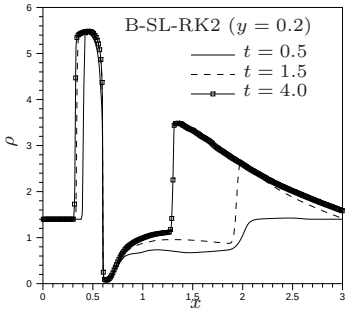


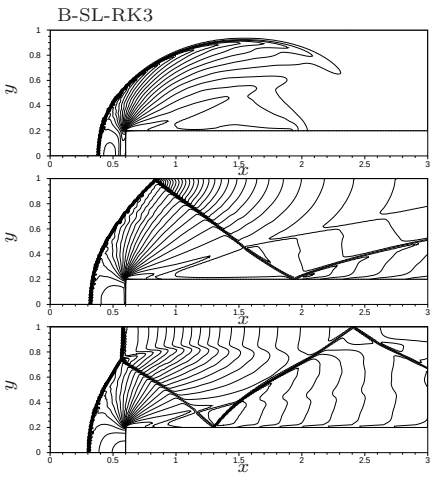
Figure 33: Mach 3 wind tunnel : close-up view of the mesh around the corner ($h = 1/80$ far from the corner, $h = 10^{-3}$ at the corner)

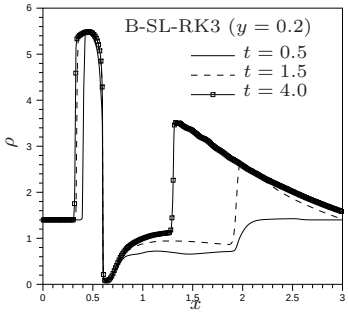




	$\rho_{\min}$	$\rho_{\max}$
$t = 0.5$	0.065	5.74
$t = 1.5$	0.067	8.29
$t = 4.0$	0.071	6.32

Figure 34: Mach 3 wind tunnel : B-SL-RK2 scheme. Left : density contours at time $t = 0.5$ (top), $t = 1.5$ (middle), and $t = 4.0$ (bottom) ; 30 equally spaced contours between 0.5 and 8. Right : density distribution along the line $y = 0.2$, and minimum and maximum values of the density.



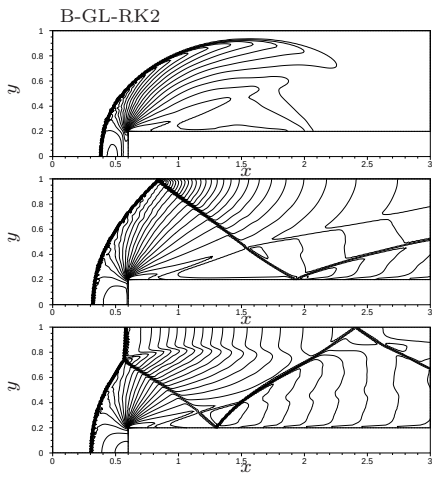


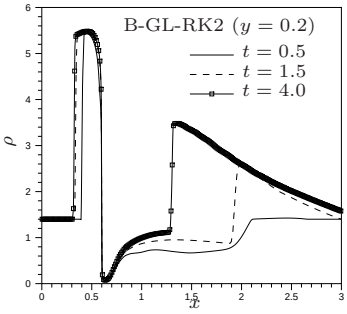
	$\rho_{\min}$	$\rho_{\max}$
$t = 0.5$	0.065	5.75
$t = 1.5$	0.067	8.19
$t = 4.0$	0.071	6.31

Figure 35: Mach 3 wind tunnel : B-SL-RK3 scheme. Left : density contours at time $t = 0.5$ (top), $t = 1.5$ (middle), and $t = 4.0$ (bottom) ; 30 equally spaced contours between 0.5 and 8. Right : density distribution along the line $y = 0.2$, and minimum and maximum values of the density.

8 Conclusions

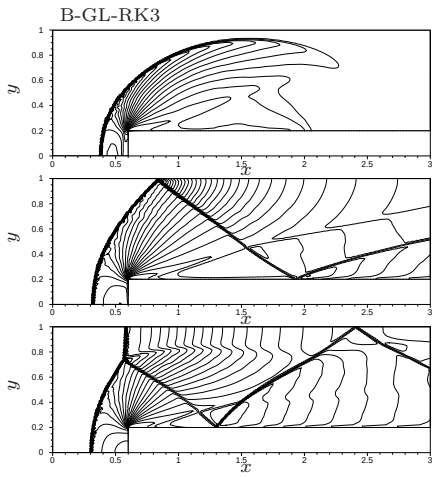
In this paper we have provided a construction of explicit second order Residual Distribution schemes based on Runge-Kutta time integration. We used second order mass lumping and a finite element interpretation to achieve a discretization where the consistent RD mass matrix does not multiply the solution at the

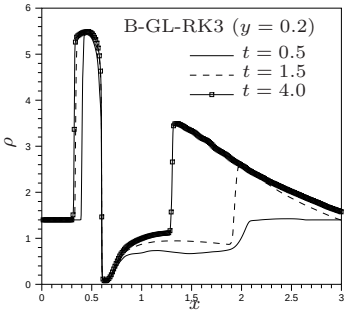




	$\rho_{\min}$	$\rho_{\max}$
$t = 0.5$	0.062	5.73
$t = 1.5$	0.067	7.89
$t = 4.0$	0.072	6.29

Figure 36: Mach 3 wind tunnel : B-GL-RK2 scheme. Left : density contours at time $t = 0.5$ (top), $t = 1.5$ (middle), and $t = 4.0$ (bottom) ; 30 equally spaced contours between 0.5 and 8. Right : density distribution along the line $y = 0.2$, and minimum and maximum values of the density.



