HAL
open science

Couplage fluide-structure d'ordre (très) élevé pour des schémas volumes finis 2D Lagrange-projection

Gautier Dakin

To cite this version:

Gautier Dakin. Couplage fluide-structure d'ordre (très) élevé pour des schémas volumes finis 2D Lagrange-projection. General Mathematics [math.GM]. Université Pierre et Marie Curie - Paris VI, 2017. English. NNT : 2017PA066404 . tel-01701774v2

HAL Id: tel-01701774

https://hal.science/tel-01701774v2

Submitted on 5 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Pierre et Marie Curie

École doctorale ED386

Laboratoire Jacques-Louis Lions

Couplage fluide-structure d'ordre (très) élevé pour des schémas volumes finis 2D Lagrange-projection

Présentée pour l'obtention du grade de DOCTEUR

DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

par

Gautier DAKIN

Thèse de doctorat de Mathématiques appliquées

Dirigée par **Bruno DESPRÉS**

Présentée et soutenue publiquement le 9 novembre 2017

Devant un jury composé de :

Rapporteur	**M. Jean-Claude Latché**	(IRSN)
Rapporteur	**Prof. Stéphane Louis Clain**	(Universidade do Minho)
Examinateur	**Prof. Edwige Godlewski**	(UPMC)
Examinateur	**Prof. Frédéric Coquel**	(École polytechnique)
Directeur de thèse	**Prof. Bruno Després**	(UPMC)
Co-encadrant de thèse	**Dr. Stéphane Jaouen**	(CEA)
Président du jury	**Prof. Christophe Chalons**	(UVSQ)

Thèse effectuée au sein du **Laboratoire Jacques-Louis Lions**

de l'Université Pierre et Marie Curie

4, place Jussieu, 75005 Paris, France

ainsi qu'au sein du **CEA, DAM, DIF**

F-91297 Arpajon, France

Résumé

Ce travail est consacré à l'étude numérique de l'interaction entre un fluide compressible et une structure indéformable, en adaptant une famille récente de schémas d'ordre très élevé à la prise en compte de conditions aux bords particulières entre le fluide et la structure. Plus précisément, on évalue l'apport de schémas d'ordre strictement supérieur à 3 par rapport à des stratégies plus classiques dans la littérature restreintes aux ordres 1 et 2. Un résultat important est qu'il est possible de réaliser le couplage à tout ordre et qu'il existe des configurations pour lesquelles on observe un gain important pour les ordres élevés. Une revue bibliographique est faite rappelant les résultats théoriques concernant les systèmes hyperboliques et décrivant les méthodes utilisées dans la littérature pour la simulation de la dynamique des gaz et la prise en compte des conditions aux bords. Un schéma sur grilles cartésiennes décalées et d'ordre très élevé est proposé pour la résolution des équations d'Euler en 1D et 2D. Ce schéma est basé sur le formalisme Lagrange-projection et bien que formulé en énergie interne assure conservation et consistance faible grâce à un correctif en énergie interne. Parallèlement, l'étude pour les systèmes hyperboliques linéaires de discrétisation à l'ordre très élevé des conditions aux bords est faite. Elle met en évidence la nécessité pour l'ordre élevé de s'intéresser à la stabilité des schémas ainsi obtenus. À partir de ces travaux, la prise en compte de conditions aux bords en vitesse normale imposée est réalisée pour les équations d'Euler en 1D et 2D. Enfin, une procédure de couplage entre fluide compressible et structure indéformable est proposée.

Mots-clé :

Équations d'Euler, volumes finis, Lagrange-projection, grilles décalées, ordre très élevé, conditions aux bords, couplage fluide-structure, stabilité.

Remerciements

Je souhaiterais en tout premier lieu remercier Bruno Després pour son encadrement, son accompagnement et sa présence pendant ces trois années de thèse au laboratoire Jacques-Louis Lions. Je te suis gré, Bruno, d'avoir pris le temps et la patience de venir régulièrement au CEA et d'avoir pu ainsi toujours être disponible. Je tiens aussi à montrer ma reconnaissance à Stéphane Jaouen qui a encadré cette thèse au sein du CEA, DAM, DIF et qui a eu l'immense courage de corriger mes proses rarement très lyriques.

Je suis extrêmement reconnaissant à Stéphane Clain et Jean-Claude Latché pour avoir accepté la lourde tâche de relire ce manuscrit, qui apparemment fut volumineux, et d'avoir, sans la moindre coertion, accepté d'être rapporteurs de cette thèse. Je tenais également à vous remercier pour vos retours qui permettront sans doute d'améliorer la qualité du manuscrit.

Un immense merci également à Christophe Chalons pour avoir accepté le rôle de président du jury, ainsi qu'à Frédéric Coquel et Edwige Godlewski pour en avoir été les examinateurs. Quel plaisir ce fut que de vous revoir. Merci encore.

Je tiens à adresser un immense merci à Hervé Jourdren pour m'avoir conduit pendant mon stage au CEA à poursuivr en thèse. Je ne l'ai pas regretté un seul instant et je dois dire que ton enthousiasme légendaire n'est pas étranger au succès de cette dernière. C'est un réel plaisir que de continuer à travailler avec toi pour la suite.

Maintenant, passons au Teratec et à la véritable tribu qui y vit et qui parfois y survit péniblement en l'absence de l'or noir, connu également sous le nom de café. Je tiens à préciser avant toute forme de récrimination que l'ordre est aléatoire et que la bienséance, comme vous le savez, ne s'applique pas avec moi. Voilà, vous êtes prévenus.

Mais quand même. En premier lieu, les chefs. Alors je dis merci à Thao, Denis et plus récemment Yves d'avoir rendu le Teratec comme un lieu accueillant et chaleureux dans lequel ma thèse s'est déroulé. Sans être l'antre du doctorant, c'est vraiment un lieu à part, vraiment idéal. Donc merci pour ça.

Je tiens à remercier les anciens, ceux avec qui j'ai commencé mon aventure au CEA, ceux avec qui je l'ai continué, et ceux surtout que j'ai laissé tout seuls pour la terminer. Donc à vous deux, Rémi et Hoby je tiens à dire merci pour votre compagnie, votre gentillesse (si si!), votre aide et surtout vos petites blagues qui permettent de rendre une journée d'autant plus lumineuse que la blague est celle d'Hoby (et là, on s'est fait..... *carottes*).

Je tiens à remercier les deux zigotos qui étaient là avant moi et sans y être, qui ont soutenus avant moi et avec qui je vais avoir le plaisir de pouvoir continuer à faire des pauses cafés intempestives. Merci à Xavier (Mimoon pour les intimes) et à Sébastien (pour des raisons de politesse et dans le but de ne pas froisser un public sensible, je m'abstiendrai d'indiquer toute forme de surnom).

Merci à vous deux pour les fous rires, pour les soutiens. Surtout pour vous avoir vu galérer avec la rédaction et tout le tintouan et pour m'être rendu compte qu'effectivement, non, ce ne serait pas facile. Loin de là.

Si je continue à faire un paragraphe entier pour remercier à chaque fois uniquement deux personnes, le manuscrit risque de refaire une crise d'embonpoint. Je vais donc opter pour la technique du tir groupé. Merci à Bhugo, Thugo et Arthur pour leurs mines fraîches matinales et pour leurs entrains permanents. Merci à Théo, Éloïse pour m'avoir supporté pendant la rédaction, pour avoir vu sans paniquer pour votre première fois un doctorant en pleine rédaction et ce, sans prendre réellement peur et sans fuir. Bravo à vous. Merci à Guillaume, le seul numéricien qu'il me restait sous la main pour venir embêter quand j'ai des questions sur de l'algèbre linéaire.

Et maintenant les stagiaires. La plaie. Mais sans eux, on rirait nettement moins. Alors merci pour vos phrases pleines d'innocence ou pour votre humour parfois plus que décapant (ça change des carottes...). Je remercie Alexandre, Adrien, Kiki, Quentin, Ludovic, Ewan (sans "r"), Florent, Maxime, Sébastien (apprenti... stagiaire... aucune différence). Et en dernier, merci à Clément. Le premier stagiaire que j'ai eu la chance de pouvoir encadrer. Je te dois te dire que ce fut, je dirais même que c'est un réel plaisir que de travailler avec toi.

Je souhaite ensuite remercier ma famille et plus particulièrement mon père et ma mère. Et ouais, vous êtes vieux. Il y a 26 ans, vous aviez une sorte de petite monstre joufflu et chevelu dans les bras. Et maintenant... il a 26 ans de plus, il est toujours relativement joufflu et chevelu, mais bon courage pour le porter dans vos bras. Merci à vous d'avoir fait en partie ce que je suis aujourd'hui. Les éléments de base étaient correctement manufacturés, merci.

Enfin, le meilleur pour la fin. Je dirais même la meilleure pour la fin. Un immense, incommensurablement grand merci à Laura. Tu as été là tout le long, tu as été là pendant les moments d'euphorie et tu as été là pendant les moments "Mais ça marche pas, ça marche pas, ça marche paaaaaaaaaaaas". Je te remercierai jamais assez de tout ce que tu m'as apporté, et que tu m'apportes encore. Tu es unique en ton genre, et je suis bien heureux et bien chanceux de t'avoir avec moi, et un peu pour moi aussi. Merci Laura.

High-order fluid-structure coupling with conservative Lagrange-remap Finite Volume schemes on Cartesian grids

Abstract

This work is devoted to the construction of stable and high-order numerical methods in order to simulate fluid - rigid body interactions. In this manuscript, a bibliographic overview is done, which highlights theoretical results about hyperbolic system of conservation laws, as well as the methods available in the literature for the hydrodynamics simulation and the numerical boundary treatment. A high-order accurate scheme is proposed on staggered Cartesian grids to approximate the solution of Euler equations in 1D and 2D. The scheme relies on Lagrange-remap formalism, and although formulated in internal energy, ensures both conservation and weak consistency thanks to an internal energy corrector. In the same time, the study of high-order numerical boundary treatment for linear hyperbolic system is done. It highlights the necessity to focus especially on the linear stability of the effective scheme. Starting from the linear results, the numerical boundary treatment with imposed normal velocity is done for Euler equations in 1D and 2D. Last, the coupling between a compressible fluid and a rigid body is realized, using the designed procedure for numerical boundary treatment.

Keywords:

Euler equations, finite volume, Lagrange-remap, staggered grids, high-order accuracy, numerical boundary treatment, fluid-structure coupling, stability.

Contents

Introduction **1**

I Hyperbolic systems of conservation laws and fluid-structure interaction **9**

I-1 Hyperbolic systems of conservation laws and their numerical approximations . . 10

I-1.1 Hyperbolic system of conservation laws in one dimension 10

I-1.1.1 Smooth solutions of conservation laws 11

I-1.1.2 Weak solutions of conservation laws 12

I-1.1.3 Entropic solutions of conservation laws 13

I-1.1.4 The initial boundary value problem 15

I-1.2 Numerical methods for conservation laws and their properties 17

I-1.2.1 Space discretization for conservation laws 17

I-1.2.2 Convergence and consistency of numerical schemes 20

I-1.2.3 Linear stability analysis of numerical schemes 22

I-1.2.4 Convergence toward a weak solution 26

I-1.2.5 Convergence toward the entropic solution for scalar conservation laws 27

I-2 Numerical methods for compressible hydrodynamics 28

I-2.1 Euler and Lagrange equations for compressible hydrodynamics 28

I-2.1.1 Euler and Lagrange systems in 1D 28

I-2.1.2 Entropic relations for the 1D Lagrange system 30

I-2.1.3 General Lagrangian formulation for multi-dimensional problem 31

I-2.2 Lagrangian and ALE methods for compressible hydrodynamics 32

I-2.2.1 Natural derivation of staggered grids for hydrodynamics 32

I-2.2.2 Internal energy formulated numerical schemes 33

I-2.2.3 Total energy Lagrangian methods for compressible hydrodynamics 35

I-2.2.4 ALE formalism for compressible hydrodynamics 35
I-2.3 High-order direct Eulerian and Lagrange-Remap numerical schemes . . . 36
I-2.3.1 High-order space interpolation on Cartesian grids and spurious oscillations 36
I-2.3.2 High-order integration in time 38
I-2.4 Artificial viscosities and hyperviscosities 39
I-2.4.1 Internal energy weak formulation 39
I-2.4.2 Standard expressions of viscosities 40
I-2.4.3 Hyperviscosities 41
I-3 Numerical methods for fluid-structure interaction 41
I-3.1 Time coupling method for fluid-structure interaction 42
I-3.1.1 Loose coupling 43
I-3.1.2 Strong coupling 44
I-3.1.3 Semi-strong coupling 44
I-3.2 Space coupling method for fluid-structure interaction 45
I-3.2.1 Mixed cells methods 45
I-3.2.2 Body-fitted methods 45
I-3.2.3 Fictitious domain methods 47

II High order 2D finite volume conservative Lagrange-Remap schemes for compressible hydrodynamics on staggered Cartesian grids 55
II-1 Structure of schemes on Arakawa C-type grids 57
II-1.1 Example of the BBC scheme 57
II-1.2 Discretized variables on Arakawa C-type grid 59
II-1.3 Definition of average and pointwise values 60
II-2 High order 1D Lagrange-Remap schemes on staggered Cartesian grids 60
II-2.1 Formulation of Runge–Kutta based Lagrangian finite volume schemes . 61
II-2.1.1 Semi-discrete formulation of the Lagrangian finite volume schemes 61
II-2.1.2 High-order in spatial reconstruction of pointwise values from averages ones and *vice versa* and of space derivatives 62
II-2.1.3 Runge–Kutta based time discretization 62
II-2.1.4 Properties of the staggered schemes (II.13)-(II.14) 64
II-2.2 A new local internal energy corrector 72
II-2.2.1 Internal energy corrector 73

II-2.2.2 Properties of the internal energy corrector 74
II-2.3 The remapping stage 81
II-2.3.1 Lagrange polynomials based conservative projection 82
II-2.3.2 Properties of the remap step 83
II-2.4 Numerical validation of the 1D conservative Lagrange-Remap schemes on staggered Cartesian grids 84
II-2.4.1 Cook–Cabot breaking wave test-case [28] 84
II-2.4.2 Non-perfect gas breaking wave test-case 85
II-2.4.3 Acoustic propagation test-case 86
II-2.4.4 Sod test-case [146] 87
II-2.4.5 Noh test-case [127] 89
II-2.4.6 Shu-Osher test-case [144] 89
II-2.4.7 Interacting blast-waves test-case [171] 90
II-3 Extension to 2D Lagrange-remap schemes on staggered Cartesian grids 91
II-3.1 Derivation of the subsystems using the operator splitting technique . . . 92
II-3.2 Modifications of the 1D schemes for the 2D finite volume case 93
II-3.2.1 nD distribution of variables on the modified Arakawa C-type grids 93
II-3.2.2 Derivation of a procedure to apply the 1D schemes in one direction using the 2D finite volume formalism 93
II-3.2.3 Properties of the 2D schemes 95
II-3.3 Numerical validation of the 2D conservative Lagrange-Remap schemes on staggered Cartesian grids 96
II-3.3.1 Isentropic vortex advection [174] 96
II-3.3.2 Vortex-pairing test-case [166] 97
II-3.3.3 Five states Riemann problems [139, 104, 108] 98
II-3.3.4 Sedov test-case [140] 105
II-3.3.5 Noh test-case [127] 106
II-3.3.6 Attenborough test-case [8] 107
II-4 Extension to the 2D compressible Navier–Stokes equations with gravity 110
II-4.1 Distribution of viscous terms on the modified Arakawa grid 110
II-4.1.1 Space distribution and discretization of the viscosity and gravity terms in 1D 111

II-4.1.2 Space distribution and discretization of the viscosity and gravity terms in 2D . . . 112
II-4.2 2D viscous staggered Lagrange-Remap schemes with gravity force 113
II-4.2.1 1D staggered Lagrange-Remap scheme to the compressible Navier-Stokes equations . . . 114
II-4.2.2 2D Extension of the 1D staggered Lagrange-remap schemes . . 116
II-4.2.3 Gravity source terms integration . . . 116
II-4.3 Numerical validation of the 2D staggered Lagrange-Remap schemes . . . 117
II-4.3.1 1D atmosphere at rest [92] . . . 117
II-4.3.2 Taylor–Green vortex [160] . . . 118
II-4.3.3 Rayleigh–Taylor instability [151, 159, 108] . . . 119

III Stable high-order methods for linear hyperbolic systems with arbitrary boundary conditions 123

III-1 Inverse Lax–Wendroff procedure for linear hyperbolic systems . . . 125
III-1.1 Derivation of high-order reconstruction operators for the advection problem 126
III-1.1.1 Derivation of high-order reconstruction operators for the finite volume approximation . . . 127
III-1.1.2 Experimental order of accuracy of the procedure . . . 130
III-1.2 Derivation of high-order reconstruction operators for the wave equations 131
III-1.2.1 Runge–Kutta based staggered schemes for the wave equations 132
III-1.2.2 Reconstruction operators for the wave equations with boundary conditions on velocity . . . 136
III-1.2.3 Reconstruction operators for the wave equations with mixed boundary conditions on both velocity and pressure . . . 138
III-1.2.4 Experimental order of accuracy for a wave problem . . . 141
III-1.3 High-order reconstruction operator for general linear system . . . 143
III-2 Stability of the inverse Lax–Wendroff procedure . . . 144
III-2.1 GKS stability for IBVP using second order reconstruction for the Lax–Wendroff scheme . . . 145
III-2.2 Reduced stability for IBVP discretization . . . 146
III-2.2.1 Analytic reduced stability of the Beam–Warming scheme . . . 147
III-2.2.2 Numerical reduced stability results for the high-order Strang projection schemes . . . 148
III-2.2.3 Numerical reduced stability results for the Runge–Kutta based staggered scheme for the wave equations . . . 149

IV Discretization of boundary conditions for compressible hydrodynamics 155

IV-1 ILW procedure for the 1D Lagrangian system 157

IV-1.1 An instructive second-order boundary treatment 158

IV-1.1.1 First method: the spatially isentropic flow hypothesis 159

IV-1.1.2 Second method: the larger stencil reconstruction 162

IV-1.2 General procedure, and characterization of the solution for the system at the boundary 164

IV-1.2.1 Well-posedness at the boundary for spatially isentropic flow hypothesis 165

IV-1.2.2 Well-posedness at the boundary for enlarged stencil 166

IV-1.3 Stabilization procedure for shocks and very high-order reconstruction . . 167

IV-1.3.1 MOOD procedure 167

IV-1.3.2 Least-square methods for very high-order methods 167

IV-1.4 1D validation and comparisons 168

IV-1.4.1 Kidder isentropic compression test-case [95] 169

IV-1.4.2 Harmonic piston test-case 169

IV-1.4.3 Sod piston test-case [146] 171

IV-2 Extension of the ILW procedure to the 2D Euler system 171

IV-2.1 Formulation of the ILW procedure using directionnal splitting 173

IV-2.1.1 Dimensional splitting technique 174

IV-2.1.2 Methodology for a given sweep 175

IV-2.2 2D numerical validation 177

IV-2.2.1 2D isentropic vortex test-case [174] 177

IV-2.2.2 Acoustic diffraction of a plane wave around a cylinder [15] . . 178

IV-2.2.3 Reflected shock wave 179

IV-2.2.4 Double Mach Reflection [171] 180

IV-2.2.5 Mach shock on a cylinder – Whitham test-case [23] 180

IV-2.2.6 Mach shock on a prism – Schardin test-case [23] 180

IV-2.2.7 Mach shock on a NACA0018 profile [88] 181

V Extension to fluid-rigid body interaction 191

V-1 Rigid body motion and dynamics 192

V-1.1 Description of a rigid body 193

V-1.1.1 Invariant of rigid body motion 193

V-1.1.2 Definition of physical quantities 194
V-1.2 Immersed rigid body dynamics 194
V-2 High-order Lagrangian schemes for rigid body dynamics 195
V-2.1 High-order schemes for rigid body dynamics in 1D 195
V-2.1.1 Runge–Kutta based approach 195
V-2.1.2 Cauchy–Kovalevskaya based approach 196
V-2.2 High-order schemes for rigid body dynamics in 2D 196
V-2.2.1 Rigid body space discretization 196
V-2.2.2 Irrotational rigid body semi-discrete scheme 199
V-2.2.3 General rigid body semi-discrete scheme 200
V-2.2.4 Runge–Kutta based approach 201
V-2.2.5 Cauchy–Kovalevskaya based approach 202
V-3 Fluid - Rigid body coupling 203
V-3.1 Description of the algorithm 203
V-3.2 Numerical results 204
V-3.2.1 Pressure motion driven piston in 1D [120] 204
V-3.2.2 Lift-Off of a cylinder [6, 88, 120] 205
V-3.2.3 Lift-Off of an ellipse 207
V-3.2.4 Lift-Off of a rhombus 207

Conclusions and perspectives **211**

A Butcher tables and weights for directional splitting methods **217**
A.1 Butcher table for usual Runge–Kutta sequences 218
A.2 Directional splitting weights sequences 220

References **223**

List of Figures

I Hyperbolic systems of conservation laws and fluid-structure interaction **9**

I.1 Space discretization for centered finite difference schemes 18

I.2 Space discretization for finite volume schemes on a Cartesian grid 19

I.3 Space discretization for finite volume schemes on an unstructured grid 20

I.4 Arakawa grid system 34

I.5 A fully explicit fluid-structure coupling algorithm on same time discretization 43

I.6 A fully explicit fluid-structure coupling algorithm on staggered time discretization 44

I.7 A fully implicit fluid-structure coupling algorithm on same time discretization 44

I.8 Direct forcing method 49

I.9 Embedded boundary method 51

I.10 Embedded boundary method - Cell merging 51

I.11 Mirroring techniques 53

II High order 2D finite volume conservative Lagrange-Remap schemes for compressible hydrodynamics on staggered Cartesian grids **55**

II.1 Staggered finite volume space discretization on Cartesian grids 60

II.2 Illustration of the interest and importance of the internal energy corrector. . 81

II.3 Non-convex equation of state for a breaking-wave test-case 86

II.4 Acoustic wave with harmonic source - Difference between the cell-centered GoHy [50] (blue, cross) and GAD schemes [84] (gray, filled triangle), the staggered BBC scheme [171] (orange, triangle) and the new staggered schemes denoted here YHORK (black, filled circle). Analytic solution is represented by the red curve. 87

II.5 Density and internal energy profiles for the Sod test-case problem [146] at time $t = 0.2$ with 100 cells for the 3rd, 4th and 6th order staggered schemes. 88

II.6 Density and pressure profiles for the Noh test-case problem [127] at time $t = 0.6$ with 400 cells for the 3rd, 4th and 6th order staggered schemes. 90

II.7 Density and pressure profiles for the Shu-Osher test-case problem [144] at time $t = 1.8$ with 200 cells for the 3rd, 4th and 6th order staggered schemes. 90

II.8 Density and pressure profiles for the Woodward test-case problem [171] at time $t = 0.038$ with 300 cells for the 3rd, 4th and 6th order staggered schemes. 91

II.9 Staggered finite volume space discretization on Cartesian grids 94

II.10 Flow chart for the 2D scheme 95

II.11 Profiles of density by colors and ϕ using 6 contours from 0 to 1 for the Vortex-Pairing test-case, CFL=0.7, for times $t = 1$, $t = 2$, $t = 3$, $t = 4$ and $t = 5$, 128 cells in each direction. 99

II.12 Results at time $t = 0.3$ for the first Riemann problem with the first and second order cell-centered scheme (top, CFL=0.5) as well as the third and fourth order staggered schemes (bottom, CFL=0.7) with 200 cells in each direction. Pressure is displayed by colors, and density using 32 contours from 0.16 to 1.71. 100

II.13 Results at time $t = 0.3$ for the second Riemann problem with the first and second order cell-centered scheme (top, CFL=0.5) as well as the third and fourth order staggered schemes (bottom, CFL=0.7) with 200 cells in each direction. Pressure is displayed by colors, and density using 29 contours from 0.25 to 3.05. 101

II.14 Results at time $t = 0.25$ for the third Riemann problem with the first and second order cell-centered scheme (top, CFL=0.5) as well as the third and fourth order staggered schemes (bottom, CFL=0.7) with 200 cells in each direction. Pressure is displayed by colors, and density using 30 contours from 0.54 to 1.7. 102

II.15 Results at time $t = 0.25$ for the fourth Riemann problem with the first and second order cell-centered scheme (top, CFL=0.5) as well as the third and fourth order staggered schemes (bottom, CFL=0.7) with 200 cells in each direction. Pressure is displayed by colors, and density using 29 contours from 0.43 to 0.99. 103

II.16 Results at time $t = 0.25$ for the fifth Riemann problem with the first and second order cell-centered scheme (top, CFL=0.5) as well as the third and fourth order staggered schemes (bottom, CFL=0.7) with 200 cells in each direction. Pressure is displayed by colors, and density using 30 contours from 0.53 to 1.98. 104

II.17 Scatter plot of density profiles for the Sedov blast-wave test-case using the third, fourth and sixth order staggered schemes (CFL=0.7) and the first and second order cell-centered schemes (CFL=0.5) at $t = 1.0$; 100 cells in each direction. 106

II.18 Scatter plot of density profiles for the 2D Noh compression test-case using the third, fourth order staggered schemes (CFL=0.7) and for the first and second order cell-centered schemes (CFL=0.5) at $t = 0.6$, 400 cells in each direction. Axis effect are present for the first and second order cell-centered schemes . . 107

II.19 Difference between pressure and atmospheric pressure p_{atm} following x at $y = 1$, for the third order scheme, with circa 10 cells per wavelength 108

II.20 Absorption (dB) of the pressure following x at $y = 1$, without rectification, for the third order scheme, with circa 10 cells per wavelength 109

II.21 Absorption (dB) of the pressure following x at $y = 1$, with geometric corrector, for the third order scheme, with circa 10 cells per wavelength 109

II.22 Arakawa C-type like grid for the compressible Navier–Stokes equation 113

II.23 Density profiles on the Rayleigh–Taylor mono-mode instability for the Euler equations (top) and for the Compressible Navier–Stokes (CNS) equations with $\mu = 10^{-4}$ and $\lambda = -\frac{2}{3}\mu$ (bottom) using third, fourth and sixth order schemes, at time $t = 9.5$ (left) and $t = 12.75$ (right) with 200 cells in the x-direction and 600 in the y-direction. 120

II.24 Density profiles on the Rayleigh–Taylor multi-mode instability for the Euler equations (top) and for the Compressible Navier–Stokes (CNS) equations with $\mu = 10^{-4}$ and $\lambda = -\frac{2}{3}\mu$ (bottom) using third, fourth and sixth order schemes, at time $t = 6$, $t = 9$, $t = 12$, $t = 15$ from left to right and top to bottom, with 200 cells in the x-direction and 300 in the y-direction 122

III Stable high-order methods for linear hyperbolic systems with arbitrary boundary conditions 123

III.1 1D Boundary between outside and inside computational domain 124

III.2 Stability area $\{(\nu, \sigma) \ / \ \rho(N_{nc}) \leq 1\}$ (in white) for the Lax–Wendroff (second order) scheme with $n_c = 20$ for the $\underline{\mathrm{R}}^{2,0}$ (left), $\underline{\mathrm{R}}^{2,1}$ (right) reconstruction operators. The whole domain is stable. 149

III.3 Stability area $\{(\nu, \sigma) \ / \ \rho(N_{nc}) \leq 1\}$ (in white) for the Beam–Warming (second order) scheme with $n_c = 20$ for the $\underline{\mathrm{R}}^{2,0}$ (left), $\underline{\mathrm{R}}^{2,1}$ (right) reconstruction operators. The whole domain is stable. 150

III.4 Stability area $\{(\nu, \sigma) \ / \ \rho(N_{nc}) \leq 1\}$ (in white) for the third-order projection scheme with $n_c = 20$ for the $\underline{\mathrm{R}}^{3,0}$ (top, left), $\underline{\mathrm{R}}^{3,1}$ (top, right) and $\underline{\mathrm{R}}^{3,2}$ (bottom) reconstruction operators. As a contrary to figs. III.2 and III.3, one notices a region of numerical instability for $\underline{\mathrm{R}}^{3,0}$. 151

III.5 Stability area $\{(\nu,\sigma)\ /\ \rho(N_{nc}) \leq 1\}$ (in white) for the fourth-order projection scheme with $n_c = 30$ for the $\mathcal{R}^{4,0}$ (top, left), $\mathcal{R}^{4,1}$ (top, right), $\mathcal{R}^{4,2}$ (bottom, left), $\mathcal{R}^{4,3}$ (bottom, right) reconstruction operators. An additional behaviour is observed w.r.t. fig. III.4 which is that the domain of instability contains a layer for small value of ν ($\underline{\mathrm{R}}^{4,0}$ and $\underline{\mathrm{R}}^{4,2}$) 152

III.6 Stability area $\{(\nu,\sigma)\ /\ \rho(N_{nc}) \leq 1\}$ (in white) for the third-order staggered scheme for the wave equations with $n_c = 40$ for the $\mathcal{R}^{3,0}$ (top, left), $\mathcal{R}^{3,1}$ (top, right) and $\mathcal{R}^{3,2}$ (bottom) reconstruction operators. 153

IV Discretization of boundary conditions for compressible hydrodynamics 155

IV.1 Discretization $\Gamma_{\Delta s}$ of $\Gamma(t)$ and decomposition of the whole domain between Ω_- (ghost-cells) and Ω_+ (fluid cells). Ω is the domain outside the ellipse. . . 157

IV.2 Graph of $x \to f(\tau_{+1}x)$ using different value of ΔX for a positive $\mathrm{D}_t g$ on the left, and a negative one on the right. 161

IV.3 Flow chart for the MOOD procedure applied at the boundary 168

IV.4 Velocity profiles with 10 cells per wavelength for the 2nd, 3rd, 4th and 6th-order GoHy schemes for the harmonic piston problem at $T = 9$. 170

IV.5 Velocity profiles with 10 cells per wavelength for the 3rd, 4th and 6th-order GoHy schemes for the harmonic piston problem at $T = 9$. On the left, results with appropriate order of reconstruction is depicted, whereas on the right results are shown with second order reconstruction. 171

IV.6 Density profiles with initially 100 cells for the 2nd, 3rd, 4th and 6th-order GoHy schemes for the Sod piston problem. 172

IV.7 Zoom on a point P_s on the discretized boundary with local coordinate system. The colored zone corresponds to a six points stencil for 3rd order reconstruction. 173

IV.8 Zoom on a point P_s on the discretized boundary with local coordinate system. The color zone corresponds to a least-squares stencil for 3rd order reconstruction. 177

IV.9 Pressure variations $|p - p_0|$ around the cylinder as a function of θ for $f = 0.5$ (top), $f = 1.0$ (middle), $f = 2.0$ (bottom) for 1st, 2nd and 3rd-order accurate schemes with $\Delta x = \Delta y = \frac{1}{20}$ (left) and $\Delta x = \Delta y = \frac{1}{40}$ (right). 182

IV.10 Pressure variations $|p - p_0|$ around the cylinder as a function of θ for $f = 0.5$ (top), $f = 1.0$ (middle), $f = 2.0$ (bottom) for the GoHy-3 interior scheme and 1st, 2nd and 3rd-order accurate ILW methods with $\Delta x = \Delta y = \frac{1}{20}$ (left) and $\Delta x = \Delta y = \frac{1}{40}$ (right). 183

IV.11 Density colors of a reflected shock wave on a wedge at CFL=0.5 with 100 cells in each direction. The expected angle of the oblique shock, depicted by the white line, is recovered by the schemes. 184

IV.12 Density contours of double Mach reflection for 1st (top), 2nd (middle) and 3rd-order (bottom) ILW-GoHy schemes with $\Delta x = \Delta y = \frac{1}{200}$; 30 contours from 1.731 to 20.92 as in [155]. 185

IV.13 Density contours of Mach 2.81 flow past a cylinder for 1st (top), 2nd (middle) and 3rd-order (bottom) ILW-GoHy schemes with $\Delta x = \Delta y = 4.10^{-4}$ at $t = 3.10^{-5}$ (left) and $= 6.10^{-5}$ (right); 30 contours from 0.3 to 8. 186

IV.14 Density contours of Mach 1.3 flow past a prism for 1st (top, left), 2nd (top, right) and 3rd-order (bottom) ILW-GoHy schemes with $\Delta x = \Delta y = 4.10^{-4}$ at $t = 1.5.10{-}4$, CFL=0.5; 30 contours from 0.5 to 1.8 187

IV.15 Pressure contours of a Mach shock on a NACA0018 for 1st (top, left), 2nd (top, right) and 3rd-order (bottom) ILW-GoHy schemes with 400 cells in each direction, CFL=0.5; 35 contours from 0.0 to 3.5 188

IV.16 Lift and drag coefficients as a function of time for the Mach shock on the NACA0018 profile considering 100, 200 and 400 cells in each direction for 1st (top, left), 2nd (top, right) and 3rd-order (bottom) ILW-GoHy schemes. . . . 189

IV.17 Pressure variations $|p - p_0|$ around the cylinder as a function of θ for $f = 0.5$ (top), $f = 1.0$ (middle), $f = 2.0$ (bottom) for the third order cell-centered scheme (GoHy-3, blue) and for the third order staggered scheme (STAG-3, black) with $\Delta x = \Delta y = \frac{1}{20}$ (left) and $\Delta x = \Delta y = \frac{1}{40}$ (right). 190

V Extension to fluid-rigid body interaction **191**

V.1 Regular curvilinear discretization of an ellipse with $\Gamma : s \to (5\cos(2\pi s), \sin(2\pi s))^t$ using 20 pearls (blue dots) 197

V.2 Using the Inverse Lax–Wendroff procedure as a time and space coupling for rigid body interaction. 204

V.3 Pressure profiles at time t=3 ms with 800 cells for the pressure motion driven piston in 1D for second, third, fourth and sixth order ILW-GoHy schemes. . . 205

V.4 60 contours of fluid pressure from 0 to 28 at times t=0.14 (top) and t=0.255 (bottom) for the third order scheme, $\Delta x = \Delta y = 6.25 \times 10^{-4}$. 206

V.5 60 contours of fluid density from 0 to 12 at times t=0.14 (top) and t=0.255 (bottom) for the third order scheme, $\Delta x = \Delta y = 6.25 \times 10^{-4}$. 207

V.6 60 contours of fluid pressure from 0 to 28 at times t=0.14 (top) and t=0.255 (bottom) for the third order scheme, $\Delta x = \Delta y = 6.25 \times 10^{-4}$. 208

V.7 60 contours of fluid density from 0 to 12 at times t=0.14 (top) and t=0.255 (bottom) for the third order scheme, $\Delta x = \Delta y = 6.25 \times 10^{-4}$. 209

V.8 60 contours of fluid pressure from 0 to 28 at times t=0.14 (top) and t=0.255 (bottom) for the third order scheme, $\Delta x = \Delta y = 6.25 \times 10^{-4}$. 209

V.9 60 contours of fluid density from 0 to 12 at times t=0.14 (top) and t=0.255 (bottom) for the third order scheme, $\Delta x = \Delta y = 6.25 \times 10^{-4}$. 210

List of Tables

I Hyperbolic systems of conservation laws and fluid-structure interaction **9**

I.1 A Butcher table for an explicit Runge–Kutta sequence 38

II High order 2D finite volume conservative Lagrange-Remap schemes for compressible hydrodynamics on staggered Cartesian grids **55**

II.1 Coefficients for the finite volume computation of point-wise values from cell-average ones. 62

II.2 Coefficients for the finite volume computation of average values from point-wise ones. 63

II.3 Coefficients for the δ operator. 63

II.4 Coefficients for the interpolation of cell-centered values from staggered ones and *vice-versa*. 64

II.5 Example of Butcher table for explicit Runge–Kutta sequence with s sub-cycles. 64

II.6 CFL conditions for linear stability of the staggered schemes 72

II.7 Illustration of the interest and importance of the internal energy corrector. Without the internal energy corrector, the term $\|\rho_0 e_{\text{kin}} - (\frac{1}{2}\rho_0 u^2)\|_{l^1([0:T]\times\Omega)}$ does not tend to 0 as ΔX and Δt tends to zero. 81

II.8 l^1-error in momentum and experimental order of convergence for the Lagrange-remap staggered scheme taken on the Cook-Cabot breaking wave test problem [28] 85

II.9 l^1-error in momentum and experimental order of convergence for the Lagrange-remap staggered scheme taken on the modified breaking wave test problem . 86

II.10 l^1-error in density for the Lagrange-remap staggered scheme taken on the Sod test problem [146] 88

II.11 l^1-error in density and experimental order of convergence for the Lagrange-remap staggered scheme taken on the 2D isentropic vortex advection test problem [174] 97

II.12 Initial states for the four quadrants of 2D Riemann problem for density, pressure and x and y velocity u and v. 105

II.13 l^1-error in density and experimental order of convergence for the Lagrange-remap staggered scheme with gravity forces taken on the atmosphere at hydrostatic equilibrium [92] 118

II.14 l^1-error in momentum and experimental order of convergence for the compressible Navier–Stokes Lagrange-remap staggered scheme for the Taylor–Green vortex [160] 119

III Stable high-order methods for linear hyperbolic systems with arbitrary boundary conditions 123

III.1 l^1-error and experimental order of convergence for the 3rd-order scheme together with the $\underline{\mathrm{R}}^{3,n}$ finite-volume reconstruction polynomial at $t = 1.5$. . . . 131

III.2 l^1-error and experimental order of convergence for the 4th-order scheme together with the $\underline{\mathrm{R}}^{4,n}$ finite-volume reconstruction polynomial at $t = 1.5$. . . . 131

III.3 Example of Butcher table for explicit Runge–Kutta sequence with p sub-cycles. 133

III.4 l^1-error and experimental order of convergence for the 3rd-order scheme together with the $\underline{\mathrm{R}}^{3,n}$ finite-volume reconstruction polynomial at $t = 0.3$ for boundary condition on the velocity. $\star$ are indications of unstable behaviour of the scheme. 142

III.5 l^1-error and experimental order of convergence for the 3rd-order scheme together with the $\underline{\mathrm{R}}^{3,n}$ finite-volume reconstruction polynomial at $t = 0.3$ for mixed boundary condition ($\lambda = 1747$). $\star$ are indications of unstable behaviour of the scheme. 142

IV Discretization of boundary conditions for compressible hydrodynamics 155

IV.1 l^1-error and experimental order of convergence (EOC) for ILW-GoHy schemes at $t = 0.01$ with a CFL of 0.9. EOC indexed with * are reduced due to double precision. For stability issues, least-squares method is used for 4th, 5th and 6th-order. 169

IV.2 l^1-error on density in both time and space, experimental order of convergence and cost in % of the ILW procedure for GoHy schemes on the 2D isentropic vortex at $t = 1.0$. 178

V Extension to fluid-rigid body interaction 191

V.1 Number of variables for rigid body motion as a function of given space dimensions 193

V.2 Comparisons of the position of the cylinder's center at $t = 0.255$. $\star$ denotes results for $\Delta x = \Delta y = 10^{-3}$. 206

V.3 Conservation on mass and total energy at $t = 0.255$ for the lift-off cylinder test-case. 207

A Butcher tables and weights for directional splitting methods 217

A.1 Generic second order Runge–Kutta sequence 218

A.2 Third order TVD Runge–Kutta sequence [70] 218

A.3 Original Kutta sequence [99] 218

A.4 The $\frac{3}{8}$-Kutta sequence [99] 219

A.5 Dormand–Prince Runge–Kutta sequence [49] 219

A.6 First order Godunov splitting weights ω_k 220

A.7 Second order Strang splitting weights ω_k 220

A.8 Third order directional splitting weights ω_k 220

A.9 Fourth order directional splitting weights ω_k 220

A.10 Sixth order directional splitting weights ω_k [175] 221

A.11 Eighth order directional splitting weights ω_k [175] 222

Introduction

En français

Les phénomènes d'interactions fluide-structure sont cruciaux pour les problèmes multi-physiques. Deux matériaux, de lois de comportement différentes, interagissent entre eux. Ici, un fluide compressible et un corps rigide sont considérés. L'écoulement du fluide est fortement conditionné par la forme de la structure ainsi que par son déplacement, tandis que le déplacement du solide est régi par les forces et moments de pression exercés à sa surface par le fluide. C'est un problème fortement couplé. Le couplage impacte directement la stabilité et la précision de la méthode numérique employée. En outre, l'utilisation de méthodes numériques sur grilles cartésiennes ajoute de la complexité à la discrétisation liée au fait que l'interface entre le fluide et la structure coupe arbitrairement la grille cartésienne.

En 1964, Noh crée le premier schéma explicite lagrangien et eulérien pour l'interaction entre un fluide et un corps rigide immobile [125]. Il propose un traitement conservatif de l'interface par plan orthogonal à la direction de balayage. Ce traitement a permis pour la première fois de retrouver par la simulation les effets d'un obstacle sur un écoulement compressible. Néanmoins, la géométrie est discrétisée de manière peu précise, ce qui induit des effets de marche sur les chocs réfléchis. En outre, ces effets de marche entraînent des erreurs d'ordre 1 qui deviennent prépondérantes pour des écoulements complexes, et nuisent conséquemment à la fiabilité de la méthode. De plus, sa discrétisation particulière de l'interface impacte directement sur la CFL, les pas de temps peuvent être infiniment petit en fonction de la position de l'interface dans la maille, ce qui peut provoquer l'intractabilité des calculs.

En 2003, Berger et al. proposent une technique de recombinaisons des mailles tronquées le long de la frontière, dénommée le h-algorithme [12]. Ce travail est basé sur des critères purement géométriques et fusionne des mailles adjacentes dans le cas où elles impacteraient la CFL. Ce travail a permis de réduire fortement l'impact des mailles tronquées sur le calcul du pas de temps. Néanmoins, la recombinaison des mailles tronquées est au plus d'ordre 2, et ne permet pas de suivre efficacement les quantités conservatives à l'intérieur de ces mailles, particulièrement dans le cas d'interfaces mobiles. La complexité de la forme de l'interface peut aussi induire des erreurs importantes voire même empêcher la convergence de l'algorithme proposé. En 3D, le coût du h-algorithme devient prohibitif et ne permet donc pas de gérer les frontières quelconques.

En 2006, Colella et al. proposent une nouvelle façon de reconstruire l'interface basée sur les

fractions volumiques de présence [26]. Contrairement à Noh, cette méthode permet de réduire considérablement les effets de marche à l'interface et reste conservative. Conjointement à l'utilisation du h-algorithme, il n'y a pas d'impact sur la condition CFL. Néanmoins, la reconstruction faite des interfaces ne permet pas d'excéder l'ordre 2 en espace.

Plus récemment, Tan et Shu proposent une méthode basée sur une procédure de Lax–Wendroff inverse pour les conditions aux bords [155]. Cette méthode est a priori sans restriction CFL et sans limitation quand à l'ordre de convergence de la méthode. Néanmoins, l'algèbre impliquée dans la méthode est extrêmement lourde et devient prépondérante en terme de coût de calcul. Elle n'est appliquée dans le cadre de leurs études qu'au cas du gaz parfait et aux schémas eulériens. De plus, certaines instabilités apparaissent et, sans contrôle, empêchent la convergence des schémas utilisés. Contrairement aux méthodes précédemment citées, il n'y a pas de preuve de conservation de la masse, quantité de mouvement et de l'énergie totale à l'interface. Dans le cas des géométries non-lipschitziennes, il est impossible, sans modification et détérioration, d'appliquer la méthode.

Partant de considérations générales concernant les systèmes hyperboliques de lois de conservation, une étude est faite d'un ensemble de méthodes numériques pour simuler les équations d'un fluide non-visqueux et compressible. L'accent est mis durant cette étude sur les schémas formulés en énergie interne et sur maillages décalés. Enfin, une revue bibliographique fait apparaître qu'il existe une multitude de méthodes permettant de simuler l'interaction entre un fluide compressible et un corps rigide indéformable de manière stable. Cette revue est présentée dans le chapitre I. Des méthodes d'ordre 2, stables et conservatives ont été créées. Des algorithmes géométriques de fusion de mailles permettent d'éviter toute contrainte sur la CFL liée à la taille des mailles tronquées. En outre, la procédure de Lax–Wendroff inverse permet de prendre en compte n'importe quelle condition aux bords à l'ordre élevé. Néanmoins la difficulté principale réside dans la discrétisation de la géométrie de l'interface qui impacte la montée en ordre ainsi que dans la stabilité de la méthode. Les méthodes de type ordre élevé proposé par Tan et Shu s'impliquent dans le cadre de schéma eulérien pur uniquement pour un gaz parfait. Elles sont en outre particulièrement onéreuses algébriquement. Enfin, cette méthode n'est pas toujours stable et peut être inapplicable dans le cas de certaines configurations géométriques.

C'est dans ce contexte que s'inscrit l'étude proposée ici. Elle consiste à développer une méthode numérique stable, d'ordre arbitrairement élevé capable de modéliser et simuler les interactions entre un fluide compressible et un corps rigide indéformable pour des schémas de type Volumes Finis Lagrange-Projection d'ordre très élevé et conservatifs sur grilles cartésiennes ainsi qu'à évaluer les gains en précision apportés par cette stratégie de couplage numérique.

La démarche a consisté dans un premier temps à étendre à l'ordre très élevé un schéma 2D Lagrange-Projection conservatif pour l'hydrodynamique compressible sur grilles cartésiennes décalées. En se basant sur le système des grilles Arakawa, les variables ont été redistribuées afin de faciliter l'intégration lagrangienne. Pour la première fois, la variable de masse est dédoublée sur la grille décalée afin d'assurer conservation, robustesse et consistance. Le schéma 1D d'ordre élevé Lagrange-projection sur grille décalée est basé sur une intégration en temps de type Runge–Kutta et en espace de type Volumes Finis pour la phase lagrangienne. Afin d'assurer la capture correcte

des chocs, pour la première fois, un correctif en énergie interne conservatif et d'ordre très élevé est proposé. Ce correctif est rendu possible par l'idée nouvelle de discrétiser l'équation d'évolution de l'énergie cinétique. La projection est basée sur l'intégration analytique par polynomes de Lagrange et est adaptée ici aux particularités des grilles cartésiennes décalées. Ainsi un schéma 1D conservatif et d'ordre élevé est obtenu. Son extension dans un cadre multi-dimensionnel par l'utilisation de séquences de balayage directionnel d'ordre élevé est faite. L'ordre très élevé est atteint expérimentalement (cf table 1). Enfin une extension aux fluides visqueux compressibles est proposée. Ce travail est présenté dans le chapitre II et a fait l'objet d'une publication [35].

La démarche a consisté dans un second temps à prendre en compte dans le cas des systèmes linéaires n'importe quelles conditions aux bords. Pour cela, on a développé une famille de méthodes d'ordre très élevé et stable pour des conditions aux bords sur frontières quelconques. Partant d'un système linéaire simplifié qu'est l'advection à vitesse constante, on développe la construction des opérateurs dits de reconstruction permettant de prendre en compte la condition aux bords imposée. Ces opérateurs de reconstruction sont d'ordre arbitrairement élevé et leur stabilité est étudiée. Dans l'idée de pouvoir déterminer a priori la stabilité d'opérateurs pour des systèmes plus complexes que l'advection, on crée la notion de stabilité réduite. Cette notion est ensuite utilisée dans le cas des systèmes linéaires hyperboliques. En particulier, une étude numérique est faite pour déterminer la stabilité réduite des opérateurs de reconstruction pour le cas du systèmes des équations des ondes. Ce travail est présenté dans le chapitre III et a fait l'objet d'une publication [34].

À partir des caractéristiques de stabilité des opérateurs de reconstruction dans le cas linéaire, la démarche a consisté dans un troisième temps à étendre les méthodes stables au cas non-linéaire des équations d'Euler 1D. Le caractère sous-déterminé du système obtenu conduit à prendre en compte une équation supplémentaire. Deux choix sont effectués. Le premier choix est basé sur une hypothèse faite sur le jeu d'équations aux dérivées partielles. Le second choix est lui basé sur l'utilisation d'un stencil plus large, afin d'éviter toute hypothèse sur les propriétés de l'écoulement. Enfin, on a étendu la méthode 1D au cas multi-dimensionnel, en se basant sur une méthode de balayage directionnel. La méthode ainsi développée permet de prendre en compte les conditions aux bords imposées en vitesse. En particulier, on a montré que c'était équivalent à réaliser le couplage entre un fluide compressible et un corps rigide indéformable de masse infinie. Ce travail est présenté dans le chapitre IV et a fait l'objet d'une publication [34].

Enfin, à partir de la discrétisation des conditions aux bords pour les équations d'Euler, le couplage entre un fluide compressible et un corps rigide de masse finie est réalisé. La méthode précédemment développée permet de calculer à l'ordre élevé les intégrales des moments et forces exercés sur le solide. Par conséquent, le couplage en temps comme en espace entre le fluide compressible et le corps rigide est naturel du fait de la discrétisation spatiale choisie pour l'interface. Pour ce faire, deux nouveaux schémas sont proposés, un premier basé sur une procédure de type Cauchy–Kovalevski et un second basé sur une procédure de type Runge–Kutta. Les propriétés de mouvement de corps rigides sont vérifiées. Enfin, on illustre numériquement la consistance, la convergence et la stabilité de la méthode. Ce travail est présenté dans le chapitre V.

Au terme de cette étude, on arrive à la conclusion que le couplage proposé est possible à l'ordre élevé (cf figure 2) et qu'il existe des configurations pour lesquelles un gain en précision est obtenu (cf figure 1).

In english

Fluid-structure interaction phenomena are important in multi-physics problems. It involves two materials that have different behaviours, different constitutive laws, but that are coupled one to another. Here, a compressible fluid and a rigid body are considered. The fluid flow is strongly conditioned by the shape of the solid but also by its displacement, and the solid motion is triggered by pressure forces and torques exerted on its boundary. This is a strongly coupled problem, which can be a predicament for the stability and accuracy of numerical methods. Indeed, for the development of numerical methods for fluid-structure interaction, the main difficulty is to obtain, without further CFL restriction, a stable and high-order accurate coupling between fluid and structure solvers. An additional difficulty is that for general problems, it is quite impossible to determine *a priori* how the coupling behaves, if the fluid forces and torques are predominant or if it is rather the displacement of the rigid body. This difficulty increases furthermore if one considers that the fluid solver is based on Cartesian grids. Indeed, the boundary intersects in an arbitrary fashion the grids. Increasing the order of accuracy leads to unstable methods, which prevent most uses of the coupling algorithm, as the schemes do not converge.

In 1964, Noh builds the first explicit Lagrangian and Eulerian scheme for the fluid-structure interaction in [125]. The structure is considered motionless and without deformations. As the scheme is based on directional splitting, he proposes a conservative treatment of the interface, considering that the boundary of the structure is always orthogonal or parallel to the cells interfaces. The numerical treatment detailed by Noh enables, for the first time, to recover using simulations, the effects of an obstacle on a fluid flow. However, the obstacle boundary is discretized abruptly, which induces "step effects" on reflected shocks. Moreover, the CFL restriction is directly impacted by the discretization proposed by Noh. Indeed, cells near interfaces are considered to be cut and then, the smaller the cut-cells, the stronger the CFL restriction. Note also that due to the geometrical approximation, the method is at most first order accurate. In 2003, Berger and al. propose a technique in order to mix cells near the boundary, called the *h-algorithm* [12]. This work relies on purely geometrical criteria to mix adjacent cells, if their size lead to CFL restriction. This work tends to reduce drastically the impact of small cut-cells on the time-step given by the CFL restriction. Nonetheless, the cut-cells mixing is at most second order accurate. For moving obstacles, special procedure must be developed to dispatch quantities inside mix-cells into the neighbourhood. Moreover, and especially in 3D, the complexity of the rigid body geometric shape induces large errors (and eventually prevent the scheme from converging). The more complex the geometric shape, the more difficult it is to deal with their numerical treatment. In 2006, Colella and al. in [26] develop an innovative way of tracking the interface based on volume fractions. As a contrary to Noh, this method reduces considerably the "step effects" due to the geometrical approximations and it is still conservative. However, due to the geometric approximation of the interface, the scheme is at most second order accurate in space. More recently, Tan and Shu propose a method based on the inverse Lax–Wendroff procedure for numerical boundary treatment in [155]. This method is *a priori* without any CFL restriction and can be very high-order accurate. However, the algebra used to design the method is extremely

heavy and the method in itself is only applied for perfect gases and for Eulerian schemes. As a contrary to the previous method, the procedure is not conservative in mass, momentum and total energy. For non-Lipschitz geometrical shapes, it is impossible to maintain high-order accuracy without modification of the procedure.

Starting from general considerations on hyperbolic systems of conservation laws, a review is done concerning the numerical methods available in the literature to approximate the compressible Euler equations. The emphasis is laid on schemes formulated in internal energy and on staggered grids. Last, an overview of the numerical methods available in the literature for fluid-structure interaction is done. Fictitious domain methods are extensively detailed. This work is presented in chapter I. Stable, conservative and second order accurate numerical methods have been designed to tackle fluid-structure interaction. Most are based on geometric approximations of the interface, as well as physical considerations concerning the behaviour of the fluid near the boundary. A focus is especially done on the possible CFL restriction induced by the chosen numerical boundary treatment.

It is in this very context that lies the work proposed in this manuscript. It consists in developing a stable and high-order accurate numerical method for fluid-structure interactions. The method is designed for conservative and high-order accurate finite volume schemes based on the Lagrange-remap formalism for Cartesian grids.

Firstly, the extension to high-order accuracy in both time and space of a hydrodynamics scheme on staggered Cartesian grids is done. The scheme is based on a Lagrange-remap formalism and is formulated in internal energy. Starting from the Arakawa grids system, variables are distributed on the staggered grids to ease the resolution of the Lagrangian system. The 1D scheme is based on a Runge–Kutta for the time integration and uses finite volume formalism. The scheme is conservative in mass, momentum and total energy (see lemmas II.2 and II.8) and weakly consistent for the compressible Euler equations (see theorem II.9). An internal energy corrector is developed and is the key for both conservation and weak consistency. Such a corrector derives from the discretization of the kinetic energy, independently of the momentum. The remapping phase is based on standard polynomial projection, but adapted here to the special case of staggered grids. The extension to multi-dimensions is made possible thanks to high-order accurate directional splitting methods. Results concerning the accuracy and the order of convergence are displayed in table 1. Then, an extension of the scheme for compressible Navier–Stokes equations is proposed. A part of this works has been published in "Comptes Rendus Mathématique" [35] and is extensively detailed in chapter II.

Secondly, for linear hyperbolic system of conservation laws, a numerical boundary treatment is developed. For any well-posed boundary conditions, a stable and high-order accurate discretization of boundary condition is proposed. Starting from the advection equation problem, a generic way of building operators to take into account the boundary condition is detailed. Those operators, called reconstruction operators, enable to build ghost-cells values outside the fluid domain without impacting CFL restriction. In order to determine if a scheme with a given numerical boundary treatment is stable, the notion of reduced stability is introduced in definition III.1.

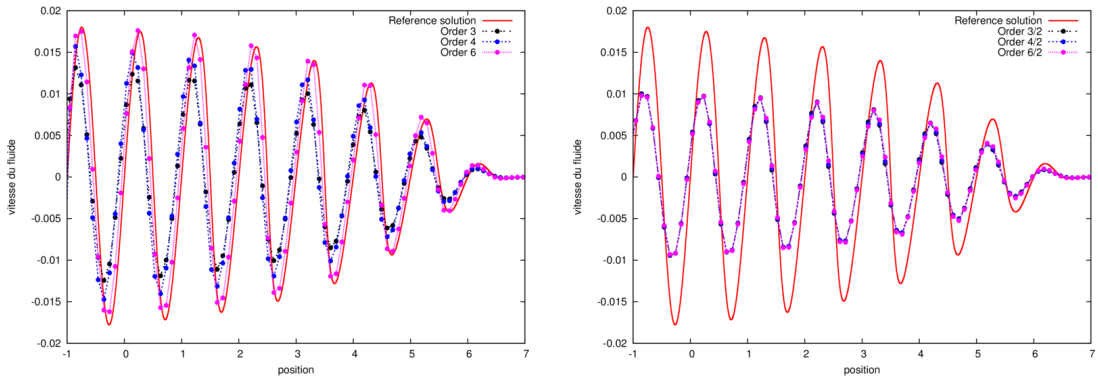
N_x	STAG-3		STAG-4		STAG-5		STAG-6		STAG-7		STAG-8	
50	3.3e-1	·	1.5e-1	·	2.6e-1	·	1.7e-1	·	1.5e-1	·	1.1e-1	·
100	9.5e-2	1.79	1.9e-2	3.01	4.9e-2	2.41	8.9e-3	4.27	1.2e-2	3.70	2.0e-3	5.83
200	1.6e-2	2.54	1.0e-3	4.19	1.9e-3	4.68	6.5e-5	7.10	8.0e-5	7.20	5.2e-6	8.59
400	2.2e-3	2.89	6.1e-5	4.06	6.1e-5	4.96	7.2e-7	6.48	6.3e-7	7.00	1.6e-8	8.37
800	2.8e-4	2.97	3.9e-6	3.99	1.9e-6	4.98	9.9e-9	6.18	5.0e-9	6.97	1.1e-10	7.17
1600	3.5e-5	2.99	2.4e-7	3.99	5.98e-8	4.99	1.5e-10	6.02	3.9e-11	6.99	3.4e-12	$\star$

Table 1 – Illustration of the high-order accuracy of the staggered schemes: l^1-error in density and experimental order of convergence for the 2D Lagrange-remap staggered scheme taken on the isentropic vortex advection test problem [174], until $t = 20$, CFL=0.9. $\star$ indicates machine precision reached.

This notion provides practical informations about the scheme stability and is used to determine *a priori* if a scheme is stable or not. It is then applied on the wave equations problem and later to generic linear hyperbolic systems. This work is presented in chapter III and has been submitted to a journal [34].

Thirdly, using results obtained in chapter III for the linear case, the method is extended for the numerical boundary treatment of Euler equations. Works are first performed in 1D case, considering the boundary condition to be imposed on the normal velocity. Interest of high-order boundary treatment is highlighted in fig. 1. For this special case, the global accuracy is mostly due to the numerical boundary treatment accuracy. It highlights the interest of having a high order discretization of boundary conditions, particularly for high order fluid solver. The procedure is first detailed for a simple second order accurate example. One identifies that the non-inversibility of the Lagrangian system Jacobian matrix requires another equation to be added. Two methods are derived. The first one consists in adding an equation that describes a peculiar feature of the flow. The flow is considered to be spatially isentropic near the boundary. A theoretical result is given in lemma IV.1 which characterizes conditions for existence and uniqueness of the reconstruction near the boundary. The second method consists in enlarging the stencil on which the reconstruction is based without any hypothesis on the flow structure near the boundary. Theoretical results are available in lemmas IV.2 and IV.3. They characterize once again conditions for existence and uniqueness of the reconstruction. Then, the method is extended to the multidimensional case, using directional splitting method. To prevent any numerical instabilities from occuring, a least-square procedure is developed, as well as a MOOD one in case of strong shocks. This is explained and illustrated in chapter IV and has also been submitted to a journal [34].

Fourthly and lastly, using the reconstruction method proposed in chapter IV, the coupling between a compressible fluid and a rigid body is done. A semi-discrete scheme for rigid body dynamics is derived to compute with high-order accuracy the forces and torques resultants exerted on the rigid body boundary. The coupling is straightforward using the reconstruction method. The time integration is done to match the one of the interior scheme, whether with a Runge–Kutta one or with a Cauchy–Kovalevskaya one. For multidimensional problems, directional splitting method is applied. As illustrated in fig. 2, the proposed coupling is able to



(a) High-order reconstruction (b) 2^{nd}order reconstruction

Figure 1 – An oscillating boundary conditions is prescribed on the left boundary. It highlights the impact of high-order accurate numerical boundary treatment for the restitution of physical oscillations. Velocity profiles are depicted with 10 cells per wavelength at $T = 9$, for 3rd, 4th and 6th-order inner schemes, with a 2nd-order (left) or with respectively the same orders (right) boundary reconstructions. High-order accurate boundary treatment outperforms 2nd-order accurate ones in the whole domain, because the gain of accuracy propagates *in* the domain (we expect this kind of behaviour to occur when considering fluid / vibrating structures interactions.).

recover complex fluid flow structures.

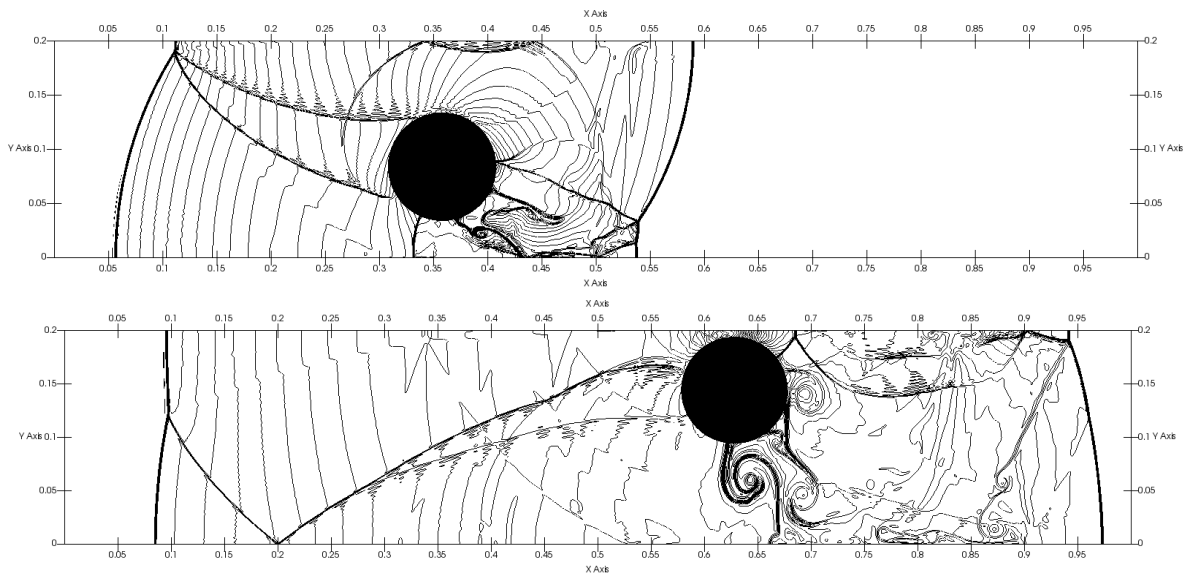


Figure 2 – Rigid-body and compressible fluid coupling. The cylinder is lift off by an incoming shock wave. In return, the shock wave is reflected on the cylinder, and the fluid is displaced by the structure. Complex structures are developed due to the reflection on the top and bottom channel as well as the rigid cylinder. 60 contours are displayed representing fluid density from 0 to 12 at times t=0.14 (top) and t=0.255 (bottom) for the third order scheme, $\Delta x = \Delta y = 6.25 \times 10^{-4}$.

Chapter I

Hyperbolic systems of conservation laws and fluid-structure interaction

---

Ce chapitre est une introduction aux méthodes numériques pour l'approximation de problèmes multiphysiques complexes. Le mode de présentation consiste à réunir dans un cadre commun des éléments classiques de la littérature, mais qui sont souvent présentés dans des contextes très différents. Dans un premier temps, des considérations générales sur les systèmes hyperboliques de lois de conservations sont rappelées. Dans un second temps, la présentation de différentes méthodes pour approcher numériquement la solution de ce type de système est faite : le cas du système hydrodynamique compressible ou des équations d'Euler est plus particulièrement étudié. Ces méthodes seront rangées dans deux familles distinctes. La première famille recense les méthodes basées sur un maillage d'éléments permettant d'approcher au mieux la déformation et/ou les bords du domaine. La seconde famille rassemble les méthodes d'ordre élevé, qu'elles soient sur grilles cartésiennes ou sur grilles non-structurées. Enfin, dans un troisième temps, une revue sera faite des différentes méthodes numériques présentes dans la littérature concernant le problème de la discrétisation et de l'approximation pour l'interaction fluide-structure. L'accent sera particulièrement mis sur le couplage en espace comme en temps de la méthode numérique pour le fluide avec celle pour la structure. Le couplage en espace portera essentiellement sur l'utilisation de méthodes de type domaine fictif.

This chapter is dedicated to an overview of numerical methods for the approximation of complex multi-physics problems. First, general considerations on hyperbolic systems of conservation laws are given. Second, the emphasis is laid on numerical approximations of such problems, with a special focus and care for the compressible hydrodynamics system. Numerical methods are classified into two families. The first family is for mesh-based method to approximate the deformation and/or the boundary with geometric elements. The second one is for the high-order accurate Direct Eulerian or Lagrange-remap methods on Cartesian grids as well as unstructured ones. Third, a focus is made on discretizations and approximations methods for the fluid-structure interaction problem. A special interest is made in the time and space coupling between the numerical method for the fluid part and the one for the structure part. A focus for the space coupling is made on fictitious domain methods.

I-1 Hyperbolic systems of conservation laws and their numerical approximations . . 10
I-1.1 Hyperbolic system of conservation laws in one dimension 10
I-1.2 Numerical methods for conservation laws and their properties 17
I-2 Numerical methods for compressible hydrodynamics 28
I-2.1 Euler and Lagrange equations for compressible hydrodynamics 28
I-2.2 Lagrangian and ALE methods for compressible hydrodynamics 32
I-2.3 High-order direct Eulerian and Lagrange-Remap numerical schemes . . . 36
I-2.4 Artificial viscosities and hyperviscosities 39
I-3 Numerical methods for fluid-structure interaction 41
I-3.1 Time coupling method for fluid-structure interaction 42
I-3.2 Space coupling method for fluid-structure interaction 45

I-1 Hyperbolic systems of conservation laws and their numerical approximations

This section is dedicated to the study of hyperbolic systems of conservation laws in one dimension and to their numerical approximations. First, mathematical properties of such systems are detailed. Second, a short overview of numerical approximations for such problems is depicted. Last, stability, consistency and convergence properties of the numerical schemes are presented as well as the analytic tools to analyze those properties for a given scheme.

I-1.1 Hyperbolic system of conservation laws in one dimension

For general non-linear conservation laws, assuming the data to be smooth over time, one may use the method of characteristics to determine smooth solutions to the hyperbolic system. But, the non-linearity introduces generally discontinuity in a finite time, even for smooth initial data.

Using the concept of weak solutions for conservation laws [102, 106, 103, 141, 61, 47, 45] and especially the Rankine-Hugoniot jump conditions, one may still define solutions to the hyperbolic system. However, uniqueness for the Cauchy problem is lost in the process. Adding the concept of entropic solutions, uniqueness for the Cauchy problem is proven in the special case of scalar conservation laws. In the special case of fluid dynamics, the thermodynamics yield a natural mathematical entropy.

Consider an hyperbolic system of conservation laws in one space dimension under the form

$$\partial_t \boldsymbol{U} + \partial_x \boldsymbol{f}(\boldsymbol{U}) = \boldsymbol{0}, \quad x \in \Omega, \quad t > 0. \tag{I.1}$$

Assuming that Ω is a bounded domain of $\mathbb{R}$, one gets

$$\partial_t \int_\Omega \boldsymbol{U} + \int_{\partial\Omega} \boldsymbol{f}(\boldsymbol{U}) = \boldsymbol{0}, \quad t > 0. \tag{I.2}$$

For special condition of no-exchange with the exterior, i.e. $\boldsymbol{f}(\boldsymbol{U})$ is null along the boundary of Ω, using eq. (I.2) one gets the global conservation of $\boldsymbol{U}$

$$\partial_t \int_\Omega \boldsymbol{U} = \boldsymbol{0}, \quad t > 0. \tag{I.3}$$

Using eq. (I.3), the average value of $\boldsymbol{U}$ over Ω defined as

$$\overline{\boldsymbol{U}} := \frac{1}{|\Omega|} \int_\Omega \boldsymbol{U}(x,t)\mathrm{d}x$$

is constant in time for no-exchange boundary conditions. It is usual to consider that the unknown $\boldsymbol{U}(x,t)$ belongs to a convex open set $\mathscr{U} \subset \mathbb{R}^N$. The flux function $\boldsymbol{f}$ is defined as a smooth enough function, typically $\boldsymbol{f} \in \mathscr{C}^1$

$$\begin{array}{rccc} \boldsymbol{f}: & \mathscr{U} & \longrightarrow & \mathbb{R}^N \\ & \boldsymbol{U} & \longmapsto & \boldsymbol{f}(\boldsymbol{U}) \end{array}$$

Less constrictive hypothesis of regularity on the flux function $\boldsymbol{f}$ are possible [61], but not detailed hereafter. In the peculiar case, where $N = 1$, one gets a scalar conservation law. For a scalar conservation law, one drops the vectorial notation and use u instead of $\boldsymbol{U}$ and f rather than $\boldsymbol{f}$.

I-1.1.1 Smooth solutions of conservation laws

First, consider that $\boldsymbol{U} \in \mathscr{C}^1(\mathbb{R} \times \mathbb{R}^{+,*}, \mathscr{U})$ and $\boldsymbol{U}$ satisfies eq. (I.1). Then $\boldsymbol{U}$ is said to be a *classical solution*. In peculiar as $\boldsymbol{U} \in \mathscr{C}^1(\mathbb{R} \times \mathbb{R}^{+,*}, \mathscr{U})$, it yields that $\partial_t \boldsymbol{U}$ and $\partial_x \boldsymbol{f}(\boldsymbol{U})$ are well-defined for any point $(x,t) \in \mathbb{R} \times \mathbb{R}^{+,*}$.

For a scalar conservation law, let $a(u) = f'(u)$ then the Cauchy problem written in non-conservative form writes

$$\begin{cases} \partial_t u + a(u)\partial_x u &= 0, \quad x \in \mathbb{R}, \quad t > 0, \\ u(x,0) &= u_0(x) \end{cases} \tag{I.4}$$

Theorem I.1 (Classical solution to the Cauchy problem [61]). *Let $f \in \mathscr{C}^2(\mathbb{R})$, $u_0 \in \mathscr{C}^1(\mathbb{R})$, and $a \in \mathscr{C}^1(\mathbb{R})$. Assume D defined as*

$$D = \inf_{x\in\mathbb{R}} \{\partial_x(a(u_0(x))\} \tag{I.5}$$

is real. Let

$$T^\star = \begin{cases} +\infty, & \text{for } D \geq 0 \\ -\frac{1}{D}, & \text{otherwise.} \end{cases} \tag{I.6}$$

If $T^\star$ is not zero, then the Cauchy problem in eq. (I.4) *has a unique solution $u \in \mathscr{C}^1(\mathbb{R}\times[0,T^\star[\,,\mathbb{R})$.*

The theorem I.1 gives the existence of a smooth solution for $0 < t < T^\star$. If D is positive, then it yields the existence for all time $t > 0$. But otherwise, it is all but natural to want to define u for time greater than $T^\star$. In fact, for a non-positive value of D, as t increases toward $T^\star$, the profile of u is going steeper until it reaches a discontinuity. At this point, the solution is no-longer in $\mathscr{C}^1$. Then the definition of classical solution as introduced previously is too narrow. For such cases, the *weak solutions* are introduced in order to allow discontinuities.

I-1.1.2 Weak solutions of conservation laws

Assume that $\boldsymbol{U}$ satisfies the initial conditions

$$\boldsymbol{U}(x,0) = \boldsymbol{U}_0(x), \quad x \in \mathbb{R}. \tag{I.7}$$

The following definition extends the definition of classical solution presented in the theorem I.1 to the case of functions with discontinuities.

Definition I.1 (Weak solution to the Cauchy problem [47]). Let $\boldsymbol{U}_0 \in L^\infty_{\text{loc}}(\mathbb{R})^N$. A function $\boldsymbol{U}$ is a weak solution of eqs. (I.1) and (I.7) if $\boldsymbol{U}(x,t) \in \mathscr{U}$ almost everywhere and if for any $\boldsymbol{\phi} \in \mathscr{C}^1_0(\mathbb{R}\times\mathbb{R}^{+,*})^N$ compactly supported

$$\int_{\mathbb{R}}\int_0^\infty (\boldsymbol{U}(x,t)\partial_t\boldsymbol{\phi} + \boldsymbol{f}(\boldsymbol{U}(x,t))\partial_x\boldsymbol{\phi})\,\mathrm{d}x\mathrm{d}t + \int_{\mathbb{R}} \boldsymbol{U}_0(x)\boldsymbol{\phi}(x,0)\mathrm{d}x = 0 \tag{I.8}$$

As a contrary to the original writing of eqs. (I.1) and (I.7), eq. (I.8) does not require the definition of the terms $\partial_t\boldsymbol{U}$ and $\partial_x\boldsymbol{f}(\boldsymbol{U})$. Moreover it contains intrinsically the initial conditions $\boldsymbol{U}_0$. In practice, a weak solution $\boldsymbol{U}$ in the sense of definition I.1 is said to satisfy eq. (I.1) in the sense of distributions. Moreover if a function $\boldsymbol{U}$ is a weak solution and is smooth, then it is a classical solution. It is stated in proposition I.2.

Proposition I.2 (A smooth weak solution is a classical solution [47]). *Let $\boldsymbol{U}$ be a weak solution in the sense of definition I.1. Assume $\boldsymbol{U} \in \mathscr{C}^1(\mathbb{R}\times\mathbb{R}^{+,*},\mathscr{U})$. Then $\boldsymbol{U}$ is a classical solution to the Cauchy problem stated in eqs.* (I.1) *and* (I.7).

For physical states of $\boldsymbol{U}$, it is interesting to focus on piecewise continuous functions. Those are functions that contains a finite number of discontinuities and are otherwise continuous on intervals. A very important result is the theorem I.3.

Theorem I.3 (Rankine–Hugoniot conditions [61]). *Let initial condition $\boldsymbol{U}_0$ be piecewise $\mathscr{C}^1$. $\boldsymbol{U} \in L^\infty_{loc}(\mathbb{R} \times \mathbb{R}^{+,*})^N$ a piecewise $\mathscr{C}^1$ function is a weak solution of eqs.* (I.1) *and* (I.7) *if and only if*

i) $\boldsymbol{U}$ is a classical solution of eqs. (I.1) *and* (I.7) *on intervals where $\boldsymbol{U}$ is smooth.*

ii) $\boldsymbol{U}$ satisfies the Rankine–Hugoniot jump conditions on the discontinuity points x_c

$$\boldsymbol{f}(\boldsymbol{U}(x_c^r, t)) - \boldsymbol{f}(\boldsymbol{U}(x_c^l, t)) = \sigma(\boldsymbol{U}(x_c^r, t) - \boldsymbol{U}(x_c^l, t)) \tag{I.9}$$

where σ is the discontinuity velocity, i.e. $\sigma = \frac{\mathrm{d}x_c}{\mathrm{d}t}$.

So far, we have exposed the notion of weak solutions to the Cauchy problem defined in eqs. (I.1) and (I.7). Using the Rankine–Hugoniot conditions defined in theorem I.3, one may build discontinuous solutions. However, it occurs that both solutions may coexist. The uniqueness of the Cauchy problem is then not satisfied. To get uniqueness back, and only in the special case of scalar conservation laws, the concepts of mathematical entropy and therefore entropic solutions are introduced.

I-1.1.3 Entropic solutions of conservation laws

For physical systems, the second law of the thermodynamics states that the entropy of a system increases over time or stays constant for an isolated system. The increase of entropy is synonym of irreversibility of processes. On the partial differential system, it yields another equation, eg. for smooth flows satisfying the Euler equations the entropy is advected.

Definition I.2 (Mathematical entropy [47]). Let Ω a open bounded subset of $\mathbb{R}^N$. Consider a flux function $\boldsymbol{f}$ of the form

$$\begin{aligned} \boldsymbol{f}: \quad \Omega &\longrightarrow \quad \mathbb{R}^N \\ \boldsymbol{U} &\longmapsto \quad \boldsymbol{f}(\boldsymbol{U}). \end{aligned}$$

A strictly convex function η such that

$$\begin{aligned} \eta: \quad \Omega &\longrightarrow \quad \mathbb{R} \\ \boldsymbol{U} &\longmapsto \quad \eta(\boldsymbol{U}) \end{aligned}$$

is a mathematical entropy for the conservation laws presented in eq. (I.1) if and only if there is an entropy flux ζ satisfying

$$\mathrm{d}\zeta(\boldsymbol{U}) = \mathrm{d}\eta(\boldsymbol{U}) \cdot \mathrm{d}\boldsymbol{f}(\boldsymbol{U}). \tag{I.10}$$

Any classical solution of eq. (I.1) satisfies

$$\partial_t \eta(\boldsymbol{U}) + \partial_x \zeta(\boldsymbol{U}) = 0 \tag{I.11}$$

The definition of the entropy flux based on eq. (I.10) gives immediately the following propriety.

