POLITECNICO DI MILANO

Facoltà di Ingegneria Industriale

Corso di Laurea in Ingegneria Aeronautica

Metodi di Distribuzione del Residuo per le equazioni di conservazione in formulazione Arbitraria Lagrangiana Euleriana con applicazione alle Equazioni di Eulero

Relatore:
Prof. Alberto Guardone

Co-relatore:
Ing. Mario Ricchiuto

Tesi di Laurea di:
Luca Arpaia Matr.755751

Anno Accademico 2012-2013

Contents

Abstract

In this work we develop a numerical method for unsteady hyperbolic conservation laws in Arbitrary Lagrangian Eulerian (ALE) formulation, in order to use moving grids without any time-consuming interpolation step. Hyperbolic equations are discretized in time with explicit Runge Kutta 2 scheme and in space with a Residual Distribution approximation. The numerical method has been tested on scalar problems and on Euler Equations: when computing smooth solutions second order of accuracy is achieved, moreover no oscillations appear when computing discontinuous solutions. For every test case, comparisons with numerical solutions provided by the Eulerian counterpart, confirmed theoretical expectations.

Sommario

In questo lavoro costruiamo una metodo numerico per problemi iperbolici non stazionari scritti nella formulazione Arbitraria Lagrangiana Euleriana (ALE), in modo da poter utilizzare griglie mobili senza alcuna necessità di interpolare la soluzione ad ogni istante di tempo. Le equazioni iperboliche sono discretizzate nel tempo con un metodo di Runge Kutta 2 esplicito e nello spazio con i metodi di Distribuzione del Residuo. Lo schema numerico è stato testato sia su problemi scalari che sulle Equazioni di Eulero: laddove atteso, il metodo è accurato al secondo ordine su soluzioni regolari ed è capace di approssimare gli urti senza oscillazioni. Ovunque vengono riportati i confronti con le soluzioni numeriche fornite dal medesimo algoritmo nella versione Euleriana, ed un buon accordo è riscontrato.

1 Introduction

In this thesis a method for the solution of hyperbolic equations in Arbitrary Lagrangian Eulerian (ALE) formulation through a Residual Distribution space approximation is presented.

First order hyperbolic partial differential equations (PDEs) govern a wide spectrum of phenomena where advection of some information is involved, such as the conservation of the fundamental quantities in geophysics, gasdynamics, acoustics, solid mechanics. Among them, the field that provided a first understanding and that led to development of new methods and ideas, was gasdynamics. The study of compressible flow started to be fundamental for aerospace applications when the speed of airplanes and missiles dramatically increased. A simplified model called Potential Flow was first developed, then, the constant growth in computer power, made possible to solve complete equations for inviscid compressible flows: the Euler Equations. These constitute a nonlinear hyperbolic system of partial differential equations and to find numerical solutions is a very challanging objective due to many reasons. Firstly, they are a set of coupled equations for d+2 variables (in d dimensions), namely the velocity components, density and total energy. Moreover, nonlinearity is revealed through the appereance of shocks and great carefulness in their treatment is required. Finally, all practical applications regards at least two dimensional phenomena. All of this points add a certain degree of difficulty, when moving from scalar equations to systems, from one dimension to two dimensions, from linear to nonlinear, new problems arise for computational methods. The resulting complexity may obscure the basic concepts behind hyperbolic equations and, in addition, it makes the development of new ideas very difficult. Working on simple problems such as scalar advection and Burgers'equation would represent an advantage: ideas are more immediate to understand and more easy to implement. They are just model problems without applications but many results, once understood in such simple cases, can be applied to the into Euler equations. In this thesis we follow this approach; we develop a method for the solution of hyperbolic equations, reasoning on the mentioned model problems. Only when things becames clearer, step by step, we move to more difficult issues, with the application to two dimensional Euler Equation being the ultimate goal.

In this introduction we give a brief overview about the topics addressed in each chapter.

1.1 Residual Distribution

Finite Volume ($\mathcal{FV}$) are the most popular schemes for the discretization of hyperbolic PDEs. They arise from the discretization of integral conservation laws which are the most fundamental equations when the solution contains discontinuities. In these cases differential form does not hold anymore but the integral one still admits solutions. The basic ideas is to break the domain into many volumes and for each of them write the integral conservation law. The solution is averaged within every cell and is updated, at every time step according to the conservation law, hence by a balance of numerical fluxes at the interfaces. The problem is to choose good numerical fluxes. An important aspect is that the solution is mimicking what the exact solution does and, even in presence of discontinuities, these class of methods does not break down. The major drawback of first order $\mathcal{FV}$ scheme is that, in more dimensions, results have proven to be inaccurate due to the large amount of diffusion introduced in the crosswind direction. This is related to the fact that the computation of fluxes at the interfaces is extended directly form one dimension where the normals at the interfaces are always aligned to the wind direction. In two dimensions this could be no longer true and upwinding appears also in crosswind direction. Even if high resolution methods are used this drawback remains.

Stabilized Finite Elements method ($\mathcal{FE}$) rapresents another class of methods which could be applied to hyperbolic equations. The Galerkin approximation for advection dominated problems yields numerical solution with strong oscillations. Thus a stabilization term is added in order to provide the method with some form of dissipation. This corresponds to use test functions belonging to a different space respect to the one in which the solution is searched and for this reason stabilized $\mathcal{FE}$ are also called Petrov Galerkin $\mathcal{FE}$. An estimation for the error shows that, depending on the degree of polynomial used, orders of accuracy higher then one are achieved. On the other hand, a well known weak point is that $\mathcal{FE}$, in the form discussed above, are not suitable for computing discontinuos solutions. Close to discontinuities the numerical solution shows again oscillations that spoils convergence order and, in this regions, dissipation through a tuning parameter has to be introduced.

At the beginning of the eigthies a new class of methods has been proposed by P.L. Roe. He reformalized $\mathcal{FV}$ into a form called Fluctuation Splitting schemes. Here the solution is modified at each time step by the balance of abstract quantities called fluctuations instead of fluxes at the interfaces. This has paved the way to the development of what today is referred to as Residual Distribution ($\mathcal{RD}$). The domain is decomposed into many elements and for each element the flux balance, called residual, is computed. Residual is then splitted through an appropriate distribution, in many contributions, one for each node of the element and, finally, the solution at each node is updated by the contribution of the elements sharing that node. In one dimension it is possible to demonstrate that this abstract passeges are just a reformulation of the classical $\mathcal{FV}$ Godunov method.

The promising aspect was that an abstract geometrical interpretation of upwinding naturally arises. When passing to multidimensional problems this fact is used to introduce upwinding in a more clever way. Indeed early experiments demonstrated that, for multidimensional problems, first order $\mathcal{RD}$ were more accurate then first order $\mathcal{FV}$. Meanwhile the method was provided of a mathematical basis for stability and accuracy analysis, the possibilty of constructing second order scheme was investigated. The results on paper were again very promising because second order and positive schemes were defined without any blending, thus completely parameter free. This rapresents an important advantage with respect to $\mathcal{FE}$. Unfortunately further studies proved that a certain amount of dissipation is needed in order to have a second order and positive accurate solution accurate solution, leading to several different forms of stable nonlinear discretizations.

Chapter 2 is an introduction to Residual Distribution method for simple scalar problems. The link between $\mathcal{FV}$, $\mathcal{RD}$ and $\mathcal{FE}$ in one dimension is pointed out. Then $\mathcal{RD}$ discretization is extended to two dimensions and unsteady problems. The basic elements to perform a stabilty and accuracy analysis on $\mathcal{RD}$ schemes are given. Finally many distributions are presented.

1.2 Time discretization

Since we are interested in unsteady phenomena, in Chapter 3 the time discretization is discussed. The extension to time dependent simulations has been a critic point in the development of $\mathcal{RD}$. An analogy between $\mathcal{RD}$ and stabilized $\mathcal{FE}$ is invoked in order to treat correctly the time part. As in $\mathcal{FE}$ a mass matrix appears. Lumping the mass matrix decouples the equations but spoils the accuracy properties while keeping consistent mass matrix requires its time-consuming inversion and moreover spoils positivity. An explicit Runge Kutta two scheme ($\mathcal{ERK}2$) has been implemented successfully on fixed grid by Ricchiuto and Abgrall: the mathematical fundation of the method is again the analogy with stabilized $\mathcal{FE}$. Firstly, PDE are discretized in time, then they are recast in variational form and fully discretized with a stabilized $\mathcal{FE}$ approximation. The advective part is put in $\mathcal{RD}$ form while the time part is lumped. Positive, second-order accurate solutions are obtained for different problems through a fast, fully explicit scheme.

1.3 Residual Distribution schemes for moving grids

In Chapter 4 we present three coordinate frameworks in which conservation laws can be written. In the Eulerian approach conservation laws are written for a control volume fixed in space, in the Lagrangian approach, the control volume is moving following particles motion, in the Arbitrary Lagrangian Eulerian (ALE) formulation, the control volume is

moving arbitrarly with a prescribed motion law. The Eulerian approach is well suited for fluid dynamics because the volume is fixed and we state conservation for particles that are flowing inside. On the opposite, the Lagrangian approach would involve big distortions of the grid that could lead to instabilties.

If the classical Eulerian approach is used together with moving grids, at every time step we have to write conservation laws for new volumes. On the updated grid, the solution at the previous time step has never been calculated. Thus an interpolation step of the old solution on the updated grid becomes necessary at every time step. With the ALE approach we write conservation laws for the same control volumes and no time-consuming interpolation step is needed. If an ALE framework is used a condition regarding the preservation of the volumes arises. This is referred to as Geometric Conservation Law (GCL). Reformulation of Eulerian scheme into ALE framwork requires minor modification in the algorithm, however great care has to be put in order to satisfy GCL.

All the space discretizations have been put into ALE formulation. Regarding $\mathcal{RD}$, in the last decade, many papers have been written, due to the contribution of Deconinck and coworkers at Von Karman Institute. First-order accuracy has been reached with Explicit Euler time-approximation, second-order has been achieved only using implicit time schemes. In this thesis a novel $\mathcal{RD}$ discretization is proposed based on the reformulation of the genuinely explicit $\mathcal{RD}$ scheme described in chapter 3, in ALE form. The resulting scheme is rearranged in order to resamble its Eulerian counterpart, mantaining its nice stability and accuracy properties. The GCL is verified by construction through an appropriate choice of the grid velocity and of time instants on which integrals are performed. We still expect to end up with an explicit algorithm that gives positive and second-order accurate solutions. The method is tested firstly on simple two dimensional scalar problems: a linear advection test case is used to see accuracy property when computing smooth solutions while the Burger's equation test case should demonstrates positivity property when computing shock waves.

1.4 Residual Distribution schemes for Euler Equations with moving grids

The necessity of moving grids comes, for unsteady phenomena, in order to adapt to moving boundaries. This demand is particularly important in aeronautical applications, where many problems of interest involve the motion of boundaries (i.e. aeroelasticity, turbomachinery and helicopter applications). Also in unsteady aerodynamics, since accuracy depends on spatial discretization, it is desirable to change the grid according to the solution itself for example refining it in regions where strong gradients appears. For this reasons in Chapter 5 we extend the scalar scheme to systems of hyperbolic equations,

in particular we focus our attention on Euler equations of gasdynamics. Again, accuracy and positivity is shown through two test cases: the advection of a vortex and a two dimensional Riemann problem. Finally a very simple application with moving boundaries is presented.

2 Residual Distribution

In this second chapter we present Residual Distribution $\mathcal{RD}$ schemes which rapresent nowdays a nice alternative for the solution of hyperbolic problems to either Finite Volume $\mathcal{FV}$ and Finite Elements $\mathcal{FE}$. The main goal will be to design a class of conservative schemes which could compute nice solutions both in smooth regions and near discontinities, hence guaranteeing accuracy and a non-oscillatory beahviour at the same time. The road to $\mathcal{RD}$ was paved by the work of Roe on Flux-Difference Splitting Finite Volume schemes and after on Fluctuation Splitting schemes [1][2]. At that time, among the two-dimensional upwind methods, many schemes were known to be less diffusive respect to the first order upwind $\mathcal{FV}$ ones [3][4][5]. The idea that brought to $\mathcal{RD}$ was to incorporate upwinding in a genuinely multidimensional way, always remaining with a compact stencil. This leads to the design of a class of optimal upwind schemes. From the pioniereeng work in the eighties of Roe at the University of Michigan and the immediate development of Deconinck and coworkers at Von Karman Institute, contributions has been given by many groups which pushed step by step to new issues: high order of accuracy, solution of unsteady problems, extension to viscous problems. Among them we mention Hubbard and coworkers (Leeds University), Napolitano and coworkers (Politecnico di Bari) and Abgrall and coworkers (INRIA Bordeaux).

2.1 Basic concept in 1D

2.1.1 An upwind $\mathcal{RD}$ scheme

We start, for its semplicity, from the approximation of the one-dimensional homogeneus hyperbolic equation called conservation law which is expressed by the following partial differential equation

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = 0 \quad x \in [0,1],\, t \in [0,T] \tag{2.1}$$

where $u = u(x,t)$ is the solution or the conserved quantity and $f = f(u)$ is the flux function. (2.1) can be rewritten in quasilinear form

$$\frac{\partial u}{\partial t} + a(u)\frac{\partial u}{\partial x} = 0 \tag{2.2}$$

where $a(u) = \frac{\partial f(u)}{\partial u}$ is called the flux jacobian and has to be real to provide the hypebolicity. Since in this case (2.2) models the simple transport or advection of the conserved quantity at local speed $a(x,t) = a(u(x,t))$, then a is rightly called advective speed.

Finite Volume and Residual Distribution

The starting point for the development of $\mathcal{RD}$ schemes by Roe was the introduction of the concept of *fluctuations* for the 1D upwind Finite Volume method. The slightly different viewpoint opened the way to this new methods. Starting from the integral form (2.1) written for the time slab $[t^n, t^{n+1}]$ and control volume h_i, which rapresent the medial dual cell area surrounding node i, we obtain

$$\int_{h_i} u(x,t^{n+1})\,dx = \int_{h_i} u(x,t^n)\,dx - \int_{t^n}^{t^{n+1}} f(u(x_{i+\frac{1}{2}},t))\,dt + \int_{t^n}^{t^{n+1}} f(u(x_{i-\frac{1}{2}},t))\,dt \quad (2.3)$$

The steps for a $\mathcal{FV}$-Godunov method are the followings

1. Approximate the solution at time t^n over the cell with the cell average $u_i^n = \frac{1}{h_i}\int_{h_i} u(x,t^n)\,dx$
2. Such an approximation define a sequence of Riemann problem at every interface. Evolve (2.3) with the above initial data to obtain the exact solution at time t^{n+1}, $u^{n+1}(x)$
3. Average the solution over the cell $u_i^{n+1} = \frac{1}{h_i}\int_{h_i} u(x,t^{n+1})\,dx$

At the end of this three steps, (2.3) reduce to

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{h_i}\left(F_{i+1/2}(u_i^n, u_{i+1}^n) - F_{i-1/2}(u_{i-1}^n, u_i^n)\right) \quad (2.4)$$

where numerical flux functions F are computed with $u^\curlyvee$ obtained by solving a Riemann problems at each cell interface

$$F_{i-1/2} = f(u^\curlyvee(u_{i-1}, u_i))$$

$$F_{i+1/2} = f(u^\curlyvee(u_i, u_{i+1}))$$

If, instead of the exact Riemann problem defined by (2.1) together piecewise constant initial data, we solve an approximate linearized problem at the interfaces using a Roe linearization, we can express the fluxes as

$$F_{i-1/2} = \frac{1}{2}(f_i + f_{i-1}) - \frac{1}{2}|a_{i-1/2}|(u_i - u_{i-1}) \quad (2.5)$$

$$F_{i+1/2} = \frac{1}{2}(f_{i+1} + f_i) - \frac{1}{2}|a_{i+1/2}|(u_{i+1} - u_i) \quad (2.6)$$

for example, at at the interface $x_{i-1/2}$, the original problem has been replaced by a linear one with constant advection speed fixed at an appropriate averege $a_{i-1/2}$ such that $f_i - f_{i-1} = a_{i-1/2}(u_i - u_{i-1})$.

The numerical fluxes, substituted into (2.4), provide a $\mathcal{FV}$ method which has an upwind property because the approximation of the numerical flux function is one-sided in the direction of the advective speed. This method can be rewritten in the following form

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{h_i}\left(a^+_{i-1/2}(u_i - u_{i-1}) + a^-_{i+1/2}(u_{i+1} - u_i)\right) = 0 \tag{2.7}$$

where

$$a^+ = \max(0, a) \quad \text{and} \quad a^- = \min(0, a)$$

Moreover for the cell (x_{i-1}, x_i) we can also write

$$f_i - f_{i-1} = a^+_{i-1/2}(u_i - u_{i-1}) + a^-_{i-1/2}(u_i - u_{i-1}) \tag{2.8}$$

The flux difference is splitted between a quantity that rapresents the effect of the right-going wave $a^+_{i-1/2}(u_i - u_{i-1})$ entering at the interface $x_{i-1/2}$ and a quantity $a^-_{i-1/2}(u_i - u_{i-1})$ rapresenting the effect of the left-going wave entering at the same interface from the opposite direction. These contributions are called *fluctuations* and give more insight about the $\mathcal{FV}$ scheme (2.7): the updated solution results from a balance of fluctuations instead that of fluxes. Infact the flux difference is splitted between a left-going fluctuation that updates u_{i-1} and a right-going fluctuation that updates u_i. A full parallelism between $\mathcal{FV}$ equation (2.4) and such method is obtained if in (2.4) one uses the following conservative numerical fluxes

$$F_{i-1/2} = f(u_i) - a^+_{i-1/2}(u_i - u_{i-1}) \tag{2.9}$$

From Finite Volumes a more general class of method arises, such methods, called Flux-Difference Splitting schemes, are based on some splitting of the flux difference, in the same fashion of (2.8), followed by application of (2.7).

$$f_i - f_{i-1} = \mathcal{F}^+_{i-1/2} + \mathcal{F}^-_{i-1/2} \tag{2.10}$$

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{h_i}\left(\mathcal{F}^+_{i-1/2} + \mathcal{F}^-_{i+1/2}\right) = 0 \tag{2.11}$$

From Flux-Difference Splitting to Residual Distribution the step is short. We have to formalize the method in terms of residuals instead that in terms of fluctuations. In the steady case the passages to construct a $\mathcal{RD}$ method are the followings:

1. On every cell define the cell residual using a continuous piecewise linear approximation of the solution $u^h(x,t) = \sum_j \varphi_j(x) u_j(t)$

$$\phi^{i-1/2} = \int_{x_{i-1}}^{x_i} \frac{\partial f(u^h)}{\partial x} dx = f_i - f_{i-1} \tag{2.12}$$

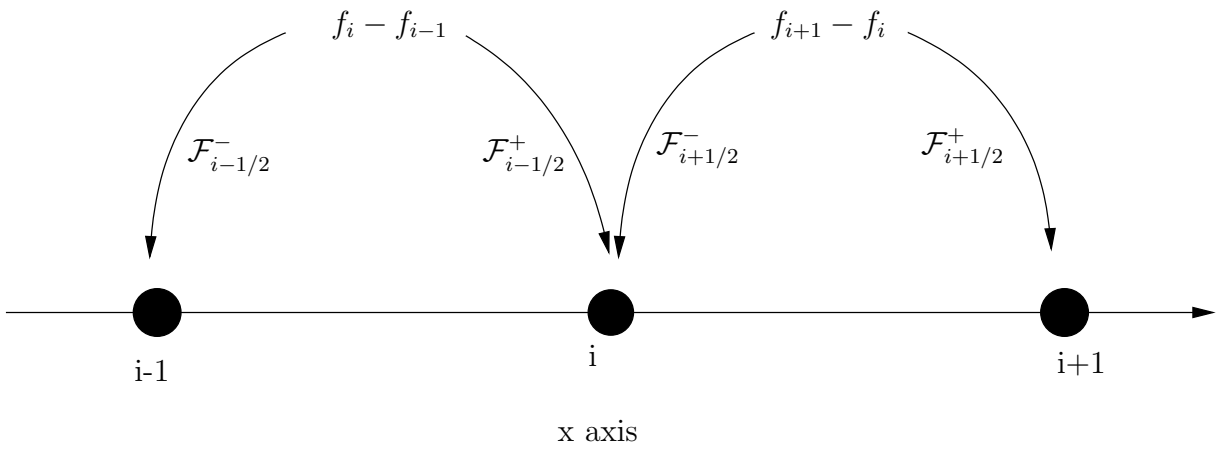

Figure 2.1: Flux-Difference Splitting

2. Split the cell residual (nothing else then a flux difference) between the two nodes of the cell through the distribution coefficients $\beta_i^{i-1/2}$, $\beta_{i+1}^{i-1/2}$

$$\phi^{i-1/2} = \underbrace{\beta_{i-1}^{i-1/2}\phi^{i-1/2}}_{\phi_{i-1}^{i-1/2}} + \underbrace{\beta_i^{i-1/2}\phi^{i-1/2}}_{\phi_i^{i-1/2}} \tag{2.13}$$

   In order to compute consistently the residual, the distribution coefficients has to sum to one for each cell

$$\beta_{i-1}^{i-1/2} + \beta_i^{i-1/2} = 1$$

   which is reffered to as consistency condition

3. Assemble the residuals at node i

$$\beta_i^{i+1/2}\phi^{i+1/2} + \beta_i^{i-1/2}\phi^{i-1/2} = 0 \tag{2.14}$$

By a proper choice of $\beta_{i-1}^{i+1/2}$, $\beta_i^{i-1/2}$ the steady version of (2.7) is recovered

$$\beta_{i-1}^{i-1/2} = \frac{a^-_{i-1/2}}{a_{i-1/2}} \quad \beta_i^{i-1/2} = \frac{a^+_{i-1/2}}{a_{i-1/2}} \tag{2.15}$$

The distribution of the residuals (2.15) is clearly upwind in the sense that we evoy all the residual to the downstream node of the cell, relatively to the advection speed. This geometrical interpretation of upwinding, very different respect to a $\mathcal{FV}$ context, will be of key importance in two-dimensions. For now, we can conclude that, in one dimension, through Flux-Difference Splitting, we have found a full parallel between upwind $\mathcal{RD}$ and upwind $\mathcal{FV}$.

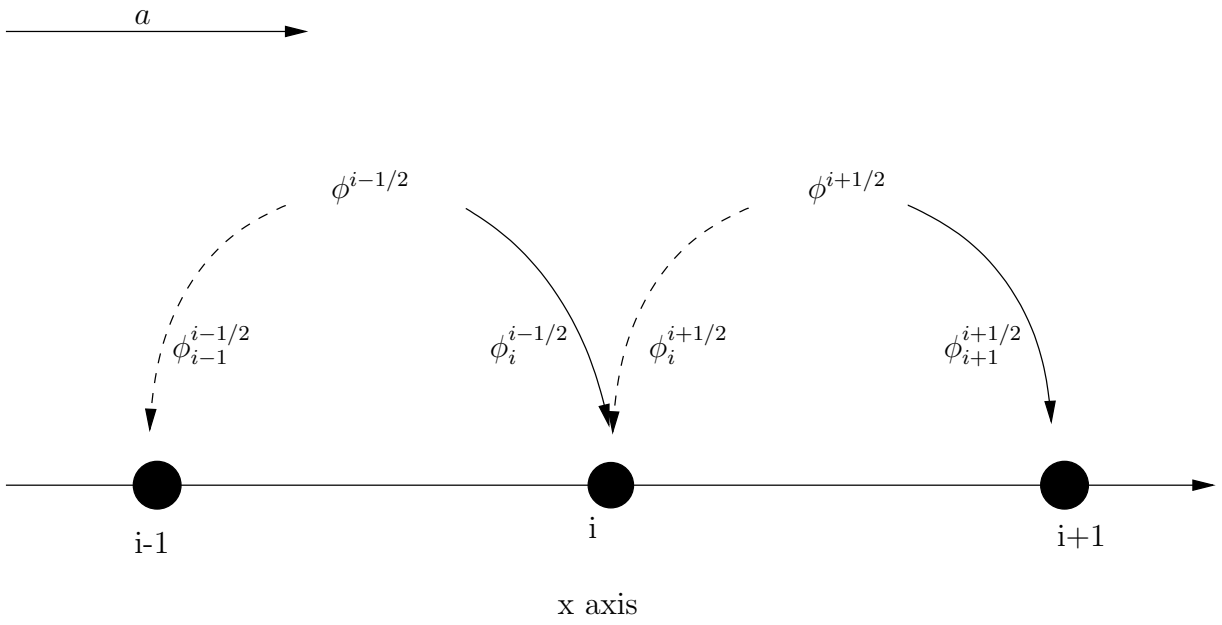

Figure 2.2: Upwinding in 1D $\mathcal{RD}$: all the residual is envoyed to the donwsteam node

Finite Elements and Residual Distribution

In order to derive the $\mathcal{RD}$ analogue of (2.4) we miss the time part which has to be treated properly. This can be done with an analogy between $\mathcal{RD}$ and Finite Elements. Starting form (2.1) we construct a stabilized $\mathcal{FE}$ method through the following steps

1. Write (2.1) in weak form with solution and test function $u, v \in V$, with V a suitable functional space, and, after the domain has been discretized, use a Galerkin Finite Element approximation with both the approximate solution and the test function belonging to the space of piecewise linear polynomial over an element $u^h, v^h \in V^h \equiv X_h^1(0,1)$. This space is described by a lagrangian basis having for elements the $N+1$ Galerkin linear shape functions $\{\varphi_i\}_{i=1,N+1}$. Neglecting boundary conditions, the approximate weak form reads

$$\text{find} \quad u^h \in V^h : \quad a(u^h, v^h) = 0, \qquad \forall v^h \in V^h \tag{2.16}$$

with the form $a(\cdot,\cdot)$ that defines the following scalar product $a(u,\cdot) = \int_0^1 L(u)(\cdot)\, dx$. $L(u)$ is the differential operator associated to the conservation law. Imposing that (2.16) is satisfied for every element of the basis of X_h^1

$$\int_0^1 \left(\frac{\partial u^h}{\partial t} + \frac{\partial f(u^h)}{\partial x} \right) \varphi_i \, dx = 0 \quad \forall \varphi_i,\ i = 1, N+1 \tag{2.17}$$

2. Since the Galerkin approximation of an hyperbolic problem unfortunatly gives unstable numerical results, we add a stabilization term modifing the approximation of the weak problem in the following way

$$\text{find} \quad u^h \in X_h^1 : \quad a(u^h, \varphi_i) + \mathcal{L}_h(u^h, \varphi_i) = 0 \quad \forall \varphi_i,\, i = 1, N+1 \tag{2.18}$$

for a suitable stabilization operator $\mathcal{L}_h$ such that it vanishes if the exact solution is used

$$\mathcal{L}_h(u, \varphi_i) = 0 \qquad \forall \varphi_i,\, i = 1, N+1 \tag{2.19}$$

One obtains a SUPG method if uses

$$\mathcal{L}(u^h, \varphi_i) = \sum_{i=1}^{N} \int_{x_i}^{x_{i+1}} L(u^h) \tau a(u^h) \frac{\partial \varphi_i}{\partial x}\, dx \tag{2.20}$$

where obviously $L(u) = 0$ and constraint (2.19) is satisfied. So far we have obtained

$$\int_0^1 \left(\frac{\partial u^h}{\partial t} + \frac{\partial f(u^h)}{\partial x} \right) \varphi_i \, dx + \sum_{i=1}^{N} \int_{x_i}^{x_{i+1}} \left(\frac{\partial u^h}{\partial t} + \frac{\partial f(u^h)}{\partial x} \right) \tau a(u^h) \frac{\partial \varphi_i}{\partial x}\, dx = 0$$

Using the fact that $\varphi_i, \frac{\partial \varphi_i}{\partial x} \neq 0$ only in the interval $[x_{i-1}, x_{i+1}]$ and assembling in a different way

$$\int_{x_{i-1}}^{x_{i+1}} \frac{\partial u^h}{\partial t} \left(\varphi_i + \tau a(u^h) \frac{\partial \varphi_i}{\partial x} \right) dx + \int_{x_{i-1}}^{x_{i+1}} \frac{\partial f(u^h)}{\partial x} \left(\varphi_i + \tau a(u^h) \frac{\partial \varphi_i}{\partial x} \right) dx = 0$$

Basically we are modifing the test function which, in general, is no more belonging to the same functional space of the solution, X_h^1, but it has an upwind bias due to the stabilization bubble, as seen figure (2.3). The method is then said Petrov-Galerkin Finite Elements. Calling

$$w_i = \varphi_i + \gamma_i, \quad \gamma_i = \tau a(u^h) \frac{\partial \varphi_i}{\partial x} \tag{2.21}$$

After step 1 and 2, for internal nodes, the scheme is written compactely

$$\int_{x_{i-1}}^{x_{i+1}} \left(\frac{\partial u^h}{\partial t} + \frac{\partial f(u^h)}{\partial x} \right) w_i \, dx = 0 \quad \forall w_i,\, i = 2, N \tag{2.22}$$

Which gives immediately the following semidiscretization

$$\sum_{j-1,j,j+1} \int_{x_{i-1}}^{x_{i+1}} \varphi_j w_i \, dx \, \frac{du_j}{dt} + \int_{x_{i-1}}^{x_{i+1}} \frac{\partial f(u^h)}{\partial x} w_i \, dx = 0 \qquad j = i,\ i = 2, N \tag{2.23}$$

Assuming a local Roe linearization of the fluxes

$$\phi^{i-1/2} = \int_{x_{i-1}}^{x_i} \frac{\partial f(u^h)}{\partial x}\, dx = a_{i-1/2}\frac{\partial u^h}{\partial x} h_{i-1/2}, \quad a_{i-1/2} = \frac{1}{h_{i-1/2}} \int_{x_{i-1}}^{x_i} a(u^h)\, dx \tag{2.24}$$

we can recast the advective part in a $\mathcal{RD}$ form as in (2.14) . Infact

$$\sum_{j-1,j,j+1} \left(\int_{x_{i-1}}^{x_i} \varphi_j w_i\, dx + \int_{x_i}^{x_{i+1}} \varphi_j w_i\, dx \right) \frac{du_j}{dt} + \int_{x_{i-1}}^{x_i} a_{i-\frac{1}{2}} \frac{\partial u^h}{\partial x} w_i\, dx + \int_{x_i}^{x_{i+1}} a_{i+\frac{1}{2}} \frac{\partial u^h}{\partial x} w_i\, dx = 0$$

In the first two terms we recognize the consistent element mass matrices for elements $i-1/2$ and $i+1/2$

$$\sum_{j-1,j} m_{ij}^{i-1/2} \frac{du_j}{dt} + \sum_{j,j+1} m_{ij}^{i+1/2} \frac{du_j}{dt} + \frac{\phi^{i-1/2}}{h_{i-1/2}} \int_{x_{i-1}}^{x_i} w_i\, dx + \frac{\phi^{i+1/2}}{h_{i+1/2}} \int_{x_i}^{x_{i+1}} w_i\, dx = 0 \tag{2.25}$$

If one calls

$$\frac{1}{h_{i-1/2}} \int_{x_{i-1}}^{x_i} w_i\, dx = \beta_i^{i-1/2} \tag{2.26}$$

That shows the equivalence between $\mathcal{FE}$ and $\mathcal{RD}$: in the first class of method the conservation law is first multiplied by a test function and then integrated over the domain while in the second, roughly speaking, first one integrates and then multiplies for some coefficients. (2.26) gives a condition to make the two operation equivalent.

With the choice $\tau = \frac{h}{2|a|}$ in (2.21), one obtains, after integration of (2.26), the upwind distribution of residuals already presented in (2.15), just written in a slightly different form

$$\beta_i^{i-1/2} = \frac{1}{2} - \frac{1}{2}\frac{a_{i-1/2}}{|a_{i-1/2}|} \tag{2.27}$$

Except for the time part where a mass matrix appears, SUPG collapses again to an upwind scheme. Stabilized $\mathcal{FE}$, $\mathcal{FV}$ and $\mathcal{RD}$ coincide and, more interesting fact, they give three different interpretations to the concept of "upwinding" as seen from the comparison of figure (2.2) and (2.3).

There are many choice of w_i that satisfy (2.26), for example two possible choices which are reffered respectively to as F1 and F2 (formulation one and two) are

$$\begin{cases} w_i^{F1} = const = \beta_i^{i-1/2} & x_{i-1} \le x \le x_i \\ w_i^{F1} = const = \beta_i^{i+1/2} & x_i \le x \le x_{i+1} \end{cases} \tag{2.28}$$

$$\begin{cases} w_i^{F2} = \varphi_i + \beta_i^{i-1/2} - \frac{1}{2} & x_{i-1} \le x \le x_i \\ w_i^{F2} = \varphi_i + \beta_i^{i+1/2} - \frac{1}{2} & x_i \le x \le x_{i+1} \end{cases} \tag{2.29}$$

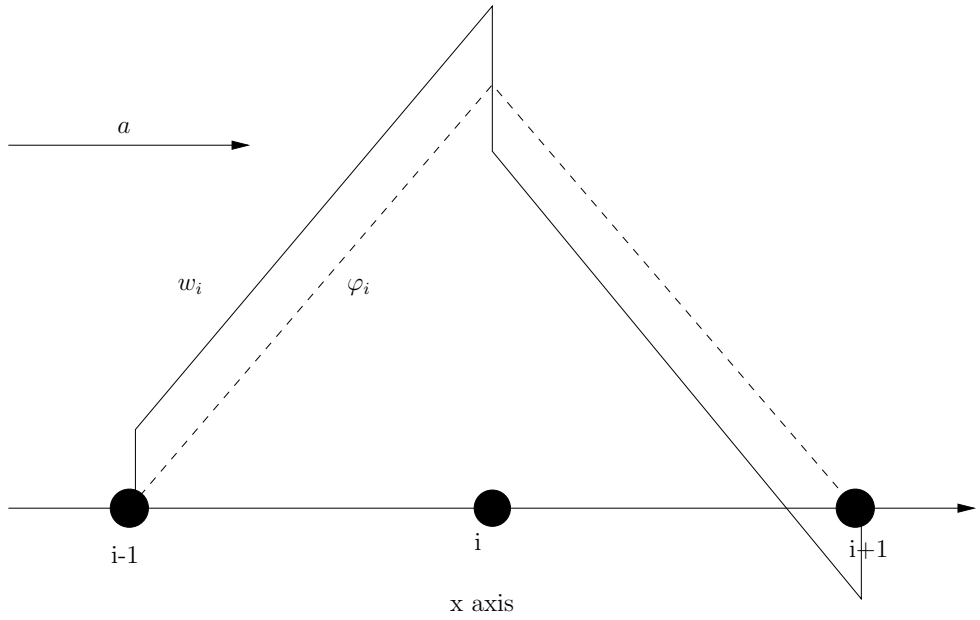

Figure 2.3: Upwinding in 1D $\mathcal{FE}$: Streamline Upwind Petrov Galerkin test function

Once a unique correspondence between β and w has been established we can compute the mass matrix which will depends from the formulation choosen. For the cell $i+1/2$ we have

$$m_{ij}^{i+1/2,F1} = \int_{x_i}^{x_{i+1}} \varphi_j w_i^{F1}\, dx = \frac{h_{i+1/2}}{2}\, \beta_i^{i+1/2} \tag{2.30}$$

$$m_{ij}^{i+1/2,F2} = \int_{x_i}^{x_{i+1}} \varphi_j w_i^{F2}\, dx = \frac{h_{i+1/2}}{2} \left(\frac{\delta_{ij}}{3} + \beta_i^{i+1/2} - \frac{1}{6} \right) \tag{2.31}$$

Finally the $\mathcal{RD}$ semidiscretization of (2.1) then reads

$$\sum_{j-1,j} m_{ij}^{i-1/2} \frac{du_j}{dt} + \sum_{j+1,j} m_{ij}^{i+1/2} \frac{du_j}{dt} = -\left(\beta_i^{i-1/2} \phi^{i-1/2} + \beta_i^{i+1/2} \phi^{i+1/2} \right) \quad j = i,\ i = 2, N \tag{2.32}$$

To get a full equivalent with the upwind $\mathcal{FV}$ method (2.4) one can do mass-lumping on the mass-matrices and, using the consistency condition, for both the formulations one gets

$$m_{ij}^{i+1/2,ML} = \frac{h_{i+1/2}}{2}\, \delta_{ij} \tag{2.33}$$

Calling $h_i = \frac{1}{2}(h_{i-1/2} + h_{i+1/2})$ the Finite Volume median dual cell area, (2.32) becomes

$$h_i \frac{du_i}{dt} = -\left(\beta_i^{i-1/2} \phi^{i-1/2} + \beta_i^{i+1/2} \phi^{i+1/2} \right) \tag{2.34}$$

Thinking to the β as more general coefficients that only has to satisfy the consistency requirment, (2.34) is the one-dimensional $\mathcal{RD}$ compact protype scheme. Regarding the accuracy and the stabilty of the method more strict conditions has to be fixed but this will be discussed later, directly in two-dimensions.