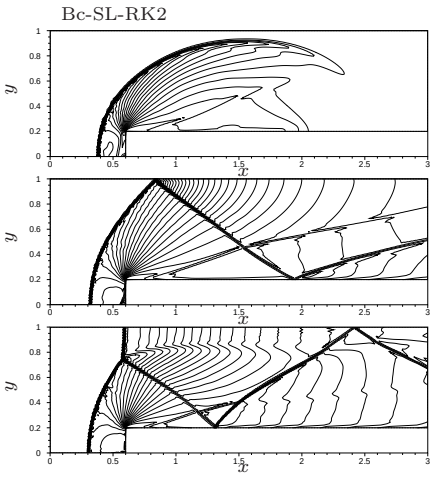


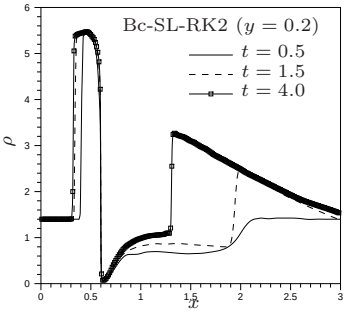
	$\rho_{\min}$	$\rho_{\max}$
$t = 0.5$	0.062	5.73
$t = 1.5$	0.067	7.86
$t = 4.0$	0.072	6.30

Figure 37: Mach 3 wind tunnel : B-GL-RK3 scheme. Left : density contours at time $t = 0.5$ (top), $t = 1.5$ (middle), and $t = 4.0$ (bottom) ; 30 equally spaced contours between 0.5 and 8. Right : density distribution along the line $y = 0.2$, and minimum and maximum values of the density.

new time level thus allowing for a truly explicit solution procedure. All the theoretical arguments justifying our contruction have been thouroughly exposed, and strong numerical evidence has been given to confirm them.

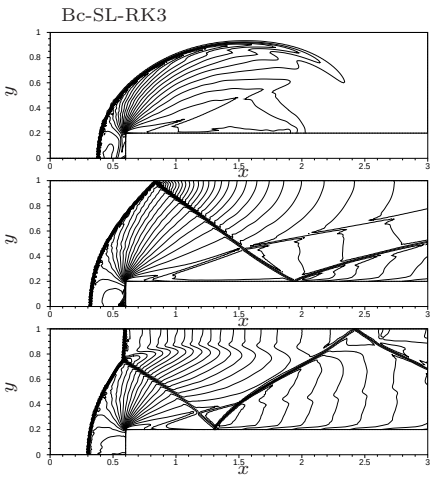
The results obtained are very encouraging both concerning accuracy, and monotonicity. We think this work paves the way for a different class of RD schemes based on explicit, or mixed, time integration where the RD mass ma-

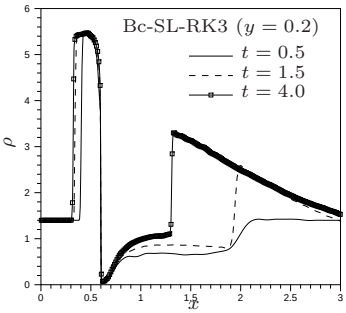




	$\rho_{\min}$	$\rho_{\max}$
$t = 0.5$	0.049	5.69
$t = 1.5$	0.065	8.04
$t = 4.0$	0.059	6.25

Figure 38: Mach 3 wind tunnel : Bc-SL-RK2 scheme. Left : density contours at time $t = 0.5$ (top), $t = 1.5$ (middle), and $t = 4.0$ (bottom) ; 30 equally spaced contours between 0.5 and 8. Right : density distribution along the line $y = 0.2$, and minimum and maximum values of the density.





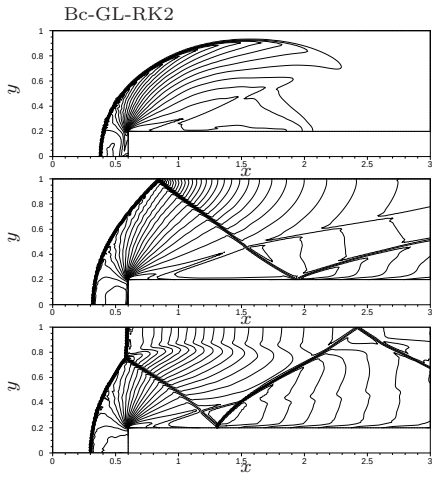
	$\rho_{\min}$	$\rho_{\max}$
$t = 0.5$	0.058	5.70
$t = 1.5$	0.054	8.17
$t = 4.0$	0.052	6.28

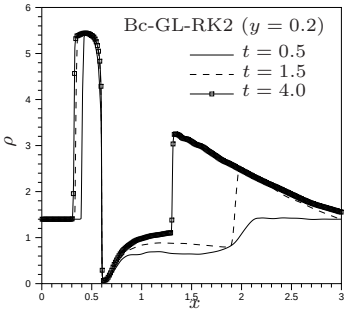
Figure 39: Mach 3 wind tunnel : Bc-SL-RK3 scheme. Left : density contours at time $t = 0.5$ (top), $t = 1.5$ (middle), and $t = 4.0$ (bottom) ; 30 equally spaced contours between 0.5 and 8. Right : density distribution along the line $y = 0.2$, and minimum and maximum values of the density.

trix does not necessarily need to be inverted.