Proposition I.4 (Hyperbolicity in 1D [47]). *Assume there exist an entropy and an entropy flux* (η, ζ) *for eq.* (I.1). *Then the system is hyperbolic. Especially the matrix* $\mathrm{d}\boldsymbol{f}(\boldsymbol{U})$ *is diagonalizable over the reals.*

Proposition I.4 can be extended to multidimensional systems. The following proposition gives hyperbolicity results for 2D systems.

Proposition I.5 (Hyperbolicity in 2D [47]). *Assume there exist an entropy and entropy fluxes* (η, ζ, ξ) *for the 2D conservation laws system*

$$\partial_t \boldsymbol{U} + \partial_x \boldsymbol{f}(\boldsymbol{U}) + \partial_y \boldsymbol{g}(\boldsymbol{U}) = \boldsymbol{0}. \tag{I.12}$$

Then the system is hyperbolic. Especially for any vector $\boldsymbol{n} = (n_x, n_y) \in \mathbb{R}^2$ *such that* $\|\boldsymbol{n}\| = 1$, *the matrix* $A = \mathrm{d}\boldsymbol{f}(\boldsymbol{U}) \cdot n_x + \mathrm{d}\boldsymbol{g}(\boldsymbol{U}) \cdot n_y$ *is diagonalizable over the reals.*

Remark I.1. Propositions I.4 and I.5 hold for three space dimensions systems.

Propositions I.4 and I.5 are particularly useful for the finite volume schemes that will be presented later on. Now, the emphasis is laid on scalar conservation law. Indeed, for such a law, any strictly convex function η is a mathematical entropy function.

Theorem I.6 (Viscous limit of a scalar conservation law [47]). *Let* η *be a mathematical entropy for the* ***scalar*** *conservation law eq.* (I.1) *with the associated entropy flux* ζ. *Let* $(u^\epsilon)_{\epsilon>0}$ *a* $\mathscr{C}^2$ *family of solution of*

$$\partial_t u^\epsilon + \partial_x f(u^\epsilon) = \epsilon \partial_{xx} u^\epsilon, \quad x \in \mathbb{R}, \quad t > 0. \tag{I.13}$$

Assume that (u^ϵ) *is uniformedly bounded in* $L^\infty(\mathbb{R} \times]0:\infty[)$ *such that*

$$\exists C > 0, \forall \epsilon > 0, \|u^\epsilon\|_{L^\infty}(\mathbb{R} \times]0:\infty[) \leq C. \tag{I.14}$$

Assume that $(u^\epsilon)_{\epsilon>0}$ *converges almost everywhere to* $u \in L^\infty(\mathbb{R} \times]0:\infty[)$. *Then* u *is solution in the sense of distributions to eq.* (I.1) *and satisfies the entropic inequality in the sense of distribution*

$$\partial_t \eta(u) + \partial_x \zeta(u) \leq 0 \text{ in the sense of distribution,} \tag{I.15}$$

which is equivalent to, for any $\phi \in \mathscr{C}^\infty(\mathbb{R} \times]0:\infty[)$ *compactly supported and* $\phi \geq 0$

$$\int_{\mathbb{R}\times]0:\infty[} (\eta(u)\partial_t \phi + \zeta(u)\partial_x \phi)\mathrm{d}x\mathrm{d}t \geq 0. \tag{I.16}$$

The theorem I.6 gives a characterization of a solution in the sense of distributions to eq. (I.1) which satisfies the entropy inequality (I.16). It seems all the more natural now, to define what is an entropic solution of a conservation law, and to determine conditions to get existence and uniqueness of such a solution.

Definition I.3 (Entropic solution of a conservation law)**.** Let $u_0 \in L^\infty(\mathbb{R})$. Let $u \in L^\infty(\mathbb{R} \times]0:\infty[)$ a weak solution to the **scalar** conservation law eq. (I.1) with the initial condition u_0. The function u is said to be an entropic solution of the Cauchy problem if for any mathematical entropy pair (η, ζ), it satisfies eq. (I.16).

Theorem I.7 (Existence and uniqueness of an entropic solution to the Cauchy problem [61])**.** *Suppose that f is a $\mathscr{C}^1$ function and that the initial condition u_0 lies in $L^\infty(\mathbb{R})$. Then the Cauchy problem with initial condition u_0 has a unique entropic solution to the **scalar** conservation law eq.* (I.1) *which satisfies the following conditions*

i) $u \in L^\infty(\mathbb{R} \times]0:\infty[)$,

ii) $\|u\|_{L^\infty(\mathbb{R}\times]0:\infty[)} \leq \|u_0\|_{L^\infty(\mathbb{R})}$

iii) Moreover, if u_0 satisfies a bounded inequality, s.t.

$$\exists \quad (\alpha, \beta) \in \mathbb{R}^2, \alpha \leq u_0(x) \leq \beta, \text{ for almost every } x \in \mathbb{R}$$

then

$$\alpha \leq u \leq \beta, \text{ for almost every } x \in \mathbb{R}, \quad \forall t > 0$$

Previous theorem only applies for the Cauchy problem with initial condition. For most cases, boundary conditions have to be prescribed. In some cases, physical considerations give natural boundary conditions, but it is not always the case, and thus, taking into account boundary conditions is both tricky and a hard problem to tackle. To understand the boundary conditions mechanism, the initial boundary value problem is introduced.

I-1.1.4 The initial boundary value problem

Consider the classical initial boundary value problem in the domain $x > 0,\ t > 0$ which writes

$$\begin{cases} \partial_t \boldsymbol{U} + \partial_x \boldsymbol{f}(\boldsymbol{U}) & = \boldsymbol{0}, & x > 0, \quad t > 0 \\ \boldsymbol{U}(x,0) & = \boldsymbol{U}_0(x), & x > 0 \\ \boldsymbol{U}(0,t) & = \boldsymbol{g}(t), & t > 0 \end{cases} \tag{I.17}$$

The problem depicted in eq. (I.17) is generally ill-posed. Boundary conditions must be prescribed accordingly to the eigenvalues of $\nabla_{\boldsymbol{U}} \boldsymbol{f}$ and not arbitrarily. The study presented here only concerns linear hyperbolic systems.

One-dimensional advection equation

The one-dimensional advection problem with prescribed boundary conditions writes as

$$\begin{cases} \partial_t u + a \partial_x u & = 0, & x > 0, \quad t > 0 \\ u(x,0) & = u_0(x), & x > 0 \\ u(0,t) & = g(t), & t > 0 \end{cases} \tag{I.18}$$

The problem is well-posed in the sense of Kreiss [96] if $a > 0$. For a negative a, no boundary conditions are required at $x = 0$, and the solution is trivially

$$u(x,t) = u_0(x - at), \quad x > 0, \quad t > 0.$$

For $a > 0$, solution to eq. (I.18) writes

$$u(x,t) = \begin{cases} u_0(x - at) & \text{for } x > at \\ g(t - \frac{x}{a}) & \text{for } x < at \end{cases} \tag{I.19}$$

Proposition I.8 (Classical solution [96]). *$u \in \mathscr{C}^1$ is a classical solution of eq. (I.18) if*

i) $u_0 \in \mathscr{C}^1$

ii) $g \in \mathscr{C}^1$

iii) *u_0 and g satisfy the compatibility relation*

$$g(0) = u_0(0), \quad \partial_t g(0) = -a\partial_x u_0(0). \tag{I.20}$$

Incrementally, u belongs to $\mathscr{C}^p$, $p > 0$ if u_0 and g belong to $\mathscr{C}^p$ and if they satisfy the compatibility relation

$$\partial_t^k g(0) = (-a)^k \partial_x^k u_0(0), \quad \text{for } 0 \leq k \leq p. \tag{I.21}$$

One-dimensional linear systems

Consider a linear hyperbolic system. Let the matrix $\underline{\boldsymbol{A}}$ satisfy $\underline{\boldsymbol{A}} = \nabla_{\boldsymbol{U}} \boldsymbol{f}(\boldsymbol{U})$ which is independent of $\boldsymbol{U}$. The initial boundary value problem for linear hyperbolic system writes

$$\begin{cases} \partial_t \boldsymbol{U} + \underline{\boldsymbol{A}} \partial_x \boldsymbol{U} & = \boldsymbol{0}, & x > 0, \quad t > 0 \\ \boldsymbol{U}(x,0) & = \boldsymbol{U}_0(x), & x > 0 \\ \underline{\boldsymbol{B}} \boldsymbol{U}(0,t) & = \underline{\boldsymbol{B}} \boldsymbol{g}(t), & t > 0 \end{cases} \tag{I.22}$$

The following theorem gives conditions for the well-posedness of eq. (I.22).

Theorem I.9 (Uniform Kreiss Condition for well-posedness [96]). *Consider the problem depicted in eq. (I.22). Let q be the number of strictly positive eigenvalues of the matrix $\underline{\boldsymbol{A}} \in \mathbb{R}^{p \times p}$. Denote the matrix $\underline{\boldsymbol{T}} \in \mathbb{R}^{p \times q}$ formed by the q eigenvectors of $\underline{\boldsymbol{A}}$ whose eigenvalues are strictly positive as columns. The initial boundary value problem is said well-posed if the matrix $\underline{\boldsymbol{B}} \in \mathbb{R}^{q \times p}$ is such that the matrix $\underline{\boldsymbol{B}\boldsymbol{T}} \in \mathbb{R}^{q \times q}$ is invertible.*

Remark I.2. In order to obtain a classical solution $\boldsymbol{U}$ to eq. (I.22), initial conditions and boundary conditions must belong to $\mathscr{C}^1$ and satisfy a compatibility relation, which writes as

$$\underline{\boldsymbol{B}} \boldsymbol{U}_0(0) = \underline{\boldsymbol{B}} \boldsymbol{g}(0), \quad \underline{\boldsymbol{B}} \partial_t \boldsymbol{g}(0) = -\underline{\boldsymbol{B}} \cdot \underline{\boldsymbol{A}} \partial_x \boldsymbol{U}_0(0). \tag{I.23}$$

The extension of theorem I.9 to multiple space-dimensions problem is known as the Uniform Kreiss-Lopantiskii Condition [97]. Non-linear hyperbolic system are not detailed here. Often one uses a quasi-linear form, assuming the matrix $\underline{\boldsymbol{A}}$ to be independent of $\boldsymbol{U}$ and applying the same theory as for linear systems. Much more can be said and proven for hyperbolic systems of conservation laws and initial boundary value problems. One may extend some of the previous definition and theorems to multiple space dimensions. Only a short overview of the main results concerning hyperbolic systems of conservation laws has been given. One may refer to [102, 106, 103, 48, 61] for more details on the subject. The problem of numerical approximations for hyperbolic system of conservation laws is now focused on.

I-1.2 Numerical methods for conservation laws and their properties

Two numerical methods for conservation laws are presented. General system of conservation laws in two dimensions on a bounded domain Ω takes the following form

$$\partial_t \boldsymbol{U} + \partial_x \boldsymbol{F}(\boldsymbol{U}) + \partial_y \boldsymbol{G}(\boldsymbol{U}) = \boldsymbol{0}, \quad t > 0, \quad (x, y) \in \Omega \tag{I.24}$$

Assume that there exists one entropy triplet (η, ζ, ξ) for the conservation laws in eq. (I.24). Two main numerical methods to solve conservation laws as in eq. (I.24) are distinguished in this part: finite difference schemes and finite volume schemes.

It is of great interest to check if a scheme satisfies a certain number of properties:

i) consistency of the scheme

ii) linear stability for the Cauchy problem,

iii) linear stability for the initial boundary value problem,

iv) discrete conservation of $\boldsymbol{U}$,

v) discrete entropy inequalities.

These properties are detailed later on.

I-1.2.1 Space discretization for conservation laws

Two space discretizations for conservation laws, commonly used in the literature [62, 106, 47, 61] are considered. The finite difference formalism consists in a regular Cartesian repartition of points to discretize the bounded domain. With such a repartition of points, it is particularly convenient to use equally-spaced polynomial reconstruction. The name originates from the fact that space derivatives are computed using finite differences of the variables placed on the nodes. A possible extension of finite difference schemes is to consider finite volume schemes on regular Cartesian grids. For this kind of schemes, the control volumes are regular, equally spaced and of same size. More generally, the finite volume formalism consists in integrating the system of partial derivatives equation on control volumes. For conservation laws, the presence of the divergence greatly simplifies the numerical computation, transforming it into a numerical computation of fluxes on the control volumes boundaries.

Finite difference schemes

First, a uniform grid $\{x_i, y_j\}$ is considered in space such that

$$\begin{aligned} x_{i+1} - x_i &= \Delta x, \quad \forall \quad i \in [0 : N_x[\,, \\ y_{j+1} - y_j &= \Delta y, \quad \forall \quad j \in [0 : N_y[\,. \end{aligned} \tag{I.25}$$

We use the notation $\boldsymbol{U}_{i,j}^n$ for an approximation of $\boldsymbol{U}$ at time $t = t^n$ and at position $(x = x_i, y = y_j)$. Such a discretization of the space is depicted on fig. I.1 with the variables $\boldsymbol{U}_{i,j}^n$ positionned at each grid nodes (x_i, y_j).

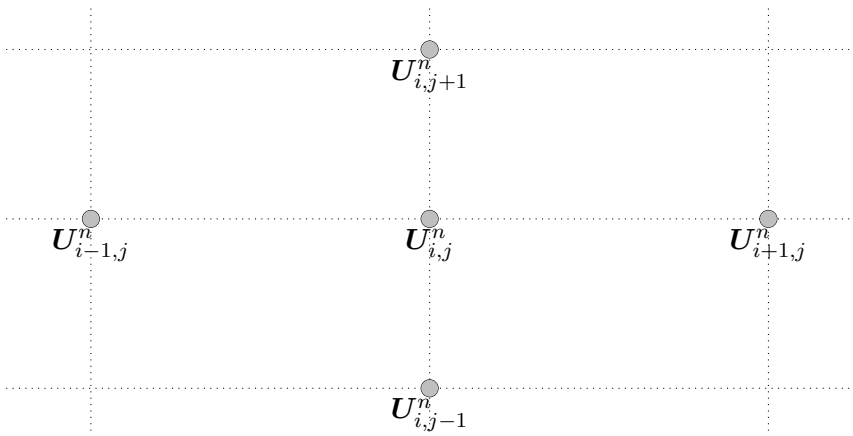


Figure I.1 – Space discretization for centered finite difference schemes on a Cartesian grid

Two kind of difference schemes are possible. The first one is based really on finite differences and the approximation of spatial derivative of $\boldsymbol{f}(\boldsymbol{U})$ and $\boldsymbol{g}(\boldsymbol{U})$. Those kind of schemes writes

$$\boldsymbol{U}_{i,j}^{n+1} = \boldsymbol{U}_{i,j}^n - \frac{t^{n+1} - t^n}{\Delta x} \underline{\boldsymbol{D_x}} \cdot \boldsymbol{f}(\boldsymbol{U}^n) - \frac{t^{n+1} - t^n}{\Delta y} \underline{\boldsymbol{D_y}} \cdot \boldsymbol{g}(\boldsymbol{U}^n) \tag{I.26}$$

where $\underline{\boldsymbol{D_x}}$ and $\underline{\boldsymbol{D_y}}$ are discrete approximations of respectively the x- and y-space derivatives. Considering hyperbolic systems of conservation laws, it is convenient to have a discrete conservation form of eq. (I.26). Indeed, for some discretization of space derivatives, one may rewrite eq. (I.26) under a conservative form as

$$\boldsymbol{U}_{i,j}^{n+1} = \boldsymbol{U}_{i,j}^n - \frac{t^{n+1} - t^n}{\Delta x} \left(\boldsymbol{f}_{i+\frac{1}{2},j}^\star - \boldsymbol{f}_{i-\frac{1}{2},j}^\star \right) - \frac{t^{n+1} - t^n}{\Delta y} \left(\boldsymbol{g}_{i,j+\frac{1}{2}}^\star - \boldsymbol{g}_{i,j-\frac{1}{2}}^\star \right). \tag{I.27}$$

Remark I.3. Any formulation as depicted in eq. (I.27) may be rewritten as in eq. (I.26). The reverse is untrue. Examples of (i, j)-dependent discretization of the space derivatives may yield to a non-conservative discretization.

Finite volume schemes on Cartesian grids

Keeping the notations for the grid, one defines a control volume with as a degree of freedom

the average value of $\boldsymbol{U}$ inside this control volume. This way, one rewrites any central difference schemes as a finite volume scheme on Cartesian grids. Finite volume schemes are based on an integration of eq. (I.24) over a control volume $\mathcal{K}$. It yields

$$\partial_t \int_{\mathcal{K}} \boldsymbol{U} \mathrm{d}V + \int_{\partial\mathcal{K}} (\boldsymbol{f}(\boldsymbol{U}) \cdot n_x + \boldsymbol{g}(\boldsymbol{U}) \cdot n_y) \mathrm{d}S = \boldsymbol{0}. \tag{I.28}$$

For finite volume schemes on Cartesian grids, one uses the following definition of the control volume denoted $\mathcal{K}_{i+\frac{1}{2},j+\frac{1}{2}}$

$$\mathcal{K}_{i+\frac{1}{2},j+\frac{1}{2}} =]x_i, x_{i+1}[\times]y_j, y_{j+1}[. \tag{I.29}$$

Denoting the average value of $\boldsymbol{U}$ over a control volume $\mathcal{K}_{i+\frac{1}{2},j+\frac{1}{2}}$ as $\overline{\boldsymbol{U}}_{i+\frac{1}{2},j+\frac{1}{2}}$ (see fig. I.2), it yields the following scheme for Cartesian grids

$$\overline{\boldsymbol{U}}^{n+1}_{i+\frac{1}{2},j+\frac{1}{2}} = \overline{\boldsymbol{U}}^{n}_{i+\frac{1}{2},j+\frac{1}{2}} - \frac{t^{n+1}-t^n}{\Delta x}\left(\boldsymbol{f}^\star_{i+1,j+\frac{1}{2}} - \boldsymbol{f}^\star_{i,j+\frac{1}{2}}\right) - \frac{t^{n+1}-t^n}{\Delta y}\left(\boldsymbol{g}^\star_{i+\frac{1}{2},j+1} - \boldsymbol{g}^\star_{i+\frac{1}{2},j}\right) \tag{I.30}$$

where $\boldsymbol{f}^\star$ and $\boldsymbol{g}^\star$ are the numerical fluxes at the boundary. Under this peculiar form, and considering vanishing fluxes at the boundary or periodic boundary conditions, by summing on every i and j, one immediately gets the conservation of $\boldsymbol{U}$.

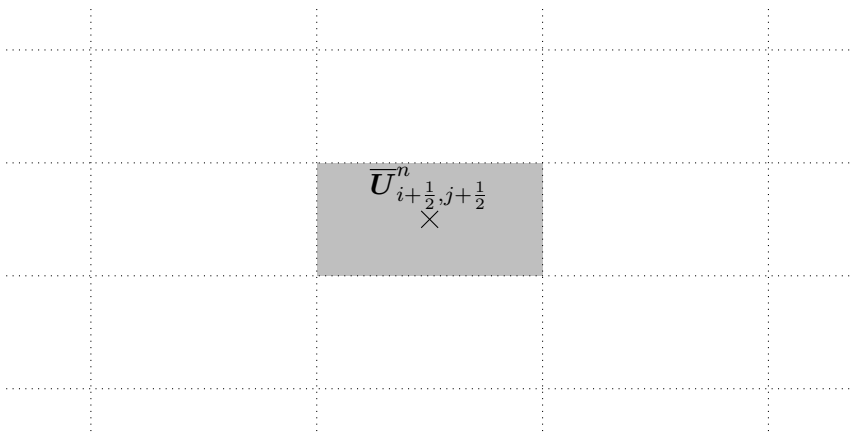


Figure I.2 – Space discretization for centered finite volume schemes on a Cartesian grid

Finite volume schemes on unstructured grids

Let $\mathscr{T}$ be a tessellation of the bounded domain in which eq. (I.24) is solved. The idea for finite volume on unstructured grids is to consider the control volumes as members of $\mathscr{T}$. An example of control volumes is depicted in fig. I.3. Using proposition I.5 and assuming that the normal outward the control volume is defined, a generic numerical conservative scheme writes

$$\overline{\boldsymbol{U}}^{n+1}_{\mathcal{K}} = \overline{\boldsymbol{U}}^{n}_{\mathcal{K}} - \frac{t^{n+1}-t^n}{|\mathcal{K}|} \sum_{\partial\mathcal{K}_q} |\partial\mathcal{K}_q| \left(\boldsymbol{f}^\star_{\partial\mathcal{K}_q}, \boldsymbol{g}^\star_{\partial\mathcal{K}_q}\right) \cdot \boldsymbol{n}_{\partial\mathcal{K}_q} \tag{I.31}$$

where $\boldsymbol{f}^\star$ and $\boldsymbol{g}^\star$ are the numerical fluxes at the boundary and $\boldsymbol{n}_{\partial\mathcal{K}_q}$ the normal to $\partial\mathcal{K}_q$ outward $\mathcal{K}$. Under this peculiar form, and considering vanishing fluxes at the boundary or periodic boundary conditions, by summing for every $\mathcal{K}$ in $\mathscr{T}$, one immediately gets the conservation of $\boldsymbol{U}$.

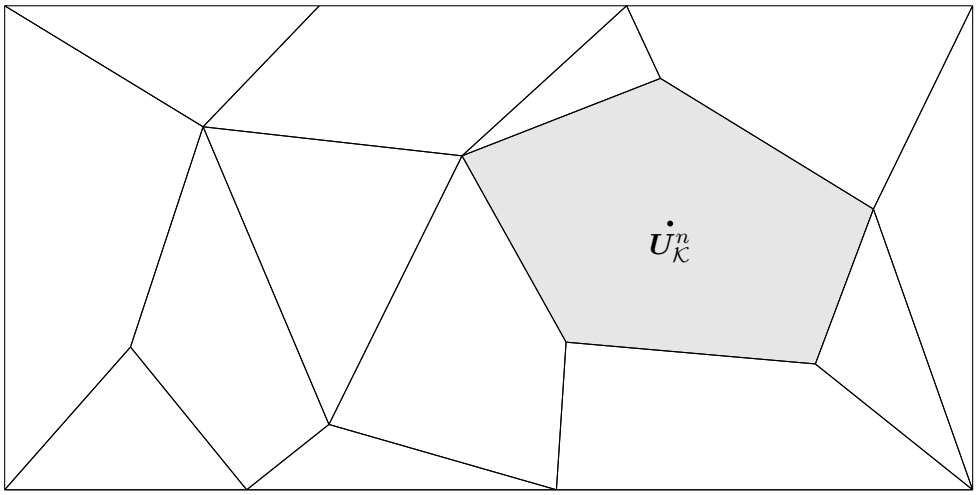


Figure I.3 – Space discretization for finite volume schemes on an unstructured grid

I-1.2.2 Convergence and consistency of numerical schemes

Convergence of a numerical scheme is a most desired property for a given scheme. Briefly, convergence means that as the time step and mesh size tend toward zero, the approximated solution gets closer to the real solution. A definition of convergence is introduced as follows

Definition I.4 (Convergence of a finite difference approximation [2])**.** A finite difference scheme approximating a partial differential system is **convergent** if for any solution to the partial differential equation $\boldsymbol{U}(x,t)$ and solutions to the finite difference schemes $\boldsymbol{U}_i^n$ such that $\boldsymbol{U}_i^0$ converges to the initial condition $\boldsymbol{U}(x,0) = \boldsymbol{U}_0(x)$, $\boldsymbol{U}_i^n$ converges to $\boldsymbol{U}(x,t)$ as $(i\Delta x, n\Delta t)$ converges toward (x,t) as $\Delta t, \Delta x$ tend to 0.

In order to get convergence of a numerical scheme, two important properties are consistency and stability. Briefly, the consistency property stands for saying that as the mesh in space and time is refined, the error between the solution to the continuous system and the approximated solution goes to zero. Consistency is defined as

Definition I.5 (Consistency of a finite difference approximation [2])**.** Let $\mathscr{P}\boldsymbol{U} = 0$ be a partial differential system approximated by a finite difference scheme denoted $\mathscr{P}_{\Delta x,\Delta t}$. The finite difference scheme is **consistent** with the partial differential system if for any smooth function $\boldsymbol{\phi}$,

$$\lim_{\Delta x,\Delta t\to 0} \mathscr{P}\boldsymbol{\phi} - \mathscr{P}_{\Delta x,\Delta t}\boldsymbol{\phi} = 0, \tag{I.32}$$

The norm (uniform convergence) is precised in [2].

Remark I.4. In order to show consistency of a numerical scheme, one often shows, using a Taylor expansion of a smooth function ϕ, that

$$\mathscr{P}\boldsymbol{\phi} - \mathscr{P}_{\Delta x,\Delta t}\boldsymbol{\phi} = \mathscr{O}(\Delta t^{\alpha} + \Delta x^{\beta}), \quad \alpha > 0, \quad \beta > 0. \tag{I.33}$$

It gives both the consistency and the accuracy of a numerical scheme.

Definition I.6 (Consistency of a flux in a finite volume approximation [61])**.** Assume a finite volume scheme which writes under the form

$$\boldsymbol{U}_i^{n+1} - \boldsymbol{U}_i^n + \frac{t^{n+1} - t^n}{\Delta x}\left[\boldsymbol{f}^{\star}_{i+\frac{1}{2}} - \boldsymbol{f}^{\star}_{i-\frac{1}{2}}\right] = \boldsymbol{0}. \tag{I.34}$$

Let $\boldsymbol{f}^{\star}_{i+\frac{1}{2},j}$ write as a vector valued function Φ, with $(r,p) \in \mathbb{N}^2$ such that

$$\boldsymbol{f}^{\star}_{i+\frac{1}{2}} = \boldsymbol{\Phi}(\boldsymbol{U}^n_{i-p+1}, ..., \boldsymbol{U}^n_{i+r}), \quad \forall i \in \mathbb{Z}, \quad \forall n \in \mathbb{N}. \tag{I.35}$$

Then if $\boldsymbol{\Phi}$ satisfies

$$\boldsymbol{\Phi}(\boldsymbol{U}, ..., \boldsymbol{U}) = \boldsymbol{f}(\boldsymbol{U}), \tag{I.36}$$

the flux is said consistent.

Definition I.7 (Weak consistency [46])**.** Consider a given numerical scheme for the discretization of eqs. (I.1) and (I.7). Assume that the numerical solution, denoted $\boldsymbol{U}_{\Delta x}$ is bounded in $(L^{\infty})^N$. Moreover assume that there exists $\widehat{\boldsymbol{U}} \in (L^{\infty})^N$ such that $\boldsymbol{U}_{\Delta x}$ converges toward $\widehat{\boldsymbol{U}}$ in $(L^1_{\text{loc}})^N$. If $\widehat{\boldsymbol{U}}$ is a weak solution in the sense of definition I.1 to eqs. (I.1) and (I.7), then the scheme is weakly consistent.

Remark I.5. A practical criterion for weak consistency is to show that the flux is consistent [106]. See also [46, 61].

Proving only consistency of a numerical scheme does not prove its convergence. As will be shown in section I-1.2.4, consistency alone is not enough. The concept of stability is needed to ensure convergence for linear systems. Although a scheme may be consistent, truncation error may stack over time and induce larger and larger errors. Stability is closely related to the property of the numerical schemes to deal with numerical errors. If a scheme has a tendency to increase at each time step the numerical errors made on the previous ones, then it is unstable. In a finite time, the numerical errors become preponderant over the approximation and the computations are not relevant anymore. As a contrary, if the numerical errors stay constant or even better if they are damped out by the finite difference schemes, it is then stable. In 1928, Courant, Frierichs and Lewy [32] formulated the fundamental CFL condition, that strongly links the time step to the mesh width to ensure quadratic stability. In order to define the notion of quadratic stability, the definition of the quadratic norms are first introduced

Definition I.8 (Discrete l^2 norms)**.** For a given sequence $\phi = (\phi_i)_{i\in\mathbb{Z}}$ on an infinite grid, the l^2 norm in space is defined as

$$\|\phi\|^2_{l^2(\mathbb{Z})} = \sum_{i\in\mathbb{Z}} \Delta x |\phi_i|^2. \tag{I.37}$$

For $\phi = (\phi^n)_{n\in\mathbb{N}} = (\phi_i^n)_{i\in\mathbb{Z},n\in\mathbb{N}}$, the l^2 norm in space and time is defined as

$$\|\phi\|^2_{l^2(\mathbb{Z}),l^2(\mathbb{N})} = \sum_{n\in\mathbb{N}} \Delta t \sum_{i\in\mathbb{Z}} \Delta x |\phi_i^n|^2. \tag{I.38}$$

Definition I.9 (Quadratic stability [2])**.** A finite difference scheme $\mathscr{P}_{\Delta x,\Delta t}$ is **stable** for the quadratic norm and for numerical parameters $(\Delta x, \Delta t) \in \Lambda \subset \mathbb{R}^2$, if there exists an integer N such that for any non-negative time T, there exists a constant C_T which depends only on T such that for $\phi \in l^2(l^2(\mathbb{Z}), \mathbb{N})$ satisfying $\mathscr{P}_{\Delta x,\Delta t}\phi = 0$

$$\|\phi^n\|_{l^2(\mathbb{Z})} \leq C_T \sum_{k=0}^{N} \|\phi^k\|_{l^2(\mathbb{Z})}, \quad \forall(\Delta x, \Delta t) \in \Lambda, 0 \leq n\Delta t \leq T \tag{I.39}$$

is satisfied.

Often, the stability criteria used for numerical scheme is stronger than the one proposed in definition I.9. Indeed, the previous stability criteria is quite difficult to prove, in general. Instead one would rather use the following one.

Definition I.10 (Von Neumann's stability [2])**.** A finite difference scheme $\mathscr{P}_{\Delta x,\Delta t}$ is **stable** in the sense of Von Neumann for numerical parameters $(\Delta x, \Delta t) \in \Lambda \subset \mathbb{R}^2$, if for any non-negative time T such that for $\phi \in l^2(l^2(\mathbb{Z}), \mathbb{N})$ satisfying $\mathscr{P}_{\Delta x,\Delta t}\phi = 0$

$$\|\phi\|_{l^2(\mathbb{Z})} \leq \|\phi^0\|_{l^2(\mathbb{Z})}, \quad \forall(\Delta x, \Delta t) \in \Lambda, \quad 0 \leq n\Delta t \leq T, \tag{I.40}$$

is satisfied. It is equivalent to

$$|||\mathscr{P}_{\Delta x,\Delta t}||| \leq 1, \quad \forall(\Delta x, \Delta t) \in \Lambda. \tag{I.41}$$

Analytic and numerical methods to check stability and determine stability regions are proposed in the next section. Studies focus only on the Von Neumann's criteria.

I-1.2.3 Linear stability analysis of numerical schemes

The Von Neumann's stability analysis has been derived to check stability regions for linear finite difference schemes. First, a stability analysis for linear finite difference schemes with periodic boundary conditions is proposed. This is the so-called Von Neumann stability analysis (also known as the Fourier stability analysis). Second, the analysis of stability for finite difference schemes with non-periodic boundary conditions is detailed.

Stability analysis for the Cauchy problem

Stability analysis for the Cauchy problem with linear partial differential equations is often performed using the Von Neumann stability analysis. The analysis is based on the Fourier decom-

position of the numerical error. It was developed by Von Neumann in the 40s, but only first briefly introduced in [33]. It was then extended in a more theoretical way in [22]. One may also refer to the textbook by Allaire [2].

Consider a finite difference scheme $\mathscr{P}_{\Delta x,\Delta t}$. The approximated solution $(u_i^n)_{i\in\mathbb{Z},n\in\mathbb{N}}$ satisfies $\mathscr{P}_{\Delta x,\Delta t}u = 0$. Considering periodic boundary conditions, one may decompose (u_i^n) as a Fourier serie in space. Up to a change of variables, one may estimate that the space interval of periodicity has a length equal to 1. Moreover, one makes the assumptions that u has an exponential growth or decay in time defined by a constant $\alpha \in \mathbb{C}$. It yields that

$$u(x,t) = \mathrm{e}^{\alpha t} \sum_{k\in\mathbb{Z}} \psi_k \, \mathrm{e}^{ik\pi x}, \psi \in l^2(\mathbb{Z}). \tag{I.42}$$

Let us define the sequence (ϵ_k) as

$$\epsilon_k(x,t) = \mathrm{e}^{\alpha t} \, \mathrm{e}^{ik\pi x}, \quad k \in \mathbb{Z}. \tag{I.43}$$

It is sufficient to consider the growth of ϵ_k for any k to get the growth of u, as the series behave as its terms. To alleviate the notation, the index k is dropped. The notation j is used for the space index in order not to introduce any confusion with the complex number i. One may notice the following relations for the discretized version of ϵ_k denoted $\epsilon_{k,j}^n$.

$$\left\{ \begin{array}{lll} \epsilon_{k,j}^n & =\epsilon_k(j\Delta x, n\Delta t) & =\mathrm{e}^{\alpha n\Delta t} \, \mathrm{e}^{ik\pi j\Delta x} \\ \epsilon_{k,j}^{n+1} & =\mathrm{e}^{\alpha(n+1)\Delta t} \, \mathrm{e}^{ik\pi j\Delta x} & =\mathrm{e}^{\alpha\Delta t} \, \epsilon_{k,j}^n \\ \epsilon_{k,j+m}^n & =\mathrm{e}^{\alpha n\Delta t} \, \mathrm{e}^{ik\pi(j+m)\Delta x} & =\mathrm{e}^{ik\pi m\Delta x} \, \epsilon_{k,j}^n \end{array} \right. \tag{I.44}$$

The amplification factor is introduced as a function of $\theta = k\pi\Delta x$, Δx and Δt as

$$G(\theta, \Delta x, \Delta t) = \frac{\epsilon_{k,j}^{n+1}}{\epsilon_{k,j}^n} = \mathrm{e}^{\alpha\Delta t}. \tag{I.45}$$

Values taken by the amplification factor G determine the stability of the schemes. Linear stability via amplification factor study is defined in definition I.11.

Definition I.11 (Amplification factor and stability [2]). A finite difference scheme $\mathscr{P}_{\Delta x,\Delta t}$ with constant coefficients is stable for numerical parameters $(\Delta x, \Delta t) \in \Lambda \subset \mathbb{R}^2$ if and only if there exists a constant C which is independent of θ, Δx, Δt such that its amplification factor satisfies

$$|G(\theta, \Delta x, \Delta t)| \leq 1 + C\Delta t, \quad \forall\theta \in [0 : 2\pi]. \tag{I.46}$$

Furthermore, the restricted stability conditions yields

$$|G(\theta, \Delta x, \Delta t)| \leq 1, \quad \forall\theta \in [0 : 2\pi]. \tag{I.47}$$

Let us take $\mathscr{P}_{\Delta x,\Delta t}$ a one-step finite difference scheme for a scalar conservation laws. Assume it

writes under the following form

$$u_j^{n+1} - u_j^n - \frac{\Delta t}{\Delta x} \sum_{m=-p}^{r} C_m u_{j+m}^n = 0.$$

which yields using eq. (I.44), that

$$G(\theta, \Delta x, \Delta t) = 1 + \frac{\Delta t}{\Delta x} \sum_{m=-p}^{r} C_m \, \mathrm{e}^{im\theta}$$

Assuming that Δt and Δx are proportional with a given constant λ, it yields

$$G(\theta, \Delta x, \lambda \Delta x) = 1 + \lambda \sum_{m=-p}^{r} C_m \, \mathrm{e}^{im\theta} \, .$$

One checks analytically or numerically that $G(\theta, \Delta x, \lambda \Delta x) \leq 1, \quad \theta \in [0 : 2\pi]$ to determine Von Neumann's stability for a given λ as for this example G is independent of Δx.

Stability analysis for the initial value boundary problem

The normal mode analysis for linear hyperbolic equation was devised and introduced in [63] and extended in [96] and [130]. The condition called the Godunov-Ryabenkii gives necessary condition for stability, and so not always sufficient. Works presented in [76] develop sufficient conditions for stability, called the GKS theory in a fully discrete version (the semi-discrete case was dealt later with [150]). The essence of their work is presented in the following propositions. Consider the problem depicted in eq. (I.22) with appropriate boundary conditions according to the uniform Kreiss condition. First, semi-discrete case for linear hyperbolic equation is considered and later extended to the fully discrete case.

Consider a semi-discrete finite difference approximation $\mathscr{Q}_{\Delta x}$ and a boundary operator $\mathscr{D}$ such that

$$\left\{ \begin{array}{ll} \partial_t u_j & = \mathscr{Q}_{\Delta x} u_j^n, \quad j \geq 1, \\ \mathscr{D} u_j & = g_j, \quad -r \leq j \leq 0. \end{array} \right. \tag{I.48}$$

Performing a Laplace transform ($u(x,t) = \mathrm{e}^{st} \, \phi(x)$) in the time variable on eq. (I.48), multiplying by Δx and using $\widehat{s} = s\Delta x$ yield

$$\left\{ \begin{array}{ll} \widehat{s}\widehat{u}_j & = \Delta x \mathscr{Q}_{\Delta x} \widehat{u}_j, \quad j \geq 1, \\ \mathscr{D} \widehat{u}_j & = g_j, \quad -r \leq j \leq 0. \end{array} \right. \tag{I.49}$$

The Godunov-Ryabenkii condition writes

Lemma I.10 (Godunov–Ryabenkii condition [62]). *Consider eq.* (I.48) *with a zero boundary condition. A* ***necessary*** *condition for stability is that there exists no nontrivial eigenvector* $\widehat{u}$ *associated to an eigenvalue* $\widehat{s}$ *with* $\Re(\widehat{s}) > 0$ *of eq.* (I.49).

In order to introduce the GKS theory in both semi-discrete and fully discrte form, the definition of generalized eigenvector if firstly given.

Definition I.12 (Generalized eigenvector for the semi-discrete problem [168]). The sequence $\{\widehat{u}_j(\widehat{s})\}$ is an eigenvector if:

1. it is not identically 0,
2. It satisfies eq. (I.49),
3. $\Re(\widehat{s}) \geq 0$ and
   — for $\Re(\widehat{s}) > 0$, the corresponding solution satisfies $\lim_{j\to\infty} \widehat{u}_j(\widehat{s}) = 0$,
   — for $\Re(\widehat{s}) = 0$, let $\widehat{s}_0 = \lim_{\epsilon\to 0^+} \widehat{s} + \epsilon$. Then $\{\widehat{u}_j(\widehat{s_0})\}$ is an eigenvector.

The GKS theory provides the following results concerning semi-discrete schemes.

Lemma I.11 (Semi-discrete GKS condition [150]). *Consider eq.* (I.49) *with a zero boundary condition. A* ***sufficient*** *condition for stability of eq.* (I.48) *is that there exists no generalized eigenvector* $\widehat{u}$ *for* $\Re(\widehat{s}) \geq 0$ *in the sense of definition I.12.*

For fully discrete case, consider a finite difference approximation $\mathscr{Q}_\nu$, with $\nu = \frac{\Delta t}{\Delta x}$ and a boundary operator $\mathscr{D}$ such that

$$\begin{cases} u_j^{n+1} - u_j^n &= \mathscr{Q}_\nu u_j^n, \quad j \geq 1, \\ \mathscr{D} u_j^n &= g_j, \quad -r \leq j \leq 0. \end{cases} \tag{I.50}$$

Then, taking the discrete Laplace as $u_j^n = z^n u_j$, one gets the fully discrete problem with Laplace transform as

$$\begin{cases} (z-1)\widehat{u}_j &= \mathscr{Q}_\nu \widehat{u}_j, \quad j \geq 1, \\ \mathscr{D}\widehat{u}_j &= g_j, \quad -r \leq j \leq 0. \end{cases} \tag{I.51}$$

We introduce the definition of generalized eigenvector for the fully discrete problems.

Definition I.13 (Generalized eigenvector for the fully discrete problem [172]). Let $|z| \geq 1$. The sequence $\{\widehat{u}_j(z)\}$ is an eigenvector if

1. it is non identically 0,
2. it satisfies eq. (I.51),
3. $\|\widehat{u}(z)\|_{l^2} < \infty$ for $|z| > 1$.

The sequence $\{\widehat{u}_j(z)\}$ is a generalized eigenvector if

1. it is non identically 0,
2. it satisfies eq. (I.51),
3. $\|\widehat{u}(z)\|_{l^2} = \infty$. Furthermore, $\widehat{u}(z) = \lim_{\theta\to z, |\theta|>1} \widehat{u}(\theta)$ and $\widehat{u}(\theta)$ satisfies $(\theta-1)\widehat{u}_j(\theta) = \mathscr{Q}_\nu \widehat{u}_j(\theta)$.

It yields in peculiar the following GKS condition for fully discrete scheme.

Lemma I.12 (Fully discrete GKS condition [76, 172]). *Consider eq.* (I.51) *with a zero boundary conditions. A* ***sufficient*** *condition for stability of eq.* (I.50) *is that there exists no generalized eigenvector* $\widehat{u}$ *for* $|z| \geq 1$ *in the sense of definition I.13.*

Further works by Wu and later by Coulombel [172, 31, 30, 29] have been done in order to change the resolvent estimates into semi-groupe stability estimates. Goldberg and Tadmor introduced stability criteria for a particular class of numerical schemes [64, 65, 66, 67]. See also [75] for a special link between the Godunov-Ryabenkii conditions for stability and the GKS theory. Last, the summation by part technique introduced by Olsson give energy estimates and hence stability using special structure of operator at the boundary [128, 129].

I-1.2.4 Convergence toward a weak solution

Convergence for linear systems using finite difference methods

The Lax–Richtmyer equivalence theorem is from [105]. Its applicability is restricted to the special case of linear numerical methods for well-posed linear partial differential equations. It states that

Theorem I.13 (Lax–Richtmyer equivalence theorem [105]). *A consistent finite difference method for a well-posed linear initial value problem is convergent if and only if it is stable.*

One can easily summarized the theorem with

$$\boxed{\text{linear,} \quad \text{consistency} \quad + \quad \text{stability} \quad \Longleftrightarrow \quad \text{convergence.}}$$

However, as indicated, the scope of applications of this theorem is restricted to linear partial differential equation systems. Stability and consistency are often not enough to imply convergence for a non-linear system. To deal with non-linearity, the Lax–Wendroff theorem for non-linear hyperbolic systems of conservation laws is introduced.

Convergence for a non-linear hyperbolic system of conservation laws

The Lax–Wendroff theorem has been presented and proved in [106]. It may be seen as an extension of the Lax–Richtmyer equivalence theorem for the non-linear hyperbolic system of conservation laws. It states about sufficient conditions to ensure convergence of the numerical scheme toward a weak solution. If a consistent, stable and conservative numerical scheme for eq. (I.1) converges toward a solution, then it converges toward a weak solution of eq. (I.1). Consider a consistent finite volume scheme in the sense of definition I.6. Consider that $(\boldsymbol{U}_j^0)_{j\in\mathbb{Z}}$ satisfies the initial condition prescribed in eq. (I.7). Then as Δt and Δx tend to zero, under certain hypothesis, the limit $\boldsymbol{U}$ is a weak solution of eq. (I.1) for the initial conditions $\boldsymbol{U}_0$.

Theorem I.14 (Lax–Wendroff theorem [106]). *Let $\boldsymbol{U}_{\Delta x}(x,t)$ be a numerical solution obtained on a given grid whose width is Δx. If*

i) $\boldsymbol{U}_{\Delta x}$ *is uniformly bounded in* Δx *in* L^∞,

ii) $\lim\limits_{\Delta x\to 0} \|\boldsymbol{U}_{\Delta x} - \boldsymbol{U}\|_{L^1}$,

iii) $\boldsymbol{U}_{\Delta x}$ *is obtained using the formulation presented in eq.* (I.34) *and* $\boldsymbol{\Phi}$ *satisfying eq.* (I.36).

Then the limit solution $\boldsymbol{U}$ *is a weak solution of eq.* (I.1) *with initial conditions prescribed in eq.* (I.7)

The theorem can be summarized as

$$\boxed{\text{convergence} + \text{consistency} + \text{stability} + \text{conservation} \implies \boldsymbol{U} \text{ is a weak solution.}}$$

As a contrary to the Lax–Richtmyer equivalence theorem, there is no equivalence in the Lax–Wendroff theorem, only an implication. A non-linear scheme which converges toward a weak solution may not be conservative or stable. Furthermore, the Lax–Richtmyer theorem gives convergence results for linear problem using stability and consistency. Whereas the Lax–Wendroff theorem assumes convergence, stability, consistency and conservation to yield convergence toward a weak solution. Theorem I.14 can be extended to unstructured grid based finite volume scheme (see [47]).

I-1.2.5 Convergence toward the entropic solution for scalar conservation laws

For scalar conservation laws, one can prove that the numerical scheme under the Lax–Wendroff hypothesis and a consistency with the entropic condition converges toward the entropic solution. The proof is done in [154]. The theorem states that if the scheme satisfies a discrete entropy inequality, then the limit solution u is the entropic solution of the scalar conservation law.

Definition I.14 (Entropy condition consistency [47])**.** A finite difference or finite volume scheme is consistent with the entropy inequality if for any entropic pair (η, ζ) there exists an entropic flux function Ξ satisfying

$$\Xi(u, ..., u) = \zeta(u), u \in \mathscr{U}.$$

such that for a scheme which writes as eq. (I.34), the discrete entropic inequality

$$\eta(u_j^{n+1}) - \eta(u_j^n) + \frac{\Delta t}{\Delta x}\left[\Xi(u_{j-p+1}^n, ..., u_{j+r}^n) - \Xi(u_{j-p}^n, ..., u_{j+r-1}^n)\right] \leq 0, \quad \forall j \in \mathbb{Z}, \quad \forall n \in \mathbb{N} \quad \text{(I.52)}$$

holds.

This definition gives a completion to the Lax–Wendroff theorem for and only for scalar conservation laws. Under entropic condition consistency, a scalar numerical scheme satisfying the hypothesis of the Lax–Wendroff theorem converges toward the entropic solution.

Theorem I.15 (Existence and uniqueness of the entropic solution [47])**.** *Let* $\boldsymbol{U}_{\Delta x}(x, t)$ *be a numerical solution obtained on a given grid whose width is* Δx *satisfying the aforementioned hypothesis of theorem I.14. If moreover the scheme presented in eq.* (I.34) *is consistent with the entropy condition (see definition I.14) and that for any entropy function* η*, the numerical flux* Ξ *is at least Lipschitz continuous, then the limit solution* u *is the unique entropic solution of the Cauchy problem formed with the scalar conservation law eq.* (I.1) *and the initial condition prescribed in eq.* (I.7)*.*

I-2 Numerical methods for compressible hydrodynamics

This section is devoted to an overview of numerical methods for the approximation of the Euler equations in multidimensional space. These numerical methods are first classified into two large families. The first one is called the Lagrangian or Arbitrary Lagrangian Eulerian family of methods. The underlying tessellation is deformed along the computation. The second family is the high-order methods on fixed grids, whether Cartesian or unstructured. Before any further details concerning numerical methods for compressible hydrodynamics, the Euler and Lagrange equations are reminded.

I-2.1 Euler and Lagrange equations for compressible hydrodynamics

Euler compressible hydrodynamics equations stand for the approximation of inviscid compressible flows. The variables are the density ρ, the velocity field $\boldsymbol{u}$ and the total energy e. Moreover it is convenient to use also the definitions of internal energy ϵ and specific volume τ as

$$\left\{\begin{array}{ll} \epsilon & = e - \frac{1}{2}\|\boldsymbol{u}\|^2 \\ \tau & = \frac{1}{\rho} \end{array}\right. \tag{I.53}$$

The Euler system writes in the absence of any source terms in $\mathbb{R}^d$

$$\partial_t \begin{pmatrix} \rho \\ \rho\boldsymbol{u} \\ \rho e \end{pmatrix} + \nabla \cdot \begin{pmatrix} \rho\boldsymbol{u} \\ \rho\boldsymbol{u}\otimes\boldsymbol{u} + p\underline{\boldsymbol{I}} \\ (\rho e + p)\boldsymbol{u} \end{pmatrix} = \mathbf{0}. \tag{I.54}$$

The convex set of states $\mathscr{U}$ writes [61]

$$\mathscr{U} = \{(\rho,\ \boldsymbol{q} = \rho\boldsymbol{u},\ E = \rho e) \text{ s.t. } \rho > 0,\ \boldsymbol{q} \in \mathbb{R}^d,\ E - \frac{\|q\|^2}{2\rho} > 0\},$$

which means that the density is non-negative as well as the internal energy. The system is closed with an equation of state which links pressure, internal energy and specific volume as

$$p = \mathrm{EOS}(\tau, \epsilon). \tag{I.55}$$

I-2.1.1 Euler and Lagrange systems in 1D

In one space dimension, the Euler system writes

$$\partial_t \begin{pmatrix} \rho \\ \rho u \\ \rho e \end{pmatrix} + \partial_x \cdot \begin{pmatrix} \rho u \\ \rho u^2 + p \\ (\rho e + p)u \end{pmatrix} = \mathbf{0}. \tag{I.56}$$

The Lagrangian system is deduced from eq. (I.56) with an appropriate change of variables.

Introducing a change of variables $(x,t) \to (X,t)$ defined as

$$\mathrm{d}x(X,t) = J(X,t)\mathrm{d}X + u(X,t)\mathrm{d}t, \tag{I.57}$$

where J is the Jacobian of the deformation and satisfies $J = \partial_X x(X,t)$. One gets the following result concerning the material derivative of J

$$\mathrm{D}_t J(X,t) = J(x,t)\partial_x u(x,t). \tag{I.58}$$

Then for any smooth enough function ϕ, one gets the following derivatives rules

$$\begin{cases} \mathrm{D}_t\phi(X,t) &= \partial_t\phi(x,t) + u(x,t)\partial_x\phi(x,t), \\ \partial_X\phi(X,t) &= J(x,t)\partial_x\phi(x,t), \\ \mathrm{D}_t(J\phi)(X,t) &= \left[J\partial_t\phi + u\partial_x\phi u\right](x,t). \end{cases} \tag{I.59}$$

Using eq. (I.59), one get the following lemma

Lemma I.16 (Euler-Lagrange change of variables)**.** *For any couple of smooth enough function (ϕ,ψ), the change of variables $(x,t) \to (X,t)$ yields*

$$\left[\mathrm{D}_t(J\phi) + \partial_X\psi\right](X,t) = \left[J\partial_t\phi + \partial_x(\phi u + \psi)\right](x,t).$$

Then using lemma I.16 in eq. (I.56), one gets the 1D Lagrange equations. It writes

$$\mathrm{D}_t \begin{pmatrix} \rho_0\tau \\ \rho_0 u \\ \rho_0 e \end{pmatrix} + \partial_X \begin{pmatrix} -u \\ p \\ pu \end{pmatrix} = \mathbf{0}. \tag{I.60}$$

Using the definition of internal energy as the difference between the total energy and the kinetic energy, it yields an hyperbolic system with a non-conservative form as

$$\begin{cases} \mathrm{D}_t\rho_0\tau - \partial_X u &= 0 \\ \mathrm{D}_t\rho_0 u + \partial_X p &= 0 \\ \mathrm{D}_t\rho_0\epsilon + p\partial_X u &= 0 \end{cases} \tag{I.61}$$

Note that eq. (I.60) is well defined in the sense of distribution for any $(\tau,u,e) \in L^\infty{}_{\mathbb{R}\times]0:T[}$. As a contrary, eq. (I.61) is not. The term $p\partial_X u$ is well-defined for smooth enough functions, but is not in general in the sense of distributions. In [36], the authors introduced a generalization of the notion of weak solution in the sense of distributions despite non-conservative products. The generalization is based on the integration along a conservative path. However, in [1], the authors produced a comment on the computation of non-conservative products. Despite the integration along the conservative path, numerical results thus obtained are not conclusive. Discretization of non-conservative products has tremendous consequences for schemes solving eq. (I.61). It is shown later for the special case of hydrodynamics.

I-2.1.2 Entropic relations for the 1D Lagrange system

Let us focus on the entropy introduced by the second principle of the thermodynamics presented in theorem I.17.

Theorem I.17 (Second principle of thermodynamics). *For a closed system, without any exchange with the exterior, the entropy of a system increases over time or stays constant.*

Remark I.6. The entropy stays constant for reversible processes. In particular, for smooth flows, the entropy is conserved.

Introducing the concave entropy function S, the temperature T, the second principle of thermodynamics writes for the compressible hydrodynamics

$$T\mathrm{d}S = \mathrm{d}\epsilon + p\mathrm{d}\tau. \tag{I.62}$$

In particular, one gets for smooth quadruplet (ϵ, p, τ, u) that

$$\begin{aligned} T\mathrm{D}_t S &= \mathrm{D}_t\epsilon + p\mathrm{D}_t\tau \\ &= -p\partial_X u + p(\partial_X u) \\ &= 0 \end{aligned} \tag{I.63}$$

meaning that for smooth flows and non-zero temperature, the entropy indeed stays constant in time.

More generally, for any flows which may include discontinuities, the entropy satisfies

$$T\mathrm{D}_t S \geq 0. \tag{I.64}$$

A first point of view, that will be detailed later on, to ensure increasing of entropy is the use of pseudo-viscous forces. On the continuous level, it forces the evolution of internal energy to satisfy $\mathrm{D}_t\rho_0\epsilon + (p+q)\partial_X u = 0$, where q is called the pseudo-viscosity or artificial viscosity. Then, if one assumes that $q = -\phi\partial_X u,\ \phi \geq 0$, then eq. (I.63) becomes formally

$$\begin{aligned} T\mathrm{D}_t S &= \mathrm{D}_t\epsilon + p\mathrm{D}_t\tau \\ &= -(p+q)\partial_X u + p(\partial_X u) \\ &= -q\partial_X u \\ &= \phi|\partial_X u|^2 \geq 0. \end{aligned} \tag{I.65}$$

The choice of artificial viscosity is detailed in section I-2.4. Note that this result is based on formal computations at the continuous level, and does not imply results on the discretized one.

In [42], Després derived a canonical formulation for Lagrangian systems of conservation laws, assuming a zero entropy flux, Galilean invariance and isentropy for smooth solutions.

Equation (I.64) often yields a natural CFL condition for the numerical scheme, in order to satisfy a correct increase of entropy. Moreover, one challenging problem for numerical simulation

containing shocks is to control the increase of entropy, but also to ensure that the entropy does not increase on smooth flows. Assuming that the function S is concave. Using similar computations as in [45], for $\boldsymbol{U} \in \mathscr{U}$, let $g(\alpha) = S(\boldsymbol{U}_j^n + \alpha(\boldsymbol{U}_j^{n+1} - \boldsymbol{U}_j^n))$. Then, there exists $\theta \in]0:1[$ such that

$$g(1) = g(0) + g'(1) - \frac{1}{2}g''(\theta).$$

By definition of g, one has that

$$\left\{ \begin{array}{lcl} g'(1) & = & \nabla_{\boldsymbol{U}} S(\boldsymbol{U}_j^{n+1}) \cdot (\boldsymbol{U}_j^{n+1} - \boldsymbol{U}_j^n), \\ g''(\theta) & = & (\boldsymbol{U}_j^{n+1} - \boldsymbol{U}_j^n) \cdot \left(\nabla^2_{\boldsymbol{U}} S(\boldsymbol{U}_j^{n+1}(\boldsymbol{U}_j^{n+1} - \boldsymbol{U}_j^n)\right). \end{array} \right. \tag{I.66}$$

Using the concavity of S, it gives that $-\frac{1}{2}g''(\theta) \geq 0$. Then, it leads to

$$S(\boldsymbol{U}_j^{n+1}) = S(\boldsymbol{U}_j^n) + \nabla_{\boldsymbol{U}} S(\boldsymbol{U}_j^{n+1}) \cdot (\boldsymbol{U}_j^{n+1} - \boldsymbol{U}_j^n) - \frac{1}{2}(\boldsymbol{U}_j^{n+1} - \boldsymbol{U}_j^n) \cdot \left(\nabla^2_{\boldsymbol{U}} S(\boldsymbol{U}_j^{n+1}(\boldsymbol{U}_j^{n+1} - \boldsymbol{U}_j^n)\right) \tag{I.67}$$

Assume (as for the example detailed in [45]) that previous equation rewrites under the form

$$S(\boldsymbol{U}_j^{n+1}) = S(\boldsymbol{U}_j^n) + (A - \frac{\Delta t}{\Delta X} B), \tag{I.68}$$

where A is a quadratic positive form evaluated on $(\boldsymbol{U}_j^{n+1} - \boldsymbol{U}_j^n)$, whereas B is also a positive quadratic form evaluated on $(\boldsymbol{\psi}_j^{n+1} - \boldsymbol{\psi}_j^n)$. Then assuming that the function $\boldsymbol{U} \mapsto \boldsymbol{\psi}$ is continuous, there exists a constant $c > 0$ such that

$$\|\boldsymbol{\psi}_j^{n+1} - \boldsymbol{\psi}_j^n\| \leq c\|\boldsymbol{U}_j^{n+1} - \boldsymbol{U}_j^n\|.$$

Then for $\nu = \frac{\Delta t}{\Delta X}$ small enough, one has $(A - \nu B) \geq 0$, and hence

$$S(\boldsymbol{U}_j^{n+1}) \geq S(\boldsymbol{U}_j^n).$$

In practice, conditions on ν to get $S(\boldsymbol{U}_j^{n+1}) \geq S(\boldsymbol{U}_j^n)$ is not easy to obtain. And, more often that not, there is no conditions on ν that gives entropic behaviour of the scheme. One should refer to [45] for further informations concerning the entropic behaviour of some numerical Lagrangian schemes.

I-2.1.3 General Lagrangian formulation for multi-dimensional problem

The multi-dimensional formulation of Lagrangian hydrodynamics [11] writes in integral form for a bounded domain $\mathcal{K}(t)$

$$\begin{cases} \mathrm{D}_t \int_{\mathcal{K}(t)} \rho \mathrm{d}V = 0, \\ \mathrm{D}_t \int_{\mathcal{K}(t)} \rho \boldsymbol{u} \mathrm{d}V = -\int_{\partial\mathcal{K}(t)} p \boldsymbol{n} \mathrm{d}S, \\ \mathrm{D}_t \int_{\mathcal{K}(t)} \rho e \mathrm{d}V = -\int_{\partial\mathcal{K}(t)} p \boldsymbol{u} \cdot \boldsymbol{n} \mathrm{d}S, \\ \mathrm{D}_t \int_{\mathcal{K}(t)} \mathrm{d}V = \int_{\partial\mathcal{K}(t)} \boldsymbol{u} \cdot \boldsymbol{n} \mathrm{d}S. \end{cases} \tag{I.69}$$

Here the domain $\mathcal{K}(t)$ may be displaced or deformed in time. D_t denotes for the material derivative, meaning $\mathrm{D}_t = \partial_t + \boldsymbol{u} \cdot \nabla$. The first three equations in system (I.69) are respectively the conservation of mass, momentum and total energy. The last one is a geometric conservation law. It links the deformation and displacement of the bounded domain $\mathcal{K}(t)$ to the normal velocity at its boundary.

I-2.2 Lagrangian and ALE methods for compressible hydrodynamics

In this section, a brief overview of Lagrangian and ALE methods for compressible hydrodynamics is given. Traditionally, Lagrangian hydrodynamics are solved using staggered schemes (see section I-2.2.1). Thermodynamics quantities and kinematic ones are not colocated. This tradition is issued from the Richtmyer and Von Neumann Richtmyer formulation for solving Lagrangian hydrodynamics. Staggered schemes were among the first to be used in fluid dynamics computation. Indeed, in the late 1940s, the first shock capturing hydrodynamic scheme by Richtmyer [137] and von Neumann and Richtmyer [124] was a time-space staggered 1D Lagrange explicit scheme, formulated in internal energy with artificial viscosity and 2^{nd}order accuracy in space and time on smooth flows. The scheme is usually called vNR (for Von Neumann–Richtmyer). Use of artificial viscosities is required to capture correctly shocks. Artificial viscosities and models of hyperviscosities are discussed later. Compatible formulations of compressible Lagrangian hydrodynamics are an improvement to such methods in which the schemes naturally preserve total energy and are consistent although being formulated in internal energy. Starting from localisation of variables on a given grid, formulation in internal energy is first extensively described as it is somehow the classical way of solving Lagrangian hydrodynamics system. Then, compatible and entropic Lagrangian methods are introduced. Last, pointing out some arising difficulties in Lagrangian simulations, ALE formalism is then introduced and detailed.

I-2.2.1 Natural derivation of staggered grids for hydrodynamics

Before addressing time and space discretizations, the localisation of the variables are important enough to be pointed out. Indeed, the disposition of the variables on a given grid can alter significantly precision and robustness of the numerical schemes. Staggered grids can be used to compute with a narrower centered stencil the spatial derivatives or pointwise values from average ones. This increases the spatial resolution. Indeed, eg. for wave propagation, it is known that staggered (grids based) schemes require less points per wavelength than cell-centered schemes. However, due to the fact that the grids are staggered, the CFL condition is often reduced compared to cell-centered schemes. There exist multiple definitions of the staggering of variables.

These definitions are gathered in [5] for the simulation of meteorology and oceanography and depicted in fig. I.4. The first one called *cell-centered* or A-type staggering is to consider that both velocity- and mass- related variables are located at the same position on the grid. The variables are placed at the cell center, or exclusively at the node delimiting the cell. Sometimes cell-centered schemes are also known as colocated ones. The second one called *node-staggering* or B-type staggering is to consider that velocity-related variables are at the nodes and the mass-related variables are at the cell centers. Equivalently velocity may be defined at the cell centers, and mass-related variables at the nodes. These kind of staggering is used for instance in [124, 176, 163, 109]. The third one called *face staggering* or C-type staggering consists in locating the x-velocity (resp y-velocity) related variables along the faces whose normals are colinear to the x-direction (resp. y-direction). The mass-related variables are positionned at cell-centers. This staggering is used in [153, 171] for the BBC scheme, and by extension to unstructured grids for the MAC schemes developed in [58, 80]. The natural extension to unstructured grids is made by positioning the normal velocity at the face on each faces of the grid's cell. This is often mostly convenient for conservation laws like Euler equations. The fourth one known as D-type is but a 90° rotation of the C-type staggering. This staggering enables both circulation and vorticity to be defined at the same location as mass-related variables. For most conservation laws, integration of the divergence is less convenient using this staggering of variables. Furthermore studies also proved that such a grid is more dispersive compared to a B- or C-type staggering. The E-type staggering is but a 45° rotation of the B-type staggering. The adjacency is no longer made on horizontal or vertical path for regular grids, but rather on a diagonal path.

I-2.2.2 Internal energy formulated numerical schemes

As aforementioned, the original vNR scheme, based on a B-type staggering is not conservative in total energy. Furthermore, without any artificial viscosity, the scheme is unable to correctly capture strong shocks. This lack of conservation is due to the choice of discretized variables made by Richtmyer. He chose to discretize the internal energy and its evolution equation. As a contrary, discretization of total energy yields naturally conservation of the discretized total energy. The main difficulty for schemes formulated in internal energy is that this is not any longer a conservation law. On a mathematical continuous level, the term appearing in the internal energy evolution is not defined in the sense of distributions, for velocity and pressure as bounded functions ($(u, p) \in L^\infty$). The use of artificial viscosities solves this problem by smoothing the pressure. With an appropriate definition of artificial viscosities terms, the internal energy evolution term becomes well-defined. The default of total energy conservation was highlighted in 1961 by Trulio and Trigger [165]. For non-constant time-steps, the vNR scheme is not conservative in total energy. They therefore proposed an implicit conservative version of the vNR scheme, still formulated in internal energy. They kept the spatial staggering of variables but without the temporal one. Similarly, works done by Popov and Samarskii [135] developed a similar staggered scheme with implicitation in time. In the early 1970s, DeBar used a Lagrange-remap formalism for the Trulio–Trigger scheme [37, 38]. At the end of each Lagrangian phase, the variables were

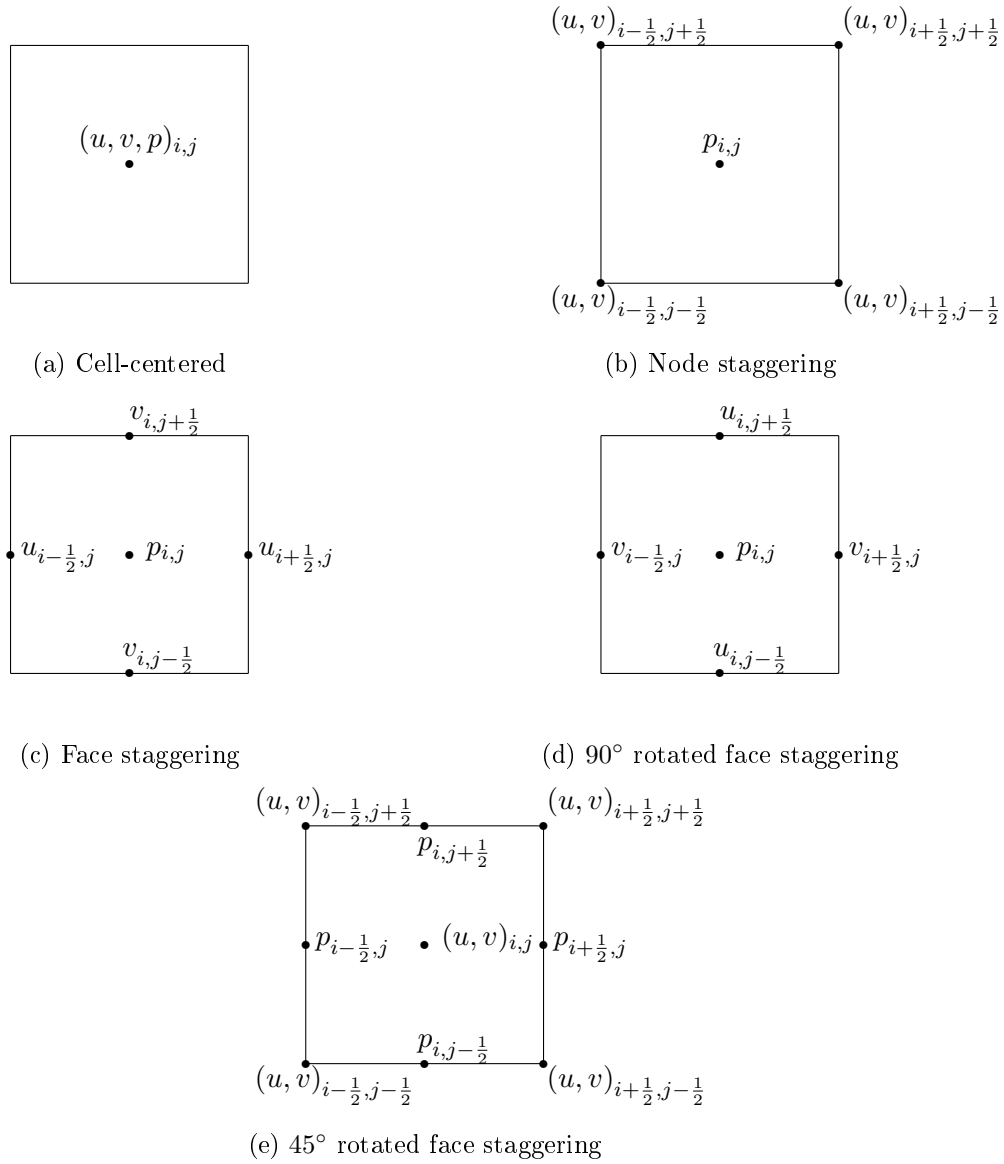


Figure I.4 – Arakawa grid system displaying the placement of variables on the grid. u denotes for x-velocity related variables, v for the y-velocity related variables and p for the mass related variables

projected on the original grids. He identified a lack of conservation due to this procedure. In fact, the projection of momentum highly dissipates kinetic energy, and so leads to a dissipation of the reconstructed total energy. He introduced a correction in internal energy to recover global total energy conservation and ensure correct shock capturing. Later, and using the earlier works by DeBar, several multifluid Eulerian hydrocodes with interface reconstruction on 2D Cartesian grids [153] were developed, based on a C-type staggering of variables. Those hydrocodes relied on the Trulio-Trigger implicit Lagrangian scheme, making use of a Lagrange-remap approach with Strang splitting. The splitting was made to consider first a 1D Lagrange-remap scheme in the x-direction, and then in the y-direction. This kind of splitting, known as directional splitting, yields the advantage of an easy extension from one dimensional problems to multi-dimensional ones.

Later, a strictly explicit predictor-corrector conservative version of the Trulio-Trigger scheme was reported by Woodward and Colella in [171]. This version was called the BBC scheme. It is a 2D Lagrange-remap scheme on staggered Cartesian grids based on a 1D Lagrange-remap setting with Strang dimensional splitting. The total energy conservation result has been credited to Noh [126]. The retained staggering of variables is the C-type one, based on Arakawa classification system. Caramana in 1998 [19] introduced the so-called compatible Lagrangian hydrodynamics for node-staggering schemes. The idea of compatible Lagrangian method is to discretize properly the internal energy evolution in order to automatically satisfy the conservation of total energy. In [9], the authors highlight the properties of such discretization. Mainly, the emphasis is laid on accuracy, consistency and stability of the compatible Lagrangian scheme. Simultaneously, on the other side of the Atlantic, Youngs developed B-type staggered schemes in which the velocity components were based on the node of the grids [176, 163, 142]. He proved his schemes, although formulated in internal energy, to be conservative in total energy, using a similar internal energy corrector as DeBar during the remapping phase. Similarly for unstructured grids, Herbin, Gallouet and al. [82, 58, 80] developed similar procedures to recover local conservation of total energy for the compressible Navier–Stokes and Euler equations for a C-type staggering. Very recently, a paper by Llor and al. proposed a conservative, compatible and entropic version of the original vNR schemes [109] staggered in both time and space. Entropic results are deduced from artificial viscosities formulation.

I-2.2.3 Total energy Lagrangian methods for compressible hydrodynamics

As a contrary to staggered scheme, the cell-centered ones naturally conserve total energy and satisfy naturally the definition of consistency for finite volume schemes as defined in definition I.6. Initial work by Després and Mazeran in [48] developed a framework in which one may easily build any cell-centered scheme to solve Lagrangian hydrodynamics. The main cell-centered total energy formulated Lagrangian schemes are EUCCLHYD developed by Maire and al. in [110] and GLACE developed by Després and al. [21, 45]. Those schemes are based on unstructured grids. GLACE builds fluxes at the boundary of each cell using an acoustic Riemann solver at each nodes in the direction given by nodes normals. EUCCLHYD builds similar fluxes but using the average of acoustic fluxes on each face around a node.

I-2.2.4 ALE formalism for compressible hydrodynamics

The Lagrangian approach can be limited due to very large deformations of the Lagrangian mesh. Indeed, the mesh deformation forces to remesh a part or the entirety of the domain, with an interface tracking in case of multi-materials simulation. For some complex and strong flows, the vorticity induced by the flows forces the remeshing regularly, which is onerous and discards partly the interest of the Lagrangian approach. A possible way to reduce this limitation is the Arbitrary Lagrangian Eulerian (or ALE) approach (see [94]). Fluid flows are computed on a domain which is deformed by a given velocity field $\boldsymbol{U}_{\text{mesh}}$. This velocity field can be chosen such that the interface between two materials is perfectly followed by the deformation of the mesh,

or such that the entire solution is smoothed by the deformation of the mesh. If one considers $\boldsymbol{U}_{\text{mesh}} = 0$, one gets back the Eulerian formulation of the scheme. And if $\boldsymbol{U}_{\text{mesh}} = \boldsymbol{u}_{\text{fluid}}$, one gets back the Lagrangian formulation of the scheme.

I-2.3 High-order direct Eulerian and Lagrange-Remap numerical schemes

In this section, an extended overview of high-order finite difference and finite volume schemes on fixed mesh for compressible hydrodynamics is given. First, the high-order space interpolation of data is presented, as well as some procedures to limit spurious oscillations in the vicinity of discontinuities. Then, multiple methods to achieve high-order integration in time are presented.

I-2.3.1 High-order space interpolation on Cartesian grids and spurious oscillations

Polynomial space interpolations

Higher-order accuracy in space is often based on high-order polynomial interpolations. Although this kind of interpolation is very accurate for smooth data, it is highly oscillatory for data with shocks or discontinuities. Indeed, Gibbs phenomenon due to polynomial interpolations generates spurious oscillations in the vicinity of discontinuities. As the mesh is refined, the Gibbs phenomenon is amplified in amplitude but bounded, and the oscillations are of lesser amplitude except near discontinuities. One possibility to reduce such oscillations is to use artificial viscosity terms (see section I-2.4). Another one is to alter the interpolation of data, considering the smoothness of the data, the average slope or the monotonicity of data. A possibility is to introduce a MUSCL-like reconstruction to damp oscillations near discontinuities. This is what is done by Nessyahu and Tadmor in [123]. Although non-oscillatory, the MUSCL reconstruction can reach beyond second order accuracy. Another point of view has been developed. The essentially non-oscillatory (aka ENO) schemes were first presented by Harten and al. in [77]. It gives a general method to build non-oscillatory interpolations for piecewise smooth functions. The main idea of ENO schemes is to select the stencil of data to perform the interpolation in function of the data smoothness inside the stencil. Originally an easy way to interpolate spatial derivatives as a function of point-wise values is to use the centered relations

$$\partial_x \phi_i = \sum_{k=0}^{r} d_k(\phi_{i+k+1} - \phi_{i-k-1}), i \in \mathbb{Z}. \tag{I.70}$$

For example, for $r = 0$, $d_0 = \frac{1}{2}$ and it yields first order of accuracy. In practice, the stencil is shifted in space in order to change the set of points on which the polynomial interpolation is performed. It yields

$$\partial_x \phi_{i,l} = \sum_{k=-r+l}^{r+l} c_{k,l}\phi_{i+k+l}, \quad l \in \{-p, \ldots, p\}, \quad i \in \mathbb{Z}. \tag{I.71}$$

Each $\partial_x \phi_{i,l}$ gives an approximation of the first space derivative of ϕ at $x = x_i$ but with a different stencil. Last, it requires to select the stencil which gives the less oscillatory interpolation. By

doing so, the interpolation thus obtained is less oscillatory than the classical one. Later, based on the ENO interpolation, the weighted essentially non-oscillatory (aka WENO) schemes were developed by Shu and Osher in [144]. The modification of the method is due to the presence of weights that tend to reduce furthermore oscillations due to the interpolation. It gives

$$\partial_x \phi_i = \sum_{l=-p}^{p} \omega_l (\partial_x \phi_{i,l}), \quad \omega_l \geq 0, \quad \sum_l \omega_l = 1, \quad i \in \mathbb{Z}. \tag{I.72}$$

One disadvantage of the WENO approach was that it was quite onerous to compute weights and smoothness indicators. Improvements of both have been developed in [90]. Last, in [143], Shu drew an analysis of the ENO/WENO schemes, as well as their evolution since the late eighties. As a contrary, the compact schemes are based on a reduced stencil reconstruction. A simple example of compact scheme is the resolution of the following system

$$\alpha \partial_x \phi_{i-1} + \partial_x \phi_i + \alpha \partial_x \phi_{i+1} = \sum_{k=-r}^{r} b_k \phi_{i+k}, \quad l \in \{-p, \ldots, p\}, \quad i \in \mathbb{Z}. \tag{I.73}$$

Compact schemes have been presented by Lele in [122]. Within this approach, the width of a stencil is reduced at the cost of a non-diagonal matrix to invert. With $\alpha = 0$, one recovers the original interpolation. A reduction of the stencil width tends to reduce interpolation oscillations. Similar procedures can be developed on unstructured grids but are more onerous than on Cartesian ones.

Discontinuous Galerkin space interpolations

Discontinuous Galerkin methods [25] assume that the discrete solution $\boldsymbol{U}_h$ lies in the finite element space of discontinuous function

$$W_h = \{\boldsymbol{V} \in (L^\infty(\Omega))^p, \ \forall \mathcal{K} \in \mathscr{T}_h, \ \boldsymbol{V}|_{\mathcal{K}} \in (\mathbb{P}(\mathcal{K}))^p\}$$

where $\mathscr{T}_h$ is a tessellation of Ω whose characteristic size is h and $\mathbb{P}(\mathcal{K})$ is the local polynomial space on $\mathcal{K}$. When computing fluxes between two members of $\mathscr{T}$, one has a discrepancy at the interface. A possible way is to solve a Riemann problem at the interface (see the ADER schemes presented in section I-2.3.2) or an interpolation between the two computed values at the interface.

Non-polynomial space interpolations

Classical interpolations are based on the assumption that locally the function is polynomial, using Taylor expansion. Another possible interpolation method is the Padé interpolation method which considers that the function is rational. Using this assumption, Padé interpolations usually reduce oscillations in the vicinity of discontinuities.