2.1.2 A Lax-Friederich $\mathcal{RD}$ scheme

Within the $\mathcal{FV}$ method (2.4), the classical Lax-Friederich (LxF) approximation of the fluxes has the form

$$F^{LxF}_{i-1/2} = \frac{1}{2}\left(f_i + f_{i-1}\right) - \frac{1}{2}\frac{h_{i-1/2}}{\Delta t}\left(u_i - u_{i-1}\right) \tag{2.35}$$

this flux looks like an unstable centered flux with an additional term that models an artificial diffusive term of type $\upsilon(x)u_x$, with artificial viscosity $\upsilon = \frac{1}{2}\frac{h^2_{i-1/2}}{\Delta t}$. We see that, as the grid is refined, this coefficient vanishes so that the method is still consistent with the original hyperbolic equation but, at the same time, we are introducing a numerical diffusion that damps instabilties. However the Lax-Frederich method introduce too much diffusion giving very poor numerical results unless a very fine grid is used.

It is very easy to recast the numerical flux difference $F^{LF}_{i+1/2} - F^{LF}_{i-1/2}$ in a $\mathcal{RD}$ framework, just reassembling the terms present in (2.35)

$$F^{LxF}_{i+1/2} - F^{LxF}_{i-1/2} = \underbrace{\frac{1}{2}\left(f_{i+1} - f_i\right) + \frac{1}{2}\frac{h_{i+1/2}}{\Delta t}\left(u_i - u_{i+1}\right)}_{\phi_i^{i+1/2,LxF}} + \underbrace{\frac{1}{2}\left(f_i - f_{i-1}\right) + \frac{1}{2}\frac{h_{i-1/2}}{\Delta t}\left(u_i - u_{i-1}\right)}_{\phi_i^{i-1/2,LxF}}$$

The residuals are distributed, according to Lax-Friederich, splitting the total residual (2.12) in a centered way and adding a diffusion term to it

$$\phi_i^{i+1/2,LxF} = \frac{1}{2}\phi^{i+1/2} + \frac{1}{2}\frac{h_{i+1/2}}{\Delta t}\left(u_i - u_{i+1}\right) \tag{2.36}$$

$$\phi_i^{i-1/2,LxF} = \frac{1}{2}\phi^{i-1/2} + \frac{1}{2}\frac{h_{i-1/2}}{\Delta t}\left(u_i - u_{i-1}\right) \tag{2.37}$$

As already seen for the upwind method, in one-dimension everything collapse to the same scheme.

2.2 2D Residual Distribution for steady scalar conservation law

Consider the steady scalar conservation law

$$\nabla \cdot \boldsymbol{f}(u) = 0, \quad \boldsymbol{x} \in \Omega \tag{2.38}$$

Once we have approximated the domain through a suitable triangulation $\mathcal{T}_h$ we propose directly a Residual Distribution approximation of (2.38). We repeat the same steps seen for the one-dimensional case. Boundary condition are neglected at this point and we imagine that every element does not share any edge with the domain boundary.

1. With a piecewice linear approximation of the solution over each triangle $u^h(\boldsymbol{x},t) = \sum_{j=1}^{N+1} \varphi_j(\boldsymbol{x}) u_j(t)$ we compute the residual on each element

$$\phi^K = \int_K \nabla \cdot \boldsymbol{f}(u^h)\, d\boldsymbol{x} = \int_{\partial K} \boldsymbol{f}(u^h) \cdot \boldsymbol{n}\, ds \tag{2.39}$$

   as in the 1D-case, the physical interpretation of the residual is a flux balance, this time, over a triangle.

2. Distribute the residuals to the nodes of the element $i, j, k \in K$ through weights that sum up to one for consistency

$$\phi^K = \beta_i^K \phi^K + \beta_j^K \phi^K + \beta_k^K \phi^K = \sum_{j \in K} \phi_j^K \tag{2.40}$$

   with

$$\beta_i^K + \beta_j^K + \beta_k^K = 1$$

3. Assembly the residuals shared by the same node. If $\mathcal{D}_i$ is the domain formed by all the elements of the triangulation that have in common node i, we have

$$\sum_{K \in \mathcal{D}_i} \beta_i^K \phi^K = \sum_{K \in \mathcal{D}_i} \phi_i^K = 0, \quad \forall i \in \mathcal{T}_h \tag{2.41}$$

Linearization

Instead of working with the conservative form of the residual (2.39), the quasi-linear form will be more suited to our purposes. For non linear problems a linearization is necessary at this point. This passage requires some carefulness since the correct computation of the residual implies to end up with a conservative method.

Being conservative is a delicate issue for every numerical method approximating conservation laws. It means that the numerical solution satisfies the integral form of conservation laws, mimicing what the exact solution does. Summing (2.41) over all the elements, using consistency condition (2.40), (2.39) and the fact that, for a conservative scheme, fluxes cancel out except at the boundaries, we have that

$$\begin{aligned} \sum_{i \in \mathcal{T}_h} \sum_{K \in \mathcal{D}_i} \phi_i^K &= \sum_{K \in \mathcal{T}_h} \sum_{j \in K} \phi_j^K = \sum_{K \in \mathcal{T}_h} \int_{\partial K} \boldsymbol{f}(u^h) \cdot \boldsymbol{n}\, ds \\ &= \int_{\partial \Omega_h} \boldsymbol{f}(u^h) \cdot \boldsymbol{n}\, ds \end{aligned}$$

which states that, imposing correctly boundary conditions, we have exact conservation over the full domain. We can reassume that a $\mathcal{RD}$ method is conservative if it is consistent and we are able to approximate the residual with a sufficient degree of precision

$$\phi^K = \int_{\partial K} \boldsymbol{f}(u^h) \cdot \boldsymbol{n}\, ds \tag{2.42}$$

This has been interpreted as a constraint on the linearization. To continue to compute correctly the residuals, the passage from a conservative form to a linearized one should be exact

$$\begin{aligned} \phi^K &= \int_K \nabla \cdot \boldsymbol{f}(u^h)\, d\boldsymbol{x} = \int_K \boldsymbol{a}(u^h)\, d\boldsymbol{x} \cdot \nabla u^h \\ &= \bar{\boldsymbol{a}} \cdot \nabla u^h |K| \end{aligned} \tag{2.43}$$

where

$$\bar{\boldsymbol{a}} = \frac{1}{|K|} \int_K \boldsymbol{a}(u^h)\, dx \tag{2.44}$$

An important result, which will be used extensively, is the following definition of the linearized residual. From the definition of the gradient of a P1 solution over the element $\nabla u^h = \frac{1}{2|K|} \sum_{j \in K} \boldsymbol{n}_j u_j$ we have that

$$\begin{aligned} \phi^K &= \bar{\boldsymbol{a}} \cdot \nabla u^h |K| = \frac{1}{2} \sum_{j \in K} \bar{\boldsymbol{a}} \cdot \boldsymbol{n}_j u_j \\ &= \sum_{j \in K} k_j u_j \end{aligned} \tag{2.45}$$

We have introduced the upwind parameter

$$k_i = \frac{1}{2} \bar{\boldsymbol{a}} \cdot \boldsymbol{n}_i \tag{2.46}$$

Equivalently

$$\phi^K = \sum_{j \in K, , j \neq i} k_j \left(u_j - u_i\right) \tag{2.47}$$

Upwinding

If in one-dimension was immediate to construct an upwind scheme able of introducing numerical diffusion to damp oscillation, in two-dimensions things are not straightforward. For $\mathcal{FV}$, one proceeds directly with a two-dimensional extension of the method that works in 1D. Connected to this procedure there is the problem of the strong cross-diffusion

introduced, which, in turn, come up directly from the fact that the normals to the faces are no more aligned with the direction of the advective velocity, as in 1D. If we recall how we have introduced upwinding for one-dimensional Residual Distribution, the extension to multidimensional problem seems to be less critic in this case. The flux balance infact, can be splitted between the nodes of the element in an "upwind way" respect to the advective speed, also if we are in more dimensions.

As in 1D it is crucial to establish if a node of an element is upstream or downstream. Through the upwind parameter it is possible to distinguish between inflow and outflow faces and upstream and downstream nodes

$$\begin{aligned} \text{inflow face} &: k_i > 0 \quad i \text{ is downstream} \\ \text{outflow face} &: k_i < 0 \quad i \text{ is upstream} \end{aligned}$$

Generalizing the $\mathcal{RD}$ one-dimensional upwind distribution, in order to have an upwind method, we want to split the residual only between downstream nodes with the upstream nodes receiving no contribution instead. We give the following definition

Definition (Upwinding) *A $\mathcal{RD}$ method is upwind if*

$$k_i < 0 \Rightarrow \beta_i = 0 \tag{2.48}$$

The one-inflow case of figure (2.4) is straightforward since all the residual is sent to the only downstream node. This distribution is said to be one-target. The two-inflow case is slightly more difficult because one has to choose how to split the residual between the two downstream nodes. Independently from the choosen criteria this distribution is said to be two-target. Different choices are possible, leading to upwind schemes with different properties.

2.3 Unsteady scalar conservation law

As in this theses unsteady problem will be under study, we consider here the integration of the time dependent conservation law $L(u) = 0$

$$\frac{\partial u}{\partial t} + \nabla \cdot \boldsymbol{f}(u) \quad in\, \Omega \in \mathbb{R}^2,\, t \in [0, T] \tag{2.49}$$

Since the advective part has been treated in the previous paragraph, here we explain how the time part can be treated. As done in 1D we use the analogy with the stabilized $\mathcal{FE}$ method. Given the unsteady scalar conservation law (2.49) the problem is approximated with the following

$$\text{find} \quad u^h \in X_h^1 : \quad a(u^h, \varphi_i) + \mathcal{L}_h(u^h, \varphi_i) = 0 \quad \forall \varphi_i,\, i = 1, N+1 \tag{2.50}$$

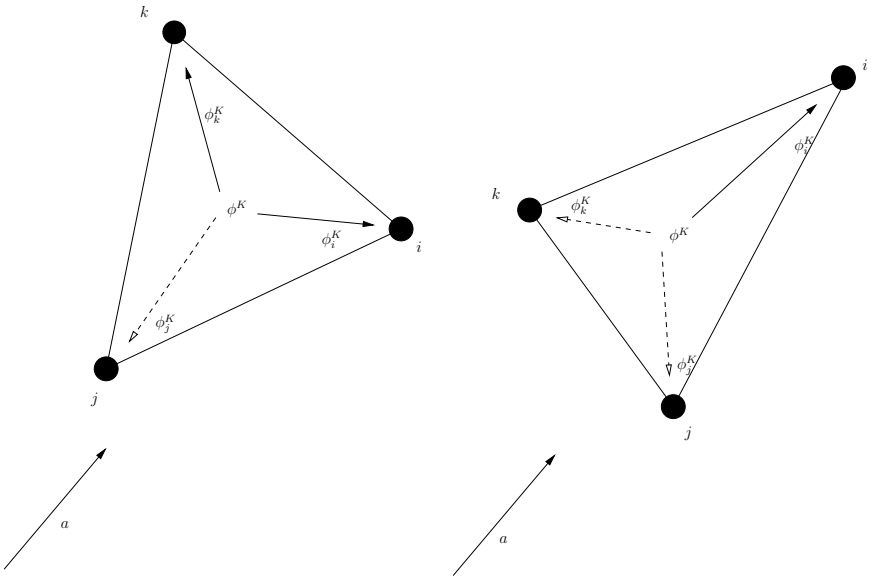

Figure 2.4: Upwinding, 2D $\mathcal{RD}$: left) two inflow case. right) one inflow case

where $a(\cdot,\cdot)$ defines the scalar product in $L^2(\Omega)$, $a(u,\cdot) = \int_\Omega L(u)(\cdot)\, d\boldsymbol{x}$, φ_i is the two-dimensional Galerkin shape function, $\mathcal{L}_h(u^h, \varphi_i)$ is a proper stabilization form that has to verify the following properties:

1. Accuracy: to not spoil the accuracy analysis of the Galerkin method it should satisfy

$$\mathcal{L}_h(u, \varphi_i) = 0 \qquad \forall \varphi_i \in X_h^1 \tag{2.51}$$

2. Stabiliy: it should introduce a diffusion-like term
3. Conservation: it should not provide any contribution to the flux balance

A possible choice, for example, is the SUPG stabilization form

$$\mathcal{L}_h(u^h, \varphi_i) = \sum_{K \in \mathcal{T}_h} \int_K L(u^h)\, \tau \boldsymbol{a}(u^h) \cdot \nabla \varphi_i \, d\boldsymbol{x}$$

where τ is a parameter. We will rapidly generalize the above operator using the bubble functions γ_i

$$\mathcal{L}_h(u^h, \varphi_i) = \sum_{K \in \mathcal{T}_h} \int_K \gamma_i \left(\frac{\partial u^h}{\partial t} + \nabla \cdot \boldsymbol{f}(u^h) \right) d\boldsymbol{x} \tag{2.52}$$

The bubble function will modify the Galerkin shape function, element by element, introducing a kind of weightening for the test function in the upwind direction. Infact if we

write the full scheme

$$\int_\Omega \varphi_i \left(\frac{\partial u^h}{\partial t} + \nabla \cdot \boldsymbol{f}(u^h) \right) d\boldsymbol{x} + \sum_{K \in \mathcal{T}_h} \int_K \gamma_i \left(\frac{\partial u^h}{\partial t} + \nabla \cdot \boldsymbol{f}(u^h) \right) d\boldsymbol{x} = 0 \tag{2.53}$$

The test function $w_i = \varphi_i + \gamma_i$ is now belonging to some other functional space different from the continuous piecewise linear one of the solution. We get back the Petrov-Galerkin approximation

$$\sum_{K \in \mathcal{T}_h} \int_K \left(\frac{\partial u^h}{\partial t} + \nabla \cdot \boldsymbol{f}(u^h) \right) w_i \, d\boldsymbol{x} = 0 \quad \forall i = 1, N+1 \tag{2.54}$$

We divide the time part from the advective one

$$\sum_{K \in \mathcal{T}_h} \int_K \varphi_j w_i \, d\boldsymbol{x} \, \frac{du_j}{dt} + \sum_{K \in \mathcal{T}_h} \int_K \nabla \cdot \boldsymbol{f}(u^h) w_i \, d\boldsymbol{x} = 0 \quad \forall i = 1, N+1 \tag{2.55}$$

Recognizing the mass-matrix $m_{ij}^K = \int_K \varphi_j w_i \, d\boldsymbol{x}$ and using the conservative linearization (2.43) we get

$$\sum_{K \in \mathcal{D}_i} \sum_{j \in K} m_{ij}^K \frac{du_j}{dt} + \sum_{K \in \mathcal{D}_i} \frac{\phi^K}{|K|} \int_K w_i \, dx = 0 \tag{2.56}$$

Calling

$$\frac{1}{|K|} \int_K w_i \, d\boldsymbol{x} = \beta_i^K \tag{2.57}$$

One gets the $\mathcal{RD}$ scheme in the classical formalism with consistent mass matrix

$$\sum_{K \in \mathcal{D}_i} \sum_{j \in K} m_{ij}^K \frac{du_j}{dt} + \sum_{K \in \mathcal{D}_i} \beta_i^K \phi^K = 0 \tag{2.58}$$

Two possible choices that satisfy (2.57) lead to the mass-matrix formulations used in our computations called respectively F1 and F2 are

$$w_i^{F1}(\boldsymbol{x}) = \beta_i^K \quad \boldsymbol{x} \in K \tag{2.59}$$

$$w_i^{F2}(\boldsymbol{x}) = \varphi_i(\boldsymbol{x}) + \beta_i^K - \frac{1}{3} \quad \boldsymbol{x} \in K \tag{2.60}$$

The mass matrix for both the formulation used in our computation, is then computed

$$\begin{aligned} m_{ij}^{K,F1} &= \int_K \varphi_j w_i^{F1} \, d\boldsymbol{x} \\ &= \frac{|K|}{3} \beta_i^K = |K| \hat{m}_{ij}^{K,F1} \end{aligned}$$

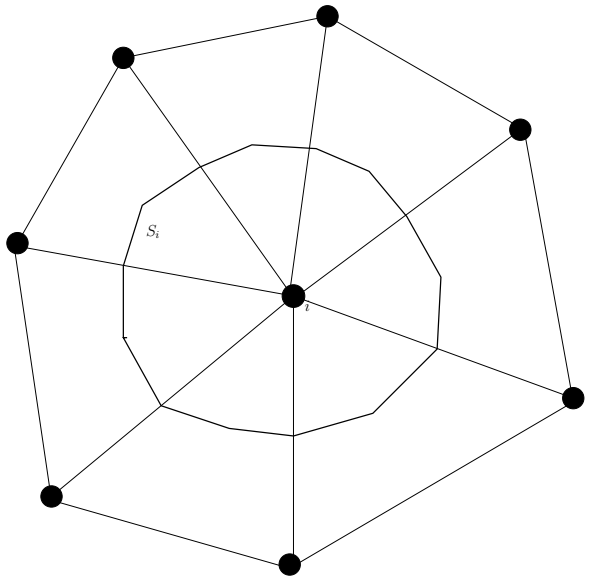

Figure 2.5: Median Duall Cell S_i

$$\begin{aligned} m_{ij}^{K,F2} &= \int_K \varphi_j w_i^{F2}\, d\boldsymbol{x} \\ &= \frac{|K|}{3}\left(\frac{\delta_{ij}}{4} + \beta_i^K - \frac{1}{12}\right) = |K|\hat{m}_{ij}^{K,F2} \end{aligned}$$

Hereinafter we will also make use extensively of the mass-lumped formulation (ML), specially in steady computations where mantainig accuracy in the time part is not important. Summing all the element on a given row, to the diagonal

$$m_{ij}^{K,ML} = \frac{|K|}{3}\delta_{ij} \tag{2.61}$$

Calling the median duall cell $|S_i| = \sum_{K\in\mathcal{D}_i} \frac{|K|}{3}$ rapresented in figure (2.5), (2.58) is then modified in the classical $\mathcal{RD}$ scheme with inconsistent mass matrix

$$|S_i|\,\frac{du_i}{dt} + \sum_{K\in\mathcal{D}_i} \phi_i^K = 0 \tag{2.62}$$

2.4 Stability and Accuracy for the compact prototype scheme

Any steady $\mathcal{RD}$ scheme can be formalized in the following abstract form

$$|S_i| \frac{du_i}{dt} = - \sum_{K \in \mathcal{D}_i} \sum_{j \in K, j \neq i} c_{ij}^K (u_i - u_j) \qquad \forall i \in \mathcal{T}_h \tag{2.63}$$

Proof. Substituting (2.47) in (2.62)

$$|S_i| \frac{du_i}{dt} = - \sum_{K \in \mathcal{D}_i} \phi_i^K = - \sum_{K \in \mathcal{D}_i} \sum_{j \in K, j \neq i} \left(-\beta_i^K k_j \right) (u_i - u_j)$$

we get the proof.

The abstract form is referred to as the *compact protoype scheme* from the compactness of the stencil, involving only the nearest neighboring nodes of node i. The compact discretization is another advantage of $\mathcal{RD}$ compared to Finite Volumes where, for high order schemes with polynomial reconstruction, wide stencil are used.

A stabilty analysis on (2.63) seems to be too restrictive, specially if unsteady phenomena are studied, as in our case. This has been a critical issue in extending $\mathcal{RD}$ to the study of unsteady problems. Instead the reference formalism will be still useful to us, because the full discretization that we are going to present in the next chapter can be recasted in a form similar to the prototype scheme.

2.4.1 Maximum Principle and Stability

In this paragraph we are interested firstly in stability based on discrete maximum principle: we search for some criteria that ensure $\mathcal{RD}$ schemes to satisfy the discrete analogue of the maximum principle which holds for the solution of (2.49)

$$\min_{\Omega} u_0(x, y) \leq u(x, y, t) \leq \max_{\Omega} u_0(x, y) \tag{2.64}$$

A very important property of the scheme written in form (2.63) is the so-called *Local Extremum Diminishing* (LED) property

Property(LED). *The prototype compact scheme (2.63) is Local Extremum Diminishing (LED) that is, in the numerical solution local maxima are non-increasing and local minima are non-decreasing, if*

$$\sum_{K \in \mathcal{D}_i \cap \mathcal{D}_j} c_{ij}^K = \tilde{c}_{ij}^K \geq 0 \quad \forall j \in \mathcal{D}_i,\ j \neq i \quad \text{and} \quad \forall i \in \mathcal{T}_h \tag{2.65}$$

Proof. It is very simple to show that

$$\begin{aligned}\frac{du_i}{dt} &= -\frac{1}{|S_i|}\sum_{K\in\mathcal{D}_i}\sum_{j\in K, j\neq i} c_{ij}^K\,(u_i-u_j)\\ &= -\frac{1}{|S_i|}\sum_{j\in\mathcal{D}_i, j\neq i}\left(\sum_{K\in\mathcal{D}_i\cap\mathcal{D}_j} c_{ij}^K\right)(u_i-u_j)\end{aligned}$$

- if u_i is a local maximum, then $u_i \geq u_j$, follows that (2.65) implies $\frac{du_i}{dt} \leq 0$
- if u_i is a local minimum, then $u_i \leq u_j$, follows that (2.65) implies $\frac{du_i}{dt} \geq 0$

A stronger requirement which is easy to control is obtained asking each c_{ij}^K to be positive, instead that their sum $\sum_{K\in\mathcal{D}_i\cap\mathcal{D}_j} c_{ij}^K$. This corresponds to ask that, in (2.67), the contribution of every element, taken separately, gives a solution which satisfy a LED property

$$\frac{du_i}{dt} = -\frac{1}{|S_i|}\sum_{j\in K, j\neq i} c_{ij}^K\,(u_i-u_j)$$

Repeating the demonstration for the above equation we get the following result

Property (sub-element LED). *The prototype compact scheme (2.63) is LED if*

$$c_{ij}^K \geq 0 \quad \forall j\in K,\ j\neq i \quad \text{and} \quad \forall K\in\mathcal{D}_i \tag{2.66}$$

Collecting for every element the coefficient $c_{ij}^K = -\beta_i^K k_j$ in a matrix, (2.66) is translated asking for the non-positivity of the off-diagonal terms.

We would like that the $\mathcal{RD}$ scheme satisfies the LED property but still this is not enough since new extrema can be created. We want a discrete version of maximum principle as it will be presented in a while. Before a fully discrete version of (2.63) is necessary. With Explicit Euler ($\mathcal{EE}$)

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{|S_i|}\sum_{K\in\mathcal{D}_i}\sum_{j\in K, j\neq i} c_{ij}^K\,(u_i-u_j) \tag{2.67}$$

A Discrete Maximum Principle is written locally for each sub-domain of Ω, $\mathcal{D}_i$, and for every time slab $[t^n, t^{n+1}]$

$$\min_{j\in\mathcal{D}_i} u_j^n \leq u_i^{n+1} \leq \max_{j\in\mathcal{D}_i} u_j^n \tag{2.68}$$

Property(Local Discrete Maximum Principle). *The protoype compact $\mathcal{RD}$-$\mathcal{EE}$ represented by (2.67) verifies a local maximum principle if the scheme is LED or sub-LED that is, (2.65) or (2.66) holds, and under the time-step restriction*

$$\triangle t \leq \frac{|S_i|}{\sum_{j\in\mathcal{D}_i, j\neq i}\tilde{c_{ij}^K}} \quad \forall i\in\mathcal{T}_h \tag{2.69}$$

Proof. The proof is immediate if u_j^{n+1} is written as a combination of the values of u_j^n

$$\begin{aligned} u_i^{n+1} &= u_i^n - \frac{\Delta t}{|S_i|} \sum_{j\in\mathcal{D}_i, j\neq i} \tilde{c}_{ij}^K \left(u_i - u_j\right) \\ &= \left(1 - \frac{\Delta t}{|S_i|} \sum_{j\in\mathcal{D}_i, j\neq i} \tilde{c}_{ij}^K\right) u_i^n + \frac{\Delta t}{|S_i|} \sum_{j\in\mathcal{D}_i, j\neq i} \tilde{c}_{ij}^K u_j^n \\ &= C_{ii} u_i^n + \sum_{j\in\mathcal{D}_i, j\neq i} C_{ij} u_j^n = \sum_{j\in\mathcal{D}_i} C_{ij} u_j^n \end{aligned}$$

LED property ensures that $C_{ij} \geq 0$, while the CFL-like condition (2.69) ensure that also $C_{ii} \geq 0$

$$\left(\sum_{j\in\mathcal{D}_i} C_{ij}\right) \min_{j\in\mathcal{D}_i} u_j^n \leq u_i^{n+1} \leq \left(\sum_{j\in\mathcal{D}_i} C_{ij}\right) \max_{j\in\mathcal{D}_i} u_j^n$$

and using the fact that $\sum_{j\in\mathcal{D}_i} C_{ij} = 1$ the result is prooved.

Property(Positivity). *A scheme that satisfy maximum principle is said to be positive ($\mathcal{P}$)*

Once we have a discrete maximum principle we can proof that the method is stable in L^∞-norm and that we have also a precise estimate for the bounds of the numerical solution

$$u_j^{n+1} = \mathcal{H}(u^n; j), \quad \|\mathcal{H}\|_{L^\infty(\Omega)} \leq C_s$$

Infact

Theorem(L^∞-stability). *If a local maximum principle is verified for all the time slabs $\{[t^n - t^{n+1}]\}_{n=0,\dots M-1}$ then the method (2.67) is L^∞-stable and the following bounds hold for its numerical solution*

$$\min_{j\in\mathcal{D}_i} u_0^n \leq u_i^{n+1} \leq \max_{j\in\mathcal{D}_i} u_0^n \quad \forall i \in \mathcal{T}_h,\ n \in [1, M] \tag{2.70}$$

We have provided ourself now of a nice criterium for the construction of schemes wich does not exhibit oscillatory behaviour near discontinuities.

2.4.2 Order of accuracy and Godunov Theorem

Apart from the stability another crucial issue for numerical methods is the order of accuracy. We search for a condition under which the solution of (2.63) is second order accurate, this being in general impossible for the lumping of the mass matrix. For this reason we only present accuracy results for steady state while for unsteady problems second order

accuracy will be achieved with a more complicated time approximation discussed in the next chapter. At steady state we want second order accuracy in some L-norm

$$\|u - u^h\|_V = \mathcal{O}(h^2) \tag{2.71}$$

We give the following result whose demontration is contained in [6]

Property (Second Order Accuracy). *A scheme in the form (2.63) produce a second order accurate solution at steady state if*

$$\phi_i^K = \mathcal{O}(h^3) \quad \forall K \in \mathcal{T}_h \text{ and } \forall i \in K \tag{2.72}$$

The following estimate for the residual holds

$$\phi^K = \mathcal{O}(h^3) \tag{2.73}$$

Proof. Starting from (2.43) the residual can be written

$$\begin{aligned} \phi^K &= \int_K \bar{\boldsymbol{a}} \cdot \nabla u^h \, d\boldsymbol{x} = \int_K \nabla \left(\bar{\boldsymbol{a}} \cdot u^h \right) \, d\boldsymbol{x} \\ &= \int_K \nabla \cdot \left(\bar{\boldsymbol{a}} u^h - \boldsymbol{a}(u) u \right) \, d\boldsymbol{x} = \int_{\partial K} \left(\bar{\boldsymbol{a}} u^h - \boldsymbol{a}(u) u \right) \cdot \boldsymbol{n} \, dl \\ &= \mathcal{O}(h^3) \end{aligned}$$

If u^h can be interpreted as the solution of the Petrov-Galerkin weak form (2.50) then, in case of P1 approximation (2.54) and smooth solution the following estimate holds

$$\|u^h - u\|_{L^2(\Omega)} = \mathcal{O}(h^2) \tag{2.74}$$

The following estimate is also true

$$\bar{\boldsymbol{a}} u^h - \boldsymbol{a} u = \mathcal{O}(u^h - u)$$

because $\boldsymbol{a}, \bar{\boldsymbol{a}}$ are bounded. Moreover $dl = \mathcal{O}(h)$. The result then is prooved.

Now the fundamental relation $\phi_i^K = \beta_i \phi^K$, together with (2.72)(2.73), lead to the following result

Property (Linearity Preserving scheme). *A $\mathcal{RD}$ scheme is linearity preserving if the distribution coefficients are uniformly bounded with respect to the solution and data of the problem, hence exists a constant C such that*

$$\max_{K \in \mathcal{T}_h} \max_{j \in K} |\beta_j^K| < C \quad \forall u^h, \boldsymbol{a}, u_0^h \tag{2.75}$$

A scheme which is linearity preserving is second order accurate at steady state.

We give the following definition:

Definition (Linearity). *A scheme in the form (2.63) is said to be linear if all c_{ij}^K are indipendent from the numerical solution.*

Unfortunatley a linear scheme cannot be positive and linearity preserving at the same time. This is stated by the following theorem [7], which is an extension of the one-dimensional Godunov theorem (a monotone scheme is at most first accurate).

Theorem (Godunov). *A linear scheme of the form (2.63) cannot be positive and second order accurate.*

This is a strong restriction that we have to face when we try to get an accurate numerical solution in smooth region without oscillations near discontinuities. To solve such a problem Godunov theorem forces us to move to non-linear schemes.

Finally we give the Lax-Wendroff theorem which confirms our hope that, once we have a method that produces a solution u^h which converges to some function u as the grid/time step is refined, then u is a weak solution.

Theorem(Lax Wendroff Theorem). Given initial data $u_0 \in L^\infty(\mathbb{R}^2)$, a function $u(\boldsymbol{x}, t) \in L^2(\mathbb{R}^2 \times \mathbb{R}^+)$ and a consistent and stable approximation $u^h(\boldsymbol{x}, t)$ of $u(\boldsymbol{x}, t)$ such that

$$\lim_{h,\Delta t \to 0} \|u_h - u\|_{L^2(\mathbb{R}^2 \times \mathbb{R}^+)} = 0$$

then u is a weak solution of the problem.

2.5 Distributions

The canonical $\mathcal{RD}$ formalism has been derived but what is inside ϕ_i^K is still missing. In this paragraph we present the different distributions that will be used in our numerical experiments in chapter 3. We start by simply extending some very known $\mathcal{FV}$ or $\mathcal{FE}$ one-dimensional schemes already introduced, showing that also in two-dimensions they can be absorbed or even recasted into $\mathcal{RD}$ form. After we investigate a new classes of schemes, the truly multidimensional upwind schemes, which can be easily construct within the $\mathcal{RD}$ philosophy and not so easily in different class of methods. Finally we provide two systematic methods to construct non linear schemes which are both positive and linearity preserving.

2.5.1 Lax Friederich scheme

A direct extension of the one-dimensional Lax-Friederich residual (2.36) is

$$\phi_i^{LxF} = \frac{1}{3}\phi^K + \alpha^K \sum_{j\in K, j\neq i} (u_i - u_j), \qquad \alpha^K \geq \max_{j\in K} |k_j| \tag{2.76}$$

The scheme is the result of a centered Galerkin scheme plus a diffusion term which introduces some form of stabilization to damp oscillations. Again a great amount of diffusion is introduced, unfortunately much more the one is effectively needed.

Using (2.45),(2.47)

$$\begin{aligned}
\phi_i^{LxF} &= \frac{1}{3}\sum_{j\in K} k_j u_j + \alpha^K \sum_{j\in K, j\neq i} (u_i - u_j) \\
&= -\frac{1}{3}\sum_{j\in K, j\neq i} k_j (u_i - u_j) + \alpha^K \sum_{j\in K, j\neq i} (u_i - u_j) \\
&= \frac{1}{3}\sum_{j\in K, j\neq i} \left(\alpha^K - k_j\right)(u_i - u_j)
\end{aligned}$$

Since, for the definition of the parameter α^K, $c_{ij}^K = \frac{1}{3}\left(\alpha^K - k_j\right) \geq 0$, the sub-element LED property (2.66) is verified. We can see the positive-stabilizing effect of the second term which overcome the possible negative contribution of the centered part. To get positivity for (2.69)

$$\Delta t \leq \frac{3|S_i|}{\sum_{K\in\mathcal{D}_i}\sum_{j\in K, j\neq i}(\alpha^K - k_j)} \tag{2.77}$$

For Godunov theorem the method is only first order accurate, moreover it does not respect property (2.48), hence it is not upwind.

2.5.2 SUPG scheme

The $\mathcal{RD}$ version of SUPG scheme is derived from the Petrov-Galerkin $\mathcal{FE}$ method if a conservative linearization is used. Considering only the advective part for simplicity (but the time part can be treated as usual with the same analogy)

$$\int_\Omega \nabla\cdot\boldsymbol{f}(u^h)\varphi_i\, d\boldsymbol{x} + \sum_{K\in\mathcal{T}_h}\int_K \nabla\cdot\boldsymbol{f}(u^h)\tau\boldsymbol{a}(u^h)\cdot\nabla\varphi_i\, d\boldsymbol{x} = 0 \quad \forall i \in \mathcal{T}_h \tag{2.78}$$

Rearranging

$$\sum_{K\in\mathcal{D}_i}\int_K \nabla\cdot\boldsymbol{f}(u^h)\left(\varphi_i + \tau\boldsymbol{a}(u^h)\cdot\nabla\varphi_i\right) d\boldsymbol{x} = \sum_{K\in\mathcal{D}_i}\int_K \nabla\cdot\boldsymbol{f}(u^h) w_i\, d\boldsymbol{x} \tag{2.79}$$

Using passage (2.45) and computing β through (2.57) one gets

$$\beta_i^{SUPG} = \frac{1}{|K|}\int_K \left(\varphi_i + \tau \boldsymbol{a}(u^h)\cdot\nabla\varphi_i\right) d\boldsymbol{x} = \frac{1}{3} + \tau\bar{\boldsymbol{a}}\cdot\frac{1}{2|K|}\boldsymbol{n}_i = \frac{1}{3} + \tau\frac{k_i}{|K|} \tag{2.80}$$

From a $\mathcal{RD}$ point of view the streamline upwind term introduces some kind of upwind bias in the centred distribution even if the scheme is not upwind. From (2.80) is not possible to prove sub-LED property while one can easily check for property (2.75) that ensures second order accuracy. In our computation we fix for each element the parameter τ as follow

$$\tau = \frac{|K|}{\left(\sum_{j\in K}|k_j|\right)} \tag{2.81}$$

The SUPG distribution then reads

$$\beta_i^{SUPG} = \frac{1}{3} + k_i T, \quad T = \left(\sum_{j\in K}|k_j|\right)^{-1} \tag{2.82}$$

2.5.3 N scheme

The Narrow (N) scheme is a successful first order positive scheme. Infact using the multidimensional upwind property of $\mathcal{RD}$ schemes, the basic idea proposed by Roe was to construct a scheme without oscillations near discontinuities but, at the same time, with as low cross-diffusion as possible. We give immediately the residual distribution for the N scheme

$$\phi_i^N = k_i^+(u_i - u_{in}) \tag{2.83}$$

where the inflow velocity is computed

$$u_{in} = -\sum_{j\in K} N k_j^- u_j, \quad N = \left(\sum_{j\in K} k_j^+\right)^{-1} \tag{2.84}$$

For the one-target case trivially all the residual is sent to the downstream node. For the two-target case of figure (2.6), where the downstream nodes are 1 and 2, one have

$$\phi_1^N = k_1(u_1 - u_3), \quad \phi_2^N = k_2(u_2 - u_3) \tag{2.85}$$

which allows a simple geometrical interpratation of the splitting philosphy between the two nodes. Infact decomposing the local averaged advective field along the parallel to the inflow faces $\bar{\boldsymbol{a}} = \bar{\boldsymbol{a}}_1 + \bar{\boldsymbol{a}}_2$

$$\begin{aligned}
\phi^K &= \int_K (\bar{\boldsymbol{a}}_1 + \bar{\boldsymbol{a}}_2)\cdot\nabla u_h \, d\boldsymbol{x} \\
&= \int_K \bar{\boldsymbol{a}}_1\cdot\nabla u_h \, d\boldsymbol{x} + \int_K \bar{\boldsymbol{a}}_2\cdot\nabla u_h \, d\boldsymbol{x} \\
&= \phi^K(\bar{\boldsymbol{a}}_1) + \phi^K(\bar{\boldsymbol{a}}_2)
\end{aligned}$$

Developing for example the first term, with the obvious fact that $\bar{\boldsymbol{a}}_1 \cdot \boldsymbol{n}_2 = 0$

$$\begin{aligned}\phi^K(\bar{\boldsymbol{a}}_1) &= \bar{\boldsymbol{a}}_1 \cdot \frac{1}{2}\sum_{j\in K} \boldsymbol{n}_j u_j = \frac{1}{2}\bar{\boldsymbol{a}}_1 \cdot (\boldsymbol{n}_1 u_1 + \boldsymbol{n}_3 u_3) \\ &= \frac{1}{2}\bar{\boldsymbol{a}}_1 \cdot \boldsymbol{n}_1 (u_1 - u_3) = \frac{1}{2}\bar{\boldsymbol{a}}_1 \cdot \boldsymbol{n}_1 (u_1 - u_3) \\ &= k_1(u_1 - u_3)\end{aligned}$$

The last passage is obtained becouse of simple geometry

$$\frac{1}{2}\bar{\boldsymbol{a}}_1 \cdot \boldsymbol{n}_1 = \frac{1}{2}(\bar{\boldsymbol{a}}_1 + \bar{\boldsymbol{a}}_2) \cdot \boldsymbol{n}_1 = \frac{1}{2}\bar{\boldsymbol{a}} \cdot \boldsymbol{n}_1 = k_1$$

The total residual can then be expressed as in (2.83)

$$\phi^K = k_1(u_1 - u_3) + k_2(u_2 - u_3) = \phi_1^N + \phi_2^N \tag{2.86}$$

Thus the scheme reduces to a one-dimensional upwind scheme along the direction of each of the two inflow edges. The N scheme introduces upwinding cleverly in two-dimensions.