Concerning the developments of the work reported in this paper, we mention the following points :

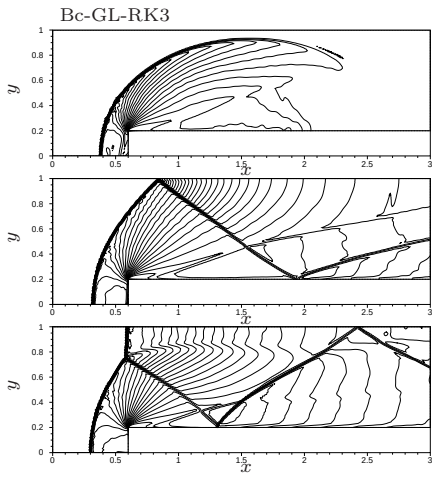
- We are currently performing a Fourier analysis on structured triangulations to better undertand the linear stability properties of the linear

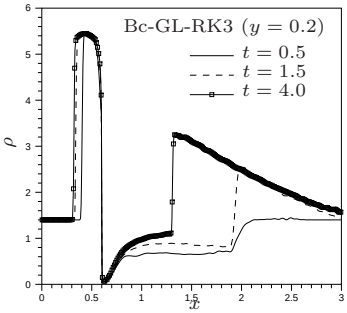




	$\rho_{\min}$	$\rho_{\max}$
$t = 0.5$	0.062	5.68
$t = 1.5$	0.059	7.88
$t = 4.0$	0.058	6.31

Figure 40: Mach 3 wind tunnel : Bc-GL-RK2 scheme. Left : density contours at time $t = 0.5$ (top), $t = 1.5$ (middle), and $t = 4.0$ (bottom) ; 30 equally spaced contours between 0.5 and 8. Right : density distribution along the line $y = 0.2$, and minimum and maximum values of the density.





	$\rho_{\min}$	$\rho_{\max}$
$t = 0.5$	0.055	5.70
$t = 1.5$	0.063	7.87
$t = 4.0$	0.045	6.33

Figure 41: Mach 3 wind tunnel : Bc-GL-RK3 scheme. Left : density contours at time $t = 0.5$ (top), $t = 1.5$ (middle), and $t = 4.0$ (bottom) ; 30 equally spaced contours between 0.5 and 8. Right : density distribution along the line $y = 0.2$, and minimum and maximum values of the density.

schemes, when using different forms of hte mass matrix, and also to understand the influence of the type of lumping on the stability of the resulting scheme ;

- Even though we judge the results presented here quite satisfactory (especially for the Euler equations) there is definitely space to improve the

nonlinear schemes, taking into account the fully discrete RK time stepping and the terms arising from mass lumping in the positivity analysis ;

- An immediate application of the explicit RK-RD schemes is given by te Shallow Water equations where the preservation of the positivity of the depth leads, for the standard implicit RD, to a strict contraint on the time step [29]. We think the RK-RD approach proposed would represent an improvement, still preserving most of the nice properties of the RD discretization ;
- The extension to more than second order should be relatively stright forward when making use of higher order elements allowing higher order mass lumping. From this point of view we will profit of the work that has been done on the wave equation (see *e.g.* [21, 36] and references therein).

Acknowledgements

We are grateful to G. Scovazzi (SANDIA NL, Albuquerque) for his numerous remarks on the solution of the nonlinear Newton loop in RD discretizations, and for pointing out the relations with predictor-multicorrector schemes used in SUPG discretizations.

Appendix 1 : proof of proposition 4.3

We derive a sufficient condition on the $\overline{r}^k$ guaranteeing that the accuracy of the Runge-Kutta Galerkin approximation is not lost when adding the bubble contribution. To do this we use a truncation error analysis, following the approach of [28]. What we want to do is to show that for a p-th order spatial approximation, and when employing a p-th order RK scheme, the solution at time t^n verifies a truncation error equation of the type $\mathcal{E}_n \leq C_n h^p$. A global space-time truncation error estimate can then be obtained by integrating over each time interval and adding up over all the time slabs.

To do this we consider a more general polynomial approximation in space. The degrees of freedom are still approximations of the values of the unknown in some nodal locations of the mesh, except that, differently from the P^1 case, these are the element vertices plus other locations, as for example in standard P^k elements, or in more "exotic" polynomial spaces, such as the ones proposed in [21, 36, 35]. As before, we denote by φ_i the basis functions spanning the polynomial space, by u_h the spatial approximation of a function u, and with K we denote the total number of degrees of freedom (DoF) contained in an element of the mesh.

We start by recalling that, given a smooth classical solution of the problem w, hypothesis 4.1 guarantees that a p-th order RK scheme verifies the truncation error estimate

$$r^{n+1}(w) = \frac{\delta w^{n+1}}{\Delta t} + f^{n+1}(w) = C_{\mathrm{RK}}(w^n)\Delta t^p$$

where, with the notation of section §4.1, we have explicitely used the fact that for the last stage of the RK scheme $r^k = r^{n+1}$. Similarly, hypothesis 4.2 ensures

that the modified semi-discrete operator used in the bubble function verifies the estimate

$$\overline{r}^{n+1}(w) = \frac{\overline{\delta w^{n+1}}}{\Delta t} + f^{n+1}(w) = \overline{C}_{\rm RK}(w^n)\Delta t^p$$

where certainly $l \leq p$. Both hypotheses are verified in appendix 2 for the RK2 and RK3 schemes considered in the paper.