α_1	$a_{1,0}$	0	0	0	$\cdots$
α_2	$a_{2,0}$	$a_{2,1}$	0	0	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\cdots$	$\cdots$
α_{s-1}	$a_{s-1,0}$	$\cdots$	$\cdots$	$a_{s-1,s-2}$	0
1	θ_0	θ_1	$\cdots$	θ_{s-2}	θ_{s-1}

Table I.1 – A Butcher table for an explicit Runge–Kutta sequence

I-2.3.2 High-order integration in time

Runge–Kutta time integration

Let us consider first an integration in time based on Runge–Kutta sequences [99]. A primary study of Runge–Kutta sequences has been done by Butcher [16, 17]. Later, multiple authors proposed up to 5th-order accurate Runge–Kutta sequences in [55, 49]. More recently, study of total variational diminishing Runge–Kutta sequences has been performed by Gottlieb and al. in [70, 71, 69]. Moreover, Runge–Kutta sequences up to 9th-order accurate are available in [167]. Runge–Kutta sequences present the interest of an easy integration in time, once the semi-discretized in space form is obtained. Assume that the semi-discretized scheme writes

$$\partial_t \boldsymbol{U}_i = (\mathscr{P}_{\Delta x}\boldsymbol{U})_i, \quad i \in \mathbb{Z} \tag{I.74}$$

Assuming an explicit Runge–Kutta sequence whose Butcher table takes the form presented in table I.1, the integrated in time scheme writes

$$\begin{aligned} \boldsymbol{U}_i^{n+\alpha_l} &= \boldsymbol{U}_i^n + \Delta t \sum_{m=0}^{l-1} a_{l,m} (\mathscr{P}_{\Delta x}\boldsymbol{U}^{n+\alpha_m})_i, \quad i \in \mathbb{Z}, \\ \boldsymbol{U}_i^{n+1} &= \boldsymbol{U}_i^n + \Delta t \sum_{m=0}^{s-1} \theta_m (\mathscr{P}_{\Delta x}\boldsymbol{U}^{n+\alpha_m})_i, \quad i \in \mathbb{Z}. \end{aligned} \tag{I.75}$$

Lax–Wendroff or Cauchy–Kovalevskaya time integration

Very high-order Lax–Wendroff or Cauchy–Kovalevskaya based schemes have been presented in [50] and are used in a CEA hydrodynamics simulation platform [91]. Originally, works have been performed for the linear case, and especially the advection and wave equations as presented in [40]. Consider an hyperbolic system of the form

$$\partial_t \boldsymbol{U} + \partial_x \boldsymbol{F}(\boldsymbol{U}) = 0.$$

Integrating in time between t^n and $t^n + \Delta t$ and space over a cell $\mathcal{K}_i = \left[x_{i-\frac{1}{2}} : x_{i+\frac{1}{2}}\right]$, it yields

$$\overline{\boldsymbol{U}}^{n+1} - \overline{\boldsymbol{U}}^{n} = -\int_{t^n}^{t^n+\Delta t} (\boldsymbol{F}(\boldsymbol{U}(x_{i+\frac{1}{2}}, \theta)) - \boldsymbol{F}(\boldsymbol{U}(x_{i-\frac{1}{2}}, \theta)))\mathrm{d}\theta \tag{I.76}$$

Performing a Taylor expansion around t^n of $\boldsymbol{F}(\boldsymbol{U}(x_{i+\frac{1}{2}}, \theta)$ and $\boldsymbol{F}(\boldsymbol{U}(x_{i-\frac{1}{2}}, \theta)$ it yields

$$\overline{\boldsymbol{U}}^{n+1} - \overline{\boldsymbol{U}}^{n} = -\sum_{k\geq 0} \partial_t^k (\boldsymbol{F}(\boldsymbol{U}(x_{i+\frac{1}{2}}, t^n)) - \boldsymbol{F}(\boldsymbol{U}(x_{i-\frac{1}{2}}, t^n)))\frac{\Delta t^{k+1}}{(k+1)!}$$

The idea is then to use the system of PDEs to replace time derivatives of $\boldsymbol{F}$ by spatial ones at time t^n. Thus, a high-order in time scheme is obtained. If one considers that space derivatives are computed with high-order accuracy in space, then it yields a high-order accurate scheme in both time and space.

ADER time integration

Arbitrary Derivative Riemann (also known as ADER) problem has been developed by Titarev and Toro in [164]. It is a high-order accurate in both time and space finite volume scheme. It uses Godunov's upwind approach and the Lax–Wendroff (or Cauchy–Kovalevskaya) procedure. For hyperbolic problem as depicted in eq. (I.56), the idea is to differentiate in time eq. (I.56) and to solve Riemann problems on each of the derivatives. Solving Riemann problems on each of the derivatives is called solving the generalized Riemann problem. Thus, it yields a high-order finite volume scheme.

I-2.4 Artificial viscosities and hyperviscosities

Artificial viscosities and hyperviscosities are a mean to damp spurious oscillations due to high-order polynomial interpolations. The main idea is to add a viscous term to prevent oscillations for occurring. The main drawback is that viscosity are tuned with user-fixed parameters, and the choice of parameters is not obvious.

I-2.4.1 Internal energy weak formulation

As aforementioned, the internal energy evolution equation has no sense for non-smooth pressure. A way to deal with this problem is to add a viscosity term such that the Lagrangian system formulated in internal energy, initially depicted in eq. (I.61), now writes

$$\begin{cases} \mathrm{D}_t \rho_0 \tau - \partial_X u &= 0, \\ \mathrm{D}_t \rho_0 u + \partial_X (p+q) &= 0, \\ \mathrm{D}_t \rho_0 \epsilon + (p+q)\partial_X u &= 0. \end{cases} \tag{I.77}$$

If q is chosen and built such that $p+q$ is smooth and non-zero then the internal energy evolution equation is defined in the sense of distributions. Moreover, from the physical point of view, q can be seen originally as the viscosity produced by the inelastic collisions between particles [138]. This can be seen as an enrichment of Euler equations. Usually artificial viscosities are used for high-order schemes and/or for schemes formulated in internal energy. Mainly, the very essence of the artificial viscosity is to reduce the Gibbs phenomenon which occurs at shocks and discontinuities due to the reconstruction of fluxes. One may refer to the paper by Benson [11] for more informations on the expression of the artificial viscosity q.

I-2.4.2 Standard expressions of viscosities

Originally in [124], the viscosity q takes the form

$$q_i = -c_q \rho_i \Delta u_i |\Delta u_i|$$

with $\Delta u_i = u_{i+\frac{1}{2}} - u_{i-\frac{1}{2}}$. This viscosity is usually called the vNR artificial viscosity or pseudo-viscosity. In [138], the Rosenbluth viscosity is proposed. It is somehow similar to the original vNR pseudo-viscosity but only activated where $\Delta u_i < 0$. Indeed, for a perfect gas, $\Delta u_i < 0$ stands for a compression, where a shock may appear. This is not the case for a non-perfect gas with a more complex EOS. The Rosenbluth viscosity writes

$$q_i = -c_q \rho_i \Delta u_i |\Delta u_i| \chi_{\{\Delta u_i < 0\}}.$$

Another legacy viscosity is denoted Landshoff pseudo-viscosity [101]. It is similar to the Rosenbluth one, with an additional linear dissipative term. It writes

$$q_i = -(c_q \rho_i \Delta u_i |\Delta u_i| + c_l \rho c_i \Delta u_i) \chi_{\{\Delta u_i < 0\}}.$$

For these viscosities, the parameters c_q and c_l are user-chosen. Many works have been performed in the literature to study the impact of viscosity as well as a way to determine *a priori* values for c_q and c_l. Wilkins developed an extension to the original von Neumann-Richtmyer viscosity to the multidimensional case in [169]. Noh in [127] showed the very limits of the use of artificial viscosity. Indeed, he showed that artificial viscosity can induce strong errors in the computation, instead of damping oscillations and smoothing pressure profiles. Caramana, Shashkov and Whalen presented in [18] a new formulation for the artificial viscosity terms. They based their works considering that the artificial viscosity should follow a certain number of conditions to be considered physically acceptable. The artificial viscosity should among other be galilean invariant and always transfer kinetic energy into internal energy. Moreover, for isentropic compressions, the artificial viscosity must not create too much dissipation or entropy. Heuzé, Jaouen and Jourdren investigated the effect of artificial viscosities for discontinuities on a non-convex EOS in [84]. More recently, Guermond and al. proposed the construction of an entropic viscosity in [74]. Last but not least, the reader may refer to the paper by Mattsson and Rider [111] about the origins of artificial viscosity terms, and the very bedrocks of pseudo-viscosities expressions

and properties.

I-2.4.3 Hyperviscosities

As said previously, the use of artificial viscosities can be seen as a necessary enrichment of Euler equations. This use enables to match better, on a physical point of view, the complex structure of flows. An idea presented by Cook and Cabot in [28] and later in [27] is to consider the compressible Navier-Stokes equations, which is nothing else but the Euler equation with a viscous term. Then the underlying viscosity coefficients in the compressible Navier-Stokes equations are set accordingly to the smoothness of the flows. For perfectly smooth flows, there is no physical, mathematical or even numerical reason to add dissipation, and thus the coefficients are set to 0. However at a discontinuity or a shock, to avoid Gibbs phenomenon, one wishes for more dissipation and thus the coefficients are no longer null.

The model is described in eq. (I.78),

$$\partial_t \begin{pmatrix} \rho \\ \rho \boldsymbol{u} \\ \rho e \end{pmatrix} + \nabla \cdot \begin{pmatrix} \rho \boldsymbol{u} \\ \rho \boldsymbol{u} \otimes \boldsymbol{u} + p\underline{\boldsymbol{I}} - \underline{\boldsymbol{\sigma}} \\ (\rho e + p)\boldsymbol{u} - \underline{\boldsymbol{\sigma}} \cdot \boldsymbol{u} \end{pmatrix} = \mathbf{0}. \tag{I.78}$$

where the viscous stress tensor is denoted by $\underline{\boldsymbol{\sigma}}$ and satisfies

$$\underline{\boldsymbol{\sigma}} = 2\mu \underline{\boldsymbol{S}}(\boldsymbol{u}) + (\beta - \frac{2}{3}\mu)(\nabla \cdot \boldsymbol{u})\underline{\boldsymbol{I}} \tag{I.79}$$

where β is the bulk viscosity, μ is the shear viscosity, and $\underline{\boldsymbol{S}}$ is the symmetric strain rate tensor $\underline{\boldsymbol{S}} = \frac{1}{2}(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^t)$. The coefficients μ and β are to be set accordingly to the smoothness of the flow. In practice, they are set as

$$\beta = C_\beta \eta_r, \quad \mu = C_\mu \eta_r, \quad \eta_r = \rho h^{r+2} G(\nabla^r \|\underline{\boldsymbol{S}}\|), \quad r \in 2\mathbb{N}, \tag{I.80}$$

where C_β and C_μ are user-specified, h is technically the typical space grid, and $\|\underline{\boldsymbol{S}}\|^2 = \underline{\boldsymbol{S}} : \underline{\boldsymbol{S}}$. Last G denotes for the application of a truncated Gaussian filter. The use of G is to smear out oscillation introduced by the differentiation of the tensor norm. This viscosity presents the advantages of maintaining high-order accuracy for smooth flows, but can be rather expensive numerically due to the differentiation of the tensor norm. Extensions and improvements of the hyperviscosity model have been presented in [13] and [93]. Essentially the authors proposed to modify the computation of η_r to yield a steeper profile for the viscosity, and so avoid undesirable dissipation in smooth areas.

I-3 Numerical methods for fluid-structure interaction

In this section, an extended review of numerical methods for fluid-structure interaction is made and especially concerning the coupling in time and space chosen for the continuity relations at

the boundary. A bounded domain Ω is divided into two parts. A fluid domain denoted Ω_f and a structure domain denoted Ω_s such that $\Omega = \Omega_f \cup \Omega_s$ and $\Omega_f \cap \Omega_s = \emptyset$. The boundary $\partial\Omega_s \cap \partial\Omega_f$ is denoted Γ in the following, and the normal to the boundary Γ going from Ω_s to Ω_f is denoted $\boldsymbol{n}_\Gamma$. The fluid (respectively structure) velocity is denoted $\boldsymbol{u}_f$ (respectively $\boldsymbol{u}_s$), and the fluid (respectively structure) stress tensor is denoted $\underline{\boldsymbol{\sigma}}_f$ (respectively $\underline{\boldsymbol{\sigma}}_s$).

For a viscous fluid, continuity relations are called the no-slip boundary conditions. The velocity and the normal stress are continuous through the boundary. It yields

$$\boldsymbol{u_f} = \boldsymbol{u_s}, \quad \underline{\boldsymbol{\sigma}}_f \cdot \boldsymbol{n}_\Gamma = \underline{\boldsymbol{\sigma}}_s \cdot \boldsymbol{n}_\Gamma, \quad \text{on} \quad \Gamma. \tag{I.81}$$

In particular, eq. (I.81) means that the displacement and velocity at the boundary are continuous. This yields in particular that the interface between fluid and solid is easier to track. For moving meshes methods (ALE) presented in the previous section, the space discretization follows perfectly the interface.

For a non-viscous fluid, the continuity relations are called slip boundary condition. It allows the fluid to slip perfectly along the structure boundary without any kind of boundary layer. It writes

$$\boldsymbol{u_f} \cdot \boldsymbol{n}_\Gamma = \boldsymbol{u_s} \cdot \boldsymbol{n}_\Gamma, \quad \underline{\boldsymbol{\sigma}}_f \cdot \boldsymbol{n}_\Gamma = \underline{\boldsymbol{\sigma}}_s \cdot \boldsymbol{n}_\Gamma, \quad \text{on} \quad \Gamma. \tag{I.82}$$

As a contrary to the no-slip boundary conditions, eq. (I.82) means that the tangential displacement is not continuous at the boundary as fluid particles may slip freely along the tangential direction of the boundary. Other models for boundary conditions may be used but in this work, the emphasis is laid on eq. (I.82). Considering two numerical methods, the coupling must be realized at the boundary in order to satisfy boundary conditions, in space as well as in time. In order to achieve that, time-coupling is first detailed. Then an overview is made on space coupling numerical methods found in the literature.

I-3.1 Time coupling method for fluid-structure interaction

There are two ways to see a fluid-structure numerical method : a partitioned domain approach or a monolithic one (see [116] for further details). The monolithic is not prone to change. Any modification in the fluid or the structure part, eg. change in the numerical flux, results in change for the whole approach. It also means that the hydrocode and structure-code must be entirely known, and may not be used as a black box. Although it gives the advantage to overview every part of the code, it is also a strong inconvenient. The partitioned/domain approach yields the advantage to perfectly decouple fluid and structure part. As an example, it allows a hydro-code to be coupled with a commercial code for structural deformations computation. The fluid and structure solver are perfectly independent and do not necessarily rely on the same space and time discretizations. Depending on the space and time coupling, boundary conditions presented in eqs. (I.81) and (I.82) are more or less satisfied at the boundary. If those conditions are perfectly satisfied at the boundary at any discrete time, the term *strong* coupling is used. However, if not, the term *loose* coupling is used for boundary conditions that are only weakly imposed. The

strong coupling often relies on a time-implicitation of terms around the boundary. This reveals quite onerous since a non-linear system is solved all along the boundary in order to perfectly satisfy the boundary conditions. Fully explicit schemes are generally considered as *loosely coupled* and may introduce large instabilities, especially when the ratio between both material masses (fluid and structure) is high. Semi-implicit coupling is a computationally compromise between implicit and explicit coupling. It is not as onerous as a full implicit one, and moreover it prevents certain instabilities present in the explicit coupling to occur. In the following, the three coupling are detailed. One may refer to [56] for an overview of the different time coupling methods for incompressible viscous flows.

I-3.1.1 Loose coupling

Loose coupling is certainly the most intuitive one in order to deal with fluid-structure interaction. The fluid and structure system of partial derivatives equations are solved in a decoupled way, with a regular exchange of information at the boundary. Mostly, one considers that the fluid part exerts a stress constraint on the structure part, and reciprocally the structure part exerts a velocity constraint on the fluid part. Velocity and stress boundary conditions are not necessarily satisfied, especially if the time discretization is not the same for both solvers. An example of loose

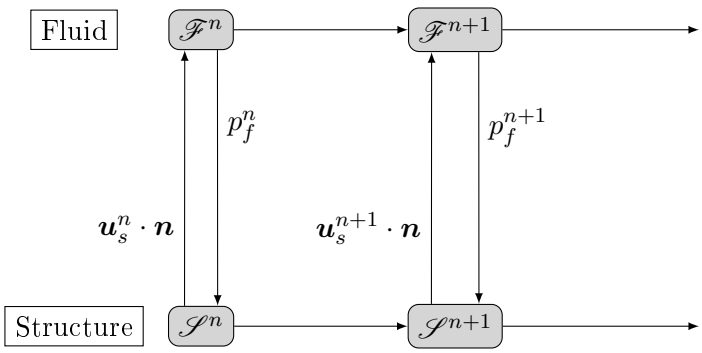


Figure I.5 – A fully explicit fluid-structure coupling algorithm on same time discretization

coupling is given in fig. I.5. At each time-step beginning, the structure imposes to the fluid the normal velocity and in return, the fluid imposes pressure stress on the structure (or reversely). In fig. I.6, the two solvers are on a different time scale, they are staggered with one another in time. Although it is still considered as a loose coupling, in practice, it proves to be slightly more stable. It is also possible to consider two different time-scales, one specific to the structure and one to the fluid. Indeed, time-step restrictions are slightly different for the fluid and structure solvers. It is then possible to achieve multi-time step of fluid evolution whereas only one is achieved for the structure part, or the reverse. In [119, 120], Monasse and al. developed a fully explicit coupling but which ensures conservation of quantities up to their algorithm precision.

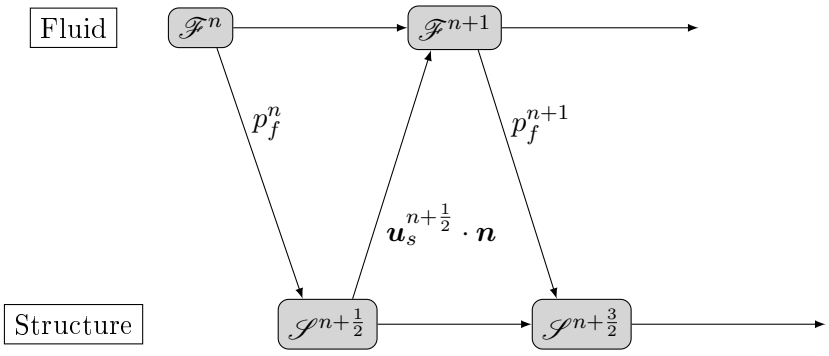


Figure I.6 – A fully explicit fluid-structure coupling algorithm on staggered time discretization

I-3.1.2 Strong coupling

Strong coupling is done in order to ensure strongly the boundary conditions. At a given time $t = t^n$, it builds boundary conditions in order to enforce that the prescribed boundary conditions hold at time $t = t^{n+1}$. The strong enforcement of boundary conditions ensures correct mass, momentum and total energy conservation at the boundary. Conservation is ensured to the limit of the convergence criteria used in system inversion algorithms. At each time-step, a non-linear system is solved to find the solution at time $t = t^{n+1}$. One uses an iterative algorithm among which fixed-point, conjugate gradient, Newton or Gauss-Seidel. A strongly coupled scheme is much more onerous than an explicit one, as the problem is solved at each iteration of the algorithm used to inverse the system. However, stability conditions on time-step are much less severe than for full explicit schemes. But in practice, the time-step must be restricted or the algorithm must use relaxation terms in order to ensure convergence. The numerical cost of such a procedure is sometimes prohibitive, and hence another class of coupling has been derived: the semi-strong coupling.

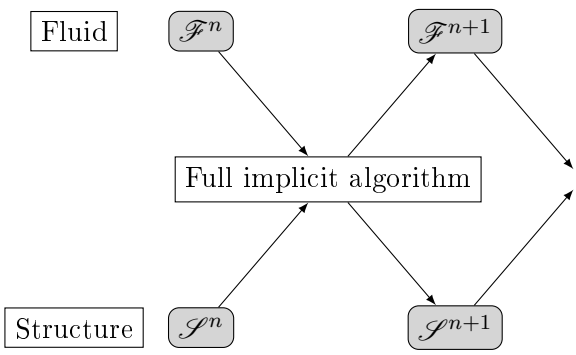


Figure I.7 – A fully implicit fluid-structure coupling algorithm on same time discretization

I-3.1.3 Semi-strong coupling

The semi-strong coupling has been derived to correct the weaknesses of both loose and strong coupling. The loose coupling is sometimes unstable and unable to track effectively the conservation of mass, momentum and total energy at the boundary. On the other hand, the implicit

coupling ensures those conservations, but is numerically onerous and the convergence is not ensured. Relaxation terms can be added to correct these defaults but to an additional numerical cost. For these reasons, semi-strong coupling has been derived. Non-oscillatory or stable terms are treated in an explicit way, whereas unstable and/or oscillatory terms are treated in an implicit way. Often, the pressure terms are treated in an implicit way, whereas advection and diffusion terms are treated in a fully explicit way. Conservation results rely strongly on the algorithm and hypothesis made previously. In [136], Puscas and al. derived a semi-implicit scheme that ensures conservation of quantities up to the algorithm precision.

I-3.2 Space coupling method for fluid-structure interaction

Independently to the time-coupling, space-coupling methods allow to spatially couple forces and torques at the boundary in order to enforce boundary conditions presented in eqs. (I.81) and (I.82). Three families of space coupling are distinguished and sorted as follows. First, the mixed cells methods which somehow average the different materials over a cell. Second, the body-fitted method which ensures a deformation and displacement of the mesh such that cells remain pure. Third and last, the fictitious domain method which uses overlapping domains to enforce boundary conditions.

I-3.2.1 Mixed cells methods

One natural way of dealing with fluid-structure interaction is to consider that a control volume for a finite volume scheme may contain both materials. A detector is then used to determine which constitutive laws are to be used. In [41], the authors proposed a unified framework to treat both solid and fluid simulations on unstructured grid. The proposed schemes can be used in a fully Lagrangian formalism or in an ALE one. The constitutive laws are then selected considering to which material the cells interfaces belong (the case of mixed interfaces is also treated). In [68], they proposed a definition of an *ad hoc* Riemann problem at solid boundaries which is formally second order accurate. Thanks to a level set method, they detect the proximity of a wall and modify the Riemann problem to take into account the boundary conditions. Although the resulting scheme is not conservative, shocks seem to be correctly captured. The scheme is based on Cartesian grids. Last, [79] introduced a full-Eulerian solid level set method in order to treat fluid-structure interaction problems for incompressible viscous flows. The method is derived by adding a solid body force and a solid-fluid interaction term for cells near the boundary. The interface tracking is realized thanks to the solid level set method. It also applies for fluid schemes based on Cartesian grids.

I-3.2.2 Body-fitted methods

The Lagrangian and ALE approaches for solving the compressible hydrodynamics system have been presented in section I-2.2. For viscous fluid, the deformation of the mesh is continuous along the boundary. It means that technically, if initially the meshes for the fluid and structure

are coincident at the boundary, then they stay coincident for any time of the simulation. It leads to an easier interface tracking as no mixed cells appear. However for inviscid flows, the deformation is no longer continuous along the boundary, only the normal deformation is. Two choices are presented in the literature. Either one uses Lagrange multipliers to transfer the forces and torques between the two meshes, either one uses the ALE formalism with a velocity of the mesh prescribed by boundary conditions.

Wall boundary conditions

For wall boundary conditions, the prescribed normal velocity at the boundary is set to 0. In [98], the author described a body-fitted discontinuous Galerkin scheme to approximate the solution to the Euler equations with solid wall boundary conditions for curved geometry. An important feature in this paper is that the boundary conditions should be prescribed on the real continuous geometry, rather that on the approximated discretized geometry obtained with the mesh. Doing so, the error due to the discretization of the geometry does not reduce the overall accuracy of the scheme. Moreover in some cases, with conditions imposed on the discretized geometry, steady flows are not reached by the schemes. The asymmetry introduced by the discretization may indeed introduce vortices or wakes that are irrelevant considering the Euler equations system.

Remeshing constrained by structure motion

The ALE method (see section I-2.2.4) relies on a periodic or cycle-based displacement of the mesh. The displacement is based on a prescribed mesh velocity field denoted $\boldsymbol{u}_{\text{mesh}}$. To ensure that the structure and fluid meshes stay coincident one may just provide the following condition on the velocity field

$$\boldsymbol{u}_{\text{mesh}} = \boldsymbol{u}_s, \quad \text{on} \quad \Gamma. \tag{I.83}$$

The Jacobian is then deduced. However, the presence of too much distorted elements or non-conformal ones, forces the algorithm to remesh partly the fluid domain and to project conservatively quantities on the new mesh. This re-meshing phase may prove quite expensive. Indeed, in 1D or 2D, the re-meshing is not problematic, but in 3D the numerical cost sometimes becomes preponderant over the cost of the fluid and structure solvers. In [86], Hu and al. presented an ALE method to couple a Navier-stokes solver with a particle one, showing in particular that an explicit coupling is not stable. Later in [87], they assessed the evolution of the ALE methods for fluid-structure coupling. The structural displacement is dealt with using eq. (I.83). The update of the meshes displacement is done in an explicit way, whereas the update of the meshes velocity is implicit, resulting in a stable scheme. Extension to visco-elastic fluid is made. In [107], Le Tallec and Mouro proposed to consider the whole space as a physical continuum. The resulting problem is then split into a fluid and a structural part, enforcing kinematic acceptable states between the two. Their method uses a Lagrangian approach for the structure and an ALE

formulation for the fluid. Mesh velocity is then imposed according to eq. (I.83) so that interfaces between fluid and structure stay coincident.

I-3.2.3 Fictitious domain methods

In order to avoid any kind of remeshing, the fictitious domain methods have been introduced for fixed mesh methods (see section I-2.3). The fluid mesh overlaps the structure and the fluid values in the overlapping cells are completely fictitious. The main problem consists in imposing the values in these overlapping cells. The main issue is how to impose these values in order to satisfy the boundary conditions. For body-fitted methods, the meshes are not overlapping, and there is no need to define such values. Many methods have been derived to tackle this problem. They gather into seven families which are listed below and described in the following:

i) Immersed boundary methods;
ii) Direct forcing methods;
iii) Penalization methods;
iv) Lagrange multipliers;
v) Embedded cut-cells methods;
vi) Reflection and mirroring ghost-cells methods;
vii) Inverse Lax–Wendroff boundary treatment.

One may refer to [117] or [147] for an extended review of the fictitious domain methods.

Immersed boundary method

The immersed boundary method (IBM) has been first proposed by Peskin [133] and later extended by Lai and Peskin [100, 134]. Originally, the method has been developed for the simulation of cardiac blood flows. The physical model used is the incompressible Navier–Stokes equations coupled with very thin elastic structures, with equivalent density. This is a very peculiar model, where for once the structural displacement is imposed by the fluid one. The method consists in forcing the movement of the structure using the fluid displacement and to weakly impose a discontinuity in the fluid constraint at the boundary. To do so, additional forces are added to the fluid near the interface. For such a fluid, one writes

$$\left\{ \begin{array}{ll} \rho\left(\partial_t \boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{u}\right) + \nabla p & = \mu\Delta\boldsymbol{u} + \boldsymbol{f}, \\ \nabla\cdot\boldsymbol{u} & = 0, \end{array} \right. \tag{I.84}$$

where the variables ρ, $\boldsymbol{u}$, p denote for the density (assumed constant), the velocity vector and the pressure. The additional force $\boldsymbol{f}$ is introduced to satisfy weakly the boundary conditions and writes

$$\boldsymbol{f}(\boldsymbol{x},t) = \int_\Omega \boldsymbol{F}(\boldsymbol{y},t)\delta_0(\boldsymbol{x}-\boldsymbol{X}(\boldsymbol{y},t))\mathrm{d}\boldsymbol{y}. \tag{I.85}$$

where $\boldsymbol{X}$ is the Lagrangian position of the elastic structure and $\boldsymbol{F}$ is the Fréchet derivative of the internal energy with respect to the Lagrangian position $\boldsymbol{X}$. δ_0 symbolizes the Dirac function. The discretization of the Dirac function is made in order to ensure mass, momentum and total energy conservation, as well as Galilean invariance. Note in particular that on the continuous level, eq. (I.85) reduces to $\boldsymbol{f} = \boldsymbol{F}$. Reciprocally, the structure part is solved thanks to the following equation

$$\partial_t \boldsymbol{X}(\boldsymbol{x}, t) = \int_\Omega \boldsymbol{u}(\boldsymbol{y}, t)\delta_0(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{y}, t))\mathrm{d}\boldsymbol{y}. \tag{I.86}$$

which yields on the continuous levels that $\partial_t \boldsymbol{X} = \boldsymbol{u}$. The fluid velocity imposes the displacement of the structure. This method is forged to deal with very thin structures, whose density is similar to the fluid one. Order of accuracy has been studied for smooth problems in [73]. The method has been modified for adapted refinement in [72] to reach second order of accuracy. For thick structure, it is rather the structure velocity that imposes the displacement of the fluid. To deal with thicker structures, direct forcing methods have been developed.

Direct forcing methods

As for the immersed boundary method, the direct forcing method consists in adding an external force in order to satisfy boundary conditions. Consider an incompressible viscous fluid flow with boundary conditions provided by eq. (I.81). A possible consistent discretization of boundary conditions is to impose near the interface the fluid velocity to be equal to the structure velocity. It is equivalent to set $\boldsymbol{f}$ such that

$$\boldsymbol{f} = \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} (\boldsymbol{u}\cdot\nabla\boldsymbol{u} + \nabla p - \mu\Delta\boldsymbol{u})\,\mathrm{d}t + \frac{1}{\Delta t}(\boldsymbol{v}^{n+1} - \boldsymbol{u}^n), \tag{I.87}$$

with $\boldsymbol{v}^{n+1}$ being the structure velocity at time t^{n+1}. Indeed substituting eq. (I.87) in eq. (I.84) integrated in time between t^n and t^{n+1}, it immediately yields that $\boldsymbol{u}^{n+1} = \boldsymbol{v}^{n+1}$. In the numerical schemes, $\boldsymbol{f}$ is not used, and the velocity directly satisfies $\boldsymbol{u}^{n+1} = \boldsymbol{v}^{n+1}$. Geometrically, the interface neighbourhood is defined as the mixed cells (partly fluid, partly solid) in addition with the cells inside the solid part. See fig. I.8 as a representative example.

The wider the stencil used by the numerical scheme, the wider the interface neighbourhood. Only mixed and fully solid cells values are to be imposed. With $\boldsymbol{f}$ defined as in eq. (I.87), the order of accuracy of the method is at most one. The method developed by Mohd-Yusof in [118] and [52] consists in doing an interpolation of the velocity relative to the interface, around the boundary. Then a antisymmetry of the relative velocity is used inside the mixed/full solid part of the domain. This method is second order accurate and *a priori* more accurate than doing a direct forcing without any kind of interpolation. It is mostly used for incompressible viscous flows but has been extended for compressible viscous flows. It is obviously not conservative in mass, momentum and total energy. In [173], the authors proposed a simplified, efficient and accurate direct forcing method for incompressible flows. It is still based on a strong coupling but without

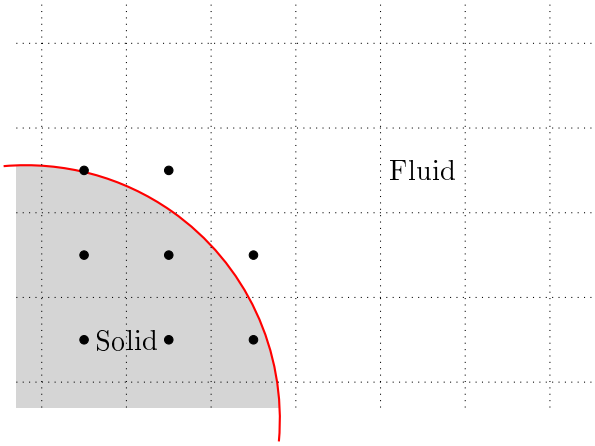


Figure I.8 – Interface neighbourhood for the direct forcing method: black dots stand for the cells where velocity values are imposed to match the structure velocity

any call to the fluid solver during the coupling, which alleviates greatly the computation. In [10], the author proposed a regularization technique for the direct forcing methods. This regularization prevents spurious force oscillations from occurring. The extension to compressible inviscid flows is presented later as the reflection and mirroring ghost-cells methods.

No-penetration and volume penalization methods

For fluid-structure interaction, considering non-porous media for the structure, the fluid mass must remain outside the structure. There is no fluid penetration inside the structure. For example, the direct forcing method does not satisfy the no-penetration condition. One possible approach to deal with this problem is to penalize any kind of fluid penetration in the structure. This is the penalization method. The method was first introduced by Arquis and Caltagirone [7] for incompressible viscous flows, with a Brinkman porosity model for the solid. It is equivalent to simulating a fluid-structure interaction with a porous media whose porosity is defined by a very small parameter. The smaller the parameter, the less porous the media is, till impermeability. In [4], Angot and al. proposed a L^2-penalization. Let η be a parameter and consider Ω_s as the solid part, it yields for incompressible viscous flows

$$\begin{cases} \partial_t \boldsymbol{u}_\eta + \boldsymbol{u}_\eta \cdot \nabla \boldsymbol{u}_\eta + \nabla p_\eta & = \frac{1}{\mathrm{Re}} \Delta \boldsymbol{u}_\eta - \frac{1}{\eta} \boldsymbol{u}_\eta \chi_{\Omega_s}, \quad t > 0, \boldsymbol{x} \in \mathbb{R}^2 \\ \nabla \cdot \boldsymbol{u}_\eta & = 0, \quad t > 0, \boldsymbol{x} \in \mathbb{R}^2 \end{cases} \tag{I.88}$$

χ denotes for the indicator function. They showed convergence when $\eta \to 0$ toward the solution of the Navier Stokes with zero-velocity boundary conditions provided on $\partial\Omega_s$. The accuracy has been proven to be at worst of order $\frac{3}{4}$ in η, but in practice 1^{st}order of accuracy is recovered. However, the CFL conditions is largely impacted due to the relaxation term $\frac{1}{\eta}\boldsymbol{u}_\eta \chi_{\Omega_s}$. Using a fully-explicit scheme yield a CFL condition as $\Delta t \leq C\eta$ which is a constrain as η should tend to zero. However, an implicit treatment of the relaxation terms entirely withdraws this condition. As the relaxation term is local, the implicit treatment is not as onerous as the implicit treatment

of complex numerical fluxes. The penalization is not generally conservative in mass, momentum and total energy. Depending on the value of η, the boundary treatment may introduce undesirable boundary layers for compressible inviscid flows. Moreover correct capture of shocks is impacted due to this special treatment. In [51], the extension to deformable obstacles is realized.

Lagrange multipliers

Fictitious domain based on Lagrange multipliers for incompressible viscous flows have been first developed by Glowinski and al. in [60]. The solid domain is filled with a fictitious fluid state. Lagrange multipliers are used to ensure rigid body motion in the Navier-Stokes variational formulation. Studies and improvements have been done to develop the method, increase robustness and alleviate the computations in [131] and [59]. Extension to visco-elastic particles/bodies has been realized in [145]. Solid and fluid problems are coupled thanks to the Lagrange multipliers. Those multipliers are seen as pseudo-forces that are exerted on both parts. A full explicit procedure is possible. As a contrary to the immersed boundary method which relies on approximate Dirac function to enforce the correct exchange of forces, here, the procedure relies on the Lagrange multipliers to exchange forces.

Embedded cut-cells methods

The first embedded cut-cells method has been introduced by Noh, while working on the coupling between a Lagrangian method for the structure part and an Eulerian finite volume method for the fluid part [125]. The proposed embedded cut-cells method provides naturally conservation of mass, momentum and total energy due to the special space discretization. The method relies on the following observation: cutting the cells near the interface and integrating forces and torque on the interface yield immediately conservation of the desired quantities. However due to the possible very small cells, the CFL condition is highly impacted. Indeed, denote by α^n the volume fraction of the structure inside a cell at time t^n, it writes

$$(1-\alpha^{n+1})\boldsymbol{U}^{n+1} = (1-\alpha^n)\boldsymbol{U}^n - \frac{t^{n+1}-t^n}{h}\Delta\boldsymbol{U} \tag{I.89}$$

where $\Delta\boldsymbol{U}$ is the flux at the boundary of the cell. Immediately, the CFL condition becomes $\Delta t < (1-\alpha)\frac{h}{c}$, which can be arbitrarily small as α tends to 1. This is the main drawback of the method. The CFL condition is proportional to the volume of a cell divided by its perimeter. Therefore one gets very small time-steps near the boundary due to the presence of cut-cells. The general principle of cut-cells methods is presented in fig. I.9. Numerical fluxes for cells around the boundary need to be modified to ensure correct boundary conditions enforcement. Two main approaches have been considered in the literature. The first one presented in [132] and [26] consists in evaluating the numerical fluxes as if there were no structure in cut-cells. Then, identifying the lack of conservation of mass, momentum and total energy, to redistribute the lacking quantities partly in the cut-cells and partly in the adjacent ones. The redistribution is

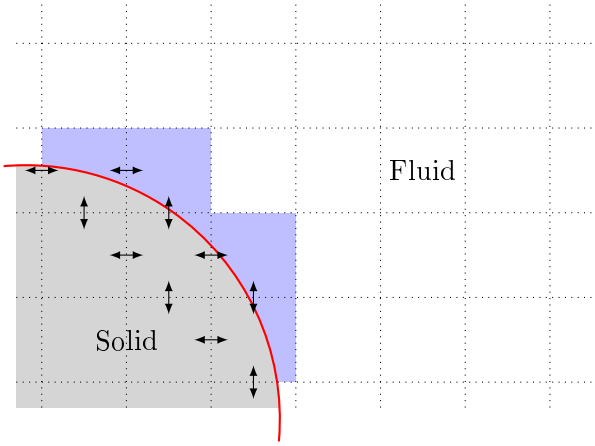


Figure I.9 – Embedded boundary methods: Double head arrows show numerical fluxes needing modification to take into account the boundary

based on mass weighting. The second one presented in [53] and [88] consists directly in drawing a conservative balance on each cut-cell. In order to avoid very small cells and so CFL restrictions, [53] made the suggestion to merge the small cells with fully fluid adjacent ones. As to [88], they proposed to mix the cells with cells aligned in the normal direction outward the solid/fluid boundary. Last [120] proposed to mix too small cells [1] with an adjacent one. This mixing is illustrated in fig. I.10.

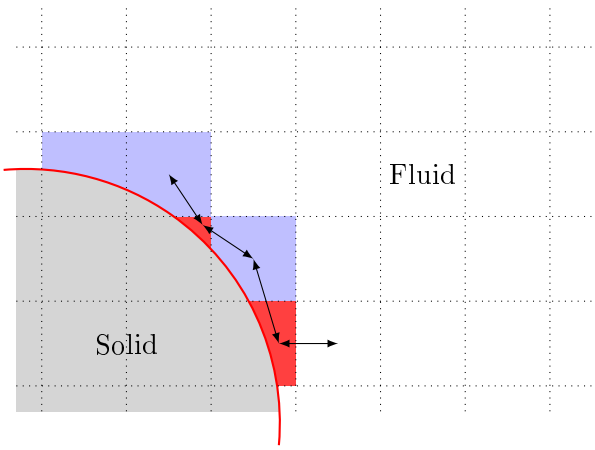


Figure I.10 – Merging of small cut-cells present along the boundary using the outward normal criteria: Arrows stand for conservative mixing of cells, red cells are mixed with the larger cells indicated by arrows

The main known limitations of cut-cells methods is that it is impossible to consider a very thin structure present in a fluid. The thickness of the structure must be at least greater that the characteristic length of the mesh. The mixing procedure can only apply if a large enough adjacent cell is found. This becomes problematic for some 3D problems. Some geometric configurations are also tricky. If two solid elements are present in the same cell, then there is no natural definition for a global outward normal. Using a finer mesh or adaptive mesh refinement (AMR)

1. The criteria if fixed to $\alpha > 0.5$, where α is the structure presence volumic fraction

solves the problem. In [115], Meyer and al. proposed a cut-cell embedded boundary method for Large-Eddy simulation of incompressible flows on staggered Cartesian grids. The interface geometry is described by a level-set method in 3D, and cells cut by this interface of small size are mixed with larger, neighbouring cells. They assessed second order of accuracy for their method. In [78], Hartmann and al. proposed a cut-cell embedded boundary method for two and three dimensional problems, on adaptive grids. The method is proved to be conservative in mass, momentum and total energy, and numerical experiments demonstrated second order of accuracy. A particularity of their method is that they considered viscous compressible flows. They used a mixing algorithm near the boundary to merge very small cells with a master cell in order not to damp the CFL condition. Muralidharan developed in [121], a new adaptive finite volume conservative cut-cell method which is third order accurate for the compressible Navier-Stokes equation. Despite a high-order geometric approximation, the robustness of their schemes is proved for viscous flows. An extension to three dimensions is proposed.

Reflection and mirroring ghost-cells methods

The reflection and mirroring ghost-cells is but an extension to the compressible hydrodynamics of the direct forcing methods using interpolation techniques. The underlying idea is that any smooth enough surface can be locally approximated by a plane. And that at the crossing of a plane, the normal velocity is anti-symmetrized whereas density, pressure, internal, kinetic and total energies are symmetrized. The mirroring method has been described by Forrer and Berger in [57]. The main idea resides in the fact that the wall acts as a mirror on the variables for a constant wall velocity. Introducing $\boldsymbol{t}$ and $\boldsymbol{n}$ as the tangential and normal vectors outward the boundary and $\boldsymbol{x_s}$ a point on the boundary, it yields for a small parameter λ that

$$\begin{cases} \rho(\boldsymbol{x_s}+\lambda\boldsymbol{n}) & = \rho(\boldsymbol{x_s}-\lambda\boldsymbol{n}) \\ p(\boldsymbol{x_s}+\lambda\boldsymbol{n}) & = p(\boldsymbol{x_s}-\lambda\boldsymbol{n}) \\ \boldsymbol{u}(\boldsymbol{x_s}+\lambda\boldsymbol{n})\cdot\boldsymbol{t} & = \boldsymbol{u}(\boldsymbol{x_s}-\lambda\boldsymbol{n})\cdot\boldsymbol{t} \\ \boldsymbol{u}(\boldsymbol{x_s}+\lambda\boldsymbol{n})\cdot\boldsymbol{n} & = 2\mathrm{D}_t\boldsymbol{x_s}-\boldsymbol{u}(\boldsymbol{x_s}+\lambda\boldsymbol{n})\cdot\boldsymbol{n} \end{cases} \tag{I.90}$$

The method is second order accurate at the boundary. Using a stencil inside the fluid domain, values of ρ, p and $\boldsymbol{u}$ are reconstructed on the blue points depicted in fig. I.11. Then, the value at the black points inside the solid domain are imposed using eq. (I.90). The fluid solver is then applied normally on the whole domain.

Similar methods have been introduced in [6], [23], [177]. As a contrary to the previously introduced cut-cells methods, the resulting scheme is not conservative in mass, momentum and total energy.

Inverse Lax–Wendroff procedure for boundary conditions

Thompson developed in [161], a high-order treatment of non-reflecting boundary conditions based

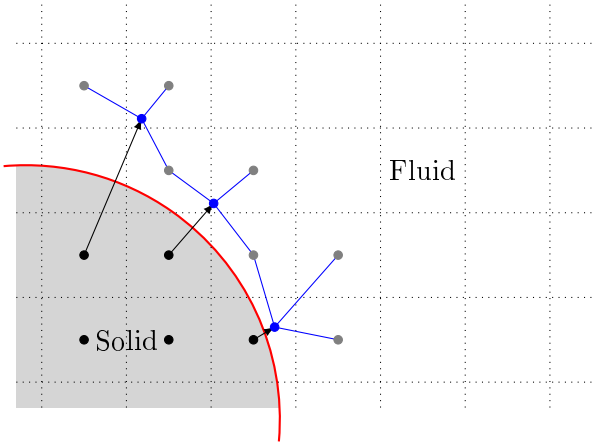


Figure I.11 – Ghost-cells values using the mirroring technique: black dots stand for the ghost-cells

on the diagonalization of the jacobian matrix. Thus getting the Riemann invariants of the system, the boundary conditions are imposed on those invariants. The values $(\rho, \boldsymbol{u}, p)$ are then set outside the fluid domain using the Riemann invariants. In [162], the author extended the method to any kind of boundary conditions and especially to the slip and no-slip wall boundary conditions. Changing the space derivatives of Riemann invariants into time derivatives, the author produces a high-order (up to 4th-order accuracy) boundary discretization. Later, Tan and Shu introduced the concept of inverse Lax–Wendroff procedure in order to treat boundary conditions in [155]. The idea is to solve repeatedly linear systems based on the jacobian matrix, in order to determine ghost-cells values outside the fluid domain. Those ghost-cells values are based on Taylor expansions of the primitive variables, using boundary conditions and values inside the fluid domain. Lax–Wendroff or Cauchy–Kovalevskaya methods are built by changing time-derivatives into space-derivatives in Taylor expansion in time of the flux function. Here the idea is to do exactly the inverse, meaning to change space-derivatives into time-derivatives in Taylor expansions in space of the primitive variables. In [156], the authors extended their previous results to the case of a moving boundary whose motion is triggered by the fluid state. This is a first step toward a fluid-structure interaction solver using Inverse Lax–Wendroff boundary treatment. The main difficulty in their articles is that the structure, considered as a rigid body, is described in a Lagrangian framework whereas the fluid solver follows an Eulerian approach. In [158], the authors attempted to reduce the numerical cost of their procedure by reducing the number of normal space derivatives changed into time and tangential space derivatives. Numerical experiments show that a certain number of normal space derivatives changes is enough to ensure *a priori* the stability of the effective scheme. In [157], the authors proposed a condensed review of their method, including applicability of the procedure. Last, Vilar and Shu in [168] developed a linear analysis of the scheme stability using the Inverse Lax–Wendroff procedure. They used the GKS theory (see lemma I.11) to analyse theoretically the stability of the effective schemes. They drew comparisons with the standard computation of the eigenvalues of the operator matrices. Similar results of required changed normal space derivatives as in [158] have been recovered. Moreover, for the considered schemes, GKS theory and the study of the eigenspectrum of the

operator matrix are similar. The method is detailed for linear systems in chapter III and then applied in the special case of Lagrange-Remap hydrodynamics schemes in chapter IV.

Chapter II

High order 2D finite volume conservative Lagrange-Remap schemes for compressible hydrodynamics on staggered Cartesian grids

---

On présente comment construire une famille de schémas volumes finis Lagrange-projection sur maillage décalé à l'ordre élevé. Ces schémas ont fait l'objet d'une note au comptes-rendus de l'Académie des Sciences [35]. Pour cela, la distribution originelle des variables sur la grille décalée Arakawa de type C est altérée pour des questions de robustesse et de conservation, tout d'abord en 1D puis en multi-dimensionnel. Pour l'extension en 1D à l'ordre élevé de ces schémas, des séquences de Runge–Kutta ont été choisies pour l'intégration en temps du système lagrangien, basé sur une formulation en énergies interne et cinétique. Une procédure conservative est développée à l'ordre élevé afin de corriger l'énergie interne et d'assurer la capture correcte des chocs. Le résultat principal de cette partie est le théorème II.9 qui prouve la consistance faible du schéma pour les équations d'Euler en référentiel lagrangien. Enfin, la projection conservative classique basée sur l'intégration analytique de polynômes de Lagrange est adaptée au cas des grilles décalées. Une extension en multidimensionnel est réalisée par l'utilisation de séquences d'ordre élevé de balayage directionnel. Enfin, la dérivation de ces schémas dans le cas des équations de Navier–Stokes compressibles, avec une distribution particulière des termes du tenseur visqueux, est faite. Des résultats numériques sont proposés tout au long du chapitre afin d'illustrer la précision et la robustesse de cette nouvelle famille de schémas.

We propose in this chapter a new class of finite volume numerical schemes on staggered Cartesian grids for solving the compressible hydrodynamics system of equations

$$\partial_t \begin{pmatrix} \rho \\ \rho \boldsymbol{u} \\ \rho e \end{pmatrix} + \nabla \cdot \begin{pmatrix} \rho \boldsymbol{u} \\ \rho \boldsymbol{u} \otimes \boldsymbol{u} + p \underline{\boldsymbol{I}} \\ (\rho e + p) \boldsymbol{u} \end{pmatrix} = \mathbf{0}. \tag{II.1}$$

The density, velocity, specific total energy and pressure are respectively denoted by ρ, $\boldsymbol{u}$, e and p. The schemes are based on 1D Lagrange-remap formalism used with directional splitting. The Lagrangian approach as well as the Lagrange-remap approach is the usual formalism used in the laboratory where my PhD thesis has taken place, as much for historical reasons as for robustness issues. They are high-order accurate in both time and space for any equation of states and are conservative in mass, momentum and total energy. The outline of the chapter is the following. First, using the Arakawa system of grids, a new grid is derived to ensure face-staggering of variables and robustness in case of shocks (section II-1). Second, the one-dimensional conservative Lagrange-remap schemes formulated in internal and kinetic energies are extended to higher order of accuracy (section II-2). The schemes may be decomposed into three steps that are detailed. The Lagrangian phase based on high-order reconstruction and interpolation of data to maintain high-order accuracy in space, and a Runge–Kutta time-integration to ensure the high-order accuracy in time. A new conservative and high-order accurate internal energy correction is proposed to ensure the correct capture of shocks. The main result of this chapter is theorem II.9 where the weak consistency of the scheme is proved. Conservative remapping phase is adapted to the staggered grids. It is based as for the cell-centered case on Lagrange polynomials. Third, the extension to two-dimensional systems is made using high-order directional splitting methods (section II-3). The 2D Lagrange-remap schemes on staggered Cartesian grids have been published in [35]. Fourth, a natural extension of the schemes in the case of Navier–Stokes compressible equation is made with gravity source terms. It is based on a special distribution of viscous terms to ensure robustness and high-order of accuracy (section II-4). Numerical results all along the chapter illustrate both the accuracy and the robustness of the schemes.

II-1 Structure of schemes on Arakawa C-type grids 57

II-1.1 Example of the BBC scheme 57

II-1.2 Discretized variables on Arakawa C-type grid 59

II-1.3 Definition of average and pointwise values 60

II-2 High order 1D Lagrange-Remap schemes on staggered Cartesian grids 60

II-2.1 Formulation of Runge–Kutta based Lagrangian finite volume schemes . 61

II-2.2 A new local internal energy corrector 72

II-2.3 The remapping stage 81

II-2.4 Numerical validation of the 1D conservative Lagrange-Remap schemes on staggered Cartesian grids 84

II-3 Extension to 2D Lagrange-remap schemes on staggered Cartesian grids 91
II-3.1 Derivation of the subsystems using the operator splitting technique . . . 92
II-3.2 Modifications of the 1D schemes for the 2D finite volume case 93
II-3.3 Numerical validation of the 2D conservative Lagrange-Remap schemes on staggered Cartesian grids 96
II-4 Extension to the 2D compressible Navier–Stokes equations with gravity 110
II-4.1 Distribution of viscous terms on the modified Arakawa grid 110
II-4.2 2D viscous staggered Lagrange-Remap schemes with gravity force 113
II-4.3 Numerical validation of the 2D staggered Lagrange-Remap schemes . . . 117

II-1 Structure of schemes on Arakawa C-type grids

In section I-2.2.1, the Arakawa system of classification for staggered grids has been presented. The BBC scheme which has been proposed in 1984 in [171] by Woodward and Collela will be extended to higher-order of accuracy in both time and space. The BBC scheme is based on a C-type Arakawa grid or face staggering. It means that the velocity are located on the face of the cells: an analysis of the space and time discretization is proposed.

II-1.1 Example of the BBC scheme

The scheme solves the Lagrange system (I.60) formulated in internal energy. On cell centers, the discretized variables are the cell mass Δm, the specific volume τ and the internal energy ϵ. On cell interfaces, the velocity u is discretized. The pressure is denoted p and artificial viscosities or bulk hyperviscosities (see section I-2.4) are denoted q. The interface mass $\Delta m_{i+\frac{1}{2}}$ is defined by

$$\Delta m_{i+\frac{1}{2}} = \frac{1}{2}(\Delta m_{i+1} + \Delta m_i). \tag{II.2}$$

The Lagrangian scheme writes in three steps:

Prediction at $t = t^{n+\frac{1}{4}}$

$$u_{i+\frac{1}{2}}^{n+\frac{1}{4}} = u_{i+\frac{1}{2}}^{n} - \frac{\Delta t}{4\Delta m_{i+\frac{1}{2}}}(p_{i+1}^n + q_{i+1}^n - p_i^n - q_i^n). \tag{II.3}$$

Prediction at $t = t^{n+\frac{1}{2}}$

$$\begin{cases} \tau_i^{n+\frac{1}{2}} = \tau_i^n + \frac{\Delta t}{2\Delta m_i}(u_{i+\frac{1}{2}}^{n+\frac{1}{4}} - u_{i-\frac{1}{2}}^{n+\frac{1}{4}}), \\ \epsilon_i^{n+\frac{1}{2}} = \epsilon_i^n - \frac{\Delta t}{2\Delta m_i}(p_i^n + q_i^n)(u_{i+\frac{1}{2}}^{n+\frac{1}{4}} - u_{i-\frac{1}{2}}^{n+\frac{1}{4}}), \\ p_i^{n+\frac{1}{2}} = EOS(\tau_i^{n+\frac{1}{2}}, \epsilon_i^{n+\frac{1}{2}}), \\ u_{i+\frac{1}{2}}^{n+\frac{1}{2}} = u_{i+\frac{1}{2}}^n - \frac{\Delta t}{2\Delta m_{i+\frac{1}{2}}}(p_{i+1}^{n+\frac{1}{2}} + q_{i+1}^n - p_i^{n+\frac{1}{2}} - q_i^n). \end{cases} \tag{II.4}$$

Prediction at $t = t^{n+1}$

$$\begin{cases} \tau_i^{n+1} = \tau_i^n + \frac{\Delta t}{\Delta m_i}(u_{i+\frac{1}{2}}^{n+\frac{1}{2}} - u_{i-\frac{1}{2}}^{n+\frac{1}{2}}), \\ \epsilon_i^{n+\frac{1}{2}} = \epsilon_i^n - \frac{\Delta t}{\Delta m_i}(p_i^{n+\frac{1}{2}} + q_i^n)(u_{i+\frac{1}{2}}^{n+\frac{1}{2}} - u_{i-\frac{1}{2}}^{n+\frac{1}{2}}), \\ p_i^{n+1} = EOS(\tau_i^{n+1}, \epsilon_i^{n+1}), \\ u_{i+\frac{1}{2}}^{n+1} = 2u_{i+\frac{1}{2}}^{n+\frac{1}{2}} - u_{i+\frac{1}{2}}^n, \\ x_{i+\frac{1}{2}}^{n+1} = x_{i+\frac{1}{2}}^n + \Delta t u_{i+\frac{1}{2}}^{n+\frac{1}{2}}. \end{cases} \tag{II.5}$$

The first prediction done in eq. (II.3) is used to stabilize the scheme. Using Arakawa C-type grids, 2^{nd}order Runge–Kutta sequences are not stable for the wave equations, as it will be shown later on. The choice to made the first predictor at $t = t^{n+\frac{1}{4}}$ on the velocity rather that on the pressure allows to reduce the number of call to the equation of state. Another interesting choice is the velocity obtained at t^{n+1}. This choice is made to obtain a compatible discretization of the kinetic energy in the sense of Caramana [19]. Doing so, it allows to get the following results

Lemma II.1 (Conservation properties of the BBC scheme). *The BBC scheme* (II.2)-(II.3)-(II.4)-(II.5) *is conservative in mass, momentum and total energy for any choice of artificial viscosities or hyperviscosities. The total energy of a cell is defined here as*

$$\Delta m_i e_i^n = \Delta m_i \epsilon_i^n + \frac{1}{2}\left(\Delta m_{i+\frac{1}{2}} e_{kin}{}_{i+\frac{1}{2}}^n + \Delta m_{i-\frac{1}{2}} e_{kin}{}_{i-\frac{1}{2}}^n\right),$$

with the kinetic energy defined as $e_{kin}{}_{i+\frac{1}{2}} = \frac{1}{2}(u_{i+\frac{1}{2}})^2$.

Proof. Mass and momentum conservation are obvious. Only total energy conservation is detailed.

$$
\begin{aligned}
\Delta m_i({e_i}^{n+1} - {e_i}^{n}) &= \Delta m_i({\epsilon_i}^{n+1} - {\epsilon_i}^{n}) \\
&\quad + \tfrac{1}{2}\left(\Delta m_{i+\frac{1}{2}}({e_{\text{kin}}}_{i+\frac{1}{2}}^{n+1} - {e_{\text{kin}}}_{i+\frac{1}{2}}^{n}) + \Delta m_{i-\frac{1}{2}}({e_{\text{kin}}}_{i-\frac{1}{2}}^{n+1} - {e_{\text{kin}}}_{i-\frac{1}{2}}^{n})\right) \\
&= -\Delta t(p_i^{n+\frac{1}{2}} + q_i^n)({u_{i+\frac{1}{2}}}^{n+\frac{1}{2}} - {u_{i-\frac{1}{2}}}^{n+\frac{1}{2}}) \\
&\quad + \tfrac{1}{4}\left(\Delta m_{i+\frac{1}{2}}(u_{i+\frac{1}{2}}^{n+1} - u_{i+\frac{1}{2}}^{n})(u_{i+\frac{1}{2}}^{n+1} + u_{i+\frac{1}{2}}^{n}) + \Delta m_{i-\frac{1}{2}}(u_{i-\frac{1}{2}}^{n+1} - u_{i-\frac{1}{2}}^{n})(u_{i-\frac{1}{2}}^{n+1} + u_{i-\frac{1}{2}}^{n})\right) \\
&= -\Delta t(p_i^{n+\frac{1}{2}} + q_i^n)(u_{i+\frac{1}{2}}^{n+\frac{1}{2}} - u_{i-\frac{1}{2}}^{n+\frac{1}{2}}) \\
&\quad - \tfrac{\Delta t}{2}(p_{i+1}^{n+\frac{1}{2}} + q_{i+1}^n - p_i^{n+\frac{1}{2}} - q_i^n)u_{i+\frac{1}{2}}^{n+\frac{1}{2}} \\
&\quad - \tfrac{\Delta t}{2}(p_i^{n+\frac{1}{2}} + q_i^n - p_{i-1}^{n+\frac{1}{2}} - q_{i-1}^n)u_{i-\frac{1}{2}}^{n+\frac{1}{2}} \\
&= -\Delta t\left(\frac{p_{i+1}^{n+\frac{1}{2}} + p_i^{n+\frac{1}{2}} + q_{i+1}^n + q_i^n}{2}u_{i+\frac{1}{2}}^{n+\frac{1}{2}} - \frac{p_i^{n+\frac{1}{2}} + p_{i-1}^{n+\frac{1}{2}} + q_i^n + q_{i-1}^n}{2}u_{i-\frac{1}{2}}^{n+\frac{1}{2}}\right)
\end{aligned}
$$

Hence total energy conservation is obtained due to the flux form of $\Delta m_i({e_i}^{n+1} - {e_i}^{n})$. ■

Remark II.1. The proof for total energy conservation uses special features of the BBC scheme, especially that $u_{i+\frac{1}{2}}^{n+1} = 2u_{i+\frac{1}{2}}^{n+\frac{1}{2}} - u_{i+\frac{1}{2}}^{n}$. Without this special relation between the velocities at different time steps, total energy conservation does not hold.

Remark II.2. In recent works by Herbin, Latché and al. [83, 80, 81], they propose an *a priori* internal energy corrector. This corrector is based on the computation of a residual term obtained using the discretization of the kinetic energy. Here, for the BBC scheme, the residual term obtained is exactly 0. Then, no special energy balance is to be performed in the Lagrangian phase.

In other words, it means that changing the time integration has a strong impact on the total energy conservation property of the scheme. The idea to be able to deal with any time-integration sequences is to discretize the kinetic energy and to evolve it using its evolution equation. This way conservation of total energy will be ensured.

II-1.2 Discretized variables on Arakawa C-type grid

In order to extend the BBC scheme at high-order in both time and space, the method used in this work is based on the analysis of the kinetic energy equation, in a way that ensures total energy conservation which appears more as a compatibility relation. The kinetic energy evolution equation writes formally

$$\partial_t \rho_0 e_{\text{kin}} + u\partial_X p = 0 \tag{II.6}$$

To form the total energy equation, it is sufficient to combine with the internal energy one which writes as

$$\partial_t \rho_0 \epsilon + p\partial_X u = 0 \tag{II.7}$$

The use of the Lagrangian kinetic energy eq. (II.6) is unusual with respect to the literature.

Another difference with the BBC scheme is that the masses will be decoupled between the centered and staggered grids, meaning that eq. (II.2) will not be satisfied. The discretization is summarized in fig. II.1. For example in 1D, two mass variables are considered, one located at the center of each cells, and one at the center of each staggered cells. To our knowledge, such a choice is also not usual in the literature.

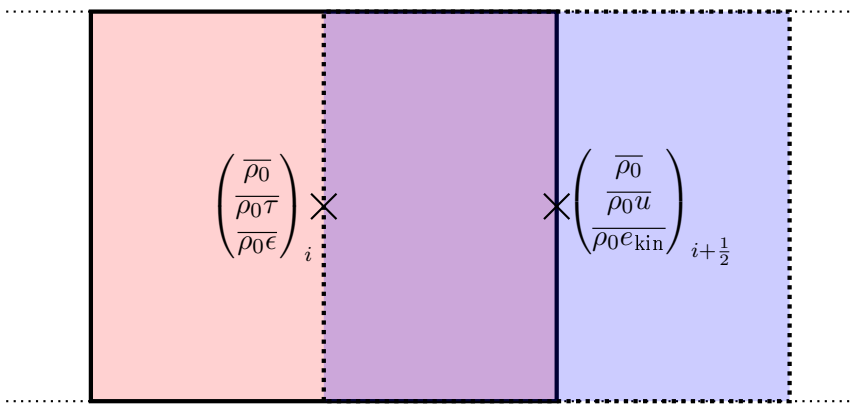


Figure II.1 – Staggered finite volume space discretization on Cartesian grids

II-1.3 Definition of average and pointwise values

Consider a *primal* uniform Cartesian grid $\{x_{i+\frac{1}{2}}\}$ with $\Delta X = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$ and a *dual* grid $\{x_i\}$ with $x_i = \frac{1}{2}(x_{i+\frac{1}{2}} + x_{i-\frac{1}{2}})$. As presented in eq. (II.8), $\overline{\phi}$ and ϕ will respectively denote the space averaged value of ϕ and its point-wise value.

$$\begin{cases} \overline{\phi}_i^n & = \frac{1}{\Delta X} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \phi(x, t^n) dx, \\ \overline{\phi}_{i+\frac{1}{2}}^n & = \frac{1}{\Delta X} \int_{x_i}^{x_{i+1}} \phi(x, t^n) dx, \\ \phi_i^n & = \phi(x_i, t^n), \\ \phi_{i+\frac{1}{2}}^n & = \phi(x_{i+\frac{1}{2}}, t^n). \end{cases} \tag{II.8}$$

II-2 High order 1D Lagrange-Remap schemes on staggered Cartesian grids

Here we propose a constructive path to build high-order 1D finite volume conservative Lagrange-remap schemes. Firstly, the formulation of Runge–Kutta based Lagrangian finite volume schemes on staggered Cartesian grids is introduced. Secondly, an internal energy corrector is detailed. This corrector is conservative, high-order accurate and yields consistency of the scheme in case of strong shocks. Thirdly, a Lagrange polynomials based conservative remapping is extended to the special case of staggered grids. Last, numerical experiments show accuracy and robustness of the method on various numerical examples presented in the literature. This section has been

the object of a publication [35] in the "Comptes-Rendus Mathématique".

II-2.1 Formulation of Runge–Kutta based Lagrangian finite volume schemes

The Lagrangian system formulated in kinetic and internal energies is reminded hereafter, using q as the artificial viscosity as detailed in section I-2.4.

$$\begin{cases} \mathrm{D}_t\rho_0\tau - \partial_X u &= 0, \\ \mathrm{D}_t\rho_0 u + \partial_X(p+q) &= 0, \\ \mathrm{D}_t\rho_0\epsilon + (p+q)\partial_X u &= 0, \\ \mathrm{D}_t\rho_0 e_{\mathrm{kin}} + u\partial_X(p+q) &= 0, \\ p &= EOS(\tau,\epsilon). \end{cases} \tag{II.9}$$

II-2.1.1 Semi-discrete formulation of the Lagrangian finite volume schemes

To get the semi-discrete formulation of the Lagrangian finite volume schemes, system depicted in eq. (II.9) is integrated in time between t^n and t^{n+1} over a cell $\left[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}\right]$ for the thermodynamics variables $\rho_0\tau$ and $\rho_0\epsilon$ and over a cell $[x_i, x_{i+1}]$ for the $\rho_0 u$ and ρe_{kin}. It yields

$$\begin{cases} \Delta X(\overline{\rho_0\tau}_i^{n+1} - \overline{\rho_0\tau}_i^{n}) &= \int_{t^n}^{t^{n+1}} u_{i+\frac{1}{2}}(\theta) - u_{i-\frac{1}{2}}(\theta)\mathrm{d}\theta, \\ \Delta X(\overline{\rho_0 u}_{i+\frac{1}{2}}^{n+1} - \overline{\rho_0 u}_{i+\frac{1}{2}}^{n}) &= \int_{t^n}^{t^{n+1}} (p+q)_{i+1}(\theta) - (p+q)_i(\theta)\mathrm{d}\theta, \\ \Delta X(\overline{\rho_0\epsilon}_i^{n+1} - \overline{\rho_0\epsilon}_i^{n}) &= \int_{t^n}^{t^{n+1}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} ((p+q)\partial_X u)(y,\theta)\mathrm{d}y\mathrm{d}\theta, \\ \Delta X(\overline{\rho_0 e_{\mathrm{kin}}}_{i+\frac{1}{2}}^{n+1} - \overline{\rho_0 e_{\mathrm{kin}}}_{i+\frac{1}{2}}^{n}) &= \int_{t^n}^{t^{n+1}} \int_{x_i}^{x_{i+1}} (u\partial_X(p+q))(y,\theta)\mathrm{d}y\mathrm{d}\theta, \\ p_i &= EOS(\tau_i,\epsilon_i). \end{cases} \tag{II.10}$$

Notations $\overline{(p+q)\delta u}_i$ and $\overline{u\delta(p+q)}_{i+\frac{1}{2}}$ are introduced as

$$\overline{(p+q)\delta u}_i = \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} ((p+q)\partial_X u)(y,\theta)\mathrm{d}y,$$
$$\overline{u\delta(p+q)}_{i+\frac{1}{2}} = \int_{x_i}^{x_{i+1}} (u\partial_X(p+q))(y,\theta)\mathrm{d}y.$$

So that eq. (II.11) rewrites

$$\begin{cases} \Delta X(\overline{\rho_0\tau}_i^{n+1} - \overline{\rho_0\tau}_i^{n}) &= \int_{t^n}^{t^{n+1}} u_{i+\frac{1}{2}}(\theta) - u_{i-\frac{1}{2}}(\theta)\mathrm{d}\theta, \\ \Delta X(\overline{\rho_0 u}_{i+\frac{1}{2}}^{n+1} - \overline{\rho_0 u}_{i+\frac{1}{2}}^{n}) &= \int_{t^n}^{t^{n+1}} (p+q)_{i+1}(\theta) - (p+q)_i(\theta)\mathrm{d}\theta, \\ \Delta X(\overline{\rho_0\epsilon}_i^{n+1} - \overline{\rho_0\epsilon}_i^{n}) &= \int_{t^n}^{t^{n+1}} \overline{(p+q)\delta u}_i(\theta)\mathrm{d}\theta, \\ \Delta X(\overline{\rho_0 e_{\mathrm{kin}}}_{i+\frac{1}{2}}^{n+1} - \overline{\rho_0 e_{\mathrm{kin}}}_{i+\frac{1}{2}}^{n}) &= \int_{t^n}^{t^{n+1}} \overline{u\delta(p+q)}_{i+\frac{1}{2}}(\theta)\mathrm{d}\theta, \\ p_i &= EOS(\tau_i,\epsilon_i). \end{cases} \tag{II.11}$$

Before performing any kind of time integration, one must first address the issue of computing with high-order accuracy the point-wise values of p, u, τ and ϵ.

II-2.1.2 High-order in spatial reconstruction of pointwise values from averages ones and *vice versa* and of space derivatives

To achieve high-order resolution, it is mandatory to compute the point-wise (resp. average) values from the average (resp. point-wise) ones with high-order accuracy. Table II.1 gives the coefficients for centered, symmetric and polynomial reconstructions using eq. (II.12). Although other reconstructions may be used, centered and symmetric ones are retained here and are sufficient for uniform Cartesian grids.

$$\left\{\begin{array}{ll}
\phi_{\xi(i)} & = \sum_k C_k \overline{\phi}_{\xi(i)+k}, \\
\overline{\phi}_{\xi(i)} & = \sum_k \widehat{C}_k \phi_{\xi(i)+k}, \\
\delta\phi_{\xi(i)} & = \sum_{k\geq 0} d_k \left(\phi_{\xi(i)+k+\frac{1}{2}} - \phi_{\xi(i)-k-\frac{1}{2}}\right), \quad \text{with } \xi(i) = \left\{\begin{array}{ll} i & \text{on primal grid,} \\ i+\frac{1}{2} & \text{on dual grid,} \end{array}\right. \\
\phi_{\xi(i)} & = \sum_k r_k(\phi_{\xi(i)+k+\frac{1}{2}} + \phi_{\xi(i)-k-\frac{1}{2}}), \\
\phi_{\xi(i)} & = \dfrac{(\rho_0\phi)_{\xi(i)}}{(\rho_0)_{\xi(i)}},
\end{array}\right. \tag{II.12}$$

The non-conservative terms $\overline{\psi\delta\phi}$ of eq. (II.11) are computed by:

1. Applying the δ operator to point-wise values of ϕ using coefficients in table II.3 and third equation of (II.12).
2. Multiplying by point-wise values of ψ, then reconstructing average values using coefficients in table II.2 and second equation of (II.12).

Order	C_0	$C_{\pm 1}$	$C_{\pm 2}$	$C_{\pm 3}$	$C_{\pm 4}$
2nd	1	0	0	0	0
3rd	$\frac{13}{12}$	$\frac{-1}{24}$	0	0	0
4th and 5th	$\frac{1067}{960}$	$\frac{-29}{480}$	$\frac{3}{640}$	0	0
6th and 7th	$\frac{30251}{26880}$	$\frac{-7621}{107520}$	$\frac{159}{17920}$	$\frac{-5}{7168}$	0
8th and 9th	$\frac{5851067}{5160960}$	$\frac{-100027}{1290240}$	$\frac{31471}{2580480}$	$\frac{-425}{258048}$	$\frac{35}{294912}$

Table II.1 – Coefficients for the finite volume computation of point-wise values from cell-average ones.

II-2.1.3 Runge–Kutta based time discretization

We consider Nth order *explicit* schemes with s sub-cycles with the following notations for Runge-Kutta sequences: α_m is the time step for the mth sub-cycle, $a_{m,l}$ the m, l term of the Butcher table and θ_l the lth reconstruction coefficient for the last step. It is represented by the table presented in table II.5. The sequences are available in appendix in section A.1. We denote the

Order	$\widehat{C}_0$	$\widehat{C}_{\pm 1}$	$\widehat{C}_{\pm 2}$	$\widehat{C}_{\pm 3}$	$\widehat{C}_{\pm 4}$
2nd	1	0	0	0	0
3rd	$\frac{11}{12}$	$\frac{1}{24}$	0	0	0
4th and 5th	$\frac{863}{960}$	$\frac{77}{1440}$	$\frac{-17}{5760}$	0	0
6th and 7th	$\frac{215641}{241920}$	$\frac{6361}{107520}$	$\frac{-281}{53760}$	$\frac{367}{967680}$	0
8th and 9th	$\frac{41208059}{46448640}$	$\frac{3629953}{58060800}$	$\frac{-801973}{116121600}$	$\frac{49879}{58060800}$	$\frac{-27859}{464486400}$

Table II.2 – Coefficients for the finite volume computation of average values from point-wise ones.

Order	d_0	d_1	d_2	d_3	d_4
2nd	1	0	0	0	0
3rd	$\frac{9}{8}$	$\frac{-1}{24}$	0	0	0
4th and 5th	$\frac{75}{64}$	$\frac{-25}{384}$	$\frac{3}{640}$	0	0
6th and 7th	$\frac{1225}{1024}$	$\frac{-245}{3072}$	$\frac{49}{5120}$	$\frac{-5}{7168}$	0
8th and 9th	$\frac{19845}{16384}$	$\frac{-735}{8192}$	$\frac{567}{40960}$	$\frac{-405}{229376}$	$\frac{35}{294912}$

Table II.3 – Coefficients for the δ operator.

sum of artificial viscosity and pressure as $\Pi = p + q$. The system (II.13) details one Runge-Kutta sub-cycle at time $t^{n+\alpha_m}$ and (II.14) details the final step at time t^{n+1}:

$$\begin{cases} \overline{\rho_0 \tau}_i^{n+\alpha_m} = \overline{\rho_0 \tau}_i^n + \frac{\Delta t}{\Delta X} \sum_{l=0}^{m-1} a_{m,l} du_i^{n+\alpha_l}, \\ \overline{\rho_0 u}_{i+\frac{1}{2}}^{n+\alpha_m} = \overline{\rho_0 u}_{i+\frac{1}{2}}^n - \frac{\Delta t}{\Delta X} \sum_{l=0}^{m-1} a_{m,l} d\Pi_{i+\frac{1}{2}}^{n+\alpha_l}, \\ \overline{\rho_0 \epsilon}_i^{n+\alpha_m} = \overline{\rho_0 \epsilon}_i^n - \frac{\Delta t}{\Delta X} \sum_{l=0}^{m-1} a_{m,l} \overline{\Pi \delta u}_i^{n+\alpha_l}, \\ p_i^{n+\alpha_m} = EOS(\tau_i^{n+\alpha_m}, \epsilon_i^{n+\alpha_m}), \end{cases} \tag{II.13}$$

Here, $d\phi$ is the difference between two consecutive point-wise values: $d\phi_i = \phi_{i+\frac{1}{2}} - \phi_{i-\frac{1}{2}}$ and $d\phi_{i+\frac{1}{2}} = \phi_{i+1} - \phi_i$. Note that in (II.13), *ie* for intermediate Runge–Kutta time-step, there is no need to compute the evolution of the kinetic energy, nor the position of the cells face $x_{i+\frac{1}{2}}$.

Order	r_0	r_1	r_2	r_3	r_4
2nd	$\frac{1}{2}$	0	0	0	0
3rd	$\frac{9}{16}$	$\frac{-1}{16}$	0	0	0
4th and 5th	$\frac{75}{128}$	$\frac{-25}{256}$	$\frac{3}{256}$	0	0
6th and 7th	$\frac{1225}{2048}$	$\frac{-245}{2048}$	$\frac{49}{2048}$	$\frac{-5}{2048}$	0
8th and 9th	$\frac{19845}{32768}$	$\frac{-2205}{16384}$	$\frac{567}{16384}$	$\frac{-405}{65536}$	$\frac{35}{65536}$

Table II.4 – Coefficients for the interpolation of cell-centered values from staggered ones and *vice-versa*.

$$\begin{array}{c|ccccc}
\alpha_1 & a_{1,0} & 0 & 0 & 0 & \cdots \\
\alpha_2 & a_{2,0} & a_{2,1} & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \ddots & \cdots & \cdots \\
\alpha_s & a_{s,0} & \cdots & \cdots & a_{s,s-1} & 0 \\
\hline
1 & \theta_0 & \theta_1 & \cdots & \theta_{s-1} & \theta_s
\end{array}$$

Table II.5 – Example of Butcher table for explicit Runge–Kutta sequence with s sub-cycles.

$$\begin{cases}
\overline{\rho_0\tau}_i^{n+1} &= \overline{\rho_0\tau}_i^n + \frac{\Delta t}{\Delta X}\sum_{l=0}^{s}\theta_l du_i^{n+\alpha_l}, \\
\overline{\rho_0 u}_{i+\frac{1}{2}}^{n+1} &= \overline{\rho_0 u}_{i+\frac{1}{2}}^n - \frac{\Delta t}{\Delta X}\sum_{l=0}^{s}\theta_l d\Pi_{i+\frac{1}{2}}^{n+\alpha_l}, \\
\overline{\rho_0\epsilon}_i^{n+1} &= \overline{\rho_0\epsilon}_i^n - \frac{\Delta t}{\Delta X}\sum_{l=0}^{s}\theta_l \overline{\Pi\delta u}_i^{n+\alpha_l}, \\
\overline{\rho_0 e_{kin}}_{i+\frac{1}{2}}^{n+1} &= \overline{\rho_0 e_{kin}}_{i+\frac{1}{2}}^n - \frac{\Delta t}{\Delta X}\sum_{l=0}^{s}\theta_l \overline{u\delta\Pi}_{i+\frac{1}{2}}^{n+\alpha_l}, \\
x_{i+\frac{1}{2}}^{n+1} &= x_{i+\frac{1}{2}}^n + \Delta t\sum_{l=0}^{s}\theta_l u_{i+\frac{1}{2}}^{n+\alpha_l}, \\
p_i^{n+1} &= EOS(\tau_i^{n+1}, \epsilon_i^{n+1}).
\end{cases} \tag{II.14}$$

II-2.1.4 Properties of the staggered schemes (II.13)-(II.14)

Two definitions of total energies are introduced in order to study the schemes properties concerning the conservation of total energy.