We can check the positivity

$$\begin{aligned}\phi_i^N &= k_i^+ u_i + k_i^+ \sum_{j\in K} N k_j^- u_j \\ &= -\sum_{j\in K} k_i^+ N k_j^- u_i + \sum_{j\in K} k_i^+ N k_j^- u_j \\ &= -\sum_{j\in K} k_i^+ N k_j^- (u_i - u_j) \\ &= -\sum_{j\in K} k_i^+ N k_j^- (u_i - u_j)\end{aligned}$$

The group $c_{ij}^K = -k_i^+ N k_j^- \geq 0$, the scheme is sub-element LED and it is also positive if time step is choosen according to (2.69)

$$\triangle t \leq \frac{|S_i|}{\sum_{K\in\mathcal{D}_i}\left(-\sum_{j\in K} k_i^+ N k_j^-\right)} = \frac{|S_i|}{\sum_{K\in\mathcal{D}_i} k_i^+} \tag{2.87}$$

2.5.4 LDA scheme

After the construction of a positive upwind scheme we search for its linearity-preserving counterpart. The Low Diffusion A scheme works very good in computing smooth numerical solution being second order in accuracy for a P1 approximation of the solution. Of

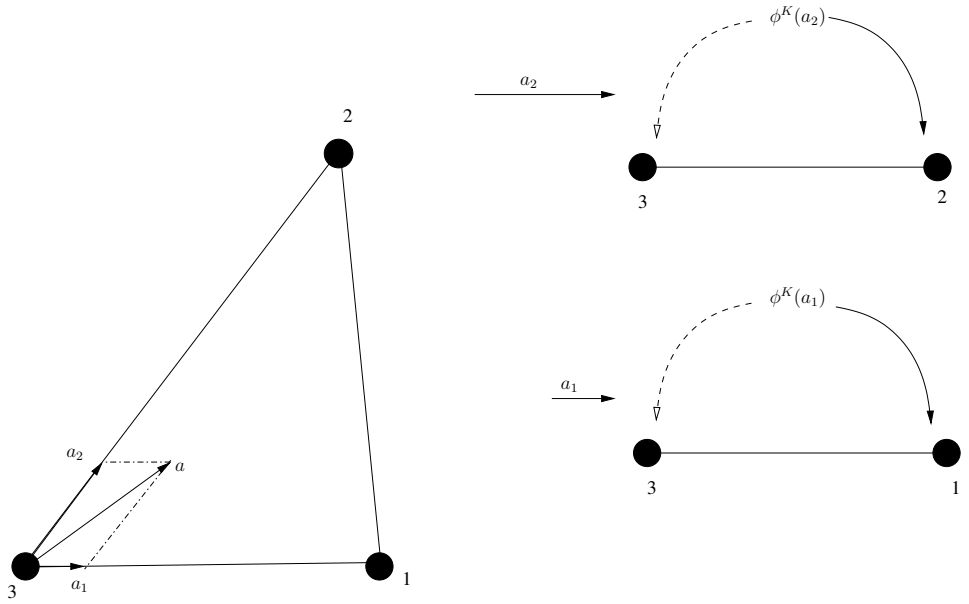

Figure 2.6: Geometrical interpretation of N scheme

course, for discontinuous solutions, one has no guarantees against the rising of oscillations since no maximum principle holds in this case.

The LDA scheme is defined by the following distribution coefficients

$$\beta_i^{LDA} = k_i^+ N \tag{2.88}$$

which is a bounded distribution independently on $\phi(u^h)$ and property (2.75) is satisfied.

The idea of a true multidimensional upwinding is to envoy the most of the residual to the most downstream node. If in one-target case the splitting is trivial and all $\mathcal{RD}$ upwind schemes coincide (LDA and N), in two-target case we need a measure to decide which one, among the two, is the most downstream nodes. The LDA scheme use a simple geometrical consideration. Using the notation of the element in figure (2.7) we compute the two areas in which the element is splitted by the line parallel to the local avereged advective and passing from the upstream node

$$|T_{423}| = \frac{l_{34} \cdot k_1}{\|\boldsymbol{a}\|}, \quad |T_{143}| = \frac{l_{34} \cdot k_2}{\|\boldsymbol{a}\|}$$

The distribution can be written

$$\beta_1^{LDA} = \frac{k_1}{k_1 + k_2} = \frac{|T_{423}|}{|T|} \quad \beta_2^{LDA} = \frac{k_2}{k_1 + k_2} = \frac{|T_{143}|}{|T|}$$

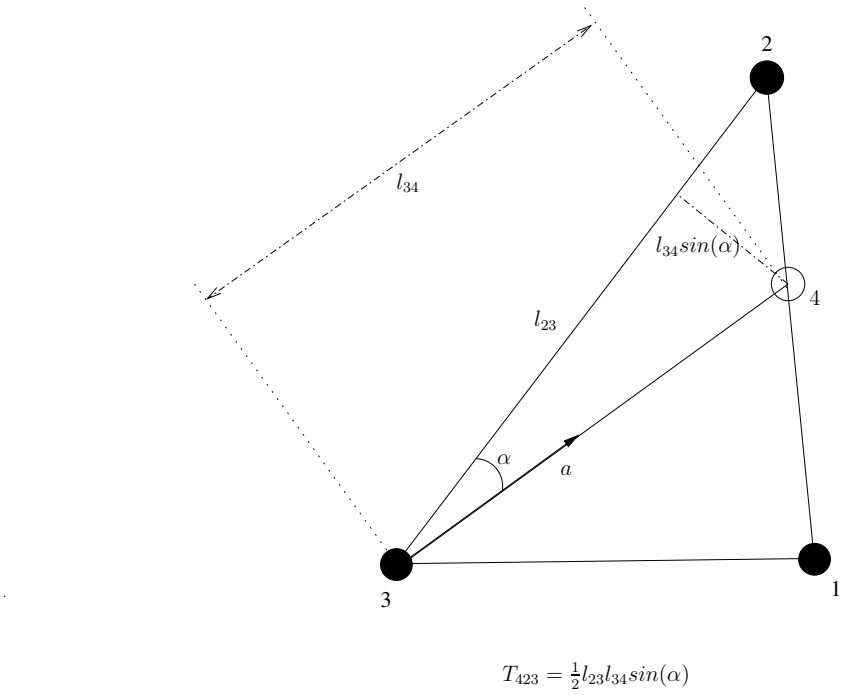

Figure 2.7: Geometrical interpratation of LDA scheme

Starting from a one target situation where all the residual is sent to node 1, as long as the outflow point moves from node 1 to node 2 more residual is sent to node 2 in a measure proportional to the area of the triangle T_{143} until a one target situation is reached again this time at node 2.

2.5.5 Blended schemes

All the linear schemes analyzed so far are or $\mathcal{P}$ (LxF,N) or $\mathcal{LP}$ (SUPG,LDA) but not both for Godunov theorem. Now we want to construct a class of schemes which ensures high accuracy without oscillations near discontinuties. In this paragraph we present the classical approach, common also to high-resolution $\mathcal{FV}$ methods, of a non-linear blending between a $\mathcal{P}$ scheme and a $\mathcal{LP}$ one. The idea is to use the $\mathcal{LP}$ scheme in regions where the numerical solution is smooth and the $\mathcal{P}$ scheme just across discontinuties. The difference beween $\mathcal{RD}$ and $\mathcal{FV}$ is that, instead of blending fluxes, this time we blend residuals

$$\phi_i^B = \left(1 - l(u^h)\right) \phi_i^{\mathcal{LP}} + l(u^h) \phi_i^{\mathcal{P}} \tag{2.89}$$

Even if the idea is simple, the design of the limiter $l(u^h)$ is not simple at all. It has to be of order $\mathcal{O}(h^2)$ when the solution is smooth and of order $\mathcal{O}(1)$ when the solution is

discontinuous but a more rigourous study of the limiter is needed in order to get a scheme which strictly guarantees the above properties. In this work we don't follow the rigorous way since there is an heuristic definition for $l(u^h)$ by Deconinck and Van der Weide which works very well in many cases

$$l = \frac{|\phi^K|}{\sum_j |\phi_j^N|} \tag{2.90}$$

This definition only ensure, for $l(u^h)$, the switch to the correct order of magnitude across discontinuities but do not satisfies the positivity requirement: imposing a sub-element LED properties for the resultant scheme (2.89) implies the satisfaction of three equations which lead to different constraints on the limiter that can't be satisfied in general by only one parameter (2.90).

In the following we will consider the blending between the LDA scheme and the N scheme both sharing the multidimensional upwind property. The resulting blended scheme is reffered to LDA-N scheme

$$\phi_i^{LDA-N} = \left(1 - l(u^h)\right) \phi_i^{LDA} + l(u^h)\phi_i^N \tag{2.91}$$

2.5.6 Limited schemes

Another root to the construction of non linear schemes which are $\mathcal{LP}$ and $\mathcal{P}$ consists in limiting the unbounded coefficients of a $\mathcal{P}$ scheme in such a way that the the resulting scheme is completely parameter free, this rapresenting an objective advantage respect to the blending approach.

The problem is the following. Consider a triangle and assume we are given a residual distribution that define a first order $\mathcal{P}$ scheme (ϕ_1, ϕ_2, ϕ_3). We want to construct a second order scheme defined by residuals $(\phi_1^\star, \phi_2^\star, \phi_3^\star)$. Calling the first order and second order weights respectively

$$x_i = \frac{\phi_i}{\phi^K} \quad \beta_i = \frac{\phi_i^\star}{\phi^K}$$

The resulting scheme has to be

1. Conservative
2. LED
3. Linearity Preserving

The above problem can be reformulated as finding a mapping $(x_1, x_2, x_3) \to (\beta_1, \beta_3, \beta_3)$ such that the scheme satisfy the properties (1,2,3). This properties can be translated directly in some constraints on β_i

1. Conservation: $\sum_{j \in K} \beta_j = 1$

2. LED: we prefer the more restricting sub-element LED

$$\phi_i^\star = \sum_{j\in K, i\neq j} c_{ij}^\star \left(u_i - u_j\right), \quad c_{ij}^\star \geq 0$$

The residual is rewritten as follows

$$\phi_i^\star = \frac{\phi_i^\star}{\phi^K}\frac{\phi^K}{\phi_i}\phi_i = \beta_i \frac{1}{x_i} \sum_{j\in K, i\neq j} c_{ij}\left(u_i - u_j\right)$$

We ask for the positivity of the coefficients $c_{ij}^\star = \frac{\beta_i}{x_i} c_{ij} \geq 0$ which, since $c_{ij} \geq 0$ by definition, is equivalent to $\beta_i x_i \geq 0$

3. Linearity Preserving: β_i bounded for any i

We provide a geometrical interpretation of the above constraints. Since $\sum_{j\in K} x_j = 1$ and $\sum_{j\in K} \beta_j = 1$ we can interprete coefficients $(x_1, x_2, x_3), (\beta_1, \beta_3, \beta_3)$ as the baricentric coordinates of the points L and H respect to an abstract reference triangle of vertices (A_1, A_2, A_3)

$$L = A_1 x_1 + A_2 x_2 + A_3 x_3 \quad H = A_1\beta_1 + A_2\beta_2 + A_3\beta_3$$

No condition are given for (x_1, x_2, x_3), except $\sum_{j\in K} x_j = 1$, so they could be unbounded with the point L everywhere in the domain. Our problem then is to find a mapping that project L onto a bounded subdomain, in this way the boundedness of the the coefficients $(\beta_1, \beta_3, \beta_3)$, is ensured. Moreover any bounded region will be a correct subdomain onto which we can carry on the projection, for example the disk in figure (2.8). A possible choice is the PSI mapping proposed by Struijs [7]

$$\beta_i = \frac{x_i^+}{\sum_{j\in K} x_j^+} \tag{2.92}$$

In this case, since β_i are not only bounded but also $0 \leq \beta_i \leq 1$, the projection subdomain is the triangle (A_1, A_2, A_3). Properties (1)(3) are clearly satisfied but also property (2) holds since

$$if \quad x_i \geq 0 \Rightarrow \beta_i \geq 0$$
$$if \quad x_i \leq 0 \Rightarrow \beta_i = 0$$

thus, the scheme is also LED.

The PSI limiting procedure can be applied to any first order linear scheme such as the Lax-Friederich scheme (LLxF or LxF-PSI) and the N-scheme (LN or N-PSI)

$$\beta_i^{LN} = \frac{\left(\beta_i^N\right)^+}{\sum_{j\in K}\left(\beta_j^N\right)^+}, \quad \beta_i^{LLxF} = \frac{\left(\beta_i^{LxF}\right)^+}{\sum_{j\in K}\left(\beta_j^{LxF}\right)^+} \tag{2.93}$$

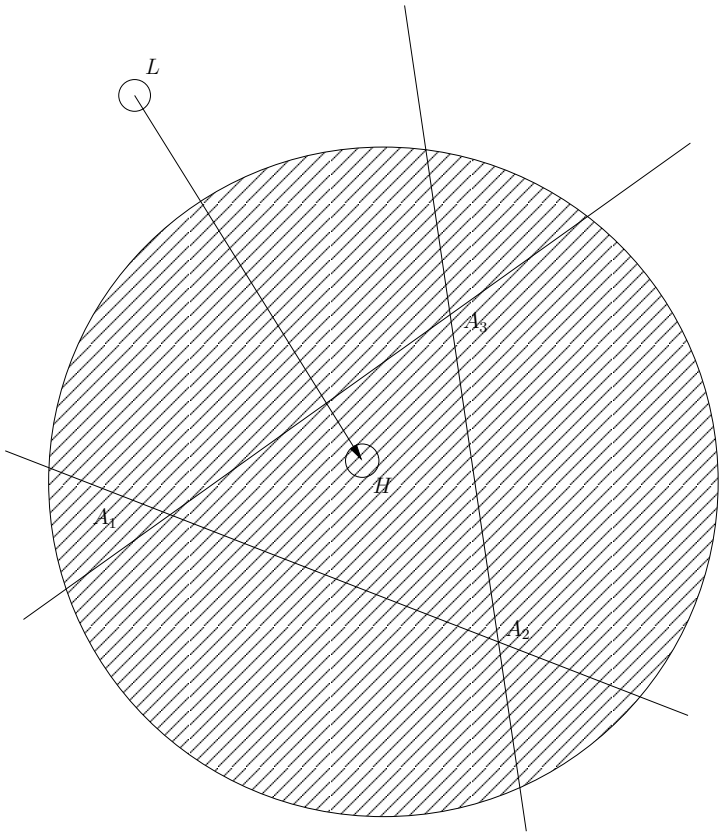

Figure 2.8: Geometrical rapresentation of the mapping $(x_1, x_2, x_3) \rightarrow (\beta_1, \beta_3, \beta_3)$

Multiplying by the total residual we get another way of rewriting the limiting. For example for for Limited Lax-Friederich

$$\beta_i^{LLxF} = \frac{\left(\phi_i^{LxF}\right)^+}{\sum_{j\in K}\left(\phi_j^{LxF}\right)^+} \tag{2.94}$$

2.5.7 Erratic convergence for LLxF scheme: LLxFs scheme

For limited schemes things, at end, turned out not so simple. Infact looking to the solution obtained with LLxF one can observe the appereance of *wiggles* on the isolines, wiggles that gives very poor results in term of accuracy and destroy the convergence property expected. This oscillations can be seen as the rising of some unexpected destibilizing phenomena even if the scheme, from the rigorous analysis followed in the previous paragraph, remains perfectly stable in L^∞-norm. A partial confirmation of this is that discontinuity are well handed while in the smooth region we encounter poor accuracy. Abgrall in [8] provides some argument to support the following hypothesis. The mapping $(x_1, x_2, x_3) \to (\beta_1, \beta_3, \beta_3)$ is done according to the sign of the distribution coefficients and not by using some consideration regarding the upwind character of the final distribution. In other words, as seen in figure (2.9), for a LxF distribution which is in general three target, the PSI limiting procedure can result in a distribution which is no more upwind, thus introducing small amplitude oscillations. Another explanation arrives from the work of Barth who performed an energy stability analysis for the PSI scheme in the simple case of scalar linear advection. Even for such a simple case, he found that source of energy instability might be introduced when the PSI limiter compresses the distribution from a three target case to two or even one target. Which is exactly our case.

A possible solution to cure the problem is suggesed in the same paper [8]. A SUPG term is added, with a limiter to tune the diffusion introduced. The final distribution for such a scheme called LLxF stabilized or briefly LLxFs reads

$$\beta_i^{LLxFs} = \beta_i^{LLxF} + \delta(u^h)\beta_i^{SUPG} \tag{2.95}$$

Now the problem is definition of the limiter. A definition calibrated for preserving the positivity is not useful since positivity is lost anyway by the addition of the SUPG term. Hence, for every triangle, we can use the following choice

$$\delta(u^h) = \min\left(1, \frac{1}{\frac{|\phi^K|}{\bar{u}h_K^2} + \varepsilon}\right) \tag{2.96}$$

with h_K the element reference size, $\bar{u} = \max_{j\in K}|u_j|$ and $\varepsilon = 10^{-10}$. It is easy to show that the definition (2.96) can detect the discontinuties. Infact $\delta(u^h)$ is of order $\mathcal{O}(1)$ in smooth region where dissipation is needed to damp oscillations and of order $\mathcal{O}(h^{-1})$ across discontinuities where the LLxF scheme behave nicely computing well-resolved profiles.

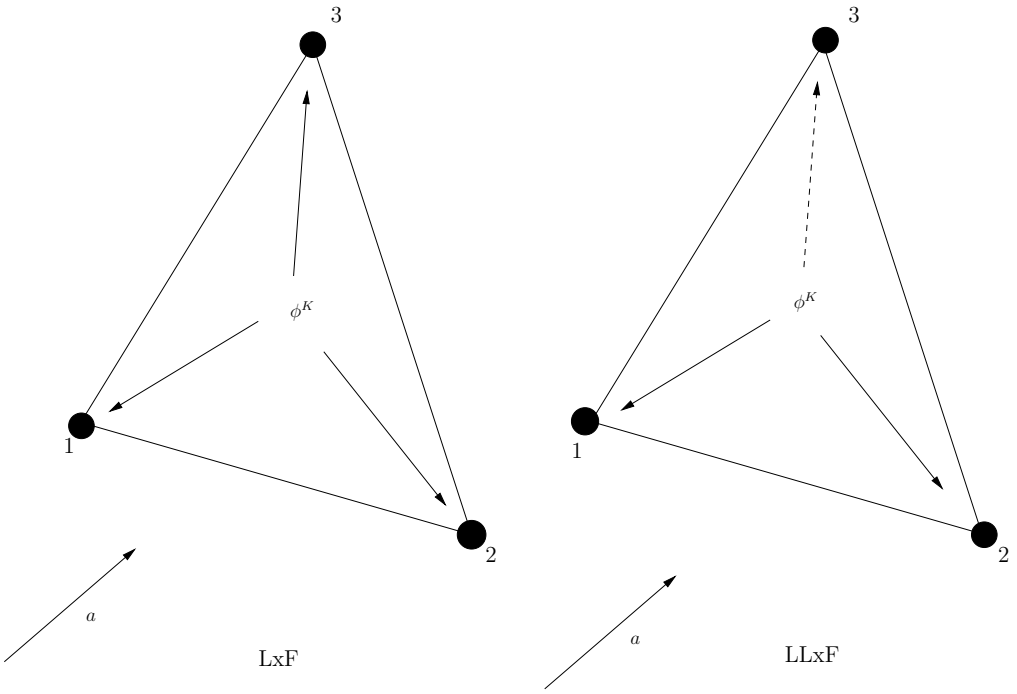

Figure 2.9: Source of erratic convergence for LLxF scheme

3 Time discretization

In this brief chapter we quickly present a time approximation to integrate conservation law (2.49). We search for a time discretization that gives second order of accuracy and, in particular that leads to an explicit scheme. The explicit Runge Kutta scheme of Ricchiuto and Abgrall [9] matches both the requirments and will fit very nicely to our application.

For a $\mathcal{RD}$ scheme written in the compact form (2.63) and discretized in time with an Explicit Euler approximation, we have seen that many results about positivity under a CFL condition exist and moreover the scheme remains compact. For steady problem an accurate discretization in time is not needed, the resulting approximation (2.67) works very good and, if accuracy requirement are matched, it gives second (or third) order accurate results. Unfortunately for unsteady problems, where we need an accurate time discretization, the lumped formulation gives only first order accurate solution even if very accurate time/space discretization are used [6]. For this reason Ferrante [10] developed a monotone and accurate scheme starting from the consistent mass-matrix formulation (2.58). The problem of this approach is that positivity is no more guaranteed directly and to correct the problem, many aspects of a $\mathcal{RD}$ methods are lost (compactness and upwinding for example). Abgrall and Mezine [11] proposed a space-time reinterpreatation of $\mathcal{RD}$ which lead them to design a second order positive scheme. The method is implicit and has to satisfy a CFL condition in order to be positive. The advantage of implicit method (not have a constraint on time step) is lost so why not try to develop a second order explicit method which will be way faster. This is discussed by Ricchiuto and Abgrall [9].

3.1 Explicit Runge Kutta

In the present section we construct, following step by step [9], a second order discretization for time dependent conservation law in the form of (2.49)

Given an ODE

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

calling the difference between the solution at the k-th step and the one at the n-th step

$\Delta u^k = u^k - u^n$, the classical Explicit Runge Kutta two reads

$$\begin{cases} \frac{\Delta u^1}{\Delta t} + e^1 = 0, & e^1 = e(u^n) \\ \frac{\Delta u^2}{\Delta t} + e^2 = 0, & e^2 = \frac{1}{2}\left(e(u^n) + e(u^1)\right) \end{cases}$$

with the generic k-th step being

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

This time approximation is second order accurate in time.

Step-shifted stabilization operator

Let us now apply the $\mathcal{RK}2$ discretization to the the unsteady conservation law approximated in space with the stabilized $\mathcal{FE}$ method, hence to the time continuous equation (2.53). We get

$$\int_\Omega \varphi_i \left(\frac{\Delta u_h^k}{\Delta t} + \nabla \cdot \boldsymbol{f}^k(u_h) \right) d\boldsymbol{x} + \sum_{K \in \mathcal{D}_i} \int_K \gamma_i \left(\frac{\Delta u_h^k}{\Delta t} + \nabla \cdot \boldsymbol{f}^k(u_h) \right) d\boldsymbol{x} = 0 \tag{3.1}$$

Such a scheme leads to a nonlinaer algebraic system because the stabilization bubble depends strongly on the solution. In order to avoid this, the term $\frac{\Delta u^k}{\Delta t}$ in the bubble contribution is replaced by a shifted approximation of the time derivative $\frac{\Delta \overline{u^k}}{\Delta t}$ where $\Delta \overline{u^k} = u^{k-1} - u^n$. This shifting guarantees to end up with an explicit linear scheme, since the term under analysis is now dependent on quantities already computed at the previous step, without deteriorating the overall accuracy. (3.1) can be rearranged in the following form

$$\int_\Omega \varphi_i \frac{\Delta u_h^k}{\Delta t}\, d\boldsymbol{x} - \int_\Omega \varphi_i \frac{\Delta \overline{u_h^k}}{\Delta t}\, d\boldsymbol{x} = -\sum_K \Phi_i^{RK(k)} = 0 \tag{3.2}$$

$$\text{with} \quad \Phi_i^{RK(k)} = \int_K w_i \left(\frac{\Delta \overline{u_h^k}}{\Delta t} + \nabla \cdot \boldsymbol{f}^k(u_h) \right) d\boldsymbol{x} \tag{3.3}$$

Putting in evidence all the mass matrices and passing to a $\mathcal{RD}$ formalism

$$\sum_{K \in \mathcal{D}_i} \sum_j m_{ij}^{GAL} \frac{\Delta u_j^k}{\Delta t} = -\sum_{K \in \mathcal{D}_i} \left(\Phi_i^{RK(k)} - \sum_j m_{ij}^{GAL} \frac{\Delta \overline{u_j^k}}{\Delta t} \right) \tag{3.4}$$

$$\Phi_i^{RK(k)} = \sum_{j \in K} m_{ij}^K \frac{\Delta \overline{u_j^k}}{\Delta t} + \phi_i^{RK(k)} \tag{3.5}$$

Exploiting the two steps

$$\Phi_i^{RK(1)} = \phi_i(u_h^n) \tag{3.6}$$

$$\Phi_i^{RK(2)} = \sum_{j\in K} m_{ij}^K \frac{\left(u_j^1 - u_j^n\right)}{\Delta t} + \frac{\Delta t}{2}\left(\phi_i(u_h^1) + \phi_i(u_h^n)\right) \tag{3.7}$$

High order mass lumping

(3.4) still requires the inversion of the Galerkin mass matrix at each Runge Kutta step. The algorithm would be very powerful if we decouple all the equations with a lumping strategy which preserve the accuracy of the method. The idea then is to replace the exact Galerkin integral by means of a quadrature formula

$$\int_K \varphi_i \frac{\Delta u_h^k}{\Delta t}\, d\boldsymbol{x} \simeq \sum_{x_q} \varphi_i(x_q) \frac{\Delta u_h^k(x_q)}{\Delta t} \omega_q |K| \tag{3.8}$$

In [12] it is shown that a sufficient condition for this approximation to keep a truncation error $\mathcal{O}(h^2)$ is for the quadrature formula to integrate exactly polynomials of degree one. The interesting application of this approach is that, when the quadrature points coincide with the degree of freedom of the element, one can use the property $\varphi_i(x_j) = \delta_{ij}$

$$\sum_{x_q} \varphi_i(x_q) \frac{\Delta u_h^k(x_q)}{\Delta t} \omega_q |K| = \sum_j \varphi_i(x_j) \frac{\Delta u_h^k(x_j)}{\Delta t} \omega_j |K| = \frac{1}{3}|K| \frac{\Delta u_i^k}{\Delta t} \tag{3.9}$$

Accuracy is mantained because the choosen quadrature formula with weights all equal to 1/3 integrates exactly polynomials of order one. The unsteady term then reads

$$\int_\Omega \varphi_i \frac{\Delta u_h^k}{\Delta t}\, d\boldsymbol{x} = \sum_K \frac{1}{3}|K| \frac{\Delta u_i^k}{\Delta t} = |S_i| \frac{\Delta u_i^k}{\Delta t} \tag{3.10}$$

Selective Lumped (SL) Formulation

If we decide to lump only the first integral, then the final $\mathcal{RK}2 - \mathcal{RD}$ approximation at the k-th step, reads

$$|S_i| \frac{\Delta u_i^k}{\Delta t} = -\sum_{K\in\mathcal{D}_i} \left(\Phi_i^{RK(k)} - \sum_{j\in K} m_{ij}^{GAL} \frac{\Delta \overline{u_j^k}}{\Delta t} \right) \tag{3.11}$$

The weak point of this formulation is that it cannot be recasted in the form (2.67) for the presence of the last term. All the results given in section (2.4.1) are no more true and, in general, we cannot expect a monotone beahviour near discontinuities. In order to recover positivity a lumping on the second Galerkin integral is needed, leading to the Global Lumped Formulation of the $\mathcal{RK}2$

Global Lumped (GL) Formulation

Lumping the second Galerkin mass matrix in (3.4) we get for the first and second step respectively

$$\begin{cases} |S_i|\frac{u_i^1-u_i^n}{\Delta t} = -\sum_{K\in\mathcal{D}_i}\Phi_i^{RK(1)} \\ |S_i|\frac{u_i^{n+1}-u_i^1}{\Delta t} = -\sum_{K\in\mathcal{D}_i}\Phi_i^{RK(2)} \end{cases} \tag{3.12}$$

(3.12) closely follows the form of the prototype comapact scheme (2.67). It can be demonstrated, using the positive coeffcient theory, that if, the residuals $\phi_i(u^h)$ are positive, (3.12) define a positive scheme.

3.1.1 Application to the different schemes

Now we explictly compute the total residual $\Phi_i^{RK(k)}$ from the formula (3.5), for each scheme presented in section (2.5). For clarity the superscript $RK(k)$ has been changed in $XXX(k)$ where XXX is the shorthand notation for the scheme. For positive schemes the definition of a mass matrix is not so clear because of the unboundness of the distribution coefficients. Moreover these schemes are only first order accurate, thus one can lump the mass matrix.

For linear schemes the total residual to be substituted in (3.11) or (3.12) is

$$\Phi_i^{LxF(k)} = \frac{|T|}{3}\frac{\Delta\overline{u^k}_i}{\Delta t} + \phi_i^{LxF(k)} \tag{3.13}$$

$$\Phi_i^{SUPG(k)} = \sum_{j\in K} m_{ij}^{SUPG}\frac{\Delta\overline{u_j^k}}{\Delta t} + \phi_i^{SUPG(k)} \tag{3.14}$$

$$\Phi_i^{LDA(k)} = \sum_{j\in K} m_{ij}^{LDA}\frac{\Delta\overline{u_j^k}}{\Delta t} + \phi_i^{LDA(k)} \tag{3.15}$$

$$\Phi_i^{N(k)} = \frac{|T|}{3}\frac{\Delta\overline{u^k}_i}{\Delta t} + \phi_i^{N(k)} \tag{3.16}$$

For the non-linear limited LLxF scheme we use directly the limiting operation (2.94) with the total residual defined in this contest

$$\beta_i^{LLxF} = \frac{\left(\Phi_i^{LxF(k)}\right)^+}{\sum_{j\in K}\left(\Phi_j^{LxF(k)}\right)^+} \tag{3.17}$$

Finally the limiter (2.96) is rewritten according to the modified form of the residual

$$\delta(u^h) = \min\left(1, \frac{1}{\frac{|\Phi^K|}{\Delta t \bar{u}|\boldsymbol{a}|h_K^2} + \varepsilon}\right) \tag{3.18}$$

4 Residual Distribution schemes for moving grids

Many unsteady problems governed by conservation laws involve the movement of the boundaries. In the numerical approximation of this phenomena additional difficulties arise because the grid must adapt at every time step to the moving boundaries. If equations are written in Eulerian framework as in (2.49), which is very well suited for fluid dynamics, this in turn makes necessary an intermediate step between the computation of two successive numerical solutions. Infact, once the grid has been adapted to the new boundaries, an interpolation of the previous solution over the new grid is essential in order to start up the computation of the new solution. On the opposite conservation laws can be written in a Lagrangian framework where we move the grid at the same velocity of the particles and we write integral conservation laws always for the same particles. The algorithm works always on the same grid, with the same nodes, and no interpolation step is needed. The problem of this approach is that the grid movement is connected to the particles paths and when strong distortion are present, like in a fluid dynamic contest, the method suffer from instabilities.

The Arbitrary Lagrangian Eulerian formulation solves the drawbacks of both the approaches since conservation laws are written for an arbitrary moving grid with respect to the particles motion. Since the grid movement is arbitrary, one of the possibilities that are exploited is the adaptation of the grid according not only to boundary motion but also to some refinement criteria, for example, with an appropriate algorithm, gathering the nodes where strong gradients of the solution appear or are expected. This issue has not been addressed in the present thesis. Our attention has been focused on the correct formulation of the ALE approach within a Residual Distribution method.

The appereance of the ALE approach dates back to the early eighties due to the contribution of J.Donea [13]. The idea was found very appealing in many field of computational continuum mechanics because of the fact that the extension of a classical Lagrangian (for solid mechanics) or Eulerian (for fluid mechanics) method into ALE formulation is straightforward and requires few lines of changes in the algorithm. In fluid dynamics the recasting of Eulerian Finite Volumes and Finite Elements has been investigated since long time with important contribution by Farhat, Geuzaine and Grandmont [14]. For Residual Distribution schemes, developments occured at Von Karman Institute with the work of

Michler and Deconinck [15] who achieved first order with an Explicit Euler time integrator and later Dobes [16] who, in his PhD thesis, moved to high order time approximation (BDF3, Crank Nicholson), this allowing him to have second order of accuracy. The aim of this thesis is to obtain a numerical solution with second order of accuracy using a faster explicit time integrator.

4.1 Actual configuration, material configuration and reference configuration

Assuming that we are given a domain Ω and field of displacements that brings every point of the domain from the reference position $\boldsymbol{X}$ to the actual one $\boldsymbol{x}(t)$ and that this field is governed by an arbitrary given motion law

$$\frac{d\boldsymbol{x}(t)}{dt} = \boldsymbol{\sigma}(\boldsymbol{x}, t), \tag{4.1}$$

Solving the ODE (4.1) starting from the reference configuration $\boldsymbol{x}(0) = \boldsymbol{X}$, gives back, at every time instant, the actual configuration through the following mapping

$$A(t) : \Omega_X \to \Omega_x(t), \qquad \boldsymbol{x} = A(\boldsymbol{X}, t) \tag{4.2}$$

We define the Jacobian matrix of the mapping as

$$\mathcal{J}_A = \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{X}}, \quad J_A = \det \mathcal{J}_A \neq 0$$

We introduce now another set of coordinates, the Lagrangian or material coordinates $\boldsymbol{\chi}$, and a mapping that describes the motion of each particle. This mapping returns the physical location, rapresented by the actual coordinate $\boldsymbol{x}$, of the particle marked with $\boldsymbol{\chi}$ at time t

$$B(t) : \Omega_\chi \to \Omega_x(t), \qquad \boldsymbol{x} = B(\boldsymbol{\chi}, t) \tag{4.3}$$

Defining the Jacobian matrix of the mappig

$$\mathcal{J}_B = \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{X}} \quad J_B = \det \mathcal{J}_B \neq 0$$

If u is a conserved quantity it can be expressed as a function of the different coordinates $\boldsymbol{x}, \boldsymbol{X}, \boldsymbol{\chi}$ and three different time derivatives can be defined. If the derivation is computed in the actual configuration one has

$$\left.\frac{\partial u(\boldsymbol{x}, t)}{\partial t}\right|_x = \frac{\partial u}{\partial t} \qquad \text{spatial derivative} \tag{4.4}$$

If it is computed following the particle motion one has

$$\left.\frac{\partial u(\boldsymbol{\chi},t)}{\partial t}\right|_{\chi} = \frac{du}{dt} \qquad \text{material derivative} \tag{4.5}$$

Finally if it is computed following the domain motion one has

$$\left.\frac{\partial u(\boldsymbol{X},t)}{\partial t}\right|_{X} \qquad \text{referential derivative} \tag{4.6}$$

Moreover two different velocities can be computed: the particle velocity and the domain velocity

$$\left.\frac{\partial x(t)}{\partial t}\right|_{X} = \frac{dA(\boldsymbol{X},t)}{dt} = \boldsymbol{\sigma} \tag{4.7}$$

$$\left.\frac{\partial x(t)}{\partial t}\right|_{\chi} = \frac{dB(\boldsymbol{\chi},t)}{dt} = \boldsymbol{a} \tag{4.8}$$

The transformation theorem provides a relation between the above derivatives and these velocities

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + \boldsymbol{a}(\boldsymbol{x},t)\cdot\nabla u(\boldsymbol{x},t) \tag{4.9}$$

$$\frac{du}{dt} = \left.\frac{\partial u}{\partial t}\right|_{X} + (\boldsymbol{a}(\boldsymbol{x},t) - \boldsymbol{\sigma}(\boldsymbol{x},t))\cdot\nabla u(\boldsymbol{x},t) \tag{4.10}$$

From continuum mechanics we need also the followings

$$\left.\frac{\partial J_B}{\partial t}\right|_{\chi} = \frac{dJ_B}{dt} = J_B\nabla\cdot\boldsymbol{a} \tag{4.11}$$

$$\left.\frac{\partial J_A}{\partial t}\right|_{X} = J_A\nabla\cdot\boldsymbol{\sigma} \tag{4.12}$$

The last one is commonly called Geometric Conservation Law (GCL) and rapresents a constraint the points of the domain have to satisfy during their arbitrary motion. This will be very important when developing a numerical method with a moving grid; up to now we only want to make clear that the movement of the domain is arbitrary but within hypothesis (4.12).