Next we define, for the stabilized Galerkin scheme, the following truncation error :

$$\mathcal{E}_n = \left| \sum_{i\in\mathcal{T}_h} \psi_i \int_\Omega \varphi_i \, r^{n+1}(w_h) \, dx\, dy + \sum_{i\in\mathcal{T}_h} \psi_i \sum_{T|i\in T} \int_T \gamma_i \, \overline{r^{n+1}}(w_h) \, dx\, dy \right| \quad (66)$$

where the ψ_is are nodal values of a $C_0^1(\Omega)$ function, which is assumed to verify [28]

$$\|\psi\|_{L^\infty(\Omega)} \leq C_\psi \,, \quad \|\psi_h\|_{L^\infty(\Omega)} \leq C_{\psi_h} \,; \quad \|\nabla\psi\|_{L^\infty(\Omega)} \leq C_{\nabla\psi} \,, \quad \|\nabla\psi_h\|_{L^\infty(\Omega)} \leq C_{\nabla\psi_h} \quad (67)$$

having denoted by ψ_h the p-th order polynomial approximation of ψ corresponding to the approximation space chosen. Similarly, w_h represents the spatial interpolant of the given smooth exact solution w. As a second step, we can immediately rewite the error as

$$\mathcal{E}_n = \left| \overbrace{\int_\Omega \psi_h \, r^{n+1}(w_h) \, dx\, dy}^{I} + \overbrace{\sum_{T\in\mathcal{T}_h} \sum_{j\in T} \int_T \gamma_j \psi_j \, \overline{r^{n+1}}(w_h) \, dx\, dy}^{II} \right|$$

Next we estimate the two terms in the error. For I we immediately make use of hypothesis 4.1 :

$$I = \int_\Omega \psi_h \left(\partial_t w_h^n + \nabla\cdot\boldsymbol{\mathcal{F}}_h(w_h^n)\right) dx\, dy + \int_\Omega \psi_h C_{\rm RK}(w_h^n)\Delta t^p \, dx\, dy$$

Now, being w a smooth exact solution, we have $\partial_t w^n + \nabla\cdot\boldsymbol{\mathcal{F}}(u^n) = 0$, hence I can be rewritten as

$$\begin{aligned} I &= \int_\Omega \psi_h \partial_t (w_h^n - w^n) dx\, dy + \int_\Omega \psi_h \nabla\cdot\left(\boldsymbol{\mathcal{F}}_h(w_h^n) - \boldsymbol{\mathcal{F}}(w^n)\right) dx\, dy + \int_\Omega \psi_h C_{\rm RK}(w_h^n)\Delta t^p \, dx\, dy \\ &= \int_\Omega \psi_h \partial_t (w_h^n - w^n) dx\, dy - \int_\Omega \left(\boldsymbol{\mathcal{F}}_h(w_h^n) - \boldsymbol{\mathcal{F}}(w^n)\right)\cdot\nabla\psi_h dx\, dy + \int_\Omega \psi_h C_{\rm RK}(w_h^n)\Delta t^p \, dx\, dy \end{aligned}$$

where, following [6, 28], we have broken the second integral over elements, integrated by parts over each element, re-assembled, and used the fact that $\psi \in C_0^1(\Omega)$. Due to the assumptions on ψ (cf. equation (67)), we can now use the properties of the approximation to estimate all terms, ending up with

$$|I| \leq \mathcal{C}_w(\mathcal{T}_h, w^n)h^p + \mathcal{C}_{\boldsymbol{\mathcal{F}}}(\mathcal{T}_h, w^n)h^p + \mathcal{C}_{\rm RK}(\mathcal{T}_h, w^n)\Delta t^p = \mathcal{C}_0(\mathcal{T}_h, w^n)h^p + \mathcal{C}_{\rm RK}(\mathcal{T}_h, w^n)\Delta t^p \quad (68)$$

This term is nothing else than the truncation error of the Galerkin scheme. As expected, it is of an order dictated purely by the spatial and temporal approximations.

We now estimate the term II. First of all we note that since $\sum_j \gamma_j = 0$, then we can write

$$II = \frac{1}{K} \sum_{T\in\mathcal{T}_h} \sum_{j\in T} \sum_{i\in T} \int_T \gamma_j(\psi_j - \psi_i)\, \overline{r^{n+1}}(w_h)\, dx\, dy$$

where we racall that K denotes the number of DoF in an element. Next we use hypothesis 4.2 to get

$$\begin{aligned} II =& \frac{1}{K} \sum_{T\in\mathcal{T}_h} \sum_{j\in T} \sum_{i\in T} \int_T \gamma_j(\psi_j - \psi_i)\partial_t w_h^n\, dx\, dy \\ &+ \frac{1}{K} \sum_{T\in\mathcal{T}_h} \sum_{j\in T} \sum_{i\in T} \int_T \gamma_j(\psi_j - \psi_i)\nabla\cdot \boldsymbol{\mathcal{F}}_h(w_h^n)\, dx\, dy \\ &+ \frac{1}{K} \sum_{T\in\mathcal{T}_h} \sum_{j\in T} \sum_{i\in T} \int_T \gamma_j(\psi_j - \psi_i)\overline{C}_{\mathrm{RK}}(w^n)\Delta t^l\, dx\, dy \end{aligned}$$

Using again the fact that w is a classical solution we have

$$\begin{aligned} II =& \frac{1}{K} \sum_{T\in\mathcal{T}_h} \sum_{j\in T} \sum_{i\in T} \int_T \gamma_j(\psi_j - \psi_i)\partial_t (w_h^n - w^n)\, dx\, dy \\ &+ \frac{1}{K} \sum_{T\in\mathcal{T}_h} \sum_{j\in T} \sum_{i\in T} \int_T \gamma_j(\psi_j - \psi_i)\nabla\cdot (\boldsymbol{\mathcal{F}}_h(w_h^n) - \boldsymbol{\mathcal{F}}(w^n))\, dx\, dy \\ &+ \frac{1}{K} \sum_{T\in\mathcal{T}_h} \sum_{j\in T} \sum_{i\in T} \int_T \gamma_j(\psi_j - \psi_i)\overline{C}_{\mathrm{RK}}(w^n)\Delta t^l\, dx\, dy \end{aligned}$$