Definition II.1. The total energy, based on the kinetic energy reconstructed from the momentum, of the system at time $t = t^n$, denoted E^n, is defined as

$$E^n = \Delta X\left(\sum_i \overline{\rho_0\epsilon}_i^n + \sum_i \overline{\rho_0 u^2}_{i+\frac{1}{2}}^n\right). \tag{II.15}$$

Definition II.2. A total energy, based on the discretized kinetic energy, of the system at time

$t = t^n$, denoted $\mathcal{E}^n$, is defined as

$$\mathcal{E}^n = \Delta X \left(\sum_i \overline{\rho_0 \epsilon_i}^n + \sum_i \overline{\rho_0 e_{\text{kin},u}}^n_{i+\frac{1}{2}} \right). \tag{II.16}$$

A desired feature is that the mass, momentum and the total energy E^n defined in definition II.1 are conserved for periodic or wall boundary conditions, meaning that $E^{n+1} - E^n = 0$. However using schemes (II.13)-(II.14), the total energy E is not conserved. Here, as mentioned in lemma II.2, the schemes conserve the total energy $\mathcal{E}^n$ defined in definition II.2.

Lemma II.2 (Conservation of the staggered schemes (II.13)-(II.14))**.** *For all explicit Runge-Kutta sequences, all artificial viscosities, all spatial reconstructions, the schemes* (II.13)-(II.14) *formulated in internal energy are conservative in mass, momentum and total energy $\mathcal{E}$ defined in definition II.2.*

Proof. Conservation of mass and momentum is straightforward. We only prove the conservation of total energy.

$$\begin{aligned}
\mathcal{E}^{n+1} - \mathcal{E}^n &= \sum_i \left(\overline{\rho_0 \epsilon_i}^{n+1} - \overline{\rho_0 \epsilon_i}^n \right) + \sum_i \left(\overline{\rho_0 e_{kin}}^{n+1}_{i+\frac{1}{2}} - \overline{\rho_0 e_{kin}}^n_{i+\frac{1}{2}} \right) \\
&= -\frac{\Delta t}{\Delta X} \sum_i \sum_{l=1}^{s} \theta_l \left(\overline{\Pi \delta u}_i^{n+\alpha_l} + \overline{u \delta \Pi}_{i+\frac{1}{2}}^{n+\alpha_l} \right) \\
&= -\frac{\Delta t}{\Delta X} \sum_i \sum_{l=1}^{s} \sum_k \sum_{k'} \theta_l \widehat{C}_k d_{k'} (\Pi_{i+k}^{n+\alpha_l} u_{i+k+k'+\frac{1}{2}}^{n+\alpha_l} + u_{i+k+\frac{1}{2}}^{n+\alpha_l} \Pi_{i+k+k'+1}^{n+\alpha_l} \\
&\qquad - \Pi_{i+k}^{n+\alpha_l} u_{i+k-k'-\frac{1}{2}}^{n+\alpha_l} - u_{i+k+\frac{1}{2}}^{n+\alpha_l} \Pi_{i+k-k'}^{n+\alpha_l}).
\end{aligned}$$

Making the change of index $i \leftarrow i + k'$ in the first term and $i \leftarrow i + k' + 1$ in the second term of the RHS we get the result for wall (with non-trivial definitions of ghost-cell values) or periodic boundary conditions.

$$\begin{aligned}
\mathcal{E}^{n+1} - \mathcal{E}^n = -\frac{\Delta t}{\Delta X} \sum_i \sum_{l=1}^{s} \sum_k \sum_{k'} \theta_l \widehat{C}_k d_{k'} (& \Pi_{i+k-k'}^{n+\alpha_l} u_{i+k+\frac{1}{2}}^{n+\alpha_l} + u_{i+k-k'-\frac{1}{2}}^{n+\alpha_l} \Pi_{i+k}^{n+\alpha_l} \\
- & \Pi_{i+k}^{n+\alpha_l} u_{i+k-k'-\frac{1}{2}}^{n+\alpha_l} - u_{i+k+\frac{1}{2}}^{n+\alpha_l} \Pi_{i+k-k'}^{n+\alpha_l}) = 0.
\end{aligned}$$

■

We introduce the barotropic version of the staggered schemes: the intermediate stages write

$$\begin{cases} \overline{\rho_0\tau}_i^{n+\alpha_m} = \overline{\rho_0\tau}_i^n + \frac{\Delta t}{\Delta X}\sum_{l=0}^{m-1} a_{m,l}(u_{i+\frac{1}{2}}^{n+\alpha_l} - u_{i-\frac{1}{2}}^{n+\alpha_l}), \\ \overline{\rho_0 u}_{i+\frac{1}{2}}^{n+\alpha_m} = \overline{\rho_0 u}_{i+\frac{1}{2}}^n - \frac{\Delta t}{\Delta X}\sum_{l=0}^{m-1} a_{m,l}(p_{i+1}^{n+\alpha_l} - p_i^{n+\alpha_l}), \\ p_i^{n+\alpha_m} = EOS(\tau_i^{n+\alpha_m}), \end{cases} \tag{II.17}$$

and the final stage writes

$$\begin{cases} \overline{\rho_0\tau}_i^{n+1} = \overline{\rho_0\tau}_i^n + \frac{\Delta t}{\Delta X}\sum_{l=0}^{s} \theta_l(u_{i+\frac{1}{2}}^{n+\alpha_l} - u_{i-\frac{1}{2}}^{n+\alpha_l}), \\ \overline{\rho_0 u}_{i+\frac{1}{2}}^{n+1} = \overline{\rho_0 u}_{i+\frac{1}{2}}^n - \frac{\Delta t}{\Delta X}\sum_{l=0}^{s} \theta_l(p_{i+1}^{n+\alpha_l} - p_i^{n+\alpha_l}), \\ x_{i+\frac{1}{2}}^{n+1} = x_{i+\frac{1}{2}}^n + \Delta t\sum_{l=0}^{s} \theta_l u_{i+\frac{1}{2}}^{n+\alpha_l}, \\ p_i^{n+1} = EOS(\tau_i^{n+1}). \end{cases} \tag{II.18}$$

Lemma II.3 (Weak consistency of the barotropic Lagrangian staggered schemes (II.17)-(II.18))**.** *For all explicit Runge–Kutta sequences, all consistent spatial reconstructions, the schemes* (II.17)-(II.18) *are weakly consistent.*

Proof. Here we use the fact that a scheme whose flux is consistent (definition I.6) is weakly consistent (definition I.7). This is why we have to verify that the scheme can be rewritten under the form (I.34)-(I.35)-(I.36).

From equation (II.18), one can define the natural flux

$$f_{i+\frac{1}{2}}^{\star} = \sum_{l=0}^{s} \theta_l \begin{pmatrix} -u_{i+\frac{1}{2}}^{n+\alpha_l} \\ p_{i+1}^{n+\alpha_l} \end{pmatrix},$$

and the intermediate fluxes are defined from (II.17)

$$f_{i+\frac{1}{2}}^{\alpha_m} = \sum_{l=0}^{m-1} a_{m,l} \begin{pmatrix} -u_{i+\frac{1}{2}}^{n+\alpha_l} \\ p_{i+1}^{n+\alpha_l} \end{pmatrix}.$$

The proof is done by induction on the intermediate time-steps. First one proves that the intermediate (resp. natural) flux can be written as $\Phi^m(\boldsymbol{U}_{i-mr+1}, ..., \boldsymbol{U}_{i+mr+1})$ (resp. $\Phi^\star(\boldsymbol{U}_{i-(s+1)r+1}, ..., \boldsymbol{U}_{i+(s+1)r+1})$)

. Second, one proves that Φ^m (resp. $\Phi^\star$) satisfies for constant state $(\rho_0\tau, \rho_0, \rho_0 u, \widehat{\rho_0})^t$

$$\left\{\begin{array}{rcl} \Phi^m(\begin{pmatrix}\rho_0\tau\\ \rho_0\\ \rho_0 u\\ \widehat{\rho_0}\end{pmatrix}, ..., \begin{pmatrix}\rho_0\tau\\ \rho_0\\ \rho_0 u\\ \widehat{\rho_0}\end{pmatrix}) & = & \alpha_m \begin{pmatrix}-u\\ p\end{pmatrix}, \\ \Phi^\star(\begin{pmatrix}\rho_0\tau\\ \rho_0\\ \rho_0 u\\ \widehat{\rho_0}\end{pmatrix}, ..., \begin{pmatrix}\rho_0\tau\\ \rho_0\\ \rho_0 u\\ \widehat{\rho_0}\end{pmatrix}) & = & \begin{pmatrix}-u\\ p\end{pmatrix}. \end{array}\right.$$

We start the proof considering the first intermediate time-step. One has

$$f^{\alpha_1}_{i+\frac{1}{2}} = a_{1,0}\begin{pmatrix}-u^n_{i+\frac{1}{2}}\\ p^n_{i+1}\end{pmatrix}, \quad \alpha_1 = a_{1,0}$$

where

$$\left\{\begin{array}{rcl} u^n_{i+\frac{1}{2}} & = & \dfrac{\displaystyle\sum_{k=-r}^{r} C_k \overline{\rho_0 u}^n_{i+\frac{1}{2}+k}}{\displaystyle\sum_{k=-r}^{r} C_k \overline{\rho_0}_{i+\frac{1}{2}+k}}, \\ p^n_i & = & p\left(\dfrac{\displaystyle\sum_{k=-r}^{r} C_k \overline{\rho_0\tau}^n_{i+k}}{\displaystyle\sum_{k=-r}^{r} C_k \overline{\rho_0}_{i+k}}\right). \end{array}\right.$$

Hence, one can write $f^{\alpha_1}_{i+\frac{1}{2}}$ as a function Φ^1 with

$$f^{\alpha_1}_{i+\frac{1}{2}} = \Phi^1(\begin{pmatrix}\overline{\rho_0\tau}^n_{i-r+1}\\ \overline{\rho_0}_{i-r+1}\\ \overline{\rho_0 u}^n_{i+\frac{1}{2}-r}\\ \overline{\rho_0}_{i+\frac{1}{2}-r}\end{pmatrix}, ..., \begin{pmatrix}\overline{\rho_0\tau}^n_{i+r+1}\\ \overline{\rho_0}_{i+r+1}\\ \overline{\rho_0 u}^n_{i+\frac{1}{2}+r}\\ \overline{\rho_0}_{i+\frac{1}{2}+r}\end{pmatrix}).$$

The function Φ^1 writes

$$\Phi^1(\begin{pmatrix}\overline{\rho_0\tau}^n_{i-r+1}\\ \overline{\rho_0}_{i-r+1}\\ \overline{\rho_0 u}^n_{i+\frac{1}{2}-r}\\ \overline{\rho_0}_{i+\frac{1}{2}-r}\end{pmatrix}, ..., \begin{pmatrix}\overline{\rho_0\tau}^n_{i+r+1}\\ \overline{\rho_0}_{i+r+1}\\ \overline{\rho_0 u}^n_{i+\frac{1}{2}+r}\\ \overline{\rho_0}_{i+\frac{1}{2}+r}\end{pmatrix}) = a_{1,0}\begin{pmatrix} -\dfrac{\displaystyle\sum_{k=-r}^{r} C_k \overline{\rho_0 u}^n_{i+\frac{1}{2}+k}}{\displaystyle\sum_{k=-r}^{r} C_k \overline{\rho_0}_{i+\frac{1}{2}+k}} \\ p\left(\dfrac{\displaystyle\sum_{k=-r}^{r} C_k \overline{\rho_0\tau}^n_{i+k}}{\displaystyle\sum_{k=-r}^{r} C_k \overline{\rho_0}_{i+k}}\right)\end{pmatrix}.$$

Hence, for constant state $(\rho_0\tau, \rho_0, \rho_0 u, \widehat{\rho_0})^t$

$$\Phi^1(\begin{pmatrix}\rho_0\tau\\ \rho_0\\ \rho_0 u\\ \widehat{\rho_0}\end{pmatrix}, ..., \begin{pmatrix}\rho_0\tau\\ \rho_0\\ \rho_0 u\\ \widehat{\rho_0}\end{pmatrix}) = a_{1,0}\begin{pmatrix} -\dfrac{\sum_{k=-r}^{r} C_k\rho_0 u}{\sum_{k=-r}^{r} C_k\widehat{\rho_0}} \\ p\left(\dfrac{\sum_{k=-r}^{r} C_k\rho_0\tau}{\sum_{k=-r}^{r} C_k\rho_0}\right)\end{pmatrix},$$

using the fact that $\sum_k C_k = 1$, it leads to

$$\Phi^1(\begin{pmatrix}\rho_0\tau\\ \rho_0\\ \rho_0 u\\ \widehat{\rho_0}\end{pmatrix}, ..., \begin{pmatrix}\rho_0\tau\\ \rho_0\\ \rho_0 u\\ \widehat{\rho_0}\end{pmatrix}) = a_{1,0}\begin{pmatrix} -\frac{\rho_0 u}{\widehat{\rho_0}} \\ p(\frac{\rho_0\tau}{\rho_0})\end{pmatrix} = a_{1,0}\begin{pmatrix} -u \\ p(\tau)\end{pmatrix}.$$

In particular, still for constant states $(\rho_0\tau, \rho_0, \rho_0 u, \widehat{\rho_0})^t$, one obtains that

$$\begin{cases} \overline{\rho_0\tau}_i^{n+\alpha_1} &= \overline{\rho_0\tau}_i^n = \rho_0\tau, \\ \overline{\rho_0 u}_{i+\frac{1}{2}}^{n+\alpha_1} &= \overline{\rho_0 u}_{i+\frac{1}{2}}^n = \rho_0 u. \end{cases}$$

Then by straightforward induction on the intermediate time-steps, any $f_{i+\frac{1}{2}}^{\alpha_m}$ writes as a function Φ^m as

$$f_{i+\frac{1}{2}}^{\alpha_m} = \Phi^m(\begin{pmatrix}\overline{\rho_0\tau}_{i-mr+1}^n\\ \overline{\rho_0}_{i-mr+1}\\ \overline{\rho_0 u}_{i+\frac{1}{2}-mr}^n\\ \overline{\rho_0}_{i+\frac{1}{2}-mr}\end{pmatrix}, ..., \begin{pmatrix}\overline{\rho_0\tau}_{i+mr+1}^n\\ \overline{\rho_0}_{i+mr+1}\\ \overline{\rho_0 u}_{i+\frac{1}{2}+mr}^n\\ \overline{\rho_0}_{i+\frac{1}{2}+mr}\end{pmatrix}).$$

The function Φ^m writes

$$\Phi^m(\begin{pmatrix}\overline{\rho_0\tau}_{i-mr+1}^n\\ \overline{\rho_0}_{i-mr+1}\\ \overline{\rho_0 u}_{i+\frac{1}{2}-mr}^n\\ \overline{\rho_0}_{i+\frac{1}{2}-mr}\end{pmatrix}, ..., \begin{pmatrix}\overline{\rho_0\tau}_{i+mr+1}^n\\ \overline{\rho_0}_{i+mr+1}\\ \overline{\rho_0 u}_{i+\frac{1}{2}+mr}^n\\ \overline{\rho_0}_{i+\frac{1}{2}+mr}\end{pmatrix}) = \sum_{l=0}^{m-1} a_{m,l}\begin{pmatrix} -\dfrac{\sum_{k=-r}^{r} C_k\overline{\rho_0 u}_{i+\frac{1}{2}+k}^{n+\alpha_l}}{\sum_{k=-r}^{r} C_k\overline{\rho_0}_{i+\frac{1}{2}+k}} \\ p\left(\dfrac{\sum_{k=-r}^{r} C_k\overline{\rho_0\tau}_{i+k}^{n+\alpha_l}}{\sum_{k=-r}^{r} C_k\overline{\rho_0}_{i+k}}\right)\end{pmatrix}$$

Then for constant state $(\rho_0\tau, \rho_0, \rho_0 u, \widehat{\rho_0})^t$ and by induction on the previous intermediate time-

steps

$$\Phi^m(\begin{pmatrix}\rho_0\tau\\ \rho_0\\ \rho_0 u\\ \widehat{\rho_0}\end{pmatrix},...,\begin{pmatrix}\rho_0\tau\\ \rho_0\\ \rho_0 u\\ \widehat{\rho_0}\end{pmatrix}) = \sum_{l=0}^{m-1} a_{m,l}\begin{pmatrix}-\frac{\rho_0 u}{\widehat{\rho_0}}\\ p(\frac{\rho_0\tau}{\rho_0})\end{pmatrix} = \sum_{l=0}^{m-1} a_{m,l}\begin{pmatrix}-u\\ p(\tau)\end{pmatrix} = \alpha_m\begin{pmatrix}-u\\ p(\tau)\end{pmatrix}.$$

And in particular, still for constant states, one obtains that

$$\begin{cases}\overline{\rho_0\tau}_i^{n+\alpha_m} &= \overline{\rho_0\tau}_i^n = \rho_0\tau,\\ \overline{\rho_0 u}_{i+\frac{1}{2}}^{n+\alpha_m} &= \overline{\rho_0 u}_{i+\frac{1}{2}}^n = \rho_0 u.\end{cases}$$

Therefore, by induction, the natural flux $f^\star_{i+\frac{1}{2}}$ writes as a vector values function $\Phi^\star$ as

$$f^\star_{i+\frac{1}{2}} = \Phi^\star(\begin{pmatrix}\overline{\rho_0\tau}^n_{i-(s+1)r+1}\\ \overline{\rho_0}_{i-(s+1)r+1}\\ \overline{\rho_0 u}^n_{i+\frac{1}{2}-(s+1)r}\\ \overline{\rho_0}_{i+\frac{1}{2}-(s+1)r}\end{pmatrix},...,\begin{pmatrix}\overline{\rho_0\tau}^n_{i+(s+1)r+1}\\ \overline{\rho_0}_{i+(s+1)r+1}\\ \overline{\rho_0 u}^n_{i+\frac{1}{2}+(s+1)r}\\ \overline{\rho_0}_{i+\frac{1}{2}+(s+1)r}\end{pmatrix}),$$

where $\Phi^\star$ satisfies

$$\Phi^\star(\begin{pmatrix}\overline{\rho_0\tau}^n_{i-(s+1)r+1}\\ \overline{\rho_0}_{i-(s+1)r+1}\\ \overline{\rho_0 u}^n_{i+\frac{1}{2}-(s+1)r}\\ \overline{\rho_0}_{i+\frac{1}{2}-(s+1)r}\end{pmatrix},...,\begin{pmatrix}\overline{\rho_0\tau}^n_{i+(s+1)r+1}\\ \overline{\rho_0}_{i+(s+1)r+1}\\ \overline{\rho_0 u}^n_{i+\frac{1}{2}+(s+1)r}\\ \overline{\rho_0}_{i+\frac{1}{2}+(s+1)r}\end{pmatrix}) = \sum_{l=0}^{s}\theta_l\begin{pmatrix}-\dfrac{\sum_{k=-r}^{r} C_k\overline{\rho_0 u}^{n+\alpha_l}_{i+\frac{1}{2}+k}}{\sum_{k=-r}^{r} C_k\overline{\rho_0}_{i+\frac{1}{2}+k}}\\ p\left(\dfrac{\sum_{k=-r}^{r} C_k\overline{\rho_0\tau}^{n+\alpha_l}_{i+k}}{\sum_{k=-r}^{r} C_k\overline{\rho_0}_{i+k}}\right)\end{pmatrix}.$$

Thus for constant states, it leads to

$$\Phi^\star(\begin{pmatrix}\rho_0\tau\\ \rho_0\\ \rho_0 u\\ \widehat{\rho_0}\end{pmatrix},...,\begin{pmatrix}\rho_0\tau\\ \rho_0\\ \rho_0 u\\ \widehat{\rho_0}\end{pmatrix}) = \sum_{l=0}^{s}\theta_l\begin{pmatrix}-\frac{\rho_0 u}{\widehat{\rho_0}}\\ p(\frac{\rho_0\tau}{\rho_0})\end{pmatrix} = \sum_{l=0}^{s}\theta_l\begin{pmatrix}-u\\ p(\tau)\end{pmatrix}.$$

Using the fact that $\sum_{l=0}^{s} \theta_l = 1$, it leads to

$$\Phi^\star(\begin{pmatrix} \rho_0\tau \\ \rho_0 \\ \rho_0 u \\ \widehat{\rho_0} \end{pmatrix}, ..., \begin{pmatrix} \rho_0\tau \\ \rho_0 \\ \rho_0 u \\ \widehat{\rho_0} \end{pmatrix}) = \begin{pmatrix} -u \\ p(\tau) \end{pmatrix}$$

Hence, the scheme is weakly consistent for the barotropic equations in the sense of definition I.6.

■

Another important property of a scheme is its linear stability. To study such a property, one considers the scheme for the linearized system of equation, which is nothing but the wave equation

$$\begin{cases} \partial_t u + \partial_x p &= 0 \\ \partial_t p + \partial_x u &= 0 \end{cases} \tag{II.19}$$

For such a linear system, the staggered scheme writes

$$\begin{cases} \overline{p}_i^{n+\alpha_m} &= \overline{p}_i^n - \frac{\Delta t}{\Delta X} \sum_{l=0}^{m-1} a_{m,l} du_i^{n+\alpha_l}, \\ \overline{u}_{i+\frac{1}{2}}^{n+\alpha_m} &= \overline{u}_{i+\frac{1}{2}}^n - \frac{\Delta t}{\Delta X} \sum_{l=0}^{m-1} a_{m,l} dp_{i+\frac{1}{2}}^{n+\alpha_l}, \end{cases} \qquad \begin{cases} \overline{p}_i^{n+1} &= \overline{p}_i^n - \frac{\Delta t}{\Delta X} \sum_{l=0}^{s-1} \theta_l du_i^{n+\alpha_l}, \\ \overline{u}_{i+\frac{1}{2}}^{n+1} &= \overline{u}_{i+\frac{1}{2}}^n - \frac{\Delta t}{\Delta X} \sum_{l=0}^{s-1} \theta_l dp_{i+\frac{1}{2}}^{n+\alpha_l}, \end{cases} \tag{II.20}$$

with a CFL condition of the form

$$\Delta t < \lambda \Delta X.$$

Using the amplification factor presented in section I-1.2.3, one deduces a CFL condition which yields linear stability for the schemes. Lemmas II.4 and II.5 give results concerning the linear stability of the staggered schemes.

Lemma II.4 (Linear instability of the second order staggered schemes)**.** *The two-steps second-order in time and space explicit Runge–Kutta schemes* (II.20) *are linearly unstable for any CFL condition.*

Remark II.3. For this reason, second order Runge–Kutta schemes are discarded. Instead the BBC scheme is used for second order accuracy in time.

Proof. A two-steps second-order explicit Runge–Kutta sequences can be parametrized using a non-zero α which leads to the following Butcher table

α	α	0
1	$1 - \frac{1}{2\alpha}$	$\frac{1}{2\alpha}$

For second-order accuracy, average and pointwise values are equivalent. The index j is used in order not to introduce confusion with the complex number i. Denoting $\nu = \frac{\Delta t}{\Delta X}$, it writes for u

$$u_{j+\frac{1}{2}}^{n+1} = u_{j+\frac{1}{2}}^{n} - \nu \left((1 - \frac{1}{2\alpha})(p_{j+1}^{n} - p_j^{n}) + \frac{1}{2\alpha}(p_{j+1}^{n+\alpha} - p_j^{n+\alpha}) \right).$$

Then plugging the terms for $p_j^{n+\alpha}$ and $p_{j+1}^{n+\alpha}$ in the previous equation, it yields

$$u_{j+\frac{1}{2}}^{n+1} = u_{j+\frac{1}{2}}^{n} - \nu \left((1 - \frac{1}{2\alpha})(p_{j+1}^{n} - p_j^{n}) + \frac{1}{2\alpha}(p_{j+1}^{n} - p_j^{n} - \alpha\nu(u_{j+\frac{3}{2}}^{n} - 2u_{j+\frac{1}{2}}^{n} + u_{j-\frac{1}{2}}^{n})) \right),$$

which can be simplified into

$$u_{j+\frac{1}{2}}^{n+1} = u_{j+\frac{1}{2}}^{n} - \nu \left((p_{j+1}^{n} - p_j^{n}) - \frac{1}{2}(\nu(u_{j+\frac{3}{2}}^{n} - 2u_{j+\frac{1}{2}}^{n} + u_{j-\frac{1}{2}}^{n})) \right).$$

The above expression is completely independent of α and thus the resulting CFL condition is as well independent of α. It writes

$$u_{j+\frac{1}{2}}^{n+1} = (1 - \nu^2)u_{j+\frac{1}{2}}^{n} - \nu \left(p_{j+1}^{n} - p_j^{n} \right) + \frac{\nu^2}{2} \left(u_{j+\frac{3}{2}}^{n} + u_{j-\frac{1}{2}}^{n} \right).$$

Denoting ϵ_j and $\epsilon_{j+\frac{1}{2}}$ the numerical errors as introduced in section I-1.2.3, it yields

$$\epsilon_{j+\frac{1}{2}}^{n+1} = (1 - \nu^2)\epsilon_{j+\frac{1}{2}}^{n} - \nu \left(\epsilon_{j+1}^{n} - \epsilon_j^{n} \right) + \frac{\nu^2}{2} \left(\epsilon_{j+\frac{3}{2}}^{n} + \epsilon_{j-\frac{1}{2}}^{n} \right).$$

Now assuming that for any n, and for any j, $\epsilon_j^n = \mathrm{e}^{\beta n \Delta t}\, \mathrm{e}^{ik\pi j \Delta X}$ with k an integer, one gets

$$\mathrm{e}^{\beta \Delta t} = (1 - \nu^2) - \nu \left(\mathrm{e}^{\frac{ik\pi\Delta X}{2}} - \mathrm{e}^{-\frac{ik\pi\Delta X}{2}} \right) + \frac{\nu^2}{2} \left(\mathrm{e}^{ik\pi\Delta X} + \mathrm{e}^{-ik\pi\Delta X} \right).$$

Using trigonometric identities, it yields

$$\mathrm{e}^{\beta \Delta t} = (1 - \nu^2) - 2i\nu \sin(\frac{k\pi\Delta X}{2}) + \nu^2 \cos(k\pi\Delta X).$$

Introducing $\theta = k\pi\Delta X$ and $g(\theta, \nu) = \mathrm{e}^{\beta\Delta t}$, one gets the following equation for the amplification factor

$$g(\theta, \nu) = (1 - \nu^2) - 2i\nu \sin(\frac{\theta}{2}) + \nu^2 \cos(\theta).$$

Then the square of the modulus of $g(\theta, \nu)$ writes

$$|g(\theta, \nu)|^2 = \left(1 - \nu^2 + \nu^2 \cos(\theta)\right)^2 + 4\nu^2 \sin^2(\frac{\theta}{2}).$$

Using the fact that $\cos(\theta) = 1 - 2\sin^2(\frac{\theta}{2})$ and after simplification one gets

$$|g(\theta, \nu)|^2 = 1 + 4\nu^4 \sin^4(\frac{\theta}{2}).$$

Then, the amplification factor satisfies for $\nu \neq 0$

$$\max_{\theta \in [0:2\pi]} |g(\theta, \nu)|^2 > 1.$$

And thus the scheme is not stable in the sense of definition I.11. ■

Similar calculations have been performed for higher-order staggered schemes. The Runge–Kutta sequences used in the following are described in section I-1.2.3. The third order Runge–Kutta sequence selected is the SSPRK3 [70, 71]. The fourth order Runge–Kutta sequence is the $\frac{3}{8}$-Kutta sequence [99]. The fifth order Runge–Kutta sequence is the Dormand–Prince sequence [49]. Last, the sixth, seventh and eighth order Runge–Kutta sequence are the robust Verner sequences available in [167]. Due to the complexity of the amplification factor, results of stability are numerical. One checks for a given value of ν that for all θ, $|g(\theta, \nu)| \leq 1$. The results are summarized in the following lemma.

Lemma II.5 (Linear stability of the staggered schemes)**.** *Higher-order schemes are stable under CFL condition*

$$\Delta t < \lambda_{\text{STAG}} \frac{\Delta X}{\max_i c_i}$$

where c_i is the speed of sound in the cell i. The coefficients λ_{STAG} are listed in table II.6 with the aforementioned sequences.

Schemes	λ_{STAG}
2nd order BBC	0.6888
3rd order SSPRK3	0.7423
4th order $\frac{3}{8}$-Kutta	1.1390
5th order Dormand-Prince	0.4015
6th order robust Verner	1.0045
7th order robust Verner	0.0134
8th order robust Verner	0.9840

Table II.6 – CFL conditions for linear stability of the staggered schemes

II-2.2 A new local internal energy corrector

Compared to the barotropic schemes, an additional theoretical difficulty shows up for the hydrodynamics case (II.13)-(II.14) with the energy equation. It is related to the fact that, even if the total energy $\mathcal{E}$ is preserved by construction, it is not the case for the total energy E. Experimentally, we also observe that the schemes (II.13)-(II.14) are unable to capture the shocks correctly, in the sense that the Rankine–Hugoniot jump relations are not recovered.

The idea is to recouple E and $\mathcal{E}$ using a correction of the internal energy at the end of the Lagrangian phase (II.13)-(II.14). The difference between the computed kinetic energy and the

kinetic energy reconstructed from the velocity is reversed in the internal energy. This is very similar to what is done in works by Herbin, Latché and al. [83, 80, 81]. The main difference is that they perform the correction *a priori*, whereas here in our case the correction is applied *a posteriori*.

II-2.2.1 Internal energy corrector

As an additional comment, the internal energy evolution equation in (II.9) is undefined classically in the sense of distributions. So, in the absence of any artificial viscosity, one expects wrong discontinuities computations. The idea of the internal energy corrector is to really solve the Lagrangian system formulated in total energy rather that in internal one.

The difference between the computed kinetic energy and the kinetic energy reconstructed from the velocity is computed. As the scheme is high-order accurate, the result is not so straightforward. It follows the steps described hereafter. First the point-wise kinetic energy reconstructed from the velocity is

$$(\frac{1}{2}\rho_0 u^2)^{n+1}_{i+\frac{1}{2}} = \frac{1}{2}\frac{\left(\sum_k C_k \overline{\rho_0 u}^{n+1}_{i+k+\frac{1}{2}}\right)^2}{\sum_k C_k \overline{\rho_0}^{n}_{i+k+\frac{1}{2}}}.$$

Second it is averaged over a cell using the coefficients $\widehat{C}_k$ presented in table II.1

$$(\frac{1}{2}\overline{\rho_0 u^2})^{n+1}_{i+\frac{1}{2}} = \sum_k \widehat{C}_k (\frac{1}{2}\rho_0 u^2)^{n+1}_{i+k+\frac{1}{2}}.$$

The difference denoted $\Delta K^{n+1}_{i+\frac{1}{2}}$ between the two kinetic energies is

$$\Delta K^{n+1}_{i+\frac{1}{2}} = \overline{\rho_0 e_{\text{kin}}}^{n+1}_{i+\frac{1}{2}} - (\frac{1}{2}\overline{\rho_0 u^2})^{n+1}_{i+\frac{1}{2}}.$$

Third, linear interpolation is made to compute ΔK^{n+1}_i

$$\Delta K^{n+1}_i = \frac{1}{2}(\Delta K^{n+1}_{i+\frac{1}{2}} + \Delta K^{n+1}_{i-\frac{1}{2}}).$$

Last, the difference ΔK^{n+1}_i is added to the internal energy $\overline{\rho_0 \epsilon}^{n+1}_i$ whereas $\Delta K^{n+1}_{i+\frac{1}{2}}$ is subtracted to the kinetic ones. It writes as an *a posteriori* correction

$$\left\{ \begin{array}{lcl} \overline{\rho_0 \epsilon}^{n+1,\star}_i & = & \overline{\rho_0 \epsilon}^{n+1}_i + \Delta K^{n+1}_i \\ \overline{\rho_0 e_{\text{kin}}}^{n+1,\star}_{i+\frac{1}{2}} & = & \overline{\rho_0 e_{\text{kin}}}^{n+1}_{i+\frac{1}{2}} - \Delta K^{n+1}_{i+\frac{1}{2}} = (\frac{1}{2}\overline{\rho_0 u^2})^{n+1}_{i+\frac{1}{2}} \end{array} \right. \tag{II.21}$$

The internal energy corrector can be applied at the end of each Runge–Kutta sub-cycle or only at the end of the time-step. Commonly, the internal energy corrector is performed only at the end of the time-step, hence the *a posteriori* correction.

II-2.2.2 Properties of the internal energy corrector

Lemma II.6 (High-order accuracy of the internal energy corrector). *The internal energy corrector is high-order accurate in both time and space.*

Proof. Assume that the solution is smooth enough. Assume that the coefficients $\widehat{C}_k$ and C_k yield N^{th} order of accuracy in space, and that the Lagrange phase is also of order N in both time and space. Then in particular, one has

$$\Delta K_{i+\frac{1}{2}}^{n+1} = \overline{\rho_0 e_{\mathrm{kin}}}_{i+\frac{1}{2}}^{n+1} - (\frac{1}{2}\overline{\rho_0 u^2})_{i+\frac{1}{2}}^{n+1} = \mathscr{O}(\Delta X^N).$$

And then trivially, one gets that

$$\Delta K_i^{n+1} = \frac{1}{2}(\Delta K_{i+\frac{1}{2}}^{n+1} + \Delta K_{i-\frac{1}{2}}^{n+1}) = \mathscr{O}(\Delta X^N).$$

As the Lagrange phase is assumed to be high-order accurate, one has that

$$\overline{\rho_0 \epsilon}_i^{n+1} = \overline{\rho_0 \epsilon}(x_i, t^{n+1}) + \mathscr{O}(\Delta X^N),$$

And then, one gets

$$\overline{\rho_0 \epsilon}_i^{n+1,\star} = \overline{\rho_0 \epsilon}_i^{n+1} + \Delta K_i = \overline{\rho_0 \epsilon}(x_i, t^{n+1}) + \mathscr{O}(\Delta X^N),$$

which concludes the proof, yielding high-order accuracy for the internal energy. ■

Moreover the following lemma gives conservation of total energy when applying the internal energy corrector.

Lemma II.7 (Conservation of the internal energy corrector). *The internal energy corrector satisfies $\mathcal{E}^{n+1,\star} = \mathcal{E}^{n+1}$.*

Proof. The internal energy corrector is conservative in $\mathcal{E}$ if and only if we have

$$\mathcal{E}^{n+1,\star} - \mathcal{E}^{n+1} = \Delta X \sum_i \Delta K_i - \sum \Delta K_{i+\frac{1}{2}} = 0.$$

As $\Delta K_i = \frac{1}{2}(\Delta K_{i+\frac{1}{2}} + \Delta K_{i-\frac{1}{2}})$, it leads to the following computations

$$\mathcal{E}^{n+1,\star} - \mathcal{E}^{n+1} = \Delta X \left(\sum_i \Delta K_i - \sum \Delta K_{i+\frac{1}{2}} \right) = \Delta X \left(\sum_i \frac{1}{2}(\Delta K_{i+\frac{1}{2}} + \Delta K_{i-\frac{1}{2}}) - \sum \Delta K_{i+\frac{1}{2}} \right).$$

Performing change of discrete variables in the first summation, and assuming wall or periodic boundary conditions, it yields $\sum_i \Delta K_i - \sum \Delta K_{i+\frac{1}{2}} = 0$. Thus, the internal energy corrector conserve the quantity $\mathcal{E}$, meaning that $\mathcal{E}^{n+1,\star} - \mathcal{E}^{n+1} = 0$. ■

Lemma II.8 (Conservation of the staggered schemes (II.13)-(II.14)-(II.21))**.** *The schemes* (II.13)-(II.14) *with the internal energy corrector* (II.21) *satisfy* $E^{n+1,\star} = E^n$ *(cf definition II.1).*

Proof. We have

$$E^{n+1,\star} - E^n = \Delta X \sum_i \left(\overline{\rho_0 \epsilon}_i^{n+1,\star} - \overline{\rho_0 \epsilon}_i^n \right) + \Delta X \sum_i \left(\overline{\rho_0 u^2}_{i+\frac{1}{2}}^{n+1,\star} - \overline{\rho_0 u^2}_{i+\frac{1}{2}}^n \right) \tag{II.22}$$

Introducing the term at time $t = t^{n+1}$, it becomes

$$\begin{aligned}
&= \Delta X \sum_i \left(\overline{\rho_0 \epsilon}_i^{n+1,\star} - \overline{\rho_0 \epsilon}_i^{n+1} + \overline{\rho_0 \epsilon}_i^{n+1} - \overline{\rho_0 \epsilon}_i^n \right) \\
&+ \Delta X \sum_i \left(\overline{\rho_0 u^2}_{i+\frac{1}{2}}^{n+1,\star} - \overline{\rho_0 e_{\text{kin},u}}_{i+\frac{1}{2}}^{n+1} + \overline{\rho_0 e_{\text{kin},u}}_{i+\frac{1}{2}}^{n+1} - \overline{\rho_0 u^2}_{i+\frac{1}{2}}^n \right) \\
&= \Delta X \sum_i \left(\overline{\rho_0 \epsilon}_i^{n+1,\star} - \overline{\rho_0 \epsilon}_i^{n+1} \right) - \Delta X \sum_i \left(\overline{\rho_0 u^2}_{i+\frac{1}{2}}^{n+1,\star} - \overline{\rho_0 e_{\text{kin},u}}_{i+\frac{1}{2}}^{n+1} \right) + \mathcal{E}^{n+1} - \mathcal{E}^n
\end{aligned}$$

Using the fact that $\overline{\rho_0 u^2}_{i+\frac{1}{2}}^{n+1,\star} = \overline{\rho_0 e_{\text{kin},u}}_{i+\frac{1}{2}}^{n+1,\star}$, it leads to

$$\begin{aligned}
&= \mathcal{E}^{n+1,\star} - \mathcal{E}^{n+1} + \mathcal{E}^{n+1} - \mathcal{E}^n \\
&= 0.
\end{aligned}$$

Thus, applying the internal corrector gives conservation of the energy E between time $t = t^{n+1,\star}$ and time $t = t^n$. ■

Theorem II.9 (Weak consistency of the staggered schemes (II.13)-(II.14)-(II.21))**.** *For all explicit Runge–Kutta sequences, for coefficients* $C_k, \widehat{C}_k, d_k, r_k$ *defined in tables II.1 to II.4, the schemes* (II.13)-(II.14)-(II.21) *are weakly consistent with the Euler equations in Lagrangian coordinates.*

Remark II.4. The proof of the weak consistency of the two first variables, specific volume and momentum, which show up in (II.13)-(II.14) is essentially similar to the one of lemma II.3 for the barotropic case, so is not detailed. Instead we focus on the energy equation. However, due to the very intricate structure of the discrete energy equation, no explicit natural fluxes for total energy have been exhibited so far. It means that the energy equation is not rewritten using the form (I.34)-(I.35)-(I.36). That is the criterion of flux consistency of definition I.6 is unfortunately not applicable, this is why the proof is detailed hereafter in full length, starting directly from definition I.7.

Proof. The assumptions presented in definition I.7 for weak consistency are done. We first detail the proof for a forward Euler, second order in space scheme because it highlights the key elements of the method. The general case with a forward Euler and any order in space will be dealt with in a second stage. The most general case with any explicit Runge–Kutta sequences will not be detailed because it would add no new technical ideas and the notations are too heavy. For the

sake of simplicity, in the following the time step $t^{n+1,\star}$ is denoted by t^{n+1}.

First stage

For a forward Euler, second order in space scheme, the internal and kinetic energies discrete equations write

$$\left\{\begin{array}{lcll} \rho_0\epsilon_i^{n+1} - \rho_0\epsilon_i^n & = & -\dfrac{\Delta t}{\Delta X} p_i^n (u_{i+\frac{1}{2}}^n - u_{i-\frac{1}{2}}^n) & + \quad \Delta K_i^{n+1}, \\ \rho_0 e_{\mathrm{kin}}{}_{i+\frac{1}{2}}^{n+1} - \rho_0 e_{\mathrm{kin}}{}_{i+\frac{1}{2}}^{n} & = & -\dfrac{\Delta t}{\Delta X} u_{i+\frac{1}{2}}^n (p_{i+1}^n - p_i^n) & - \quad \Delta K_{i+\frac{1}{2}}^{n+1}, \end{array}\right.$$

The idea is to take a test function $\phi \in \mathscr{C}_0^\infty$ with compact support. Denote $\phi_i^n = \phi(i\Delta X, t^n)$ and $\phi_{i+\frac{1}{2}}^n((i+\frac{1}{2})\Delta X, t^n)$. Multiply the first equation by $\Delta X \phi_i^{n+1}$ and the second by $\Delta X \phi_{i+\frac{1}{2}}^{n+1}$ then to sum over the n and i and to combine both. It leads to

$$\begin{array}{rl} & \displaystyle\sum_n \sum_i \Delta X \left[(\rho_0\epsilon_i^{n+1} - \rho_0\epsilon_i^n)\phi_i^{n+1} + (\rho_0 e_{\mathrm{kin}}{}_{i+\frac{1}{2}}^{n+1} - \rho_0 e_{\mathrm{kin}}{}_{i+\frac{1}{2}}^{n})\phi_{i+\frac{1}{2}}^{n+1} \right] \\ + & \displaystyle\sum_n \sum_i \Delta t \left[p_i^n \phi_i^{n+1} (u_{i+\frac{1}{2}}^n - u_{i-\frac{1}{2}}^n) + u_{i+\frac{1}{2}}^n \phi_{i+\frac{1}{2}}^{n+1} (p_{i+1}^n - p_i^n) \right] \\ - & \displaystyle\sum_n \sum_i \Delta X \left[\Delta K_i^{n+1} \phi_i^{n+1} - \Delta K_{i+\frac{1}{2}}^{n+1} \phi_{i+\frac{1}{2}}^{n+1} \right] = 0. \end{array} \tag{II.23}$$

Denote h a parameter proportional to ΔX and Δt. Introducing the notation

$$\left\{\begin{array}{lcl} \mathcal{T}_1^h & = & \displaystyle\sum_n \sum_i \Delta X \left[(\rho_0\epsilon_i^{n+1} - \rho_0\epsilon_i^n)\phi_i^{n+1} + (\rho_0 e_{\mathrm{kin}}{}_{i+\frac{1}{2}}^{n+1} - \rho_0 e_{\mathrm{kin}}{}_{i+\frac{1}{2}}^{n})\phi_{i+\frac{1}{2}}^{n+1} \right], \\ \mathcal{T}_2^h & = & \displaystyle\sum_n \sum_i \Delta t \left[p_i^n \phi_i^{n+1} (u_{i+\frac{1}{2}}^n - u_{i-\frac{1}{2}}^n) + u_{i+\frac{1}{2}}^n \phi_{i+\frac{1}{2}}^{n+1} (p_{i+1}^n - p_i^n) \right], \\ \mathcal{T}_3^h & = & \displaystyle -\sum_n \sum_i \Delta X \left[\Delta K_i^{n+1} \phi_i^{n+1} - \Delta K_{i+\frac{1}{2}}^{n+1} \phi_{i+\frac{1}{2}}^{n+1} \right], \end{array}\right. \tag{II.24}$$

eq. (II.23) rewrites simply under the form $\mathcal{T}_1^h + \mathcal{T}_2^h + \mathcal{T}_3^h = 0$. Terms $\mathcal{T}_1^h$ are reordered into

$$\mathcal{T}_1^h = -\sum_n \Delta t \sum_i \Delta X \left[\rho_0 \epsilon_i^n \frac{\phi_i^{n+1} - \phi_i^n}{\Delta t} + \rho_0 e_{\mathrm{kin}}{}_{i+\frac{1}{2}}^{n} \frac{\phi_{i+\frac{1}{2}}^{n+1} - \phi_{i+\frac{1}{2}}^{n}}{\Delta t} \right].$$

We will use the natural definition/notation for staircase functions

$$\begin{aligned} \psi_h(x,t) &= \sum_i \sum_n \chi_{]t^n, t^{n+1}[}(t) \chi_{]x_{i-1/2}, x_{i+1/2}[}(x) \psi_i^n, \\ \psi_{h,\mathrm{stag}}(x,t) &= \sum_i \sum_n \chi_{]t^n, t^{n+1}[}(t) \chi_{]x_i, x_{i+1}[}(x) \psi_{i+\frac{1}{2}}^n. \end{aligned}$$

Then, using the internal corrector, it yields that $\rho_0 e_{\text{kin}}{}^n_{i+\frac{1}{2}} = \frac{1}{2}(\rho_0 u^2)^n_{i+\frac{1}{2}}$ and so

$$\begin{aligned}\mathcal{T}_1^h = &-\int_0^T\int_\Omega (\rho_0\epsilon)_h \partial_t\phi_h \mathrm{d}x\mathrm{d}t - \int_0^T\int_\Omega (\frac{1}{2}\rho_0 u^2)_{h,\text{stag}}\partial_t\phi_{h,\text{stag}}\mathrm{d}x\mathrm{d}t\\ &+\int_\Omega (\rho_0\epsilon)^0_h\phi^0_h\mathrm{d}x + \int_\Omega (\frac{1}{2}\rho_0 u^2)^0_{h,\text{stag}}\phi^0_{h,\text{stag}}\mathrm{d}x.\end{aligned}$$

Using the convergence hypothesis of definition I.7 and the regularity of the test function ϕ, one can pass to the limit as ΔX and Δt tend to 0. It leads to

$$\begin{aligned}\lim_{h\to 0}\mathcal{T}_1^h = &-\int_0^T\int_\Omega \widehat{\rho_0\epsilon}\partial_t\phi\mathrm{d}x\mathrm{d}t - \int_0^T\int_\Omega \widehat{\frac{1}{2}\rho_0 u^2}\partial_t\phi\mathrm{d}x\mathrm{d}t\\ &+\int_\Omega \widehat{\rho_0\epsilon}(x,0)\phi(x,0)\mathrm{d}x + \int_\Omega \widehat{\frac{1}{2}\rho_0 u^2}(x,0)\phi(x,0)\mathrm{d}x.\end{aligned}$$

Using the definition of the total energy as $\rho_0 e = \rho_0\epsilon + \frac{1}{2}\rho_0 u^2$, one gets

$$\lim_{h\to 0}\mathcal{T}_1^h = -\int_0^T\int_\Omega \widehat{\rho_0 e}\partial_t\phi\mathrm{d}x\mathrm{d}t + \int_\Omega \widehat{\rho_0 e}(x,0)\phi(x,0)\mathrm{d}x.$$

Now, focus on $\mathcal{T}_3^h$ which writes

$$\mathcal{T}_3^h = -\sum_n\sum_i \Delta X\left[\Delta K_i^{n+1}\phi_i^{n+1} - \Delta K_{i+\frac{1}{2}}^{n+1}\phi_{i+\frac{1}{2}}^{n+1}\right],$$

which leads using $\Delta K_i^{n+1} = \frac{1}{2}(\Delta K_{i+\frac{1}{2}}^{n+1} + \Delta K_{i-\frac{1}{2}}^{n+1})$ to

$$\mathcal{T}_3^h = -\sum_n\sum_i \Delta X\left[\frac{1}{2}(\Delta K_{i+\frac{1}{2}}^{n+1} + \Delta K_{i-\frac{1}{2}}^{n+1})\phi_i^{n+1} - \Delta K_{i+\frac{1}{2}}^{n+1}\phi_{i+\frac{1}{2}}^{n+1}\right],$$

which gives after reordering the terms

$$\mathcal{T}_3^h = -\sum_n\sum_i \Delta X\Delta K_{i+\frac{1}{2}}^{n+1}\left(\frac{\phi_{i+1}^{n+1}+\phi_i^{n+1}}{2} - \phi_{i+\frac{1}{2}}^{n+1}\right).$$

Using the boundedness in L^∞ of $\Delta K_{i+\frac{1}{2}}^{n+1}$ and regularity of ϕ, it leads to

$$|\mathcal{T}_3^h| \le C_\phi \Delta X\|(\Delta K)_h\|_{L^\infty}, \text{ which gives immediately } \lim_{h\to 0}|\mathcal{T}_3^h| = 0.$$

The term $\mathcal{T}_2^h$ writes

$$\mathcal{T}_2^h = \sum_n\sum_i \Delta t\left[p_i^n\phi_i^{n+1}(u_{i+\frac{1}{2}}^n - u_{i-\frac{1}{2}}^n) + u_{i+\frac{1}{2}}^n\phi_{i+\frac{1}{2}}^{n+1}(p_{i+1}^n - p_i^n)\right],$$

which, once the terms reordered, writes as

$$
\begin{aligned}
\mathcal{T}_2^h &= \sum_n \Delta t \sum_i \left[p_i^n u_{i+\frac{1}{2}}^n \phi_i^{n+1} - p_{i+1}^n u_{i+\frac{1}{2}}^n \phi_{i+1}^{n+1} + u_{i+\frac{1}{2}}^n \phi_{i+\frac{1}{2}}^{n+1} (p_{i+1}^n - p_i^n) \right] \\
&= \sum_n \Delta t \sum_i \left[u_{i+\frac{1}{2}}^n p_i^n (\phi_i^{n+1} - \phi_{i+\frac{1}{2}}^{n+1}) + u_{i+\frac{1}{2}}^n p_{i+1}^n (\phi_{i+\frac{1}{2}}^{n+1} - \phi_{i+1}^{n+1}) \right] \\
&= -\sum_n \Delta t \Delta X \sum_i u_{i+\frac{1}{2}}^n \left[\frac{1}{2} p_i^n \frac{\phi_{i+\frac{1}{2}}^{n+1} - \phi_i^{n+1}}{\frac{\Delta X}{2}} + \frac{1}{2} p_{i+1}^n \frac{\phi_{i+1}^{n+1} - \phi_{i+\frac{1}{2}}^{n+1}}{\frac{\Delta X}{2}} \right] \\
&= -\sum_n \Delta t \Delta X \sum_i u_{i+\frac{1}{2}}^n \left[\frac{p_i^n + p_{i+1}^n}{4} \frac{\phi_{i+\frac{1}{2}}^{n+1} - \phi_i^{n+1}}{\frac{\Delta X}{2}} + \frac{p_i^n + p_{i+1}^n}{4} \frac{\phi_{i+1}^{n+1} - \phi_{i+\frac{1}{2}}^{n+1}}{\frac{\Delta X}{2}} \right] \\
&- \sum_n \Delta t \Delta X \sum_i u_{i+\frac{1}{2}}^n \left[\frac{p_i^n - p_{i+1}^n}{4} \frac{\phi_{i+\frac{1}{2}}^{n+1} - \phi_i^{n+1}}{\frac{\Delta X}{2}} + \frac{p_{i+1}^n - p_i^n}{4} \frac{\phi_{i+1}^{n+1} - \phi_{i+\frac{1}{2}}^{n+1}}{\frac{\Delta X}{2}} \right].
\end{aligned}
$$

The previous expression is decomposed into two terms denoted $\mathcal{T}_{2,1}^h$ and $\mathcal{T}_{2,2}^h$ with

$$
\left\{
\begin{aligned}
\mathcal{T}_{2,1}^h &= -\sum_n \Delta t \Delta X \sum_i u_{i+\frac{1}{2}}^n \left[\frac{p_i^n + p_{i+1}^n}{4} \frac{\phi_{i+\frac{1}{2}}^{n+1} - \phi_i^{n+1}}{\frac{\Delta X}{2}} + \frac{p_i^n + p_{i+1}^n}{4} \frac{\phi_{i+1}^{n+1} - \phi_{i+\frac{1}{2}}^{n+1}}{\frac{\Delta X}{2}} \right], \\
\mathcal{T}_{2,2}^h &= -\sum_n \Delta t \Delta X \sum_i u_{i+\frac{1}{2}}^n \left[\frac{p_i^n - p_{i+1}^n}{4} \frac{\phi_{i+\frac{1}{2}}^{n+1} - \phi_i^{n+1}}{\frac{\Delta X}{2}} + \frac{p_{i+1}^n - p_i^n}{4} \frac{\phi_{i+1}^{n+1} - \phi_{i+\frac{1}{2}}^{n+1}}{\frac{\Delta X}{2}} \right].
\end{aligned}
\right.
$$

The $\mathcal{T}_{2,2}^h$ is dealt with assuming that p_h is in BV, u_h is bounded in L^∞, and ϕ is in $\mathscr{C}^1$ which gives

$$
|\mathcal{T}_{2,2}^h| \leq \Delta X C_\phi \|u_h\|_{L^\infty} \|p_h\|_{BV}.
$$

Hence, passing to the limit, it leads to

$$
\lim_{h \to 0} |\mathcal{T}_{2,2}^h| = 0. \tag{II.25}
$$

On the other hand, one easily notices that $\mathcal{T}_{2,1}^h$ rewrites as

$$
\begin{aligned}
\mathcal{T}_{2,1}^h &= -\sum_n \Delta t \Delta X \sum_i u_{i+\frac{1}{2}}^n \frac{p_i^n + p_{i+1}^n}{2} \left[\frac{1}{2} \frac{\phi_{i+\frac{1}{2}}^{n+1} - \phi_i^{n+1}}{\frac{\Delta X}{2}} + \frac{1}{2} \frac{\phi_{i+1}^{n+1} - \phi_{i+\frac{1}{2}}^{n+1}}{\frac{\Delta X}{2}} \right], \\
&= -\int_0^T \int_\Omega (pu)_h \partial_X \phi_h \mathrm{d}x \mathrm{d}t
\end{aligned}
$$

Using the regularity of ϕ it leads, passing to the limit, to

$$\lim_{h\to 0} \mathcal{T}_{2,1}^h = -\int_0^T \int_\Omega \widehat{pu}\partial_X \phi \mathrm{d}x\mathrm{d}t.$$

Reassembling all the terms, it yields that

$$\lim_{h\to 0} \mathcal{T}_1^h + \mathcal{T}_2^h + \mathcal{T}_3^h = -\int_0^T \int_\Omega \widehat{\rho_0 e}\partial_t \phi \mathrm{d}x\mathrm{d}t - \int_0^T \int_\Omega \widehat{pu}\partial_X \phi \mathrm{d}x\mathrm{d}t + \int_\Omega \widehat{\rho_0 e}(x,0)\phi(x,0)\mathrm{d}x.$$

And, hence, it leads to

$$\int_0^T \int_\Omega \widehat{\rho_0 e}\partial_t \phi \mathrm{d}x\mathrm{d}t + \int_0^T \int_\Omega \widehat{pu}\partial_X \phi \mathrm{d}x\mathrm{d}t = \int_\Omega \widehat{\rho_0 e}(x,0)\phi(x,0)\mathrm{d}x.$$

Previous equation gives weak consistency for the second order in space, forward Euler staggered scheme with internal energy corrector.

Second stage

Now, the problem of high-order in space is tackled. It does not yield to any difficulty for the terms $\mathcal{T}_1^h$ and $\mathcal{T}_3^h$, but this is not the case for $\mathcal{T}_2^h$, where the desired results is not obvious. For the sake of simplicity here, we consider that $\widehat{C}_0 = 1,\ \widehat{C}_k = 0, \forall |k| > 0$. The results does not change, if $\sum_k \widehat{C}_k = 1$ but it greatly alleviates the algebra of the proof.

One has that

$$\mathcal{T}_2^h = -\sum_n \Delta t \sum_i \sum_{k\geq 0} d_k \left[p_i^n \phi_i^{n+1}(u_{i+k+\frac{1}{2}}^n - u_{i-k-\frac{1}{2}}^n) + u_{i+\frac{1}{2}}^n \phi_{i+\frac{1}{2}}^{n+1}(p_{i+k+1}^n - p_{i-k}^n) \right].$$

Reordering the terms, so that only $u_{i+\frac{1}{2}}^n$ shows up, leads to

$$\mathcal{T}_2^h = -\sum_n \Delta t \sum_i u_{i+\frac{1}{2}}^n \sum_{k\geq 0} d_k \left[p_{i-k}^n(\phi_{i-k}^{n+1} - \phi_{i+\frac{1}{2}}^{n+1}) + p_{i+k+1}^n(\phi_{i+\frac{1}{2}}^{n+1} - \phi_{i+k+1}^{n+1}) \right].$$

Highlighting the space derivatives of ϕ gives

$$\mathcal{T}_2^h = +\sum_n \Delta t \sum_i \Delta X u_{i+\frac{1}{2}}^n \sum_{k\geq 0} (k+\frac{1}{2}) d_k \left[p_{i-k}^n \frac{\phi_{i+\frac{1}{2}}^{n+1} - \phi_{i-k}^{n+1}}{\Delta X(k+\frac{1}{2})} + p_{i+k+1}^n \frac{\phi_{i+k+1}^{n+1} - \phi_{i+\frac{1}{2}}^{n+1}}{\Delta X(k+\frac{1}{2})} \right].$$

Noticing that $(k+\frac{1}{2})d_k = r_k,\ k \geq 0$, it yields

$$\mathcal{T}_2^h = +\sum_n \Delta t \sum_i \Delta X u_{i+\frac{1}{2}}^n \sum_{k\geq 0} r_k \left[p_{i-k}^n \frac{\phi_{i+\frac{1}{2}}^{n+1} - \phi_{i-k}^{n+1}}{\Delta X(k+\frac{1}{2})} + p_{i+k+1}^n \frac{\phi_{i+k+1}^{n+1} - \phi_{i+\frac{1}{2}}^{n+1}}{\Delta X(k+\frac{1}{2})} \right].$$

As previously for the case of second order in space accuracy, the conclusion is reached using the assumption that p_h is in BV, u_h is bounded in L^∞, ϕ is in $\mathscr{C}^1$ and $\sum_k r_k = 1$. One has

$$\lim_{h\to 0} \mathcal{T}^h_{2,1} = -\int_0^T \int_\Omega \widehat{pu}\partial_X \phi \mathrm{d}x\mathrm{d}t.$$

And, hence, it leads to

$$\int_0^T \int_\Omega \widehat{\rho_0 e}\partial_t \phi \mathrm{d}x\mathrm{d}t + \int_0^T \int_\Omega \widehat{pu}\partial_X \phi \mathrm{d}x\mathrm{d}t = \int_\Omega \widehat{\rho_0 e}(x,0)\phi(x,0)\mathrm{d}x.$$

Previous equation gives weak consistency for forward Euler staggered scheme with internal energy corrector. Using Runge–Kutta sequences adds only more technical difficulty in the algebra, but does not alter the weak consistency result. Idem for the use of the coefficients $\widehat{C}_k$. The key point for consistency is to use the same coefficients d_k and $\widehat{C}_k$ for both the internal and kinetic energies equations. ■

Remark II.5. Without internal energy corrector, for a forward Euler second order in space scheme, the first term writes

$$\begin{aligned}
\widehat{\mathcal{T}}^h_1 &= -\sum_n \Delta t \sum_i \Delta X \left[\rho_0 \epsilon_i^n \frac{\phi_i^{n+1} - \phi_i^n}{\Delta t} + \rho_0 e_{\mathrm{kin}}{}^n_{i+\frac{1}{2}} \frac{\phi_{i+\frac{1}{2}}^{n+1} - \phi_{i+\frac{1}{2}}^n}{\Delta t}\right] \\
&= -\sum_n \Delta t \sum_i \Delta X \left[\rho_0 \epsilon_i^n \frac{\phi_i^{n+1} - \phi_i^n}{\Delta t} + (\frac{1}{2}\rho_0 u^2)^n_{i+\frac{1}{2}} \frac{\phi_{i+\frac{1}{2}}^{n+1} - \phi_{i+\frac{1}{2}}^n}{\Delta t}\right] \\
&- \sum_n \Delta t \sum_i \Delta X \left[\left(\rho_0 e_{\mathrm{kin}}{}^n_{i+\frac{1}{2}} - (\frac{1}{2}\rho_0 u^2)^n_{i+\frac{1}{2}}\right) \frac{\phi_{i+\frac{1}{2}}^{n+1} - \phi_{i+\frac{1}{2}}^n}{\Delta t}\right] \\
&= \mathcal{T}^h_{1,1} + \mathcal{T}^h_{1,2}.
\end{aligned}$$

where $\mathcal{T}^h_{1,1}$ and $\mathcal{T}^h_{1,2}$ are defined by

$$\left\{\begin{aligned}
\mathcal{T}^h_{1,1} &= -\sum_n \Delta t \sum_i \Delta X \left[\rho_0 \epsilon_i^n \frac{\phi_i^{n+1} - \phi_i^n}{\Delta t} + (\frac{1}{2}\rho_0 u^2)^n_{i+\frac{1}{2}} \frac{\phi_{i+\frac{1}{2}}^{n+1} - \phi_{i+\frac{1}{2}}^n}{\Delta t}\right], \\
\mathcal{T}^h_{1,2} &= -\sum_n \Delta t \sum_i \Delta X \left[\left(\rho_0 e_{\mathrm{kin}}{}^n_{i+\frac{1}{2}} - (\frac{1}{2}\rho_0 u^2)^n_{i+\frac{1}{2}}\right) \frac{\phi_{i+\frac{1}{2}}^{n+1} - \phi_{i+\frac{1}{2}}^n}{\Delta t}\right].
\end{aligned}\right. \tag{II.26}$$

The term $\mathcal{T}^h_{1,1}$ has been dealt with as it is equal to the term $\mathcal{T}^h_1$ of the proof. Now, consider the term $\mathcal{T}^h_{1,2}$. Then under regularity hypothesis on the test function, one obtains something of the form

$$|\mathcal{T}^h_{1,2}| \leq C_\phi \|\rho_0 e_{\mathrm{kin}} - (\frac{1}{2}\rho_0 u^2)\|_{l^1([0:T]\times\Omega)}.$$

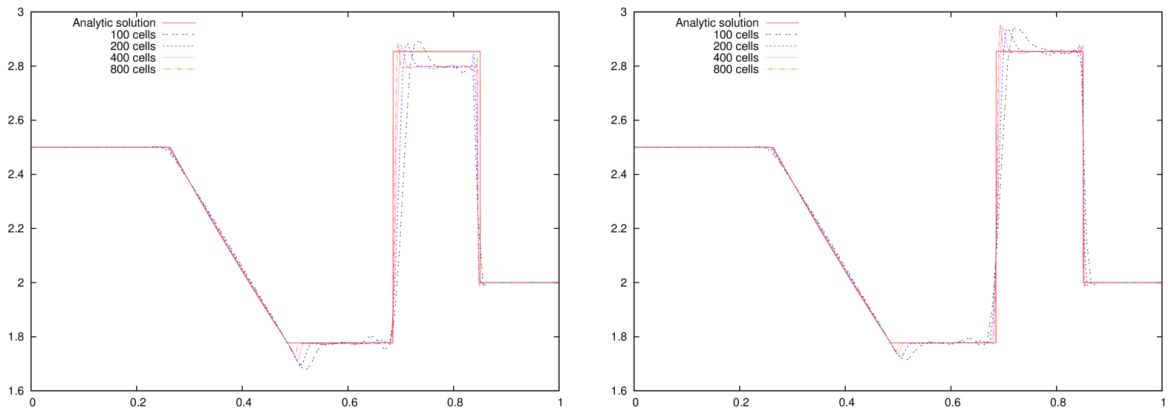


Figure II.2 – Illustration of the interest and importance of the internal energy corrector. Without the internal energy corrector (left), the scheme does not converge toward the weak solution for the Sod shock tube. Using the internal energy corrector (right), although oscillatory, the jump relations are recovered.

ΔX	$\frac{1}{100}$	$\frac{1}{200}$	$\frac{1}{400}$	$\frac{1}{800}$	$\frac{1}{1600}$
$\lVert\rho_0 e_{\text{kin}} - (\frac{1}{2}\rho_0 u^2)\rVert_{l^1([0:T]\times\Omega)}$	6.7e-3	4.7e-3	3.7e-3	3.3e-3	3.1e-3

Table II.7 – Illustration of the interest and importance of the internal energy corrector. Without the internal energy corrector, the term $\lVert\rho_0 e_{\text{kin}} - (\frac{1}{2}\rho_0 u^2)\rVert_{l^1([0:T]\times\Omega)}$ does not tend to 0 as ΔX and Δt tends to zero.

Experimentally, one observes that without internal energy corrector, $\lVert\rho_0 e_{\text{kin}} - (\frac{1}{2}\rho_0 u^2)\rVert_{l^1([0:T]\times\Omega)}$ does not tend to 0 as ΔX and Δt tend to 0. Indeed, here we present a short example where the internal energy corrector is most wanted to ensure correct capture of shocks. This example called the Sod shock tube [146] is presented later on. It is run with and without the energy corrector. Results are displayed in fig. II.2. On the left picture, we show that the scheme does not converge toward the analytical solution without the internal energy corrector. On the right one, we show that adding the internal energy corrector, the profile obtained in internal energy is much more satisfactory. In table II.7, we present the values of $\lVert\rho_0 e_{\text{kin}} - (\frac{1}{2}\rho_0 u^2)\rVert_{l^1([0:T]\times\Omega)}$, for the scheme without internal energy corrector, to assess that it does not tend to 0 experimentally. Further studies on the Sod shock tube are presented later on.

II-2.3 The remapping stage

The remapping is the algorithm designed to project the Lagrangian quantities on the original Cartesian grids, so that one gets a Cartesian Euler scheme. The quantities to be remapped at the end of the Lagrangian phase (II.13)-(II.14)-(II.21) are $\rho_0, \rho_0\epsilon$ on the primal grid $\{x_{i+\frac{1}{2}}\}$ and $\rho_0, \rho_0 u, \rho_0 e_{\text{kin}}$ on the dual one $\{x_i\}$. The projection detailed hereafter is equal to the one proposed in [50, 170] but adapted here to the staggered grids.

II-2.3.1 Lagrange polynomials based conservative projection

At the end of the Lagrangian phase, the primal deformed grid $\{x_{i+\frac{1}{2}}^{n+1}\}$ is known. In order to be able to project the staggered variables $\rho_0, \rho_0 u, \rho_0 e_{\text{kin}}$, one must be able to deduce the deformation of the dual grid $\{x_i^{n+1}\}$. This is done by using the coefficients r_k presented in table II.4, and using

$$x_i^{n+1} = \sum_k r_k (x_{i+k+\frac{1}{2}} + x_{i-k-\frac{1}{2}}),$$

which leads to locations of cell centers at high-order accuracy provided $\{x_{i+\frac{1}{2}}^{n+1}\}$ is also known at high-order accuracy. We consider any function $\phi \in L^\infty$, then the finite volume of $\overline{(\rho_0\phi)}_{\xi(i)}^{n+1}$ leads to

$$\Delta X \overline{(\rho_0\phi)}_{\xi(i)}^{n+1} = \int_{X_{\xi(i)-\frac{1}{2}}}^{X_{\xi(i)+\frac{1}{2}}} (\rho_0\phi)(Y, t^{n+1}) \mathrm{d}X + \mathscr{O}(\Delta X^N).$$

Using the definition of the change of variables $(x,t) \to (X,t)$, the integral computation yields

$$\Delta X \overline{(\rho_0\phi)}_{\xi(i)}^{n+1} = \int_{x_{\xi(i)-\frac{1}{2}}^{n+1}}^{x_{\xi(i)+\frac{1}{2}}^{n+1}} (\rho\phi)(y, t^{n+1}) \mathrm{d}y + \mathscr{O}(\Delta X^N). \tag{II.27}$$

Then, on the other hand, one has the following identity

$$\Delta X \overline{(\rho\phi)}_{\xi(i)}^{n+1} = \int_{x_{\xi(i)-\frac{1}{2}}}^{x_{\xi(i)+\frac{1}{2}}} (\rho\phi)(y, t^{n+1}) \mathrm{d}y + \mathscr{O}(\Delta X^N).$$

Using the integral linearity, it gives

$$\begin{aligned}\Delta X \overline{(\rho\phi)}_{\xi(i)}^{n+1} &= \int_{x_{\xi(i)-\frac{1}{2}}}^{x_{\xi(i)-\frac{1}{2}}^{n+1}} (\rho\phi)(y, t^{n+1}) \mathrm{d}y + \int_{x_{\xi(i)-\frac{1}{2}}^{n+1}}^{x_{\xi(i)+\frac{1}{2}}^{n+1}} (\rho\phi)(y, t^{n+1}) \mathrm{d}y \\ &+ \int_{x_{\xi(i)+\frac{1}{2}}^{n+1}}^{x_{\xi(i)+\frac{1}{2}}} (\rho\phi)(y, t^{n+1}) \mathrm{d}y + \mathscr{O}(\Delta X^N).\end{aligned}$$

Plugging eq. (II.27) into the previous equation, it yields

$$\Delta X \overline{(\rho\phi)}_{\xi(i)}^{n+1} = \int_{x_{\xi(i)-\frac{1}{2}}}^{x_{\xi(i)-\frac{1}{2}}^{n+1}} (\rho\phi)(y, t^{n+1}) \mathrm{d}y + \Delta X \overline{(\rho_0\phi)}_{\xi(i)}^{n+1} + \int_{x_{\xi(i)+\frac{1}{2}}^{n+1}}^{x_{\xi(i)+\frac{1}{2}}} (\rho\phi)(y, t^{n+1}) \mathrm{d}y + \mathscr{O}(\Delta X^N),$$

which written under conservative form, dropping the $\mathscr{O}(\Delta X^N)$, gives

$$\overline{(\rho\phi)}_{\xi(i)}^{n+1} = \overline{(\rho_0\phi)}_{\xi(i)}^{n+1} - \left[\frac{x_{\xi(i)+\frac{1}{2}}^{n+1} - x_{\xi(i)+\frac{1}{2}}}{\Delta X} (\rho_0\phi)_{\xi(i)+\frac{1}{2}}^{\star} - \frac{x_{\xi(i)-\frac{1}{2}}^{n+1} - x_{\xi(i)-\frac{1}{2}}}{\Delta X} (\rho_0\phi)_{\xi(i)-\frac{1}{2}}^{\star} \right], \tag{II.28}$$

where $(\rho_0\phi)^\star_{\xi(i)+\frac{1}{2}}$ satisfies

$$(\rho_0\phi)^\star_{\xi(i)+\frac{1}{2}} = \frac{1}{x^{n+1}_{\xi(i)+\frac{1}{2}} - x_{\xi(i)+\frac{1}{2}}} \int_{x_{\xi(i)-\frac{1}{2}}}^{x^{n+1}_{\xi(i)-\frac{1}{2}}} (\rho\phi)(y, t^{n+1})\mathrm{d}y. \tag{II.29}$$

One easily notices that $(\rho_0\phi)^\star_{\xi(i)+\frac{1}{2}}$ can be written as

$$(\rho_0\phi)^\star_{\xi(i)+\frac{1}{2}} = \frac{1}{x^{n+1}_{\xi(i)+\frac{1}{2}} - x_{\xi(i)+\frac{1}{2}}} \left(\int_{x_{\xi(i^\star)-\frac{1}{2}}}^{x^{n+1}_{\xi(i)-\frac{1}{2}}} (\rho\phi)(y, t^{n+1})\mathrm{d}y - \int_{x_{\xi(i^\star)-\frac{1}{2}}}^{x_{\xi(i)-\frac{1}{2}}} (\rho\phi)(y, t^{n+1})\mathrm{d}y \right)$$

with $i^\star$ an integer still to be determined to ensure both accuracy and stability. Then introducing the function $H^{\rho\phi}_{\xi(i^\star)}(x) = \int_{x_{\xi(i^\star)-\frac{1}{2}}}^{x} (\rho\phi)(y, t^{n+1})\mathrm{d}y$, one gets

$$(\rho_0\phi)^\star_{\xi(i)+\frac{1}{2}} = \frac{1}{x^{n+1}_{\xi(i)+\frac{1}{2}} - x_{\xi(i)+\frac{1}{2}}} \left(H^{\rho\phi}_{\xi(i^\star)}(x^{n+1}_{\xi(i)-\frac{1}{2}}) - H^{\rho\phi}_{\xi(i^\star)}(x_{\xi(i)-\frac{1}{2}}) \right). \tag{II.30}$$

Here, upwinded centered Lagrange polynomials are used to interpolate value of $H^{\rho\phi}_{\xi(i^\star)}$. The upwinding is done in function of sign of $x^{n+1}_{\xi(i)+\frac{1}{2}} - x_{\xi(i)+\frac{1}{2}}$. It yields natural value for $i^\star$ as a function of the upwinding and the order of the scheme N. In practice, one has

$$i^\star = \begin{cases} i - 1 - \lfloor \frac{N}{2} \rfloor & \text{if} \quad x^{n+1}_{\xi(i)+\frac{1}{2}} > x_{\xi(i)+\frac{1}{2}}, \\ i - \lfloor \frac{N-1}{2} \rfloor & \text{otherwise.} \end{cases} \tag{II.31}$$

II-2.3.2 Properties of the remap step

Lemma II.10. *The remap step* (II.28) *is conservative in mass, momentum, internal and kinetic energies. It conserves in particular the total energy $\mathcal{E}$ defined in definition II.2.*

Proof. The proof is straightforward. Indeed due to the conservative form depicted in eq. (II.28), the projection is conservative in mass, momentum, internal and kinetic energies. Thus, as $\mathcal{E}$ is the sum of both internal and kinetic energies, it is also conserved. ■

For the same motives mentioned in section II-2.2, the conservation of E is a desired feature. The dissipation of total energy during the remap phase is mentionned in the early literature. Indeed, as pointed out by DeBar [37, 38] "kinetic energy disappears in the momentum advection process, and must be compensated for in the internal energy if total energy conservation is to be maintained". It was also formulated similarly later by Youngs [176, 163].

Using the conservation of $\mathcal{E}$, the internal energy corrector eq. (II.21) is applied at the end of the remapping stage. It thus yields straightforwardly conservation of both $\mathcal{E}$ and E. Hence, three algorithms are available.

1. $\boxed{\text{Lagrange phase} \rightarrow \text{Internal energy corrector,}}$

2. $\boxed{\text{Lagrange phase} \to \text{Internal energy corrector} \to \text{Remap phase} \to \text{Internal energy corrector,}}$
3. $\boxed{\text{Lagrange phase} \to \text{Remap phase} \to \text{Internal energy corrector,}}$

The first algorithm is used to solve the Euler equations in Lagrangian coordinates, whereas the other two are used for the standard Euler equations. One can show that 2. and 3. are equivalent. In the following, the third algorithm is used.

Moreover, another CFL condition is imposed on the scheme, where now the time-step must satisfy

$$\Delta t < \frac{\Delta X}{\max_i |u_{i+\frac{1}{2}}|}.$$

This CFL condition comes directly from the stability of the Strang schemes derived in [43, 44]. A possible modification of the projection is to use monotonicity limiters in order to ensure the monotonic behaviour of the projection. In practice, one may apply the monotonicity preserving limiters [152] for more robustness during the remap phase. If not mentioned in numerical examples, limiters are not activated.

II-2.4 Numerical validation of the 1D conservative Lagrange-Remap schemes on staggered Cartesian grids

The numerical test-suite for validation contains among others three smooth test-problems which are the Cook–Cabot breaking wave test-case proposed in 2004 [28], a slight modification of the breaking wave using a non-convex equation of state and last an acoustic propagation which highlights the advantages concerning staggered grids schemes over cell-centered ones concerning the propagation of waves. Then, four shock test-problems are shown to illustrate the correct capture of shocks, among which the Sod test-case, the Woodward–Colella double blast wave and the Noh compression. The idea is to validate the schemes on a very large variety of test-cases to assess both accuracy and robustness. This is the real difficulty of the proposed test-suite. Recall that for all shock problems, additional artificial viscosities or hyperviscosities are never used. The dissipation induced by the time and space discretization is enough for the proposed test-suite.