4.2 ALE equations

The conservation of the scalar quantity u can be written, dependig on convenience, in the different coordinate frameworks. If we choose a material control volume $C(t)$ which

contains always the same particles, following them throughout all the domain, the conservation is simply stated in actual coordinates

$$\frac{d}{dt}\int_{C(t)} u(\boldsymbol{x},t)\,d\boldsymbol{x} = 0 \tag{4.13}$$

Passing to material coordinates and using (4.11) together with the transformation theorem (4.9)

$$\begin{aligned}\frac{d}{dt}\int_{C_\chi} u(\boldsymbol{X},t)J_B\,d\boldsymbol{X} &= \int_{C_\chi}\frac{d}{dt}\left(u(\boldsymbol{\chi},t)J_B\right)\,d\boldsymbol{\chi}\\ &= \int_{C_\chi}\left(\frac{du}{dt}J_B + u\frac{dJ_B}{dt}\right)\,d\boldsymbol{\chi}\\ &= \int_{C_\chi}\left(\frac{\partial u}{\partial t} + \boldsymbol{a}\cdot\nabla u + u\nabla\cdot\boldsymbol{a}\right)J_B\,d\boldsymbol{\chi}\end{aligned} \tag{4.14}$$

We have derived the conservation law in integral Eulerian form

$$\int_C\left(\frac{\partial u}{\partial t} + \nabla\cdot\boldsymbol{f}\right)\,d\boldsymbol{x} = 0 \tag{4.15}$$

Now, in the intermediate passage (4.14), we use (4.10) instead of (4.9)

$$\begin{aligned}\int_{C_\chi}\left(\frac{du}{dt}J_B + u\frac{dJ_B}{dt}\right)\,d\boldsymbol{\chi} &= \int_{C_\chi}\left(\left.\frac{\partial u}{\partial t}\right|_X + (\boldsymbol{a}-\boldsymbol{\sigma})\cdot\nabla u + u\nabla\cdot\boldsymbol{a}\right)J_B\,d\boldsymbol{\chi}\\ &= \int_{C(t)}\left(\left.\frac{\partial u}{\partial t}\right|_X + \nabla\cdot\boldsymbol{f} - \boldsymbol{\sigma}\cdot\nabla u\right)\,d\boldsymbol{x}\end{aligned} \tag{4.16}$$

The first term can be rewritten if we compute the derivative of the conserved quantity inside a control volume $C(t)$, which is following the motion of the points of the domain. Note that there is a little abuse in the notation since $C(t)$ has been already used to rapresent a material volume. Transforming into referential coordinate and using the fact that C_X does not depend on time

$$\begin{aligned}\left.\frac{\partial}{\partial t}\right|_X\int_{C(t)} u(\boldsymbol{x},t)\,d\boldsymbol{x} &= \left.\frac{\partial}{\partial t}\right|_X\int_{C_X} u(\boldsymbol{X},t)J_A\,d\boldsymbol{X} = \int_{C_X}\left.\frac{\partial\left(J_A u\right)}{\partial t}\right|_X\,d\boldsymbol{X}\\ &= \int_{C_X}\left.\frac{\partial u}{\partial t}\right|_X J_A\,d\boldsymbol{X} + \int_{C_X}\left.\frac{\partial J_A}{\partial t}\right|_X u\,d\boldsymbol{X}\\ &= \int_{C_X}\left.\frac{\partial u}{\partial t}\right|_X J_A\,d\boldsymbol{X} + \int_{C_X} J_A u\nabla\cdot\boldsymbol{\sigma}\,d\boldsymbol{X}\end{aligned} \tag{4.17}$$

So we have

$$\int_{C(t)} \left.\frac{\partial u}{\partial t}\right|_X d\boldsymbol{x} = \left.\frac{\partial}{\partial t}\right|_X \int_{C(t)} u\, d\boldsymbol{x} - \int_{C(t)} u\nabla\cdot\boldsymbol{\sigma}\, d\boldsymbol{x} \tag{4.18}$$

Substituting (4.18) in (4.16) we get the integral form of conservation law written in Arbitrary Lagrangian Eulerian Formulation (ALE)

$$\left.\frac{\partial}{\partial t}\right|_X \int_{C(t)} u\, d\boldsymbol{x} + \int_{C(t)} \nabla\cdot(\boldsymbol{f} - u\boldsymbol{\sigma})\; d\boldsymbol{x} = 0 \tag{4.19}$$

which express the conservation of u contained in a control volume which is moving arbitrarly. The equilibrium is reached by the relative flux of u entering and leaving the volume with velocity $\boldsymbol{a} - \boldsymbol{\sigma}$.

It is interesting that, for the arbitrariety of the movement, the ALE formulation rapresents a generalization of both the Eulerian and Lagrangian formulations. Infact in (4.19)

1. If $\boldsymbol{\sigma} = \boldsymbol{0}$, the control volume is fixed in space (from $C(t)$ to C) and we get the Eulerian form (4.15)
2. If $\boldsymbol{\sigma} = \boldsymbol{a}$, the control volume is moving with the particle motion and we get the Lagrangian form (4.13)

A differential form of conservation law in ALE formulation is needed but its derivation is simple if we start from the integral form (4.19) and we use (4.12)

$$\left.\frac{\partial}{\partial t}\right|_X \int_{C_X} uJ_A\, d\boldsymbol{X} + \int_{C_X} J_A\nabla\cdot(\boldsymbol{f} - u\boldsymbol{\sigma})\; d\boldsymbol{X} =$$

$$= \int_{C_X} \left(\left.\frac{\partial\,(J_A u)}{\partial t}\right|_X + J_A\nabla\cdot(\boldsymbol{f} - u\boldsymbol{\sigma}) \right) d\boldsymbol{X} = 0$$

Using the localization principle, the differential form of conservation law in ALE formulation is derived

$$\left.\frac{\partial\,(J_A u)}{\partial t}\right|_X + J_A\nabla\cdot(\boldsymbol{f} - u\boldsymbol{\sigma}) = 0 \tag{4.20}$$

It is easy to see that the requirement for volume conservation (4.12) we have previously done can be derived simply by imposing a state of uniform flow in (4.20). In this case we are modelling a situation in which the flow is uniform and the domain is moving from behind.

Developing the derivative in (4.20) and then substituting (4.12)

$$J_A\left.\frac{\partial u}{\partial t}\right|_X + J_A u\nabla\cdot\boldsymbol{\sigma} + J_A\nabla\cdot(\boldsymbol{f} - u\boldsymbol{\sigma}) = 0$$

which lead to the following equation that we will use extensively hereinafter

$$\left.\frac{\partial u}{\partial t}\right|_X + \nabla\cdot\boldsymbol{f} - \boldsymbol{\sigma}\cdot\nabla u = 0 \tag{4.21}$$

4.3 Galerkin Finite Element

We start with the approximation of (4.20) both in time and space. The domain is initially approximated with an unstructured triangulation Ω_h^X, then mapping (4.2) produce a time-continuous transformation of the grid $\Omega_h^X \to \Omega_h(t)$. The time discretization will make us evaluate the grid at instants t^n generating a set of grid $\Omega_h(t^n) = \Omega_h^n$.

Moreover we ask our numerical method to satisfy a discrete version of the GCL condition (4.12), often referred to as Discrete Geometric Conservation Law (DGCL). Referring to the interpretation previously given, we are asking the method to preserve the state of uniform flow, which is a natural requirement for every nice scheme. We start with the simple Galerkin Finite Element space approximation which allows us a simple satisfaction of the GCL at a discrete level.

4.3.1 Galerkin method

We proceed in building the classical Galerkin method on the conservation law in ALE framework (4.20), with solution u^h, test function φ_i and grid velocity $\boldsymbol{\sigma}^h = \sum_{j \in K} \varphi_j \boldsymbol{\sigma}_j$ belonging to the space of piecewise linear polynomials

$$\int_{\Omega_h^X} \left(\left.\frac{\partial (J_A u^h)}{\partial t}\right|_X + J_A \nabla \cdot (\boldsymbol{f}(u^h) - \boldsymbol{\sigma}^h u^h) \right) \varphi_i \, d\boldsymbol{X} = 0 \tag{4.22}$$

Since the the configuration Ω_X does not depend on time and assuming $\left.\frac{\partial \varphi_i}{\partial t}\right|_X = 0$ we can take the time derivative outside the integral

$$\left.\frac{\partial}{\partial t}\right|_X \int_{\Omega_h^X} \varphi_i J_A u^h \, d\boldsymbol{X} + \int_{\Omega_X} \varphi_i J_A \nabla \cdot (\boldsymbol{f}(u^h) - \boldsymbol{\sigma}^h u^h) \, d\boldsymbol{X} = 0$$

Passing to the current coordinates $\boldsymbol{x}$ we have the Galerkin approximation for (4.20)

$$\left.\frac{\partial}{\partial t}\right|_X \int_{\Omega_h(t)} \varphi_i u^h \, d\boldsymbol{x} + \int_{\Omega_h(t)} \varphi_i \nabla \cdot (\boldsymbol{f}(u^h) - \boldsymbol{\sigma}^h u^h) \, d\boldsymbol{x} = 0 \tag{4.23}$$

If the flow is uniform we get the time continuous and space discrete approximation of (4.12)

$$\begin{aligned} \left.\frac{\partial}{\partial t}\right|_X \int_{\Omega_h(t)} \varphi_i \, d\boldsymbol{x} &= \int_{\Omega_h(t)} \varphi_i \nabla \cdot \boldsymbol{\sigma}^h \, d\boldsymbol{x} \\ \sum_{K \in \mathcal{D}_i} \frac{1}{3} \left.\frac{\partial K(t)}{\partial t}\right|_X &= \sum_{K \in \mathcal{D}_i} \frac{1}{3} \int_{K(t)} \nabla \cdot \boldsymbol{\sigma}^h \, d\boldsymbol{x} \end{aligned}$$

which reads as follows

$$\left.\frac{\partial K(t)}{\partial t}\right|_X = \int_{K(t)} \nabla \cdot \boldsymbol{\sigma}^h \, d\boldsymbol{x} \tag{4.24}$$

From (4.24) we clearly see that the satisfaction of the the GCL at a discrete level is related to the time scheme that one is using to integrate the conservation law. If one uses $\mathcal{BDF}3$ rather then $\mathcal{RK}2$ then, the different approximations of the left-hand side will lead to different ways of verifying exactly (4.24). We stress the fact that the DGCL is specific to the time scheme. If equation (4.24), approximated in time with the same scheme used to integrate the conservation law, is exactly satisfied, then the method is said to satisfy the *Discrete Geometric Conservation Law*.

4.3.2 An example of a DGCL satisfing scheme

A very useful time integrator is $\mathcal{BDF}2$ which provide second order accuracy in time. We explain how to satisfy the DGCL for $\mathcal{BDF}2$ following a method proposed by Dobes in a $\mathcal{RD}$ framework [16]. The weak form is obtained starting from a sligthly different form of (4.20) obtained by splitting the ALE flux term and using the fact that $\boldsymbol{a} = \boldsymbol{a}(u)$

$$\left.\frac{\partial \left(J_A u\right)}{\partial t}\right|_X + J_A \left(\boldsymbol{a} - \boldsymbol{\sigma}\right) \nabla \cdot u - J_A u \nabla \cdot \boldsymbol{\sigma} = 0 \tag{4.25}$$

Since we want to satisfy the GCL condition we substitute (4.12) into the above equation

$$\left.\frac{\partial}{\partial t}\right|_X \int_{\Omega_X} J_A \varphi_i u_h \, d\boldsymbol{X} + \int_{\Omega_X} J_A \varphi_i \left(\boldsymbol{a}(u^h) - \boldsymbol{\sigma}^h\right) \cdot \nabla u^h \, d\boldsymbol{X} - \left.\frac{\partial J_A}{\partial t}\right|_X \int_{\Omega_X} \varphi_i u_h \, d\boldsymbol{X} = 0$$

The last term is usually reffered to as Geometric Source Term. After some calculation the algorithm is rewritten

$$\sum_{K \in \mathcal{D}_i} \sum_{j \in K} \hat{m}_{ij}^{GAL} \left.\frac{\partial \left(|K| u_j\right)}{\partial t}\right|_X + \int_{\Omega_h(t)} \varphi_i \left(\boldsymbol{a}(u^h) - \boldsymbol{\sigma}^h\right) \cdot \nabla u^h \, d\boldsymbol{x} - \sum_{K \in \mathcal{D}_i} \left.\frac{\partial |K|}{\partial t}\right|_X \sum_{j \in K} \hat{m}_{ij}^{GAL} u_j = 0 \tag{4.26}$$

where $\hat{m}_{ij}^{GAL} = \frac{\delta_{ij}}{4} + \frac{1}{4}$ is the Galerkin mass matrix. As we can see, satisfying the GCL is completely different from satisfing the DGCL for which we have to discretize both the derivatives with the same time discrete operator. Proceeding in this fashion we are sure to balance, element by element, the volume variation in the time step with the integral of the grid velocity flux along the boundaries of the element (4.24).

This approach has a nice recasting into a $\mathcal{RD}$ framework. Infact the second term in (4.26) is already in a quasi-linear form, so it can be written in a $\mathcal{RD}$ form through a conservative

linearization. Since the grid velocity is approximated with P1 interpolation the correct conservative linearization of the ALE part is immediate

$$\int_{\Omega(t)} \varphi_i \boldsymbol{\sigma}^h \cdot \nabla u^h \, d\boldsymbol{x} = \sum_{j \in K} \frac{\boldsymbol{\sigma}_j}{3} \nabla u^h |K| = \bar{\boldsymbol{\sigma}} \nabla u^h |K|$$

The upwind parameter with the ALE correction naturally becomes

$$k_i = \frac{1}{2} (\bar{\boldsymbol{a}} - \bar{\boldsymbol{\sigma}}) \cdot \boldsymbol{n}_i \tag{4.27}$$

If $\alpha^{n+1}, \alpha^n, \alpha^{n-1}$are the coefficients of $\mathcal{BDF}2$, the Galerkin $\mathcal{RD}$ scheme then reads

$$\sum_{K \in \mathcal{D}_i} \sum_{j \in K} \hat{m}_{ij}^{GAL} \frac{\alpha^{n+1}|K^{n+1}|u_j^{n+1} + \alpha^n |K^n| u_j^n + \alpha^{n-1}|K^{n-1}|u_j^{n-1}}{\Delta t} + \sum_{K \in \mathcal{D}_i} \beta_i^{GAL} \sum_{j \in K} k_j^{n+1} u_j^{n+1} +$$

$$- \sum_{K \in \mathcal{D}_i} \frac{\alpha^{n+1}|K^{n+1}| + \alpha^n |K^n| + \alpha^{n-1}|K^{n-1}|}{\Delta t} \sum_{j \in K} \hat{m}_{ij}^{GAL} u_j = 0 \tag{4.28}$$

where $\beta^{GAL} = \frac{1}{3}$ is the distribution coefficient for the Galerkin method. (4.28) satisfies the DGCL by construction. This is supposed to be just an example since Galerkin method for hyperbolic problems is unstable.

4.3.3 A Farhat approach

Keeping in mind that our objective is a method verifying the time discrete counterpart of (4.24), in this paragraph we proceed in a different way, according to what suggested by Farhat in [17]. The main idea is that many of the most used time discretizations satisfy naturally the DGCL condition by the choice of a proper grid velocity and of a proper quadrature rule for the integrals.

First, we present some useful results that will be use everywhere hereinafter. Integrating (4.24) in the timeslab $[t^n, t^{n+1}]$ provides

$$K^{n+1} - K^n = \int_{t^n}^{t^{n+1}} \int_{\partial K(t)} \nabla \cdot \boldsymbol{\sigma}^h \, d\boldsymbol{x} dt \tag{4.29}$$

We have already observed that great care has to be put, when building the numerical method, in order to satisfy exactly the above equation. This can be done with simple geometry and algebra. Since the triangle area can be computed as $|K| = \frac{1}{2} \sum_{j \in K} \boldsymbol{x}_j \cdot \boldsymbol{k}_j$

with

$$\begin{aligned} |K|^{n+1} - |K|^n &= \frac{1}{2}\sum_{j\in K}\left[(\boldsymbol{x}_j\cdot\boldsymbol{k}_j)^{n+1} - (\boldsymbol{x}_j\cdot\boldsymbol{k}_j)^n\right] \\ &= \frac{1}{2}\sum_{j\in K}\left[\boldsymbol{x}_j^{n+1/2}\cdot\left(\boldsymbol{k}_j^{n+1}-\boldsymbol{k}_j^n\right) + \boldsymbol{k}_j^{n+1/2}\cdot\left(\boldsymbol{x}_j^{n+1}-\boldsymbol{x}_j^n\right)\right] \\ &= \sum_{j\in K}\boldsymbol{k}_j^{n+1/2}\cdot\left(\boldsymbol{x}_j^{n+1}-\boldsymbol{x}_j^n\right) \end{aligned} \tag{4.30}$$

If we set the grid velocity

$$\boldsymbol{\sigma}_j^* = \frac{\boldsymbol{x}_j^{n+1}-\boldsymbol{x}_j^n}{\Delta t} \tag{4.31}$$

We can recast (4.30) in the following form

$$|K|^{n+1} - |K|^n = \Delta t \int\limits_{K^{n+1/2}} \nabla\cdot\boldsymbol{\sigma}_h^*\, d\boldsymbol{x} \tag{4.32}$$

We have proved that, in order to satisfy (4.29), a natural choice for the grid velocity is (4.31) while the configuration on which we perform integrations should be the midpoint one between t^n and t^{n+1}.

We found the result of Farhat for which it is crucial to establish in (4.29) where the time integral must be computed and the same question arise for the grid velocity. Since the left-end side is always computed exactly, an appropriate scheme for evaluating exactly $\int_{t^n}^{t^{n+1}}\int_{\Omega_h}\nabla\cdot\boldsymbol{\sigma}_h\, d\boldsymbol{x}\, dt$ is presented

$$\int\limits_{t^n}^{t^{n+1}}\int_K \nabla\cdot\boldsymbol{\sigma}_h\, d\boldsymbol{x}\, dt = \Delta t\int_{K^{n+\frac{1}{2}}}\nabla\cdot\boldsymbol{\sigma}_h^*\, d\boldsymbol{x} \tag{4.33}$$

This result is very useful once we have discretized in time (4.23).

Explicit Euler $\mathcal{EE}$

Discretizing in time (4.23) with $\mathcal{EE}$ we have

$$\frac{\Delta}{\Delta t}\int_{\Omega_h(t)}\varphi_i u_h\, d\boldsymbol{x} + \int_{\Omega_h(t)}\varphi_i\nabla\cdot\left(\boldsymbol{f}(u_h^n) - \boldsymbol{\sigma}^h u_h^n\right)\, d\boldsymbol{x} = 0$$

We have still to face the problem of satisfying the DGCL, both $\boldsymbol{\sigma}$ and $\Omega_h(t)$ are undefined infact. Imposing a uniform flow, one sees that the satisfaction of the DGCL conditon passes through (4.32). Setting $\boldsymbol{\sigma}_j = \boldsymbol{\sigma}_j^*$ and $\Omega_h(t) = \Omega_h^{n+1}$ we close the problem

$$\int\limits_{\Omega_h^{n+1}}\varphi_i u_h^{n+1}\, d\boldsymbol{x} - \int\limits_{\Omega_h^n}\varphi_i u_h^n\, d\boldsymbol{x} + \Delta t\int\limits_{\Omega_h^{n+1/2}}\varphi_i\nabla\cdot(\boldsymbol{f}(u_h^n) - \boldsymbol{\sigma}_h^* u_h^n)\, d\boldsymbol{x} = 0 \tag{4.34}$$

Implicit Euler $\mathcal{IE}$

We proceed as before, setting $\boldsymbol{\sigma}_j = \boldsymbol{\sigma}_j^*$ and $\Omega_h(t) = \Omega_h^{n+1}$ and we found that the DGCL condition is again satisfied.

$$\int\limits_{\Omega_h^{n+1}} \varphi_i u_h^{n+1}\, d\boldsymbol{x} - \int\limits_{\Omega_h^n} \varphi_i u_h^n\, d\boldsymbol{x} + \Delta t \int\limits_{\Omega_h^{n+1}} \varphi_i \nabla \cdot \left(\boldsymbol{f}(u_h^{n+1}) - \boldsymbol{\sigma}_h^* u_h^{n+1} \right) d\boldsymbol{x} = 0 \tag{4.35}$$

Infact imposing a uniform flow one finds (4.32).

Crank-Nicholson $\mathcal{CN}$

Also in this case, the imposition of a uniform flow leads to equation (4.32) which is exactly satisfied for $\boldsymbol{\sigma}_j = \boldsymbol{\sigma}_j^*$ and $\Omega_h(t) = \Omega_h^{n+1}$. Hence we have

$$\int\limits_{\Omega_h^{n+1}} \varphi_i u_h^{n+1} - \int\limits_{\Omega_h^n} \varphi_i u_h^n +$$

$$+\frac{\Delta t}{2} \left(\int\limits_{\Omega_h^{n+1}} \varphi_i \nabla \cdot (\boldsymbol{f}(u_h^{n+1}) - \boldsymbol{\sigma}_h^* u_h^{n+1}) + \int\limits_{\Omega_h^{n+1}} \varphi_i \nabla \cdot (\boldsymbol{f}(u_h^n) - \boldsymbol{\sigma}_h^* u_h^n) \right) = 0$$

4.4 Stabilized Finite Element and Residual Distribution

The Galerkin method provide a centered approximation of the advective part leading to an unstable numerical solution. To cure this problem the stabilized Finite Element method is invoked in this paragraph.

In the stabilization form $\mathcal{L}_h(u^h, \varphi_i)$, many choices that satisfies (2.51) are possible for the operator $L(u)$:

1. Using the conservation law in conservative form (4.20) one gets

$$\mathcal{L}_h = \sum_K \int_{K_X} \gamma_i \left(\left. \frac{\partial \left(J_A u \right)}{\partial t} \right|_X + J_A \nabla \cdot (\boldsymbol{f} - u\boldsymbol{\sigma}) \right) d\boldsymbol{X} \tag{4.36}$$

2. Using the mixed formulation (4.21)

$$\mathcal{L}_h = \sum_K \int_K \gamma_i \left(\left. \frac{\partial u}{\partial t} \right|_X + \nabla \cdot \boldsymbol{f} - \boldsymbol{\sigma} \cdot \nabla u \right) d\boldsymbol{x} \tag{4.37}$$

3. Farhat within a Finite Element method in [17] uses instead the Eulerian formulation

$$\mathcal{L}_h = \sum_K \int_K \gamma_i \left(\frac{\partial u}{\partial t} + \nabla \cdot \boldsymbol{f} \right) d\boldsymbol{x} \tag{4.38}$$

4.4.1 Explicit Euler

In the next section we show two different ways of formulating a GCL-satisfying stabilized $\mathcal{FE}$-$\mathcal{RD}$ using (4.36) as stabilization term.

Dobes Closure

We can rearrange the weak form as

$$\int_{\Omega_h^X} \left(\left.\frac{\partial (J_A u^h)}{\partial t}\right|_X + J_A \nabla \cdot (\boldsymbol{f}(u^h) - \boldsymbol{\sigma}^h u^h) \right) w_i \, d\boldsymbol{X} = 0 \quad \forall i \in \mathcal{T}_h \tag{4.39}$$

where the test function is $w_i = \varphi_i + \gamma_i$.

Reapeating the passages of paragraph (2.2.2) with the only difference that now we have w_i instead of φ_i, we get

$$\sum_{K \in \mathcal{D}_i} \sum_{j \in K} \hat{m}_{ij}^K \left. \frac{\partial \left(|K| u_j \right)}{\partial t} \right|_X + \int_{\Omega(t)} w_i \left(\boldsymbol{a}(u^h) - \boldsymbol{\sigma}^h \right) \cdot \nabla u^h \, d\boldsymbol{x} - \sum_{K \in \mathcal{D}_i} \left. \frac{\partial |K|}{\partial t} \right|_X \sum_{j \in K} \hat{m}_{ij}^K u_j = 0$$

where $\hat{m}_{ij}^K$ is the general mass matrix that depends on the test function, introduced in section (2.3). Discretizing the time derivative with Explict Euler, lumping the mass-matrix, and using the $\mathcal{FEM} - \mathcal{RD}$ analogy

$$\frac{|S_i^{n+1}| u_i^{n+1} - |S_i^n| u_i^n}{\Delta t} + \sum_{K \in \mathcal{D}_i} \beta_i \sum_{j \in K} k_j^n u_j^n - \frac{|S_i^{n+1}| - |S_i^n|}{\Delta t} u_i^n = 0 \tag{4.40}$$

For the presence of the Geometric Source Term, the above scheme is not written in the compact prototype form but one can prove that a sub-element positivity property still holds. The scheme for a single element is written

$$\frac{|K^{n+1}| u_i^{n+1} - |K^n| u_i^n}{3\Delta t} + \sum_{j \in K} c_{ij}^K u_j^n - \frac{|K^{n+1}| - |K^n|}{3\Delta t} u_i^n = 0$$

$$u_i^{n+1} = \left(\frac{|K^n| + |K^{n+1}|}{|K^{n+1}|} - \frac{3\Delta t}{|K^{n+1}|} c_{ii}^K \right) u_i^n - \frac{3\Delta t}{|K^{n+1}|} \sum_{j \in K, j \neq i} c_{ij}^K u_j^n$$

Positivity is ensured with the following CFL-like condition

$$\Delta t \leq \frac{|K^n| + |K^{n+1}|}{3 c_{ii}^K} \quad \forall i, K \in \mathcal{T}_h$$

Dobes used this approach together with second order implicit time schemes, in particular $\mathcal{BDF}2$ with consistent mass-matrix - getting the stabilized version of the algorithm of paragraph (2.2.2) - and $\mathcal{CN}$ with lumped mass matrix, obtaining very good results. We have to mention only that, if a consistent mass-matrix formulation is used, then positivity is spoiled.

Another closure

In this thesis we suggest another closure to the problem which is somewhat simple. We start again from equation (4.39) discretized in time with $\mathcal{EE}$, we split again the ALE flux term, finally we use grid velocity (4.31) and midpoint configuration.

$$\frac{\Delta}{\Delta t}\int_{\Omega(t)} w_i u^h \, d\boldsymbol{x} + \int_{\Omega^{n+1/2}} w_i \left(\nabla \cdot \boldsymbol{f}(u_h^n) - \boldsymbol{\sigma}_h^* \cdot \nabla u_h^n\right) d\mathbf{x} - \int_{\Omega^{n+1/2}} w_i u_h^n \nabla \cdot \boldsymbol{\sigma}_h^* \, d\boldsymbol{x} = 0 \tag{4.41}$$

If a uniform flow is imposed one gets

$$\int_{\Omega_h^{n+1}} w_i \, d\boldsymbol{x} - \int_{\Omega_h^n} w_i \, d\boldsymbol{x} = \Delta t \int_{\Omega_h^{n+1/2}} w_i \nabla \cdot \boldsymbol{\sigma}_h^* \, d\boldsymbol{x}$$

It is easy to check that the above equation is verified exactely. Infact for property (2.57) we have

$$\sum_{K\in\mathcal{D}_i} \beta_i^K \left(|K|^{n+1} - |K|^n\right) = \Delta t \sum_{K\in\mathcal{D}_i} \beta_i^K \int_{\Omega_h^{n+1/2}} \nabla \cdot \boldsymbol{\sigma}_h^* \, d\boldsymbol{x}$$

Thus, we get again (4.32) which is an identity.

We recall the following property which come from the fact that, given two quantity A,B, then $\Delta(AB) = \overline{B}\Delta A + \overline{A}\Delta B$

$$\Delta|_n^{n+1}\left(\int_{\Omega(t)} w_i u^h \, d\boldsymbol{x}\right) =$$

$$= \int_{\Omega^{n+1/2}} w_i \left(u_h^{n+1} - u_h^n\right) d\boldsymbol{x} + \Delta t \int_{\Omega^{n+1/2}} w_i \frac{\left(u_h^{n+1} + u_h^n\right)}{2} \nabla \cdot \boldsymbol{\sigma}_h^* \, d\boldsymbol{x} \tag{4.42}$$

Second, we substitute (4.42) in (4.41) and we sum the last term of the above equation with the last one in (4.41)

$$\left(1 + \frac{\Delta t}{2}\nabla \cdot \boldsymbol{\sigma}_h^*\right) \int_{\Omega^{n+1/2}} w_i \left(u_h^{n+1} - u_h^n\right) d\boldsymbol{x} +$$

$$+\Delta t \int_{\Omega^{n+1/2}} w_i \left(\nabla \cdot \boldsymbol{f}(u_h^n) - \boldsymbol{\sigma}_h^* \cdot \nabla u_h^n\right) d\boldsymbol{x} = 0 \tag{4.43}$$

Third we invoke the analogy with Residual Distribution method, we do mass-lumping

$$\sum_{K\in\mathcal{D}_i} \left(1 + \frac{\Delta t}{2}\nabla \cdot \boldsymbol{\sigma}_h^*\right) \frac{|K^{n+1/2}|}{3} \left(u_i^{n+1} - u_i^n\right) = -\Delta t \sum_{K\in\mathcal{D}_i} \beta_i \sum_j k_j^n u_j^n$$

And the final algorithm reads

$$|\tilde{S}_i^{n+1/2}| \left(u_i^{n+1} - u_i^n\right) = -\Delta t \sum_{K \in \mathcal{D}_i} \beta_i \sum_j k_j^n u_j^n \tag{4.44}$$

where the median dual cell area of (2.63), here evaluated at midpoint configuartion, has to be modified to take into account the grid distortion

$$|\tilde{S}_i^{n+1/2}| = \sum_{K \in \mathcal{D}_i} \left(1 + \frac{\Delta t}{2} \nabla \cdot \boldsymbol{\sigma}_h^*\right) \frac{|K^{n+1/2}|}{3} \tag{4.45}$$

The method satisfy the DGCL by construction but it is extremely easy to prove this again, by simply imposing a uniform state in the method presented so far.

Apart from the ALE flux part in the upwind parameter k_j, the formulation follows very closely the prototype scheme in Eulerian formulation (2.67), an extension of the results regarding positivity should be straightforward. A modified median dual cell area appear to take into account mesh distortion $\left(1 + \frac{\Delta t}{2} \nabla \cdot \boldsymbol{\sigma}_h^*\right)$. Strictly speaking this scalar quantity can be also negative (in a critical situation of very fast compression for the mesh) spoiling every positive coefficient anlysis. If the grid displacements within the timestep are of order h then this term is of order $\mathcal{O}(h^2)$ and does not affect the result provided in section (1.4.1). In all the computation that we did, even the ones involving big distortion of the grid, the positivity of $|\tilde{S}_i^{n+1/2}|$ was mantained.

Unfortunately this scheme converge only with first order of accuracy. To achieve second order accuracy and stay explicit we can use the explicit Runge-Kutta presented in section (3.1)

4.4.2 Runge Kutta two

This method has been discussed in the Eulerian version in section (2.1). In order to end up with the correspondent ALE scheme that verify a DGCL condition we use the idea developed in the previous paragraph, even if a direct extension, will be a little bit tricky. The problem that we have to face is that the shifting cause, in the first step, the disappereance of the time part associated to the stabilization bubble and this breaks the terms balance for the satisfaction of the DGCL. A simple way to fix this inconvenient is to choose for the stabilization term equation (4.37). The Galerkin part, discretized with a $\mathcal{RK}2$, with $\boldsymbol{\sigma}^h = \boldsymbol{\sigma}_h^*$ and midpoint configuartion, writes

$$\frac{\Delta}{\Delta t} \int_{\Omega(t)} \varphi_i u_h^k \, d\boldsymbol{x} + \int_{\Omega^{n+1/2}} \varphi_i \left(\nabla \cdot \boldsymbol{f}(u_h) - \boldsymbol{\sigma}_h^* \cdot \nabla u_h\right)^k \, d\boldsymbol{x} +$$

$$- \int_{\Omega^{n+1/2}} \varphi_i \left(u_h \nabla \cdot \boldsymbol{\sigma}_h^*\right)^k \, d\boldsymbol{x} = 0 \tag{4.46}$$

While for the stabilization one, also computed at midpoint configuration, we have

$$\sum_{K\in\mathcal{D}_i}\int_{K^{n+1/2}}\gamma_i\frac{\Delta\overline{u_h^k}}{\Delta t}\,d\boldsymbol{x}+\sum_{K\in\mathcal{D}_i}\int_{K^{n+1/2}}\gamma_i\left(\nabla\cdot\boldsymbol{f}(u_h^n)-\boldsymbol{\sigma}_h^*\cdot\nabla u_h^n\right)^k\,d\boldsymbol{x}=0 \tag{4.47}$$

Now, in analogy with (3.3), we give the following definition

$$\Phi_i^{RK(k)}=\int_{K^{n+1/2}}w_i\left(\frac{\Delta\overline{u_h^k}}{\Delta t}+\left(\nabla\cdot\boldsymbol{f}(u^h)-\boldsymbol{\sigma}_h^*\nabla u^h\right)^k\right)d\boldsymbol{x} \tag{4.48}$$

or, as in (3.5)

$$\Phi_i^{RK(k)}=\sum_{j\in K}m_{ij}^K\frac{\Delta\overline{u_j^k}}{\Delta t}+\phi_i^{RK(k)} \tag{4.49}$$

In particular exploiting the two steps

$$\begin{cases}\Phi_i^{RK(1)}= & \phi_i(u^h)\\ \Phi_i^{RK(2)}=\sum_{j\in K}m_{ij}^K\frac{u_j^1-u_j^n}{\Delta t} & +\frac{1}{2}\left(\phi_i(u_h^1)+\phi_i(u_h^n)\right)\end{cases} \tag{4.50}$$

Selective Lumped

$$|\tilde{S}_i^{n+1/2}|\frac{\Delta u_i^k}{\Delta t}=-\sum_K\left(\Phi_i^{RK(k)}-\sum_j\tilde{m}_{ij}^{GAL}\frac{\Delta\overline{u_j^k}}{\Delta t}\right) \tag{4.51}$$

Apart from the ALE part in the definition of the upwind parameter, the only differences respect to Eulerian version (3.11) is the presence of the modified median dual cell area (4.45) and of a modified Galerkin mass matrix

$$\tilde{m}_{ij}^{GAL}=\left(1+\frac{\Delta t}{2}\nabla\cdot\boldsymbol{\sigma}_h^*\right)m_{ij}^{GAL} \tag{4.52}$$

Proof. For the first step ($k=1$) assembling (4.46)(4.47) and at the same time using (4.42)

$$\int_{\Omega^{n+1/2}}\varphi_i\frac{u_h^1-u_h^n}{\Delta t}\,d\boldsymbol{x}+\frac{\Delta t}{2}\int_{\Omega^{n+1/2}}\varphi_i\nabla\cdot\boldsymbol{\sigma}_h^*\left(u_h^1-u_h^n\right)\,d\boldsymbol{x}+$$

$$+\int_{\Omega^{n+1/2}}w_i\left(\nabla\cdot f(u_h^n)-\boldsymbol{\sigma}_h^*\cdot\nabla u_h^n\right)\,d\boldsymbol{x}=0$$

In a $\mathcal{RD}$ formalism

$$\sum_{K\in\mathcal{D}_i}\left(1+\nabla\cdot\boldsymbol{\sigma}_h^*\right)\frac{|K^{n+1/2}|}{3}\frac{u_h^1-u_h^n}{\Delta t}=-\sum_{K\in\mathcal{D}_i}\beta_i\sum_j k_j^n u_j^n=-\Delta t\sum_{K\in\mathcal{D}_i}\phi_i(u^h)$$

Using the definition (4.45), then the first line of (4.51) is prooved.

For the second step ($k = 2$) the algebra is a little longer. We put in evidence the clear fact

$$\sum_{K\in\mathcal{D}_i}\int_{K^{n+1/2}}\gamma_i\frac{\Delta\overline{u_h^k}}{\Delta t}\,d\boldsymbol{x}=\sum_{K\in\mathcal{D}_i}\int_{K^{n+1/2}}w_i\frac{\Delta\overline{u_h^k}}{\Delta t}\,d\boldsymbol{x}-\int_{\Omega^{n+1/2}}\varphi_i\frac{\Delta\overline{u_h^k}}{\Delta t}\,d\boldsymbol{x} \tag{4.53}$$

Again assembling (4.46)(4.47) and at the same time using (4.42)

$$\int_{\Omega^{n+1/2}}\varphi_i\frac{u_h^{n+1}-u_h^n}{\Delta t}\,d\boldsymbol{x}+\frac{\Delta t}{2}\int_{\Omega^{n+1/2}}\varphi_i\nabla\cdot\underline{\sigma}_h^*\frac{u_h^{n+1}+u_h^n}{\Delta t}\,d\boldsymbol{x}+$$

$$-\frac{1}{2}\int_{\Omega^{n+1/2}}\varphi_i\nabla\cdot\boldsymbol{\sigma}_h^*\left(u_h^1+u_h^n\right)\,d\boldsymbol{x}-\int_{\Omega^{n+1/2}}\varphi_i\frac{u_h^1-u_h^n}{\Delta t}\,d\boldsymbol{x}+$$

$$+\sum_{K\in\mathcal{D}_i}\int_{K^{n+1/2}}w_i\frac{u_h^1-u_h^n}{\Delta t}\,d\boldsymbol{x}+\frac{1}{2}\sum_{K\in\mathcal{D}_i}\int_{K^{n+1/2}}w_i\left(\nabla\cdot f(u_h^n)-\boldsymbol{\sigma}_h^*\cdot\nabla u_h^n\right)\,d\boldsymbol{x}+$$

$$+\frac{1}{2}\sum_{K\in\mathcal{D}_i}\int_{K^{n+1/2}}w_i\left(\nabla\cdot f(u_h^1)-\boldsymbol{\sigma}_h^*\cdot\nabla u_h^1\right)\,d\boldsymbol{x}=0$$

Now we sum and subtract the quantity $\frac{\Delta t}{2}\int_{\Omega^{n+1/2}}\varphi_i\nabla\cdot\boldsymbol{\sigma}_h^*\frac{u_h^{n+1}-u_h^n}{\Delta t}\,d\boldsymbol{x}$. The term with plus sum with the first term of the above equation, the term with minus sum with the second, the third and the fourth ones

$$\int_{\Omega^{n+1/2}}\left(1+\frac{\Delta t}{2}\nabla\cdot\underline{\sigma}_h^*\right)\varphi_i\frac{u_h^{n+1}-u_h^n}{\Delta t}\,d\boldsymbol{x}-\int_{\Omega^{n+1/2}}\left(1+\frac{\Delta t}{2}\nabla\cdot\boldsymbol{\sigma}_h^*\right)\varphi_i\frac{u_h^1-u_h^n}{\Delta t}\,d\boldsymbol{x}$$

$$+\sum_{K\in\mathcal{D}_i}\int_{K^{n+1/2}}w_i\frac{u_h^1-u_h^n}{\Delta t}\,d\boldsymbol{x}+\frac{1}{2}\sum_{K\in\mathcal{D}_i}\int_{K^{n+1/2}}w_i\left(\nabla\cdot\boldsymbol{f}(u_h^n)-\boldsymbol{\sigma}_h^*\cdot\nabla u_h^n\right)\,d\boldsymbol{x}+$$

$$+\frac{1}{2}\sum_{K\in\mathcal{D}_i}\int_{K^{n+1/2}}w_i\left(\nabla\cdot\boldsymbol{f}(u_h^1)-\boldsymbol{\sigma}_h^*\cdot\nabla u_h^1\right)\,d\boldsymbol{x}=0$$

The last three terms can be rewritten compactly with (4.48),

$$\int_{\Omega^{n+1/2}}\left(1+\frac{\Delta t}{2}\nabla\cdot\boldsymbol{\sigma}_h^*\right)\varphi_i\frac{u_h^{n+1}-u_h^n}{\Delta t}\,d\boldsymbol{x}=$$

$$=-\sum_{K\in\mathcal{D}_i}\Phi_i^{RK(2)}+\int_{\Omega^{n+1/2}}\left(1+\frac{\Delta t}{2}\nabla\cdot\boldsymbol{\sigma}_h^*\right)\varphi_i\frac{u_h^1-u_h^n}{\Delta t}\,d\boldsymbol{x}$$

Developing both the modified mass matrices but lumping only the one on the right-hand side, then using definitions (4.45) and (4.52), (4.51) is finally prooved

Global Lumped

$$\begin{cases} |\tilde{S}_i^{\,n+1/2}| \frac{u_i^1 - u_i^n}{\Delta t} = -\sum_K \Phi_i^{RK(1)} \\ |\tilde{S}_i^{\,n+1/2}| \frac{u_i^{n+1} - u_i^1}{\Delta t} = -\sum_K \Phi_i^{RK(2)} \end{cases} \tag{4.54}$$

The scheme closely resemble the Eulerian one (3.12). The presence of the modified median dual cell area does not destroy the positivity.