In order to give an upper bound to the last expression, we make use of the fact that in 2D the total number of elements in the mesh can be bounded by h^{-2}, the properties of ψ to deduce that $\psi_j - \psi_i$ can be bounded by $\|\nabla\psi\|_{L^\infty(\Omega)} h$, the fact that $|T| \le C_0 h^2$, and the properties of the approximation. This leads to

$$\begin{aligned} |II| \le& \mathcal{C}(\Omega,\mathcal{T}_h)\, h^{-2}\, \|\gamma\|_{L^\infty(\Omega)}\, C_0 h^2\, \|\nabla\psi\|_{L^\infty(\Omega)}\, h\, \mathcal{C}_w(\mathcal{T}_h, w^n)\, h^p \\ &+ \mathcal{C}(\Omega,\mathcal{T}_h)\, h^{-2}\, \|\gamma\|_{L^\infty(\Omega)}\, C_0 h^2\, \|\nabla\psi\|_{L^\infty(\Omega)}\, h\, \mathcal{C}_{\nabla\boldsymbol{\mathcal{F}}}(\mathcal{T}_h, w^n)\, h^{p-1} \\ &+ \mathcal{C}(\Omega,\mathcal{T}_h)\, h^{-2}\, \|\gamma\|_{L^\infty(\Omega)}\, C_0 h^2\, \|\nabla\psi\|_{L^\infty(\Omega)}\, h\, \overline{C}_{\mathrm{RK}}(w^n)\Delta t^l \end{aligned}$$

With $\|\gamma\|_{L^\infty(\Omega)} = \max_{T\in\mathcal{T}_h} \max_{j\in T} \|\gamma_j\|_{L^\infty(T)}$. Setting $\mathcal{C}_1(\Omega,\mathcal{T}_h,w^n) = \mathcal{C}(\Omega,\mathcal{T}_h) C_0 C_{\nabla\psi} \max(\mathcal{C}_w(\mathcal{T}_h,w^n), \mathcal{C}_{\nabla\boldsymbol{\mathcal{F}}}(\mathcal{T}_h,w^n$ and $\overline{\mathcal{C}}_{\mathrm{RK}}(\Omega,\mathcal{T}_h,w^n) = \mathcal{C}(\Omega,\mathcal{T}_h) C_0 C_{\nabla\psi} \overline{C}_{\mathrm{RK}}(w^n)$ we get the estimate

$$|II| \le \|\gamma\|_{L^\infty(\Omega)} \left(\mathcal{C}_1(\Omega,\mathcal{T}_h,w^n)\, h^p + \overline{\mathcal{C}}_{\mathrm{RK}}(\Omega,\mathcal{T}_h,w^n)\, h\, \Delta t^l\right)$$

Assembling the Galerkin error and the error associated to the bubble, we get finally

$$\mathcal{E}_n \le \mathcal{C}_0(\mathcal{T}_h,w^n)\, h^p + \mathcal{C}_{\mathrm{RK}}(\mathcal{T}_h,w^n)\, \Delta t^p + \|\gamma\|_{L^\infty(\Omega)}\, \mathcal{C}_1(\Omega,\mathcal{T}_h,w^n)\, h^p + \|\gamma\|_{L^\infty(\Omega)}\, \overline{\mathcal{C}}_{\mathrm{RK}}(\Omega,\mathcal{T}_h,w^n)\, h\, \Delta t^l \tag{69}$$

This immediately shows that, provided that the bubble functions are uniformly bounded, we are allowed to have $l \le p$, in particular, it is enough to take $l = p-1$ to retain the accuracy of the Galerkin approximation.

In particular, if, as it is always the case for explicit schemes, we can find two positive bounded constants $C_{h/\Delta t}$ and $C_{\Delta t/h}$ such that

$$C_{h/\Delta t} \leq \frac{\Delta t}{h} \leq C_{\Delta t/h}$$

then we have for $l = p - 1$

$$\mathcal{E}_n \leq \mathcal{C}\ h^p \tag{70}$$

with

$$\mathcal{C} = \mathcal{C}_0(\mathcal{T}_h, w^n) + \mathcal{C}_{\rm RK}(\mathcal{T}_h, w^n)\, C^p_{\Delta t/h} + \|\gamma\|_{L^\infty(\Omega)} \left(\mathcal{C}_1(\Omega, \mathcal{T}_h, w^n) + \overline{\mathcal{C}}_{\rm RK}(\Omega, \mathcal{T}_h, w^n) C^{p-1}_{\Delta t/h}\right)$$

Note that, mass lumping is kept out of the analysis. However, as shown in section §3, it can be included in the definition of the (bounded) bubble function, at least in the P^1 case. For the higher order case, we refer to [21, 36, 35] for more.

Appendix 2 : hypotheses 4.1 and 4.2

In this appendix we justify the choice of the approximate time incerements $\overline{\delta u^k}$ (cf. section §4.1 and appendix 1) for the RK2 and RK3 schemes of section §4.1. We recall that the constraint to respect is that for the last RK step

$$\overline{r}^k = \frac{\overline{\delta u^k}}{\Delta t} + \overline{f}^k = \mathcal{O}(\Delta t^l)$$

with $l \geq p - 1$, where with p we denote the (desired) overall accuracy of the scheme. The analysis will be performed for the autonomous ODE :

$$\partial_t u + f(u) = 0 \tag{71}$$

RK2 scheme. Let us start by verifying hypothesis 4.1 for the RK2 scheme defined by

$$\begin{aligned} u^1 =& u^n - \Delta t f(u^n) \\ u^{n+1} =& u^n - \frac{\Delta t}{2} f(u^n) - \frac{\Delta t}{2} f(u^1) \end{aligned}$$