II-2.4.1 Cook–Cabot breaking wave test-case [28]

The Cook–Cabot test-case is designed to assess numerically the order of accuracy of the schemes as the variables profiles are smooth until a given time T_{shock} where a discontinuity occurs. The breaking wave [28] initial data are set as follows:

$$\begin{cases} \rho = \rho_0(1+\alpha\sin(2\pi x)), \\ p = \quad p_0\left(\frac{\rho}{\rho_0}\right)^{\gamma}, \\ c = \quad c_0\left(\frac{\rho}{\rho_0}\right)^{(\gamma-1)/2}, \\ u = \quad \frac{2}{\gamma-1}(c_0 - c), \end{cases} \quad \text{for } -0.5 \le x \le 0.5 \tag{II.32}$$

with the constants defined as $\rho_0 = 10^{-3}$, $p_0 = 10^6$, $\gamma = \frac{5}{3}$ and $\alpha = 0.1$. T_{shock} is defined as

$$T_{\text{shock}} = \frac{1}{(\gamma+1)\pi\alpha c_0}.$$

The fluid is supposed to be a perfect gas. "For this set of initial conditions, two of the three caracteristics are initially constant, with the third satisfying a Burgers-like equation" [28]. The exact solution until T_{shock} is the initial profile advected with velocity $u-c$. The momentum error in l^1-norm as well as the experimental order of convergence are displayed in table II.8. Expected order of convergence are almost reached. For very high-order methods, the machine precision is already reached for 200 cells.

N_x	STAG-3		STAG-4		STAG-5		STAG-6		STAG-7		STAG-8	
50	9.3e-5	·	6.4e-6	·	5.3e-7	·	1.0e-7	·	3.1e-8	·	5.6e-9	·
100	1.2e-5	2.91	4.3e-7	3.89	2.0e-8	4.68	2.1e-9	5.64	2.6e-10	6.88	5.1e-11	6.79
200	1.6e-6	2.95	3.0e-8	3.86	7.7e-10	4.73	4.1e-11	5.69	2.8e-12	6.59	5.4e-13	6.56
400	2.0e-7	2.98	2.0e-9	3.93	2.6e-11	4.87	1.2e-12	5.1	8.2e-13	$\star$	8.6e-13	$\star$
800	2.6e-8	2.99	1.2e-10	3.96	1.8e-12	3.87	1.4e-12	$\star$	1.7e-12	$\star$	1.7e-12	$\star$
1600	3.2e-9	2.99	8.7e-12	3.85	3.6e-12	$\star$	1.5e-12	$\star$	3.0e-12	$\star$	2.8e-12	$\star$
3200	4.0e-10	3.00	6.2e-12	$\star$	3.8e-12	$\star$	2.2e-12	$\star$	3.3e-12	$\star$	3.1e-12	$\star$

Table II.8 – l^1-error in momentum and experimental order of convergence for the Lagrange-remap staggered scheme taken on the Cook-Cabot breaking wave test problem [28], until $t = 0.9T_{\text{shock}}$. $\star$ indicates machine precision reached.

II-2.4.2 Non-perfect gas breaking wave test-case

The previous test-case is designed for a perfect gas. A similar test-case but for arbitrary EOS gas can be defined. This time, the EOS is not convex and the initial data are set in such a way that the inflexion point is present in the computational domain. The initial data are

$$\left\{\begin{array}{rcl} \rho &=& \rho_0(1+\alpha\sin(2\pi x)), \\ c(\rho) &=& \sqrt{\gamma}\rho^{(\gamma-1)/2} + \beta_1\rho^{\beta_2}, \\ p(\rho) &=& \displaystyle\int c(\rho)^2 \mathrm{d}\rho, \\ u(\rho) &=& \displaystyle\int \frac{c(\rho)}{\rho}\mathrm{d}\rho, \end{array}\right. \quad \text{with} \quad \left\{\begin{array}{rcl} \alpha &=& 0.7, \\ \beta_1 &=& 0.03\sqrt{\gamma}\rho_0^{(\gamma-1)/2-\beta_2)}, \\ \beta_2 &=& -4, \\ \rho_0 &=& 1.4, \\ p_0 &=& 10^3. \end{array}\right. \tag{II.33}$$

The exact solution until T_{shock} is the initial profile advected with velocity $u-c$. The velocity error in l^1-norm as well as the experimental order of convergence are displayed in table II.9. Although the equation of state is not convex, expected order of convergence are reached by the staggered schemes.

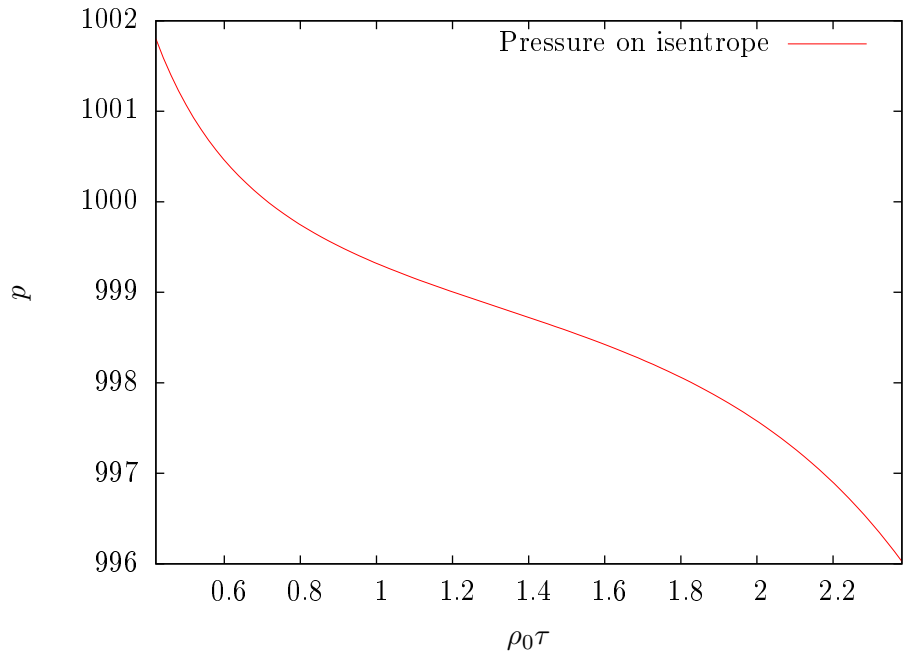


Figure II.3 – Non-convex equation of state for a breaking-wave test-case

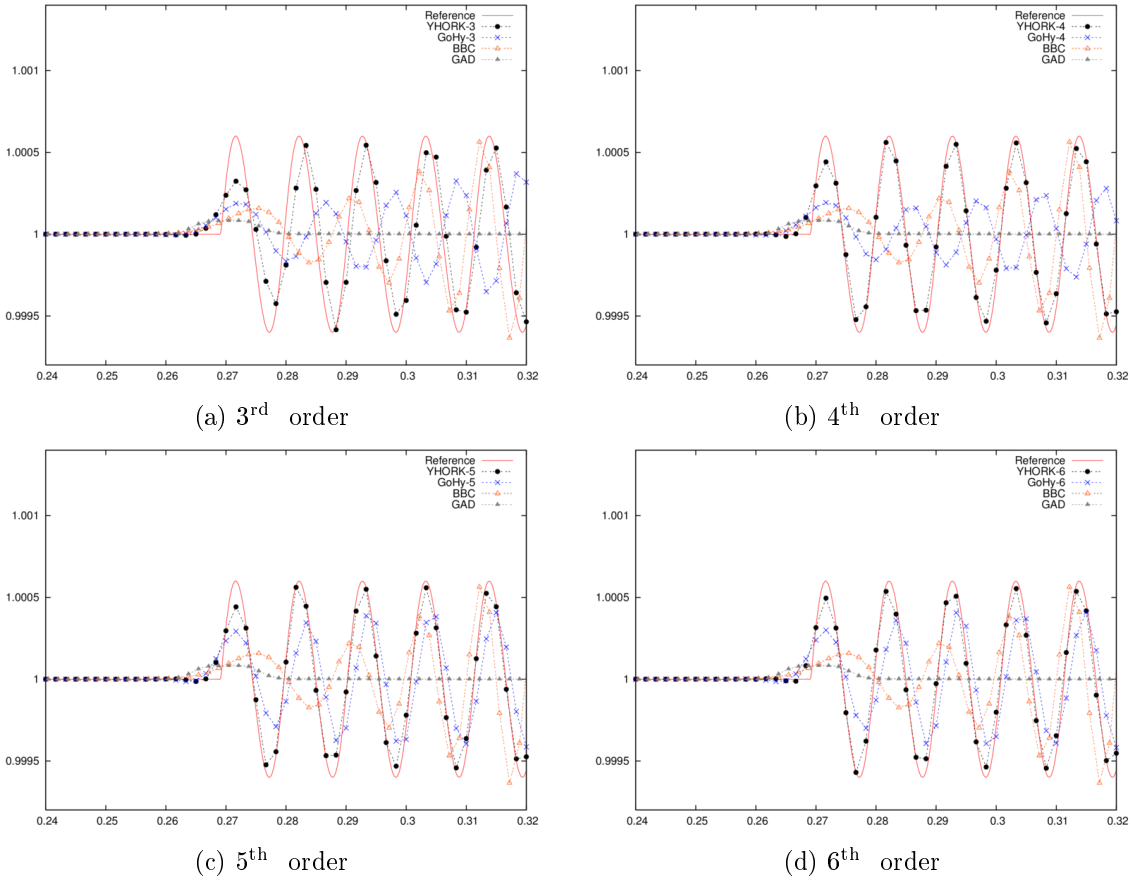
N_x	STAG-3		STAG-4		STAG-5		STAG-6		STAG-7		STAG-8	
50	4.2e-4	·	6.5e-4	·	2.6e-4	·	3.4e-4	·	2.9e-4	·	2.3e-4	·
100	1.9e-4	1.14	2.1e-4	1.64	5.7e-5	2.20	1.1e-4	1.67	4.4e-5	2.69	6.8e-5	1.73
200	4.5e-5	2.07	4.3e-5	2.29	1.5e-5	1.90	1.4e-5	2.91	1.3e-5	1.80	9.8e-6	2.80
400	9.3e-6	2.27	5.6e-6	2.94	1.7e-6	3.20	1.5e-6	3.25	6.5e-7	4.30	6.9e-7	3.83
800	1.5e-6	2.66	4.8e-7	3.54	9.9e-8	4.07	5.9e-8	4.64	2.0e-8	5.04	1.5e-8	5.56
1600	2.0e-7	2.89	3.1e-8	3.91	3.8e-9	4.69	1.3e-9	5.53	2.6e-10	6.24	1.2e-10	6.90
3200	2.6e-8	2.96	2.0e-9	3.98	1.3e-10	4.84	2.5e-11	5.68	7.5e-12	5.12	5.7e-12	4.43

Table II.9 – l^1-error in momentum and experimental order of convergence for the Lagrange-remap staggered scheme taken on the modified breaking wave test problem, until $t = 0.9T_{\text{shock}}$. $\star$ indicates machine precision reached.

II-2.4.3 Acoustic propagation test-case

This test-case is an acoustic oscillator. It is similar to a plate acting as a pressure harmonic source at $x = 1$. The mesh is chosen such that there are 7 cells by wavelength. The sound speed is set to 1. Slight modifications of pressure are imposed by the plate, such that the system of equations can be linearised. Comparisons between cell-centered schemes (GAD [84] and GoHy [50]) with the presented staggered schemes and the BBC schemes are drawn. Pressure profiles are depicted in fig. II.4 with a zoom on the wave front. As expected when the order of accuracy is increased, signal phase and amplitude are better restored by the schemes. Moreover, at equivalent order, the staggered schemes demonstrate a better restitution of both phase and amplitude of the signal. This one of the main advantages of B-type and C-type staggered schemes as pointed out in [5]. The initial data are

$$\begin{cases} \rho = \gamma, \\ p = 1, \quad \text{for } 0 \leq x \leq 0.5 \\ u = 0, \end{cases} \tag{II.34}$$



(a) 3rd order (b) 4th order

(c) 5th order (d) 6th order

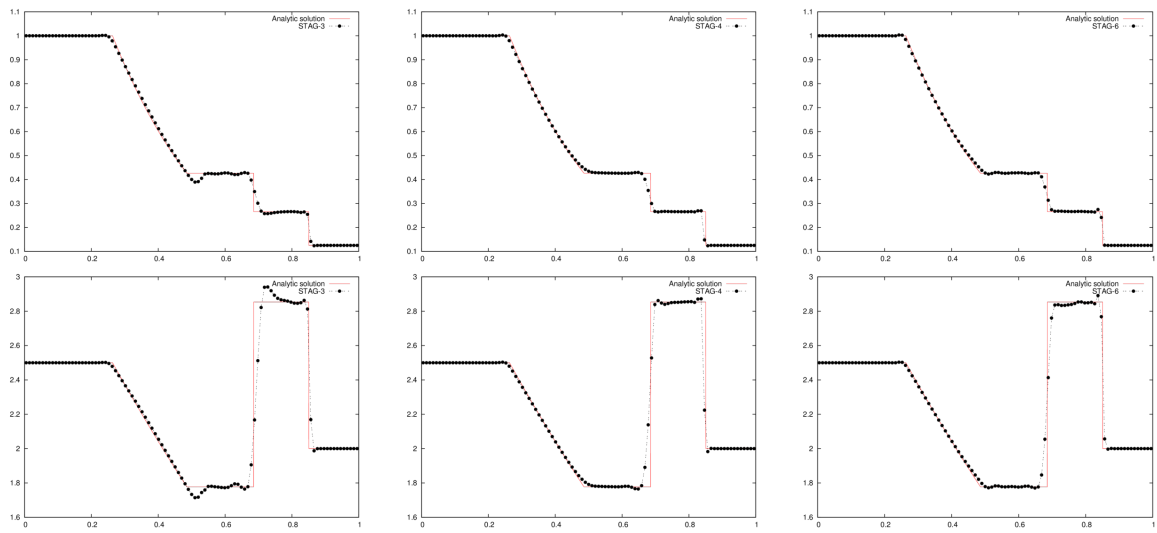
Figure II.4 – Acoustic wave with harmonic source - Difference between the cell-centered GoHy [50] (blue, cross) and GAD schemes [84] (gray, filled triangle), the staggered BBC scheme [171] (orange, triangle) and the new staggered schemes denoted here YHORK (black, filled circle). Analytic solution is represented by the red curve.

II-2.4.4 Sod test-case [146]

The Sod shock tube [146] is very common in the literature as a simple Riemann problem for the Euler equations. This test-case proves useful to determine the ability of the scheme to handle shocks and especially the capacity to recover correct discrete Rankine-Hugoniot relations at the shock using the proposed internal energy corrector. Initially, a left state and a right state trigger a rarefaction, contact discontinuity and shock. The domain is $[0:1]$ and the initial data are

$$\begin{cases} \rho_0(x) &= 1.0\chi_{\{x<0.5\}} + 0.125\chi_{\{x>0.5\}}, \\ p_0(x) &= 1.0\chi_{\{x<0.5\}} + 0.1\chi_{\{x>0.5\}}, \\ u_0(x) &= 0, \\ \gamma &= 1.4. \end{cases} \tag{II.35}$$

Wall boundary conditions are imposed at $x = 0$ and at $x = 1$. In fig. II.5, profiles of density and internal energy are depicted with the analytic solution for a mesh containing 100 cells. In table II.10, convergence results on density in norm l^1 are proposed. Although oscillatory due to the absence of artificial viscosities, convergence in the l^1-norm is achieved. As presented in fig. II.2, the scheme is not consistent without the internal energy corrector, and thus, the l^1 error does not converge to 0.



(a) 3rd order (b) 4th order (c) 6th order

Figure II.5 – Density (top) and internal energy (bottom) profiles on $[0:1]$ for the Sod test-case problem [146] at time $t = 0.2$, CFL=0.7, 100 cells, monotonicity limiters used during the remap phase, no artificial viscosities during the Lagrangian phase, for the 3rd, 4th and 6th order staggered schemes.

N_x	GAD	GoHy-3	BBC	STAG-3	STAG-4	STAG-5	STAG-6	STAG-7	STAG-8
50	2.92e-2	1.32e-2	1.81e-2	1.16e-2	1.00e-2	9.70e-3	1.03e-2	1.02e-2	8.69e-3
200	1.12e-2	3.91e-3	5.75e-3	3.47e-3	2.57e-3	2.50e-3	3.08e-3	5.64e-3	2.53e-3
800	3.96e-3	9.73e-4	1.51e-3	8.65e-4	7.82e-4	7.51e-4	7.09e-4	6.95e-4	6.59e-4
3200	1.37e-3	2.99e-4	4.86e-4	2.82e-4	2.38e-4	2.17e-4	2.24e-4	2.17e-4	2.02e-4
12800	4.56e-4	9.41e-5	1.67e-4	1.02e-4	6.86e-5	7.02e-5	9.43e-5	8.67e-5	6.02e-5
25600	2.61e-4	5.54e-5	1.00e-4	6.20e-5	3.80e-5	3.72e-5	7.07e-5	6.20e-5	4.94e-5

Table II.10 – l^1-error in density for the Lagrange-remap staggered scheme taken on the Sod test problem [146], until $t = 0.2$.

II-2.4.5 Noh test-case [127]

The Noh test-case [127] is a compression with a complete conversion of kinetic energy into internal energy. The domain is fixed at $[0:1]$. A continuous incoming flux of gas at $x=1$ is entering the computational domain with a constant speed and compress the gas located around $x=0$. We consider an incoming constant state of gas at $x=1$ and a wall boundary at $x=0$. The initial data are

$$\begin{cases} \rho_0 &= 1.0, \\ u_0 &= -1.0, \\ p_0 &= 10^{-8}, \\ \gamma &= \frac{5}{3}. \end{cases} \tag{II.36}$$

The analytical solution writes

$$\begin{cases} \rho(x,t) &= 4.0\chi_{\{x<\frac{t}{3}\}} + 1.0\chi_{\{x>\frac{t}{3}\}}, \\ u &= -1.0\chi_{\{x>\frac{t}{3}\}}, \\ p &= \frac{4}{3}\chi_{\{x<\frac{t}{3}\}} + 10^{-8}\chi_{\{x>\frac{t}{3}\}}, \end{cases} \tag{II.37}$$

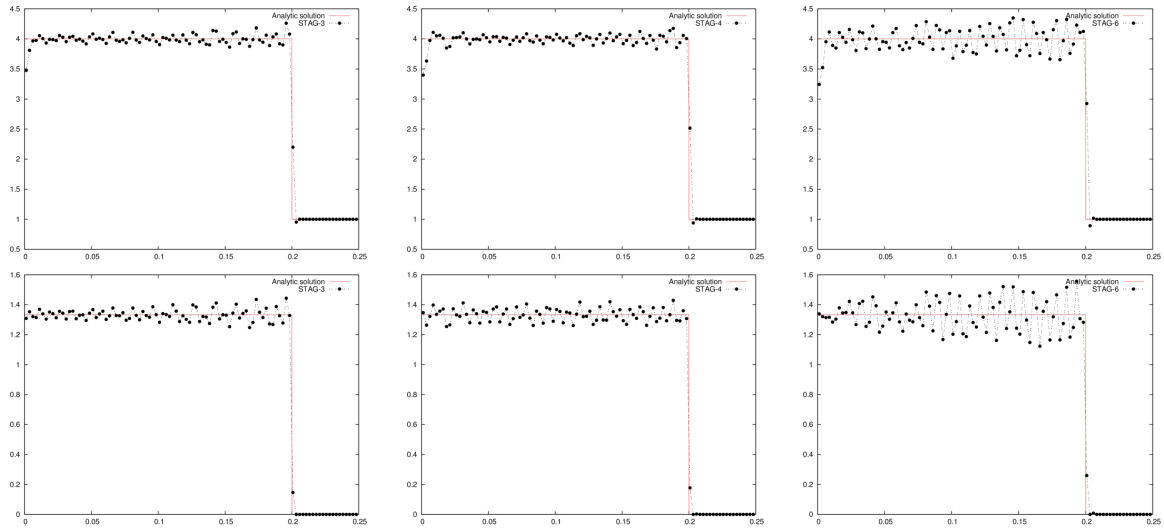
which gives an infinite shock intensity. This is a real difficulty for most schemes as highlighted in [127]. With this test-case, the robustness of the schemes is studied, without any artifical viscosity or hyperviscosity. In fig. II.6, profiles of density and pressure are depicted with the analytic solution for a mesh containing 400 cells over $[0:1]$. Zoom is made on $[0:0.25]$. The higher the order, the more oscillatory the profile is. This is due to the high-order approximations done in the scheme. Adding artificial viscosity with appropriate coefficients should smear out these oscillations. The important point is that even without artificial viscosity, the schemes even at very high-order are able to handle such a difficult test-case with an infinite shock intensity.

II-2.4.6 Shu-Osher test-case [144]

The Shu-Osher test-case [144] initial data are depicted in eq. (II.38)) on a $[-5:5]$ domain with a Mach 3 shock wave interacting with a sinusoidal density field. Computations till $t=1.8$ with CFL=0.7 are reported in fig. II.7. This test-case highlights the interest of high-order accuracy even on a shock problem, and especially the restitution of the density profile with high-order accurate schemes.

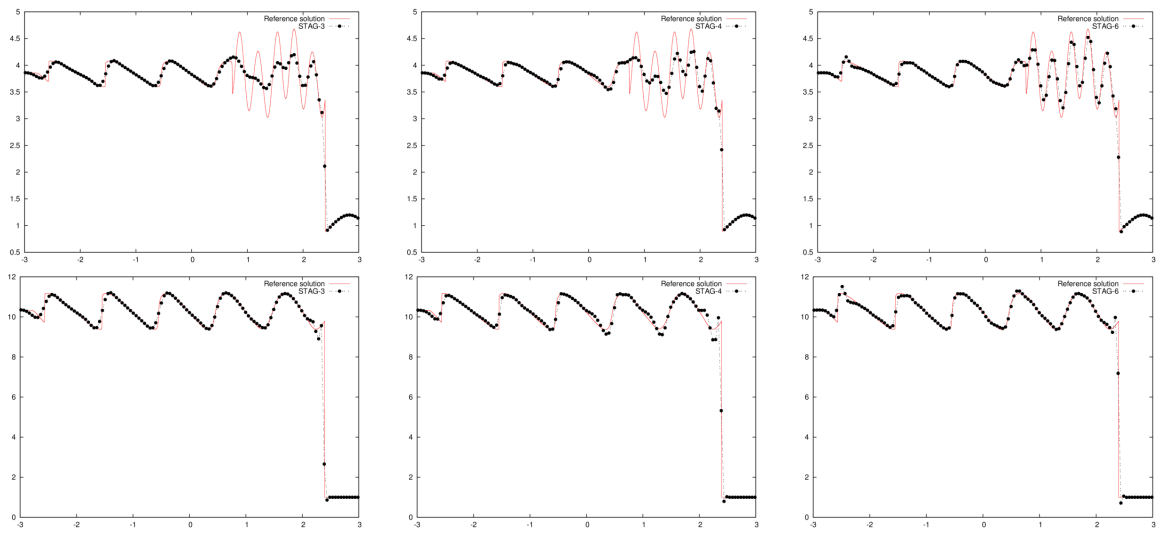
$$\begin{cases} \rho_0(x) &= \frac{27}{7}\chi_{\{x<-4\}} + (1+\frac{\sin(5x)}{5})\chi_{\{x>-4\}}, \\ p_0(x) &= \frac{31}{3}\chi_{\{x<-4\}} + 1\chi_{\{x>-4\}}, \\ u_0(x) &= \frac{4\sqrt{35}}{9}\chi_{\{x<-4\}}, \\ \gamma &= 1.4. \end{cases} \tag{II.38}$$

Reference solution is obtained using the GAD scheme with CFL=0.5 and 50000 cells.



(a) 3rd order (b) 4th order (c) 6th order

Figure II.6 – Density (top) and pressure (bottom) profiles on $[0:0.25]$ for the Noh test-case problem [127] at time $t = 0.6$, CFL=0.7, 400 cells, monotonicity limiters used during the remap phase, no artificial viscosities during the Lagrangian phase, for the 3rd, 4th and 6th order staggered schemes.



(a) 3rd order (b) 4th order (c) 6th order

Figure II.7 – Density (top) and pressure (bottom) profiles on $[-3:3]$ for the Shu-Osher test-case problem [144] at time $t = 1.8$, CFL=0.7, 200 cells, monotonicity limiters used during the remap phase, no artificial viscosities during the Lagrangian phase, for the 3rd, 4th and 6th order staggered schemes.

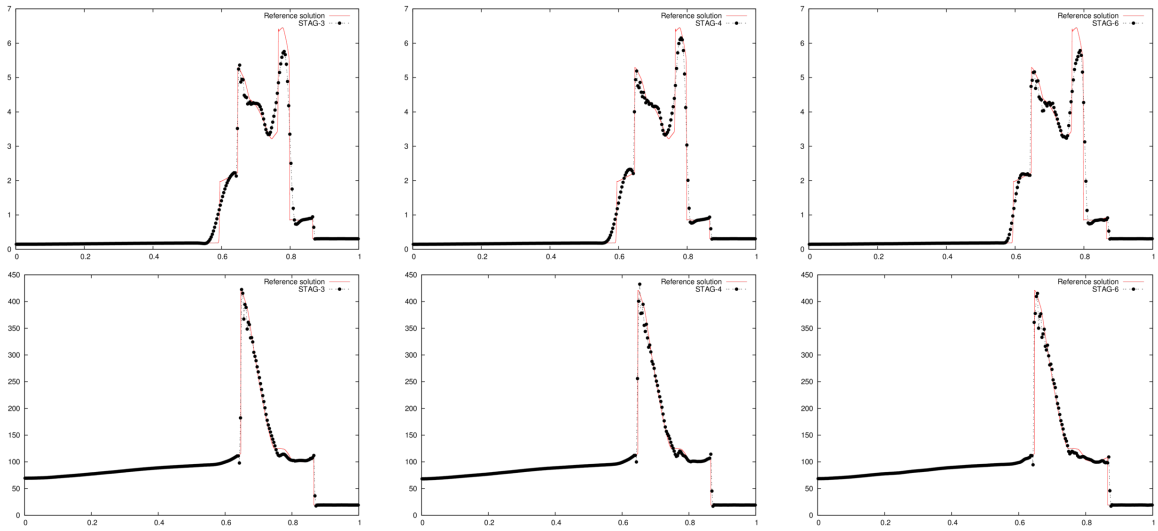
II-2.4.7 Interacting blast-waves test-case [171]

The interacting blast-waves test-case was proposed in [171]. It is a three states shock tube. The left blast will propagate to the right and the right one to the left till interaction between both. This test-case highlights the robustness of the schemes. The initial data are depicted in eq. (II.39). The domain is set to $[0:1]$. Wall boundary conditions are imposed at $x = 0$ and

$x = 1$.

$$\begin{cases} \rho_0(x) &= 1, \\ p_0(x) &= 1000\chi_{\{x<0.1\}} + 0.01\chi_{\{0.1<x<0.9\}} + 100\chi_{\{0.9<x\}}, \\ u_0(x) &= 0, \\ \gamma &= 1.4. \end{cases} \tag{II.39}$$

Density and pressure profiles are shown in fig. II.8. Reference solution is obtained using the GAD scheme with CFL=0.5 and 50000 cells. This interest of this test-case comes from the fact that both shocks are interacting which is a technical difficulty for low dissipative schemes as the one proposed without artificial viscosity or hyperviscosity.



(a) 3rd order (b) 4th order (c) 6th order

Figure II.8 – Density (top) and pressure (bottom) profiles on $[0:1]$ for the Woodward test-case problem [171] at time $t = 0.038$, CFL=0.7, 300 cells, monotonicity limiters used during the remap phase, no artificial viscosities during the Lagrangian phase, for the 3rd, 4th and 6th order staggered schemes.

II-3 Extension to 2D Lagrange-remap schemes on staggered Cartesian grids

As presented in [50, 170], the extension to the multidimensional case is realized using directional splitting. The Euler system in 2D writes

$$\begin{cases} \partial_t\rho + \partial_x(\rho u) + \partial_y(\rho v) = 0, \\ \partial_t(\rho u) + \partial_x(\rho u^2 + p) + \partial_y(\rho uv) = 0, \\ \partial_t(\rho v) + \partial_x(\rho uv) + \partial_y(\rho v^2 + p) = 0, \\ \partial_t(\rho e) + \partial_x(\rho ue + pu) + \partial_y(\rho ve + pv) = 0. \end{cases} \tag{II.40}$$

System in eq. (II.40) can be rewritten under the operator form

$$\partial_t(\boldsymbol{U}) + \boldsymbol{\mathcal{A}}(\boldsymbol{U}) = \boldsymbol{0}, \tag{II.41}$$

using $\boldsymbol{U} = \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ \rho e \end{pmatrix}$. The idea of the operator splitting is to find two operators $\boldsymbol{\mathcal{A}}_1$ and $\boldsymbol{\mathcal{A}}_2$ such that $\boldsymbol{\mathcal{A}}(\boldsymbol{U}) = \boldsymbol{\mathcal{A}}_1(\boldsymbol{U}) + \boldsymbol{\mathcal{A}}_2(\boldsymbol{U})$. For directional splitting, which is a peculiar class of operator splitting, the idea is to split $\boldsymbol{\mathcal{A}}$ such that all $x-$derivative are contained in $\boldsymbol{\mathcal{A}}_1$, and all $y-$derivatives are contained in $\boldsymbol{\mathcal{A}}_2$. First, derivation of the subsystems using the directional splitting method is made. Then, special distribution of variables is detailed for the staggered grids in 2D and 3D. This distribution allows then to apply the derived 1D staggered schemes to the nD cases. The schemes properties derived for the 1D case are then extended to the nD case. A numerical test suite is proposed to assess both accuracy and robustness of the schemes.

II-3.1 Derivation of the subsystems using the operator splitting technique

The main idea is to split system presented in eq. (II.40) according to the $x-$ and $y-$direction. It writes

$$\begin{cases} \partial_t\rho + \partial_x(\rho u) = 0 \\ \partial_t(\rho u) + \partial_x(\rho u^2 + p) = 0 \\ \partial_t(\rho v) + \partial_x(\rho u v) = 0 \\ \partial_t(\rho e) + \partial_x(\rho u e + p u) = 0 \end{cases} \qquad \begin{cases} \partial_t\rho + \partial_y(\rho v) = 0 \\ \partial_t(\rho u) + \partial_y(\rho u v) = 0 \\ \partial_t(\rho v) + \partial_y(\rho v^2 + p) = 0 \\ \partial_t(\rho e) + \partial_y(\rho v e + p v) = 0 \end{cases} \tag{II.42}$$

The above system in eq. (II.42) can be rewritten under a similar form as in eq. (II.41)

$$\partial_t(\boldsymbol{U}) + \boldsymbol{\mathcal{A}}_1(\boldsymbol{U}) = \boldsymbol{0}, \qquad \partial_t(\boldsymbol{U}) + \boldsymbol{\mathcal{A}}_2(\boldsymbol{U}) = \boldsymbol{0}. \tag{II.43}$$

Splitting techniques relies on solving alternatively first and second equation of eq. (II.43) with weighted time-steps in order to reach high-order accuracy. For Δt small enough, one can write

$$\boldsymbol{U}(t + \Delta t) = \mathrm{e}(\Delta t(\boldsymbol{\mathcal{A}}_1 + \boldsymbol{\mathcal{A}}_2))(\boldsymbol{U})(t). \tag{II.44}$$

Solving first equation of eq. (II.43) then the second one, one gets that

$$\widehat{\boldsymbol{U}}(t + \Delta t) = \mathrm{e}(\Delta t \boldsymbol{\mathcal{A}}_2)\, \mathrm{e}(\Delta t \boldsymbol{\mathcal{A}}_1)(\boldsymbol{U})(t). \tag{II.45}$$

Assuming that the operators $\mathcal{A}_1$ and $\mathcal{A}_2$ are commutative, the solution is then equivalent. If both are non commutative, then it is not. A simple Taylor expansion of both expressions yields

at most first order accuracy in time. The idea is then to set

$$\widehat{\boldsymbol{U}}(t+\Delta t) = \prod_{k=1}^{q} \mathrm{e}(\omega_{2k}\Delta t\boldsymbol{\mathcal{A}}_2)\,\mathrm{e}(\omega_{2k-1}\Delta t\boldsymbol{\mathcal{A}}_1)(\boldsymbol{U})(t), \tag{II.46}$$

where $(\omega_k)_{k\in[1:q]}$ is a sequence of parameters which are set to reach high-order accuracy in time. The theory of operator splitting and especially of high-order splitting sequences are extensively detailed by McLachlan in [113, 112, 114] and very high-order splitting methods are described by Yoshida in [175]. The weights ω_k are available in appendix, section A.2. Using directionnal splitting methods, each subsystems of eq. (II.43) is solved using the 1D schemes proposed in section II-2. However, slight modifications must be first performed. Indeed, as one wishes for global conservation of mass, momentum and total energy, use of values averaged in both directions is required, using rectangle control volumes. This is explained hereafter.

II-3.2 Modifications of the 1D schemes for the 2D finite volume case

The first important point to mention is the special distribution of variables on the staggered grids in both 2D and 3D. The extension of the internal energy corrector proposed for the 1D schemes is straightforward for multidimensional case.

II-3.2.1 nD distribution of variables on the modified Arakawa C-type grids

The distribution of variables on the modified Arakawa C-type grids is very similar to the one for the 1D case. The x-velocity u is staggered along the x-direction as well as the density and the kinetic energy related to the x-velocity u. It will be denoted in the following by $e_{\mathrm{kin},u}$. Then similarly, the y-velocity is staggered along the y-direction as well as the density and the kinetic energy $e_{\mathrm{kin},v}$ related to the y-velocity v. If one wishes to extend the schemes to the 3D case, then the z-velocity denoted w should be staggered along the z-direction along with the density and the kinetic energy $e_{\mathrm{kin},w}$. Distribution of variables is depicted on fig. II.9.

Then, for such a choice of variables, the total energy is the sum of the internal energy and the kinetic energies in each direction. This a key ingredient to yield conservation as will be shown hereafter.

II-3.2.2 Derivation of a procedure to apply the 1D schemes in one direction using the 2D finite volume formalism

The aim here is to apply with slight modifications the 1D schemes for two dimensions problem using directional splitting method. For two dimensions problem, the degree of freedom are the 2D-average value inside a cell. Thus it is mandatory at the beginning of a sweep, to deduce from the 2D average values the values average in only one direction. The procedure originates from [50, 170] and is extended here to staggered grids. A sweep along the x-direction proceeds as follows:

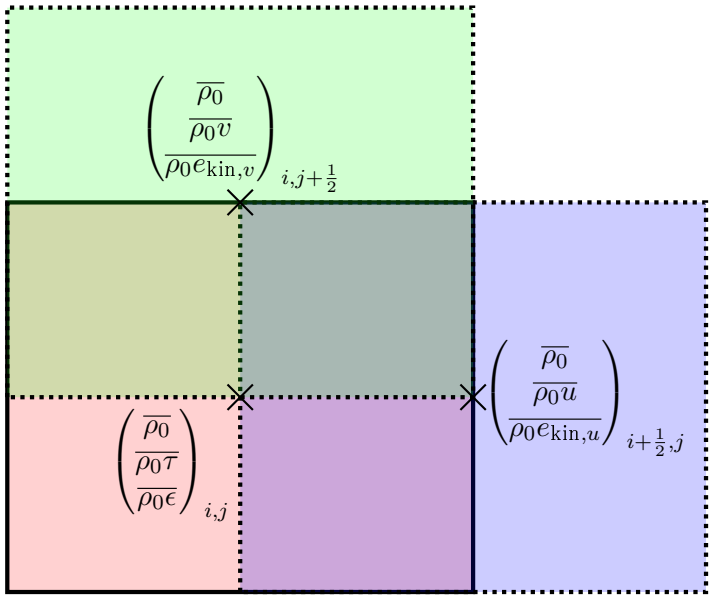


Figure II.9 – Staggered finite volume space discretization on Cartesian grids

1. Interpolate the 2D-values average values $\overline{\overline{\boldsymbol{U}}}$ along the y-direction to get 1D-cell-average values $\overline{\boldsymbol{U}}$ of the variables according to eq. (II.12). It writes for cell-centered variables

$$\overline{\boldsymbol{U}}^n_{i,j} = \sum_k C_k \overline{\overline{\boldsymbol{U}}}^n_{i,j+k}.$$

   This way, we only get 1D-cell-average values along the x-direction. This is exactly the values needed to use the 1D scheme.

2. Compute the 1D Lagrange evolution terms using $\overline{\boldsymbol{U}}$. Note that the velocity in the y-direction as well as its related kinetic energy do not change. The Lagrange evolution step gives values of the deformed grid $\{x_{i+\frac{1}{2},j}\}$. Interpolation gives value for the $\{x_{i,j}\}$ and $\{x_{i+\frac{1}{2},j+\frac{1}{2}}\}$ grids. The first grid is used to compute remap fluxes of the centered variables $(\rho_0, \rho_0\tau, \rho_0\epsilon)$, the second for the variables $(\rho_0, \rho_0 u, \rho_0 e_{\mathrm{kin},u})$ staggered along the x-direction, and the third one for the variables $(\rho_0, \rho_0 v, \rho_0 e_{\mathrm{kin},v})$ staggered along the y-direction.

3. Denote by $\Delta\boldsymbol{U}$ the evolution terms (see fig. II.10). Reconstruct the average values of $\Delta\boldsymbol{U}$ in the y-direction using eq. (II.12) denoted $\overline{\Delta\boldsymbol{U}}$. It writes for cell-centered variables

$$\overline{\Delta\boldsymbol{U}}^n_{i,j} = \sum_k \widehat{C}_k \Delta\boldsymbol{U}^n_{i,j+k}.$$

4. Apply the reconstructed 2D Lagrange-remap terms $\overline{\Delta\boldsymbol{U}}$ on the 2D-cell-average values. It leads for cell-centered variables to

$$\overline{\overline{\boldsymbol{U}}}^{n+1}_{i,j} = \overline{\overline{\boldsymbol{U}}}^n_{i,j} + \overline{\Delta\boldsymbol{U}}^n_{i,j}.$$

The procedure is summarized in fig. II.10.

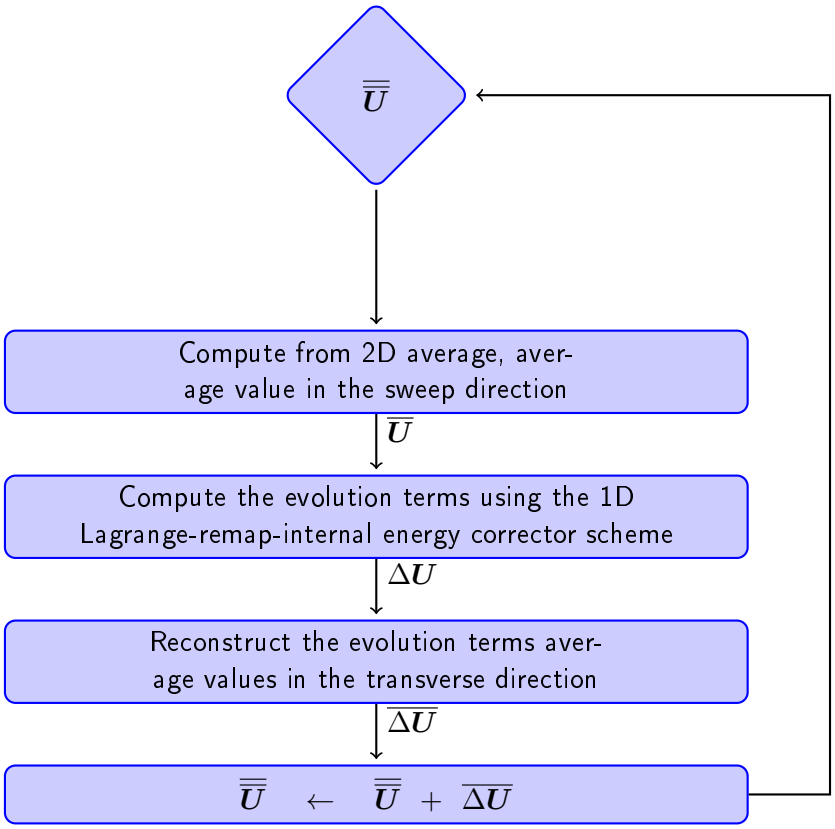


Figure II.10 – Flow chart of the 2D scheme

II-3.2.3 Properties of the 2D schemes

Lemma II.11. *The 2D staggered schemes* (II.13)-(II.14)-(II.21)-(II.28) *are conservative in mass, momentum and total energy E.*

Proof. With the proposed C-type staggering of variables, the 2D schemes satisfy lemmas II.2, II.8 and II.10 direction by direction and so are globally conservative in mass, momentum and total energy for any dimensional splitting sequences. ■

Remark II.6. Extension to the 3D case is straightforward.

Lemma II.12. *For a given directionnal splitting sequence $\{\omega_k\}$, the resulting 2D Cartesian grid schemes are linearly stable under the condition*

$$\Delta t < \frac{1}{\max_k \omega_k} \min\left(\Delta X \min(\frac{\lambda_{\text{STAG}}}{\max_{(i,j)} c_{i,j}}, \frac{1}{\max_{(i,j)} |u_{i+\frac{1}{2},j}|}), \Delta Y \min(\frac{\lambda_{\text{STAG}}}{\max_{(i,j)} c_{i,j}}, \frac{1}{\max_{(i,j)} |v_{i,j+\frac{1}{2}}|})\right).$$

Proof. Using lemma II.5 and stability of the remapping phase, one gets that $\mathcal{A}_1$ is linearly stable under the condition

$$\Delta t < \Delta X \left(\min(\frac{\lambda_{\text{STAG}}}{\max_{(i,j)} c_{i,j}}, \frac{1}{\max_{(i,j)} |u_{i+\frac{1}{2},j}|})\right)$$

and $\mathcal{A}_2$ under the condition

$$\Delta t < \Delta Y \left(\min(\frac{\lambda_{\text{STAG}}}{\max_{(i,j)} c_{i,j}}, \frac{1}{\max_{(i,j)} |v_{i,j+\frac{1}{2}}|}) \right).$$

Using the special structure of the operator splitting, one trivially gets the result. ■

II-3.3 Numerical validation of the 2D conservative Lagrange-Remap schemes on staggered Cartesian grids

A test-suite is proposed to assess both accuracy and robustness of the 2D staggered schemes. Once again, the wide range of problems is a high difficulty for numerical schemes. The idea here is to demonstrate the effectiveness of such schemes for such a variety of problems. First, numerical order of convergence of the method is assessed using the isentropic vortex advection [174]. Then, further vortex dynamics is studied with the vortex pairing problem [166]. Considering classical problems with strong discontinuities, five 2D Riemann problems are studied [139, 104, 108] to assess robustness and respect toward symmetry of the staggered schemes. Then two strong shocks problems are proposed: a strong blast-wave [140] and the 2D Noh compression problem [127]. Last, an extension of the 1D acoustic propagation problem is proposed with a 2D set up of acoustic propagation with a sound speed gradient in the vertical direction. It is derived from the works by Attenborough and al. [8].

II-3.3.1 Isentropic vortex advection [174]

We assess high-order accuracy on the 2D vortex test [174] whose initial data are given by (with $r^2 = x^2 + y^2$)

$$\begin{cases} \rho_0(x,y) &= \left(1 - \dfrac{(\gamma-1)\beta^2}{8\gamma\pi^2} e^{1-r^2}\right)^{\frac{1}{\gamma-1}}, \\ \boldsymbol{u}_0(x,y) &= (2,1)^t + \dfrac{\beta}{2\pi} e^{\frac{1-r^2}{2}} \cdot (-y,x)^t, \\ p_0(x,y) &= \rho_0(x,y)^\gamma, \\ \gamma &= 1.4 \end{cases} \tag{II.47}$$

with $\gamma = 1.4$ and $\beta = 5$. Computations are performed till $t = 20$ with a CFL number of 0.9 on the computational domain $\Omega = [-10, 10]^2$. Periodic boundary conditions are imposed. The l^1-error in both space and time is computed as

$$Err_{l^1} = \sum_n (t^{n+1} - t^n) \cdot \Delta x \cdot \Delta y \sum_{i,j} ||\overline{\rho}_{i,j}^n - \overline{\rho}_{i,j}^{exact}(t^n)||.$$

The l^1-error as well as experimental order of convergence are presented in table II.11. Expected orders of accuracy of the schemes are reached.

N_x	STAG-3		STAG-4		STAG-5		STAG-6		STAG-7		STAG-8	
50	3.3e-1	·	1.5e-1	·	2.6e-1	·	1.7e-1	·	1.5e-1	·	1.1e-1	·
100	9.5e-2	1.79	1.9e-2	3.01	4.9e-2	2.41	8.9e-3	4.27	1.2e-2	3.70	2.0e-3	5.83
200	1.6e-2	2.54	1.0e-3	4.19	1.9e-3	4.68	6.5e-5	7.10	8.0e-5	7.20	5.2e-6	8.59
400	2.2e-3	2.89	6.1e-5	4.06	6.1e-5	4.96	7.2e-7	6.48	6.3e-7	7.00	1.6e-8	8.37
800	2.8e-4	2.97	3.9e-6	3.99	1.9e-6	4.98	9.9e-9	6.18	5.0e-9	6.97	1.1e-10	7.17
1600	3.5e-5	2.99	2.4e-7	3.99	5.98e-8	4.99	1.5e-10	6.02	3.9e-11	6.99	3.4e-12	$\star$

Table II.11 – l^1-error in density and experimental order of convergence for the Lagrange-remap staggered scheme taken on the isentropic vortex advection test problem [174], until $t = 20$, CFL=0.9. $\star$ indicates machine precision reached.

II-3.3.2 Vortex-pairing test-case [166]

We assess here the ability of the staggered schemes to handle vortex dynamics with the vortex pairing test-case [166]. We first introduce the equation satisfied by a function ϕ advected by the velocity field $\boldsymbol{u}$,

$$\partial_t \phi + \nabla \cdot (\phi \boldsymbol{u}) = 0$$

In order to define the initial states, a perturbation function ψ is introduced as the sum of two Kelvin–Helmotz instability eigenmodes as

$$\psi(x,y) = A_1(y)\frac{\nu_1}{k_1}\cos(k_1 x)\,\mathrm{e}^{-k_1|y|} + A_2(y)\frac{\nu_2}{k_2}\cos(k_2 x)\,\mathrm{e}^{-k_2|y|}$$

with

$$A_i(y) = \frac{1-\mathrm{e}^{-2k_i(\frac{L}{2}-|y|)}}{1-\mathrm{e}^{-k_i L}}, i \in \{1,2\}.$$

Last, the initial data are given by

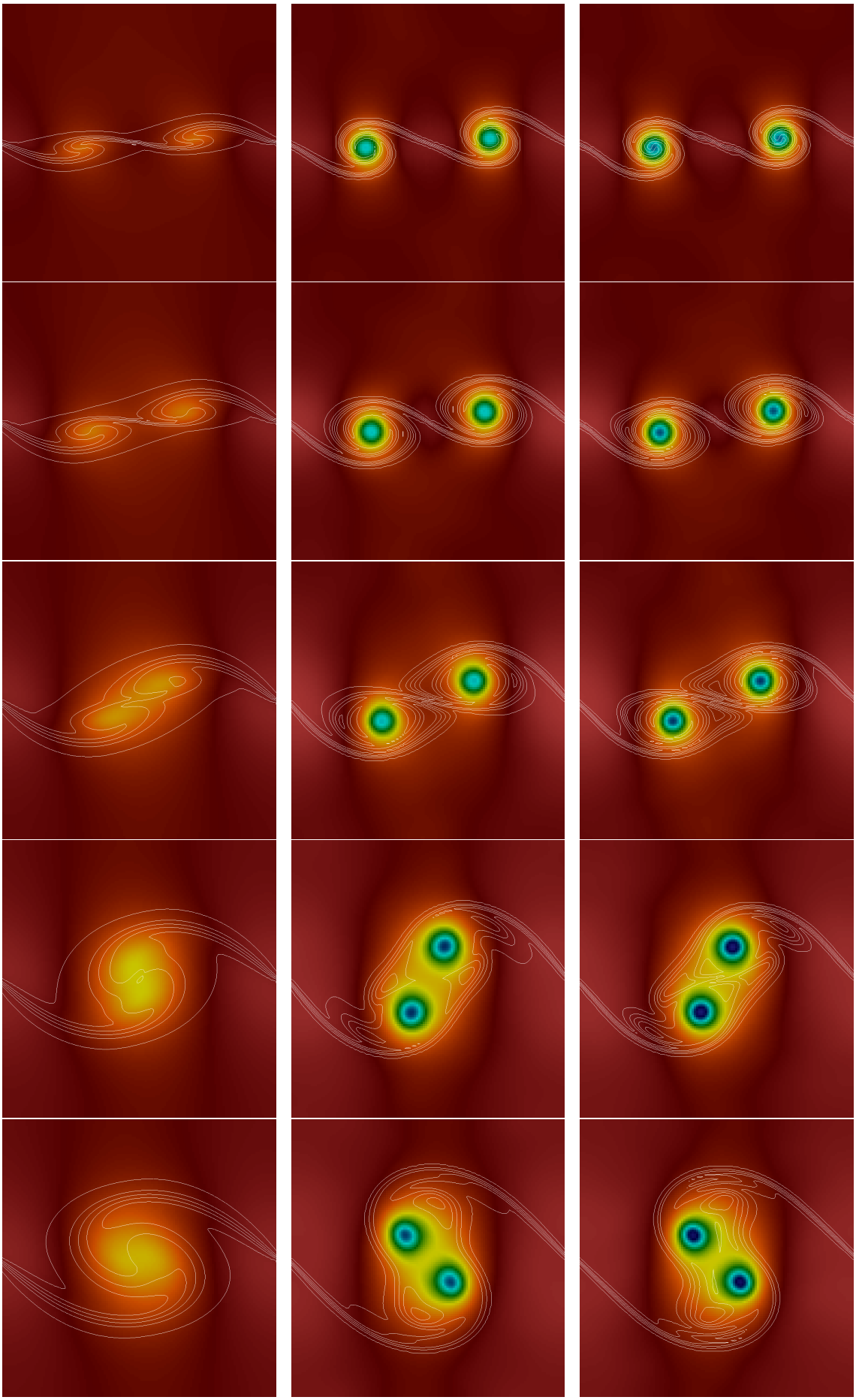
$$\begin{cases} \rho_0(x,y) &= 1.0, \\ \boldsymbol{u}_0(x,y) &= \begin{pmatrix} -\frac{1}{2}\Delta U \tanh(\frac{y}{2\theta_0}) - \partial_y \psi \\ \partial_x \psi \end{pmatrix}, \\ p_0(x,y) &= \rho_0(x,y)^\gamma, \\ \gamma &= \frac{5}{3}, \\ \phi_0(x,y) &= \chi_{\{y>0\}}. \end{cases} \tag{II.48}$$

Parameters are $k_1 = \frac{2\pi}{L}$, $k_2 = \frac{4\pi}{L}$, $\nu_1 = 0.025\Delta U$, $\nu_2 = 0.05\Delta U$, $\Delta U = 2.62$, $\theta_0 = 0.03$. Computations are performed till $t = 6.0$ with a CFL number of 0.9 on the computational domain $\Omega = [0,6] \times [-3,3]$. Periodic boundary conditions are imposed on left and right boundaries, and wall boundary conditions are imposed on top and bottom boundaries. In fig. II.11, the profile of density is depicted as well as 6 contours of ϕ from 0 to 1 on a coarse mesh with 128 cells along each direction. We present results using a first and second order cell-centered schemes and the proposed third order staggered scheme. First order scheme, as expected, struggles to restitute

the vortex dynamics. The second order scheme is more dissipative on the profile, but is still able to recover the vortex dynamics. Using high-order schemes gives a steeper profile for both the density and ϕ and hence yields a better restitution of vortex dynamics.

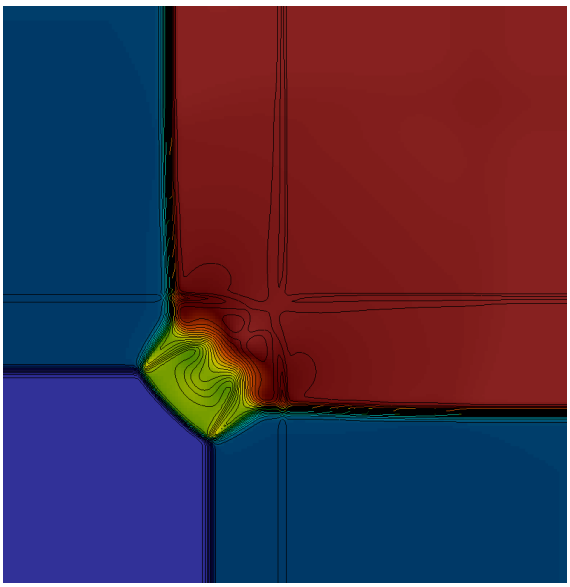
II-3.3.3 Five states Riemann problems [139, 104, 108]

We assess the robustness of the staggered schemes for 5 different 2D Riemann problems. The domain $\Omega = [0:1]^2$ is divided into four quadrants formed with the line $x = 1/2$ and $y = 1/2$. The Riemann problems are defined by constant states in each quadrant, with a perfect gas with $\gamma = 1.4$. These initial states in each quadrants are the density ρ_0, the pressure p_0, the x and y velocity u_0 and v_0. The selected Riemann problems are such that the solutions of all four 1D Riemann problems between quadrants have exactly one wave, which are whether a shock-wave (S), a rarefaction one (R) or a contact-slip (J) (see [104]). All initial data are gathered in table II.12 with the initial values of $(\rho, p, u, v)^t$ as well as the structure between two consecutive quadrants. Constant inflow boundary conditions are imposed. Computations are run with CFL=0.7 for the staggered schemes and with CFL=0.5 for the cell-centered ones (GAD available in [84], GoHy-2 available in [50, 170]). Monotonicity limiters are applied during the remap phase. No artificial viscosities are used. Results are depicted in figs. II.12 to II.16 with pressure profiles displayed using colors, and density using contours. Profiles are in accordance with those found in the literature [139, 104, 108] for all Riemann problems. In fig. II.12, some artefacts are present on two segments of the initial discontinuities between upper left, upper right and lower right quadrants. Moreover, oscillations are present due to the lack of artificial viscosities and dissipation. The symmetry along the axis $x = y$ is better recovered using the third order scheme than for the fourth order one. In fig. II.13, the main difference between results is that, as expected, the higher the order, the more oscillatory it is, but also the steeper is the profile concerning the contact-slip. This is expected due to high-order polynomial integration. A small density artefact is present in the lower right quadrant, but is also present in the literature. Pressure artefacts are present in the high-pressure areas, certainly due to the non-aligned grids. In fig. II.14, discontinuities are steeper as the order of accuracy is higher. The symmetry along the axis $x = y$ is quite well recovered. The stationary contacts bordering the lower left quadrant are well recovered. In fig. II.15, the stress is laid on the resolution of slowly moving contact discontinuities bordering the lower left quadrant. Vortex dynamics is already recovered with a coarse mesh using third and fourth order schemes. However, the second order cell-centered scheme shows a peculiar behaviour between the bottom quadrants with the formation of a small vortex. This is so far still unexplained. In fig. II.16, contact discontinuities are recovered on the line $x = \frac{1}{2}$. Moreover, the vortex induced by the interacting states is well recovered by high-order staggered methods, not so for the low order ones. Some artefacts are present in the right bottom quadrants, certainly due to boundary conditions that induce oscillations.

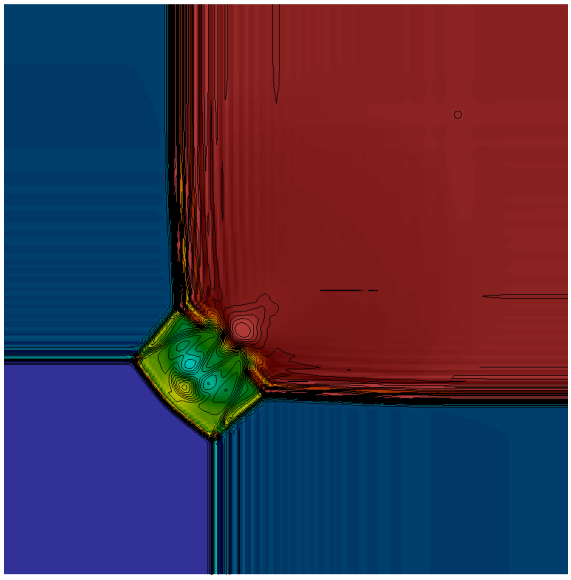


(a) 1st order (b) 2nd order (c) 3rd order

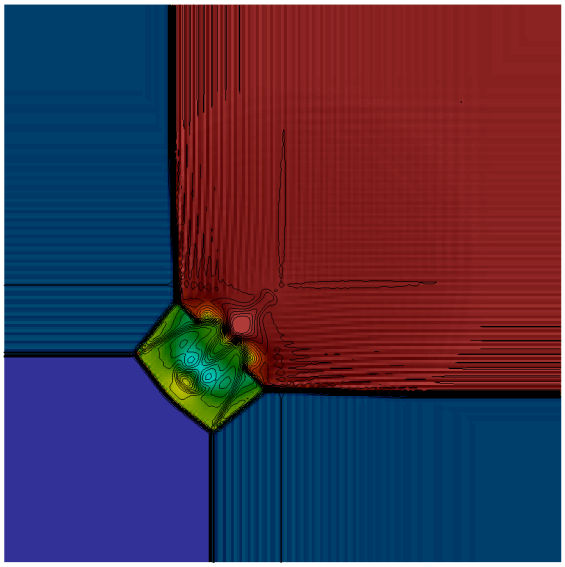
Figure II.11 – Profiles of density by colors and ϕ using 6 contours from 0 to 1 for the Vortex-Pairing test-case, CFL=0.7, for times $t = 1$, $t = 2$, $t = 3$, $t = 4$ and $t = 5$, 128 cells in each direction.



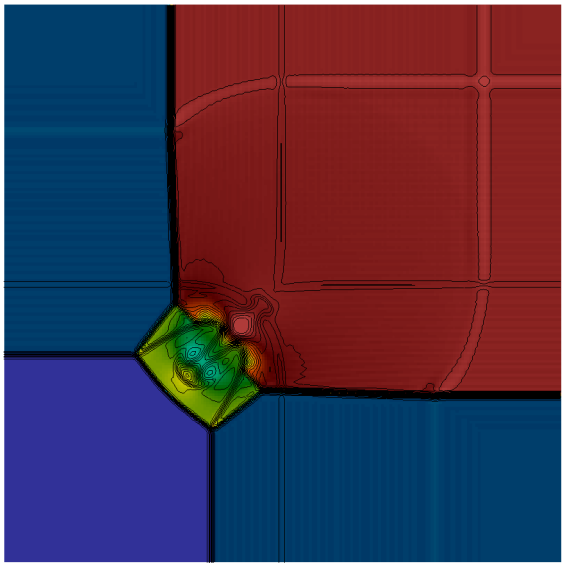
(a) 1st order



(b) 2nd order

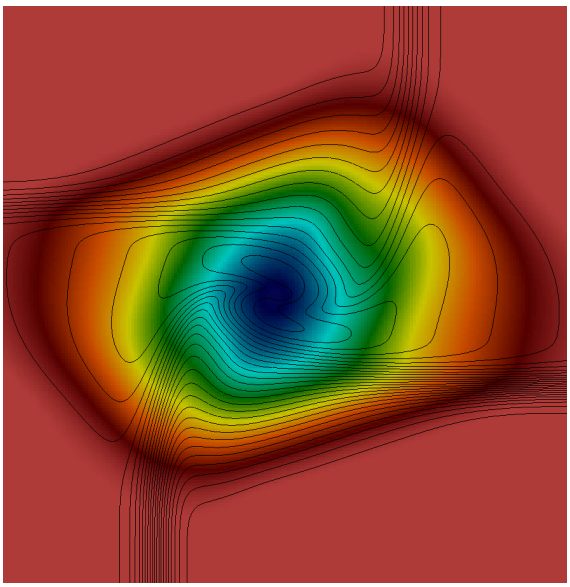


(c) 3rd order

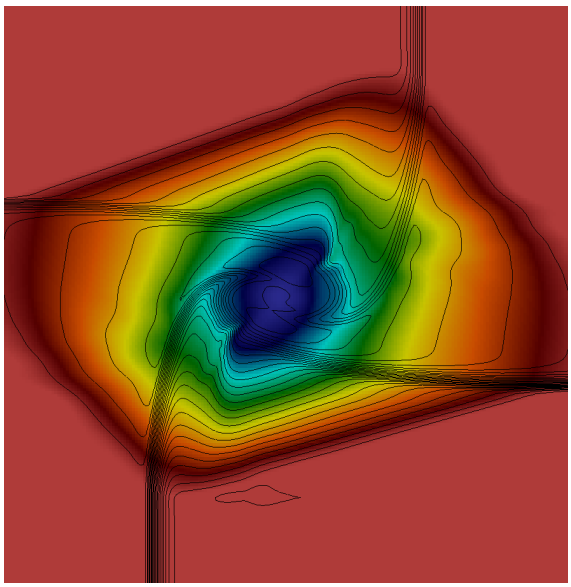


(d) 4th order

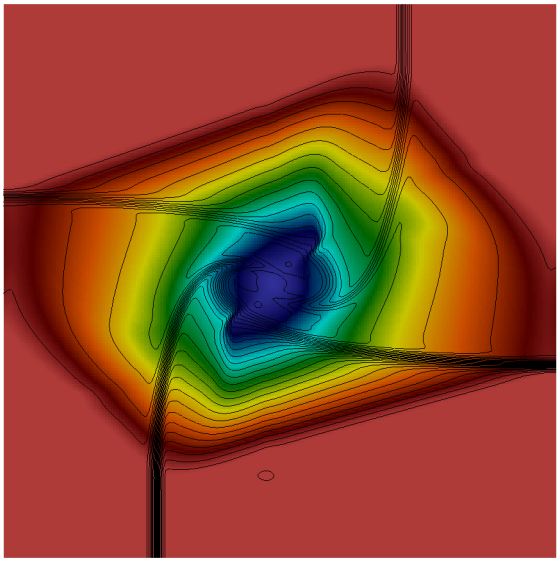
Figure II.12 – Results at time $t = 0.3$ for the first Riemann problem with the first and second order cell-centered scheme (top, CFL=0.5) as well as the third and fourth order staggered schemes (bottom, CFL=0.7) with 200 cells in each direction. Pressure is displayed by colors, and density using 32 contours from 0.16 to 1.71.



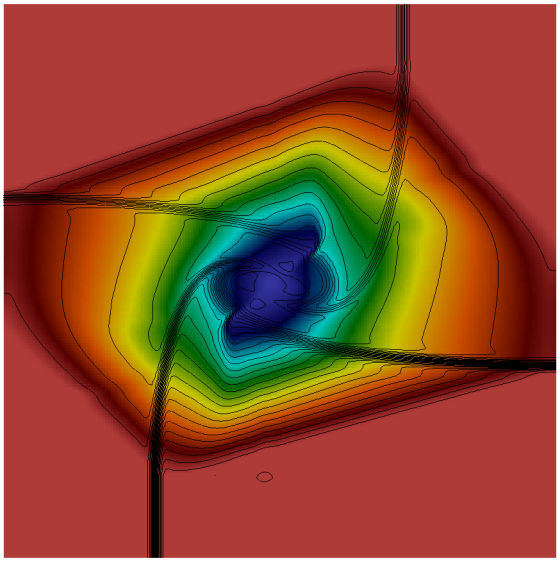
(a) 1st order



(b) 2nd order

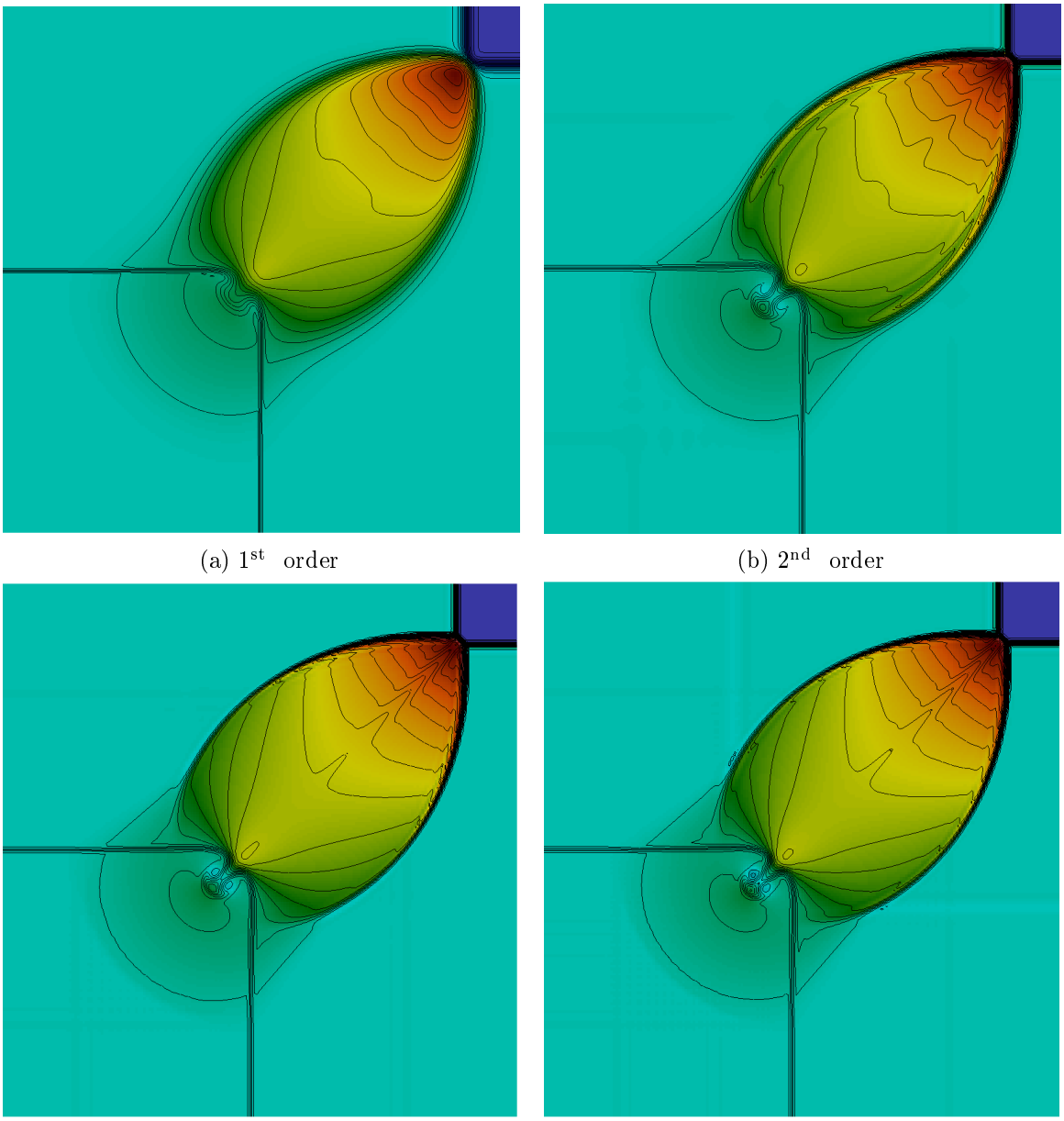


(c) 3rd order



(d) 4th order

Figure II.13 – Results at time $t = 0.3$ for the second Riemann problem with the first and second order cell-centered scheme (top, CFL=0.5) as well as the third and fourth order staggered schemes (bottom, CFL=0.7) with 200 cells in each direction. Pressure is displayed by colors, and density using 29 contours from 0.25 to 3.05.

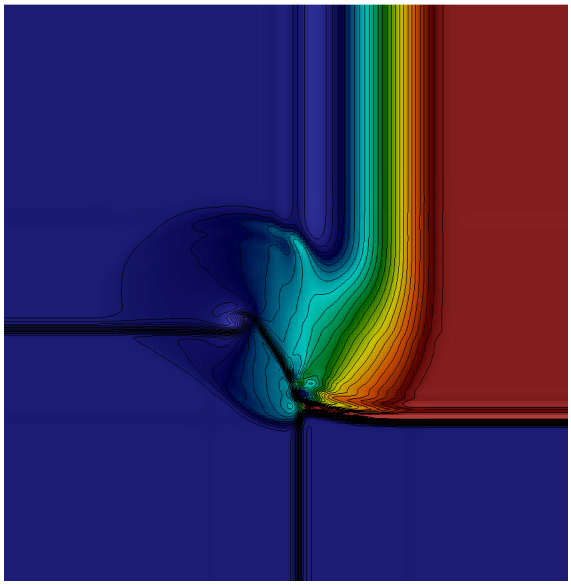


(a) 1st order (b) 2nd order

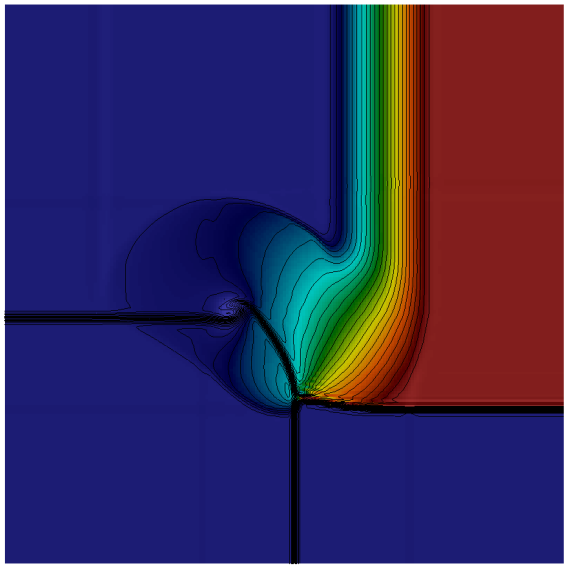
(c) 3rd order (d) 4th order

Figure II.14 – Results at time $t = 0.25$ for the third Riemann problem with the first and second order cell-centered scheme (top, CFL=0.5) as well as the third and fourth order staggered schemes (bottom, CFL=0.7) with 200 cells in each direction. Pressure is displayed by colors, and density using 30 contours from 0.54 to 1.7.

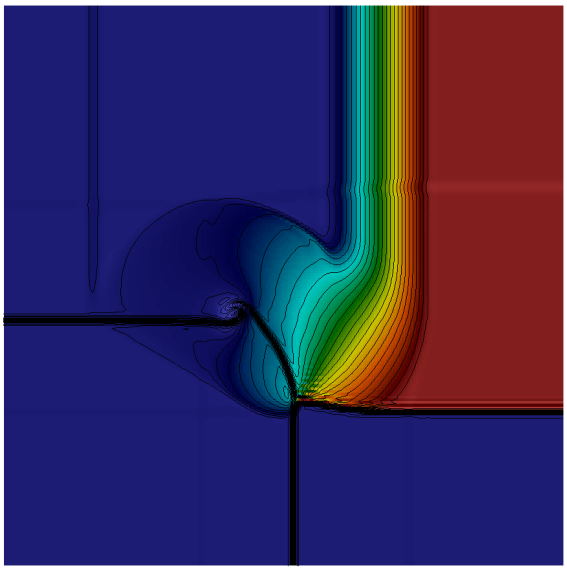
(a) 1st order



(b) 2nd order

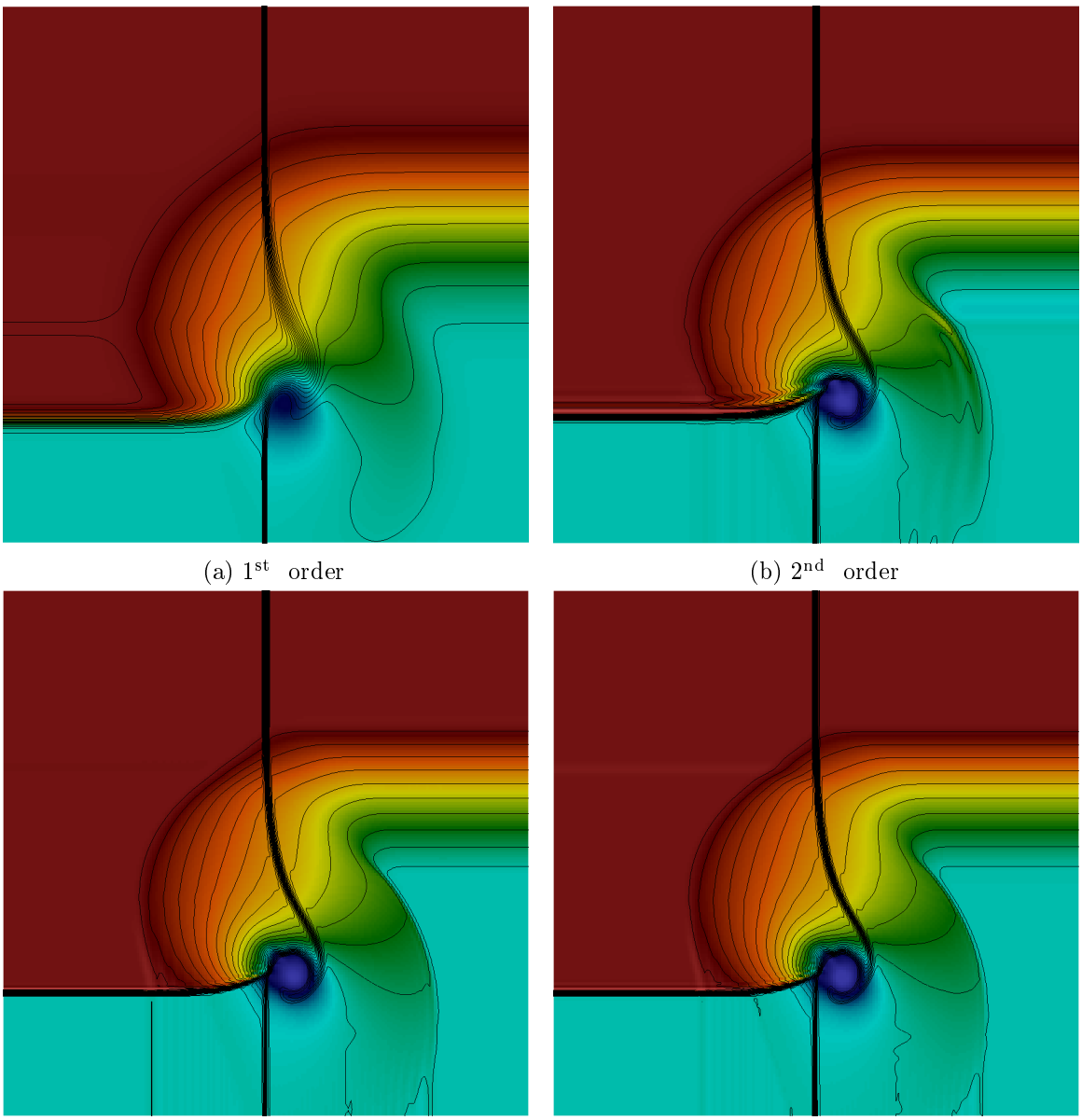


(c) 3rd order



(d) 4th order

Figure II.15 – Results at time $t = 0.25$ for the fourth Riemann problem with the first and second order cell-centered scheme (top, CFL=0.5) as well as the third and fourth order staggered schemes (bottom, CFL=0.7) with 200 cells in each direction. Pressure is displayed by colors, and density using 29 contours from 0.43 to 0.99.



(a) 1st order (b) 2nd order

(c) 3rd order (d) 4th order

Figure II.16 – Results at time $t = 0.25$ for the fifth Riemann problem with the first and second order cell-centered scheme (top, CFL=0.5) as well as the third and fourth order staggered schemes (bottom, CFL=0.7) with 200 cells in each direction. Pressure is displayed by colors, and density using 30 contours from 0.53 to 1.98.