Proof. The first step remain the same and has been already prooved

For the second step ($k = 2$) assembling (4.46)(4.47), togheter with (4.42)(4.53)

$$\int_{\Omega^{n+1/2}} \varphi_i \frac{u_h^{n+1} - u_h^n}{\Delta t}\, d\boldsymbol{x} + \frac{\Delta t}{2} \int_{\Omega^{n+1/2}} \varphi_i \nabla \cdot \boldsymbol{\sigma}_h^* \frac{u_h^{n+1} + u_h^n}{\Delta t}\, d\boldsymbol{x} +$$

$$-\frac{1}{2} \int_{\Omega^{n+1/2}} \varphi_i \nabla \cdot \boldsymbol{\sigma}_h^* \left(u_h^1 + u_h^n\right)\, d\boldsymbol{x} - \int_{\Omega^{n+1/2}} \varphi_i \frac{u_h^1 - u_h^n}{\Delta t}\, d\underline{x} +$$

$$+ \sum_{K \in \mathcal{D}_i} \int_{K^{n+1/2}} w_i \frac{u_h^1 - u_h^n}{\Delta t}\, d\boldsymbol{x} + \frac{1}{2} \sum_{K \in \mathcal{D}_i} \int_{K^{n+1/2}} w_i \left(\nabla \cdot \boldsymbol{f}(u_h^n) - \boldsymbol{\sigma}_h^* \cdot \nabla u_h^n\right)\, d\boldsymbol{x} +$$

$$\frac{1}{2} \sum_{K \in \mathcal{D}_i} \int_{K^{n+1/2}} w_i \left(\nabla \cdot \boldsymbol{f}(u_h^1) - \boldsymbol{\sigma}_h^* \cdot \nabla u_h^1\right)\, d\boldsymbol{x} = 0$$

Summing the first and the fourth term togheter and the second and third too

$$\int_{\Omega^{n+1/2}} \varphi_i \frac{u_h^{n+1} - u_h^1}{\Delta t}\, d\boldsymbol{x} + \frac{\Delta t}{2} \int_{\Omega^{n+1/2}} \varphi_i \nabla \cdot \boldsymbol{\sigma}_h^* \frac{u_h^{n+1} - u_h^1}{\Delta t}\, d\boldsymbol{x} +$$

$$+ \sum_{K \in \mathcal{D}_i} \int_{K^{n+1/2}} w_i \frac{u_h^1 - u_h^n}{\Delta t}\, d\boldsymbol{x} + \frac{1}{2} \sum_{K \in \mathcal{D}_i} \int_{K^{n+1/2}} w_i \left(\nabla \cdot \boldsymbol{f}(u_h^n) - \boldsymbol{\sigma}_h^* \cdot \nabla u_h^n\right)\, d\boldsymbol{x} +$$

$$+\frac{1}{2} \sum_{K \in \mathcal{D}_i} \int_{K^{n+1/2}} w_i \left(\nabla \cdot \boldsymbol{f}(u_h^1) - \boldsymbol{\sigma}_h^* \cdot \nabla u_h^1\right)\, d\boldsymbol{x} = 0$$

The last three terms can be rewritten compactly with (4.48), while the first two terms sum up

$$\int_{\Omega^{n+1/2}} \left(1 + \frac{\Delta t}{2} \nabla \cdot \boldsymbol{\sigma}_h^*\right) \varphi_i \frac{u_h^{n+1} - u_h^1}{\Delta t}\, d\boldsymbol{x} = -\sum_{K \in \mathcal{D}_i} \Phi_i^{RK(2)}$$

Developing the mass matrix, lumping it and using (4.45) we get the second line of (4.54).

4.5 Numerical results

The scalar experiments we show here are used to test the formulation presented so far. The test cases are exactely the ones performed by [9] with a the explicit $\mathcal{RK}2$ of chapter 2: we hope to recover the same results, in terms of accuracy and non-oscillatory behaviour, when the grid is moving with an arbitrary motion and the RK2-ALE scheme presented in the previous chapter is used.

All the schemes, modified in the proper way for $\mathcal{RK}2$ time integrator in section (3.1.1), are used here. We have just to remember that, for ALE computaions, the upwind parameter takes into account the grid movement and follows the definition (4.27) which is recalled below

$$k_i = \frac{1}{2}\left(\bar{\boldsymbol{a}} - \bar{\boldsymbol{\sigma}}\right) \cdot \boldsymbol{n}_i$$

All the definition which involves k_i has been revisted.

For all the experiments the time step is computed in order to verify the CFL condition

$$\Delta t = \mathrm{CFL} \min_{i \in \mathcal{T}_h} \frac{|S_i|}{\sum_{K \in \mathcal{D}_i} 3\alpha^K} \tag{4.55}$$

where CFL = 0.8. Inequalities (2.77) and (2.87) are verified hence the LxF and N schemes are positive.

4.5.1 Convergence properties

To test the accuracy of the method we use the simple case of linear advection of a smooth sinusoidal hill

$$\begin{cases} \frac{\partial u}{\partial t} + \boldsymbol{a} \cdot \nabla u = 0, & \boldsymbol{a} = [0,1]\,,\ \boldsymbol{x} \in [0,1] \times [0,2],\ t \in [0,1] \\ u_0 = cos\left(2\pi r\right) & \text{if } r \leq 0.25,\ r = \sqrt{(x-0.5)^2 + (y-0.5)^2} \\ u_0 = 0 & \text{otherwise} \end{cases}$$

We choose 4 unstructured grid with characteristic lenghts shown in table. Given a reference domain (X, Y), it is mapped according to the following law

$$\begin{cases} x(t) &= X + 0.1 \sin\left(2\pi X\right) \sin\left(\pi Y\right) \sin\left(2\pi t\right) \\ y(t) &= Y + 0.2 \sin\left(2\pi X\right) \sin\left(\pi Y\right) \sin\left(4\pi t\right) \end{cases} \tag{4.56}$$

At $t = 1$, the mapping is the identity $\boldsymbol{x} = \boldsymbol{X}$, so we can compare the ALE solution with the Eulerian one easily. In figure (4.1) grid number 3 is shown with the corresondent mapping.

All the results collected in figure (4.2) shows that second order of accuracy, when expected, is achieved. The ALE convergence curve almost collapse on the classical one. The Blended LDA-N on smooth solution should collapse to the LDA scheme but it converges more slowly, only with order 1.5 instead. This is due to the fact that the advecting hill is very narrow and the presence of strong gradients cause the switch to a first order N scheme.

grid	h
2	1/30
3	1/50
4	1/80
5	1/160

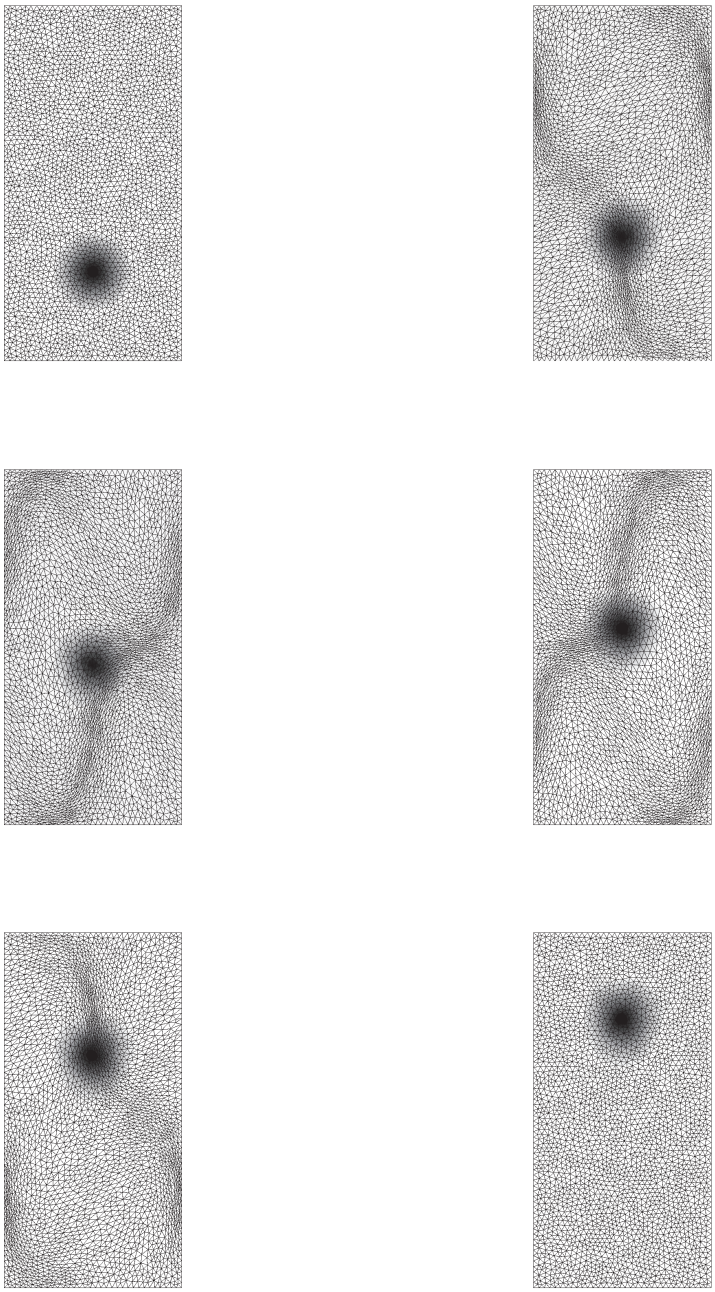

Figure 4.1: Linear Advection. Mapping for the grid and example of the numerical solution

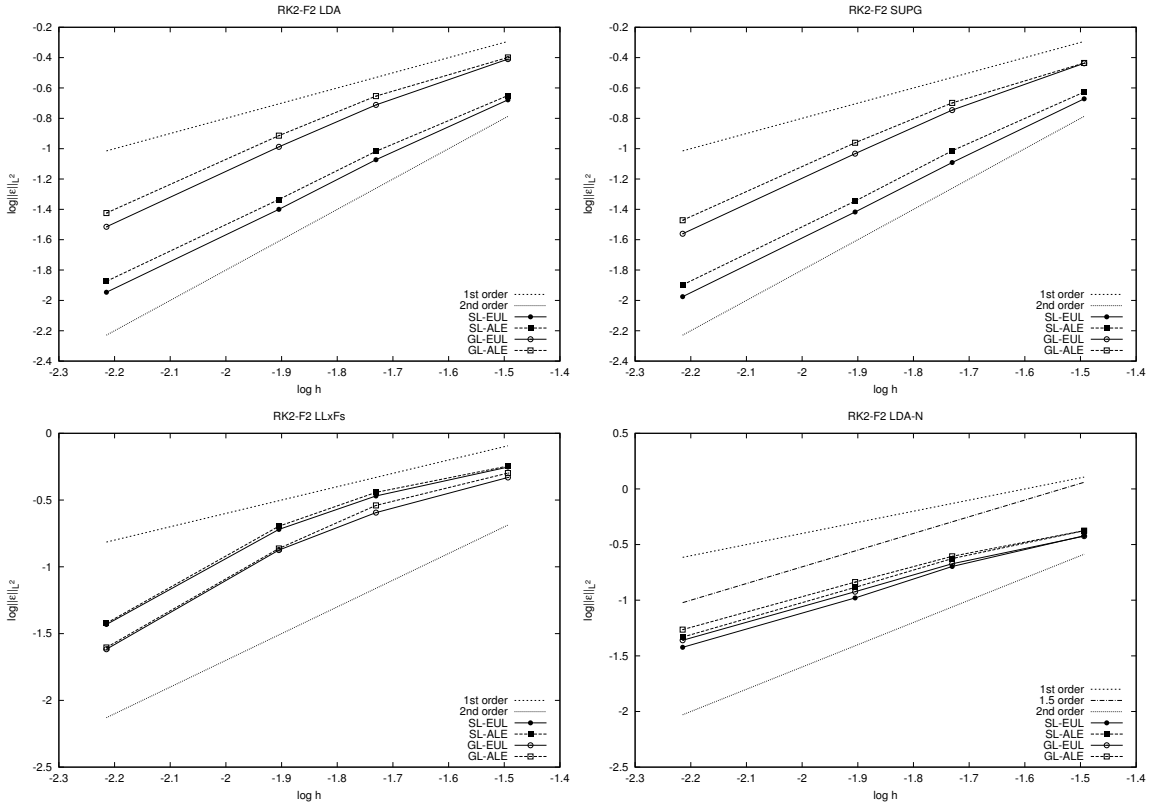

Figure 4.2: Linear Advection. Order of Convergence

4.5.2 2D Burger equation

The non-linear Burger equation is a good test to see how the schemes behaves near discontinuities

$$\begin{cases} \frac{\partial u}{\partial t} + \boldsymbol{a} \cdot \nabla u = 0, & \boldsymbol{a} = [u, u]\,,\ x \in [-1,1] \times [-1,1],\ t \in [0,1] \\ u_0 = 1 & \text{if } x \in [-0.6, -0.1] \times [-0.35, 0.15] \\ u_0 = 0 & \text{otherwise} \end{cases}$$

The reference grid size is $h = 1/80$. The domain is mapped in a similar way to (4.56), according to

$$\begin{cases} x(t) &= X + 0.2 \sin(\pi X) \sin(\pi Y) \sin(2\pi t) \\ y(t) &= Y + 0.2 \sin(\pi X) \sin(\pi Y) \sin(4\pi t) \end{cases}$$

Only results with formulation F1 are shown but the use of formulation F2 lead to very similar results.

First we consider the linearity preserving LDA and SUPG scheme in figures (4.3), (4.4), (4.5), (4.6). As expected, these two schemes that gives very good results when computing smooth solutions, fail when computing discontinuities. Oscillations appears on the shock and at the tail of the rarefaction wave. The important observation is that the ALE results, far from the discontinuity, are very close to the Eulerian ones, on the tail of the rarefaction even better.

The non-linear schemes LLxFs and LDA-N are designed to capture very well discontinuities. This is shown if (4.7), (4.8), (4.9), (4.10). We have seen that SL formulation, unfortunately, does not allow us to mantein positivity hence oscillation still appear but are less pronunced respect to linear schemes. With GL formulation this problem is cured provided that some dissipation is introduced due to the mass-lumping [9]. The ALE results trace very good the Eulerian one and they are even better smeared out when computing the tail of the rarefaction wave.

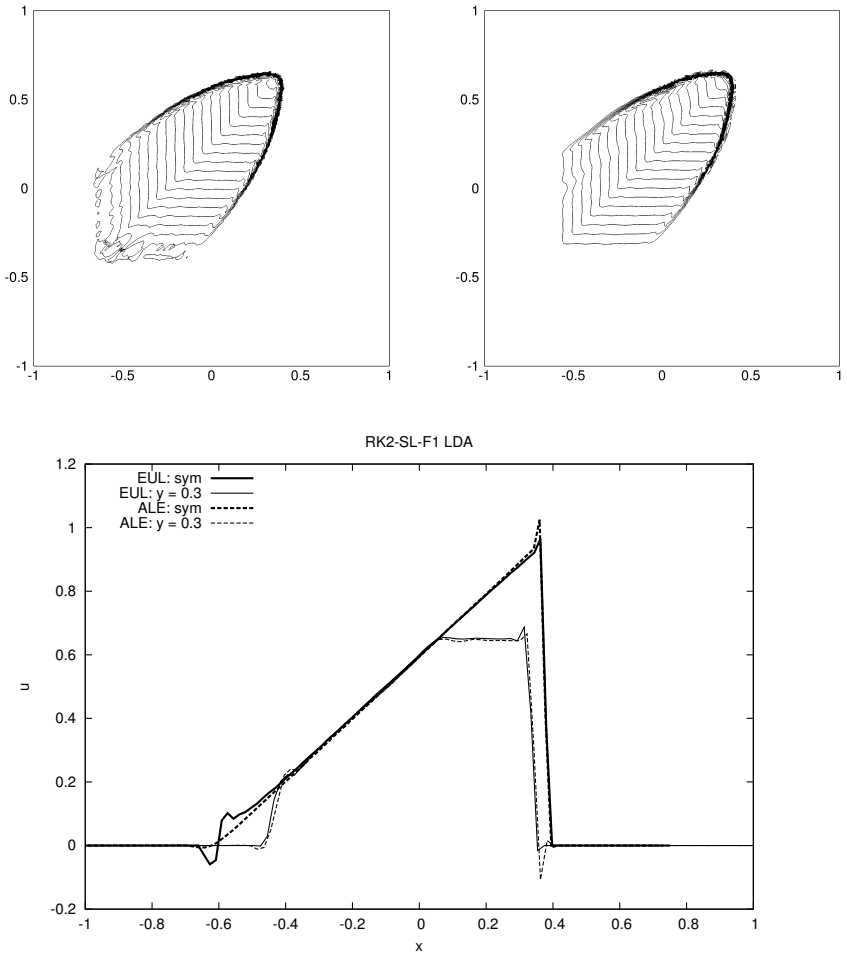

Figure 4.3: 2D Burger Equation, F1-SL LDA scheme. Top: 20 equispaced isolines between 0 and 1 at time $t = 1$. left, Eulerian. right ALE. Bottom: comparison of the solution along the symmetry line and along the line $y = 0.3$

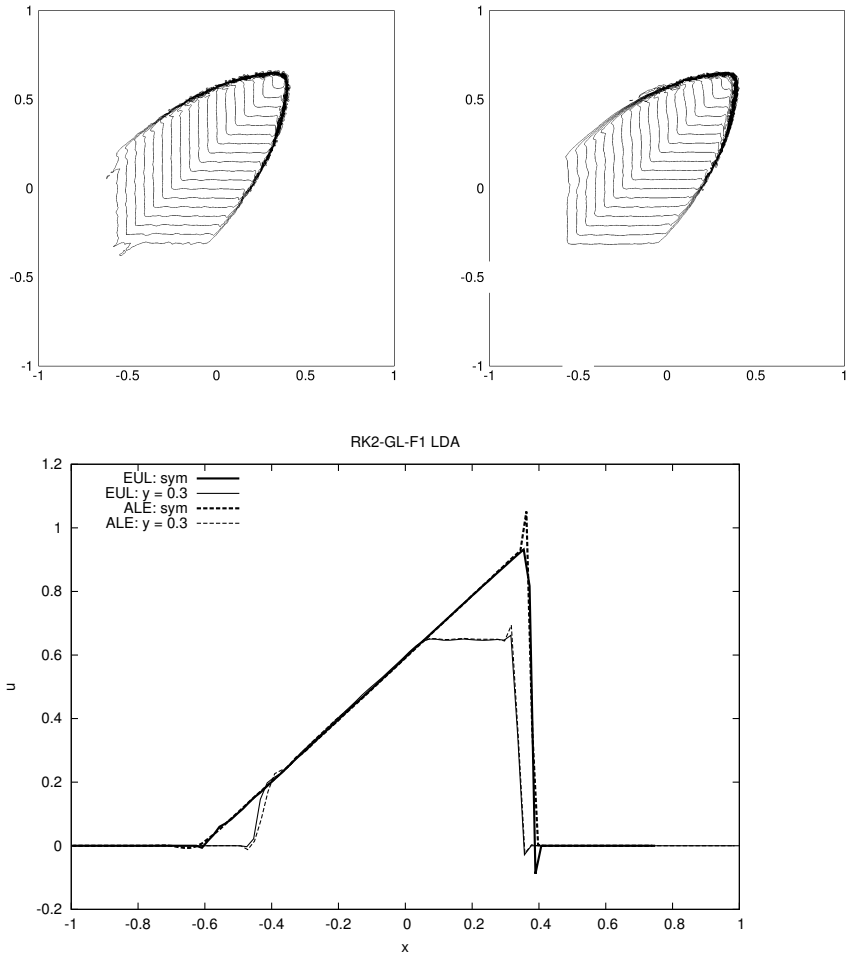

Figure 4.4: 2D Burger Equation, F1-GL LDA scheme. Top: 20 equispaced isolines between 0 and 1 at time $t = 1$. left, Eulerian. right ALE. Bottom: comparison of the solution along the symmetry line and along the line $y = 0.3$

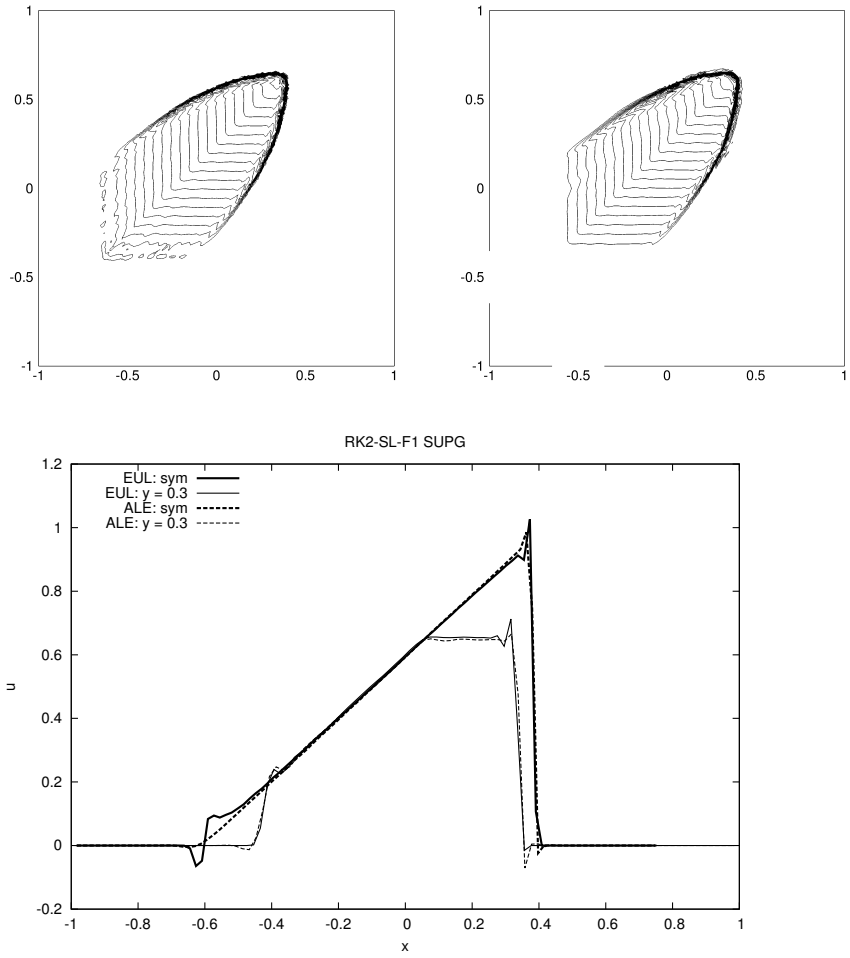

Figure 4.5: 2D Burger Equation, F1-SL SUPG scheme. Top: 20 equispaced isolines between 0 and 1 at time $t = 1$. left, Eulerian. right ALE. Bottom: comparison of the solution along the symmetry line and along the line $y = 0.3$

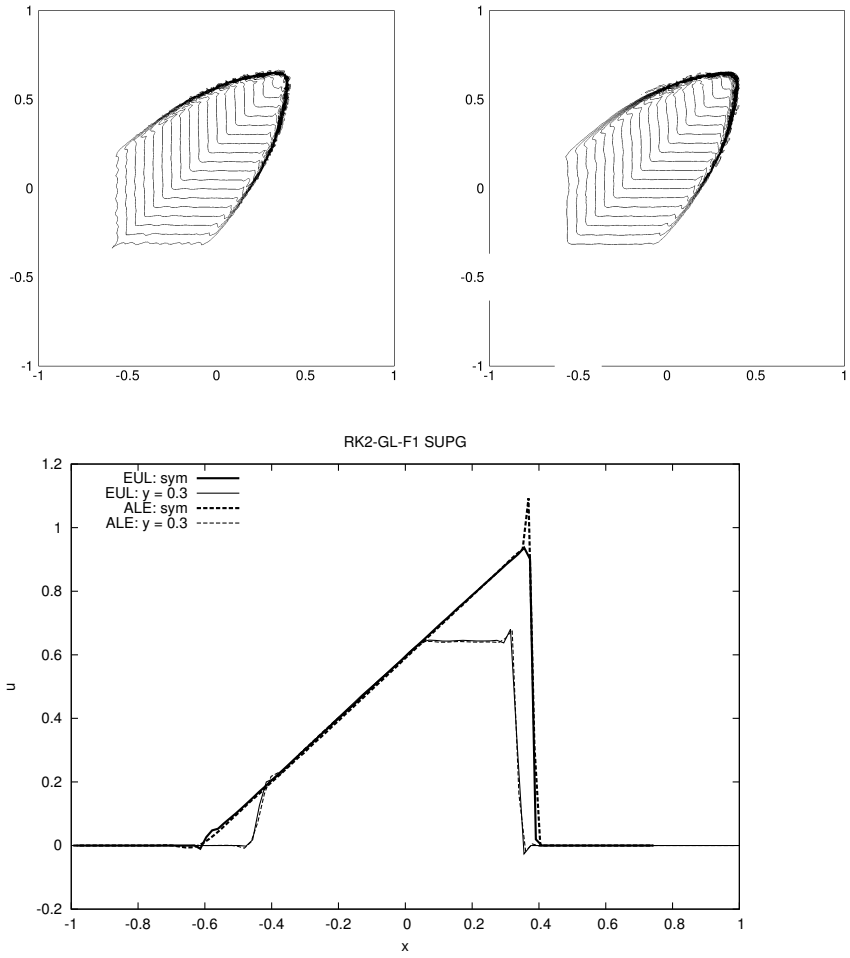

Figure 4.6: 2D Burger Equation, F1-GL SUPG scheme. Top: 20 equispaced isolines between 0 and 1 at time $t = 1$. left, Eulerian. right ALE. Bottom: comparison of the solution along the symmetry line and along the line $y = 0.3$

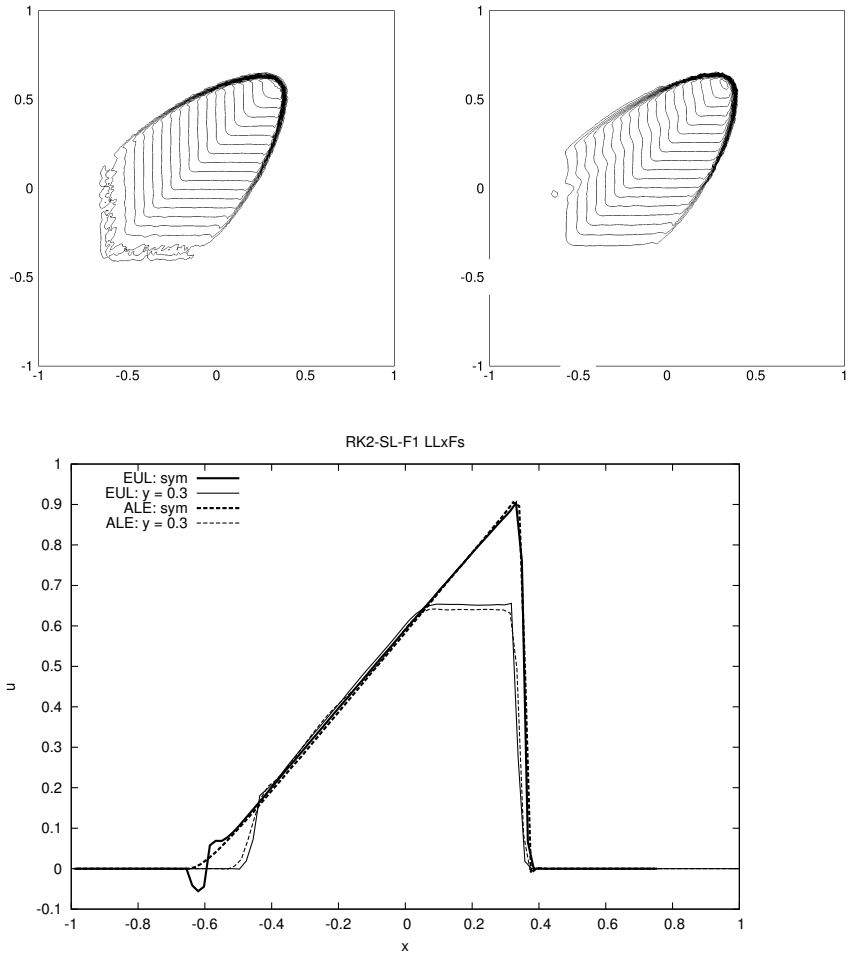

Figure 4.7: 2D Burger Equation, F1-SL LLxFs scheme. Top: 20 equispaced isolines between 0 and 1 at time $t = 1$. left, Eulerian. right ALE. Bottom: comparison of the solution along the symmetry line and along the line $y = 0.3$

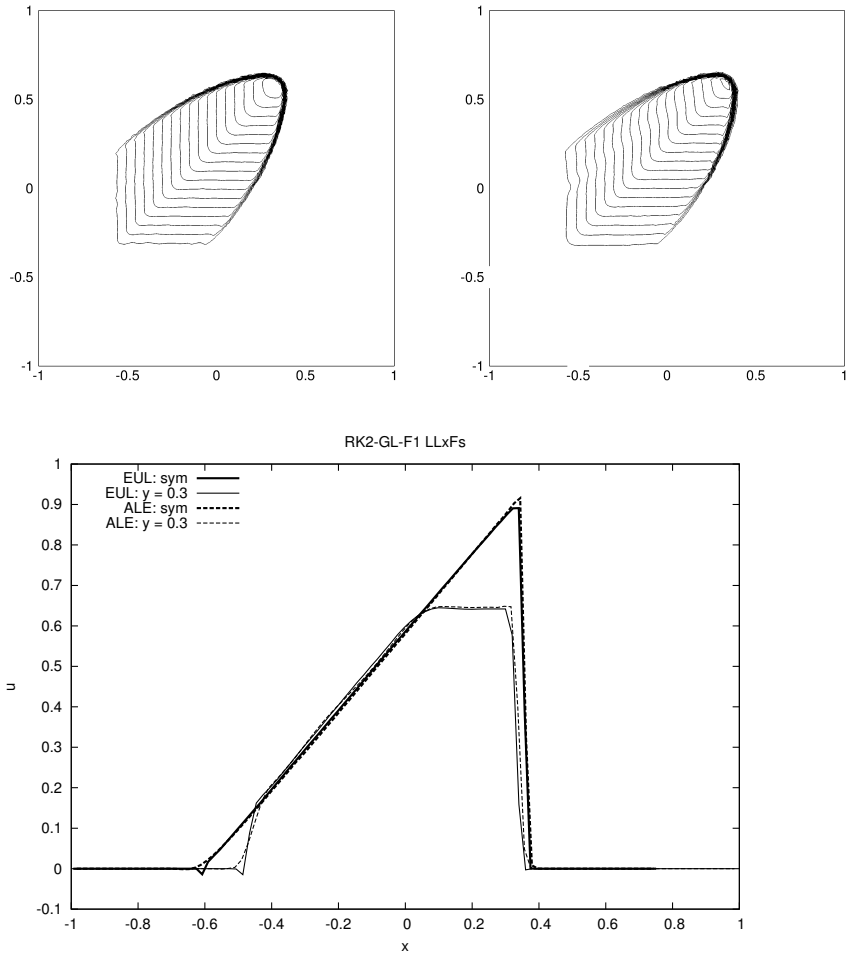

Figure 4.8: 2D Burger Equation, F1-GL LLxFs scheme. Top: 20 equispaced isolines between 0 and 1 at time $t = 1$. left, Eulerian. right ALE. Bottom: comparison of the solution along the symmetry line and along the line $y = 0.3$

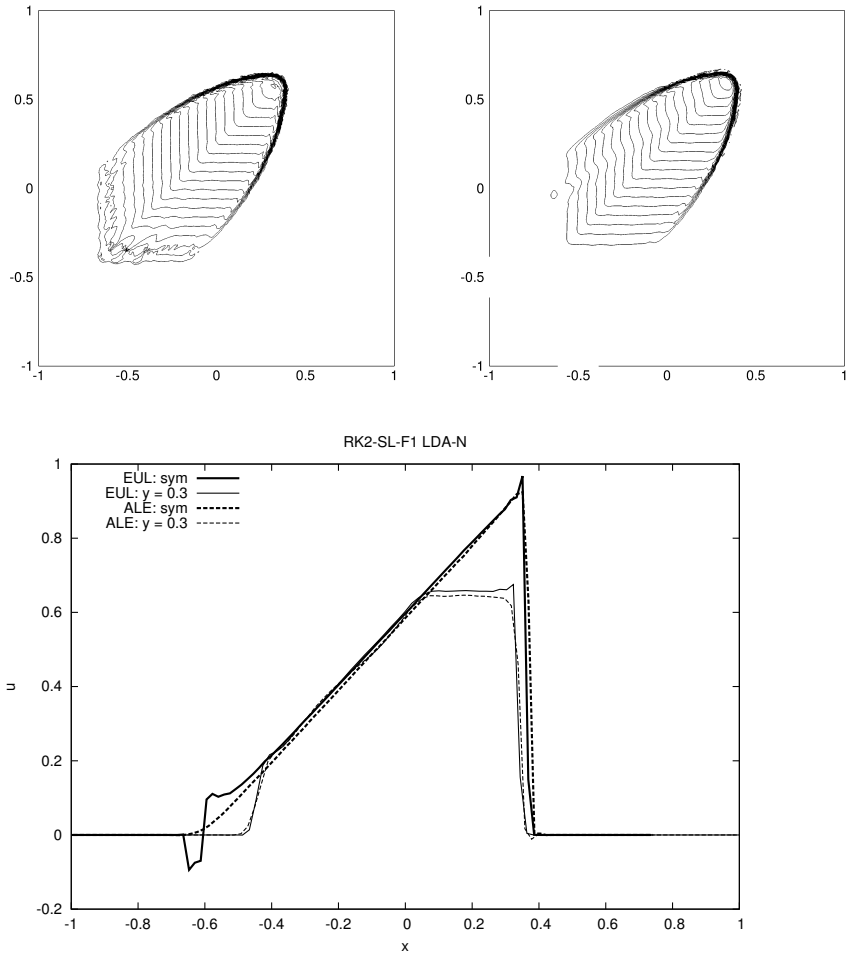

Figure 4.9: 2D Burger Equation, F1-SL LDA-N scheme. Top: 20 equispaced isolines between 0 and 1 at time $t = 1$. left, Eulerian. right ALE. Bottom: comparison of the solution along the symmetry line and along the line $y = 0.3$

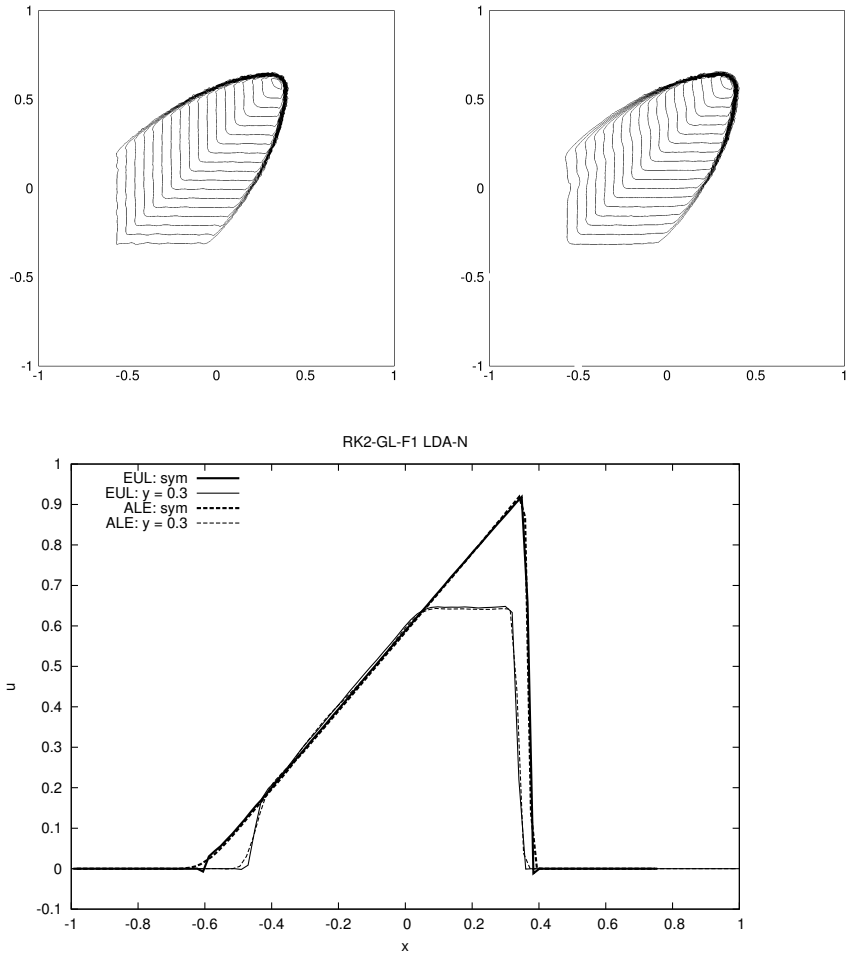

Figure 4.10: 2D Burger Equation, F1-GL LDA-N scheme. Top: 20 equispaced isolines between 0 and 1 at time $t = 1$. left, Eulerian. right ALE. Bottom: comparison of the solution along the symmetry line and along the line $y = 0.3$

5 Residual Distribution for Euler Equations with moving grids

5.1 Basic concept for 1D systems

Once Residual Distribution schemes for scalar conservation laws have been understood, their extension to systems is quite simple. A nice exercise to see this, can be the reformalization of a Finite Volume first order upwind approximation of a 1D-linear system into a Residual Distribution method. Given the hyperbolic linear system

$$\frac{\partial \mathsf{u}}{\partial t} + \mathsf{A}\frac{\partial \mathsf{u}}{\partial x} = 0 \tag{5.1}$$

with $\mathsf{u} \in \mathbb{R}^m$ vector of conserved quantities, we recall that the system is hyperbolic if $\mathsf{A} \in \mathbb{R}^{m\times m}$ is diagonalizable with real eigenvalues. We denote the eigenvalues by

$$\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_m$$

The matrix A is diagonalizable if a complete set of linearly independent eigenvectors exists

$$\mathsf{r}_1, \mathsf{r}_2, ..., \mathsf{r}_m$$

In this case the matrix having for coloumns the eigenvectors

$$\mathsf{R} = [\, \mathsf{r}_1 \, \mathsf{r}_2 \, ... \, \mathsf{r}_m \,]$$

is non-singular and has an inverse so we can put the matrix A in a diagonal form by the following tranformation

$$\Lambda = \mathsf{R}^{-1}\mathsf{A}\mathsf{R}, \quad \text{where} \quad \Lambda = \mathrm{Diag}\,\{\lambda_1, \lambda_2, ..., \lambda_m\} \tag{5.2}$$

If we define the vectors of characteristic variables $\mathsf{w} = \mathsf{R}^{-1}\mathsf{u}$, we can rewrite the linear system (5.1) as a set of m decoupled linear advection equations

$$\frac{\partial \mathsf{w}}{\partial t} + \Lambda\frac{\partial \mathsf{w}}{\partial x} = 0 \tag{5.3}$$

whose solution, in turn, can be used to solve the original system of equation.