When replacing u^n and u^{n+1} by the values at t^n and t^{n+1} of an exact solution $w(t)$, and using the fact that $w^1 = w^n - \Delta t f(w^n)$, we can write

$$\begin{aligned} f(w^1) =& f(w^n + (w^1 - w^n)) = f(w^n) + (w^1 - w^n)\partial_u f(u^n) + \frac{(w^1 - w^n)^2}{2}\partial_{uu} f(u^n) + \mathcal{O}\big((w^1 - w^n)^3\big) \\ =& f(w^n) - \Delta t f(u^n)\partial_u f(u^n) + \frac{\Delta t^2}{2} f(u^n)^2 \partial_{uu} f(u^n) + \mathcal{O}(\Delta t^3) \end{aligned}$$

which immediately leads to

$$\begin{aligned} \frac{w^{n+1} - w^n}{\Delta t} + \frac{1}{2}\Big(f(u^n) + f(u^1)\Big) =& \partial_t w^n + \frac{\Delta t}{2}\partial_{tt} w^n + \frac{\Delta t^2}{6}\partial_{ttt} w^n \\ &+ f(w^n) - \frac{\Delta t}{2} f(u^n)\partial_u f(u^n) + \frac{\Delta t^2}{4} f(u^n)^2 \partial_{uu} f(u^n) + \mathcal{O}(\Delta t^3) \\ =& \partial_t w^n + f(w^n) - \frac{\Delta t^2}{3}\left(\frac{1}{2} f(w^n)\partial_u f(w^n)^2 + f(w^n)^2 \partial_{uu} f(w^n)\right) + \mathcal{O}(\Delta t^3) \end{aligned}$$

having used the relations

$$\begin{aligned}
\partial_t w^n =& - f(w^n)\\
\partial_{tt} w^n =& f(w^n)\partial_u f(w^n)\\
\partial_{ttt} w^n =& - f(w^n)\partial_u f(w^n)^2 - f(w^n)^2\partial_{uu} f(w^n)\\
\partial_{tttt} w^n =& f(w^n)\partial_u f(w^n)^3 + 4f(w^n)^2\partial_u f(w^n)\partial_{uu} f(w^n) + f(w^n)^3\partial_{uuu} f(w^n)
\end{aligned} \tag{72}$$

To verify hypothesis 4.2, we perform a similar exercise :

$$\begin{aligned}
\overline{r}^{n+1} =& \frac{w^1 - w^n}{\Delta t} + \frac{1}{2}\Big(f(w^n) + f(w^1)\Big) = -f(w^n) + \frac{1}{2}\Big(f(w^n) + f(w^1)\Big) =\\
& - \frac{1}{2}f(w^n) + \frac{1}{2}f(w^n) - \frac{\Delta t}{2}f(w^n)\partial_u f(w^n) + \mathcal{O}(\Delta t^2) = -\frac{\Delta t}{2}f(w^n)\partial_u f(w^n) + \mathcal{O}(\Delta t^2) = \mathcal{O}(\Delta t)
\end{aligned}$$

proving that this definition of $\overline{r}^{n+1}$ is enough for use in the stabilization term in second order schemes.

RK3 scheme. We repeat the same exercise for the RK3 scheme defined by

$$\begin{aligned}
u^1 =& u^n - \Delta t f(u^n)\\
u^2 =& u^n - \frac{\Delta t}{4}f(u^n) - \frac{\Delta t}{4}f(u^1)\\
u^{n+1} =& u^n - \frac{\Delta t}{6}f(u^n) - \frac{2\Delta t}{3}f(u^2) - \frac{\Delta t}{6}f(u^1)
\end{aligned}$$

When replacing u^n and u^{n+1} by the values at t^n and t^{n+1} of an exact solution $w(t)$, we can easily prove the following developments

$$\begin{aligned}
f(w^1) =& f(w^n) - \Delta t f(w^n)\partial_u f(w^n) + \frac{\Delta t^2}{2}f(w^n)^2\partial_{uu} f(w^n) - \frac{\Delta t^3}{6}f(w^n)^3\partial_{uuu} f(w^n) + \mathcal{O}(\Delta t^4)\\
f(w^2) =& f(w^n) - \frac{\Delta t}{2}f(w^n)\partial_u f(w^n) + \frac{\Delta t^2}{4}\left(f(w^n)\partial_u f(w^n)^2 + \frac{1}{2}f(w^n)^2\partial_{uu} f(w^n)\right)\\
& - \frac{\Delta t^3}{8}\left(2f(w^n)^2\partial_u f(w^n)\partial_{uu} f(w^n) + \frac{1}{6}f(w^n)^3\partial_{uuu} f(w^n)\right) + \mathcal{O}(\Delta t^4)
\end{aligned}$$

These developments can be readily used to show that

$$\begin{aligned}
\frac{w^{n+1} - w^n}{\Delta t} +& \frac{1}{6}f(w^n) + \frac{1}{6}f(w^1) + \frac{2}{3}f(w^2) = \partial_t w^n + \frac{\Delta t}{2}\partial_{tt} w^n + \frac{\Delta t^2}{6}\partial_{ttt} w^n + \frac{\Delta t^3}{24}\partial_{tttt} w^n + \frac{1}{6}f(w^n)\\
& + \frac{1}{6}f(w^n) - \frac{\Delta t}{6}f(w^n)\partial_u f(w^n) + \frac{\Delta t^2}{12}f(w^n)^2\partial_{uu} f(w^n) - \frac{\Delta t^3}{36}f(w^n)^3\partial_{uuu} f(w^n)\\
& + \frac{2}{3}f(w^n) - \frac{\Delta t}{3}f(w^n)\partial_u f(w^n) + \frac{\Delta t^2}{6}\left(\frac{1}{2}f(w^n)^2\partial_{uu} f(w^n) + f(w^n)\partial_u f(w^n)^2\right)\\
& - \frac{\Delta t^3}{12}\left(2f(w^n)^2\partial_u f(w^n)\partial_{uu} f(w^n) + \frac{1}{6}f(w^n)^3\partial_{uuu} f(w^n)\right) + \mathcal{O}(\Delta t^4)
\end{aligned}$$