Problem	upper left	$\rightsquigarrow$	upper right	$\rightsquigarrow$	bottom right	$\rightsquigarrow$	bottom left	$\rightsquigarrow$
1	$\begin{pmatrix}0.5323\\0.3\\1.206\\0\end{pmatrix}$	S	$\begin{pmatrix}1.5\\1.5\\0\\0\end{pmatrix}$	S	$\begin{pmatrix}0.5323\\0.3\\0\\1.206\end{pmatrix}$	S	$\begin{pmatrix}0.138\\0.029\\1.206\\1.206\end{pmatrix}$	S
2	$\begin{pmatrix}2\\1\\0.75\\0.5\end{pmatrix}$	J	$\begin{pmatrix}1\\1\\0.75\\-0.5\end{pmatrix}$	J	$\begin{pmatrix}3\\1\\-0.75\\-0.5\end{pmatrix}$	J	$\begin{pmatrix}1\\1\\-0.75\\0.5\end{pmatrix}$	J
3	$\begin{pmatrix}1\\1\\0.7276\\0\end{pmatrix}$	J	$\begin{pmatrix}0.5343\\0.4\\0\\0\end{pmatrix}$	S	$\begin{pmatrix}1\\1\\0\\0.7276\end{pmatrix}$	S	$\begin{pmatrix}0.8\\1.0\\0\\0\end{pmatrix}$	J
4	$\begin{pmatrix}0.5197\\0.4\\-0.6259\\-0.3\end{pmatrix}$	J	$\begin{pmatrix}1\\1\\0.1\\-0.3\end{pmatrix}$	R	$\begin{pmatrix}0.5313\\0.4\\0.1\\0.4276\end{pmatrix}$	S	$\begin{pmatrix}0.8\\0.4\\0.1\\-0.3\end{pmatrix}$	J
5	$\begin{pmatrix}2\\1\\0\\-0.3\end{pmatrix}$	S	$\begin{pmatrix}1\\1\\0\\-0.4\end{pmatrix}$	J	$\begin{pmatrix}0.5197\\0.4\\0\\-1.1259\end{pmatrix}$	R	$\begin{pmatrix}1.0625\\0.4\\0\\0.2145\end{pmatrix}$	J

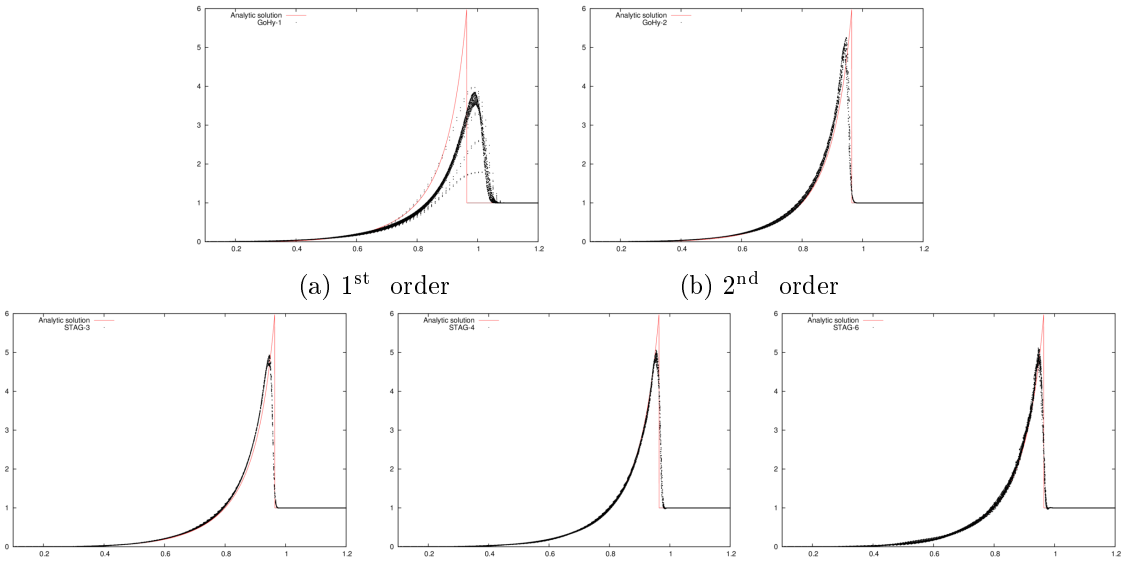
Table II.12 – Initial states for the four quadrants of 2D Riemann problem for density, pressure and x and y velocity u and v.

II-3.3.4 Sedov test-case [140]

With the Sedov test-case, we assess the robustness of the staggered schemes as well as the ability to restitue correct cylindrical symmetry. Let $r_{\text{Sedov}} = \frac{1}{\sqrt{2}}\sqrt{\Delta X^2 + \Delta Y^2}$. Initial data are

$$\begin{cases} \rho_0(x,y) &= 1, \\ \boldsymbol{u}_0(x,y) &= \mathbf{0}, \\ p_0(x,y) &= \dfrac{(\gamma-1)\epsilon_{\text{Sedov}}}{\pi r_{\text{Sedov}}^2}\chi_{\{x^2+y^2<r_{\text{Sedov}}^2\}} + 10^{-14}\chi_{\{x^2+y^2>r_{\text{Sedov}}^2\}}, \\ \gamma &= 1.4, \end{cases} \tag{II.49}$$

where $\epsilon_{\text{Sedov}} = 0.851072$. A scatter plot is realized to display profiles of density along each radius in fig. II.17 using 100 cells in each direction. Even without the use of artificial viscosities, the density profile is quite smooth for each scheme. The higher the order of the staggered schemes, the better the maximum of density near the shock is recovered. The shock position is in good agreement with the analytic solution for the three staggered schemes. Results for first and second order cell-centered schemes are presented for comparison.



(a) 1st order (b) 2nd order

(c) 3rd order (d) 4th order (e) 6th order

Figure II.17 – Scatter plot of density profiles for the Sedov blast-wave test-case using the third, fourth and sixth order staggered schemes (CFL=0.7) and the first and second order cell-centered schemes (CFL=0.5) at $t = 1.0$; 100 cells in each direction.

II-3.3.5 Noh test-case [127]

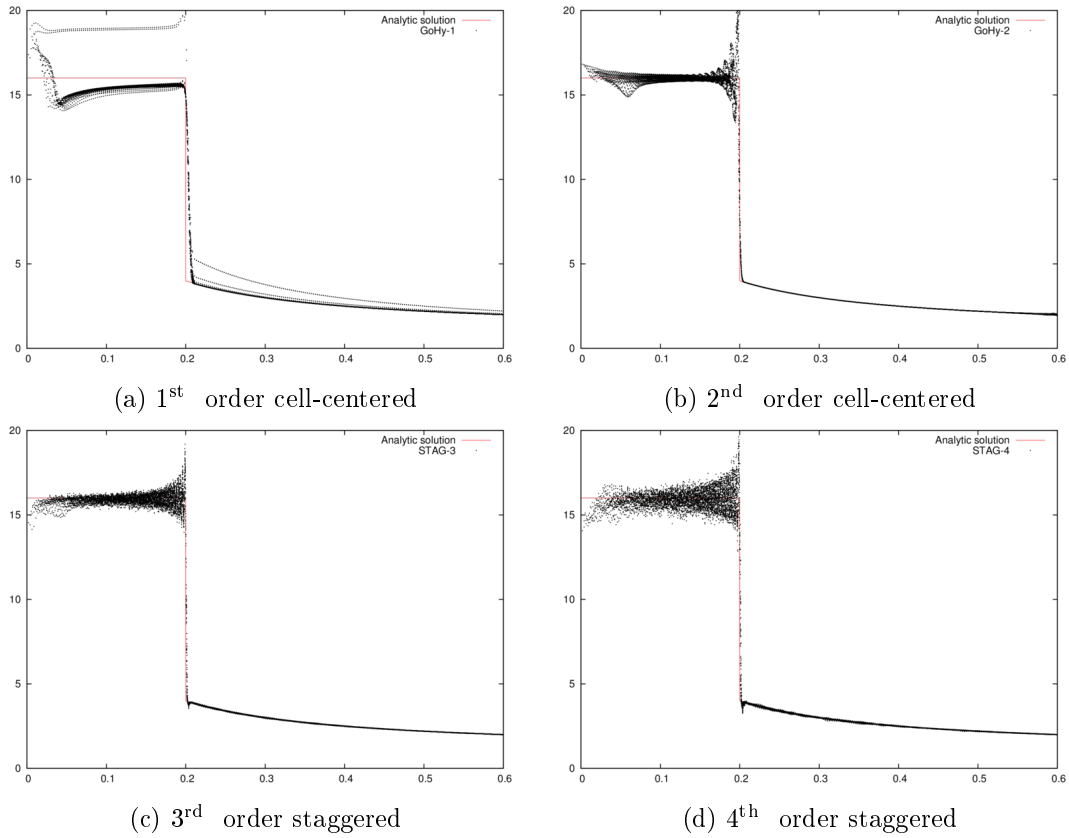
Exactly as in the 1D case, the kinetic energy is transformed into internal energy, giving a compression of the gas by a factor 16. Denote $r = \sqrt{x^2 + y^2}$, initial data are

$$\begin{cases} \rho_0(x,y) &= 1, \\ \boldsymbol{u}_0(x,y) &= \dfrac{1}{r}\begin{pmatrix} -x \\ -y \end{pmatrix}, \\ p_0(x,y) &= 10^{-8}, \\ \gamma &= \dfrac{5}{3}. \end{cases} \tag{II.50}$$

Considering free inflow boundary conditions the analytic solution writes, with $r_s(t) = \frac{\gamma-1}{2}t$,

$$\begin{cases} \rho(x,y,t) &= \left(\dfrac{\gamma+1}{\gamma-1}\right)^2 \chi_{r<r_s(t)} + (1+\dfrac{t}{r})\chi_{r>r_s(t)}, \\ \boldsymbol{u}_0(x,y,t) &= \dfrac{1}{r}\begin{pmatrix} -x \\ -y \end{pmatrix}\chi_{r>r_s(t)}, \\ p_0(x,y,t) &= \dfrac{1}{2}\dfrac{(\gamma+1)^2}{\gamma-1}\chi_{r<r_s(t)} + 10^{-8}\chi_{r>r_s(t)}, \\ \gamma &= \dfrac{5}{3}. \end{cases} \tag{II.51}$$

A scatter plot is realized to display profiles of density along each radius in fig. II.18 using 400 cells in each direction. Without artificial viscosities, the sixth order scheme fails, and therefore is not



(a) 1st order cell-centered (b) 2nd order cell-centered

(c) 3rd order staggered (d) 4th order staggered

Figure II.18 – Scatter plot of density profiles for the 2D Noh compression test-case using the third, fourth order staggered schemes (CFL=0.7) and for the first and second order cell-centered schemes (CFL=0.5) at $t = 0.6$, 400 cells in each direction. Axis effect are present for the first and second order cell-centered schemes

presented in the results. Obviously, the fourth order scheme is much more oscillatory than the third order one. Otherwise, even without the use of artificial dissipation, the compression by a factor 16 is recovered by the staggered schemes, except near the point $(0, 0)$ due to wall heating. The artefacts present for the first and second order cell-centered schemes are not present with the staggered ones. Those are certainly due to wall boundary conditions (as highlighted by Noh in [127]). However, due to high-order polynomial interpolation, results are more oscillatory.

II-3.3.6 Attenborough test-case [8]

We assess here the ability of the staggered schemes to recover correctly long-range acoustic propagation with the Attenborough test-case [8, 39] which has been designed by the geoacoustic community. In 1D, it has been highlighted during numerical experiments that the high-order staggered schemes require less cells per wavelength compared to same order cell-centered schemes. We here want to check that this result still holds in 2D and see if the signal is correctly recovered by the schemes. Comparisons are drawn with results available in the literature [39]. The computational domain is $\Omega = [0, 5000] \times [0, 4000]$. Initially the domain is filled with a perfect gas at rest, with $\gamma = 1.4$ and at the atmospheric pressure ($p_{\text{atm}} = 10^5$ Pa). A gradient in the

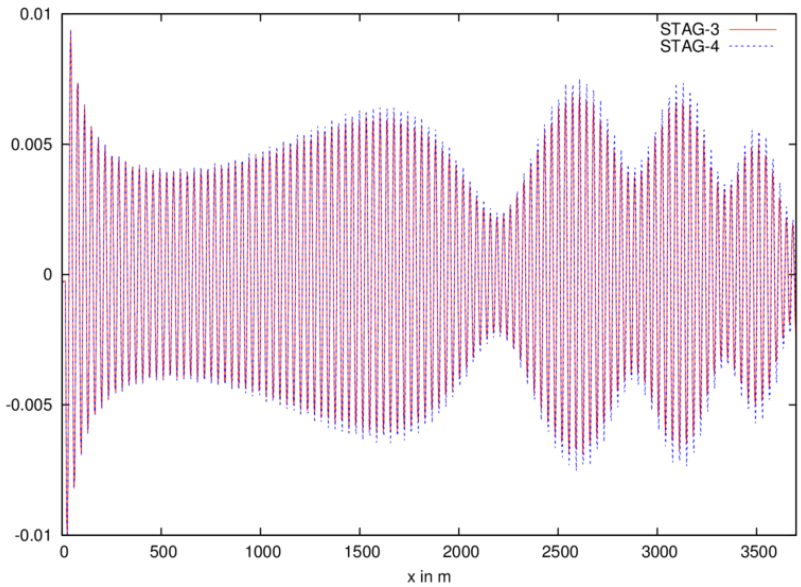


Figure II.19 – Difference between pressure and atmospheric pressure p_{atm} following x at $y = 1$, for the third order scheme, with circa 10 cells per wavelength

sound speed is set in the vertical direction. It writes

$$\begin{cases} c(x,y) &= 343.23 + 0.1y, \\ p_0(x,y) &= p_{\text{atm}}, \\ \rho_0(x,y) &= \gamma \dfrac{p_{\text{atm}}}{c(x,y)^2}, \\ \boldsymbol{u}_0(x,y) &= \mathbf{0}, \\ \gamma &= 1.4. \end{cases} \tag{II.52}$$

Wall boundary conditions are imposed. A harmonic source is placed at point $\boldsymbol{P}_{\text{source}} = (0,5)^t$ and the pressure at this point is set such that $p(\boldsymbol{P}_{\text{source}}, t) = p_{\text{atm}} + \sin(2\pi f t)$ with $f = 10$ Hz. Computations are run until $t = 10$ s. In fig. II.19, the pressure profile is depicted along the line $y = 1$, $x \in [0 : 3700]$ at $t = 10$ s. In fig. II.20, the attenuation in dB of the pressure along the line $y = 1$, $x \in [0 : 3700]$ is depicted. In order to recover a 2D-axisymmetric results, a geometric corrector is applied, which consists in dividing the normalized pressure profile by a factor $\sqrt{r}$, where r is the radius. Result is displayed in fig. II.21 and is in good agreements with the one presented in the literature [8, 39]. Indeed, the staggered schemes require less cells per wavelength (circa 8) compared to cell-centered ones (circa 12) to correctly recover phase and amplitude of the signal.

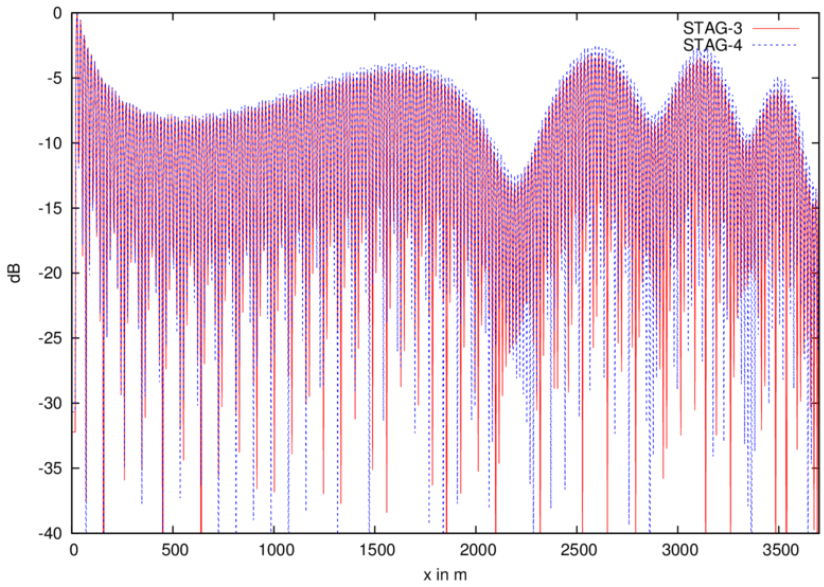


Figure II.20 – Absorption (dB) of the pressure following x at $y = 1$, without rectification, for the third order scheme, with circa 10 cells per wavelength

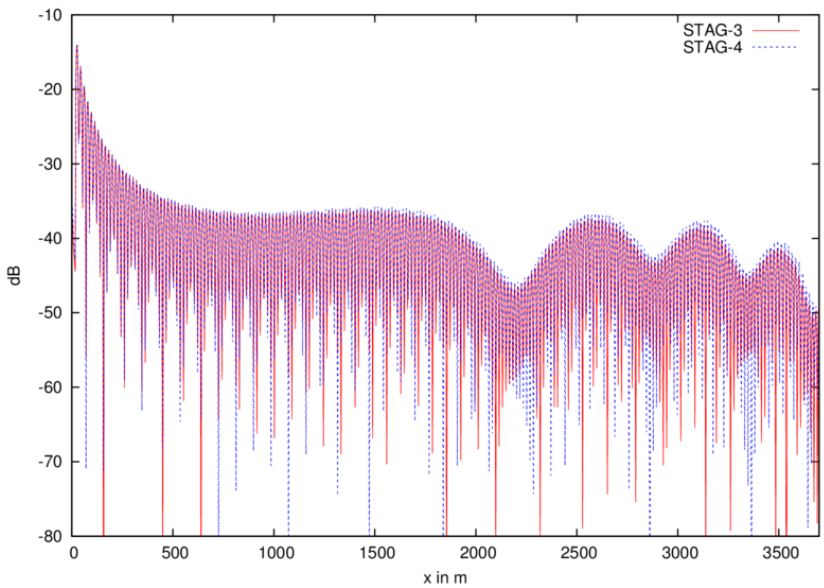


Figure II.21 – Absorption (dB) of the pressure following x at $y = 1$, with geometric corrector, for the third order scheme, with circa 10 cells per wavelength

II-4 Extension to the 2D compressible Navier–Stokes equations with gravity

The compressible Navier–Stokes equations are similar to the Euler equations with an additive viscous stress tensor usually denoted by $\underline{\boldsymbol{\tau}}$. In order to avoid any confusion with the specific volume already denoted τ, it will be denoted by the letter $\underline{\boldsymbol{\Upsilon}}$ in this manuscript. The system of equations in 2D writes in conservative form as

$$\left\{\begin{array}{lcll} \partial_t\rho & + & \nabla\cdot\rho\boldsymbol{u} & = 0, \\ \partial_t\rho\boldsymbol{u} & + & \nabla\cdot(\rho\boldsymbol{u}\otimes\boldsymbol{u}+p\underline{\boldsymbol{I}}-\underline{\boldsymbol{\Upsilon}}) & = 0, \\ \partial_t\rho e & + & \nabla\cdot(((\rho e+p)\underline{\boldsymbol{I}}-\underline{\boldsymbol{\Upsilon}})\cdot\boldsymbol{u}) & = 0, \end{array}\right. \tag{II.53}$$

where $\underline{\boldsymbol{\Upsilon}} = \mu\left(\underline{\boldsymbol{\nabla}}\boldsymbol{u}+(\underline{\boldsymbol{\nabla}}\boldsymbol{u})^t\right)+\lambda\left(\nabla\cdot\boldsymbol{u}\right)\underline{\boldsymbol{I}}$, μ and λ being two parameters which described the viscous properties of the considered fluid. From now on, μ and λ are assumed constant. Adding a constant gravity source-term $\boldsymbol{g}$, it yields

$$\left\{\begin{array}{lcll} \partial_t\rho & + & \nabla\cdot\rho\boldsymbol{u} & = 0, \\ \partial_t\rho\boldsymbol{u} & + & \nabla\cdot(\rho\boldsymbol{u}\otimes\boldsymbol{u}+p\underline{\boldsymbol{I}}-\underline{\boldsymbol{\Upsilon}}) & = \boldsymbol{g}, \\ \partial_t\rho e & + & \nabla\cdot(((\rho e+p)\underline{\boldsymbol{I}}-\underline{\boldsymbol{\Upsilon}})\cdot\boldsymbol{u}) & = \boldsymbol{g}\cdot\boldsymbol{u}. \end{array}\right. \tag{II.54}$$

In the following, a discretization of the viscous terms is proposed on the staggered grids, as well as the discretization of the gravity terms.

II-4.1 Distribution of viscous terms on the modified Arakawa grid

In section II-1.2, an C-type Arakawa grid designed expressly for a diagonal stress tensor has been derived. Due to the presence of the viscous stress tensor, it is necessary to address non-diagonal terms. System without gravity presented in eq. (II.53) rewrites

$$\left\{\begin{array}{llll} \partial_t\rho & +\partial_x\rho u & +\partial_y\rho v & = 0, \\ \partial_t\rho u & +\partial_x(\rho u^2+p-\Upsilon_{1,1}) & +\partial_y(\rho uv-\Upsilon_{2,1}) & = 0, \\ \partial_t\rho v & +\partial_x(\rho uv-\Upsilon_{1,2}) & +\partial_y(\rho v^2+p-\Upsilon_{2,2}) & = 0, \\ \partial_t\rho e & +\partial_x(\rho ue+(p-\Upsilon_{1,1})u-\Upsilon_{1,2}v) & +\partial_y(\rho ve+(p-\Upsilon_{2,2})v-\Upsilon_{2,1}u) & = 0, \end{array}\right. \tag{II.55}$$

where the $\underline{\boldsymbol{\Upsilon}}$ is a symmetric viscous stress tensor which satisfies

$$\left\{\begin{array}{lcl} \Upsilon_{1,1} & = & 2\mu\partial_x u+\lambda\left(\partial_x u+\partial_y v\right), \\ \Upsilon_{2,1} & = & \mu\left(\partial_y u+\partial_x v\right), \\ \Upsilon_{1,2} & = & \Upsilon_{2,1}, \\ \Upsilon_{2,2} & = & 2\mu\partial_y v+\lambda\left(\partial_x u+\partial_y v\right). \end{array}\right. \tag{II.56}$$

II-4.1.1 Space distribution and discretization of the viscosity and gravity terms in 1D

Let address first the discretization of viscous stress and gravitiy terms in one space dimension. The 1D problem ignoring the y-velocity v, and for now the gravity terms writes

$$\left\{\begin{array}{lcll} \partial_t\rho & + & \partial_x\rho u & = & 0, \\ \partial_t\rho u & + & \partial_x(\rho u^2+p-\Upsilon_{1,1}) & = & 0, \\ \partial_t\rho e & + & \partial_x(\rho u e+(p-\Upsilon_{1,1})u) & = & 0, \end{array}\right. \tag{II.57}$$

which rewrites in Lagrangian form as

$$\left\{\begin{array}{lcll} \mathrm{D}_t\rho_0\tau & - & \partial_x u & = & 0, \\ \mathrm{D}_t\rho_0 u & + & \partial_x(p-\Upsilon_{1,1}) & = & 0, \\ \mathrm{D}_t\rho e & + & \partial_x((p-\Upsilon_{1,1})u) & = & 0, \end{array}\right. \tag{II.58}$$

then, using the formulation in both kinetic and internal energies, it yields formally

$$\left\{\begin{array}{lcll} \mathrm{D}_t\rho_0\tau & - & \partial_x u & = & 0, \\ \mathrm{D}_t\rho_0 u & + & \partial_x(p-\Upsilon_{1,1}) & = & 0, \\ \mathrm{D}_t\rho\epsilon & + & (p-\Upsilon_{1,1})\partial_x u & = & 0, \\ \mathrm{D}_t\rho e_{\mathrm{kin}} & + & u\partial_x(p-\Upsilon_{1,1}) & = & 0. \end{array}\right. \tag{II.59}$$

The choice has been made to discretize $\Upsilon_{1,1}$ in the same location as the pressure. It yields that $\Upsilon_{1,1}$ lies on the primal grid. As $\Upsilon_{1,1}=(2\mu+\lambda)\partial_x u$, and as the velocity is staggered, it yields a centered discretization of the space derivative in x of u. Such a discretization is exactly the one obtained by the δ operator defined in the third equation of (II.12).

$$\Upsilon_{1,1i}=(2\mu+\lambda)\frac{1}{\Delta X}\delta u_i.$$

Consider now a uniform gravity field g such that now, eq. (II.59) writes

$$\left\{\begin{array}{lcll} \mathrm{D}_t\rho_0\tau & - & \partial_x u & = & 0, \\ \mathrm{D}_t\rho_0 u & + & \partial_x(p-\Upsilon_{1,1}) & = & g\rho_0, \\ \mathrm{D}_t\rho\epsilon & + & (p-\Upsilon_{1,1})\partial_x u & = & 0, \\ \mathrm{D}_t\rho e_{\mathrm{kin}} & + & u\partial_x(p-\Upsilon_{1,1}) & = & g\rho_0 u. \end{array}\right. \tag{II.60}$$

Integrating in space over a dual cell equations for momentum and kinetic energy leads to

$$\left\{\begin{array}{lcllcl} \mathrm{D}_t\overline{\rho_0 u}_{i+\frac{1}{2}} & = & g\overline{\rho_0}_{i+\frac{1}{2}} & - & ((p-\Upsilon_{1,1})_{i+1}-(p-\Upsilon_{1,1})_i)\,, \\ \mathrm{D}_t\overline{\rho_0 e_{\mathrm{kin}}}_{i+\frac{1}{2}} & = & g\overline{\rho_0 u}_{i+\frac{1}{2}} & - & \frac{1}{\Delta X}\int_{x_i}^{x_{i+1}} u\partial_x(p-\Upsilon_{1,1}). \end{array}\right. \tag{II.61}$$

The formulation in both kinetic and internal energies yields a simple computation for the gravity terms. This is in particular due to the choice to discretize the average density $\overline{\rho_0}$ on both the

primal and dual mesh, initially for robustness issues. Moreover, it does not alter either the internal energy corrector nor the remapping phase. The extension in two dimensions is now discussed.

II-4.1.2 Space distribution and discretization of the viscosity and gravity terms in 2D

The 2D staggered hydrodynamics schemes are based on directional splitting. Here, the choice of splitting, mainly due to memory alignment is the following for the x-direction

$$\begin{cases} \partial_t \rho \quad +\partial_x(\rho u) & =0, \\ \partial_t(\rho u)+\partial_x(\rho u^2 + p - \Upsilon_{1,1}) & =0, \\ \partial_t(\rho v)+\partial_x(\rho uv - \Upsilon_{2,1}) & =0, \\ \partial_t(\rho e)+\partial_x(\rho ue + (p - \Upsilon_{1,1})u - \Upsilon_{2,1}v) & =0, \end{cases} \tag{II.62}$$

and in the y-direction

$$\begin{cases} \partial_t \rho \quad +\partial_y(\rho v) & =0, \\ \partial_t(\rho u)+\partial_y(\rho uv - \Upsilon_{1,2}) & =0, \\ \partial_t(\rho v)+\partial_y(\rho v^2 + p - \Upsilon_{2,2}) & =0, \\ \partial_t(\rho e)+\partial_y(\rho ve + (p - \Upsilon_{2,2})v - \Upsilon_{1,2}u) & =0. \end{cases} \tag{II.63}$$

As aforementioned, the term $\Upsilon_{1,1}$ is discretized on the same position as the pressure, meaning at the center of each primal cell. Symmetrically, it also holds for $\Upsilon_{2,2}$. Consider now eq. (II.62) which formally writes in Lagrangian form

$$\begin{cases} \mathrm{D}_t(\rho_0 \tau)+\partial_X u & =0, \\ \mathrm{D}_t(\rho_0 u)+\partial_X(p - \Upsilon_{1,1}) & =0, \\ \mathrm{D}_t(\rho_0 v)+\partial_X(-\Upsilon_{2,1}) & =0, \\ \mathrm{D}_t(\rho_0 e)+\partial_X((p - \Upsilon_{1,1})u - \Upsilon_{2,1}v) & =0. \end{cases} \tag{II.64}$$

Reminding that $\Upsilon_{1,1} = (2\mu + \lambda)\partial_x u + \lambda \partial_y v$, the choice has been made to discretize $\partial_x u$ and $\partial_y v$ at each cell centers. Since u and v are staggered respectively in the x- and y-directions, centered discretizations of space derivatives give the desired results. Once again, the use of the δ operator yields high-order accuracy in space for the terms $\Upsilon_{1,1}$ and $\Upsilon_{2,2}$. Furthermore, reminding that the momentum $\rho_0 v$ lies on the third grid, and is formally indexed $\rho_0 v_{i,j+\frac{1}{2}}$ and integrating over a dual cells in the x-direction it yields

$$\mathrm{D}_t \overline{\rho_0 v}_{i,j+\frac{1}{2}} = \frac{1}{\Delta X} \Upsilon_{2,1_{i+\frac{1}{2},j+\frac{1}{2}}} - \Upsilon_{2,1_{i-\frac{1}{2},j+\frac{1}{2}}}. \tag{II.65}$$

The choice has been made to discretize the non-diagonal terms of the viscous stress tensor on a grid staggered in both directions. Similar analysis performed on eq. (II.63) gives the same results for $\Upsilon_{1,2}$. Reminding that $\Upsilon_{1,2} = \Upsilon_{2,1} = \mu\left(\partial_y u + \partial_x v\right)$, it remains to discretize the terms

$\partial_y u$ and $\partial_x v$. Since u and v are respectively staggered in the x-direction and in the y-directionn considering centered approximations of the derivatives naturally leads to approximations of $\partial_y u$ and $\partial_x v$ staggered in both directions as expected. Then, one can use the previsously introduced δ operator. It yields high-order accuracy in space for the terms $\Upsilon_{2,1}$ and $\Upsilon_{2,2}$. Finally, using the δ operator, we have

$$\left\{\begin{array}{lcl} \Upsilon_{1,1_{i,j}} & = & \frac{2\mu+\lambda}{\Delta X}\delta_x u_{i,j} + \frac{\lambda}{\Delta Y}\delta_y v_{i,j}, \\ \Upsilon_{2,1_{i+\frac{1}{2},j+\frac{1}{2}}} & = & \mu\left(\frac{1}{\Delta Y}\delta_y u_{i+\frac{1}{2},j+\frac{1}{2}} + \frac{1}{\Delta X}\delta_x v_{i+\frac{1}{2},j+\frac{1}{2}}\right), \\ \Upsilon_{1,2_{i+\frac{1}{2},j+\frac{1}{2}}} & = & \mu\left(\frac{1}{\Delta Y}\delta_y u_{i+\frac{1}{2},j+\frac{1}{2}} + \frac{1}{\Delta X}\delta_x v_{i+\frac{1}{2},j+\frac{1}{2}}\right), \\ \Upsilon_{2,2_{i,j}} & = & \frac{2\mu+\lambda}{\Delta Y}\delta_y v_{i,j} + \frac{\lambda}{\Delta X}\delta_x u_{i,j}. \end{array}\right. \tag{II.66}$$

That way, a natural distribution of the viscous terms is summarized in fig. II.22. This discretization holds for non-symmetric tensor $\underline{\boldsymbol{\Upsilon}}$.

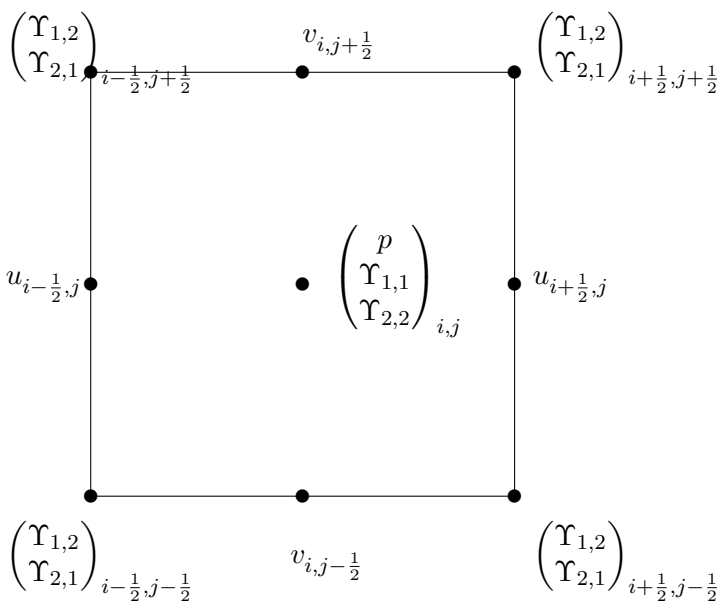


Figure II.22 – Arakawa C-type like grid for the compressible Navier–Stokes equation with a special distribution for the non-diagonal viscous terms

The gravity terms are not explicited here, as this is very similar to the 1D case considering a constant gravity field $\boldsymbol{g} = (g_x, g_y)^t$.

II-4.2 2D viscous staggered Lagrange-Remap schemes with gravity force

First, the 1D staggered scheme is derived using explicit Runge–Kutta time-integration. Then the extension to the multidimensional case is detailed using directional splitting. Gravity terms are then introduced.

II-4.2.1 1D staggered Lagrange-Remap scheme to the compressible Navier-Stokes equations

Consider the 1D compressible Navier–Stokes equation in Lagrangian coordinates as depicted in eq. (II.64). The total energy is then split into internal and kinetic energies. It formally yields

$$\begin{cases} \mathrm{D}_t(\rho_0\tau) & - & \partial_X u & =0, \\ \mathrm{D}_t(\rho_0 u) & + & \partial_X(p-\Upsilon_{1,1}) & =0, \\ \mathrm{D}_t(\rho_0 v) & + & \partial_X(-\Upsilon_{2,1}) & =0, \\ \mathrm{D}_t(\rho_0\epsilon) & + & (p-\Upsilon_{1,1})\partial_X u - \Upsilon_{2,1}\partial_X v & =0, \\ \mathrm{D}_t(\rho_0 e_{\mathrm{kin},u}) & + & u\partial_X(p-\Upsilon_{1,1}) & =0, \\ \mathrm{D}_t(\rho_0 e_{\mathrm{kin},v}) & + & v\partial_X(-\Upsilon_{2,1}) & =0. \end{cases} \tag{II.67}$$

The intermediate steps for the staggered scheme write for the compressible Navier–Stokes

$$\begin{cases} \overline{\rho_0\tau}^{n+\alpha_m}_{i,j} = \overline{\rho_0\tau}^{n}_{i,j} + \frac{\Delta t}{\Delta X}\sum_{l=0}^{m-1} a_{m,l} du^{n+\alpha_l}_{i,j}, \\ \overline{\rho_0 u}^{n+\alpha_m}_{i+\frac{1}{2},j} = \overline{\rho_0 u}^{n}_{i+\frac{1}{2},j} - \frac{\Delta t}{\Delta X}\sum_{l=0}^{m-1} a_{m,l}(dp - d\Upsilon_{1,1})^{n+\alpha_l}_{i+\frac{1}{2},j}, \\ \overline{\rho_0 v}^{n+\alpha_m}_{i,j+\frac{1}{2}} = \overline{\rho_0 v}^{n}_{i,j+\frac{1}{2}} - \frac{\Delta t}{\Delta X}\sum_{l=0}^{m-1} a_{m,l}(-d\Upsilon_{2,1})^{n+\alpha_l}_{i,j+\frac{1}{2}}, \\ \overline{\rho_0\epsilon}^{n+\alpha_m}_{i,j} = \overline{\rho_0\epsilon}^{n}_{i,j} - \frac{\Delta t}{\Delta X}\sum_{l=0}^{m-1} a_{m,l}\overline{(p-\Upsilon_{1,1})\delta u}^{n+\alpha_l}_{i,j} + \overline{(-\Upsilon_{2,1})\delta v}^{n+\alpha_l}_{i,j}, \\ p^{n+\alpha_m}_{i,j} = EOS(\tau^{n+\alpha_m}_{i,j}, \epsilon^{n+\alpha_m}_{i,j}), \end{cases} \tag{II.68}$$

and the final step writes

$$
\begin{cases}
\overline{\rho_0\tau}_{i,j}^{n+1} & = \overline{\rho_0\tau}_{i,j}^{n} + \frac{\Delta t}{\Delta X}\sum_{l=0}^{s-1}\theta_l du_{i,j}^{n+\alpha_l}, \\
\overline{\rho_0 u}_{i+\frac{1}{2},j}^{n+1} & = \overline{\rho_0 u}_{i+\frac{1}{2},j}^{n} - \frac{\Delta t}{\Delta X}\sum_{l=0}^{s-1}\theta_l (dp - d\Upsilon_{1,1})_{i+\frac{1}{2},j}^{n+\alpha_l}, \\
\overline{\rho_0 v}_{i,j+\frac{1}{2}}^{n+1} & = \overline{\rho_0 v}_{i,j+\frac{1}{2}}^{n} - \frac{\Delta t}{\Delta X}\sum_{l=0}^{s-1}\theta_l (-d\Upsilon_{2,1})_{i,j+\frac{1}{2}}^{n+\alpha_l}, \\
\overline{\rho_0 \epsilon}_{i,j}^{n+1} & = \overline{\rho_0 \epsilon}_{i,j}^{n} - \frac{\Delta t}{\Delta X}\sum_{l=0}^{s-1}\theta_l \overline{(p-\Upsilon_{1,1})\delta u}_{i,j}^{n+\alpha_l} + \overline{(-\Upsilon_{2,1})\delta v}_{i,j}^{n+\alpha_l}, \\
\overline{\rho_0 e_{\mathrm{kin},u}}_{i+\frac{1}{2},j}^{n+1} & = \overline{\rho_0 e_{\mathrm{kin},u}}_{i+\frac{1}{2},j}^{n} - \frac{\Delta t}{\Delta X}\sum_{l=0}^{s-1}\theta_l \overline{u\delta(p-\Upsilon_{1,1})}_{i+\frac{1}{2},j}^{n+\alpha_l}, \\
\overline{\rho_0 e_{\mathrm{kin},v}}_{i,j+\frac{1}{2}}^{n+1} & = \overline{\rho_0 e_{\mathrm{kin},v}}_{i,j+\frac{1}{2}}^{n} - \frac{\Delta t}{\Delta X}\sum_{l=0}^{s-1}\theta_l \overline{v\delta(-\Upsilon_{2,1})}_{i,j+\frac{1}{2}}^{n+\alpha_l}, \\
x_{i+\frac{1}{2}}^{n+1} & = x_{i+\frac{1}{2}}^{n} + \Delta t\sum_{l=0}^{s-1}\theta_l u_{i+\frac{1}{2}}^{n+\alpha_l}, \\
p_i^{n+1} & = EOS(\tau_i^{n+1}, \epsilon_i^{n+1}).
\end{cases}
\tag{II.69}
$$

As for the 1D Euler staggered schemes, the kinetic energies need only to be updated at the end of the Lagrangian phase. Conservation properties of the staggered schemes for the compressible Navier–Stokes formulated in both internal and kinetic energies are summarized in the following lemma.

Lemma II.13 (Conservation of the staggered schemes (II.68)-(II.69))**.** *For all explicit Runge-Kutta sequences and all consistent spatial reconstructions, the schemes* (II.68)-(II.69) *are conservative in mass, momentum and total energy $\mathcal{E}$ definition II.2.*

Proof. The proof is identical to the one for (II.13)-(II.14) schemes. ■

As for the 1D Euler scheme, the scheme does not conserve the total energy E. The idea is to recouple E and $\mathcal{E}$ using the internal energy corrector proposed in eq. (II.21). It leads to the following lemma.

Lemma II.14 (Conservation of the staggered schemes (II.68)-(II.69)-(II.21))**.** *For all explicit Runge-Kutta sequences and all spatial reconstructions, the schemes* (II.68)-(II.69)-(II.21) *are conservative in mass, momentum and total energy E (see definition II.1).*

Proof. The proof is straightforward using lemmas II.7 and II.13. ■

The remapping stage is identical to the one for the 1D Euler staggered schemes. Once again in practice, the Lagrangian phase is performed, then quantities are remapped and at last the internal energy corrector is applied.

II-4.2.2 2D Extension of the 1D staggered Lagrange-remap schemes

Equations (II.62) and (II.63) can be rewritten under a similar form as in eq. (II.41)

$$\partial_t \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ \rho e \end{pmatrix} + \mathcal{B}_1 \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ \rho e \end{pmatrix} = \mathbf{0}, \qquad \partial_t \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ \rho e \end{pmatrix} + \mathcal{B}_2 \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ \rho e \end{pmatrix} = \mathbf{0}. \tag{II.70}$$

Splitting techniques relies on solving alternatively first and second equation of eq. (II.70) with weighted time-step in order to reach high-order accuracy. This procedure is identical as for the Euler equations.

Lemma II.15 (Conservation of the 2D schemes (II.68)-(II.69)-(II.21)-(II.28))**.** *The resulting 2D Cartesian grid schemes for the compressible Navier–Stokes equations are conservative in mass, momentum and total energy E (see definition II.1).*

Proof. With the proposed C-type staggering of variables, the 2D schemes satisfy lemmas II.10 and II.14 direction by direction and are therefore globally conservative in mass, momentum and total energy for any dimensional splitting sequence. ■

II-4.2.3 Gravity source terms integration

In this part, the 2D schemes with gravity source terms are proposed. There is no special modifications for the gravity source terms integration compared to the 1D case. Consider a constant gravity field $\boldsymbol{g} = (g_x, g_y)^t$. Then the proposed integration of gravity source terms writes in the x-direction as

$$\begin{cases} \overline{\rho_0 \tau}_{i,j}^{n+\alpha_m} = \overline{\rho_0 \tau}_{i,j}^{n} + \frac{\Delta t}{\Delta X} \sum_{l=0}^{m-1} a_{m,l} du_{i,j}^{n+\alpha_l}, \\ \overline{\rho_0 u}_{i+\frac{1}{2},j}^{n+\alpha_m} = \overline{\rho_0 u}_{i+\frac{1}{2},j}^{n} - \frac{\Delta t}{\Delta X} \sum_{l=0}^{m-1} a_{m,l} (dp - d\Upsilon_{1,1})_{i+\frac{1}{2},j}^{n+\alpha_l} + \alpha_m \Delta t g_x \overline{\rho_0}_{i+\frac{1}{2},j}^{n}, \\ \overline{\rho_0 v}_{i,j+\frac{1}{2}}^{n+\alpha_m} = \overline{\rho_0 v}_{i,j+\frac{1}{2}}^{n} - \frac{\Delta t}{\Delta X} \sum_{l=0}^{m-1} a_{m,l} (-d\Upsilon_{2,1})_{i,j+\frac{1}{2}}^{n+\alpha_l}, \\ \overline{\rho_0 \epsilon}_{i,j}^{n+\alpha_m} = \overline{\rho_0 \epsilon}_{i,j}^{n} - \frac{\Delta t}{\Delta X} \sum_{l=0}^{m-1} a_{m,l} \overline{(p - \Upsilon_{1,1}) \delta u}_{i,j}^{n+\alpha_l} + \overline{(-\Upsilon_{2,1}) \delta v}_{i,j}^{n+\alpha_l}, \\ p_{i,j}^{n+\alpha_m} = EOS(\tau_{i,j}^{n+\alpha_m}, \epsilon_{i,j}^{n+\alpha_m}), \end{cases} \tag{II.71}$$

$$
\begin{cases}
\overline{\rho_0\tau}_{i,j}^{n+1} & = \overline{\rho_0\tau}_{i,j}^{n} + \frac{\Delta t}{\Delta X}\sum_{l=0}^{s-1}\theta_l du_{i,j}^{n+\alpha_l}, \\
\overline{\rho_0 u}_{i+\frac{1}{2},j}^{n+1} & = \overline{\rho_0 u}_{i+\frac{1}{2},j}^{n} - \frac{\Delta t}{\Delta X}\sum_{l=0}^{s-1}\theta_l (dp - d\Upsilon_{1,1})_{i+\frac{1}{2},j}^{n+\alpha_l} + \Delta t g_x \overline{\rho_0}_{i+\frac{1}{2},j}^{n}, \\
\overline{\rho_0 v}_{i,j+\frac{1}{2}}^{n+1} & = \overline{\rho_0 v}_{i,j+\frac{1}{2}}^{n} - \frac{\Delta t}{\Delta X}\sum_{l=0}^{s-1}\theta_l (-d\Upsilon_{2,1})_{i,j+\frac{1}{2}}^{n+\alpha_l}, \\
\overline{\rho_0\epsilon}_{i,j}^{n+1} & = \overline{\rho_0\epsilon}_{i,j}^{n} - \frac{\Delta t}{\Delta X}\sum_{l=0}^{s-1}\theta_l \overline{(p-\Upsilon_{1,1})\delta u}_{i,j}^{n+\alpha_l} + \overline{(-\Upsilon_{2,1})\delta v}_{i,j}^{n+\alpha_l}, \\
\overline{\rho_0 e_{\text{kin},u}}_{i+\frac{1}{2},j}^{n+1} & = \overline{\rho_0 e_{\text{kin},u}}_{i+\frac{1}{2},j}^{n} - \frac{\Delta t}{\Delta X}\sum_{l=0}^{s-1}\theta_l \left(\overline{u\delta(p-\Upsilon_{1,1})}_{i+\frac{1}{2},j}^{n+\alpha_l} + g_x \overline{\rho_0 u}_{i+\frac{1}{2},j}^{n+\alpha_l}\right), \\
\overline{\rho_0 e_{\text{kin},v}}_{i,j+\frac{1}{2}}^{n+1} & = \overline{\rho_0 e_{\text{kin},v}}_{i,j+\frac{1}{2}}^{n} - \frac{\Delta t}{\Delta X}\sum_{l=0}^{s-1}\theta_l \overline{v\delta(-\Upsilon_{2,1})}_{i,j+\frac{1}{2}}^{n+\alpha_l}, \\
x_{i+\frac{1}{2}}^{n+1} & = x_{i+\frac{1}{2}}^{n} + \Delta t\sum_{l=0}^{s-1}\theta_l u_{i+\frac{1}{2}}^{n+\alpha_l}, \\
p_i^{n+1} & = EOS(\tau_i^{n+1}, \epsilon_i^{n+1}).
\end{cases}
\tag{II.72}
$$

II-4.3 Numerical validation of the 2D staggered Lagrange-Remap schemes

Three test-cases are proposed to assess the accuracy and robustness of the 2D staggered schemes for the compressible Navier–Stokes equations. The first test-case is in 1D, with no viscous terms, which assesses schemes ability to recover hydrostatic equilibrium. Then, a 2D test-case without gravity forces, the Taylor–Green vortex, is presented. Last, a Rayleigh–Taylor instability is studied with and without viscous terms.

II-4.3.1 1D atmosphere at rest [92]

This test-case has been proposed in [92] by Mishra and Kappeli. It consists of a hydrostatic equilibrium between pressure and gravitational forces. Initial conditions are

$$
\begin{cases}
\rho_0(x) & = & \left(\rho_0^{\gamma-1} + K_0\dfrac{\gamma-1}{\gamma}gx\right)^{\frac{1}{\gamma-1}}, \\
p_0(x) & = & K_0\rho_0(x)^{\gamma}, \\
u_0(x) & = & 0, \\
\gamma & = & \frac{5}{3},
\end{cases}
\tag{II.73}
$$

with here $K_0 = \frac{1}{\rho_0}^{\gamma}$, $\rho_0 = 1$, $g = -1$. As the proposed schemes are not well-balanced, it challengingly assesses the ability of the schemes to recover hydrostatic equilibrium as well as to see if waves induced by numerical errors are amplified or dumped by the schemes. In table II.13, the l^1 error in density is displayed for the staggered schemes. The third order scheme reaches

machine-precision – and so hydrostatic equilibrium – using approximately 560 cells. Fourth and fifth order schemes reach hydrostatic equilibrium at about 140 cells, and the higher-order schemes have already reached hydrostatic equilibrium with only 35 cells. In practice, it means that for such a problem, high-order accuracy is able to recover the smooth hydrostatic equilibrium up to a relatively small number of cells.

N_x	STAG-3		STAG-4		STAG-5		STAG-6		STAG-7		STAG-8	
35	2.2e-9	·	2.0e-11	·	2.0e-11	·	7.3e-13	·	7.2e-13	·	2.8e-14	·
70	1.5e-10	3.88	3.2e-13	6.0	3.2e-13	6.0	5.1e-14	$\star$	5.0e-14	$\star$	1.1e-14	$\star$
140	9.8e-12	3.93	1.5e-14	$\star$	1.9e-14	$\star$	6.1e-14	$\star$	6.4e-14	$\star$	5.0e-14	$\star$
280	6.2e-13	3.98	2.8e-14	$\star$	6.1e-14	$\star$	7.2e-14	$\star$	8.1e-14	$\star$	7.4e-14	$\star$
560	1.1e-14	5.77	6.8e-14	$\star$	9.1e-14	$\star$	6.7e-14	$\star$	9.6e-14	$\star$	1.1e-13	$\star$

Table II.13 – l^1-error in density and experimental order of convergence for the Lagrange-remap staggered scheme with gravity forces taken on the atmosphere at hydrostatic equilibrium [92], until $t = 20$, CFL=0.7. $\star$ indicates machine precision reached.

II-4.3.2 Taylor–Green vortex [160]

The Taylor–Green vortex is used to assess the accuracy of the proposed schemes. It is usually studied by the incompressible Navier–Stokes community. Here, enforcing a very high sound speed, the compressible Navier–Stokes equations are in near incompressible regime.

$$\begin{cases} \rho_0(x,y) &= 1, \\ u_0(x,y) &= \sin(x)\cos(y), \\ v_0(x,y) &= \cos(x)\sin(y), \\ p_0(x,y) &= p_0 - \frac{1}{4}\left(\cos(2x) + \sin(2y)\right). \end{cases} \tag{II.74}$$

The analytical solution for incompressible flows writes

$$\begin{cases} \rho(x,y,t) &= 1, \\ u(x,y,t) &= \sin(x)\cos(y)\,\mathrm{e}^{-2\mu t}, \\ v(x,y,t) &= \cos(x)\sin(y)\,\mathrm{e}^{-2\mu t}, \\ p(x,y,t) &= p_0 - \frac{1}{4}\left(\cos(2x) + \sin(2y)\right)\mathrm{e}^{-4\mu t}, \end{cases} \tag{II.75}$$

with $p_0 = 10$. The pressure is set such that the regime is nearly incompressible, using a stiffened gas EOS which writes

$$p = (\gamma - 1)\rho\epsilon - \gamma p^\star.$$

Here $p^\star = 10^8$. The viscosity parameters are set to $\mu = 10$, $\lambda = 0$. Computations are performed till $t = 10^{-3}$ with a CFL set to 0.9 on the computational domain $\Omega = [-\pi, \pi]^2$. Periodic boundary conditions are imposed. The limitation on the final time is due to the use of explicit Runge–Kutta sequences combined with the very high sound speed number. l^1-error in momentum as well as experimental order of convergence are presented in table II.14. Machine precision is reached quickly on the every variables due to the large difference existing between the numerical

values of momentum, density, pressure with the values of internal energy. Indeed the error are not taken as relative errors but as absolute ones. Magnitude differs by a factor 10^8. Hence, for relative errors, one should divide by at least 10^8. We believe double precision is not sufficient to reach smaller absolute error.

N_x	STAG-3		STAG-4		STAG-5		STAG-6		STAG-7		STAG-8	
10	5.0e-1	·	1.8e-4	·	3.1e-3	·	1.5e-4	·	2.4e-4	·	6.1e-5	·
20	7.6e-2	3.88	1.2e-5	6.0	1.0e-4	6.0	1.1e-5	$\star$	1.1e-5	$\star$	1.1e-5	$\star$
40	1.0e-2	3.93	1.1e-5	$\star$	1.2e-5	$\star$	1.3e-5	$\star$	1.3e-5	$\star$	1.3e-5	$\star$
80	1.4e-3	3.98	1.3e-5	$\star$	1.1e-5	$\star$	1.2e-5	$\star$	1.2e-5	$\star$	1.2e-5	$\star$
160	2.2e-4	5.77	1.4e-5	$\star$	1.4e-5	$\star$	1.3e-5	$\star$	1.4e-5	$\star$	1.2e-5	$\star$
320	3.1e-5	2.87	1.3e-5	$\star$	1.5e-5	$\star$	1.4e-5	$\star$	1.6e-5	$\star$	1.4e-5	$\star$

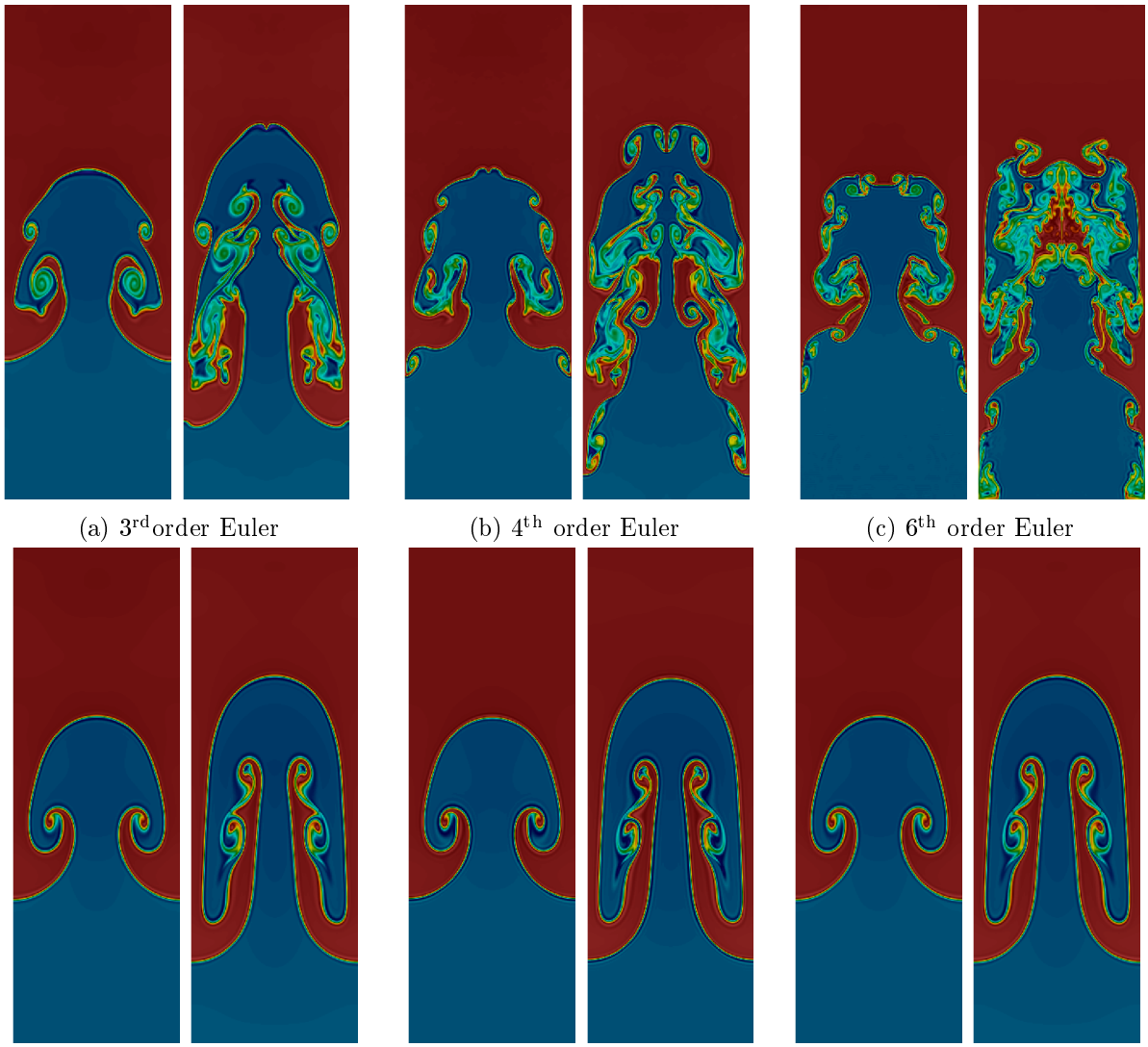
Table II.14 – l^1-error in density and experimental order of convergence for the compressible Navier–Stokes Lagrange-remap staggered scheme for the Taylor–Green vortex [160], until $t = 2.10^{-3}$, CFL=0.9. Machine precision is reduced to 10^{-5} as error are taken in absolute. For relative errors, one should divide by 10^8. $\star$ indicates machine precision reached.

II-4.3.3 Rayleigh–Taylor instability [151, 159, 108]

The Rayleigh–Taylor instability is used to assess the ability of the schemes to handle instability, and if those instabilities are accentuated by the high-order accuracy. The initial data for the single perturbation mode are

$$\begin{cases} \rho_0(x,y) &= 2\chi_{\{y>0\}} + 1\chi_{\{y<0\}}, \\ u_0(x,y) &= 0, \\ v_0(x,y) &= 0.25a(1+\cos(4\pi x))(1+\cos(3\pi y))\chi\{|y|<1/6\}, \\ p_0(x,y) &= K_0 + \rho_0(x,y)gy, \end{cases} \tag{II.76}$$

where $g = -0.1$, $K_0 = 2.5$, $a = 10^{-2}$. The viscous parameters are chosen very small with $\mu = 10^{-4}$ and $\lambda = -\frac{2}{3}\mu$. In order to highlight the role of viscosity, computations are run first with the Euler schemes and then with the Compressible Navier–Stokes (CNS) schemes. Periodic boundary conditions are set on the left and right boundaries, whereas wall boundary conditions are imposed on the top and bottom boundaries. The computation domain is set to $[-0.25 : 0.25] \times [-0.75 : 0.75]$. Since the hydrostatic equilibrium is not perfectly recovered, additional noise is added, but still small compared to the perturbations inducing the instability. Results are depicted in fig. II.23. Without viscous stress tensor, the higher the order, the more modes develop. As a contrary, using even a small coefficient of viscosity prevents such modes from developing, and leads to the expected results. Without dissipation, Euler schemes are unable to recover correctly the Rayleigh–Taylor expected profiles, and do not seem to converge. This is not a new result since it has been highlighted among others in [108].



(a) 3rd order Euler (b) 4th order Euler (c) 6th order Euler

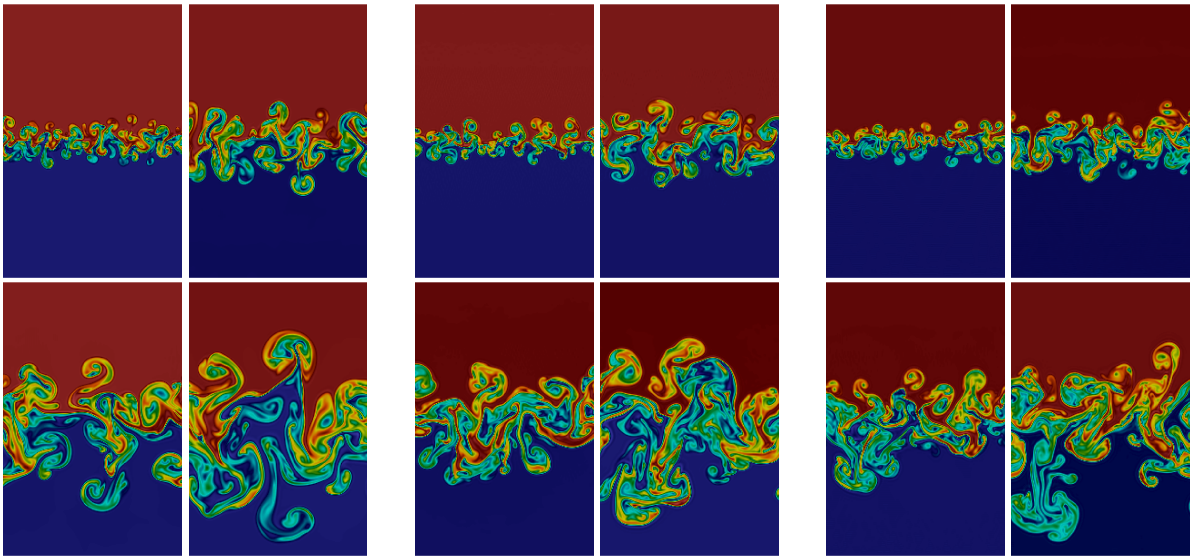
(d) 3rd order CNS (e) 4th order CNS (f) 6th order CNS

Figure II.23 – Density profiles on the Rayleigh–Taylor mono-mode instability for the Euler equations (top) and for the Compressible Navier–Stokes (CNS) equations with $\mu = 10^{-4}$ and $\lambda = -\frac{2}{3}\mu$ (bottom) using third, fourth and sixth order schemes, at time $t = 9.5$ (left) and $t = 12.75$ (right) with 200 cells in the x-direction and 600 in the y-direction.

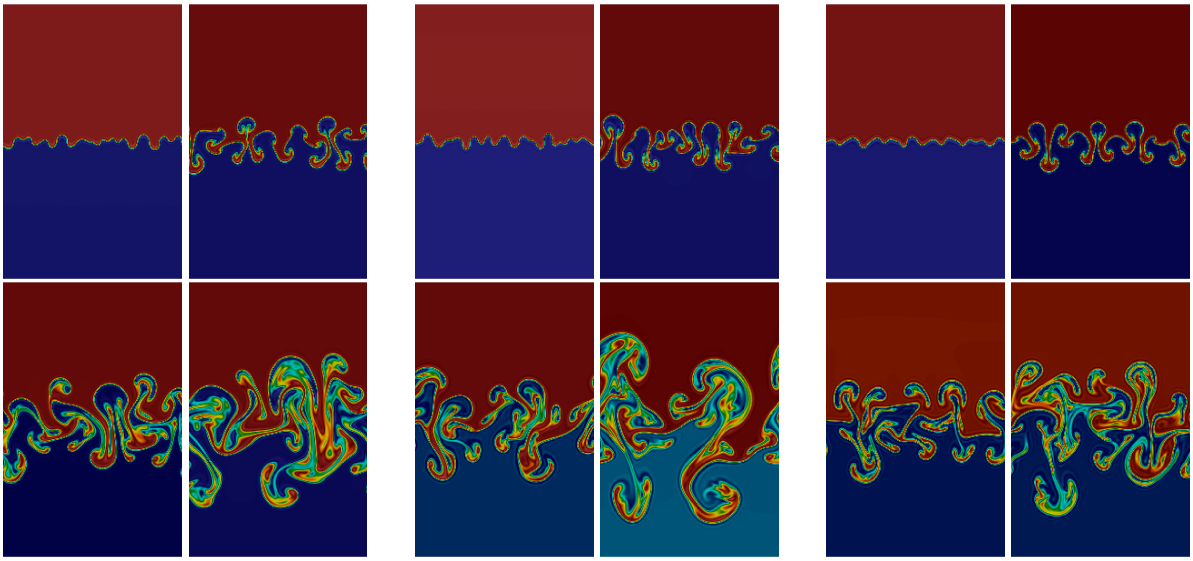
For the multi-mode perturbation, the initialization is slightly modified as

$$\begin{cases} \rho_0(x,y) &= 2\chi_{\{y>0\}} + 1\chi_{\{y<0\}}, \\ u_0(x,y) &= 0, \\ v_0(x,y) &= A(x)(1+\cos(3\pi y))\chi\{|y|<1/6\}, \\ p_0(x,y) &= K_0 + \rho_0(x,y)gy, \end{cases} \tag{II.77}$$

where $A(x)$ is chosen as a random number belonging to $\left[0 : 10^{-2}\right]$. The parameters are left unchanged. The computation domain is set to $[-0.25 : 0.25] \times [-0.375 : 0.375]$. Results are depicted in fig. II.24.



(a) 3rd order Euler (b) 4th order Euler (c) 6th order Euler



(d) 3rd order CNS (e) 4th order CNS (f) 6th order CNS

Figure II.24 – Density profiles on the Rayleigh–Taylor multi-mode instability for the Euler equations (top) and for the Compressible Navier–Stokes (CNS) equations with $\mu = 10^{-4}$ and $\lambda = -\frac{2}{3}\mu$ (bottom) using third, fourth and sixth order schemes, at time $t = 6$, $t = 9$, $t = 12$, $t = 15$ from left to right and top to bottom, with 200 cells in the x-direction and 300 in the y-direction

Chapter III

Stable high-order methods for linear hyperbolic systems with arbitrary boundary conditions

---

L'étude d'une nouvelle famille de schémas numériques pour des systèmes linéaires hyperboliques avec conditions aux bords est réalisée au cours de ce chapitre. On présente dans un premier temps la procédure afin de construire les opérateurs d'intégration des conditions aux bords dans le cas de l'équation de l'advection pour des approximations de type différences finies et volumes finis. Ensuite, cette procédure est étendue au cas du système des équations des ondes avec deux conditions aux bords différentes. La méthode est alors étendue au cas général des systèmes hyperboliques linéaires avec conditions aux bords. Afin de pouvoir caractériser la stabilité des schémas ainsi obtenus par l'ajout de ces opérateurs, une étude de type GKS est proposée. Afin de permettre de disposer d'un aperçu de la stabilité du schéma effectif, une définition de stabilité dite réduite est introduite. Des résultats numériques sont proposés tout au long du chapitre afin d'illustrer la précision ainsi que la pertinence de la définition de stabilité réduite introduite. Une partie de ce travail a été soumise à une revue scientifique [34].

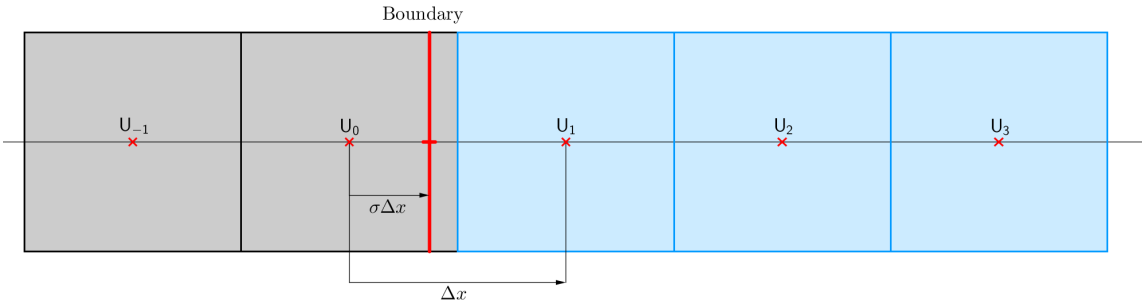


Figure III.1 – 1D Boundary between outside and inside computational domain

In this chapter, a way to impose boundary conditions building ghost-cells values for linear conservation laws is proposed and studied. It is very close to the method developed in [155, 156, 158, 157, 168]. In order to deal with the discretization of boundary conditions in the special case of Lagrange-remap schemes, the case of a simplified linear 1D hyperbolic system of conservation laws on $\Omega = \{x \in \mathbb{R}, x > x_s\}$ is studied as

$$\begin{cases} \partial_t \boldsymbol{U} + \underline{\boldsymbol{A}} \partial_x \boldsymbol{U} &= 0, & t > 0,\ x > x_s,\ \boldsymbol{U}(x,t) \in \mathbb{R}^p \\ \underline{\boldsymbol{B}} \boldsymbol{U}(x_s,t) &= \underline{\boldsymbol{B}} \boldsymbol{G}(t), & t > 0, \\ \boldsymbol{U}(x,0) &= \boldsymbol{U_0}(x), & x > x_s. \end{cases} \tag{III.1}$$

The geometry is depicted in fig. III.1. Put aside temporarily the peculiar shape of $\Omega = [x_s, \infty[$ and consider the whole domain. The 1D domain is discretized in regular cells $I_j = [x_{j-\frac{1}{2}},\ x_{j+\frac{1}{2}}]$, with $\Delta x = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}$ the constant space between two consecutive cell interfaces. Only finite-differences and finite-volume-type schemes will be considered (see section I-1.2.1). At time t^n, the discrete solution writes $\boldsymbol{\mathcal{U}}^n = (\boldsymbol{U}_j^n)_{j\in\mathbb{Z}}$. Consider now that $\Omega = \{x \in \mathbb{R}, x > x_s\}$ and that boundary conditions are specified at $x = x_s = \sigma\Delta x \in I_0$ with $\sigma \in [-\frac{1}{2}, \frac{1}{2}[$. The most interesting case is when the boundary location does not coincide with the discretized grid (see fig. III.1).

Only fully interior cells (depicted in blue in the figure) are considered to be part of the "interior" computational domain denoted $\Omega_+ \subset \Omega$. Cells in gray are considered as part of the "ghost" computational domain denoted Ω_-. In practice, one has $\Omega \subset \Omega_+ \cup \Omega_- \subset \mathbb{R}$. The algorithm proposed in this work builds ghost values in Ω_- such that the resulting scheme is both high-order accurate and stable. For this peculiar value of x_s, one has that $\Omega_+ = \{x \in \mathbb{R}, x \geq x_{\frac{1}{2}}\}$. It implies that only interior values $\boldsymbol{\mathcal{U}}_+ = (\boldsymbol{U}_j^n)_{j\geq 1}$ are known at the beginning of the time-step. Boundary conditions specified at $x = x_s$ are provided according to the incoming/outgoing characteristics of $\underline{\boldsymbol{A}} = \nabla_U \boldsymbol{F}(\boldsymbol{U})$. Moreover, the matrix $\underline{\boldsymbol{B}}$ satisfies the condition of theorem I.9. In the whole chapter, the matrix $\underline{\boldsymbol{A}}$ is assumed invertible to alleviate computations. To build ghost values, which is ultimately the real problem, one has in hands the boundary conditions and any kind of extrapolation technique to reconstruct $\boldsymbol{\mathcal{U}}_- = (\boldsymbol{U}_j)_{j\leq 0}$ from $\boldsymbol{\mathcal{U}}_+ = (\boldsymbol{U}_j)_{j\geq 1}$. Therefore the problem discussed hereafter can be formulated as follows

Problem III.1. Build an operator $\underline{\boldsymbol{\mathcal{R}}}$

$$\begin{aligned} \underline{\boldsymbol{\mathcal{R}}}: \quad (\mathbb{R}^p)^{\mathbf{card}(\Omega_+)} &\to (\mathbb{R}^p)^{\mathbf{card}(\Omega_-)} \\ \underline{\boldsymbol{\mathcal{R}}}(\boldsymbol{\mathcal{U}}_+) &= \boldsymbol{\mathcal{U}}_- \end{aligned} \tag{III.2}$$

and such that the coupling with the internal scheme (in Ω_+) is stable and a high-order approximation of eq. (III.1).

To numerically solve the initial boundary value problem (III.1), it remains to build averaged ghost-cell values $\boldsymbol{\mathcal{U}}_- = (\boldsymbol{U}_j^n)_{j\leq 0}$ from $\boldsymbol{\mathcal{U}}_+$, on a stencil which depends on the interior scheme. In this chapter, first the focus is made on the scalar advection problem, and a method is derived to reach high-order accuracy. Then, a generalization is made to linear hyperbolic system of conservation laws, and especially for the wave equations. Numerical results illustrate the accuracy of the method all along the chapter. Our findings highlight the need to tackle stability issues due to the reconstruction. Hence, stability results are first obtained using the GKS theory (using lemma I.11), and then the concept of reduced stability is introduced to alleviate part of the computation to obtain stability. The practical interest of the reduced stability definition is confirmed by numerical results. This work is part of a submitted publication [34].

III-1 Inverse Lax–Wendroff procedure for linear hyperbolic systems 125
III-1.1 Derivation of high-order reconstruction operators for the advection problem 126
III-1.2 Derivation of high-order reconstruction operators for the wave equations 131
III-1.3 High-order reconstruction operator for general linear system 143
III-2 Stability of the inverse Lax–Wendroff procedure 144
III-2.1 GKS stability for IBVP using second order reconstruction for the Lax–Wendroff scheme 145
III-2.2 Reduced stability for IBVP discretization 146

III-1 Inverse Lax–Wendroff procedure for linear hyperbolic systems

The Inverse Lax–Wendroff (ILW) method is first detailed for the special case of the scalar advection equation. It is used to build high-order accurate values $\mathcal{U}_-$ using $\mathcal{U}_+$ and the boundary conditions. Numerical experiments illustrate the accuracy of the method. Later on, the procedure is extended to the wave equations, considering two different boundary conditions satisfying the Kreiss condition. At last, a generic procedure is introduced to deal with general linear hyperbolic system with boundary conditions.

III-1.1 Derivation of high-order reconstruction operators for the advection problem

Guiding lines of the method are first explained on the scalar version of (III.1), *ie* the advection equation. Let $a > 0$, the model is

$$\begin{cases} \partial_t u + a\partial_x u = 0, & t \geq 0,\ x > x_s, \\ u(t, x_s) = g(t), & t \geq 0, \\ u(0, x) = u_0(x), & x > x_s. \end{cases} \tag{III.3}$$

As $a > 0$ a boundary condition must be provided at the left boundary. Obviously (III.3) satisfies the Uniform Kreiss conditions (theorem I.9). Using either a finite difference or a finite volume formalism and denoting $\nu = a\frac{\Delta t}{\Delta x}$, numerical schemes under conservative form to solve (III.3) write

$$\overline{u}_j^{n+1} = \overline{u}_j^n - \nu\left(\overline{u}^*_{j+\frac{1}{2}} - \overline{u}^*_{j-\frac{1}{2}}\right). \tag{III.4}$$

Since u is constant along characteristics $x = at$ it is straightforward to show that the numerical flux rewrites

$$\begin{aligned} \overline{u}^*_{j+\frac{1}{2}} &= \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} u(x_{j+\frac{1}{2}}, \theta)d\theta = \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} u(x_{j+\frac{1}{2}} - a\,(\theta - t^n), t^n)\,d\theta, \\ &= \frac{1}{\nu\Delta x}\int_{x_{j+\frac{1}{2}}-\nu\Delta x}^{x_{j+\frac{1}{2}}} u^n(y)dy. \end{aligned}$$

A possible way to compute the fluxes $\overline{u}^*_{j+\frac{1}{2}}$ is to use polynomials P interpolating the primitive of u^n *ie*

$$\overline{u}^*_{j+\frac{1}{2}} = \frac{1}{\nu\Delta x}\left(P(x_{j+\frac{1}{2}}, j) - P(x_{j+\frac{1}{2}} - \nu\Delta x, j)\right). \tag{III.5}$$

Let m be the order of the scheme. Let also r and p be two positive integers such that $r + p = m$. Interpolating polynomials write

$$P(x, j) = \sum_{k=-r}^{p}\left(\prod_{\substack{i=-r \\ i\neq k}}^{p} \frac{x - x_{j+i+\frac{1}{2}}}{x_{j+k+\frac{1}{2}} - x_{j+i+\frac{1}{2}}}\right)\sum_{l=-r}^{k}\overline{u}_{j+l}\Delta x. \tag{III.6}$$

As examples, for $(p, r) = (1, 1)$ we get the Lax–Wendroff scheme

$$\overline{u}^*_{j+\frac{1}{2}} = \frac{1}{2}(\overline{u}_j + \overline{u}_{j+1}) + \frac{\nu}{2}(\overline{u}_j - \overline{u}_{j+1}), \tag{III.7}$$

for $(p, r) = (0, 2)$ the Beam–Warming scheme

$$\overline{u}^*_{j+\frac{1}{2}} = \frac{1}{3}(3\overline{u}_j - \overline{u}_{j-1}) - \frac{\nu}{2}(\overline{u}_j - \overline{u}_{j-1}), \tag{III.8}$$

and for $(p,r) = (1,2)$ we get the third order upwinded scheme (O3):

$$\overline{u}^*_{j+\frac{1}{2}} = \frac{1}{6}(5\overline{u}^n_j + 2\overline{u}^n_{j+1} - \overline{u}^n_{j-1}) + \frac{\nu}{2}(\overline{u}^n_j - \overline{u}^n_{j-1}) + \frac{\nu^2}{6}(\overline{u}^n_{j+1} - 2\overline{u}^n_j + \overline{u}^n_{j-1}). \tag{III.9}$$

The three aforementioned schemes are used in the sequel, whether as examples or for numerical experiments. Such schemes, also described in [148, 149, 44] are very close to those that will be used to solve Euler equations during the remapping phase as in [50, 170, 35] and in section II-2.3. Introducing the floor $\lfloor\rfloor$ and the ceil $\lceil\rceil$ functions

$$\begin{aligned} \lfloor x \rfloor &= m \in \mathbb{Z}, \quad \text{where m is the largest integer less than or equal to x,} \\ \lceil x \rceil &= m \in \mathbb{Z}, \quad \text{where m is the smallest integer greater than or equal to x,} \end{aligned}$$

it is proved in [44] that for $\nu \leq 1$ these schemes are stable for $p = \lfloor \frac{m}{2} \rfloor$ and $r = \lceil \frac{m}{2} \rceil$.

The main idea in the Inverse Lax–Wendroff is to use the system of partial differential equations to change space derivatives into time derivatives in Taylor expansions. For the scalar advection problem, it writes

$$\partial_t u = -a\partial_x u,$$

and since a is assumed to be non-negative, it becomes

$$\partial_x u = (-a)^{-1}\partial_t u.$$

Differentiating in time an arbitrary number of times the previous equation, and changing time derivatives into space derivatives, it writes

$$\partial_x^k u = (-a)^{-k}\partial_t^k u, \quad k \in \mathbb{N}.$$

We present hereafter the formal computations to introduce the previous equality in Taylor expansions. The emphasis is laid on the construction of high-order reconstruction operators for the finite volume approximation.