The Godunov approximation of (5.1) writes

$$\mathsf{u}_i^{n+1} = \mathsf{u}_i^n - \frac{\Delta t}{h_i}\left(\mathsf{F}_{i+1/2}(\mathsf{u}_i^n, \mathsf{u}_{i+1}^n) - \mathsf{F}_{i-1/2}(\mathsf{u}_{i-1}^n, \mathsf{u}_i^n)\right) \tag{5.4}$$

where the numerical flux function, indicated with little abuse in the notation as F, is computed according to

$$\mathsf{F}_{i-1/2} = \mathsf{A}(\mathrm{u}^\curlyvee(\mathrm{u}_{i-1}^n, \mathrm{u}_i^n)) = \frac{1}{2}\mathsf{A}(\mathsf{u}_i^n + \mathsf{u}_{i-1}^n) - \frac{1}{2}|\mathsf{A}|(\mathsf{u}_i^n - \mathsf{u}_{i-1}^n) \tag{5.5}$$

As already seen for the scalar case, the bridge between Finite Volume and Residual Distribution consists in rewriting (5.5) in terms of fluctuations. Using the following numerical flux

$$\mathsf{F}_{i-1/2} = \mathsf{A}\mathsf{u}_i^n - \mathsf{A}^+(\mathsf{u}_i^n - \mathsf{u}_{i-1}^n)$$

in analogy with the scalar reformulation (2.7), the $\mathcal{FV}$ approximation can be rewritten as a Flux-Difference Splitting approximation

$$\mathsf{u}_i^{n+1} = \mathsf{u}_i^n - \frac{\Delta t}{h_i}\left(\mathsf{A}^-(\mathsf{u}_{i+1}^n - \mathsf{u}_i^n) + \mathsf{A}^+(\mathsf{u}_i^n - \mathsf{u}_{i-1}^n)\right) \tag{5.6}$$

where

$$\mathsf{A}^+ = \mathsf{R}\Lambda^+\mathsf{R}^{-1} \quad \text{and} \quad \mathsf{A}^- = \mathsf{R}\Lambda^-\mathsf{R}^{-1}$$

$$\Lambda^+ = \begin{bmatrix} (\lambda_1)^+ & & & \\ & (\lambda_2)^+ & & \\ & & \ddots & \\ & & & (\lambda_m)^+ \end{bmatrix}, \quad \Lambda^- = \begin{bmatrix} (\lambda_1)^- & & & \\ & (\lambda_2)^- & & \\ & & \ddots & \\ & & & (\lambda_m)^- \end{bmatrix}$$

The method as a nice interpratation, infact the flux difference across the boundaries of the cell $[x_{i-1}, x_i]$ can be splitted in the following way

$$\mathsf{F}_i - \mathsf{F}_{i-1} = \mathsf{A}^-(\mathsf{u}_i^n - \mathsf{u}_{i-1}^n) + \mathsf{A}^+(\mathsf{u}_i^n - \mathsf{u}_{i-1}^n) \tag{5.7}$$

Two contributions called *fluctuations* appears, one coming from the net effect of all left-going waves and one coming from the net effect of all right-going waves. The Flux-Difference splitting algorithm (5.6) follows immediately after the application of (5.7) for every cell. The value of the solution at x_i is updated summing up a right-going fluctuation $\mathsf{A}^+(\mathsf{u}_i^n - \mathsf{u}_{i-1}^n)$ entering from the interface $x_{i-1/2}$ and a left-going fluctuation $\mathsf{A}^-(\mathsf{u}_{i+1}^n - \mathsf{u}_i^n)$ entering from the opposite interface $x_{i+1/2}$.

We have already seen in the scalar case that the Residual Distribution method is a Flux-Difference Splitting method with an arbitrary splitting of the flux difference, which in turn is called, in a $\mathcal{RD}$ contest, *residual*. Through the usual steps, a true $\mathcal{RD}$ scheme is constructed:

1. Using a piecewise linear approximation of the solution, compute the residual over every cell. For the cell $[x_{i-1}, x_i]$
$$\Phi^{i-1/2} = \int_{x_{i-1}}^{x_i} \mathsf{A}\frac{\partial \mathsf{u}}{\partial x}\, dx = \mathsf{A}(\mathsf{u}_i^n - \mathsf{u}_{i-1}^n)$$
with residual $\Phi^{i-1/2} \in \mathbb{R}^m$.
2. Split the residual between the two nodes of the element through the distribution matrix $\boldsymbol{\beta}_i^{i-1/2}$, $\boldsymbol{\beta}_{i-1}^{i-1/2} \in \mathbb{R}^m \times \mathbb{R}^m$
$$\Phi^{i-1/2} = \boldsymbol{\beta}_i^{i-1/2}\Phi^{i-1/2} + \boldsymbol{\beta}_{i-1}^{i-1/2}\Phi^{i-1/2}$$
3. Assembly all the contributions belonging to node i to update the solution at the new time step
$$\mathsf{u}_i^{n+1} = \mathsf{u}_i^n - \frac{\Delta t}{h_i}\left(\boldsymbol{\beta}_i^{i-1/2}\Phi^{i-1/2} + \boldsymbol{\beta}_i^{i+1/2}\Phi^{i+1/2}\right) \tag{5.8}$$

It is easy to see that the following choice for the distribution matrices leads to an equivalence with upwind $\mathcal{FV}$/Flux-Difference Splitting (5.6)
$$\boldsymbol{\beta}_i^{i-1/2} = \mathsf{R}\mathsf{B}_i^{i-1/2}\mathsf{R}^{-1}, \quad \text{with} \quad \mathsf{B}_i^{i-1/2} = \Lambda^+\Lambda^{-1}$$
$$\boldsymbol{\beta}_{i-1}^{i-1/2} = \mathsf{R}\mathsf{B}_{i-1}^{i-1/2}\mathsf{R}^{-1}, \quad \text{with} \quad \mathsf{B}_{i-1}^{i-1/2} = \Lambda^-\Lambda^{-1}$$
It is interesting to note that this particular distribution corresponds to apply an upwind Residual Distribution scheme on the transformed system (5.3). Infact applying the one-dimensional scalar upwind $\mathcal{RD}$ method to each scalar decoupled equation we would have obtained the same result. Briefly:

1. Compute the residual
$$\Phi^{i-1/2}(\mathsf{w}^h) = \Lambda\left(\mathsf{w}_i - \mathsf{w}_{i-1}\right) = \Lambda\mathsf{R}^{-1}\left(\mathsf{u}_i - \mathsf{u}_{i-1}\right)$$
remembering that the following transformation holds
$$\Phi^{i-1/2}(\mathsf{w}^h) = \mathsf{R}^{-1}\Phi^{i-1/2}(\mathsf{u}^h) \tag{5.9}$$
2. Split the residual according to the one-dimensional upwind distribution (2.15), hence with distribution matrices $\mathsf{B}_i^{i-1/2}, \mathsf{B}_{i-1}^{i-1/2}$
$$\Phi^{i-1/2}(\mathsf{w}_h^n) = \mathsf{B}_i^{i-1/2}\Phi^{i-1/2}(\mathsf{w}_h^n) + \mathsf{B}_{i-1}^{i-1/2}\Phi^{i-1/2}(\mathsf{w}_h^n)$$
3. Assembly
$$\mathsf{w}_i^{n+1} = \mathsf{w}_i^n - \frac{\Delta t}{h_i}\left(\mathsf{B}_i^{i-1/2}\Phi^{i-1/2}(\mathsf{w}_h^n) + \mathsf{B}_i^{i+1/2}\Phi^{i+1/2}(\mathsf{w}_h^n)\right) \tag{5.10}$$

If one transforms back in conserved variables premultipyling (5.10) by R and then using (5.9), gets (5.8).

5.2 Residual Distribution for non-linear systems

A system of coupled conservation laws can be written, in a general form, as an hyperbolic system of equations. We will consider only two dimensional problems

$$\frac{\partial \mathsf{u}}{\partial t} + \nabla \cdot \mathbf{f} = 0 \quad \text{in } \Omega \in \mathbb{R}^2,\, t \in [0,T] \tag{5.11}$$

with solution $\mathsf{u}(\boldsymbol{x},t) \in \mathbb{R}^m$, flux function $\mathbf{f}(\mathsf{u}) = [\,\mathsf{f}_x\,\mathsf{f}_y\,] \in \mathbb{R}^m \times \mathbb{R}^2$, with $\mathsf{f}_x, \mathsf{f}_y \in \mathbb{R}^m$. In a quasi linear form

$$\frac{\partial \mathsf{u}}{\partial t} + \mathsf{A}_x(\mathsf{u})\frac{\partial \mathsf{u}}{\partial x} + \mathsf{A}_y(\mathsf{u})\frac{\partial \mathsf{u}}{\partial y} = 0 \tag{5.12}$$

where $\mathsf{A}_x = \frac{d\mathsf{f}_x}{d\mathsf{u}}$ and $\mathsf{A}_y = \frac{d\mathsf{f}_y}{d\mathsf{u}}$. Collecting both in a three dimensional array $\mathbf{A}(\mathsf{u}) = \frac{d\mathbf{f}}{d\mathsf{u}} = \left[\begin{array}{cc} \mathsf{A}_x & \mathsf{A}_y \end{array}\right] \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^2$ we obtain the Jacobian of the flux. The system is called hypebolic if, for any vector $\boldsymbol{\xi} = (\xi_x, \xi_y) \in \mathbb{R}^2$, the $m \times m$ matrix K

$$\mathsf{K}(\boldsymbol{\xi}, \mathsf{u}) = \mathbf{A}(\mathsf{u}) \cdot \boldsymbol{\xi} = \mathsf{A}_x(\mathsf{u})\,\xi_x + \mathsf{A}_y(\mathsf{u})\,\xi_y \tag{5.13}$$

is diagonalizable with real eigenvalues. Everywhere in the text we shall denotes with $\Lambda(\boldsymbol{\xi}, \mathrm{u}) = \mathrm{Diag}\,\{\lambda_1, \lambda_2, ...\lambda_p\}$ the diagonal matrix of the eigenvalues of K, while with $\mathsf{R}(\boldsymbol{\xi}, \mathrm{u})$ the matrix of right eigenvectors of K. Thus we can bring K to a diagonal form by a similarity transformation

$$\Lambda(\boldsymbol{\xi}, \mathsf{u}) = \mathsf{R}^{-1}(\boldsymbol{\xi}, \mathsf{u})\,\mathsf{K}(\boldsymbol{\xi}, \mathsf{u})\,\mathsf{R}(\boldsymbol{\xi}, \mathsf{u}) \tag{5.14}$$

A Residual Distribution approximation of (5.11) is obtained throught the following step

1. Once we have approximated the solution with a P1 approximation u^h, compute the residual over every element

$$\Phi^K = \int_K \nabla \cdot \mathbf{f}(\mathsf{u}^h)\, d\boldsymbol{x} = \int_{\partial K} \mathbf{f}(\mathsf{u}^h) \cdot \boldsymbol{n}\, d\boldsymbol{x} \quad \forall K \in \mathcal{T}^h \tag{5.15}$$

2. Distribute the residual to the nodes i, j, k of the element through the distribution matrix $\boldsymbol{\beta}_i^K, \boldsymbol{\beta}_j^K, \boldsymbol{\beta}_k^K$

$$\Phi^K = \boldsymbol{\beta}_i^K \Phi^K + \boldsymbol{\beta}_j^K \Phi^K + \boldsymbol{\beta}_k^K \Phi^K = \sum_{j \in K} \Phi_j^K \tag{5.16}$$

If I^m is the $m \times m$ identity matrix, the consistency condition is translated as follows

$$\sum_{j \in K} \boldsymbol{\beta}_j^K = \mathrm{I}^m \tag{5.17}$$

3. For every node i assemby the contribution from all $K \in \mathcal{D}_i$ and evolve u_i in time

$$|S_i|\frac{d\mathsf{u}_i}{dt} + \sum_{K\in\mathcal{D}_i} \Phi_i^K = 0 \quad \forall i \in \mathcal{T}^h \tag{5.18}$$

The steps define a generalization of paragraph 1.2.

In the splitting step, in order to end up with an upwind distribution, we will need to know some information about the eigenstructure of the jacobian K (as in 1D we needed the scalar parameter k), the access to the quasi-linear or better to a linearized form becomes thus necessary. Regarding the linearization we have to face the same problem of the scalar case: for conservativity reasons the linearization should be exact and this fixes the correct averaged Jacobian $\bar{\mathbf{A}}$

$$\begin{aligned}\int_K \nabla\cdot\mathbf{f}(\mathsf{u}^h)\,d\boldsymbol{x} &= \int_K \left(\mathsf{A}_x(\mathsf{u}^h)\frac{\partial \mathsf{u}^h}{\partial x} + \mathsf{A}_y(\mathsf{u}^h)\frac{\partial \mathsf{u}^h}{\partial y}\right) d\boldsymbol{x}\\ &= \int_K \mathsf{A}_x(\mathsf{u}^h)\,d\boldsymbol{x}\frac{\partial \mathsf{u}^h}{\partial x} + \int_K \mathsf{A}_y(\mathsf{u}^h)\,d\boldsymbol{x}\frac{\partial \mathsf{u}^h}{\partial y}\\ &= \left(\bar{\mathsf{A}}_x\frac{\partial \mathsf{u}^h}{\partial x} + \bar{\mathsf{A}}_y\frac{\partial \mathsf{u}^h}{\partial y}\right)|K|\end{aligned} \tag{5.19}$$

with

$$\bar{\mathsf{A}}_x = \frac{1}{|K|}\int_K \mathsf{A}_x(\mathsf{u}^h)\,d\boldsymbol{x} \quad \text{and} \quad \bar{\mathsf{A}}_y = \frac{1}{|K|}\int_K \mathsf{A}_y(\mathsf{u}^h)\,d\boldsymbol{x} \tag{5.20}$$

Using this linearization we can write the residual in the following form, very similar to the scalar analogue (2.45)

$$\begin{aligned}\int_K \nabla\cdot\mathbf{f}(\mathsf{u}^h)\,d\boldsymbol{x} &= \frac{1}{2}\sum_{j\in K}\left(\bar{\mathsf{A}}_x n_{x,j} + \bar{\mathsf{A}}_y n_{y,j}\right)\mathsf{u}_j\\ &= \sum_{j\in K}\mathsf{K}_j\mathsf{u}_j\end{aligned}$$

with upwind matrix defined by

$$\mathsf{K}_j = \frac{1}{2}\left(\bar{\mathsf{A}}_x n_{x,j} + \bar{\mathsf{A}}_y n_{y,j}\right) \tag{5.21}$$

Moreover we define the following matrix

$$\bar{\mathsf{K}}(\boldsymbol{\xi}) = \frac{1}{2}\left(\bar{\mathsf{A}}_x \xi_{x,} + \bar{\mathsf{A}}_y \xi_{y,}\right)$$

which admits the usual decomposition

$$\bar{\mathsf{K}}(\boldsymbol{\xi}) = \bar{\mathsf{R}}(\boldsymbol{\xi})\,\bar{\Lambda}(\boldsymbol{\xi})\,\bar{\mathsf{R}}^{-1}(\boldsymbol{\xi})$$

We conclude this section with the hope that, almost all the nice properties of scalar $\mathcal{RD}$, could be extended to systems. Even in the cases when this turns to be true, the extension is not trivial at all.

Multidimensional Upwinding

In the first example we have seen that, simply working on characteristic variables instead of the physical ones, it is possible to design a one-dimensional upwind $\mathcal{RD}$ scheme for systems. This is certainley true also in multidimensional problems where we can imply an upwind scheme for every simple wave, thus coming to a scheme which inherits from the scalar case the property of a genuinely multidimensional upwinding. Then, it seems logical to use the eigenvalues of K_j as a reference to decide whether node i receives a larger or smaller amount of residual.

On the other hand, for systems, the concepts of inflow/outflow faces, downstream/upstream nodes, one/two-targert cells have no more sense. A sense clearly holds for each simple wave, but for the system in conserved variables the coupling of the equations makes the cell generally three-target. The definition of upwinding given in section (2.2) has to be revisited for systems.

Definition (Matrix $\mathcal{RD}$, Upwinding). *A $\mathcal{RD}$ method in the form (5.18) is upwind if*

$$\mathsf{K}_i^+ = 0 \Rightarrow \Phi_i^K = 0 \tag{5.22}$$

Stability and Accuracy

The direct extension of the results provied for the scalar case in paragraph (2.4.1) could be misleading. First of all the existence of a maximum principle for a system of non-linear conservation laws is not a trivial task. This does not mean that a non oscillatory behaviour near discontinuities is desired/expected. Even if, heuristically, we extend the positive analysis to systems getting to a compact form like

$$\mathsf{u}_i^{n+1} = \mathsf{u}_i^n - \frac{\Delta t}{|S_i|} \sum_{j \in K} \mathsf{C}_{ij} \mathsf{u}_j$$

it is not easy to handle with the LED condition, this time applied to a matrix $\mathsf{C}_{ij} \leq 0$. To avoid this difficulties, someone searches for a condition relying on entropy consideration but this is beyond the scope of the thesis [6]. We will extend directly the scalar schemes to systems without providing a rigorous demonstration of why they are oscillation-free or not.

We can only say that, given a direction $\boldsymbol{\xi}$, the solution can be decomposed as sum of simple waves

$$\mathsf{u} = \mathsf{R}(\boldsymbol{\xi}, \mathsf{u})\mathsf{w} \Rightarrow \mathsf{u}^h = \bar{\mathsf{R}}(\boldsymbol{\xi})\mathsf{w}^h \tag{5.23}$$

and the linearized residual can be splitted into a linear combination of simple residuals ϕ_l^K, each of them acting on a single simple wave

$$\begin{aligned}\Phi^K(\mathsf{u}^h) &= \Phi^K(\bar{\mathsf{R}}(\boldsymbol{\xi})\mathsf{w}^h) \\ &= \bar{\mathsf{R}}(\boldsymbol{\xi})\Phi^K(\mathsf{w}^h) = \sum_{l=1,m} \phi_l^K(w_h^l)\mathsf{r}_l\end{aligned}$$

From this point to the end, in order to lighten a little bit the notation we skip the apex K in the defintion of residual. Using (2.47), every distributed sub-residual $\phi_i^l(w^l)$ can be written in the compact form

$$\phi_i^l(w_h^l) = \sum_{j\in K} c_{ij}^l \left(w_j^l - w_i^l\right) \tag{5.24}$$

Hence, playing with the coefficient c_{ij}^l, we can still define a positive scheme for every wave, for example using a N distribution for ϕ_i^l. Heuristically we can say also that, many schemes we shall present in a while, the LDA and N schemes for example, once written in characteristic variables, correspond exactly to their respective scalar versions applied to each simple wave and, somehow, we expect that the scheme acting on conserved variables inherits the properties (LED , positivity) from each scalar scheme.

On the contrary accuracy results presented in paragraph 1.4.2 can be extended very easily to systems. We give directly the result

Definition (Matrix $\mathcal{RD}$, Second Order Accuracy and Linearity Preserving schemes). *A $\mathcal{RD}$ method in the form (5.18) is second order accurate at steady state if*

$$\Phi_i = \mathcal{O}(h^3) \tag{5.25}$$

Moreover the matrix scheme is said to be linearity preseving if the distribution matrix $\boldsymbol{\beta}_i^K$ is uniformly bounded with respect to the solution and data of the problem. Linearity preserving schemes are second order accurate at steady state.

5.2.1 Lax Friederich scheme

The Lax-Friederich scheme for systems write

$$\Phi_i^{LxF} = \frac{1}{3}\Phi_i^K - \alpha^K \sum_{j\in K, j\neq i} (\mathsf{u}_i - \mathsf{u}_j) \qquad \alpha^K = \max_{j\in K} \det \mathsf{K}_j \tag{5.26}$$

The parameter α has to give the correct dimensionalization and the correct order of magnitude to the dissipation operator. (5.26) is a generalization of (2.76). The scheme is positive and first order accurate

5.2.2 SUPG scheme

The SUPG scheme is also derived directly from a formal generalization of its scalar counterpart (2.82). The correspondent distribution matrix is

$$\boldsymbol{\beta}_i^{SUPG} = \frac{1}{3}\mathsf{I}^m + \mathsf{K}_i\mathsf{T}, \qquad \mathsf{T} = \left(\sum_{j\in K} |\mathsf{K}_j|\right)^{-1} \tag{5.27}$$

which lead to a second order non-positive scheme.

5.2.3 LDA scheme

Extending the LDA scalar scheme (2.88)

$$\boldsymbol{\beta}_i^{LDA} = \mathsf{K}_i^+\mathsf{N}, \qquad \mathsf{N} = \left(\sum_{j\in K} \mathsf{K}_j^+\right)^{-1} \tag{5.28}$$

we get a second order upwind scheme.

5.2.4 N scheme

And finally, starting from the scalar counterpart (2.83) we have the system N scheme

$$\boldsymbol{\Phi}_i^N = \mathsf{K}_i^+ \left(\mathsf{u}_i - \mathsf{u}_{in}\right), \qquad \mathsf{u}_{in} = -\sum_{j\in K} \mathsf{N}\mathsf{K}_j^- \mathsf{u}_j \tag{5.29}$$

which is positive and upwind but only first order accurate.

5.2.5 Blended schemes

Once a second order and a first order residual $\boldsymbol{\Phi}_i^{\mathcal{LP}}, \boldsymbol{\Phi}_i^{\mathcal{P}}$ has been computed, the issue is how to implement the limiter. The idea of Abgrall and Mezine suggested in [18] is to decompose the problem in simple waves and define a scalar limiter acting on each wave. We define the right and left eigenvectors associated to the flow direction (but the result is rather independent of this choice)

$$\boldsymbol{\Phi}^K(\mathsf{w}^h) = \boldsymbol{\Phi}^K(\bar{\mathsf{L}}\mathsf{u}^h) = \bar{\mathsf{L}}\boldsymbol{\Phi}^K, \quad \phi_l^K(w^l) = \mathsf{l}_l^T \cdot \boldsymbol{\Phi}^K \tag{5.30}$$

Hence using the same limiter as in (2.90) for every l-wave

$$\mu_l = \frac{|\phi_l^K(w^l)|}{\sum_{j\in K} |\phi_{l,j}^N(w^l)|} = \frac{|\mathsf{l}_l^T \cdot \boldsymbol{\Phi}^K|}{\sum_{j\in K} |\mathsf{l}_l^T \cdot \boldsymbol{\Phi}_j^N|} \quad \forall l = 1, 2, ..m \tag{5.31}$$

Blending residuals in the domain of characteristic variables we get, with scalar notation

$$\phi^B_{l,i}(\mathsf{w}) = \mathsf{l}^T_l \cdot \mathbf{\Phi}^{\mathcal{LP}}_i + \mu_l \left(\mathsf{l}^T_l \cdot \mathbf{\Phi}^{\mathcal{P}}_i - \mathsf{l}^T_l \cdot \mathbf{\Phi}^{\mathcal{LP}}_i \right) \tag{5.32}$$

Projecting back into conserved variables

$$\mathbf{\Phi}^B_i = \bar{\mathsf{R}} \mathbf{\Phi}^B_i(\mathsf{w}^h), \quad \mathbf{\Phi}^B_i = \sum_{l=1,p} \phi^B_{l,i} \mathsf{r}_l \tag{5.33}$$

Substituting (5.32) in (5.33) we get

$$\mathbf{\Phi}^B_i = \sum_{l=1,p} \left(\mathsf{l}^T_l \cdot \mathbf{\Phi}^{\mathcal{LP}}_i + \mu_l \left(\mathsf{l}^T_l \cdot \mathbf{\Phi}^{\mathcal{P}}_i - \mathsf{l}^T_l \cdot \mathbf{\Phi}^{\mathcal{LP}}_i \right) \right) \mathsf{r}_l \tag{5.34}$$

where μ_l is computed with (5.31).

5.2.6 Limited schemes

When extending the limiting procedure to system, we face the same problem of the blending approach. We have the PSI limiter that works very good for the scalar case and we would like to extend it to systems. The idea of defining a limiter on each wave allows us to apply a PSI limiter on every positive residual distribution $\mathbf{\Phi}^{\mathcal{P}}_i$. Infact

$$\beta^l_i = \frac{\left(\phi^{\mathcal{P}}_{l,i}(w^l) \right)^+}{\sum_{j \in K} \left(\phi^{\mathcal{P}}_{l,j}(w^l) \right)^+} = \frac{\left(\mathsf{l}^T_l \cdot \mathbf{\Phi}^{\mathcal{P}}_i \right)^+}{\sum_{j \in K} \left(\mathsf{l}^T_l \cdot \mathbf{\Phi}^{\mathcal{P}}_j \right)^+} \quad \forall l = 1, 2, ..m \tag{5.35}$$

Limiting the positive residual projected in the domain of characteristic variables results in getting m limited residuals

$$\phi^{LIM}_{l,i} = \beta^l_i \phi^{\mathcal{P}}_l = \beta^l_i \left(\mathsf{l}^T_l \cdot \mathbf{\Phi}^{\mathcal{P}} \right) \forall l = 1, 2, ..m \tag{5.36}$$

Finally we project back into conserved variables

$$\mathbf{\Phi}^{LIM}_i = \bar{\mathsf{R}} \mathbf{\Phi}^{LIM}_i(\mathsf{w}^h), \quad \mathbf{\Phi}^{LIM}_i = \sum_{l=1,p} \phi^{LIM}_{l,i} \mathsf{r}_l \tag{5.37}$$

Substituting (5.36) in (5.37) we get

$$\mathbf{\Phi}^{LIM}_i = \sum_{l=1,m} \beta^l_i \left(\mathsf{l}^T_l \cdot \mathbf{\Phi}^{\mathcal{P}} \right) \mathsf{r}_l \tag{5.38}$$

with β^l_i computed with (5.35).

5.2.7 LLxFs and LLxF-SUPG schemes

Also for systems the LLxF scheme suffers from spurious oscillation that, eventually, can reduce the accuracy to first order. As in section (2.5.7), a SUPG term, tuned by some parameter, is added. For Euler equations this parameter is introduced by Abgrall in [8] in a way very similar to the scalar analogue (2.96). He reasons that the residual, projected in characteristic variables, has an entropy component which is defined through the usual projection (5.30), this time on the correct left eigenvector corresponding to the entropy wave, l_s^T

$$\phi_s^K = \mathsf{l}_s^T \cdot \mathbf{\Phi}^K$$

The limiter is then computed

$$\delta(\mathsf{u}^h) = \min\left(1, \frac{1}{\frac{|\phi_s^K|}{h_K^2} + \varepsilon}\right) \tag{5.39}$$

It has the correct order of magnitude, hence $\delta(\mathsf{u}^h) \simeq 1$ in smooth region and $\delta(\mathsf{u}^h) \simeq 0$ near discontinuities. An explanation for the steady case is given in the same paper [8] but (5.39) captures very well discontinuities also in unsteady cases. The matrix distribution for the LLxFs reads

$$\boldsymbol{\beta}_i^{LLxFs} = \boldsymbol{\beta}_i^{LLxF} + \delta(\mathsf{u}^h)\boldsymbol{\beta}_i^{SUPG} \tag{5.40}$$

Indeed another version, for which has been demonstrated in [9] that better results are obtained, consists in a full blending between a LLxF and a SUPG scheme through the limiter (5.39). It is defined by the following matrix distribution

$$\boldsymbol{\beta}_i^{LLxF-SUPG} = \left(1 - \delta(\mathsf{u}^h)\right)\boldsymbol{\beta}_i^{LLxF} + \delta(\mathsf{u}^h)\boldsymbol{\beta}_i^{SUPG}$$

The scheme is called in shorthand notation LLxF-SUPG and it is the one used in the computation.

5.3 Residual Distribution for Euler equations

Through the chapter we have built a powerful tool (matrix $\mathcal{RD}$) to solve non-linear hyperbolic systems of conservation laws in the form (5.11). We will test it on the Euler equations governing inviscid compressible flows. The system states the conservation of mass, momentum and energy for a perfect gas. The conserved variables and flux matrix are given by

$$\mathsf{u} = \begin{bmatrix} \rho \\ \rho u \\ \rho v \\ \rho E \end{bmatrix}, \quad \mathbf{f}(\mathsf{u}) = \begin{bmatrix} \rho u & \rho v \\ \rho u^2 + p & \rho uv \\ \rho uv & \rho v^2 + p \\ \rho H u & \rho H v \end{bmatrix} \tag{5.41}$$

where ρ is the fluid density, $\boldsymbol{u} = (u, v)$ is the flow speed, E is the total energy per unit mass, H is the total entalpy per unit mass

$$H = E + \frac{p}{\rho}$$

For polytropic ideal gas thermodynamic properties are completely defined by the following pair of equations of state

$$p(T, \rho) = RT\rho, \quad e(T) = \frac{RT}{(\gamma - 1)} \tag{5.42}$$

where we have introduced the internal energy per unit mass e and the temperature of the fluid T. Combining the two equations we obtain p as a function of e and ρ

$$p(e, \rho) = (\gamma - 1)\, e\rho \tag{5.43}$$

The Euler equations are closed with the definition of internal energy

$$e = E - \frac{1}{2}\|\boldsymbol{u}\| \tag{5.44}$$

since we are able to express the pressure as function of the unknown ρ, u, v, E.

To define the eigenstructure of the Euler equations we need the definition of another thermodynamic variable, the speed of sound. For polytropic ideal gas

$$c = \gamma\,(\gamma - 1)\, e$$

We compute the jacobian of the flux $\mathbf{A}(\mathsf{u}) = \frac{d\mathbf{f}}{d\mathsf{u}}$, then the matrix $\mathsf{K}(\mathsf{u}, \xi)$ according to (5.13). The eigenvalues of $\mathsf{K}(\mathsf{u}, \boldsymbol{\xi})$ are

$$\lambda_{1,4}(\mathsf{u}, \boldsymbol{\xi}) = \boldsymbol{u} \cdot \boldsymbol{\xi} \mp c(\mathsf{u})\|\boldsymbol{\xi}\|, \quad \lambda_{2,3}(\mathsf{u}, \boldsymbol{\xi}) = \boldsymbol{u} \cdot \boldsymbol{\xi} \tag{5.45}$$

which are all real, making the system hyperbolic and $\mathsf{K}(\mathsf{u}, \boldsymbol{\xi})$ diagonalizable with diagonal matrix Λ.