which, using (72), leads immediately to

$$\frac{w^{n+1} - w^n}{\Delta t} + \frac{1}{6}f(w^n) + \frac{1}{6}f(w^1) + \frac{2}{3}f(w^2) = \partial_t w^n + f(w^n) + \frac{\Delta t^3}{12}f(w^n)\partial_u f(w^n)^3 + \mathcal{O}(\Delta t^4) = \mathcal{O}(\Delta t^3)$$

A similar exercise can be used now to show that

$$\begin{aligned}\overline{r}^2(w) =& \frac{w^1-w^n}{2\Delta t}+\frac{1}{4}f(w^n)+\frac{1}{4}f(w^1)= \\ & -\frac{1}{2}f(w^n)+\frac{1}{4}f(w^n)+\frac{1}{4}f(w^n)-\frac{\Delta t}{4}f(w^n)\partial_u f(w^n)+\mathcal{O}(\Delta t^2)= \\ & -\frac{\Delta t}{4}f(w^n)\partial_u f(w^n)+\mathcal{O}(\Delta t^2)=\mathcal{O}(\Delta t)\end{aligned}$$

and more importantly that

$$\begin{aligned}\overline{r}^{n+1}(w) =& 2\frac{w^2-w^n}{\Delta t}+\frac{1}{6}f(w^n)+\frac{1}{6}f(w^1)+\frac{2}{3}f(w^2)= \\ & \frac{2}{\Delta t}\left(-\frac{\Delta t}{2}f(w^n)+\frac{\Delta t^2}{4}f(w^n)\partial_u f(w^n)-\frac{\Delta t^3}{8}f(w^n)^2\partial_{uu}f(w^n)\right)+ \\ & \frac{1}{6}f(w^n)+\frac{1}{6}f(w^n)-\frac{\Delta t}{6}f(w^n)\partial_u f(w^n)+\frac{\Delta t^2}{12}f(w^n)^2\partial_{uu}f(w^n)+ \\ & \frac{2}{3}f(w^n)-\frac{\Delta t}{3}f(w^n)\partial_u f(w^n)+\frac{\Delta t^2}{6}\left(f(w^n)\partial_u f(w^n)^2+\frac{1}{2}f(w^n)^2\partial_{uu}f(w^n)\right)+\mathcal{O}(\Delta t^3)= \\ & \frac{\Delta t^2}{6}\left(f(w^n)\partial_u f(w^n)^2-\frac{1}{2}f(w^n)^2\partial_{uu}f(w^n)\right)+\mathcal{O}(\Delta t^3)=\mathcal{O}(\Delta t^2)\end{aligned}$$

which shows that also for the RK3 scheme, our definitions of the $\overline{r}^k$ do verify the accuracy constraint.

References

[1] R. Abgrall. Toward the ultimate conservative scheme : Following the quest. *J. Comput. Phys*, 167(2):277–315, 2001.

[2] R. Abgrall. Essentially non oscillatory residual distribution schemes for hyperbolic problems. *J. Comput. Phys*, 214(2):773–808, 2006.

[3] R. Abgrall. Residual distribution schemes: Current status and future trends. *Computers and Fluids*, 35(7):641–669, 2006.

[4] R. Abgrall and M. Mezine. Construction of second order accurate monotone and stable residual distribution schemes for unsteady flow problems. *J. Comput. Phys.*, 188:16–55, 2003.

[5] R. Abgrall and M. Mezine. Construction of second-order accurate monotone and stable residual distribution schemes for steady flow problems. *J. Comput. Phys.*, 195:474–507, 2004.

[6] R. Abgrall and P.L. Roe. High order fluctuation schemes on triangular meshes. *J. Sci. Comput.*, 19(3):3–36, 2003.

[7] C. Bolley and M. Crouzeix. Conservation de la positivité lors de la discétization des problèmes d'èvolution paraboliques. *R.A.I.R.O. Analyse Numérique*, 12:237–254, 1978.

[8] D. Caraeni and L. Fuchs. Compact third–order multidimensional upwind scheme for navier stokes simulations. *Theoretical and Computational Fluid Dynamics*, 15:373–401, 2002.

[9] D.A. Caraeni. *Development of a Multidimensional Upwind Residual Distribution Solver for Large Eddy Simulation of Industrial Turbulent Flows*. PhD thesis, Lund Institute of Technology, 2000.

[10] C.Corre and X.Du. A residual-based scheme for computing compressible flows on unstructured grids. *Computers and Fluids*, 38(7):1338–1347, 2009.

[11] C.-S. Chou and C.-W. Shu. High order residual distribution conservative finite difference weno schemes for steady state problems on non-smooth meshes. *J. Comput. Phys.*, 214(3):698–724, 2006.

[12] C. Corre, G. Hanss, and A. Lerat. A residual-based compact scheme for the unsteady compressible Navier-Stokes equations. *Comput. Fluids*, 34(4-5):561–580, 2005.

[13] Á. Csík and H. Deconinck. Space time residual distribution schemes for hyperbolic conservation laws on unstructured linear finite elements. *International Journal for Numerical Methods in Fluids*, 40:573–581, 2002.

[14] Á. Csík, M. Ricchiuto, and H. Deconinck. A conservative formulation of the multidimensional upwind residual distribution schemes for general nonlinear conservation laws. *J. Comput. Phys*, 179(2):286–312, 2002.