III-1.1.1 Derivation of high-order reconstruction operators for the finite volume approximation

Ghost-cell methods rely on the determination of the $\mathcal{U}_- = (\overline{u}_0, \overline{u}_{-1}, ...)$ values that are to be set from the boundary condition $g(t)$ and the *interior* values $\mathcal{U}_+ = (\overline{u}_1, \overline{u}_2, \overline{u}_3, ...)$. For x in a

neighborhood of x_s, a formal Taylor expansion leads to

$$\overline{u}(x,t) = \frac{1}{\Delta x}\int_{x-\frac{\Delta x}{2}}^{x+\frac{\Delta x}{2}} u(y,t)\mathrm{d}y = \frac{1}{\Delta x}\int_{x-\frac{\Delta x}{2}}^{x+\frac{\Delta x}{2}} \sum_{k\geq 0} \partial_x^k u(x_s,t)\frac{(y-x_s)^k}{k!}\mathrm{d}y$$
$$= \frac{1}{\Delta x}\sum_{k\geq 0}\partial_x^k u(x_s,t)\left(\frac{(x+\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!} - \frac{(x-\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!}\right)$$

Reminding that for $a \neq 0$ one has $\partial_x^k u = (-a)^{-k}\partial_t^k u$ for the advection equation (III.3)

$$= \frac{1}{\Delta x}\sum_{0\leq k\leq n}(-a)^{-k}\partial_t^k u(x_s,t)\left(\frac{(x+\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!} - \frac{(x-\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!}\right)$$
$$+ \frac{1}{\Delta x}\sum_{k\geq n+1}\partial_x^k u(x_s,t)\left(\frac{(x+\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!} - \frac{(x-\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!}\right).$$

Truncating up to order m, previous equation leads to

$$\overline{u}(x,t) = \frac{1}{\Delta x}\sum_{0\leq k\leq n}(-a)^{-k}\partial_t^k u(x_s,t)\left(\frac{(x+\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!} - \frac{(x-\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!}\right)$$
$$+ \frac{1}{\Delta x}\sum_{n+1\leq k<m}\partial_x^k u(x_s,t)\left(\frac{(x+\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!} - \frac{(x-\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!}\right) + \mathscr{O}(\Delta x^m).$$

Consider a mth-order scheme, and consider we only use the n first time derivatives of g, with $n < m$. Using $u(x_s,t) = g(t)$, one therefore gets

$$\overline{u}(x,t) = \frac{1}{\Delta x}\sum_{0\leq k\leq n}(-a)^{-k}\partial_t^k g(t)\left(\frac{(x+\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!} - \frac{(x-\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!}\right)$$
$$+ \frac{1}{\Delta x}\sum_{n+1\leq k<m}\partial_x^k u(x_s,t)\left(\frac{(x+\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!} - \frac{(x-\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!}\right) + \mathscr{O}(\Delta x^m). \tag{III.10}$$

Consider a scheme that requires n_g ghost-cell values. We introduce two set of points $x^- = \{x_0, x_{-1}, ..., x_{-n_g+1}\}$ and $x^+ = \{x_1, x_2, ..., x_{m-n-1}\}$ which are finite sets of points respectively included in Ω_- and in Ω_+. Using the identity (III.10) and dropping the $\mathscr{O}(\Delta x^m)$ for $x \in x^+$, one builds a system of unknowns $\partial_x^k u(x_s,t)$ with $n+1 \leq k < m$. Solving this system allows then to build averaged ghost-cell values $\overline{u}(x,t)$ for $x \in x^-$.

As an example we consider the O3 scheme ($m = 3$) whose flux is given by (III.9) and whose total

stencil is $S_j = \{j-2, j-1, j, j+1\}$. It therefore requires $n_g = 2$ ghost-cells ($x^- = \{x_0, x_{-1}\}$). For this example and for the sake of simplicity, we assume $g = 0$ and we take $n = 1$ (*ie* $g(t)$ and $\partial_t g(t)$ are known at the boundary). We therefore get $x^+ = \{x_1\}$ and relation (III.10) writes

$$\begin{aligned}\overline{u}(x,t) &= \frac{1}{\Delta x}\partial_x^2 u(x_s,t)\left(\frac{(x+\frac{\Delta x}{2}-x_s)^3}{3!} - \frac{(x-\frac{\Delta x}{2}-x_s)^3}{3!}\right) + \mathscr{O}(\Delta x^3)\\ &= \partial_x^2 u(x_s,t)\left(\frac{12x^2 - 24x\sigma\Delta x + 12\Delta x^2\sigma^2 + \Delta x^2}{24}\right) + \mathscr{O}(\Delta x^3). \qquad \text{(III.11)}\end{aligned}$$

Dropping the $\mathscr{O}(\Delta x^3)$ and using the first interior cell $\overline{u}_1 = \overline{u}(\Delta x, t)$ allows to compute the unknown $\partial_x^2 u(x_s,t)$

$$\partial_x^2 u(x_s,t) = \left(\frac{24}{12\Delta x^2\sigma^2 - 24\sigma\Delta x^2 + 13\Delta x^2}\right)\overline{u}_1. \qquad \text{(III.12)}$$

Ghost-cell values $\overline{u}_0 = \overline{u}(0,t)$ and $\overline{u}_{-1} = \overline{u}(-\Delta x, t)$ can now be explicitely computed from (III.11-III.12)

$$\left\{\begin{aligned}\overline{u}_0 &= \left(\tfrac{12\Delta x^2\sigma^2+\Delta x^2}{24}\right)\partial_x^2 u(x_s,t),\\ \overline{u}_{-1} &= \left(\tfrac{12\Delta x^2\sigma^2+24\sigma\Delta x^2+13\Delta x^2}{24}\right)\partial_x^2 u(x_s,t),\end{aligned}\right. \quad ie \quad \left\{\begin{aligned}\overline{u}_0 &= \tfrac{12\sigma^2+1}{12\sigma^2-24\sigma+13}\overline{u}_1,\\ \overline{u}_{-1} &= \tfrac{12\sigma^2+24\sigma+13}{12\sigma^2-24\sigma+13}\overline{u}_1.\end{aligned}\right.$$

These staightforward computations can be formalized by introducing the Taylor coefficients matrices $\underline{\boldsymbol{\mathcal{Y}}}_+ \in \mathbb{R}^{1\times 1}$ and $\underline{\boldsymbol{\mathcal{Y}}}_- \in \mathbb{R}^{2\times 1}$,

$$\underline{\boldsymbol{\mathcal{Y}}}_+ = \left(\frac{12\Delta x^2\sigma^2 - 24\sigma\Delta x^2 + 13\Delta x^2}{24}\right) \text{ and } \underline{\boldsymbol{\mathcal{Y}}}_- = \begin{pmatrix}\frac{12\Delta x^2\sigma^2 + \Delta x^2}{24}\\ \frac{12\Delta x^2\sigma^2 + 24\sigma\Delta x^2 + 13\Delta x^2}{24}\end{pmatrix}. \qquad \text{(III.13)}$$

Note that for any σ, $\underline{\boldsymbol{\mathcal{Y}}}_+ \geq 0$. Then, under the assumption that $\Delta x \neq 0$, $\underline{\boldsymbol{\mathcal{Y}}}_+$ is invertible. We set $\underline{\boldsymbol{\mathcal{R}}} = \underline{\boldsymbol{\mathcal{Y}}}_-(\underline{\boldsymbol{\mathcal{Y}}}_+)^{-1}$ and get $\boldsymbol{\mathcal{U}}_- = \underline{\boldsymbol{\mathcal{R}}}(\boldsymbol{\mathcal{U}}_+)$, *ie*

$$\begin{pmatrix}\overline{u}_0\\ \overline{u}_{-1}\end{pmatrix} = \begin{pmatrix}\frac{12\sigma^2+1}{12\sigma^2-24\sigma+13}\\ \frac{12\sigma^2+24\sigma+13}{12\sigma^2-24\sigma+13}\end{pmatrix}\overline{u}_1. \qquad \text{(III.14)}$$

We now extend this procedure to the general case. Let m be the order of the reconstruction. Let n denote the number of time derivatives of the boundary condition used in the reconstruction and assume the numerical scheme requires n_g ghost-cells. We build matrices $\underline{\boldsymbol{\mathcal{Y}}}_-^{m,n} \in \mathbb{R}^{n_g\times(m-n-1)}$ and $\underline{\boldsymbol{\mathcal{Y}}}_+^{m,n} \in \mathbb{R}^{(m-n-1)\times(m-n-1)}$

$$\left\{\begin{aligned}(\underline{\boldsymbol{\mathcal{Y}}}_-^{m,n})_{i,j} &= \frac{(x_{1-i}+\frac{\Delta x}{2}-x_s)^{n+j+1} - (x_{1-i}-\frac{\Delta x}{2}-x_s)^{n+j+1}}{\Delta x(n+j+1)!},\\ (\underline{\boldsymbol{\mathcal{Y}}}_+^{m,n})_{i,j} &= \frac{(x_i+\frac{\Delta x}{2}-x_s)^{n+j+1} - (x_i-\frac{\Delta x}{2}-x_s)^{n+j+1}}{\Delta x(n+j+1)!}.\end{aligned}\right. \qquad \text{(III.15)}$$

The boundary condition g, previously assumed to be zero is reintroduced in $\boldsymbol{\mathcal{S}}_-^n \in \mathbb{R}^{n_g}$ and $\boldsymbol{\mathcal{S}}_+^n \in \mathbb{R}^{(m-n-1)}$ defined as

$$\left\{\begin{array}{l} (\boldsymbol{\mathcal{S}}_-^n)_i = \displaystyle\sum_{k=0}^{n}(-a)^k\partial_t^k g(t)\frac{(x_{1-i}+\frac{\Delta x}{2}-x_s)^{k+1}-(x_{1-i}-\frac{\Delta x}{2}-x_s)^{k+1}}{\Delta x(k+1)!}, \\ (\boldsymbol{\mathcal{S}}_+^n)_i = \displaystyle\sum_{k=0}^{n}(-a)^k\partial_t^k g(t)\frac{(x_{i}+\frac{\Delta x}{2}-x_s)^{k+1}-(x_{i}-\frac{\Delta x}{2}-x_s)^{k+1}}{\Delta x(k+1)!}. \end{array}\right. \tag{III.16}$$

Let $\boldsymbol{\Theta} = (\partial_x^{n+1}u, ..., \partial_x^{m-1}u)^t$. Relation (III.10) can be rewritten

$$\left\{\begin{array}{l} \boldsymbol{\mathcal{U}}_- = \boldsymbol{\mathcal{S}}_-^n + \underline{\boldsymbol{\mathcal{Y}}}_-^{m,n}\cdot\boldsymbol{\Theta}, \\ \boldsymbol{\mathcal{U}}_+ = \boldsymbol{\mathcal{S}}_+^n + \underline{\boldsymbol{\mathcal{Y}}}_+^{m,n}\cdot\boldsymbol{\Theta}. \end{array}\right. \tag{III.17}$$

A similar proof as for Vandermonde matrices shows that $\underline{\boldsymbol{\mathcal{Y}}}_+^{m,n}$ is invertible for any (m,n) if $0 \le n < m$. Elimination of $\boldsymbol{\Theta}$ in (III.17) leads to

$$\boldsymbol{\mathcal{U}}_- = \boldsymbol{\mathcal{S}}_-^n + \underline{\boldsymbol{\mathcal{Y}}}_-^{m,n}\cdot(\underline{\boldsymbol{\mathcal{Y}}}_+^{m,n})^{-1}\cdot(\boldsymbol{\mathcal{U}}_+ - \boldsymbol{\mathcal{S}}_+^n). \tag{III.18}$$

This relation gives a reconstruction up to m^{th}-order of $\overline{u}$ outside the computational domain using the n first time derivatives of g. It defines the so-called $\underline{\boldsymbol{\mathcal{R}}}^{m,n}$ reconstruction operator

$$\underline{\boldsymbol{\mathcal{R}}}^{m,n} = \underline{\boldsymbol{\mathcal{Y}}}_-^{m,n}\cdot(\underline{\boldsymbol{\mathcal{Y}}}_+^{m,n})^{-1}. \tag{III.19}$$

Remark III.1. The previous formal computations also apply straightforwardly in the case of finite difference schemes. Terms of the form $\left(\frac{(x+\frac{\Delta x}{2}-x_s)^{k+1}}{\Delta x(k+1)!} - \frac{(x-\frac{\Delta x}{2}-x_s)^{k+1}}{\Delta x(k+1)!}\right)$ become $\frac{(x-x_s)^k}{k!}$ in the finite difference case.

III-1.1.2 Experimental order of accuracy of the procedure

Previous computations yield high-order accurate numerical methods to solve eq. (III.3). Consider the initial boundary value problem (III.3) with $a = 1$ and the following C^∞ data

$$\left\{\begin{array}{lllll} u(0,x) & = & u_0(x) & = & 0, \\ u(t,x_s) & = & g(t) & = & \mathrm{e}^{-\frac{0.1}{t^2}}\sin(4\pi t). \end{array}\right. \tag{III.20}$$

Equation (III.3) is solved on $\Omega = \{x \in \mathbb{R},\ x_s < x < 2\}$, with a classical outflow boundary condition in $x = 2$ and the high-order accurate boundary treatment previously proposed at $x = x_s = \frac{\sqrt{3}}{2}\ 10^{-3}$. The computational domain, discretized in N_x cells, is $[0,2]$ so that the left boundary lies in the first cell. The CFL is set to 0.8. Computations are done in order to assess the accuracy of the proposed methods. In Table III.1, we present the l_1-error with respect to the number of cells for the $\underline{\boldsymbol{\mathcal{R}}}^{3,0}$,$\underline{\boldsymbol{\mathcal{R}}}^{3,1}$, and $\underline{\boldsymbol{\mathcal{R}}}^{3,2}$ reconstructions using the 3rd-order interior

N_x	$\underline{\mathcal{R}}^{3,0}$		$\underline{\mathcal{R}}^{3,1}$		$\underline{\mathcal{R}}^{3,2}$	
20	3.1e-2	·	2.8e-2	·	2.9e-2	·
40	5.9e-3	2.39	5.6e-3	2.32	5.6e-3	2.35
80	8.0e-4	2.88	7.7e-4	2.86	7.7e-4	2.86
160	1.0e-4	2.93	1.0e-4	2.92	1.0e-4	2.92
320	1.3e-5	2.97	1.3e-5	2.97	1.3e-5	2.97
640	1.7e-6	2.99	1.6e-6	2.99	1.6e-6	2.99
1280	2.1e-7	2.99	2.1e-7	2.99	2.1e-7	2.99

Table III.1 – l^1-error and experimental order of convergence for the 3rd-order scheme together with the $\underline{\mathcal{R}}^{3,n}$ finite-volume reconstruction polynomial at $t = 1.5$.

N_x	$\underline{\mathcal{R}}^{4,0}$		$\underline{\mathcal{R}}^{4,1}$		$\underline{\mathcal{R}}^{4,2}$		$\underline{\mathcal{R}}^{4,3}$	
20	2.0e-2	·	1.9e-2	·	2.0e-2	·	2.1e-2	·
40	2.4e-3	3.12	2.3e-3	3.10	2.3e-3	3.15	2.3e-3	3.21
80	1.7e-4	3.80	1.7e-4	3.76	1.7e-4	3.76	1.7e-4	3.76
160	1.1e-5	3.90	1.1e-5	3.89	1.1e-5	3.89	1.1e-5	3.89
320	7.4e-7	3.96	7.3e-7	3.96	7.3e-7	3.96	7.2e-7	3.96
640	4.7e-8	3.98	4.6e-8	3.98	4.6e-8	3.98	4.6e-8	3.98
1280	2.9e-9	3.99	2.9e-9	3.99	2.9e-9	3.99	2.9e-9	3.99

Table III.2 – l^1-error and experimental order of convergence for the 4th-order scheme together with the $\underline{\mathcal{R}}^{4,n}$ finite-volume reconstruction polynomial at $t = 1.5$.

scheme (III.4), (III.9). In Table III.2, we present the l_1-error with respect to the number of cells for the $\underline{\mathcal{R}}^{4,0}$,$\underline{\mathcal{R}}^{4,1}$, $\underline{\mathcal{R}}^{4,2}$, and $\underline{\mathcal{R}}^{4,3}$ reconstructions using the 4th-order interior scheme (III.4). The expected order of convergence for both schemes is reached for all reconstructions. We also have checked that modifying x_s does not alter the order of accuracy but slightly changes the initial error level (for $N_x = 20$). Similar experimental orders of convergence for finite difference reconstruction operators have been recovered. An important feature of the reconstruction operator is its impact on the final scheme stability. This will be discussed hereafter in section III-2.1.

III-1.2 Derivation of high-order reconstruction operators for the wave equations

The wave equations have already been detailed for the linear stability analysis of the staggered schemes in section II-2. The system of equations is

$$\begin{cases} \partial_t u + \partial_x p &= 0, \\ \partial_t p + \partial_x u &= 0, \end{cases} \tag{III.21}$$

which can be written, for $\boldsymbol{U} = (u, p)^t \in \mathbb{R}^2$ as

$$\partial_t \boldsymbol{U} + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \partial_x \boldsymbol{U} = 0. \tag{III.22}$$

In the following, we introduce the matrix $\underline{\boldsymbol{A}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, and obviously previous equation rewrites

$$\partial_t \boldsymbol{U} + \underline{\boldsymbol{A}} \partial_x \boldsymbol{U} = 0. \tag{III.23}$$

The initial value boundary problem that we are interested in therefore writes

$$\begin{cases} \partial_t \boldsymbol{U} + \underline{\boldsymbol{A}} \partial_x \boldsymbol{U} &= 0, & t > 0, \quad x > x_s \\ \underline{\boldsymbol{B}} \boldsymbol{U} &= \underline{\boldsymbol{B}} \boldsymbol{G}, & t > 0, \quad x = x_s \end{cases} \tag{III.24}$$

Lemma III.1. *The initial value boundary problem* (III.24) *is well-posed in the sense of theorem I.9 if* $\underline{\boldsymbol{B}} \in \mathbb{R}^{1\times 2}$ *and satisfies* $\underline{\boldsymbol{B}} = \begin{pmatrix} b_1 & b_2 \end{pmatrix}$ *with* $b_1 + b_2 \neq 0$.

Proof. Trivially, one has that the spectrum of $\underline{\boldsymbol{A}}$ satisfies $Sp(\underline{\boldsymbol{A}}) = \{-1, 1\}$ and the eigenvectors are

$$\boldsymbol{v}_+ = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{v}_- = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Using the notation introduced in theorem I.9, it yields that $\underline{\boldsymbol{T}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Then, we get that $\underline{\boldsymbol{B}} \in \mathbb{R}^{1\times 2}$ with $\underline{\boldsymbol{B}} = \begin{pmatrix} b_1 & b_2 \end{pmatrix}$. Thus $\underline{\boldsymbol{B}}\underline{\boldsymbol{T}} = b_1 + b_2$. To ensure invertibility of $\underline{\boldsymbol{B}}\underline{\boldsymbol{T}}$, one requires that $b_1 + b_2 \neq 0$, which concludes the proof. ■

In the following, two different matrices $\underline{\boldsymbol{B}}$ are proposed which satisfy lemma III.1. The emphasis is laid on how the boundary condition impacts the reconstruction operator. Before studying specifically the boundary condition, the interior schemes are introduce to solve eq. (III.21). Those schemes are the linear version of the one proposed for the Euler system introduced and detailed in section II-2.

III-1.2.1 Runge–Kutta based staggered schemes for the wave equations

The Runge–Kutta based staggered schemes for the wave equations (already introduced in section II-2) are eq. (III.25), denoting $\nu = \frac{\Delta t}{\Delta X}$,

$$\begin{cases} \overline{p}_i^{n+\alpha_m} &= \overline{p}_i^n - \nu \sum_{l=0}^{m-1} a_{m,l} du_i^{n+\alpha_l}, \\ \overline{u}_{i+\frac{1}{2}}^{n+\alpha_m} &= \overline{u}_{i+\frac{1}{2}}^n - \nu \sum_{l=0}^{m-1} a_{m,l} dp_{i+\frac{1}{2}}^{n+\alpha_l}, \end{cases} \qquad \begin{cases} \overline{p}_i^{n+1} &= \overline{p}_i^n - \nu \sum_{l=0}^{s-1} \theta_l du_i^{n+\alpha_l}, \\ \overline{u}_{i+\frac{1}{2}}^{n+1} &= \overline{u}_{i+\frac{1}{2}}^n - \nu \sum_{l=0}^{s-1} \theta_l dp_{i+\frac{1}{2}}^{n+\alpha_l}, \end{cases} \tag{III.25}$$

The explicit Runge–Kutta coefficients are given in table III.3. A possible way would be to build the reconstruction operator only at time $t = t^n$, exactly as for the advection case with one-step schemes. However, considering as an example that there are 3 ghost-cells values to be built at each sub-cycle, and that the scheme requires 6 Runge–Kutta sub-cycles, then one must build

α_1	$a_{1,0}$	0	0	0	$\cdots$
α_2	$a_{2,0}$	$a_{2,1}$	0	0	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\cdots$	$\cdots$
α_p	$a_{p,0}$	$\cdots$	$\cdots$	$a_{p,p-1}$	0
1	θ_0	θ_1	$\cdots$	θ_{p-1}	θ_p

Table III.3 – Example of Butcher table for explicit Runge–Kutta sequence with p sub-cycles.

18 ghost-cells values at time $t = t^n$. This will probably be a predicament for the stability of the effective schemes. Thus, the choice has been made here to build ghost-cells values at each Runge–Kutta sub-cycles. However, as explained and illustrated by Carpenter and al. in [20], prescribing "naively" boundary conditions at each Runge–Kutta sub-cycle yields only second order of accuracy. Lemma III.2 gives results concerning a way to impose high-order accurate values of a given function at intermediary fictitous time-step.

Lemma III.2 (High-order accurate in time for function values at intermediary fictitious time)**.** *Consider a qth-order explicit Runge–Kutta sequences whose coefficients are given by a Butcher table as table III.3. In order to impose high-order accurate values of a function $g : t \to g(t)$ at intermediary fictitious time, one sets*

$$g^{n+\alpha_l} = g(t^n) + \sum_{r=1}^{q} \beta_l^r \partial_t^r g(t^n) \Delta t^r,$$

where the β coefficients satisfy

$$\begin{cases} \beta_l^1 &= \displaystyle\sum_{m=0}^{l-1} a_{l,m}, \\ \beta_l^r &= \displaystyle\sum_{m=0}^{l-1} a_{l,m} \beta_m^{r-1}, \\ \beta_{p+1}^r &= \displaystyle\sum_{m=0}^{p} \theta_m \beta_m^{r-1}. \end{cases}$$

Proof. To build high-order accurate boundary conditions, we consider the following system (III.26):

$$\begin{cases} \partial_t g_0(t) = g_1(t) \\ \vdots \\ \partial_t g_q(t) = g_{q+1}(t) \\ \vdots \end{cases} \tag{III.26}$$

System (III.26) needs closure to be well posed. We close the system considering that for a fixed $q \in \mathbb{N}$ (linked to the order of the Runge–Kutta sequence), we have $\partial_t g_{q+1}(t) = 0$. This way, we

get the following system (III.27).

$$\begin{cases} \partial_t g_0(t) & = g_1(t) \\ \vdots & \\ \partial_t g_q(t) & = g_{q+1}(t) \\ \partial_t g_{q+1}(t) & = 0 \end{cases} \tag{III.27}$$

We consider q^{th} order *explicit* Runge–Kutta schemes with the following notations for Runge–Kutta sequences: α_l is the time step for the lth sub-cycle, $a_{l,m}$ the l, m term of the Butcher table and θ_m the mth reconstruction coefficient for the last step. We consider p sub-cycles schemes (see table III.3).

Using Runge–Kutta integration in time with time-step Δt and considering that $g_k^n = \frac{\mathrm{d}^k g}{\mathrm{d}t^k}(t^n)$ we will get the following schemes, for $l \in \{1, ..., p+1\}$

$$\begin{cases} g_0^{n+\alpha_l} = g_0^n + \Delta t \sum_{m=0}^{l-1} a_{l,m} g_1^{n+\alpha_m} \\ g_1^{n+\alpha_l} = g_1^n + \Delta t \sum_{m=0}^{l-1} a_{l,m} g_2^{n+\alpha_m} \\ \vdots \quad = \vdots \\ g_q^{n+\alpha_l} = g_q^n \end{cases}, \tag{III.28}$$

Developing system (III.28) to keep only terms with $g_0^n, g_1^n, ..., g_q^n$, we get for $k \in \{0, ..., q\}$

$$g_k^{n+\alpha_l} = g_k^n + \sum_{r=1}^{r+k \leq q} \beta_l^r g_{k+r}^n \Delta t^r, \tag{III.29}$$

where the β_l^m coefficients satisfy the following equation:

$$\begin{cases} \beta_l^1 & = \sum_{m=0}^{l-1} a_{l,m}, \\ \beta_l^r & = \sum_{m=0}^{l-1} a_{l,m} \beta_m^{r-1}, \\ \beta_{p+1}^r & = \sum_{m=0}^{p} \theta_m \beta_m^{r-1}. \end{cases} \tag{III.30}$$

which concludes the proof using $k = 0$ into eq. (III.29). ■

Once the Butcher table of a Runge–Kutta sequence is given, the β_l^r can easily be computed once and for all. Then, it allows to impose the value of the $g^{n+\alpha_l}$ function only of Δt and of the values of g and its time-derivatives at time $t = t^n$. Let us prove that the "time matching" method which consists of imposing $g^{n+\alpha l} = g(t^n + \alpha_l \Delta t)$ is only second order accurate in time.

Lemma III.3 (Low order accuracy of the "time matching" method)**.** *For general Butcher coef-*

ficients, the "time matching" method is only second order accurate. It satisfies

$$g(t^n + \alpha_l \Delta t) = g^{n+\alpha_l} + \mathscr{O}(\Delta t^2).$$

Remark III.2. This is a generalization to any Runge–Kutta sequences of the results given by Carpenter and al. in [20].

Proof. Recall that

$$g^{n+\alpha_l} = g(t^n) + \sum_{r=1}^{q} \beta_l^r \partial_t^r g(t^n) \Delta t^r.$$

The Taylor expansion in Δt of $g(t^n + \alpha_l \Delta t)$ writes

$$g(t^n + \alpha_l \Delta t) = g(t^n) + \sum_{r=1}^{q} \partial_t^r g(t^n) \frac{(\alpha_l \Delta t)^r}{r!} + \mathscr{O}(\Delta t^{q+1})$$

Then it leads to

$$g^{n+\alpha_l} - g(t^n + \alpha_l \Delta t) = \sum_{r=1}^{q} \partial_t^r g(t^n) \Delta t^r \left(\beta_l^r - \frac{(\alpha_l)^r}{r!} \right) + \mathscr{O}(\Delta t^{q+1})$$

Introducing the notations $\gamma_r = \beta_l^r - \dfrac{(\alpha_l)^r}{r!}$, one gets that

$$\gamma_1 = \beta_l^1 - \alpha_l = \sum_{m=0}^{l-1} a_{l,m} - \alpha_l = 0,$$

since $\alpha_l = \displaystyle\sum_{m=0}^{l-1} a_{l,m}$ for any Butcher table. Now, let us consider γ_2, it writes

$$\begin{aligned} \gamma_2 &= \beta_l^2 - \frac{1}{2}\alpha_l^2 \\ &= \sum_{m=0}^{l-1} a_{l,m} \beta_m^1 - \frac{1}{2}\alpha_l^2 \\ &= \sum_{m=0}^{l-1} a_{l,m} \alpha_m - \frac{1}{2}\alpha_l^2, \end{aligned}$$

which is not equal to zero for general coefficients $a_{l,m}$. Hence, it yields that

$$g^{n+\alpha_l} - g(t^n + \alpha_l \Delta t) = \mathscr{O}(\Delta t^2).$$

■

Using the β coefficients, let us now deal with building appropriate reconstruction operators depending on the boundary condition. A method has been devised to deal with such a problem,

and then to build high-order boundary conditions for any explicit Runge–Kutta sequences. It has been done for the Runge–Kutta sequences presented in appendix, section A.1.

III-1.2.2 Reconstruction operators for the wave equations with boundary conditions on velocity

First, we consider that the matrix $\underline{\boldsymbol{B}}$ takes the simple form $\underline{\boldsymbol{B}} = \begin{pmatrix} 1 & 0 \end{pmatrix}$, which obviously satisfies lemma III.1. The system rewrites as

$$\left\{ \begin{array}{lcll} \partial_t u + \partial_x p & = & 0, & x \geq x_s, \quad t > 0, \\ \partial_t p + \partial_x u & = & 0, & x \geq x_s, \quad t > 0, \\ u(x_s, t) & = & g(t), & t > 0. \end{array} \right. \tag{III.31}$$

Then using the eq. (III.31), one gets in particular that for any $q \in \mathbb{N}$

$$\left\{ \begin{array}{lcl} \partial_t^{2q+1} u & = & -\partial_x^{2q+1} p, \\ \partial_t^{2q} u & = & \partial_x^{2q} u, \end{array} \right. \tag{III.32}$$

which yields

$$\left\{ \begin{array}{lcl} \partial_x^{2q+1} p(x_s, t) & = & -\partial_t^{2q+1} g(t), \\ \partial_x^{2q} u(x_s, t) & = & \partial_t^{2q} g(t). \end{array} \right. \tag{III.33}$$

For x in a neighborhood of x_s, a formal Taylor expansion leads to

$$\begin{aligned} \begin{pmatrix} \overline{u} \\ \overline{p} \end{pmatrix}(x,t) &= \frac{1}{\Delta x} \int_{x-\frac{\Delta x}{2}}^{x+\frac{\Delta x}{2}} \begin{pmatrix} u \\ p \end{pmatrix}(y,t) \mathrm{d}y = \frac{1}{\Delta x} \int_{x-\frac{\Delta x}{2}}^{x+\frac{\Delta x}{2}} \sum_{k \geq 0} \partial_x^k \begin{pmatrix} u \\ p \end{pmatrix}(x_s,t) \frac{(y-x_s)^k}{k!} \mathrm{d}y \\ &= \frac{1}{\Delta x} \sum_{k \geq 0} \partial_x^k \begin{pmatrix} u \\ p \end{pmatrix}(x_s,t) \left(\frac{(x + \frac{\Delta x}{2} - x_s)^{k+1}}{(k+1)!} - \frac{(x - \frac{\Delta x}{2} - x_s)^{k+1}}{(k+1)!} \right). \end{aligned}$$

Introducing the notation

$$\psi_k(x) = \left(\frac{(x + \frac{\Delta x}{2} - x_s)^{k+1}}{(k+1)!} - \frac{(x - \frac{\Delta x}{2} - x_s)^{k+1}}{(k+1)!} \right),$$

it rewrites as

$$\begin{aligned}\begin{pmatrix}\overline{u}\\ \overline{p}\end{pmatrix}(x,t) &= \frac{1}{\Delta x}\sum_{k\geq 0}\begin{pmatrix}\partial_x^k u\\ \partial_x^k p\end{pmatrix}(x_s,t)\psi_k(x),\\ &= \frac{1}{\Delta x}\left[\sum_{k\geq 0}^{\lfloor\frac{n}{2}\rfloor}\begin{pmatrix}\partial_x^{2k} u\\ \partial_x^{2k} p\end{pmatrix}(x_s,t)\psi_{2k}(x) + \sum_{k\geq 0}^{\lfloor\frac{n-1}{2}\rfloor}\begin{pmatrix}\partial_x^{2k+1} u\\ \partial_x^{2k+1} p\end{pmatrix}(x_s,t)\psi_{2k+1}(x)\right]\\ &+ \frac{1}{\Delta x}\sum_{k\geq n+1}\begin{pmatrix}\partial_x^k u\\ \partial_x^k p\end{pmatrix}(x_s,t)\psi_k(x).\end{aligned}$$

Reminding that $\partial_x^{2k}u = \partial_t^{2k}u$ and that $\partial_x^{2k+1}p = -\partial_t^{2k+1}u$

$$\begin{aligned}\begin{pmatrix}\overline{u}\\ \overline{p}\end{pmatrix}(x,t) &= \frac{1}{\Delta x}\left[\sum_{k\geq 0}^{\lfloor\frac{n}{2}\rfloor}\begin{pmatrix}\partial_t^{2k} u\\ \partial_x^{2k} p\end{pmatrix}(x_s,t)\psi_{2k}(x) + \sum_{k\geq 0}^{\lfloor\frac{n-1}{2}\rfloor}\begin{pmatrix}\partial_x^{2k+1} u\\ -\partial_t^{2k+1} u\end{pmatrix}(x_s,t)\psi_{2k+1}(x)\right]\\ &+ \frac{1}{\Delta x}\sum_{k\geq n+1}\begin{pmatrix}\partial_x^k u\\ \partial_x^k p\end{pmatrix}(x_s,t)\psi_k(x).\end{aligned}$$

Truncating up to order m, previous equation gives

$$\begin{aligned}\begin{pmatrix}\overline{u}\\ \overline{p}\end{pmatrix}(x,t) &= \frac{1}{\Delta x}\left[\sum_{k\geq 0}^{\lfloor\frac{n}{2}\rfloor}\begin{pmatrix}\partial_t^{2k} u\\ \partial_x^{2k} p\end{pmatrix}(x_s,t)\psi_{2k}(x) + \sum_{k\geq 0}^{\lfloor\frac{n-1}{2}\rfloor}\begin{pmatrix}\partial_x^{2k+1} u\\ -\partial_t^{2k+1} u\end{pmatrix}(x_s,t)\psi_{2k+1}(x)\right]\\ &+ \frac{1}{\Delta x}\sum_{k=n+1}^{m-1}\begin{pmatrix}\partial_x^k u\\ \partial_x^k p\end{pmatrix}(x_s,t)\psi_k(x) + \mathscr{O}(\Delta x^m).\end{aligned}$$

Inserting boundary condition and dropping the $\mathscr{O}(\Delta x^m)$, one gets

$$\begin{aligned}\begin{pmatrix}\overline{u}\\ \overline{p}\end{pmatrix}(x,t) &= \frac{1}{\Delta x}\left[\sum_{k\geq 0}^{\lfloor\frac{n}{2}\rfloor}\begin{pmatrix}\partial_t^{2k} g(t)\\ \partial_x^{2k} p(x_s,t)\end{pmatrix}\psi_{2k}(x) + \sum_{k\geq 0}^{\lfloor\frac{n-1}{2}\rfloor}\begin{pmatrix}\partial_x^{2k+1} u(x_s,t)\\ -\partial_t^{2k+1} g(t)\end{pmatrix}\psi_{2k+1}(x)\right]\\ &+ \frac{1}{\Delta x}\sum_{k\geq n+1}^{m-1}\begin{pmatrix}\partial_x^k u(x_s,t)\\ \partial_x^k p(x_s,t)\end{pmatrix}\psi_k(x).\end{aligned}$$

Getting the terms in g in the left hand side, it rewrites as

$$\begin{cases}\overline{u}(x,t) - \frac{1}{\Delta x}\sum_{k\geq 0}^{\lfloor\frac{n}{2}\rfloor}\partial_t^{2k} g(t)\psi_{2k}(x) = \dfrac{1}{\Delta x}\left[\sum_{k\geq 0}^{\lfloor\frac{n-1}{2}\rfloor}\partial_x^{2k+1} u(x_s,t)\psi_{2k+1}(x) + \sum_{k\geq n+1}^{m-1}\partial_x^k u(x_s,t)\psi_k(x)\right],\\ \overline{p}(x,t) + \frac{1}{\Delta x}\sum_{k\geq 0}^{\lfloor\frac{n-1}{2}\rfloor}\partial_t^{2k+1} g(t)\psi_{2k+1}(x) = \dfrac{1}{\Delta x}\left[\sum_{k\geq 0}^{\lfloor\frac{n}{2}\rfloor}\partial_x^{2k} p(x_s,t)\psi_{2k}(x) + \sum_{k\geq n+1}^{m-1}\partial_x^k p(x_s,t)\psi_k(x)\right].\end{cases}$$

Then, it enables to establish a similar procedure to the one presented in section III-1.1. It writes

$$\begin{cases} \boldsymbol{\mathcal{U}}_- - \boldsymbol{\mathcal{S}}_-^n = \underline{\boldsymbol{\mathcal{Y}}}_-^{m,n} \cdot \boldsymbol{\Theta}, \\ \boldsymbol{\mathcal{U}}_+ - \boldsymbol{\mathcal{S}}_+^n = \underline{\boldsymbol{\mathcal{Y}}}_+^{m,n} \cdot \boldsymbol{\Theta}, \end{cases} \tag{III.34}$$

A similar proof as for Vandermonde matrices shows that $\underline{\boldsymbol{\mathcal{Y}}}_+^{m,n}$ is invertible for any (m,n) with $0 \leq n < m$. Then eq. (III.34) gives after elimination of $\boldsymbol{\Theta}$ formed with spatial derivatives of u and p,

$$\boldsymbol{\mathcal{U}}_- = \boldsymbol{\mathcal{S}}_-^n + \underline{\boldsymbol{\mathcal{Y}}}_-^{m,n} \cdot (\underline{\boldsymbol{\mathcal{Y}}}_+^{m,n})^{-1} \cdot (\boldsymbol{\mathcal{U}}_+ - \boldsymbol{\mathcal{S}}_+^n). \tag{III.35}$$

Here again, the reconstruction operator writes $\underline{\boldsymbol{\mathcal{R}}}^{m,n} = \underline{\boldsymbol{\mathcal{Y}}}_-^{m,n} \cdot (\underline{\boldsymbol{\mathcal{Y}}}_+^{m,n})^{-1}$.

Remark III.3. Straightforwardly, as u and p play a symetric role, one deduces the reconstruction operator for the following IBVP problem

$$\begin{cases} \partial_t u + \partial_x p &= 0, & x \geq x_s, & t > 0, \\ \partial_t p + \partial_x u &= 0, & x \geq x_s, & t > 0, \\ p(x_s,t) &= g(t), & t > 0. \end{cases}$$

III-1.2.3 Reconstruction operators for the wave equations with mixed boundary conditions on both velocity and pressure

First, we consider that the matrix $\underline{\boldsymbol{B}}$ takes the form $\underline{\boldsymbol{B}} = \begin{pmatrix} 1 & \lambda \end{pmatrix}$, where λ is chosen in order to satisfy lemma III.1. It yields a condition on λ which writes $\lambda \neq -1$. The special case where $\lambda = 0$ has been dealt with previously. The system rewrites as

$$\begin{cases} \partial_t u + \partial_x p &= 0, & x \geq x_s, & t > 0, \\ \partial_t p + \partial_x u &= 0, & x \geq x_s, & t > 0, \\ u(x_s,t) + \lambda p(x_s,t) &= g(t), & t > 0. \end{cases} \tag{III.36}$$

In particular, one has

$$\partial_t^q \begin{pmatrix} u \\ p \end{pmatrix} = (-A)^q \partial_x^q \begin{pmatrix} u \\ p \end{pmatrix}, \tag{III.37}$$

and since $\underline{\boldsymbol{A}}$ is invertible, it leads to

$$\partial_x^q \begin{pmatrix} u \\ p \end{pmatrix} = (-A)^{-q} \partial_t^q \begin{pmatrix} u \\ p \end{pmatrix}. \tag{III.38}$$

The matrix $\underline{\widehat{\boldsymbol{B}}} \in \mathbb{R}^{p \times p}$ is introduced as

$$\underline{\widehat{\boldsymbol{B}}} = \begin{pmatrix} \underline{\boldsymbol{B}} \\ \underline{\boldsymbol{0}} \end{pmatrix}.$$

Keeping the notation previously introduced, for x in a neighborhood of x_s, a formal Taylor

expansion gives

$$\begin{pmatrix}\overline{u}\\ \overline{p}\end{pmatrix}(x,t) = \frac{1}{\Delta x}\sum_{k\geq 0}\begin{pmatrix}\partial_x^k u\\ \partial_x^k p\end{pmatrix}(x_s,t)\psi_k(x),$$

which is split into two terms

$$\begin{pmatrix}\overline{u}\\ \overline{p}\end{pmatrix}(x,t) = \frac{1}{\Delta x}\sum_{0\leq k\leq n}(-1)^k\psi_k(x)(\underline{\boldsymbol{A}}^k)^{-1}\partial_t^k\begin{pmatrix}u\\ p\end{pmatrix}(x_s,t) + \frac{1}{\Delta x}\sum_{n+1\leq k}\psi_k(x)\partial_x^k\begin{pmatrix}u\\ p\end{pmatrix}(x_s,t),$$

Decomposing along $\underline{\widehat{\boldsymbol{B}}}$ and $\underline{\boldsymbol{I}} - \underline{\widehat{\boldsymbol{B}}}$, it leads to

$$\begin{aligned}\begin{pmatrix}\overline{u}\\ \overline{p}\end{pmatrix}(x,t) &= \frac{1}{\Delta x}\sum_{0\leq k\leq n}\psi_k(x)(\underline{\boldsymbol{A}}^k)^{-1}(-1)^k\underline{\widehat{\boldsymbol{B}}}\partial_t^k\begin{pmatrix}u\\ p\end{pmatrix}(x_s,t) + \frac{1}{\Delta x}\sum_{0\leq k\leq n}\psi_k(x)\underline{\boldsymbol{A}}^{-k}(\underline{\boldsymbol{I}} - \underline{\widehat{\boldsymbol{B}}})\underline{\boldsymbol{A}}^k\partial_x^k\begin{pmatrix}u\\ p\end{pmatrix}(x_s,t)\\ &+ \frac{1}{\Delta x}\sum_{n+1\leq k}\psi_k(x)\partial_x^k\begin{pmatrix}u\\ p\end{pmatrix}(x_s,t).\end{aligned}$$

Truncating up to m^{th}-order, dropping the $\mathscr{O}(\Delta x^m)$ and using $\underline{\widehat{\boldsymbol{B}}}\begin{pmatrix}u\\ p\end{pmatrix}(x_s,t) = \underline{\widehat{\boldsymbol{B}}}\boldsymbol{G}(t)$, we get

$$\begin{aligned}\begin{pmatrix}\overline{u}\\ \overline{p}\end{pmatrix}(x,t) &= \frac{1}{\Delta x}\sum_{0\leq k\leq n}\psi_k(x)\underline{\boldsymbol{A}}^{-k}(-1)^k\underline{\widehat{\boldsymbol{B}}}\partial_t^k\boldsymbol{G}(t)\\ &+ \frac{1}{\Delta x}\sum_{0\leq k\leq n}\psi_k(x)\underline{\boldsymbol{A}}^{-k}(\underline{\boldsymbol{I}} - \underline{\widehat{\boldsymbol{B}}})\underline{\boldsymbol{A}}^k\partial_x^k\begin{pmatrix}u\\ p\end{pmatrix}(x_s,t)\\ &+ \frac{1}{\Delta x}\sum_{n+1\leq k<m}\psi_k(x)\partial_x^k\begin{pmatrix}u\\ p\end{pmatrix}(x_s,t).\end{aligned}$$

Noticing in particular that $\underline{\boldsymbol{A}}^2 = \underline{\boldsymbol{I}}$, thus $A^{-1} = A$, one gets

$$\begin{aligned}\begin{pmatrix}\overline{u}\\ \overline{p}\end{pmatrix}(x,t) &= -\frac{1}{\Delta x}\sum_{0\leq k\leq\lfloor\frac{n-1}{2}\rfloor}\psi_{2k+1}(x)\underline{\boldsymbol{A}\widehat{\boldsymbol{B}}}\partial_t^{2k+1}\boldsymbol{G}(t)\\ &+ \frac{1}{\Delta x}\sum_{0\leq k\leq\lfloor\frac{n}{2}\rfloor}\psi_{2k}(x)\underline{\widehat{\boldsymbol{B}}}\partial_t^{2k}\boldsymbol{G}(t)\\ &+ \frac{1}{\Delta x}\sum_{0\leq k\leq\lfloor\frac{n-1}{2}\rfloor}\psi_{2k+1}(x)\underline{\boldsymbol{A}}(\underline{\boldsymbol{I}} - \underline{\widehat{\boldsymbol{B}}})\underline{\boldsymbol{A}}\partial_x^{2k+1}\begin{pmatrix}u\\ p\end{pmatrix}(x_s,t)\\ &+ \frac{1}{\Delta x}\sum_{0\leq k\leq\lfloor\frac{n}{2}\rfloor}\psi_{2k}(x)(\underline{\boldsymbol{I}} - \underline{\widehat{\boldsymbol{B}}})\partial_x^{2k}\begin{pmatrix}u\\ p\end{pmatrix}(x_s,t)\\ &+ \frac{1}{\Delta x}\sum_{n+1\leq k<m}\psi_k(x)\partial_x^k\begin{pmatrix}u\\ p\end{pmatrix}(x_s,t).\end{aligned}$$

Then, computing the values of $(\underline{\boldsymbol{I}}-\underline{\widehat{\boldsymbol{B}}})$,$\underline{\boldsymbol{A}}(\underline{\boldsymbol{I}}-\underline{\widehat{\boldsymbol{B}}})\underline{\boldsymbol{A}}$ and $\underline{\boldsymbol{A}\widehat{\boldsymbol{B}}}$, one gets the following results

$$\begin{cases} \underline{\boldsymbol{I}}-\underline{\widehat{\boldsymbol{B}}} & = \begin{pmatrix} 0 & -\lambda \\ 0 & 1 \end{pmatrix}, \\ \underline{\boldsymbol{A}\widehat{\boldsymbol{B}}} & = \begin{pmatrix} \lambda & 1 \\ 0 & 0 \end{pmatrix}, \\ \underline{\boldsymbol{A}}(\underline{\boldsymbol{I}}-\underline{\widehat{\boldsymbol{B}}})\underline{\boldsymbol{A}} & = \begin{pmatrix} 1 & 0 \\ -\lambda & 0 \end{pmatrix}, \end{cases}$$

which leads to, denoting $g=\underline{\boldsymbol{B}}\boldsymbol{G}$ and inserting in the previous expression

$$\begin{aligned} \begin{pmatrix} \overline{u} \\ \overline{p} \end{pmatrix}(x,t) = & -\frac{1}{\Delta x}\sum_{0\le k\le\lfloor\frac{n-1}{2}\rfloor}\psi_{2k+1}(x)\begin{pmatrix} 0 \\ \partial_t^{2k+1}g(t) \end{pmatrix} \\ & +\frac{1}{\Delta x}\sum_{0\le k\le\lfloor\frac{n}{2}\rfloor}\psi_{2k}(x)\begin{pmatrix} \partial_t^{2k}g(t) \\ 0 \end{pmatrix} \\ & +\frac{1}{\Delta x}\sum_{0\le k\le\lfloor\frac{n-1}{2}\rfloor}\psi_{2k+1}(x)\begin{pmatrix} \partial_x^{2k+1}u(x_s,t) \\ -\lambda\partial_x^{2k+1}u(x_s,t) \end{pmatrix} \\ & +\frac{1}{\Delta x}\sum_{0\le k\le\lfloor\frac{n}{2}\rfloor}\psi_{2k}(x)\begin{pmatrix} -\lambda\partial_x^{2k}p(x_s,t) \\ \partial_x^{2k}p(x_s,t) \end{pmatrix} \\ & +\frac{1}{\Delta x}\sum_{n+1\le k<m}\psi_k(x)\partial_x^k\begin{pmatrix} u \\ p \end{pmatrix}(x_s,t). \end{aligned}$$

Getting the terms in g in the left side, one gets

$$\begin{cases} \overline{u}(x,t)-\frac{1}{\Delta x}\sum_{0\le k\le\lfloor\frac{n}{2}\rfloor}\psi_{2k}(x)\partial_t^{2k}g(t) & = \frac{1}{\Delta x}\sum_{0\le k\le\lfloor\frac{n-1}{2}\rfloor}\psi_{2k+1}(x)\partial_x^{2k+1}u(x_s,t) \\ & \quad -\frac{1}{\Delta x}\sum_{0\le k\le\lfloor\frac{n}{2}\rfloor}\lambda\psi_{2k}(x)\partial_x^{2k}p(x_s,t) \\ & \quad +\frac{1}{\Delta x}\sum_{n+1\le k<m}\psi_k(x)\partial_x^k u(x_s,t), \\ \overline{p}(x,t)+\frac{1}{\Delta x}\sum_{0\le k\le\lfloor\frac{n-1}{2}\rfloor}\psi_{2k+1}(x)\partial_t^{2k+1}g(t) & = \frac{1}{\Delta x}\sum_{0\le k\le\lfloor\frac{n}{2}\rfloor}\psi_{2k}(x)\partial_x^{2k}p(x_s,t) \\ & \quad -\frac{1}{\Delta x}\sum_{0\le k\le\lfloor\frac{n-1}{2}\rfloor}\lambda\psi_{2k+1}(x)\partial_x^{2k+1}u(x_s,t) \\ & \quad +\frac{1}{\Delta x}\sum_{n+1\le k<m}\psi_k(x)\partial_x^k p(x_s,t). \end{cases}$$

Then, it enables to establish a similar procedure to the one presented in section III-1.1. It writes

$$\begin{cases} \boldsymbol{\mathcal{U}}_- - \boldsymbol{\mathcal{S}}_-^n = \underline{\boldsymbol{\mathcal{Y}}}_-^{m,n}\cdot\boldsymbol{\Theta}, \\ \boldsymbol{\mathcal{U}}_+ - \boldsymbol{\mathcal{S}}_+^n = \underline{\boldsymbol{\mathcal{Y}}}_+^{m,n}\cdot\boldsymbol{\Theta}, \end{cases} \tag{III.39}$$

A similar proof as for Vandermonde matrices shows that $\underline{\boldsymbol{\mathcal{Y}}}_+^{m,n}$ is invertible for any (m,n) with

$0 \leq n < m$. Then after elimination of $\boldsymbol{\Theta}$ formed with spatial derivatives of u and p,

$$\boldsymbol{\mathcal{U}}_- = \boldsymbol{\mathcal{S}}_-^n + \underline{\boldsymbol{\mathcal{Y}}}_-^{m,n} \cdot (\underline{\boldsymbol{\mathcal{Y}}}_+^{m,n})^{-1} \cdot (\boldsymbol{\mathcal{U}}_+ - \boldsymbol{\mathcal{S}}_+^n). \tag{III.40}$$

The reconstruction operator writes $\underline{\boldsymbol{\mathcal{R}}}^{m,n} = \underline{\boldsymbol{\mathcal{Y}}}_-^{m,n} \cdot (\underline{\boldsymbol{\mathcal{Y}}}_+^{m,n})^{-1}$. One notices several differences compared with the previous example. For a velocity based boundary condition, the problem could be decoupled between solving a problem on p and later on u (or *vice versa*). Here, due to the particular boundary condition coupling both u and p, the obtained problem is solved simultaneously on both u and p and their derivatives.

Remark III.4. Straightforwardly, one deduces the reconstruction operator for boundary conditions imposed as $\lambda u + p = g$, or conditions imposed as $\mu u + \lambda p = g$ with $\lambda + \mu \neq 0$.

III-1.2.4 Experimental order of accuracy for a wave problem

We consider a C^∞ data solution to eq. (III.21) as

$$\begin{cases} u(t,x) &= a\sin(\omega(t+x)) + b\sin(\omega(t-x)), \\ p(t,x) &= -a\sin(\omega(t+x)) + b\sin(\omega(t-x)). \end{cases} \tag{III.41}$$

where arbitrarily, $a = 1$, $b = -1$, $\omega = 2\pi$. We consider a domain $\Omega = \{x \in \mathbb{R},\ x_s < x < 10\}$ where the boundary conditions on the right are imposed using the exact solution presented in eq. (III.41), and on the left, using the high-order accurate boundary treatment (according to the boundary condition) for $x = x_s = \frac{\sqrt{3}}{2}\ 10^{-3}$, so that the left boundary lies in the first cell. The CFL is set to 0.5. Computations are done in order to assess the accuracy of the proposed methods. First, the boundary treatment for boundary conditions on velocity is detailed, and its accuracy assessed with numerical experiments. Second, the boundary treatment for mixed boundary conditions is detailed, and the error as well as experimental order of convergence are presented.

Using boundary conditions on velocity

We consider here the initial data and boundary conditions on velocity for the IBVP as

$$\begin{cases} u(x,0) &= 2\sin(\omega x), \\ p(x,0) &= 0, \\ u(x_s,t) &= 2\sin(\omega x_s)\cos(\omega t). \end{cases} \tag{III.42}$$

In Table III.4, we present the l_1-error with respect to the number of cells for the $\underline{\boldsymbol{\mathcal{R}}}^{3,0}$,$\underline{\boldsymbol{\mathcal{R}}}^{3,1}$, and $\underline{\boldsymbol{\mathcal{R}}}^{3,2}$ reconstructions using the 3rd-order interior scheme presented in section II-2. The expected order of convergence for the third order staggered scheme is reached for the $\underline{\boldsymbol{\mathcal{R}}}^{3,1}$ and $\underline{\boldsymbol{\mathcal{R}}}^{3,2}$ reconstructions. Indeed, one can see that using $\underline{\boldsymbol{\mathcal{R}}}^{3,0}$ leads to an unstable effective scheme. We also have checked that modifying x_s does not alter the order of accuracy but slightly changes the

N_x	$\underline{\mathcal{R}}^{3,0}$		$\underline{\mathcal{R}}^{3,1}$		$\underline{\mathcal{R}}^{3,2}$	
20	1.3e-2	·	1.1e-3	·	1.5e-3	·
40	7.9e-4	3.99	8.9e-5	3.66	1.5e-4	3.39
80	4.9e-5	4.02	6.5e-6	3.78	1.3e-5	3.49
160	1.1e-5	2.14	5.1e-7	3.65	1.2e-6	3.43
320	8.5e-5	$\star$	5.1e-8	3.32	1.2e-7	3.31
640	1.1e-1	$\star$	6.2e-9	3.05	1.4e-8	3.18
1280	2.8e6	$\star$	7.9e-10	2.97	1.6e-9	3.08

Table III.4 – l^1-error and experimental order of convergence for the 3rd-order scheme together with the $\underline{\mathcal{R}}^{3,n}$ finite-volume reconstruction polynomial at $t = 0.3$ for boundary condition on the velocity. $\star$ are indications of unstable behaviour of the scheme.

N_x	$\underline{\mathcal{R}}^{3,0}$		$\underline{\mathcal{R}}^{3,1}$		$\underline{\mathcal{R}}^{3,2}$	
20	2.4e-2	·	1.4e-3	·	2.2e-3	·
40	2.3e-3	3.40	9.9e-5	3.84	2.4e-4	3.17
80	7.9e-5	4.88	8.6e-6	3.52	2.5e-5	3.24
160	1.1e-4	$\star$	8.3e-7	3.37	2.8e-6	3.16
320	2.5e-3	$\star$	8.1e-8	3.36	2.9e-7	3.27
640	1.1e5	$\star$	8.4e-9	3.28	3.0e-8	3.26
1280	$\star$	$\star$	9.3e-10	3.18	3.3e-9	3.21

Table III.5 – l^1-error and experimental order of convergence for the 3rd-order scheme together with the $\underline{\mathcal{R}}^{3,n}$ finite-volume reconstruction polynomial at $t = 0.3$ for mixed boundary condition ($\lambda = 1747$). $\star$ are indications of unstable behaviour of the scheme.

initial error level (for $N_x = 20$). Similar experimental orders of convergence for finite difference reconstruction operators have been recovered.

Using mixed boundary conditions

The initial data and mixed boundary conditions for the IBVP are

$$\begin{cases} u(x,0) & = & 2\sin(\omega x), \\ p(x,0) & = & 0 \\ u(x_s,t) + \lambda p(x_s,t) & = & (1-\lambda)\sin(\omega(t+x_s)) - (1+\lambda)\sin(\omega(t-x_s)), \end{cases} \tag{III.43}$$

with arbitrarily fix the parameter λ to $\lambda = 1747$. In Table III.5, we present the l_1-error with respect to the number of cells for the $\underline{\mathcal{R}}^{3,0}$,$\underline{\mathcal{R}}^{3,1}$, and $\underline{\mathcal{R}}^{3,2}$ reconstructions using the 3rd-order interior scheme presented in section II-2. The expected order of convergence for the third order staggered scheme is reached for the $\underline{\mathcal{R}}^{3,1}$ and $\underline{\mathcal{R}}^{3,2}$ reconstructions. Indeed, one can see that using $\underline{\mathcal{R}}^{3,0}$ leads to an unstable effective scheme. We also have checked that modifying x_s does not alter the order of accuracy but slightly changes the initial error level (for $N_x = 20$). Similar experimental orders of convergence for finite difference reconstruction operator have been recovered.

III-1.3 High-order reconstruction operator for general linear system

We extend the previous case to general hyperbolic linear system with boundary conditions. For linear hyperbolic system (III.1), one gets the following equality, assuming that $\underline{\boldsymbol{A}}$ is invertible,

$$\begin{cases} \partial_t^k \boldsymbol{U} = (-1)^k \underline{\boldsymbol{A}}^k \partial_x^k \boldsymbol{U}, \\ \partial_x^k \boldsymbol{U} = (-1)^k \underline{\boldsymbol{A}}^{-k} \partial_t^k \boldsymbol{U}. \end{cases} \tag{III.44}$$

Consider a mth-order scheme in both time and space and consider we use only the first n time derivatives of the boundary conditions $\boldsymbol{G}$, with $n < m$. Relation (III.44) is used to change time derivatives into space derivatives and *vice versa*. The matrix $\underline{\widehat{\boldsymbol{B}}} \in \mathbb{R}^{p\times p}$ is introduced as

$$\underline{\widehat{\boldsymbol{B}}} = \begin{pmatrix} \underline{\boldsymbol{B}} \\ \underline{\boldsymbol{0}} \end{pmatrix}.$$

Taylor expansion of $\overline{\boldsymbol{U}}$ for x in a neighborhood of x_s leads to

$$\begin{aligned} \overline{\boldsymbol{U}}(x,t) &= \frac{1}{\Delta x}\int_{x-\frac{\Delta x}{2}}^{x+\frac{\Delta x}{2}} \boldsymbol{U}(y,t)\mathrm{d}y = \frac{1}{\Delta x}\int_{x-\frac{\Delta x}{2}}^{x+\frac{\Delta x}{2}} \sum_{k\geq 0} \partial_x^k \boldsymbol{U}(x_s,t)\frac{(y-x_s)^k}{k!}\mathrm{d}y, \\ &= \frac{1}{\Delta x}\sum_{k\geq 0} \partial_x^k \boldsymbol{U}(x_s,t)\left(\frac{(x+\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!} - \frac{(x-\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!}\right). \end{aligned}$$

To alleviate the notations, let us introduce $\psi_k(x) = \dfrac{(x+\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!} - \dfrac{(x-\frac{\Delta x}{2}-x_s)^{k+1}}{(k+1)!}$. We have

$$\begin{aligned} \overline{\boldsymbol{U}}(x,t) &= \frac{1}{\Delta x}\sum_{0\leq k\leq n} (-1)^k \psi_k(x)(\underline{\boldsymbol{A}}^k)^{-1}\partial_t^k \boldsymbol{U}(x_s,t) + \frac{1}{\Delta x}\sum_{n+1\leq k} \psi_k(x)\partial_x^k \boldsymbol{U}(x_s,t), \\ &= \frac{1}{\Delta x}\sum_{0\leq k\leq n} \psi_k(x)(\underline{\boldsymbol{A}}^k)^{-1}(-1)^k \underline{\widehat{\boldsymbol{B}}}\partial_t^k \boldsymbol{U}(x_s,t) + \frac{1}{\Delta x}\sum_{0\leq k\leq n} \psi_k(x)\underline{\boldsymbol{A}}^{-k}(\underline{\boldsymbol{I}}-\underline{\widehat{\boldsymbol{B}}})\underline{\boldsymbol{A}}^k\partial_x^k \boldsymbol{U}(x_s,t) \\ &+ \frac{1}{\Delta x}\sum_{n+1\leq k} \psi_k(x)\partial_x^k \boldsymbol{U}(x_s,t). \end{aligned}$$

Truncating up to m^{th}-order, dropping the $\mathscr{O}(\Delta x^m)$ and using $\underline{\widehat{\boldsymbol{B}}}\boldsymbol{U}(x_s,t) = \underline{\widehat{\boldsymbol{B}}}\boldsymbol{G}(t)$, we get

$$\begin{aligned} \overline{\boldsymbol{U}}(x,t) &= \frac{1}{\Delta x}\sum_{0\leq k\leq n} \psi_k(x)(\underline{\boldsymbol{A}}^k)^{-1}(-1)^k \underline{\widehat{\boldsymbol{B}}}\partial_t^k \boldsymbol{G}(t) + \frac{1}{\Delta x}\sum_{0\leq k\leq n} \psi_k(x)\underline{\boldsymbol{A}}^{-k}(\underline{\boldsymbol{I}}-\underline{\widehat{\boldsymbol{B}}})\underline{\boldsymbol{A}}^k\partial_x^k \boldsymbol{U}(x_s,t) \\ &+ \frac{1}{\Delta x}\sum_{n+1\leq k<m} \psi_k(x)\partial_x^k \boldsymbol{U}(x_s,t). \end{aligned}$$

that is rewritten the following way

$$\overline{\boldsymbol{U}}(x,t) - \frac{1}{\Delta x}\sum_{0\le k\le n}\psi_k(x)(\underline{\boldsymbol{A}}^k)^{-1}(-1)^k\underline{\widehat{\boldsymbol{B}}}\partial_t^k\boldsymbol{G}(t) = \frac{1}{\Delta x}\sum_{0\le k\le n}\psi_k(x)\underline{\boldsymbol{A}}^{-k}(\underline{\boldsymbol{I}}-\underline{\widehat{\boldsymbol{B}}})\underline{\boldsymbol{A}}^k\partial_x^k\boldsymbol{U}(x_s,t)$$
$$+\frac{1}{\Delta x}\sum_{n+1\le k<m}\psi_k(x)\partial_x^k\boldsymbol{U}(x_s,t),$$

to establish a similar procedure to the one presented in section III-1.1. It writes

$$\begin{cases}\boldsymbol{\mathcal{U}}_- - \boldsymbol{\mathcal{S}}_-^n = \underline{\boldsymbol{\mathcal{Y}}}_-^{m,n}\cdot\boldsymbol{\Theta},\\ \boldsymbol{\mathcal{U}}_+ - \boldsymbol{\mathcal{S}}_+^n = \underline{\boldsymbol{\mathcal{Y}}}_+^{m,n}\cdot\boldsymbol{\Theta},\end{cases} \tag{III.45}$$

A similar proof as for Vandermonde matrices shows that $\underline{\boldsymbol{\mathcal{Y}}}_+^{m,n}$ is invertible for any (m,n) with $0\le n<m$. Then after elimination of $\boldsymbol{\Theta}$ formed with spatial derivatives of $\boldsymbol{U}$,

$$\boldsymbol{\mathcal{U}}_- = \boldsymbol{\mathcal{S}}_-^n + \underline{\boldsymbol{\mathcal{Y}}}_-^{m,n}\cdot(\underline{\boldsymbol{\mathcal{Y}}}_+^{m,n})^{-1}\cdot(\boldsymbol{\mathcal{U}}_+ - \boldsymbol{\mathcal{S}}_+^n). \tag{III.46}$$

Here again, the reconstruction operator writes $\underline{\boldsymbol{\mathcal{R}}}^{m,n} = \underline{\boldsymbol{\mathcal{Y}}}_-^{m,n}\cdot(\underline{\boldsymbol{\mathcal{Y}}}_+^{m,n})^{-1}$.

III-2 Stability of the inverse Lax–Wendroff procedure

We have seen in tables III.4 and III.5 that the third order scheme for the wave equation with the $\underline{\boldsymbol{\mathcal{R}}}^{3,0}$ is unstable, at least for the set of parameters used during the computations. Our purpose in this section is to establish the stable or unstable behaviour of the effective schemes.

In this section a procedure to study the stability of the reconstruction operator is developed. For any matrix $\underline{\boldsymbol{M}}$, $\rho(\underline{\boldsymbol{M}})$ denotes the spectral radius of $\underline{\boldsymbol{M}}$. Let $\underline{\boldsymbol{\mathcal{Z}}}$ denote the interior numerical scheme operator such that $\boldsymbol{\mathcal{U}}^{n+1} = \underline{\boldsymbol{\mathcal{Z}}}\boldsymbol{\mathcal{U}}^n$ solves (III.1). Let $\underline{\boldsymbol{\mathcal{R}}}$ denote the reconstruction operator such that $\boldsymbol{\mathcal{U}}_- = \underline{\boldsymbol{\mathcal{R}}}\boldsymbol{\mathcal{U}}_+$. The scheme writes

$$\begin{pmatrix}\boldsymbol{\mathcal{U}}_+\\ \boldsymbol{\mathcal{U}}_-\end{pmatrix}^{n+1} = \begin{pmatrix}\underline{\boldsymbol{\mathcal{Z}}}_{1,1} & \underline{\boldsymbol{\mathcal{Z}}}_{1,2}\\ \underline{\boldsymbol{\mathcal{Z}}}_{2,1} & \underline{\boldsymbol{\mathcal{Z}}}_{2,2}\end{pmatrix}\cdot\begin{pmatrix}\boldsymbol{\mathcal{U}}_+\\ \boldsymbol{\mathcal{U}}_-\end{pmatrix}^{n} = \begin{pmatrix}(\underline{\boldsymbol{\mathcal{Z}}}_{1,1}+\underline{\boldsymbol{\mathcal{Z}}}_{1,2}\underline{\boldsymbol{\mathcal{R}}})\boldsymbol{\mathcal{U}}_+^n\\ (\underline{\boldsymbol{\mathcal{Z}}}_{2,1}+\underline{\boldsymbol{\mathcal{Z}}}_{2,2}\underline{\boldsymbol{\mathcal{R}}})\boldsymbol{\mathcal{U}}_+^n\end{pmatrix}. \tag{III.47}$$

The reduced version where only $\boldsymbol{\mathcal{U}}_+^{n+1}$ shows up writes

$$\boldsymbol{\mathcal{U}}_+^{n+1} = (\underline{\boldsymbol{\mathcal{Z}}}_{1,1}+\underline{\boldsymbol{\mathcal{Z}}}_{1,2}\underline{\boldsymbol{\mathcal{R}}})\boldsymbol{\mathcal{U}}_+^n = \underline{\boldsymbol{\mathcal{N}}}\boldsymbol{\mathcal{U}}_+^n, \tag{III.48}$$

where $\underline{\boldsymbol{\mathcal{N}}} = (\underline{\boldsymbol{\mathcal{Z}}}_{1,1}+\underline{\boldsymbol{\mathcal{Z}}}_{1,2}\underline{\boldsymbol{\mathcal{R}}})$ is called the effective operator. The purpose of this section is first to study the stability of such an effective scheme, and later on to design a special criteria to characterize in a reduced sense the stability of this scheme.

III-2.1 GKS stability for IBVP using second order reconstruction for the Lax–Wendroff scheme

For this section, we consider the special case of the advection problem with boundary conditions written in eq. (III.3). As presented in section I-1.2.3, considering IBVP, an important feature of the effective scheme is its stability. The Cauchy stability analysis has already been mentioned for the interior schemes. To perform the GKS stability [76] analysis of a scheme, we first consider the second-order Lax–Wendroff projection scheme (presented in in eq. (III.7)) with the two proposed second-order reconstructions. We consider also that $g = 0$ which does not impact the linear stability analysis. The Lax–Wendroff scheme requires only one ghost-cell value to the left of the boundary. The reconstructions are

$$\underline{\boldsymbol{\mathcal{R}}}^{2,0} = \mathcal{R}_0 = \frac{\sigma}{\sigma+1} \quad \text{and} \quad \underline{\boldsymbol{\mathcal{R}}}^{2,1} = \mathcal{R}_1 = 0. \tag{III.49}$$

Proposition III.4 (GKS stability of the Lax–Wendroff scheme)**.** *The Lax–Wendroff scheme is stable in the sense of lemma I.12 using $\mathcal{R}_0$ or $\mathcal{R}_1$ defined in eq.* (III.49) *for $\nu \in [0:1]$, $\sigma \in \left[-\frac{1}{2}, \frac{1}{2}\right[$.*

Proof. From linear stability analysis, one gets the characteristic equation for the Lax–Wendroff equation which is

$$z\kappa = \frac{\nu^2+\nu}{2} + (1-\nu^2)\kappa + \frac{\nu^2-\nu}{2}\kappa^2. \tag{III.50}$$

Let $f(\kappa)$ defined as

$$f(\kappa) = \frac{\nu^2+\nu}{2} + (1-\nu^2-z)\kappa + \frac{\nu^2-\nu}{2}\kappa^2. \tag{III.51}$$

One gets from linear stability analysis of the interior scheme that for κ satisfying $f(\kappa) = 0$ and $|\kappa| = 1$ that $|z| \leq 1$ for $\nu \in [0:1]$. Then, the number of roots with $|\mathcal{K}| < 1$ of the characteristics equation is independent of the value of z. Thus, one may choose any z such that $|z| > 1$ to determine the number of roots κ such that $|\kappa| < 1$. Arbitrarily we set $z = 2$, it yields that

$$f(\kappa) = \frac{\nu^2+\nu}{2} + (-1-\nu^2)\kappa + \frac{\nu^2-\nu}{2}\kappa^2,$$

from which one deduces that

$$\begin{cases} \kappa_1(\nu) = \dfrac{1+\nu^2-\sqrt{1+3\nu^2}}{\nu^2-\nu}, \\ \kappa_2(\nu) = \dfrac{1+\nu^2+\sqrt{1+3\nu^2}}{\nu^2-\nu}. \end{cases} \tag{III.52}$$

In particular, one gets that for

$$\nu \in [0:1] \begin{cases} \kappa_1(\nu) \in \left[0:\frac{1}{2}\right], \\ \kappa_2(\nu) \notin [-1:1]. \end{cases} \tag{III.53}$$

It thus implies that trivially the roots are distinct. If one consider now that $z = \mathrm{e}^{ik}$ with $k \in \mathbb{R}$,

one get two roots κ_1, κ_2, with for certain values of k that $|\kappa_1| = 1$. A perturbation analysis is then performed. To illustrate the perturbation analysis, assume that $z = 1$, then one gets that $\kappa_1 = 1$, $\kappa_2 = -\frac{1+\nu}{1-\nu}$. Then the perturbation analysis consists of considering that now $z = \delta$ and $\kappa = 1 + \epsilon$ inside the characteristic equation (III.50). One obtains

$$\delta = 1 - \frac{\nu(\epsilon^2 + 2\epsilon) - \epsilon^2\nu^2}{2\epsilon + 2},$$

which leads to, performing a Taylor expansion at $\epsilon = 0$,

$$\delta = 1 - \epsilon\nu + \mathscr{O}(\epsilon^2),$$

which proves that $\kappa = 1$ is stable under perturbation as for ϵ small enough, $\delta < 1$. Then to get the non-existence of generalized eigensolution, one must verify that there is no solution to

$$\left(\frac{\nu^2 + \nu}{2}\mathcal{R} + (1 - \nu^2 - z)\right)\kappa^2 + \frac{\nu^2 - \nu}{2}\kappa^2 = 0, \tag{III.54}$$

for $\nu \in [0 : 1]$, $\sigma \in \left[-\frac{1}{2}, \frac{1}{2}\right[$, $|z| \geq 1$, $\mathcal{R} = \mathcal{R}_0$ or $\mathcal{R} = \mathcal{R}_1$ as defined in eq. (III.49) and κ satisfying the characteristic equation (III.50). The system has no solution. Thus, there is no generalized eigensolution and the scheme is linearly stable. ■

Similar studies can be perform for the Beam–Warming scheme. Increasing the order of the scheme and of the reconstruction yields more and more complexity of the fully discrete GKS analysis. Thus, a criteria is introduced (very similar to the one proposed in [168]) to alleviate the algebra of the GKS stability. The cost of such a criteria is that it does not give strong results concerning the linear stability of the effective scheme.

III-2.2 Reduced stability for IBVP discretization

Let us consider now general linear hyperbolic system with appropriate boundary conditions written ineq. (III.1). Here, we add an *a priori* requirement of this stability. We will set $N_{n_c} \in \mathbb{R}^{n_c^2}, N_{n_c} = \mathcal{P}_{n_c}\mathcal{N}\mathcal{P}_{n_c}^t$ where $\mathcal{P}_{n_c}$ is the natural projection such that for $\mathcal{X} \in l^2$, $\mathcal{P}_{n_c}\mathcal{X} = (X_1, ..., X_{n_c}) \in \mathbb{R}^{n_c}$.

Definition III.1 (Reduced stability)**.** Let $\mathcal{Z}$ be the interior scheme, and $\mathcal{R}$ the reconstruction operator. The operator $\mathcal{N} = (\mathcal{Z}_{1,1} + \mathcal{Z}_{1,2}\mathcal{R})$ is stable in a reduced sense if

1. $\mathcal{Z}$ is stable using normal mode analysis [22, 2],
2. There exists $n_c \in \mathbb{N}^*$ such that $\rho(N_{n_c}) \leq 1$.

Remark III.5. Definition III.1 provides practical information concerning the stability of the final scheme and is used to determine *a priori* if a reconstruction is unstable by taking n_c large enough.

III-2.2.1 Analytic reduced stability of the Beam–Warming scheme

The Beam–Warming scheme presented in eq. (III.8), linearly stable for $\nu \in [0:2]$ writes as

$$u_i^{n+1} = (1 + \frac{\nu^2 - 3\nu}{2})u_i^n + (2\nu - \nu^2)u_{i-1}^n + \frac{\nu^2 - \nu}{2}u_{i-2}^n, \quad i \in \mathbb{Z}.$$

Considering that the boundary condition g satisfies $g = 0$, and taking $m = 2, n = 0$, it yields that

$$\underline{\boldsymbol{\mathcal{R}}} = \begin{pmatrix} \frac{\sigma}{\sigma - 1} \\ \frac{\sigma + 1}{\sigma - 1} \end{pmatrix}.$$

Then, the effective scheme writes

$$\begin{cases} u_1^{n+1} &= \dfrac{\sigma + \nu - 1}{\sigma - 1}u_1^n, \\ u_2^{n+1} &= \dfrac{3\sigma\nu - \sigma\nu^2 - 4\nu + 2\nu^2}{2\sigma - 2}u_1^n + (1 + \dfrac{\nu^2 - 3\nu}{2})u_2^n, \\ u_i^{n+1} &= (1 + \dfrac{\nu^2 - 3\nu}{2})u_i^n + (2\nu - \nu^2)u_{i-1}^n + \dfrac{\nu^2 - \nu}{2}u_{i-2}^n, \quad i > 2. \end{cases} \tag{III.55}$$

It is possible to rewrite the previous system under the form $\boldsymbol{\mathcal{U}}_+^{n+1} = \underline{\boldsymbol{\mathcal{N}}}\boldsymbol{\mathcal{U}}_+^n$ where the operator $\underline{\boldsymbol{\mathcal{N}}}$ satisfies

$$\underline{\boldsymbol{\mathcal{N}}} = \begin{pmatrix} \frac{\sigma+\nu-1}{\sigma-1} & 0 & 0 & \dots \\ \frac{3\sigma\nu-\sigma\nu^2-4\nu+2\nu^2}{2\sigma-2} & (1+\frac{\nu^2-3\nu}{2}) & 0 & \dots \\ \frac{\nu^2-\nu}{2} & (2\nu-\nu^2) & (1+\frac{\nu^2-3\nu}{2}) & 0 \\ 0 & \ddots & \ddots & \ddots \end{pmatrix}. \tag{III.56}$$

It leads to the following proposition

Proposition III.5. *The operator $\underline{\boldsymbol{\mathcal{N}}}$ given in eq. (III.56) is stable in the sense of the reduced stability defined in definition III.1.*

Proof. Let p be an integer. Let us introduce the operator $\underline{\boldsymbol{\mathcal{P}}}_p$ such that

$$\forall \boldsymbol{\mathcal{U}} \in l^2, \underline{\boldsymbol{\mathcal{P}}}_p\boldsymbol{\mathcal{U}} = (u_1, ..., u_p)^t \in \mathbb{R}^p,$$

and the operator $\underline{\boldsymbol{\mathcal{Q}}}_p$ such that

$$\forall (u_1, ..., u_p)^t \in \mathbb{R}^p, \underline{\boldsymbol{\mathcal{Q}}}(u_1, ..., u_p)^t = (u_1, ..., u_p, 0, ...)^t \in l^2.$$

Let the matrix $\underline{\boldsymbol{N}}_p = \underline{\boldsymbol{\mathcal{P}}}_p\underline{\boldsymbol{\mathcal{N}}}\underline{\boldsymbol{\mathcal{Q}}}_p \in \mathbb{R}^{p\times p}$. The spectrum of the matrix $\underline{\boldsymbol{N}}_p$ writes

$$Sp(\underline{\boldsymbol{N}}_p) = \{1 + \frac{\nu^2 - 3\nu}{2}, \frac{\sigma + \nu - 1}{\sigma - 1}\}.$$

Now, we wish to exhibit condition on ν depending on σ such that

$$|\frac{\sigma+\nu-1}{\sigma-1}| \leq 1.$$

As $\sigma \in \left[-\frac{1}{2} : \frac{1}{2}\right[$, it yields that $\sigma - 1 < 0$, and thus it writes

$$\sigma - 1 \leq \sigma + \nu - 1 \leq 1 - \sigma, \quad \sigma \in \left[-\frac{1}{2} : \frac{1}{2}\right[,$$

which yields

$$0 \leq \nu \leq 2 - 2\sigma, \quad \sigma \in \left[-\frac{1}{2} : \frac{1}{2}\right[.$$

Taking the minimum over σ on the right hand side, it yields

$$0 \leq \nu \leq 1.$$

Hence the result. ■

III-2.2.2 Numerical reduced stability results for the high-order Strang projection schemes

We illustrate this definition by taking the O3 scheme (III.9) with the reconstruction (III.14). The interior operator $\mathcal{Z}$ writes as a band matrix whose coefficients are for any $i \in \mathbb{Z}$

$$\mathcal{Z}_{i,i-2} = \frac{\nu^3}{6} - \frac{\nu}{6}, \quad \mathcal{Z}_{i,i-1} = \nu + \frac{\nu^2}{2} - \frac{\nu^3}{2}, \quad \mathcal{Z}_{i,i} = \frac{\nu^3}{2} - \nu^2 - \frac{\nu}{2} + 1, \quad \mathcal{Z}_{i,i+1} = -\frac{\nu^3}{6} + \frac{\nu^2}{2} - \frac{\nu}{3}.$$

The reconstruction for $m = 2$ and $n = 1$ writes

$$\mathcal{R}_{1,1} = \frac{12\sigma^2 + 1}{12\sigma^2 - 24\sigma + 13}, \quad \mathcal{R}_{2,1} = \frac{12\sigma^2 + 24\sigma + 13}{12\sigma^2 - 24\sigma + 13}.$$

To alleviate notations and since the interior operator is a band matrix, we denote $C_j = \mathcal{Z}_{i,i+j}$ for any $j \in \mathbb{Z}$. Combining both, operator $\mathcal{N}$ writes

$$\mathcal{N} = \begin{pmatrix} C_{-2}\mathcal{R}_{2,1} + C_{-1}\mathcal{R}_{1,1} + C_0 & C_1 & 0 & 0 & 0 \\ C_{-2}\mathcal{R}_{1,1} + C_{-1} & C_0 & C_1 & 0 & 0 \\ C_{-2} & C_{-1} & C_0 & C_1 & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots \end{pmatrix}. \tag{III.57}$$

One then checks numerically if the spectral radius of N_{n_c} is less or equal to one. Remind that $\mathcal{R}^{m,n}$ denotes the mth-order reconstruction operator that takes into account the n first time derivatives of the boundary condition. Results for the Lax–Wendroff (fig. III.2), the Beam–Warming (fig. III.3), the third order projection (fig. III.4) and the fourth order one (fig. III.5) are depicted. Those results highlight the areas of reduced stabilities. Parts of the considered space (ν, σ) where the scheme is stable in a reduced sense are in white and in black otherwise.