Roe-Struijs-Deconinck linearization

In order to put the residual in a linearized form (5.50) the correct conservative linearization of the flux now is performed. We use the fact that the flux components are quadratic function of the the Roe parameter vector

$$\mathsf{z} = \sqrt{\rho}\begin{bmatrix} 1 & u & v & H \end{bmatrix}^T \tag{5.46}$$

First we express the residual as a function of z^h. Assuming a piecewise linear variation of z^h

$$\int_K \nabla \cdot \mathbf{f}(\mathsf{z}^h)\, d\boldsymbol{x} = \left(\frac{d\mathsf{f}_x}{d\mathsf{z}}(\bar{\mathsf{z}}) \frac{\partial \mathsf{z}^h}{\partial x} + \frac{d\mathsf{f}_y}{d\mathsf{z}}(\bar{\mathsf{z}}) \frac{\partial \mathsf{z}^h}{\partial y} \right) |K| \tag{5.47}$$

The passage from the conservative form to the linearized one is exact only if

$$\frac{d\mathsf{f}_x}{d\mathsf{z}}(\bar{\mathsf{z}}) = \frac{1}{|K|} \int_K \mathsf{A}_x(\mathsf{z}^h)\, d\boldsymbol{x} \quad \text{and} \quad \frac{d\mathsf{f}_y}{d\mathsf{z}}(\bar{\mathsf{z}}) = \frac{1}{|K|} \int_K \mathsf{A}_y(\mathsf{z}^h)\, d\boldsymbol{x} \tag{5.48}$$

The choice of the variable z seems the only possible one leading to a full conservative scheme, in this way infact the above integrals exist and are very easy to compute. $\mathbf{A}(\mathsf{z}^h)$ is linear and follows immediately that we have to evaluate the jacobian at a proper average state which correspond to the arithmetical average of Roe variables over the element

$$\bar{\mathbf{A}} = \mathbf{A}(\bar{\mathsf{z}}), \quad \text{with} \quad \bar{\mathsf{z}} = \sum_{j \in K} \frac{\mathsf{z}_j}{3} \tag{5.49}$$

This averageing, for one dimensional problems, reduce to the classical 1D Roe conservative linearization [19] and rapresents its multidimensional extension as suggested by [20]. Substituting (5.49) in (5.47), the linearized residual is computed

$$\begin{aligned}
\int_K \nabla \cdot \mathbf{f}\, d\boldsymbol{x} &= \left(\frac{d\mathsf{f}_x}{d\mathsf{z}}(\bar{\mathsf{z}}) \frac{\partial \mathsf{z}^h}{\partial x} + \frac{d\mathsf{f}_y}{d\mathsf{z}}(\bar{\mathsf{z}}) \frac{\partial \mathsf{z}^h}{\partial y} \right) |K| \\
&= \left(\frac{d\mathsf{f}_x}{d\mathsf{u}}(\bar{\mathsf{z}}) \frac{d\mathsf{u}}{d\mathsf{z}}(\bar{\mathsf{z}}) \frac{\partial \mathsf{z}^h}{\partial x} + \frac{d\mathsf{f}_y}{d\mathsf{u}}(\bar{\mathsf{z}}) \frac{d\mathsf{u}}{d\mathsf{z}}(\bar{\mathsf{z}}) \frac{\partial \mathsf{z}^h}{\partial y} \right) |K| \\
&= \left(\mathsf{A}_x(\bar{\mathsf{z}}) \frac{d\mathsf{u}}{d\mathsf{z}}(\bar{\mathsf{z}}) \frac{\partial \mathsf{z}^h}{\partial x} + \mathsf{A}_y(\bar{\mathsf{z}}) \frac{d\mathsf{u}}{d\mathsf{z}}(\bar{\mathsf{z}}) \frac{\partial \mathsf{z}^h}{\partial y} \right) |K| \\
&= \left(\mathsf{A}_x(\bar{\mathsf{z}}) \frac{\partial \mathsf{u}^h}{\partial x} + \mathsf{A}_y(\bar{\mathsf{z}}) \frac{\partial \mathsf{u}^h}{\partial y} \right) |K|
\end{aligned} \tag{5.50}$$

Matrix K_j follows from (5.21)

$$\mathsf{K}_j = \frac{1}{2} \left(\mathsf{A}_x(\bar{\mathsf{z}})\, n_{x,j} + \mathsf{A}_y(\bar{\mathsf{z}})\, n_{y,j} \right) \tag{5.51}$$

5.4 Euler equations in ALE framework

We don't derive again the system of conservation laws in ALE formulation but we extend it directly from the scalar case. Extending the scalar equation (4.19) we get the integral form of Euler equations in ALE framework

$$\left. \frac{\partial}{\partial t} \right|_X \int_{C(t)} \mathsf{u}\, d\boldsymbol{x} + \int_{C(t)} \nabla \cdot (\mathbf{f} - \mathsf{u}\boldsymbol{\sigma})\, d\boldsymbol{x} = 0 \tag{5.52}$$

Extending (4.20) brings to the differential form of Euler equations written in ALE framework

$$\left.\frac{\partial\left(J_A\mathsf{u}\right)}{\partial t}\right|_X + J_A\nabla\cdot\left(\mathbf{f}-\mathsf{u}\boldsymbol{\sigma}\right)=0 \tag{5.53}$$

Given a vector $\boldsymbol{\xi}\in\mathbb{R}^2$ the jacobian of the fluxes in ALE framework writes

$$\begin{aligned}\mathsf{K}(\mathsf{u},\boldsymbol{\xi},\boldsymbol{\sigma}) &= \frac{d}{d\mathsf{u}}\left(\mathbf{f}-\mathsf{u}\boldsymbol{\sigma}\right)\cdot\boldsymbol{\xi}=\frac{d\mathbf{f}}{d\mathsf{u}}\cdot\boldsymbol{\xi}-\frac{d}{d\mathsf{u}}\left(\mathsf{u}\boldsymbol{\sigma}\right)\cdot\boldsymbol{\xi}\\ &= \mathbf{A}\cdot\boldsymbol{\xi}-\frac{d}{d\mathsf{u}}\left(\left(\boldsymbol{\sigma}\cdot\boldsymbol{\xi}\right)\mathsf{u}\right)=\mathbf{A}\cdot\boldsymbol{\xi}-\boldsymbol{\sigma}\cdot\boldsymbol{\xi}\frac{d\mathsf{u}}{d\mathsf{u}}\\ &= \mathsf{A}_x(\mathsf{u})\,\xi_x+\mathsf{A}_y(\mathsf{u})\,\xi_y-\left(\sigma_x\xi_x+\sigma_y\xi_y\right)\mathsf{I}^4\end{aligned} \tag{5.54}$$

There is a local modification of the jacobian matrix due to the ALE part in the flux. The new eigenvalues, all real, are grouped into the diagonal matrix

$$\Lambda(\mathsf{u},\boldsymbol{\xi},\boldsymbol{\sigma})=(\boldsymbol{u}-\boldsymbol{\sigma})\cdot\boldsymbol{\xi}\,\mathsf{I}^4-\mathrm{Diag}\left\{c,00,-c\right\} \tag{5.55}$$

Right and left eigenvectors does not change with the reference system and remains the one computed in the Eulerian framework $\mathsf{R}(\mathsf{u},\boldsymbol{\xi}),\mathsf{L}(\mathsf{u},\boldsymbol{\xi})$.

5.5 $\mathcal{RD}$-$\mathcal{RK}$2 for Euler Equation in ALE framework

This paragraph rapresents the last step of the thesis: the matrix $\mathcal{RD}$ schemes presented in paragraph (5.2) together with the explicit Runge-Kutta time integrator of paragraph (4.4.2) are used to approximate Euler Equation in ALE formulation.

The extension to systems of the $\mathcal{RD}$ explicit Runge Kutta two approximation for scalar ALE conservation laws expressed by (4.51)(4.54) is straightforward. Using the same notation introduced for that formulas, the selective lumped formulation writes

$$|\tilde{S}_i^{n+1/2}|\frac{\Delta\mathsf{u}_i^k}{\Delta t}=-\sum_K\left(\Phi_i^{RK(k)}-\sum_j\tilde{m}_{ij}^{GAL}\frac{\Delta\overline{\mathsf{u}_j^k}}{\Delta t}\right) \tag{5.56}$$

where the definition for the scalar $|\tilde{S}_i^{n+1/2}|$, $\tilde{m}_{ij}^{GAL}$ are the same of paragraph (4.4.2) while for the vector $\Delta\mathsf{u}_i^k$, $\Delta\overline{\mathsf{u}_j^k}$ we use simple extensions of the scalar definitions already given. For the residual the scalar definition (4.50) is rewritten in a matrix $\mathcal{RD}$ formalism

$$\begin{cases}\Phi_i^{RK(1)}=\Phi_i(\mathsf{u}^h)\\ \Phi_i^{RK(2)}=\sum_{j\in K}m_{ij}^K\frac{\mathsf{u}_j^1-\mathsf{u}_j^n}{\Delta t}+\frac{1}{2}\left(\Phi_i(\mathsf{u}_h^1)+\Phi_i(\mathsf{u}_h^n)\right)\end{cases}$$

For the global lumped formulation one has

$$\begin{cases} |\tilde{S}_i^{\,n+1/2}| \frac{\mathsf{u}_i^1 - \mathsf{u}_i^n}{\Delta t} = -\sum_K \Phi_i^{RK(1)} \\ |\tilde{S}_i^{\,n+1/2}| \frac{\mathsf{u}_i^{n+1} - \mathsf{u}_i^1}{\Delta t} = -\sum_K \Phi_i^{RK(2)} \end{cases} \tag{5.57}$$

The only passage where some care is needed is the computation of the total residual which becomes

$$\Phi^K(\mathsf{u}^h) = \int_K \left(\nabla \cdot \mathbf{f}(\mathsf{u}^h) - \boldsymbol{\sigma}^h \nabla \mathsf{u}^h \right) d\boldsymbol{x}$$

Using, for the advective part, the linearized form of the residual (5.50), the total residual for Euler equation in ALE framework writes

$$\int_K \left(\nabla \cdot \mathbf{f}(\mathsf{u}^h) - \boldsymbol{\sigma}^h \nabla \mathsf{u}^h \right) d\boldsymbol{x} = \left(\mathsf{A}_x\left(\bar{\mathsf{z}}\right) \frac{\partial \mathsf{u}^h}{\partial x} + \mathsf{A}_y\left(\bar{\mathsf{z}}\right) \frac{\partial \mathsf{u}^h}{\partial y} \right) |K| - \int_K \boldsymbol{\sigma}^h \, d\boldsymbol{x} \cdot \nabla \mathsf{u}^h \tag{5.58}$$

Since $\boldsymbol{\sigma}^h$ is linear over the element we can evaluate the integral at the second term

$$\int_K \boldsymbol{\sigma}^h \, d\boldsymbol{x} = \sum_{j \in K} \frac{\boldsymbol{\sigma}_j}{3} |K| = \bar{\boldsymbol{\sigma}} |K| \tag{5.59}$$

Thus the linearization remains exact also in ALE framework, hence the conservativity of the method is preserved. Remember how we compute the gradient of a linear function over a triangle

$$\nabla \mathsf{u}^h = \frac{1}{2|K|} \sum_{j \in K} \boldsymbol{n}_j \mathsf{u}_j \tag{5.60}$$

With some calculation the total residual (5.58) becomes

$$\begin{aligned} \int_K \left(\nabla \cdot \mathbf{f}(\mathsf{u}^h) - \boldsymbol{\sigma}^h \nabla \mathsf{u}^h \right) d\boldsymbol{x} &= \left(\mathsf{A}_x\left(\bar{\mathsf{z}}\right) \frac{\partial \mathsf{u}^h}{\partial x} + \mathsf{A}_y\left(\bar{\mathsf{z}}\right) \frac{\partial \mathsf{u}^h}{\partial y} \right) |K| - \bar{\boldsymbol{\sigma}} \cdot \nabla \mathsf{u}^h |K| \\ &= \frac{1}{2} \left(\mathsf{A}_x\left(\bar{\mathsf{z}}\right) \sum_{j \in K} n_{x,j} \mathsf{u}_j + \mathsf{A}_y\left(\bar{\mathsf{z}}\right) \sum_{j \in K} n_{y,j} \mathsf{u}_j \right) - \frac{1}{2} \bar{\boldsymbol{\sigma}} \sum_{j \in K} \boldsymbol{n}_j \mathsf{u}_j \\ &= \sum_{j \in K} \frac{1}{2} \left(\mathsf{A}_x\left(\bar{\mathsf{z}}\right) n_{x,j} \mathsf{u}_j + \mathsf{A}_y\left(\bar{\mathsf{z}}\right) n_{y,j} \mathsf{u}_j \right) - \sum_{j \in K} \left(\frac{1}{2} \bar{\boldsymbol{\sigma}} \cdot \boldsymbol{n}_j \, \mathsf{I}^4 \right) \mathsf{u}_j \\ &= \sum_{j \in K} \left(\frac{1}{2} \mathbf{A}(\bar{\mathsf{z}}) \cdot \boldsymbol{n}_j - \frac{1}{2} \bar{\boldsymbol{\sigma}} \cdot \boldsymbol{n}_j \, \mathbf{I}^4 \right) \mathsf{u}_j \\ &= \sum_{j \in K} \mathsf{K}_j^{ALE} \mathsf{u}_j \end{aligned} \tag{5.61}$$

This results in the following modifcation of the upwind matrix which closely resamble the upwind parameter obtained in the scalar case (4.27). Infact from the above equation

$$\mathsf{K}_j^{ALE} = \mathsf{K}_j - \frac{1}{2} \bar{\boldsymbol{\sigma}} \cdot \boldsymbol{n}_j \, \mathsf{I}^4 \tag{5.62}$$

The upwind matrix is used to define the various matrix linear schemes, LxF (5.26), LDA (5.28), N (5.29), SUPG (5.27). The non linear schemes, both the limited (5.38) and the blended one (5.34), are built starting from the linear ones, where the only additional information we need are the matrices of right and left eigenvectors of the Roe linearized jacobian $\mathsf{K}(\bar{\mathsf{z}}, \boldsymbol{\xi})$ which does not change respect to the Eulerian formulation.

Finally for the computation of the residual $\Phi_i^{RK(k)}$ we reminds the scalar formulas in section (3.1.1) which can be immediatly extended to systems.

5.6 Boundary conditions

Up to now the imposition of boundary conditions has not been addressed yet, it has been assumed that the elements share no edges with the boundary. Here we follow [21]. If an element K share an edge Γ_K with the boundary Γ_h, the following boundary residual has to be added to the total one computed in (5.15)

$$\Phi^{\Gamma_K} = \int_{\Gamma_K} \left(\mathbf{f}^{\partial}(\mathsf{u}^h) - \mathbf{f}(\mathsf{u}^h) \right) \cdot \boldsymbol{n} \, d\boldsymbol{x} \tag{5.63}$$

where $\mathbf{f}^{\partial}$ is the boundary flux. We have already seen that conservation implies the following relation

$$\sum_{i \in \mathcal{T}_h} \sum_{K \in \mathcal{D}_i} \Phi_i^K = \int_{\Gamma_h} \mathbf{f}^{\partial}(\mathsf{u}^h) \cdot \boldsymbol{n} \, d\boldsymbol{x} \tag{5.64}$$

Separating the contribution of the domain residuals $\Phi_i^{K,d}$ and the contribution of the boundary ones $\Phi_i^{\Gamma_K}$, the above relation writes

$$\begin{aligned} \sum_{i \in \mathcal{T}_h} \sum_{K \in \mathcal{D}_i} \Phi_i^K &= \sum_{i \in \mathcal{T}_h} \sum_{K \in \mathcal{D}_i} \Phi_i^{K,d} + \sum_{i \in \Gamma_h} \sum_{K : K \in \mathcal{D}_i, \partial K \in \Gamma_h} \Phi_i^{\Gamma_K} \\ &= \sum_{K \in \mathcal{T}_h} \sum_{j \in K} \Phi_j^{K,d} + \sum_{K : \partial K \in \Gamma_h} \sum_{j \in \Gamma_K} \Phi_j^{\Gamma_K} \end{aligned} \tag{5.65}$$

Conservation (5.64) is respected if the distribution for the boundary residual respects the following relation

$$\sum_{j \in \Gamma_K} \Phi_j^{\Gamma_K} = \Phi^{\Gamma_K}$$

This means that we have to split separately the boundary residual and the domain ones. A solution could be a centered distribution using the trace of the Galerkin shape function along the boundary $\varphi_j(x) = \varphi_j(\boldsymbol{x})|_{\Gamma_K}$

$$\Phi_j^{\Gamma_K} = \int_{\Gamma_h} \left(\mathbf{f}^{\partial}(\mathsf{u}^h) - \mathbf{f}(\mathsf{u}^h) \right) \cdot \boldsymbol{n} \, \varphi_j(x) \, d\boldsymbol{x}$$

Two types of conditions can be imposed continuosly along the boundaries

1. Wall boundary conditions: we impose the condition $\boldsymbol{u} \cdot \boldsymbol{n} = 0$ directly when we compute the product $\mathbf{f}^{\partial} \cdot \boldsymbol{n}$ in (5.63). This correspond to have

$$\mathbf{f}^{\partial}(\mathsf{u}^h) \cdot \boldsymbol{n} = \begin{pmatrix} 0 \\ p(\mathsf{u}^h) n_x \\ p(\mathsf{u}^h) n_y \\ 0 \end{pmatrix} \tag{5.66}$$

2. Inflow/outflow boundary conditons: for a scalar equation, if a boundary point is of inflow the freestream state u_∞ is imposed. For systems of equations we encounter the problem that, at every boundary points, we have to decide for which variables the same point is of inflow/outflow, hence for which variables we have to impose the boundary condition. This is determined by the sign of the eigenvalues of the local jacobian $\mathsf{K} = \mathbf{A}(\mathsf{u}^h) \cdot \boldsymbol{n}$. For the l-th eigenvalue

$$\lambda_l > 0 \quad \text{inflow point for } w_l$$

$$\lambda_l \leq 0 \quad \text{outflow point for } w_l$$

The eigenvalues defines the matrices Λ^-, Λ^+. Imposing the boundary values on the linearized residual in characteristic variables

$$\mathbf{f}^{\partial}(\mathsf{w}^h) \cdot \boldsymbol{n} = \Lambda^- \mathsf{w}^h + \Lambda^+ \mathsf{w}_\infty$$

Then transforming back in conserved variables we get the correct boundary flux

$$\mathbf{f}^{\partial}(\mathsf{u}^h) \cdot \boldsymbol{n} = \mathsf{K}^-(\mathsf{u}^h)\mathsf{u}^h + \mathsf{K}^+(\mathsf{u}^h)\mathsf{u}_\infty \tag{5.67}$$

5.7 Numerical Results

We first present a test case to prove that the ALE formulation of Euler equations, implemented as proposed in section (5.5), does not spoil the accuracy property of its Eulerian counterpart. The second test case is a Riemann problem to see if we are able to recover the Eulerian results, in terms of positivity and accuracy, when shock waves are present. For this cases, a comparison with the results obtained on a fixed grid is invoked to prove the effectiveness of the method. The following mapping involving the usual distortion of the grid with sinusoidal law is used. At the final time it reduces to the identity mapping and we can compare results on the same grid

$$\begin{cases} x = X + 0.1 \sin(2\pi X) \sin(2\pi Y) \sin(2\pi\, {}^{t}/_{t_{max}}) \\ y = Y + 0.1 \sin(2\pi X) \sin(2\pi Y) \sin(2\pi\, {}^{t}/_{t_{max}}) \end{cases} \tag{5.68}$$

We believe that, summing up these last two experiments with the scalar ones, the formulation has been tested enough. We can conclude that the ALE reformulation of the RK2-$\mathcal{RD}$ scheme of Ricchiuto and Abgrall, as it is implemented in this thesis, give quite good result. At this point a very simple experiment where moving boundaries are involved is given.

Time step is computed according to

$$\Delta t = \text{CFL} \min_{i \in \mathcal{T}_h} \frac{|S_i|}{\sum_{K \in \mathcal{D}_i} 3\alpha^K}$$

where CFL $= 0.8$.

5.7.1 Advection of a Vortex

The accuracy of the schemes is measured on the advection of a constant density vortex. The test case is the one used in [22]. Initial conditions are now presented. The flow velocity is given by the sum of a freestream velocity and a circumeferential perturbation

$$\boldsymbol{u}_0 = \begin{bmatrix} 6 \\ 0 \end{bmatrix} + \Delta \boldsymbol{u} \tag{5.69}$$

$$\begin{cases} \Delta \boldsymbol{u} = \begin{bmatrix} -y_c \\ x_c \end{bmatrix} \omega & if\ r < 0.25 \\ \Delta \boldsymbol{u} = \boldsymbol{0} & if\ r \geq 0.25 \end{cases} \tag{5.70}$$

with $x_c = x - 0.5$, $y_c = y - 0.5$, $\omega = 15\,(cos4\pi r + 1)$ and $r = \sqrt{x_c^2 + y_c^2}$. Density is choosen constant $\rho_0 = 1.4$, the pressure is a given by

$$p_0 = p_m + \Delta p \tag{5.71}$$

$$\Delta p = \frac{15^2 \rho}{(4\pi)^2} \left(2\cos(4\pi r) + 8\pi r \sin(4\pi r) + \frac{\cos(8\pi r)}{8} + \frac{4\pi r \sin(8\pi r)}{4} + 12\pi^2 r^2 \right) + C \tag{5.72}$$

The constant C is fixed such that the pressure at $r = 0.25$ is the freestream pressure $p_0 = p_m = 100$. The maximum Mach number is $M_0^{max} = 0.8$.

The problem is solved on a square domain $[0, 1] \times [0, 1]$ until a final time $t_{max} = 1/6$. The domain is approximated with 4 unstructured triangulations with element's reference size shown in table. Freestream boundary conditions are imposed at $y = \pm 0, 1$ through (5.67) and periodic boundary one are used at $x = \pm 0, 1$.

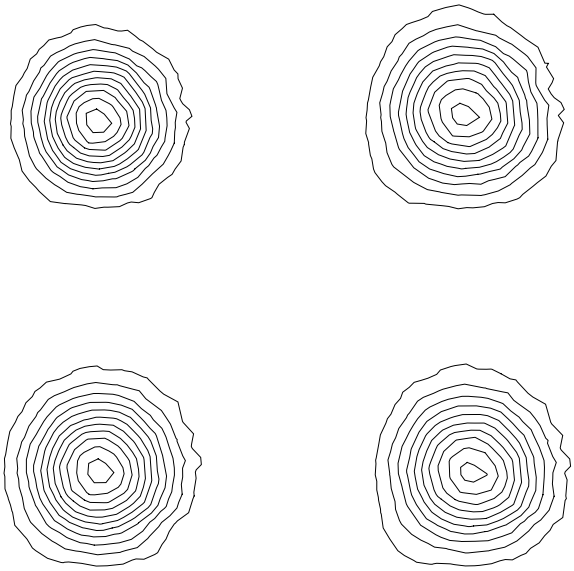

Figure 5.1: Advection of a Vortex for linear schemes (RK-F1-SL). 15 equispaced pressure isolines between level 94 and 101. Left column: Eulerian. Right column: ALE. From the top, in order of rows: LDA, SUPG

In figure (5.3) some results are presented: the convergence curves are qualitatively similar to the ones obtained for the scalar advection of a smooth profile. From figure (5.2) we see that results closely follow the one in Eulerian framework. Second order of accuracy is achieved also in ALE framework for both the lumped and the selective formulation. The lowest convergence rate (equal to 1.5) is observed for the LDA-N scheme but, again, this is due to a switch to the first order N scheme in regions where strong gradients of the vortex are present.

grid	h
1	1/40
2	1/80
3	1/160
4	1/320

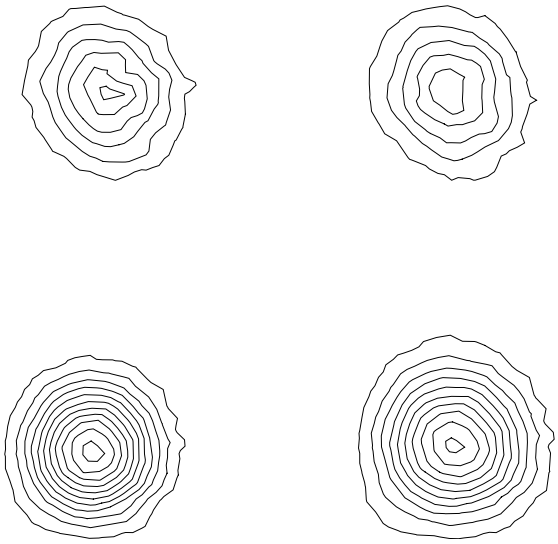

Figure 5.2: Advection of a Vortex for non linear schemes (RK-F1-SL). 15 equispaced pressure isolines between level 94 and 101. Left column: Eulerian. Right column: ALE. From the top, in order of rows: LDA-N, LLxF-SUPG

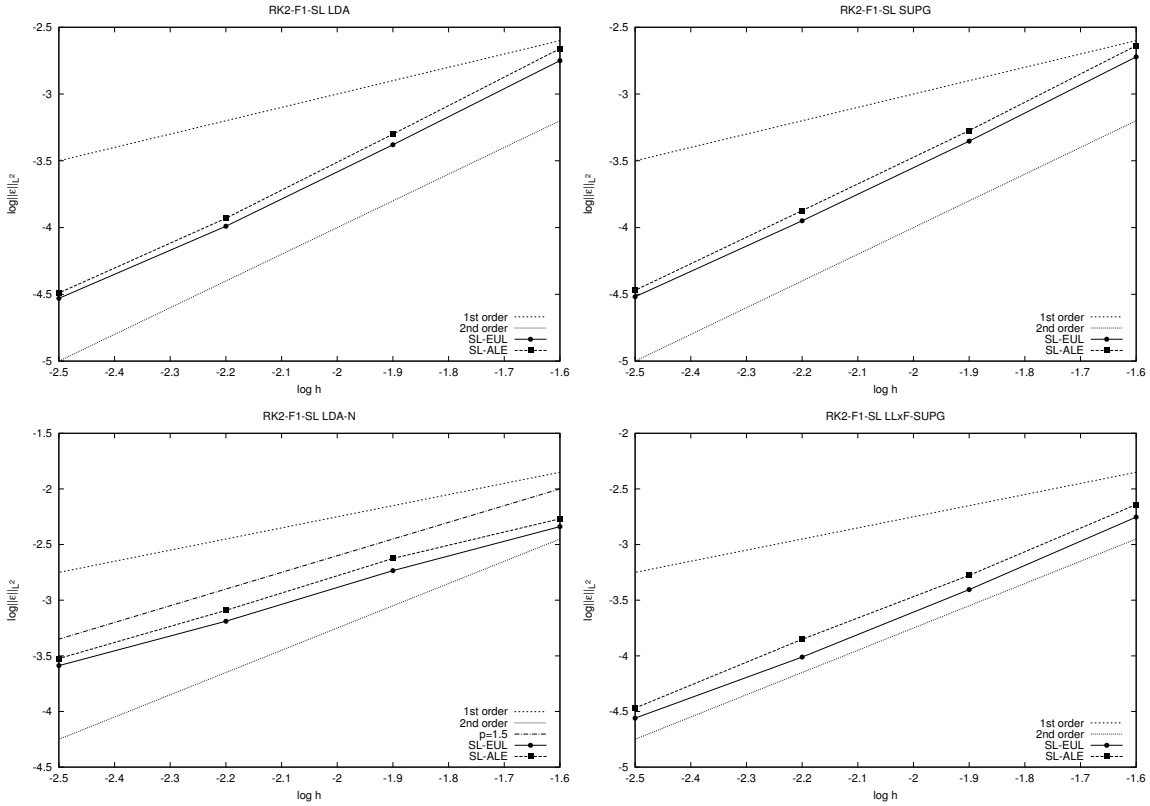

Figure 5.3: Advection of a Vortex. Order of Convergence

5.7.2 A 2D Riemann problem

This test case is contained in [23]. We use it to test the shock-capturing capabilities of the schemes. With reference to the notation of the figure (5.4), the initial solution is given by

$$\begin{pmatrix} \rho & u & v & p \end{pmatrix} = \begin{cases} \begin{pmatrix} 1.5 & 0 & 0 & 1.5 \end{pmatrix} & \text{state a} \\ \begin{pmatrix} 0.1379928 & 1.2060454 & 1.2060454 & 0.0290323 \end{pmatrix} & \text{state b} \\ \begin{pmatrix} 0.5322581 & 1.2060454 & 0 & 0.3 \end{pmatrix} & \text{state c} \\ \begin{pmatrix} 0.5322581 & 0 & 1.2060454 & 0.3 \end{pmatrix} & \text{state d} \end{cases} \tag{5.73}$$

The structure of the solution is very complex. Two normal shocks are interacting with two oblique shocks. This interaction generates two couples of symmetric lambda shocks with the appereance of contact discontinuities emanating from each of the 4 triple points. The amount of fluid that passes through the upper lambda shock structures (hence through two oblique shocks) is then pushed by the pressure gradient between state a and b into a transonic jet against the normal shock. The domain is a box $[0,1] \times [0,1]$ and it is approximated through a structured triangulation with element reference size $h = 1/200$. The final time is $t_{max} = 0.8$.

Only the non-linear schemes are expected to give postive and second order accurate results, hence results in figure are referred only to the LDA-N and LLxF-SUPG schemes. The LDA-N case is shown in figure (5.5)(5.6). The ALE results are overlapped, almost everywhere, with the ones obtained with Eulerian formulation on a fixed grid. As in that case, only when the global lumped formulation is used, we get positive results. With selective lumping, the solution is quite monotone but small oscillations appears near the discontinuities.

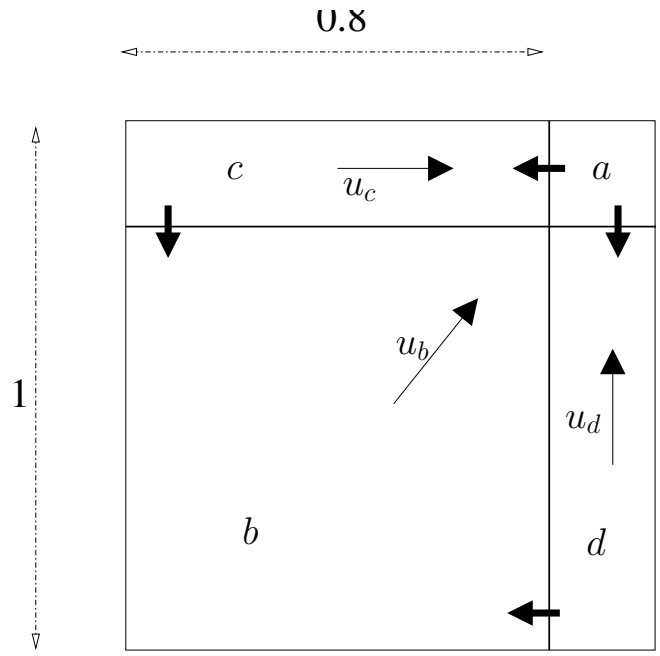

Figure 5.4: 2D Riemann Problem: initial solution

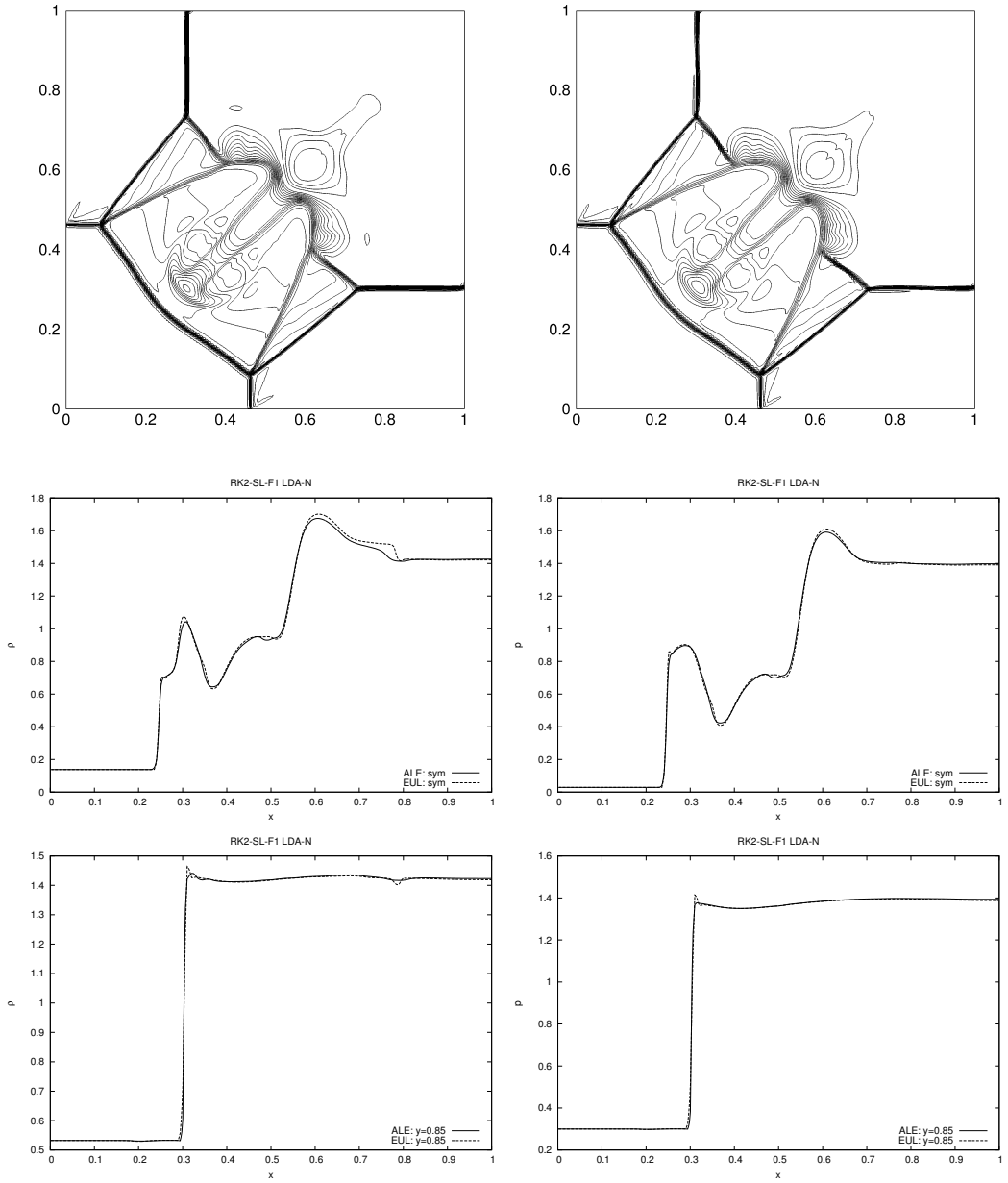

Figure 5.5: 2D Riemann problem computed with LDA-N scheme and RK2-F1-SL formulation. Top: 30 equispaced density isolines between maximum and minimum values of 1.65 and 0.1. Top left: Eulerian formulation. Top right: ALE formulation. Middle: comparison of the solutions along the symmetry line. Bottom: comparison of the solutions at $y = 0.85$

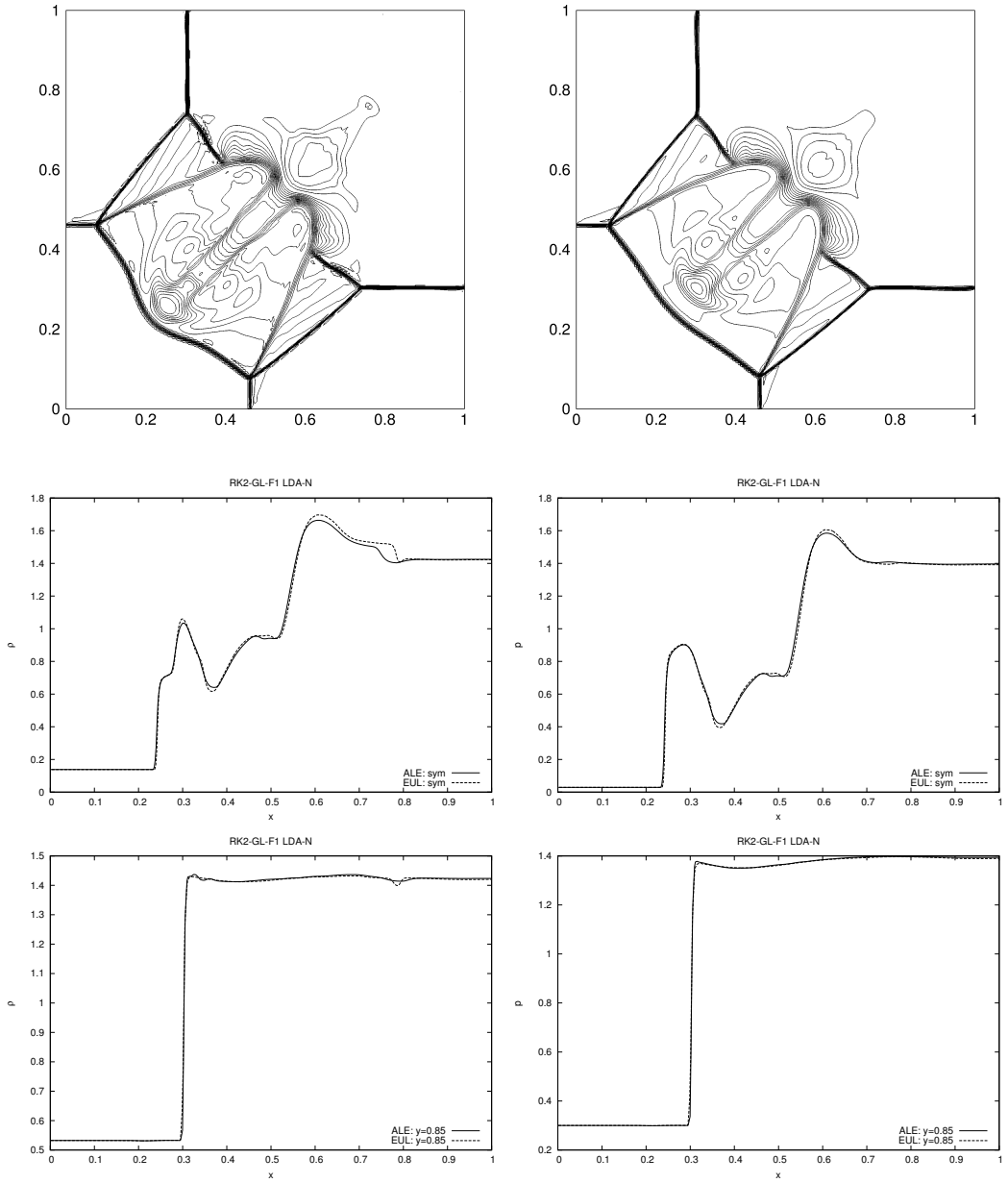

Figure 5.6: 2D Riemann problem computed with LDA-N scheme and RK2-F1-GL formulation. Top left: 35 equispaced density isolines for Eulerian formulation. Top right: 35 equispaced density isolines for ALE formulation. Middle: comparison of the solutions along the symmetry line. Bottom: comparison of the solutions at $y = 0.85$