[15] Á. Csík, M. Ricchiuto, H. Deconinck, and S. Poedts. Space-time residual distribution schemes for hyperbolic conservation laws. 15th AIAA Computational Fluid Dynamics Conference, Anahein, CA, USA, June 2001.

[16] H. Deconinck and M. Ricchiuto. Residual distribution schemes: foundation and analysis. In E. Stein, R. de Borst, and T.J.R. Hughes, editors, *Encyclopedia of Computational Mechanics*. John Wiley & Sons, Ltd., 2007. DOI: 10.1002/0470091355.ecm054.

[17] H. Deconinck, K. Sermeus, and R. Abgrall. Status of multidimensional upwind residual distribution schemes and applications in aeronautics. AIAA paper 2000-2328, June 2000. AIAA CFD Conference, Denver (USA).

[18] P. De Palma, G. Pascazio, G. Rossiello, and M. Napolitano. A second-order accurate monotone implicit fluctuation splitting scheme for unsteady problems. *J. Comput. Phys*, 208(1):1–33, 2005.

[19] J. Dobes and H. Deconinck. Second order blended multidimensional upwind residual distribution scheme for steady and unsteady computations. *J.Comput.Appl.Math*, 215(1):378–389, 2006.

[20] A. Ferrante and H. Deconinck. Solution of the unsteady Euler equations using residual distribution and flux corrected transport. Technical Report VKI-PR 97-08, von Karman Institute for Fluid Dynamics, 1997.

[21] G.Cohen, P.Joly, J.E.Roberts, and N.Tordjman. High order triangular finite elements with mass lumping for the wave equation. *SIAM J. Numer. Anal.*, 38(6):2047–2078, 2001.

[22] G. Hauke and M.H. Doweidar. Fourier analysis of semi-discrete and spacetime stabilized methods for the advectivediffusivereactive equation: I. SUPG. *Comp. Meth. Appl. Mech. Engrg.*, 194(1):45–81, 2005.

[23] G. Hauke and M.H. Doweidar. Fourier analysis of semi-discrete and space-time stabilized methods for the advectivediffusivereactive equation: II. SGS. *Comp. Meth. Appl. Mech. Engrg.*, 194(6-8):691–724, 2005.

[24] G. Hauke and M.H. Doweidar. Fourier analysis of semi-discrete and space-time stabilized methods for the advectivediffusivereactive equation: III. SGS/GSGS. *Comp. Meth. Appl. Mech. Engrg.*, 195(44-47):6158–6176, 2006.

[25] T.J.R. Hughes and T.E. Tezduyar. Development of time-accurate finite element techniques for first order hyperbolic systems with emphasys on the compressible euler equations. *Comp. Meth. Appl. Mech. Engrg.*, 45(1-3):217–284, 1984.

[26] J. Maerz and G. Degrez. Improving time accuracy of residual distribution schemes. Technical Report VKI-PR 96-17, von Karman Institute for Fluid Dynamics, 1996.

[27] M. Ricchiuto and R. Abgrall. Stable and convergent residual distribution for time-dependent conservation laws. In *ICCFD4 Proceedings*. Springer-Verlag, 2006.

[28] M. Ricchiuto, R. Abgrall, and H. Deconinck. Application of conservative residual distribution schemes to the solution of the shallow water equations on unstructured meshes,. *J. Comput. Phys.*, 222:287–331, 2007.

[29] M. Ricchiuto and A. Bollermann. Stabilized residual distribution for shallow water simulations. *J. Comput. Phys*, 228(4):1071–1115, 2009.

[30] M. Ricchiuto, Á. Csík, and H. Deconinck. Residual distribution for general time dependent conservation laws. *J. Comput. Phys*, 209(1):249–289, 2005.

[31] P.L. Roe. Fluctuations and signals - a framework for numerical evolution problems. In K.W. Morton and M.J. Baines, editors, *Numerical Methods for Fluids Dynamics*, pages 219–257. Academic Press, 1982.

[32] G. Rossiello, P. De Palma, G. Pascazio, and M. Napolitano. Second-order-accurate explicit fluctuation splitting schemes for unsteady problems. *Computer and Fluids*, 38(7):1384–1393, 2009.

[33] G. Scovazzi, E. Love, and M.J. Shashkov. Multi-scale lagrangian shock hydrodynamics on q1/p0 finite elements: Theoretical framework and two-dimensional computations. *Comp. Meth. Appl. Mech. Engrg.*, 197(9-12):1056–1079, 2007.

[34] F. Shakib and T.J.R. Hughes. A new finite element formulation for computational fluid dynamics: Ix. fourier analysis of space-time galerkin/least-squares algorithms. *Comp. Meth. Appl. Mech. Engrg.*, 87(1):35–58, 1991.

[35] S.Jund. *Méthodes d'éléments finis d'ordre élevé pour la simulations numérique de la propagation d'ondes*. PhD thesis, Université Lous Pasteur, Strasbourg, 2007.

[36] S.Jund and S.Salmon. Arbitrary high order finite element schemes and high order mass lumping. *Int.J.Appl.Math.Comput.Sci*, 17(3):375–393, 2007.

[37] T.E. Tezduyar and T.J.R. Hughes. Development of time-accurate finite element techniques for first order hyperbolic systems with emphasys on the compressible euler equations. Technical Report NASA-CR-204772, NASA-Ames, 1983.

[38] P.R. Woodward and P. Colella. The numerical simulation of two-dimensional flows with strong shocks. *J. Comput. Phys.*, 54:115–173, 1984.

Centre de recherche INRIA Bordeaux – Sud Ouest
Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex (France)

Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399