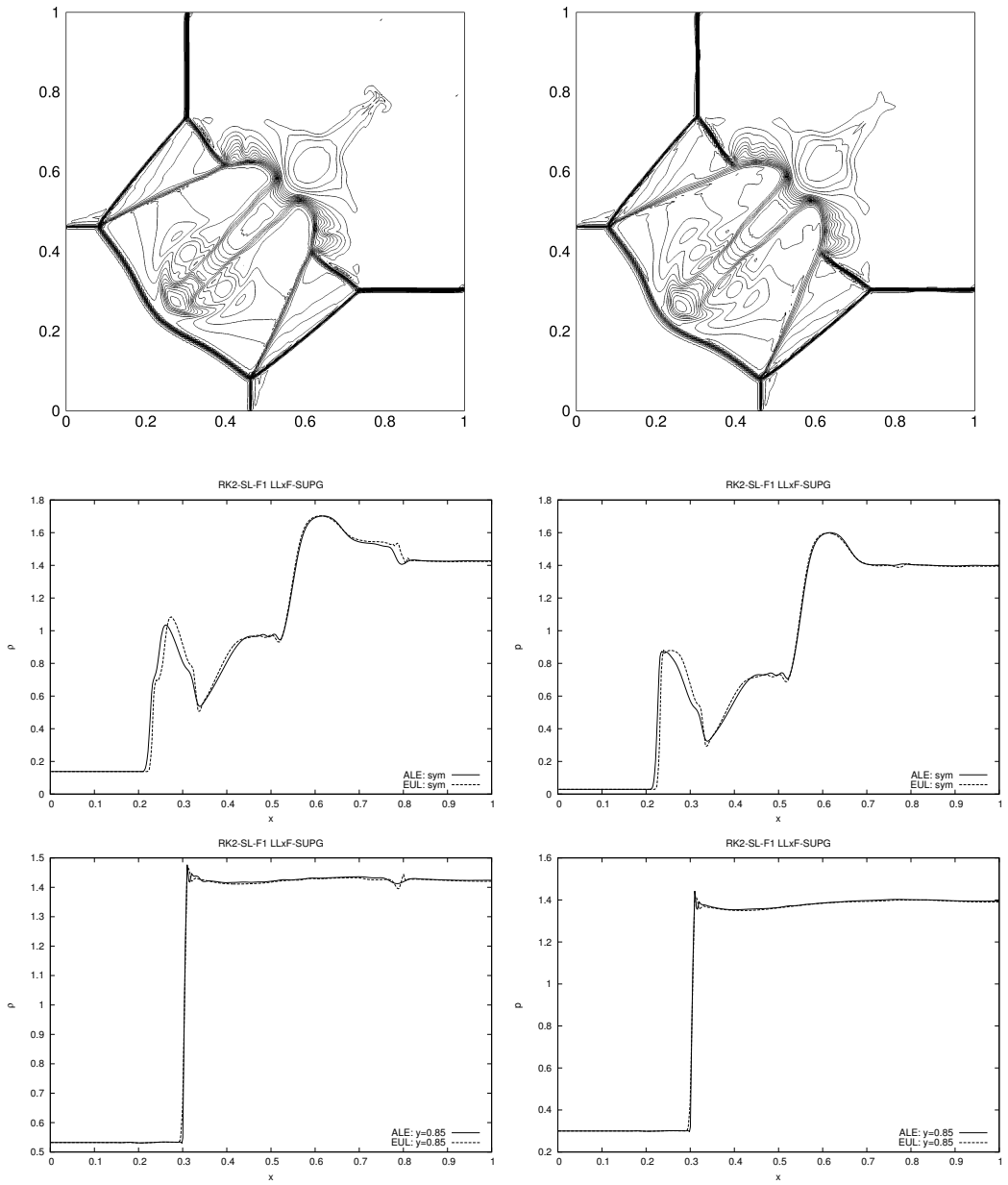

Figure 5.7: 2D Riemann problem computed with LLxF-SUPG scheme and RK2-F1-SL formulation. Top left: 35 equispaced density isolines for Eulerian formulation. Top right: 35 equispaced density isolines for ALE formulation. Middle: comparison of the solutions along the symmetry line. Bottom: comparison of the solutions at $y = 0.85$

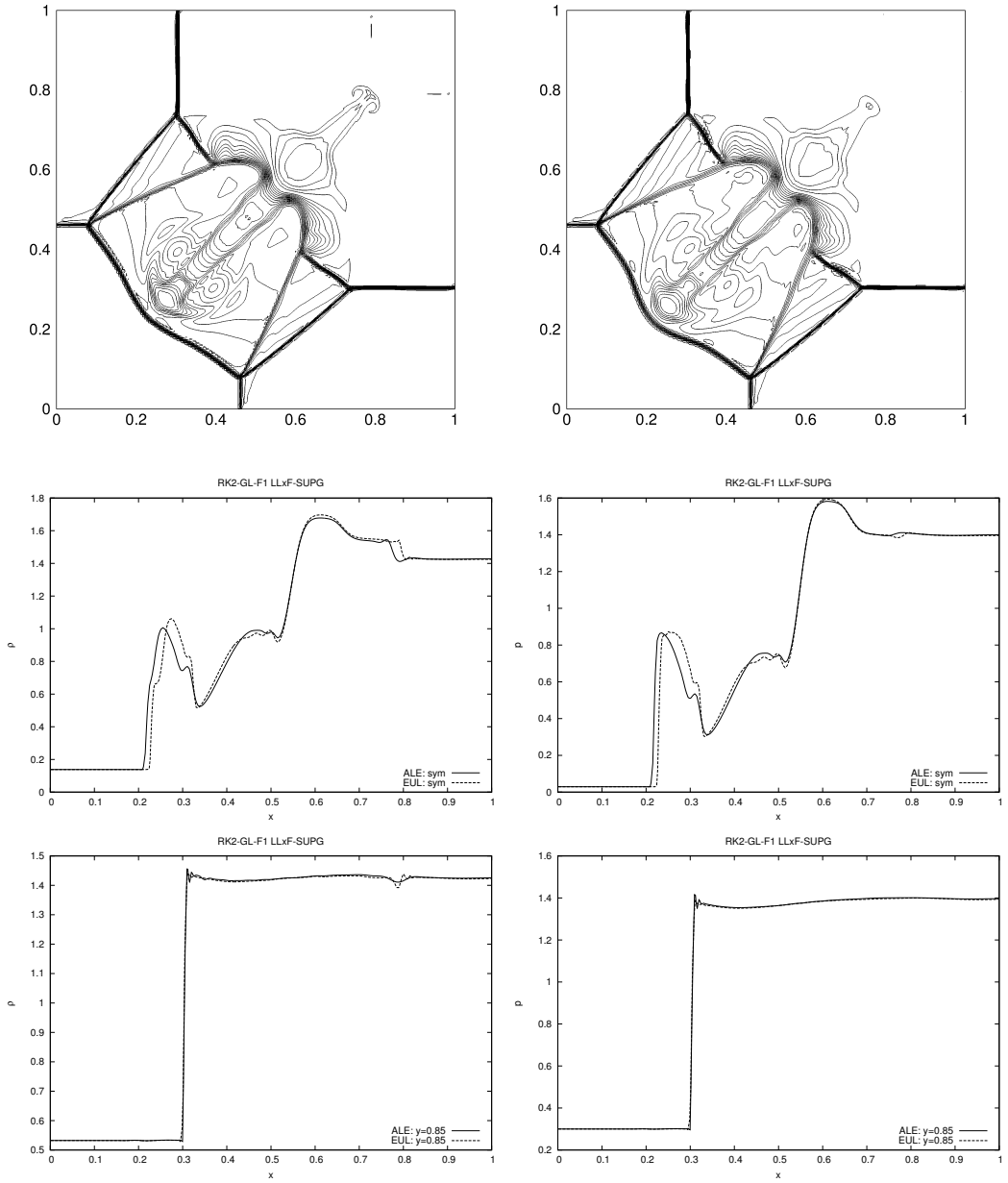

Figure 5.8: 2D Riemann problem computed with LLxF-SUPG scheme and RK2-F1-GL formulation. Top left: 35 equispaced density isolines for Eulerian formulation. Top right: 35 equispaced density isolines for ALE formulation. Middle: comparison of the solutions along the symmetry line. Bottom: comparison of the solutions at $y = 0.85$

5.7.3 A very simple application: wind tunnel with wall deflection

We have seen that, for all the test cases that we have run, the ALE formulation proposed in this thesis works quite good and we are able to recover almost the same result of Eulerian formulation. A very simple application, just to see the use of ALE formulation, is shown, involving moving boundaries. In this case Eulerian formulation cannot work without an interpolation step. The Eulerian formulation with the interpolation step has not been implemented, henco no comparison is given for this case.

We have a 2D channel $[2 \times 1]$ with an hinge on the lower surface placed at $x = 0.25$. This hinge allows a rigid deflection of the lower wall which is governed by the following exponential motion law for angle α defined from the horizontal axis

$$\begin{cases} \alpha(t) = \alpha_{max}\left(1 - e^{-t/\tau}\right) & t \leq t_{switch} \\ \alpha(t) = \alpha_{max} - 2\alpha_{max}\left(1 - e^{-(t-t_{switch})/\tau}\right) & t > t_{switch} \end{cases} \tag{5.74}$$

We choose the following values

$$t_{switch} = 1.25, \quad \tau = 0.05, \quad \alpha_{max} = 20^{\circ}$$

The final time for our simulation is $t_{max} = 2.5$. The domain is approximated with an unstructured triangulation with an element reference size $h = 1/160$. During the simulation the grid is distorted solving a Laplace equation along every abscissa with boundary conditions given by the flap displacement at that abiscissa. In figure (5.9) the mapping for the grid is shown. Since shock waves are expected, we have tested only the non-linear schemes LDA-N and LLxF-SUPG. The formulation choosen is F1-GL. The Mach number at the inlet is $M = 3$.

From the experiments we can observe that, after the transient, the shock structure finds a stable configuration close to the the analitical solution (Mach reflection of the shock at the upper surface) at $t \backsimeq 1.2$. Immediately after the wall deflects an unsteady interaction, between the shock and the expansion wave rising from the corner, is observed. The shock wave, while it is going back, takes an S-shaped configuration. In particular, in the region near the lower wall, the shock seems to be particularly strong becouse of the interaction between the accelerating flow, in expansion after the corner, and the compressed region at the outlet. Finally, at $t \backsimeq 2.5$, the supersonic Prandtl-Mayer expansion is recovered.

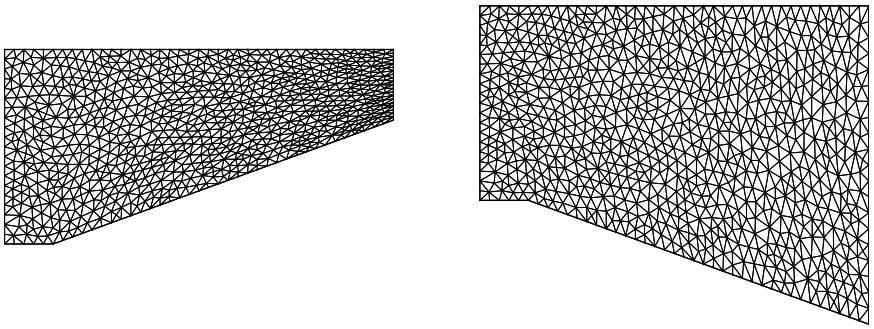

Figure 5.9: Topology for the grid. Left: compression. Right: expansion.

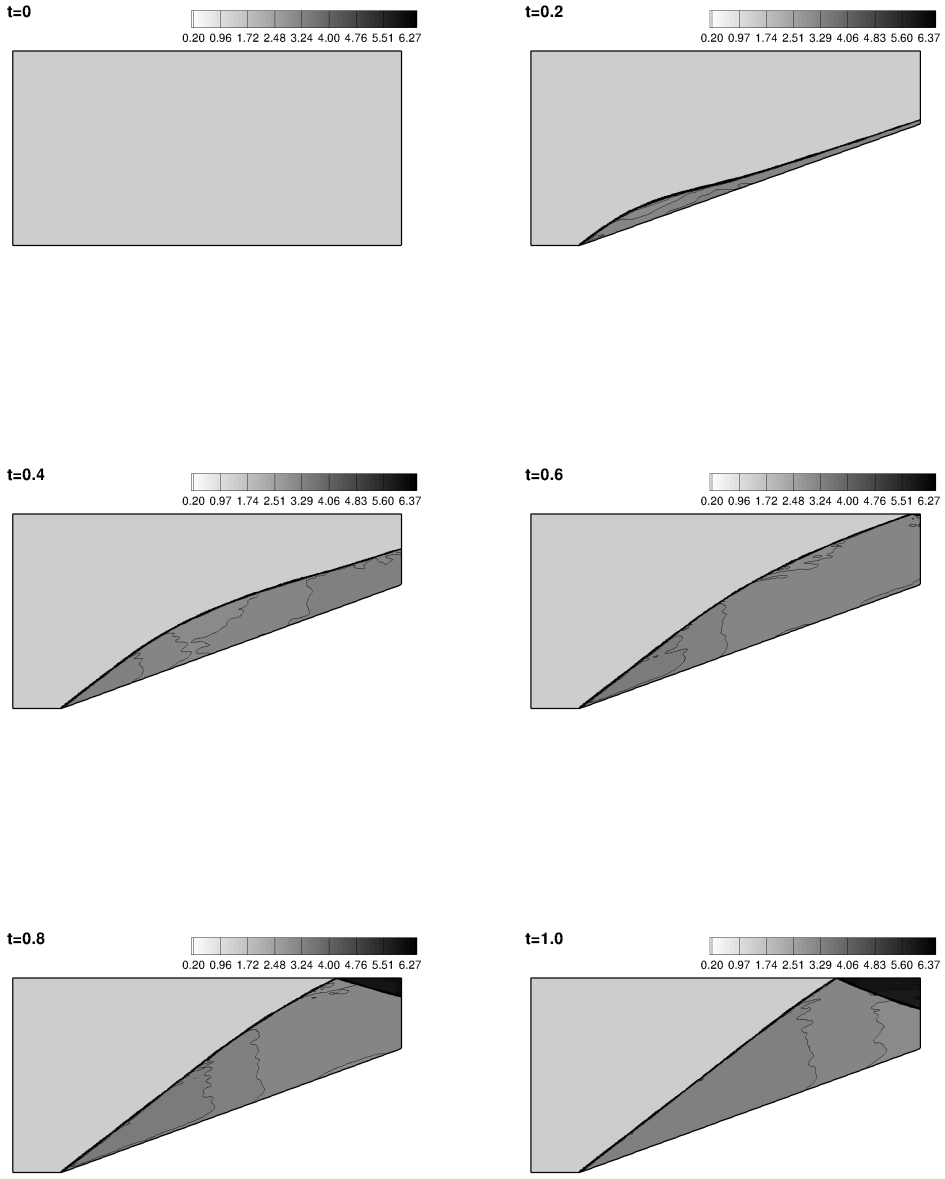

Figure 5.10: Mach 3 wind tunnel with LDA-N RK2-F1-GL: 50 equispaced density isolines between extreme values of $0.2 - 6.5$ at different time instants

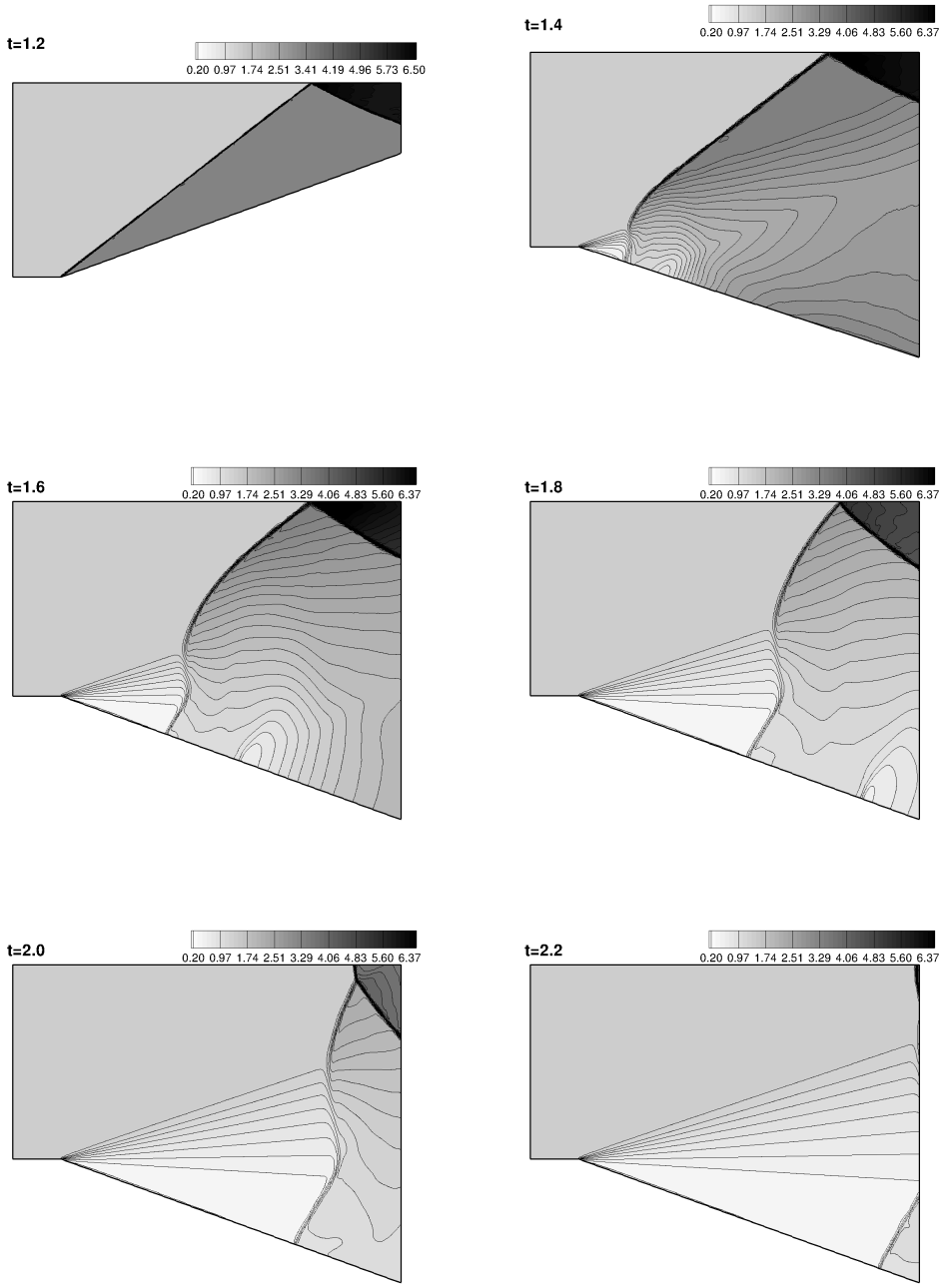

Figure 5.11: Mach 3 wind tunnel with LDA-N RK2-F1-GL: 50 equispaced density isolines between extreme values of $0.2 - 6.5$ at different time instants

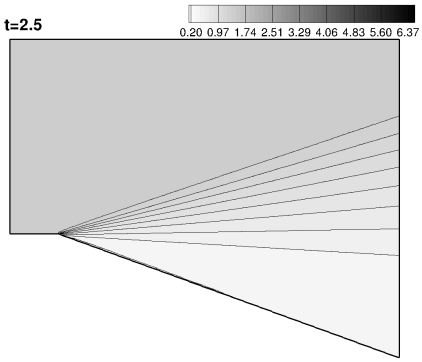

Figure 5.12: Mach 3 wind tunnel with LDA-N RK2-F1-GL: 50 equispaced density isolines between extreme values of $0.2 - 6.5$. Final time

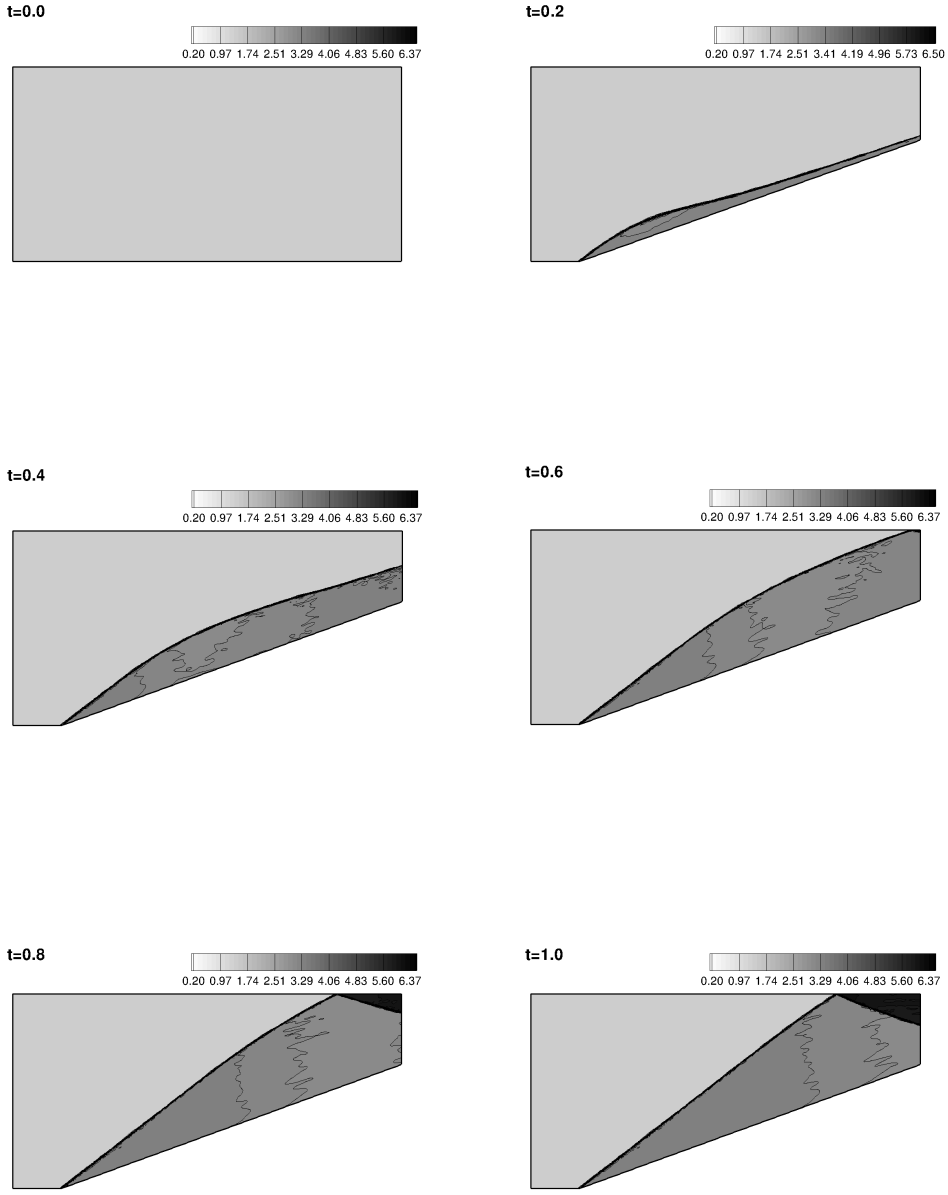

Figure 5.13: Mach 3 wind tunnel with LLxF-SUPG RK2-F1-GL: 50 equispaced density isolines between extreme values of $0.2 - 6.5$ at different time instants

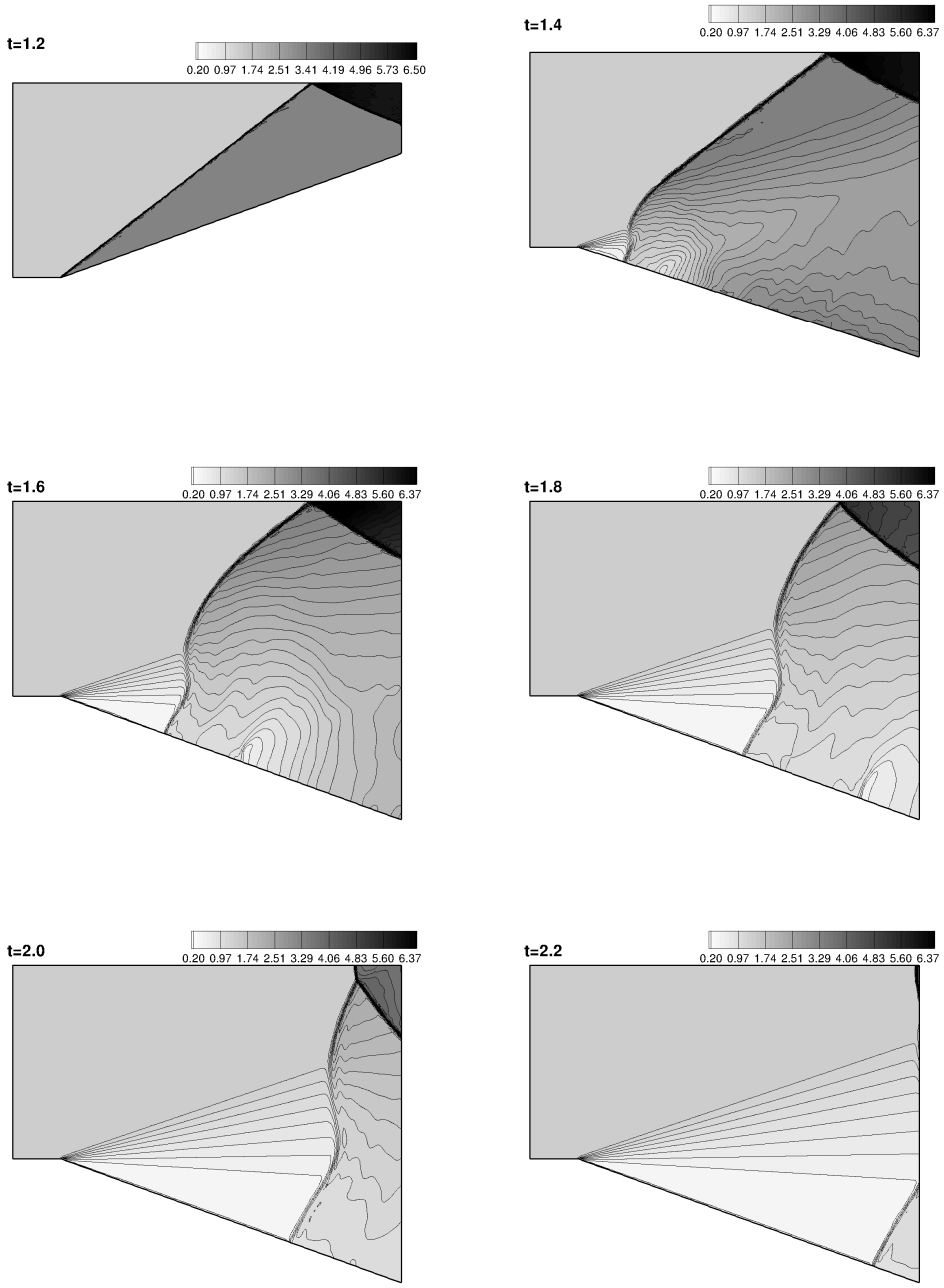

Figure 5.14: Mach 3 wind tunnel with LLxF-SUPG RK2-F1-GL: 50 equispaced density isolines between extreme values of $0.2 - 6.5$ at different time instants

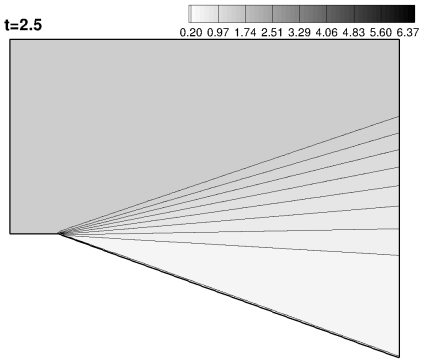

Figure 5.15: Mach 3 wind tunnel with LLxF-SUPG RK2-F1-GL: 50 equispaced density isolines between extreme values of $0.2 - 6.5$. Final time

Conclusion

In this thesis we have developed a novel class of genuinely explicit residual based discretizations for conservation laws on moving meshes. This has been achieved by proposing a new Arbitrary Lagrangian Eulerian (ALE) formulation for the explicit Runge Kutta Residual Distribution ($\mathcal{RD}$) schemes of [9]. In the manuscript we have slowly discussed all the aspects related to this study, from the basics of $\mathcal{RD}$ schemes, to the extensive numerical testing of the new proposed method.

In Chapter 2 many properties of $\mathcal{RD}$ have been introduced. Among other classes of numerical methods used to solve hyperbolic equations, $\mathcal{RD}$ are the least known. This class of schemes has an inherently geometrical interpretation of the concept of upwinding, this leading to less diffusive schemes respect, for example, to first order two dimensional Finite Volume method ($\mathcal{FV}$). The name of the two linear $\mathcal{RD}$ schemes used in computations recall to the reader this property: the first order Narrow scheme (N) and the second order Low Diffusion A scheme (LDA). Other two popular schemes have been reformalized as linear $\mathcal{RD}$ schemes: the first order Lax Friederich scheme (LxF) and the second order Streamline Upwind Petrov Galerkin scheme (SUPG). A stability criteria and an error estimation for $\mathcal{RD}$ in comptact prototype form is also given. Nonlinear schemes are invoked in order to have both positive and second order accurate solvers. Two design philosophy are followed: a strategy borrowed from $\mathcal{FV}$ is the blending of a positive scheme with a second order one through the use of an appropriate limiter, a second approach, developed directly into a $\mathcal{RD}$ contest, consists in limiting in a correct way a first order distribution. In theory a limited schemes is positive and second order accurate, but, even in simple experiments, spurious oscillations spoil the order of accuracy and a stabilization term is necessary to fix the problem. In computations both the approaches have been tested: the blended LDA-N scheme and the Limited Lax Friederich stabilized scheme (LLxFs).

PDEs are discretized in time with the explicit Runge Kutta 2 ($\mathcal{ERK}2$) tested by Ricchiuto and Abgrall on fixed grids, for conservation laws written in Eulerian framework. First conservation laws are discretized in time, a full discretization is obtained with Petrov Galerkin Finite Elements, the analogy between stabilized Finite Elements ($\mathcal{FE}$) and $\mathcal{RD}$ allows to recast the scheme into a $\mathcal{RD}$ contest, finally, through high order mass lumping, consistent mass matrix is replaced by a median dual cell area. Since two Galerkin mass matrices appears, lumping is performed with two different strategies: on both ma-

trices (Global Lumped formulation) or only on the right-end side term (Selective Lumped formulation).

Conservation laws have been finally presented in Arbitrary Lagrangian Eulerian (ALE) formulation together with the constraint of the Geometric Conservation Law (GCL). With a little more algebra it is possible to repeat the same steps listed in the previous paragraph for the new equations: the time discretization, the Petrov Galerkin space discretization, this time with an appropriate stabilization term, and finally the reformulation into $\mathcal{RD}$. The stabilization term, in particular the differential operator inside it, was choosen not only concerning accuracy, stability and conservation properties, but also in order to satisfy the discrete counterpart of the GCL. We ended up with a scheme that results in minor modifications respect to the Eulerian algorithm. The ALE part in the fluxes modifies the definition of the residual but, at the end, this is interpreted as a local modification of the advective field. The median dual cell area is modified by the presence of a term involving the divergence of the grid velocity instead. In all the computation that we did, this term that takes into account of grid distortion, never spoiled the positivity of the median dual cell area.

The method has been studied extensively through theorical investigation and numerical experiments. Numerical results were in good agreement with Eulerian ones. The two advection test cases provided numerical evidence that convergence order is not spoiled when arbitrary grid distortions are involved. On smooth solutions SUPG and LDA are second order accurate, moreover LDA-N and LLxFs are also second order accurate, thus the definition of blending and limiting procedures, even if extended directly from the Eulerian formulation, in such cases works well. The Burger's equation test case and the Riemann problem showed the ability, for LLxFs and LDA-N both in the Global Lumped version, to handle well discontinuities.

Unfortunately in this thesis we did not have the opportunity to cover some aspects that were not clear enough or that, in our opinion, deserve further studies. In the following list we enumerate them briefly

1. A rigorous stability analysis of the scheme in the ALE formulation, in particular for positive schemes (N,LxF), in order to understand some not clear phenomena. With Selective Lumped formulation oscillations appears near shock waves. The Global Lumped formulation closely recalls the compact prototype form but the presence of different terms arising from the ALE approach, makes a further investigation necessary. Numerical simulations seems to confirm that the $\mathcal{ERK}2$-$\mathcal{RD}$ N scheme, in the Global Lumped formulation, is positive.
2. A ridefinition of some parameters such as limiters and blending procedures that have been extended directly from the Eulerian scheme.
3. Even if strong and arbitrary distortions of the grid have been applied, only sinusoidal mappings have been tested. For the last experiment infact, the grid is distorted ac-

cording to a one dimensional Laplace equation but no comparison exists, apart from the analytical solution at the end of the transients. The classical two dimensional pitching airfoil test would have represented an important test.

To conclude we mention possible future developments:

1. Grid adaptation not only to moving boundaries but also through a mechanism of node insertion/removal in order to refine the grid where strong gradients of the solution are expected. A succesfull algorithm has been already implemented by Guardone and Isola in a $\mathcal{FV}$ context by [27]
2. The extension to third order accurate solutions through high order space and time approximations. $\mathcal{RD}$ schemes that converges with order higher then two have been studied extensively for the steady case by Abgrall. The extension of the present work to third order should involve higher order elements and also an high order time discretization such as $\mathcal{RK}3$.

Bibliography

[1] P.L.Roe. Fluctuations and signals, a framework for numerical evolution problems. *Numerical Methods for Fluid Dynamics*, 1982.

[2] P.L.Roe. Linear advection schemes on triangular meshs. *Numerical Methods for Fluid Dynamics*, 1986.

[3] R.H.Ni. A multiple grid scheme for solving the euler equation. *AIAA Journal, 20:1565-1571*, 1981.

[4] J.Rice and R.Schnipke. A monotone streamline upwind method for convection dominated problems. *Computer Methods in Applied Mechanics and Engineering, 48:313-327*, 1985.

[5] A.N.Brooks and T.J.R.Hughes. Streamline upwind petrov-galerkin formulation for convection dominated flows with particular emphasis on the incompressible navier-stokes equations. *Comp. Meth. Mech. Eng., 32:199-259*, 1982.

[6] M.Ricchiuto. *Construction and analysis of compact residual discretizations for conservation laws on unstructured meshes*. PhD thesis, Von Karman Institute, 2005.

[7] R.Struijs. *A multidimensional upwind discretization method for the Euler equations on unstructured grids*. PhD thesis, Von Karman Institute, 1994.

[8] R.Abgrall. Essentially non-oscillatory residual distribution schemes for hyperbolic problems. *Journal of Computational Physics*, 2005.

[9] R.Abgrall and M.Ricchiuto. Explicit runge-kutta residual distribution schemes for time dependent problem: Second order case. *Journal of Computational Physics*, 2010.

[10] A.Ferrante and H.Deconinck. Solution if the unsteady euler equation using residual distribution and flux corrected transport. Technical report, Von Karman Institute, 1997.

[11] R.Abgrall and M.Mezine. Construction of second order accurate monotone and stable residual distribution for unsteady flow problems. *Journal of Computational Physics*, 2003.

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

[13] J.Donea. Computational methods for transient analysis. chapter 10, Arbitrary Lagrangian Eulerian finite element methods. Elsevier Science Publisher, Amsterdam, 1983.

[14] C.Grandmont C.Farhat, P.Geuzaine. The discrete geometric conservation law and the nonlinaer stability of ale schemes for the solution of flow problems on moving grids. *Journal of Computational Physics*, 2000.

[15] C.Michler and H.Deconinck. An arbitrary lagrangian eulerian formulation for residual distribution schemes on moving grids. *Computers and Fluids*, 2001.

[16] J.Dobes. *Numerical Algorithms for the Computation of Steady and Unsteady Compressible Flow over Moving Geometries - Application to Fluid-Structure Interaction*. PhD thesis, Von Karman Institute, 2007.

[17] M.Lesoinne and C.Farhat. Geometric conservation laws for flow problems with moving boundaries and deformable meshes, and their impact on aeroelastic computation. *Computer methods in applied mechanics and engineering*, 1996.

[18] R.Abgrall and M.Mezine. Construction of second order accurate monotone and stable residual distribution for steady flow problems. *Journal of Computational Physics*, 2003.

[19] P.L.Roe. Approximate riemann solvers, parameter vectors and difference schemes. *Journal of Computational Physics*, 1981.

[20] P.L.Roe H.Deconinck and R.Struijs. A multidimensional generalization of roe's flux difference splitter for the euler equations. *Computers and Fluids*, 1993.

[21] R.Abgrall. A review of residual distribution schemes for hyperbolic and parabolic problems: the july 2010 state of the art. Technical report, Team Bacchus, INRIA and University of Bordeaux, 2010.

[22] J.Dobes and H.Deconinck. Second order blended multidimensional upwind residual distribution scheme for steady and unsteady computations. *Journal of Computational and Applied Mathematics*, 2005.

[23] H.Deconinck M.Ricchiuto, A.Cisk. Residual distribution for general time-dependent conservation laws. *Journal of Computational Physics*, 2005.

[24] R.J.LeVeque. *Numerical Methods for Conservation Laws*. Birkhauser.

[25] R.J.LeVeque. *Finite Volume Methods for Hyperbolic problems*. Cambridge University Press.

[26] A.Quarteroni. *Modellistica Numerica per Problemi Differenziali*. Springer-Verlag.

[27] D.Isola. *An Interpolation Free Two-Dimensional Conservative ALE scheme over Adaptive Unstructured Grids for Rotorcraf Aerodynamics*. PhD thesis, Politecnico di Milano, Dipartimento di Ingegneria Aerospaziale, 2012.

[28] A.Guardone L.Quartapelle. High resolution unstructured finite volume methods for conservation laws. Technical report, Politecnico di Milano, Dipartimento di Ingegneria Aerospaziale.