In particular, it means that second order reconstruction are unconditionally stable (in the sense of definition III.1) for the Lax–Wendroff and the Beam–Warming scheme. As a contrary, third order reconstruction, with $n = 0$ does not satisfy the reduced stability condition for certain values of ν and σ. As a matter of fact, for such values of ν and σ, a fully discrete GKS stability analysis proves the existence of generalized eigensolution. Moreover numerical experiments using values of ν and σ in this area highlight the unstable behaviour of the effective scheme.

Notice that on fig. III.2, the reduced stability results and the results obtained for the fully discrete GKS analysis presented in proposition III.4 are identical. As well, on fig. III.3, the reduced stability and the results presented in proposition III.5 are the same. It assesses the practical relevance of the reduced stability criterion.

Furthermore, an interesting feature is shown in fig. III.5, where one notices that for the $\mathcal{R}^{4,0}$, the bottom left corner of the (ν, σ)-space is unstable. It means in peculiar that the whole space (ν, σ) must be treated in order to get a complete idea of the effective scheme stability. Indeed, one may not consider that if a scheme is stable for $\nu = \nu_1$, then for any $\nu < \nu_1$ the scheme is also stable.

Drawing comparisons between reduced stability results and results obtained by performing a numerical fully discrete GKS analysis for the advection problem using inverse Lax–Wendroff procedure and projection scheme gives very similar results. As we use the reduced stability definition to choose which reconstruction operator to obtain a stable effective scheme, this is, to our opinion, a sufficient criteria. Therefore in the following for the wave equations, only reduced stability is studied, and the complete fully discrete GKS analysis is not performed.

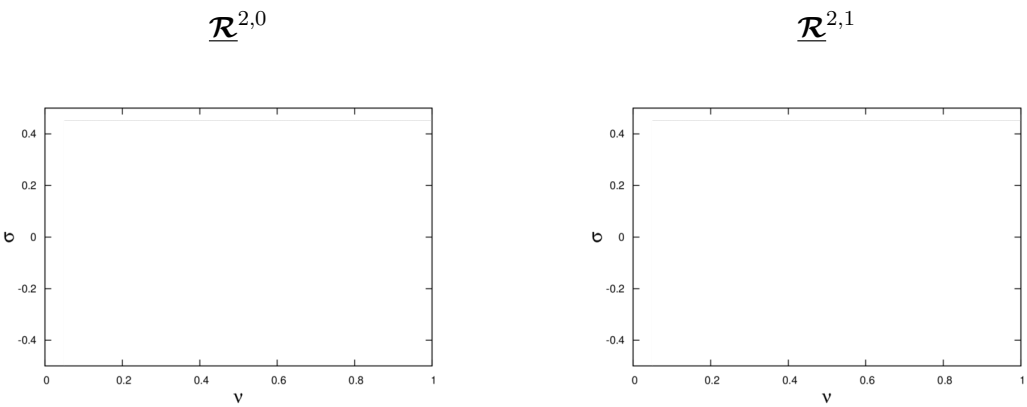


Figure III.2 – Stability area $\{(\nu, \sigma) \ / \ \rho(N_{nc}) \leq 1\}$ (in white) for the Lax–Wendroff (second order) scheme with $n_c = 20$ for the $\underline{\mathcal{R}}^{2,0}$ (left), $\underline{\mathcal{R}}^{2,1}$ (right) reconstruction operators. The whole domain is stable.

III-2.2.3 Numerical reduced stability results for the Runge–Kutta based staggered scheme for the wave equations

Similarly to the advection equation, we perform a numerical study of the reduced stability of the Runge–Kutta based staggered scheme for the wave equations for boundary condition on velocity.

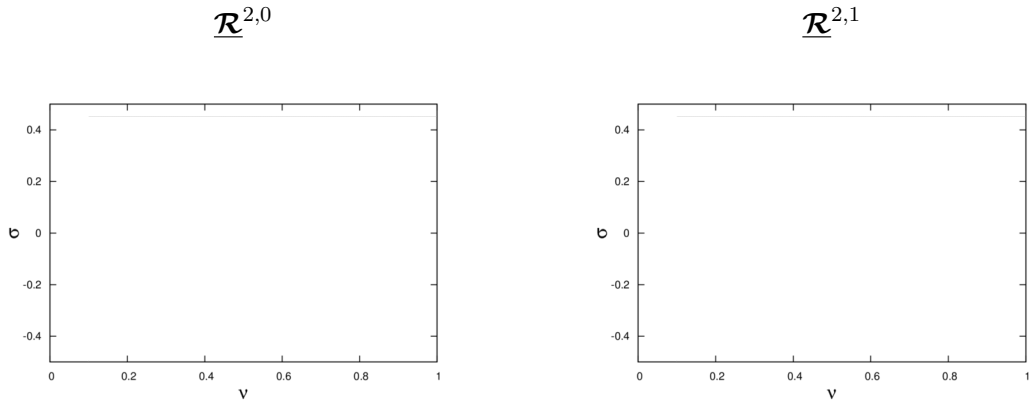


Figure III.3 – Stability area $\{(\nu, \sigma) \ / \ \rho(N_{nc}) \leq 1\}$ (in white) for the Beam–Warming (second order) scheme with $n_c = 20$ for the $\underline{\mathcal{R}}^{2,0}$ (left), $\underline{\mathcal{R}}^{2,1}$ (right) reconstruction operators. The whole domain is stable.

In fig. III.6, parts of the considered space (ν, σ) where the scheme is stable in a reduced sense are depicted in white and in black otherwise. One notices that using only g yields an effective scheme that does not satisfy the reduced stability definition. It has already been numerically checked on an example in Table III.4. As a contrary, considering more derivatives of g, the effective schemes fully satisfy the reduced stability definition.

The reduce stability study for the wave equations determines that until third order of accuracy, g, $\mathrm{D}_t g$ are required for linear stability of the initial boundary value problem. The next chapter is dedicated to the study in the case of the Lagrange-remap hydrodynamics system. Using the previous results, only g and $\mathrm{D}_t g$ are going to be used in the Inverse Lax–Wendroff procedure.

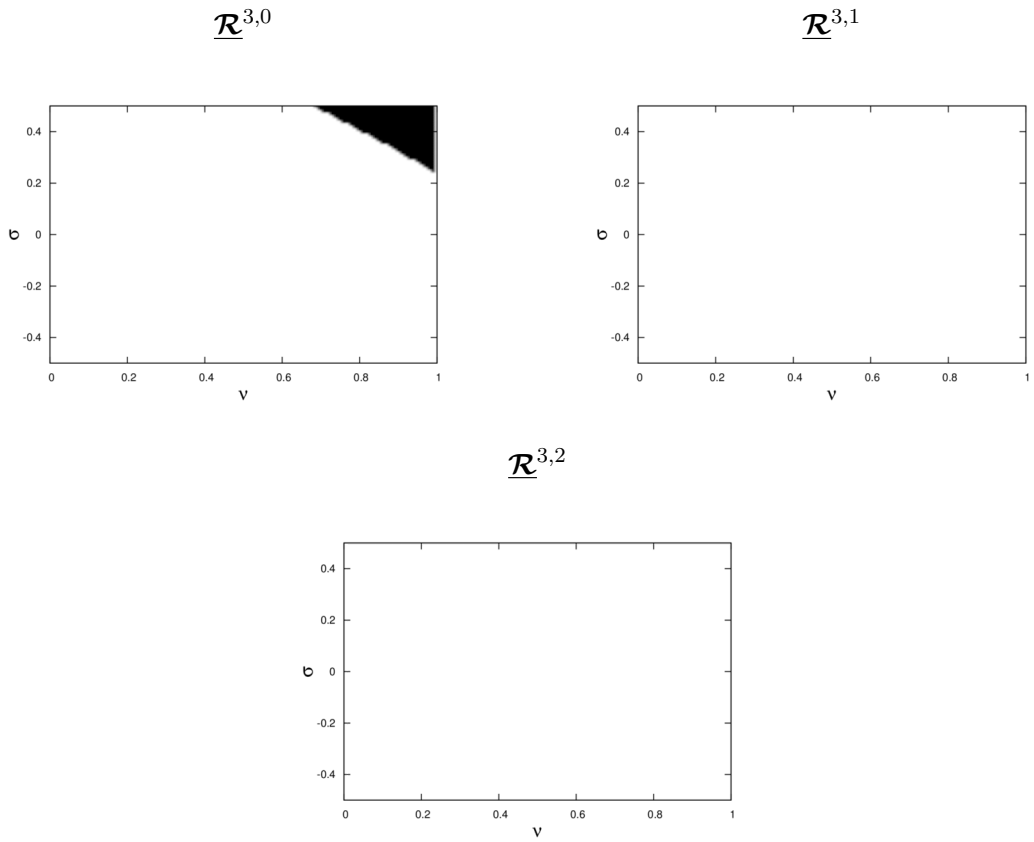


Figure III.4 – Stability area $\{(\nu, \sigma) \ / \ \rho(N_{nc}) \leq 1\}$ (in white) for the third-order projection scheme with $n_c = 20$ for the $\underline{\mathcal{R}}^{3,0}$ (top, left), $\underline{\mathcal{R}}^{3,1}$ (top, right) and $\underline{\mathcal{R}}^{3,2}$ (bottom) reconstruction operators. As a contrary to figs. III.2 and III.3, one notices a region of numerical instability for $\underline{\mathcal{R}}^{3,0}$.

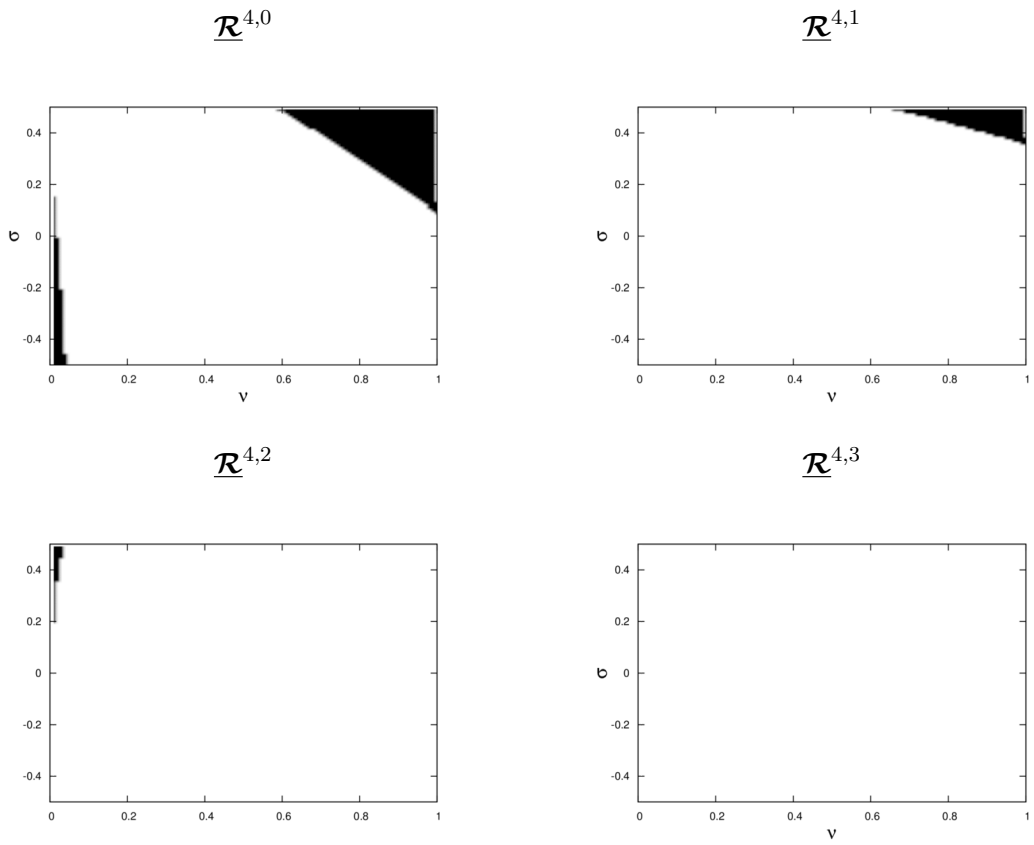


Figure III.5 – Stability area $\{(\nu, \sigma) \ / \ \rho(N_{nc}) \leq 1\}$ (in white) for the fourth-order projection scheme with $n_c = 30$ for the $\mathcal{R}^{4,0}$ (top, left), $\mathcal{R}^{4,1}$ (top, right), $\mathcal{R}^{4,2}$ (bottom, left), $\mathcal{R}^{4,3}$ (bottom, right) reconstruction operators. An additional behaviour is observed w.r.t. fig. III.4 which is that the domain of instability contains a layer for small value of ν ($\underline{\boldsymbol{\mathcal{R}}}^{4,0}$ and $\underline{\boldsymbol{\mathcal{R}}}^{4,2}$)

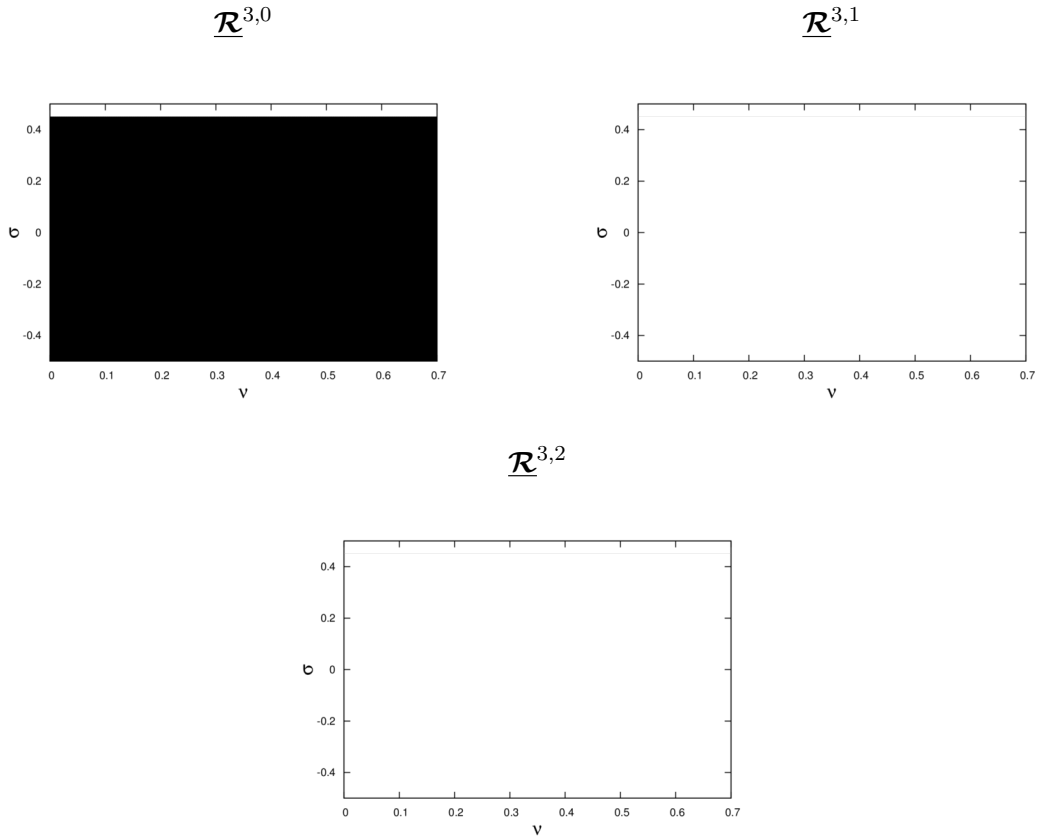


Figure III.6 – Stability area $\{(\nu, \sigma) \ / \ \rho(N_{nc}) \leq 1\}$ (in white) for the third-order staggered scheme for the wave equations with $n_c = 40$ for the $\mathcal{R}^{3,0}$ (top, left), $\mathcal{R}^{3,1}$ (top, right) and $\mathcal{R}^{3,2}$ (bottom) reconstruction operators.

Chapter IV

Discretization of boundary conditions for compressible hydrodynamics

---

En partant de la méthode de Lax–Wendroff inverse développée au chapitre précédent pour des systèmes linéaires, on propose dans ce chapitre une extension au système non linéaire de l'hydrodynamique compressible, en traitant la difficulté majeure qui est que la jacobienne du système Lagrangien possède une valeur propre nulle. Des schémas centrés sont considérés pour la résolution de l'hydrodynamique afin de simplifier la présentation et la construction de la méthode. Après une courte introduction concernant la particularité du système lagrangien 1D des équations de l'hydrodynamique compressible, un problème à l'ordre 2 et à masse constante est isolé et traité de deux façons différentes. Dans un premier temps, une hypothèse est faite sur la nature des écoulements proches de la frontière afin de se rapprocher le plus possible du cas linéaire de l'équation des ondes. Dans un second temps, aucune hypothèse n'est faite sur la nature des écoulements et l'impact sur la stabilité linéaire est étudiée numériquement. Puis, la détermination de la structure de l'opérateur de reconstruction aux bords est étendue au cas de problèmes à masse variable et à l'ordre élevé. Les résultats principaux se situent dans les lemmes IV.4, IV.5 et IV.6 qui caractérisent les conditions d'existence et d'unicité de l'opérateur de reconstruction. Une procédure de type MOOD est établie afin de garantir la robustesse de la reconstruction dans le cas de chocs forts. Enfin, une extension 2D des opérateurs de reconstruction est proposée. Des résultats numériques sont proposés tout au long du chapitre afin d'illustrer la précision, la stabilité et la robustesse de la méthode décrite. Une partie des résultats obtenus a été soumise à une revue scientifique [34]. Une dernière section est consacrée à l'adaptation de la procédure de discrétisation des conditions aux bords pour les schémas décalés.

New high-order accurate methods to take into account boundary conditions for hyperbolic equations, based on the so-called inverse Lax-Wendroff (ILW) procedure (see section I-3.2.3) have been recently published. The study addressed in this work aims at extending these methods to the Lagrange-remap discretization of the model 2D Euler system (IV.1) involving complex (eventually moving) boundaries

$$\left\{\begin{array}{rl} \partial_t\rho + \partial_x(\rho u) + \partial_y(\rho v) & =0, \\ \partial_t(\rho u) + \partial_x(\rho u^2 + p) + \partial_y(\rho uv) & =0, \\ \partial_t(\rho v) + \partial_x(\rho uv) + \partial_y(\rho v^2 + p) & =0, \\ \partial_t(\rho e) + \partial_x(\rho ue + pu) + \partial_y(\rho ve + pv) & =0. \end{array}\right. \tag{IV.1}$$

Variables ρ, $\tau = \frac{1}{\rho}$, e, p, u, v respectively denote the density, specific volume, total energy, pressure, x-velocity and y-velocity and eq. (IV.1) is closed with an arbitrary equation of state $p = EOS(\tau = 1/\rho, e, u, v)$. Introducing $\boldsymbol{U} = (\rho, \rho u, \rho v, \rho e)^t$, system (IV.1) rewrites as a general hyperbolic system of conservation laws

$$\partial_t \boldsymbol{U} + \partial_x \boldsymbol{F}(\boldsymbol{U}) + \partial_y \boldsymbol{G}(\boldsymbol{U}) = \boldsymbol{0},\ t \geq 0,\ (x, y) \in \Omega. \tag{IV.2}$$

Let $\Omega \subset \mathbb{R}^2$ be the "fluid domain". Boundary conditions are added along a curve $\Gamma(t)$, $t \geq 0$. In this paper we focus on imposed velocity boundary conditions for inviscid flows, so that only the normal velocity on $\Gamma(t)$ is prescribed

$$(u, v) \cdot \overrightarrow{n}(t, s) = g(t, s),\ t \geq 0,\ (x, y) \in \Gamma(t), \tag{IV.3}$$

where s is the curvilinear coordinate along the boundary $\Gamma(t)$, and $\overrightarrow{n}(t, s)$ denotes the normal to the curve at coordinate s and time t. The domain Ω is defined as the outside of the volume delimited by Γ. In numerical algorithms, $\Gamma(t)$ is approximated by $\Gamma_{\Delta s}$ as depicted in figure IV.1. In this work, we will consider that $\Gamma_{\Delta s}$ is formed as a necklace of pearls P_s without any hypothesis on how to link two consecutive pearls. Only full fluid cells are considered to be part of the "fluid" computational domain denoted $\Omega_+ \subset \Omega$. Cells in gray are considered as part of the "ghost" computational domain denoted Ω_-. In practice, one has $\Omega \subset \Omega_+ \cup \Omega_- \subset \mathbb{R}^2$. The algorithm proposed in this work builds ghost values in Ω_- such that the resulting scheme is both high-order accurate and stable.

To build ghost values, which is ultimately the real problem, one has in hands the boundary conditions and any kind of extrapolation technique to reconstruct $\boldsymbol{\mathcal{U}}_- = (\boldsymbol{U}_j)_{j\in\Omega_-}$ from $\boldsymbol{\mathcal{U}}_+ = (\boldsymbol{U}_j)_{j\in\Omega_+}$. Therefore the problem discussed hereafter can be formulated as follows

Problem IV.1. Build an operator $\underline{\boldsymbol{\mathcal{R}}}$

$$\begin{array}{rccc} \underline{\boldsymbol{\mathcal{R}}}: & (\mathbb{R}^4)^{\mathbf{card}(\Omega_+)} & \longrightarrow & (\mathbb{R}^4)^{\mathbf{card}(\Omega_-)} \\ & \boldsymbol{\mathcal{U}}_+ & \longmapsto & \boldsymbol{\mathcal{U}}_-, \end{array} \tag{IV.4}$$

such that the coupling with the internal scheme (in Ω_+) is stable and a high-order approximation of (IV.2-IV.3).

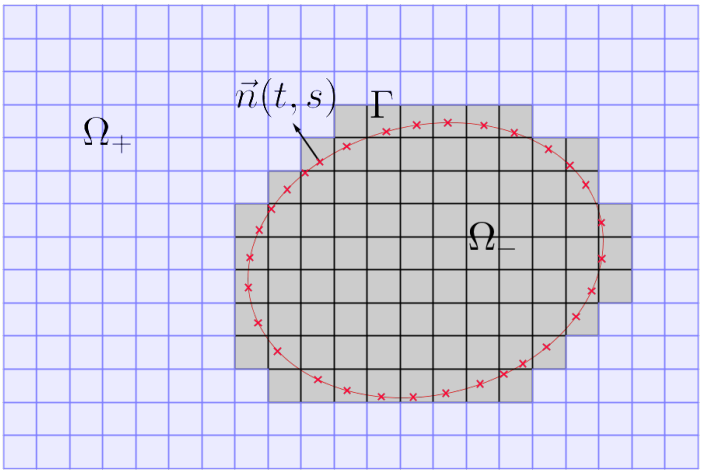


Figure IV.1 – Discretization $\Gamma_{\Delta s}$ of $\Gamma(t)$ and decomposition of the whole domain between Ω_- (ghost-cells) and Ω_+ (fluid cells). Ω is the domain outside the ellipse.

This work is part of a submitted publication [34].

IV-1 ILW procedure for the 1D Lagrangian system 157
IV-1.1 An instructive second-order boundary treatment 158
IV-1.2 General procedure, and characterization of the solution for the system at the boundary 164
IV-1.3 Stabilization procedure for shocks and very high-order reconstruction . . 167
IV-1.4 1D validation and comparisons 168
IV-2 Extension of the ILW procedure to the 2D Euler system 171
IV-2.1 Formulation of the ILW procedure using directionnal splitting 173
IV-2.2 2D numerical validation 177

IV-1 ILW procedure for the 1D Lagrangian system

So far, the reconstruction method has been described in section III-1.3 for linear hyperbolic system with $\underline{\boldsymbol{A}}$ invertible. Our interest now lies in its derivation and application for non-linear systems, and especially the 1D Euler system. We recall that ρ, τ, u, p and e respectively describe the density, specific volume, velocity, pressure and total energy. The 1D Euler system writes

$$\begin{cases} \partial_t \rho + \partial_x (\rho u) &= 0, \\ \partial_t (\rho u) + \partial_x \left(\rho u^2 + p\right) &= 0, \\ \partial_t (\rho e) + \partial_x (\rho u e + p u) &= 0, \end{cases} \tag{IV.5}$$

closed with the equation of state (EOS) $p = EOS(\tau = 1/\rho, e, u)$. System (IV.5) is solved with a Lagrange-remap scheme. Let ρ_0 denote the initial mass density. Introducing the $(x,t) \to (X,t)$

variable change such that $\rho dx = \rho_0 dX$, (IV.5) rewrites

$$\left\{\begin{array}{rcl} \mathrm{D}_t\left(\rho_0\tau\right) - \partial_X u & = & 0, \\ \mathrm{D}_t\left(\rho_0 u\right) + \partial_X p & = & 0, \\ \mathrm{D}_t\left(\rho_0 e\right) + \partial_X pu & = & 0, \end{array}\right. \tag{IV.6}$$

in Lagrangian coordinates. The Lagrange-remap method consists in the following two steps for integrating (IV.5). Let $\rho_0(x) = \rho(x, t^n)$, *ie* the regular Eulerian and Lagrangian grids $x_{i+\frac{1}{2}}$ and $X_{i+\frac{1}{2}}$ coincide at time t^n. First, system (IV.6) is time-integrated to give Lagrangian conservative variables at time t^{n+1} on a non-uniform grid. These variables are then remapped on the initial grid, leading to Eulerian conservative variables at time t^{n+1}. For the Lagrange system (IV.6), the flux is $\boldsymbol{F}(\boldsymbol{U}) = (-u, p, pu)^t$ and its jacobian $\underline{\boldsymbol{A}} = \underline{\boldsymbol{\nabla}}_U \boldsymbol{F}(\boldsymbol{U})$ writes

$$A = \begin{pmatrix} 0 & -\dfrac{1}{\rho_0} & 0 \\ \dfrac{\partial p}{\partial \rho_0 \tau} & \dfrac{\partial p}{\partial \rho_0 u} & \dfrac{\partial p}{\partial \rho_0 e} \\ u\dfrac{\partial p}{\partial \rho_0 \tau} & \dfrac{p}{\rho_0} + u\dfrac{\partial p}{\partial \rho_0 u} & u\dfrac{\partial p}{\partial \rho_0 e} \end{pmatrix}. \tag{IV.7}$$

The matrix $\underline{\boldsymbol{A}}$ admits three eigenvalues $\lambda_1 > 0, \lambda_2 = 0, \lambda_3 = -\lambda_1$ and is therefore non-invertible. Due to the sign of the eigenvalues, only one boundary condition is to be set in $x = x_s$ and we choose to prescribe the normal velocity as in eq. (IV.3). It writes

$$u(x_s(t), t) = g(t) \text{ or, in Lagrangian coordinates } u(X_s, t) = g(t). \tag{IV.8}$$

We present in the following two methods that are based on two different point of views. The first point of view is to include in the system of partial derivative equations another equation which is the entropy equation. The second one is to focus on the set of data inside the computation. But first, the emphasis is laid on a simplified second order problem at the boundary, which highlights both point of views.

IV-1.1 An instructive second-order boundary treatment

To give insights into existence of a solution and explain how we proceed, we here focus on a sample problem in which we assume a constant initial mass density $\rho_0 = 1$, a perfect gas EOS and a second-order treatment of the boundary condition. We drop the time variable to alleviate notations. Dropping also the $\mathscr{O}((X - X_s)^2)$ term, the truncated Taylor expansions of (τ, u, e) at second order writes

$$\left\{\begin{array}{rcl} \tau(X_s) + \partial_X \tau(X_s)(X - X_s) & = & \tau(X), \\ u(X_s) + \partial_X u(X_s)(X - X_s) & = & u(X), \\ e(X_s) + \partial_X e(X_s)(X - X_s) & = & e(X). \end{array}\right. \tag{IV.9}$$

In order to apply the previously described method, variables in X_s must be known. The boundary conditions writes $u(X_s) = g$ and the equation of state writes $p = p(\tau, e, u)$. Using the Euler

equation in Lagrangian coordinates for the momentum, one gets that

$$\rho_0 \mathrm{D}_t u = -\partial_x p,$$

which rewrites inserting the equation of state, and using $\rho_0 = 1$ as

$$\mathrm{D}_t u = -\partial_X p(\tau, e, u).$$

Using the chain rule, it leads to

$$\mathrm{D}_t u = -\partial_X \tau \partial_\tau p(\tau, e, u) - \partial_X e \partial_e p(\tau, e, u) - \partial_X u \partial_u p(\tau, e, u),$$

thus inserting the boundary condition $\mathrm{D}_t u(X_s) = \mathrm{D}_t g$,

$$\mathrm{D}_t g = -\partial_X \tau(X_s)\partial_\tau p(X_s) - \partial_X e(X_s)\partial_e p(X_s) - \partial_X u(X_s)\partial_u p(X_s). \tag{IV.10}$$

Then, we get the following system

$$\left\{\begin{array}{rcl} \tau(X_s) + \partial_X \tau(X_s)(X - X_s) & = & \tau(X), \\ u(X_s) + \partial_X u(X_s)(X - X_s) & = & u(X), \\ e(X_s) + \partial_X e(X_s)(X - X_s) & = & e(X), \\ u(X_s) & = & g, \\ \partial_X \tau(X_s)\partial_\tau p(X_s) + \partial_X e(X_s)\partial_e p(X_s) + \partial_X u(X_s)\partial_u p(X_s) & = & -\mathrm{D}_t g, \end{array}\right. \tag{IV.11}$$

whose unknowns are $\tau(X_s), \partial_X \tau(X_s), u(X_s), \partial_X u(X_s), e(X_s), \partial_X e(X_s)$.

IV-1.1.1 First method: the spatially isentropic flow hypothesis

The system (IV.11) needs one more equation, to get 6 equations for 6 unknowns. The first method is based on the choice of an hypothesis on the flow structure near the boundary. A spatially isentropic flow near the boundary is assumed. We use the second law of thermodynamics

$$T\mathrm{d}S = \mathrm{d}e - u\mathrm{d}u + p\mathrm{d}\tau. \tag{IV.12}$$

From (IV.12) we get using space derivation that

$$T\partial_X S = \partial_X e - u\partial_X u + p\partial_X \tau. \tag{IV.13}$$

Assuming in (IV.13) that the flow is locally isentropic $\partial_X S = 0$ and that p depends only on τ and S it yields that

$$\partial_X \tau = \left(\frac{\partial \tau}{\partial p}\right)\bigg|_S \partial_X p = -\left(\frac{\partial \tau}{\partial p}\right)\bigg|_S \rho_0 \mathrm{D}_t u. \tag{IV.14}$$

Then using (IV.14) in (IV.12), it writes

$$\partial_X e = u\partial_X u + p\left(\frac{\partial\tau}{\partial p}\right)\Big|_S \rho_0 \mathrm{D}_t u. \tag{IV.15}$$

The hypothesis of locally spatial isentropic flow is strong, it couples the space variation of total energy with the variation of both velocity and specific volume. For the sake of simplicity, we focus on perfect gas EOS and recall that $\rho_0 = 1$. But the study may be performed for any analytic EOS. Therefore we set $p = (\gamma-1)\dfrac{e-\frac{u^2}{2}}{\tau}$ and it yields

$$\left\{\begin{array}{l} \partial_X\tau = \dfrac{\tau^2}{\gamma(\gamma-1)(e-\frac{u^2}{2})}\mathrm{D}_t u, \\ \partial_X e = \quad u\partial_X u - \frac{\tau}{\gamma}\mathrm{D}_t u. \end{array}\right. \tag{IV.16}$$

The non-linear system using (IV.16) and (IV.8) writes for a perfect gas

$$\left\{\begin{array}{ll} \tau(X_s) + \dfrac{\tau(X_s)^2}{\gamma(\gamma-1)(e(X_s)-\frac{g^2}{2})}\mathrm{D}_t g(X-X_s) & = \tau(X), \\ g + \partial_X u(X_s)(X-X_s) & = u(X), \\ e(X_s) + (g\partial_X u(X_s) - \frac{\tau(X_s)}{\gamma}\mathrm{D}_t g)(X-X_s) & = e(X). \end{array}\right. \tag{IV.17}$$

Considering all values known at $X = \Delta X$ with $U(X) = U_{+1}$ and that $X_s = \sigma\Delta X$, (IV.17) writes

$$\left\{\begin{array}{ll} \tau(X_s) + \dfrac{\tau(X_s)^2}{\gamma(\gamma-1)(e(X_s)-\frac{g^2}{2})}\mathrm{D}_t g(1-\sigma)\Delta X & = \tau_{+1}, \\ g + \partial_X u(X_s)(1-\sigma)\Delta X & = u_{+1}, \\ e(X_s) + (g\partial_X u(X_s) - \frac{\tau(X_s)}{\gamma}\mathrm{D}_t g)(1-\sigma)\Delta X & = e_{+1}. \end{array}\right. \tag{IV.18}$$

From second equation of (IV.18), one easily gets $\partial_X u(X_s) = du = \frac{u_{+1}-g}{(1-\sigma)\Delta X}$. Then (IV.18) writes

$$\left\{\begin{array}{ll} \tau(X_s) + \dfrac{\tau(X_s)^2}{\gamma(\gamma-1)(e(X_s)-\frac{g^2}{2})}\mathrm{D}_t g(1-\sigma)\Delta X & = \tau_{+1}, \\ e(X_s) + (gdu - \frac{\tau(X_s)}{\gamma}\mathrm{D}_t g)(1-\sigma)\Delta X & = e_{+1}. \end{array}\right. \tag{IV.19}$$

Using second equation of (IV.19) in the first one, and using $y = (1-\sigma)\Delta X$ to alleviate the notations, it yields

$$(\tau_{+1} - \tau(X_s))\left(e_{+1} - \frac{g^2}{2} - (gdu - \frac{\tau(X_s)}{\gamma}\mathrm{D}_t g)y\right) = \frac{\tau(X_s)^2\mathrm{D}_t g}{\gamma(\gamma-1)}y. \tag{IV.20}$$

One obtains the polynomial equation (for a non-perfect gas, the equation may not be polynomial, but procedures still work)

$$\begin{aligned} f(\tau(X_s)) \quad &= \tau(X_s)^2(\frac{\mathrm{D}_t g}{\gamma-1}y) - \tau(X_s)\left((\frac{\mathrm{D}_t g \tau_{+1}}{\gamma} + gdu)y - e_{+1} + \frac{g^2}{2}\right) \\ &+\tau_{+1}\left((gdu)y - e_{+1} + \frac{g^2}{2}\right) = 0. \end{aligned} \tag{IV.21}$$

where f is a second order polynomial.

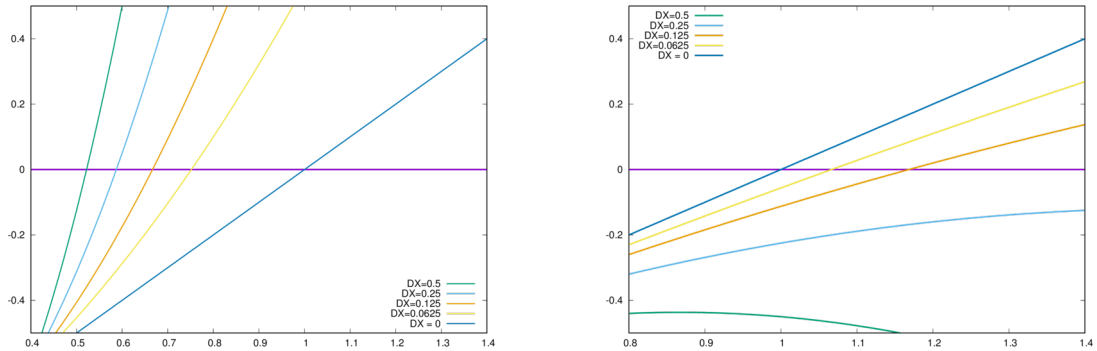
— If $\mathrm{D}_t g = 0$ then f becomes affine and the solution is $\tau(X_s) = \tau_{+1}$.

— Assume $\Delta X = 0$ then f becomes also an affine function and the solution is $\tau(X_s) = \tau_{+1}$.

— Otherwise, f has two roots β_1, β_2 with β_2 going to the infinity as ΔX goes to zero.

  — Assume $\mathrm{D}_t g > 0$, then the roots are always real.

  — Assume $\mathrm{D}_t g < 0$, then for ΔX small enough, the roots are real.



(a) $\mathrm{D}_t g > 0$ (b) $\mathrm{D}_t g < 0$

Figure IV.2 – Graph of $x \to f(\tau_{+1}x)$ using different value of ΔX for a positive $\mathrm{D}_t g$ on the left, and a negative one on the right.

On figure IV.2, values of $x \to f(\tau_{+1}x)$ as a function of x is shown for differents values of ΔX. For $\mathrm{D}_t g < 0$, we can see on the graph the non-existence of solution to $f(\tau_{+1}x) = 0$ as some curves do not cross the X-axis. But for smaller values of ΔX, real solution to $f(\tau_{+1}x) = 0$ exists.

Lemma IV.1 (Solution to the non-linear system and Lipschitz EOS gas)**.** *For any EOS such that the EOS function* $F(\tau,\epsilon) = \left(\frac{\partial \tau}{\partial p}\right)\Big|_S$ *is a Lipschitz function of* (τ,ϵ) *and such that* $p(\tau,\epsilon)$ *is locally bounded, then for* ΔX *small enough, the solution of eq.* (IV.17) *is unique, and a fixed point algorithm converges toward such a solution.*

Proof. Consider that u_s satisfies $u_s = g$. Denoting that $\epsilon_s = e_s - \frac{1}{2}u_s^2$, $\epsilon_{+1} = e_{+1} - \frac{1}{2}u_{+1}^2$, one writes the system as

$$\begin{cases} \tau_s &= \tau_{+1} - F(\tau_s,\epsilon_s)(1-\sigma)\Delta X, \\ \epsilon_s &= \epsilon_{+1} - \frac{1}{2}(u_s^2 - u_{+1}^2) - p(\tau_s,\epsilon_s)F(\tau_s,\epsilon_s)(1-\sigma)\Delta X, \end{cases} \tag{IV.22}$$

which can be easily rewritten under the form

$$\begin{pmatrix} \tau_s \\ \epsilon_s \end{pmatrix} = \boldsymbol{\psi}(\tau_s, \epsilon_s). \tag{IV.23}$$

If one shows in peculiar that the application $\boldsymbol{\psi}$ is a contraction mapping, thus using the Banach fixed point theorem, the result is proved. Using the Lipschitz hypothesis concerning F and using that p is locally bounded, and denoting $\boldsymbol{\alpha} = \begin{pmatrix} \tau \\ \epsilon \end{pmatrix}$ one gets immediately that

$$\|\boldsymbol{\psi}(\boldsymbol{\alpha}_1) - \boldsymbol{\psi}(\boldsymbol{\alpha}_2)\| \leq C(1-\sigma)\Delta X \|\boldsymbol{\alpha_1} - \boldsymbol{\alpha_2}\|. \tag{IV.24}$$

Then there exists β such that $\Delta X = \frac{\beta}{C(1-\sigma)}$, and so

$$\|\boldsymbol{\psi}(\boldsymbol{\alpha}_1) - \boldsymbol{\psi}(\boldsymbol{\alpha}_2)\| \leq \beta \|\boldsymbol{\alpha_1} - \boldsymbol{\alpha_2}\|, \tag{IV.25}$$

and for ΔX small enough, $\beta < 1$, which yields that $\boldsymbol{\psi}$ is a contraction mapping. Hence, the result. ■

Remark IV.1. The strong hypothesis $\partial_X S = 0$ is made for stabilization of the procedure. It yields high-order accuracy for smooth and isentropic flows, and gives first-order accuracy for non-isentropic flows.

Remark IV.2. One could change the procedure to compute first $\partial_X S$ doing an extrapolation of the entropy near the boundary. Then it gives high-order accuracy for smooth isentropic flows, but also for smooth non-isentropic flows.

IV-1.1.2 Second method: the larger stencil reconstruction

Here, the choice is made to use system (IV.11) written in the first cell of the computational domain ($X = X_1$) and to add a Taylor expansion of τ written in the second cell ($X = X_2$). Denoting $\varphi_s = \varphi(X_s)$ for simplicity, this leads to

$$\left\{ \begin{array}{ccc} \tau_s + (X_1 - X_s)\partial_X \tau_s & = & \tau_1, \\ \tau_s + (X_2 - X_s)\partial_X \tau_s & = & \tau_2, \\ u_s + (X_1 - X_s)\partial_X u_s & = & u_1, \\ e_s + (X_1 - X_s)\partial_X e_s & = & e_1, \\ u_s & = & g, \\ \partial_X \tau_s \partial_\tau p_s + \partial_X e_s \partial_e p_s + \partial_X u_s \partial_u p_s & = & -\mathrm{D}_t g, \end{array} \right. \tag{IV.26}$$

which rewrites

$$\begin{cases} \tau_s &= \dfrac{\tau_1(X_2-X_s)-\tau_2(X_1-X_s)}{X_2-X_1}, \\ \partial_X\tau_s &= \dfrac{\tau_2-\tau_1}{X_2-X_1}, \\ \partial_X u_s &= \dfrac{u_1-g}{X_1-X_s}, \\ e_s+(X_1-X_s)\partial_X e_s &= e_1, \\ u_s &= g, \\ \partial_X\tau_s\partial_\tau p_s+\partial_X e_s\partial_e p_s+\partial_X u_s\partial_u p_s &= -\mathrm{D}_t g, \end{cases} \tag{IV.27}$$

Indeed, since $p=(\gamma-1)\rho(e-u^2/2)$ for a perfect gas EOS, straightforward computations lead to

$$\partial_X e = u\partial_X u+\frac{\tau}{\gamma-1}\partial_X p+\frac{e-\dfrac{u^2}{2}}{\tau}\partial_X\tau. \tag{IV.28}$$

Using the second equation of (IV.6) – which here writes $D_t u+\partial_X p=0$ – together with the boundary condition, this rewrites, in $X=X_s$

$$\partial_X e_s = g\partial_X u_s-\frac{\tau_s}{\gamma-1}D_t g+\frac{e_s-\dfrac{g^2}{2}}{\tau_s}\partial_X\tau_s. \tag{IV.29}$$

Combining this equation with (IV.27) we get a linear equation for e_s and the whole system is solved if invertible. In peculiar, here, it yields $\tau_1\neq 0$. Once quantities are known in $X=X_s$, averaged ghost-cell values are computed as described in the preceding section. Results can be extended to ϵ-affine EOS as follows.

Lemma IV.2 (Linear system and ϵ-affine EOS). *If the EOS is affinely dependent on ϵ, ie $p(\epsilon,\tau)=a(\tau)\epsilon+b(\tau)$, then for $X_1\neq X_2$, $a(\tau_s)\neq(X_1-X_s)a'(\tau_s)\partial_X\tau_s$ and $a(\tau_s)\neq 0$, there exists a unique solution to* (IV.26).

Proof. Assume the EOS takes the form $p(\epsilon,\tau)=a(\tau)\epsilon+b(\tau)$. Then using that

$$\partial_X e = u\partial_X u+\left(\frac{\partial p}{\partial\epsilon}_{|\tau}\right)^{-1}\left(\partial_X p-\left(\frac{\partial p}{\partial\tau}_{|\epsilon}\right)\partial_X\tau\right),$$

it yields at the boundary that for $a(\tau_s)\neq 0$

$$\partial_X e_s = g\partial_X u_s-\frac{1}{a(\tau_s)}\left(\mathrm{D}_t g+\left(a'(\tau_s)(e_s-\frac{g^2}{2})+b'(\tau_s)\right)\partial_X\tau_s\right).$$

Inserting the previous equation in the Taylor expansion of e_s, one gets

$$e_s\left(1-(X_1-X_s)\frac{a'(\tau_s)}{a(\tau_s)}\partial_X\tau_s\right)=e_1-(X_1-X_s)\left(g\partial_X u_s-\frac{1}{a(\tau_s)}\left(\mathrm{D}_t g+\left(-a'(\tau_s)\frac{g^2}{2}+b'(\tau_s)\right)\partial_X\tau_s\right)\right).$$

Then the linear equation is solvable if

$$a(\tau_s) \neq (X_1 - X_s)a'(\tau_s)\partial_X \tau_s.$$

■

In the literature, many ϵ-affine EOS are presented. A non-exhaustive list of such EOS is presented hereafter.

— Perfect gas: $p(\epsilon, \tau) = (\gamma - 1)\frac{\epsilon}{\tau}$,

— Stiffened gas: $p(\epsilon, \tau) = (\gamma - 1)\frac{\epsilon}{\tau} - p^\star$,

— Mie-Grüneisen gas [85]: $p(\epsilon, \tau) = p^\star(\tau) + \frac{\Gamma(\tau)}{\tau}(\epsilon - \epsilon^\star(\tau))$.

For non ϵ-affine EOS, the following lemma gives result concerning existence and uniqueness of the solution

Lemma IV.3. *For any EOS such that the EOS function* $F_1(\epsilon) = \left(\frac{\partial p}{\partial \epsilon}_{|\tau}\right)^{-1}$ *is a Lipschitz function of* ϵ *and that the function* $F_2(\epsilon) = \left(\frac{\partial p}{\partial \tau}_{|\epsilon}\right)$ *is locally bounded, then for* ΔX *small enough, the solution is unique, and a fixed point algorithm converges toward such a solution.*

Proof. The proof is very similar and uses the same argument as the one for lemma IV.1. The coefficient ΔX gives the contraction mapping using the Lipschitz hypothesis of F_1, and the locally boundedness of F_2. ■

The aim of the work is now to see if lemmas IV.1 to IV.3 still holds for arbitrary orders of accuracy and non-constant masses.

IV-1.2 General procedure, and characterization of the solution for the system at the boundary

The previous study has been made for the special case of a second order boundary treatment, with constant mass. For spatially isentropic flow hypothesis, lemma IV.1 gives existence and uniqueness of the solution under Lipschitz hypothesis concerning the EOS for ΔX small enough. Similar results hold for the second approach – removing the $\partial_x S = 0$ hypothesis and using an enlarged stencil – and we morever get existence and uniqueness without any restriction for ϵ-affine EOS. We now study the general case.

The procedure is now extended without any restriction on the initial density profile. In the following we will set $n = 1$, meaning that only g and $\mathrm{D}_t g$ are known at the boundary (in practice, more material derivatives of g could be taken into account but it would lead to heavier algebra). To alleviate notations, we also introduce

$$\psi_{i,k} = \frac{1}{\Delta X}\left(\frac{(X_i + \frac{\Delta X}{2} - X_s)^{k+1} - (X_i - \frac{\Delta X}{2} - X_s)^{k+1}}{(k+1)!}\right).$$

Considering a mth-order scheme and dropping the $\mathscr{O}(\Delta X^m)$, spatial Taylor expansions of conservative variables write

$$\left\{\begin{array}{rcl} \overline{\rho_0}_i & = & \displaystyle\sum_{k<m} \partial_X^k \rho_0\Big|_{x=x_s} \psi_{i,k}, \\ \overline{\rho_0 u}_i & = & \displaystyle\sum_{k<m}\sum_{l\leq k} \binom{k}{l} \partial_X^l \rho_0\Big|_{x=x_s} \partial_X^{k-l} u\Big|_{x=x_s} \psi_{i,k}, \\ \overline{\rho_0 \tau}_i & = & \displaystyle\sum_{k<m}\sum_{l\leq k} \binom{k}{l} \partial_X^l \rho_0\Big|_{x=x_s} \partial_X^{k-l} \tau\Big|_{x=x_s} \psi_{i,k}, \\ \overline{\rho_0 e}_i & = & \displaystyle\sum_{k<m}\sum_{l\leq k} \binom{k}{l} \partial_X^l \rho_0\Big|_{x=x_s} \partial_X^{k-l} e\Big|_{x=x_s} \psi_{i,k}. \end{array}\right. \tag{IV.30}$$

IV-1.2.1 Well-posedness at the boundary for spatially isentropic flow hypothesis

Boundary condition and isentropic flow hypothesis provide the following informations :

$$\left\{\begin{array}{rcl} u|_{x=x_s} & = & g(t), \\ \partial_X \tau|_{x=x_s} & = & -\left(\dfrac{\partial \tau}{\partial p}\right)\Big|_{x=x_s} \rho_0|_{x=x_s} \mathrm{D}_t g, \\ \partial_X e & = & u\partial_X u - p\partial_X \tau, \end{array}\right. \tag{IV.31}$$

It yields three subsystems to be solved at each boundary in the following order:

— The first system is built using the first equation of (IV.30). It is a linear system whose size is $m \times m$. It allows then to build ghost cells values of $\overline{\rho_0}$.

— The second system is built using the second equation of (IV.30) and the boundary condition on the velocity. It is also a linear system whose size is $(m-1)\times(m-1)$. It allows then to build ghost cells values of $\overline{\rho_0 u}$.

— The third and last system is built using the third and fourth equations of (IV.30) and system (IV.31). The non-linearity of the system is explained by the non-linearity of (IV.31). The size of the system is $(2m-2)\times(2m-2)$. It allows then to build ghost cells values of $\overline{\rho_0 \tau}$ and $\overline{\rho_0 e}$.

Once the three systems are solved, ghost-cells values of all quantities are built by Taylor expansions.

Lemma IV.4 (Solution to the non-linear system and Lipschitz EOS gas)**.** *For any EOS such that the EOS function* $F(\tau,\epsilon) = \left(\dfrac{\partial \tau}{\partial p}\right)\Big|_S$ *is a Lipschitz function of* (τ,ϵ) *and such that* $p(\tau,\epsilon)$ *is locally bounded, then for* ΔX *small enough, the solution is unique, and a fixed point algorithm converges toward such a solution.*

Proof. The proof is identical to the one proposed for lemma IV.1. ■

Remark IV.3. One could use repeated space derivation of the third equation of eq. (IV.31), to substitute space derivatives in e into functions of $(e, \tau, \partial_X \tau, ...)$, yielding a $m \times m$ system to be solved. But for such a choice, theoretical results concerning existence and uniqueness of solution are not accessible, and requires stronger regularity hypothesis on the EOS.

IV-1.2.2 Well-posedness at the boundary for enlarged stencil

We have shown existence in lemmas IV.2 and IV.3 of a 2nd-order solution to the prescribed velocity boundary problem for Lagrangian hydrodynamics when initial mass density is uniform using the larger stencil based reconstruction. The boundary condition and the equation of state provide the following informations

$$\left\{\begin{array}{l} u|_{X=X_s}=g(t), \\ \partial_X e \ =u\partial_X u - \left(\frac{\partial p}{\partial \epsilon}_{|\tau}\right)^{-1}\left((\rho_0 \mathrm{D}_t g + \left(\frac{\partial p}{\partial \tau}_{|\epsilon}\right)\partial_X \tau)\right) \text{ in } X = X_s. \end{array}\right. \tag{IV.32}$$

Considering (IV.30-IV.32) we therefore have four subsystems to solve at each boundary. This is done the following way:

— The first system is built using the first equation of (IV.30), considering m interior cells. It leads to a $m \times m$ linear system. It allows then to build ghost-cell values of $\overline{\rho_0}$.

— The second system is built using the second equation of (IV.30), considering $m-1$ interior cells and the boundary condition on the velocity. It leads to a $(m-1)\times(m-1)$ linear system. It allows then to build ghost-cell values of $\overline{\rho_0 u}$.

— The third system is built using the third equation of (IV.30), considering m interior cells. It leads to a $m \times m$ linear system. It allows then to build ghost-cell values of $\overline{\rho_0 \tau}$.

— The fourth system is built using the fourth equation of (IV.30), considering $m-1$ interior cells and system (IV.32). This system is linear for perfect and stiffened gases EOS but may be non-linear for some EOS, thus requiring fixed-point algorithms to be solved. The size of the system is $(m-1)\times(m-1)$. Once the solution is known, it allows to build ghost-cell values of $\overline{\rho_0 e}$.

We extend lemma IV.2 to arbitrary orders and non-constant ρ_0 as

Lemma IV.5 (Linear system and ϵ-affine EOS). *If the EOS is affinely dependent on ϵ, ie $p(\epsilon,\tau) = a(\tau)\epsilon + b(\tau)$, then the system eqs.* (IV.30) *and* (IV.32) *is linear.*

Proof. Assume the EOS writes $p(\epsilon,\tau) = a(\tau)\epsilon + b(\tau)$, then using

$$\partial_X e = u\partial_X u + \left(\frac{\partial p}{\partial \epsilon}_{|\tau}\right)^{-1}\left(\partial_X p - \left(\frac{\partial p}{\partial \tau}_{|\epsilon}\right)\partial_X \tau\right),$$

it yields at the boundary that

$$\partial_X e_s = g\partial_X u_s - \frac{1}{a(\tau_s)}\left(\rho_{0s}\mathrm{D}_t g + \left(a'(\tau_s)(e_s - \frac{g^2}{2}) + b'(\tau_s)\right)\partial_X \tau_s\right).$$

Therefore $\partial_X e_s$ is a linear function of e_s, and thus the system is linear. ■

For non ϵ-affine EOS, the following lemma gives existence and uniqueness of the solution.

Lemma IV.6 (Uniqueness of solution for Lipschitz hypothesis on the EOS). *For any EOS such that the EOS function $F_1(\epsilon) = \left(\frac{\partial p}{\partial \epsilon}_{|\tau}\right)^{-1}$ is a Lipschitz function of ϵ and that the function*

$F_2(\epsilon) = \left(\frac{\partial p}{\partial \tau}_{|\epsilon}\right)$ is locally bounded, then for ΔX small enough, the solution exists and is unique, and a fixed point algorithm converges toward such a solution.

Proof. The proof is essentially the same as for the case with constant mass and second order of accuracy. ■

Remark IV.4. In practice, ghost-cells values are imposed at the beginning of each time-step or sub-cycle eg. if the scheme is based on Runge–Kutta sequences.

IV-1.3 Stabilization procedure for shocks and very high-order reconstruction

Spurious oscillations or non-physical values may result with this high-order treatment in case of discontinuous solutions near the boundary. A MOOD procedure has been developed to improve robustness. Moreover for very high-order scheme, the linearized version is not stable using only g and $\mathrm{D}_t g$. Thus a least-square method also has been developed to enforce stability.

IV-1.3.1 MOOD procedure

A MOOD procedure [24] has been added to automatically decrease the order of this inverse Lax–Wendroff method if some criteria are violated during the reconstruction of ghost cells values. It is done in order to improve stability in case of strong shocks ingoing towards the boundary. The flow chart of the procedure is depicted in fig. IV.3. The idea is to set as a criteria, the positivity of the density and internal energy. While the reconstructed density or internal energy in Ω_- are non-positive, the order of reconstruction is decreased until first order accuracy or a positive internal energy and density are reached.

IV-1.3.2 Least-square methods for very high-order methods

The problem, linear or not, to be solved at the boundary can be rewritten under the form

$$\boldsymbol{\mathcal{F}}(\boldsymbol{\Theta}) = \boldsymbol{X}. \tag{IV.33}$$

If the system is linear, there exists a matrix $\mathcal{A}$ such that $\boldsymbol{\mathcal{F}}(\boldsymbol{\Theta}) = \underline{\boldsymbol{\mathcal{A}}}\boldsymbol{\Theta}$, where $\underline{\boldsymbol{\mathcal{A}}}$ is a square matrix of size $p \times p$ and hence $\boldsymbol{X} \in \mathbb{R}^p$, $\Theta \in \mathbb{R}^p$. The idea of the least-square method is to add values in the interior domain such that the system writes

$$\underline{\widehat{\boldsymbol{\mathcal{A}}}}\boldsymbol{\Theta} = \widehat{\boldsymbol{X}} \tag{IV.34}$$

where $\underline{\widehat{\boldsymbol{\mathcal{A}}}} \in \mathbb{R}^{q\times p}$ and $\widehat{\boldsymbol{X}} \in \mathbb{R}^q$. Instead of solving directly eq. (IV.34), we introduce the functional $\mathcal{J}$ as

$$\mathcal{J} = \|\underline{\widehat{\boldsymbol{\mathcal{A}}}}\boldsymbol{\Theta} - \widehat{\boldsymbol{X}}\|, \tag{IV.35}$$

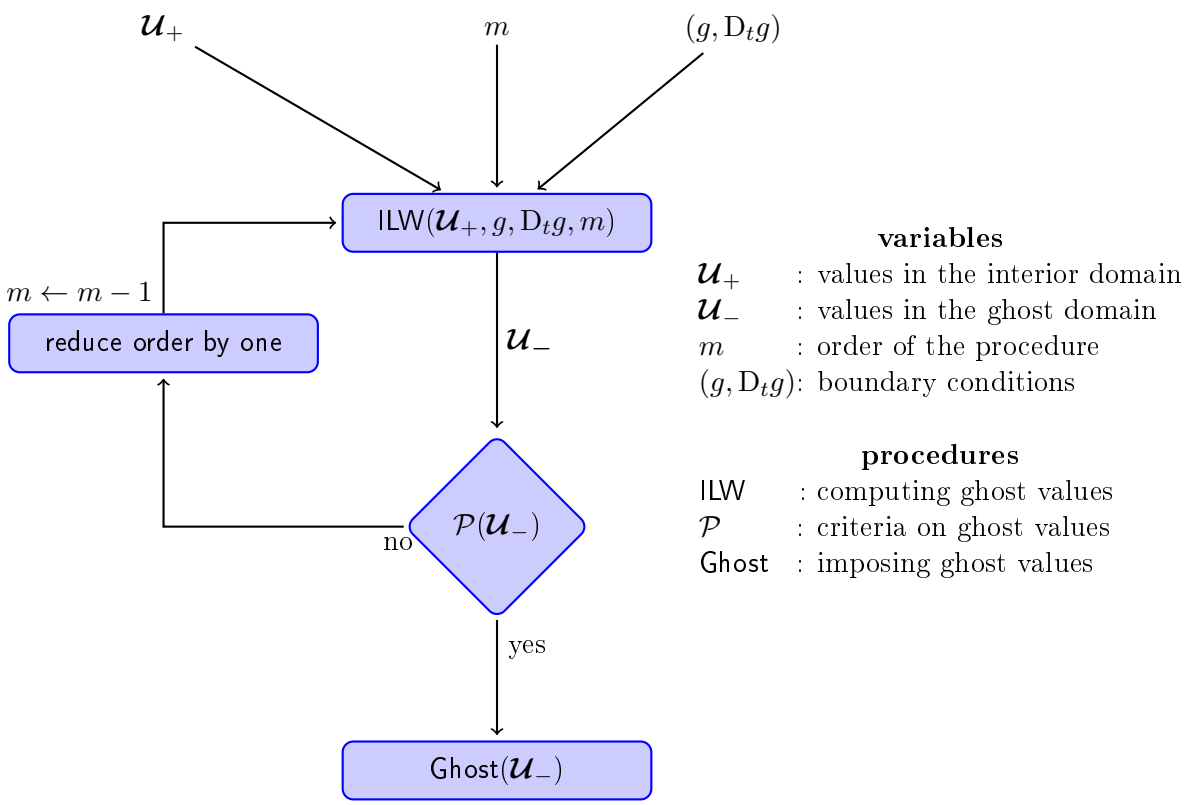


Figure IV.3 – Flow chart for the MOOD procedure applied at the boundary imposing specific criteria on the computed ghost values

where the norm here is arbitrary fixed to the Euclidean norm. The idea is to minimize the functional $\mathcal{J}$ in order to satisfy in a reduced sense the Taylor expansions. Meaning, in particular that the solution $\Theta^\star$ is defined as

$$\forall \Theta \in \mathbb{R}^p, \quad \|\underline{\widehat{\boldsymbol{\mathcal{A}}}}\boldsymbol{\Theta} - \widehat{\boldsymbol{X}}\| \geq \|\underline{\widehat{\boldsymbol{\mathcal{A}}}}\boldsymbol{\Theta}^\star - \widehat{\boldsymbol{X}}\| \tag{IV.36}$$

Such a procedure, called the least-square method (see [3]), is used to stabilize the reconstruction operator, especially for very high-order reconstructions where the classical reconstruction is proved to be linearly unstable. A classical Gauss–Newton algorithm is performed to solve eq. (IV.36). If the system is non-linear, then the solution $\Theta^\star$ is defined as

$$\forall \Theta \in \mathbb{R}^p, \quad \|\widehat{\boldsymbol{\mathcal{F}}}(\boldsymbol{\Theta}) - \widehat{\boldsymbol{X}}\| \geq \|\widehat{\boldsymbol{\mathcal{F}}}(\boldsymbol{\Theta}^\star) - \widehat{\boldsymbol{X}}\| \tag{IV.37}$$

IV-1.4 1D validation and comparisons

We assess in this part both the accuracy and the robustness of our method for the 1D Euler system. The study here is performed using the larger stencil based reconstruction applied to the GoHy schemes developed in [50, 170]. The spatially isentropic flow hypothesis based reconstruction gives similar results concerning isentropic test-cases, but dramatically reduces to first order

accuracy (using the MOOD procedure) for any non-spatially isentropic flow, as expected.

IV-1.4.1 Kidder isentropic compression test-case [95]

Kidder's test problem represents the isentropic compression of an ideal volume of gas initially at rest. For this test, the computational domain $[0,1]$ is discretized in N_x regular cells. Let (p_i, ρ_i) and (p_e, ρ_e) denote initial pressures and mass densities at $x = 0$ and $x = 1$ respectively. Initial profiles are defined by

$$\begin{cases} \rho_0(x) &= \left(x^2\rho_e^{\gamma-1} + (1-x^2)\rho_i^{\gamma-1}\right)^{\frac{1}{\gamma-1}}, \\ u_0(x) &= 0, \\ p_0(x) &= p_e\left(\dfrac{\rho(x)}{\rho_e}\right)^{\gamma}, \end{cases} \tag{IV.38}$$

with $\gamma = 3$ and here we will take $p_e = 100$, $p_i = 1$, $\rho_e = 1$ and $\rho_i = \rho_e(p_i/p_e)^{\frac{1}{\gamma}}$. Introducing the sound speed $c = \sqrt{\gamma p/\rho}$, we define the focalization time $t_c = \sqrt{\frac{\gamma-1}{2}\frac{1}{c_e^2-c_i^2}}$ which allows to write the complete analytical solution. Defining $h(t) = \sqrt{1-(t/t_c)^2}$, it is given by

$$\rho(x,t) = \rho_0\left(\frac{x}{h(t)}\right)\cdot h(t)^{\frac{2}{\gamma-1}}, \quad u(x,t) = -\frac{xt}{t_c^2 h(t)^2}, \quad p(x,t) = p_e\left(\frac{\rho(x,t)}{\rho_e}\right)^{\gamma}.$$

For this test we solve Euler equations on $\Omega = [x_l, x_r]$ and exact velocities are prescribed at left and right boundaries $x_l = 0.05 + 5\sqrt{7}\,10^{-3}$ and $x_r = 0.95 - 3.33\sqrt{5}\,10^{-3}$. The scheme GoHy-1 stands for the classic acoustic solver.

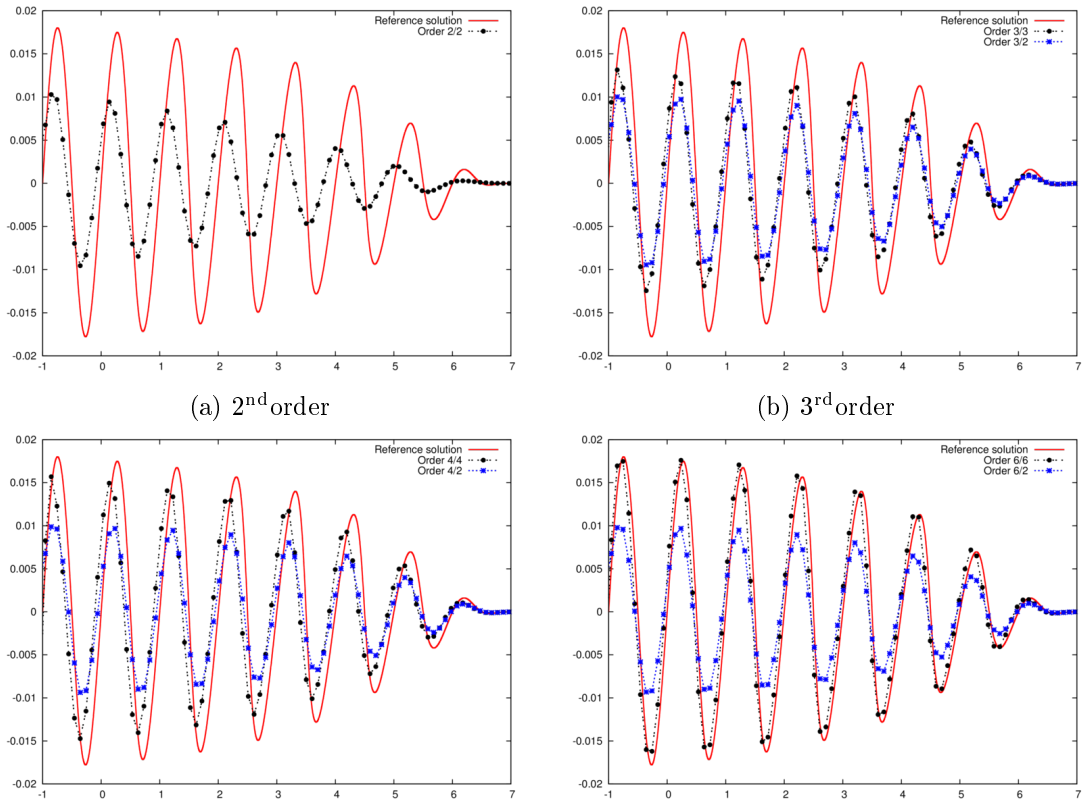
N_x	GoHy-1		GoHy-2		GoHy-3		GoHy-4		GoHy-5		GoHy-6	
25	1.6e-3	·	2.3e-4	·	8.7e-6	·	9e-6	·	3.9e-6	·	1.4e-5	·
50	7.1e-4	1.2	3.5e-5	2.7	4.4e-7	4.3	3.6e-8	8.0	6.1e-9	9.3	3.5e-7	5.3
100	3.7e-4	0.9	2.8e-5	0.3	2.71e-7	1.0	1.6e-9	4.4	1.2e-10	5.7	1.8e-12	17.6
200	1.8e-4	1.0	7.3e-6	2.0	2.7e-8	3.0	4.8e-11	5.0	1.8e-12	6.1	3.5e-14	5.7
400	9.0e-5	1.0	1.8e-6	2.0	3.4e-9	3.0	1.2e-12	5.2	9.4e-15	7.5	4.9e-15	2.9
800	4.5e-5	1.0	4.7e-7	2.0	4.3e-10	3.0	7.4e-14	4.0	2.7e-14	$\star$	3.3e-14	$\star$
1600	2.2e-5	1.0	1.2e-7	2.0	5.4e-11	3.0	8e-14	$\star$	3.6e-14	$\star$	3.7e-14	$\star$
3200	1.1e-5	1.0	2.9e-8	2.0	6.8e-12	3.0	8.3e-14	$\star$	3.9e-14	$\star$	3.5e-14	$\star$

Table IV.1 – l^1-error and experimental order of convergence (EOC) for ILW-GoHy schemes at $t = 0.01$ with a CFL of 0.9. EOC indexed with $\star$ are reduced due to double precision. For stability issues, least-squares method is used for 4th, 5th and 6th-order.

Results concerning the l^1-errors and experimental orders of convergence are given in table IV.1 for GoHy schemes up to 6th-order. For each scheme the expected order of accuracy is reached.

IV-1.4.2 Harmonic piston test-case

The harmonic piston test-case is used to assess the ability of the reconstruction to recover correct phase/amplitude profiles using a harmonic source. The initial data are those of a perfect gas



(a) 2^{nd}order (b) 3^{rd}order

(c) 4th order (d) 6th order

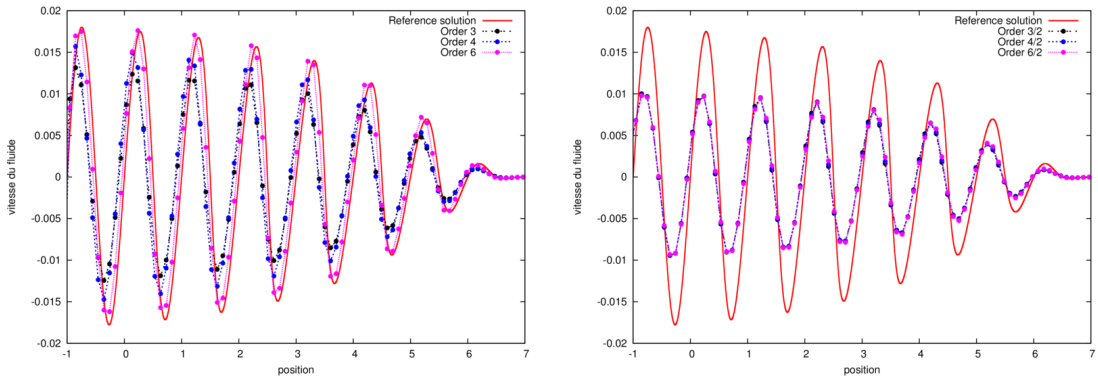
Figure IV.4 – Velocity profiles with 10 cells per wavelength for the 2nd, 3rd, 4th and 6th-order GoHy schemes for the harmonic piston problem at $T = 9$.

($\gamma = 1.4$) at rest, and the velocity at left boundary, initially located at $x_l = -1$, is imposed.

$$\begin{cases} \rho_0(x) &= \gamma, \\ u_0(x) &= 0, \\ p_0(x) &= 1, \end{cases} \quad \text{for } x \geq x_l(0) = -1 \quad \text{and} \quad u(x_l(t)) = a\,\mathrm{e}^{\frac{-8}{t^2}} \sin(2\pi t). \tag{IV.39}$$

such that the sound speed in initially set to 1, and $a = 2.10^{-2}$.

Velocity profiles are depicted on fig. IV.4. The red plain line represents the reference solution computed with the first order scheme (acoustic solver) and 100000 cells. The black dotted line represents results obtained with inner scheme and reconstruction fixed to the same order of accuracy. The blue dotted line is for inner scheme at high-order accuracy but with only a second order reconstruction procedure. As expected as the order of accuracy is increased, so is the ability of the scheme concerning the recovering of both phase and amplitude of the signal. The most significant feature lies in the difference between the blue and black dotted lines. When the order of the reconstruction is fixed to 2^{nd}order, both phase and amplitude are not so well recovered. On fig. IV.5, one can see that with a second order reconstruction, results for third, fourth and sixth order inner schemes are equivalent. This is not the case with reconstruction whose order match the one of the inner scheme.



(a) High-order reconstruction (b) 2^{nd}order reconstruction

Figure IV.5 – Velocity profiles with 10 cells per wavelength for the 3rd, 4th and 6th-order GoHy schemes for the harmonic piston problem at $T = 9$. On the left, results with appropriate order of reconstruction is depicted, whereas on the right results are shown with second order reconstruction.

IV-1.4.3 Sod piston test-case [146]

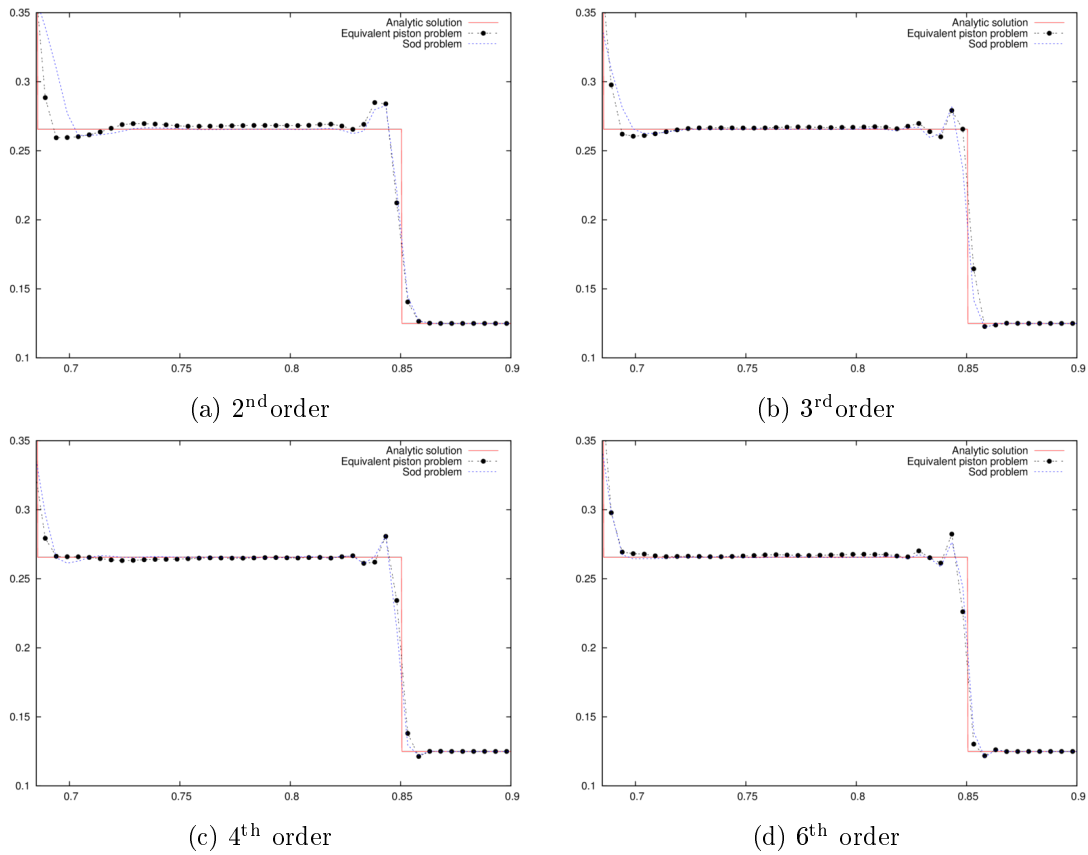
Next test-case is representative of a piston shocking a gas at rest. Initial data are provided by the right-state of the Sod's shock tube (perfect gas EOS with $\gamma = 1.4$) and at the left boundary, initially located at $x_l = 0.5$, the exact contact discontinuity velocity is prescribed:

$$\begin{cases} \rho_0(x) &= 0.125, \\ u_0(x) &= 0, \\ p_0(x) &= 0.1, \end{cases} \quad \text{for } x \geq x_l(0) = 0.5 \quad \text{and} \quad u(x_l(t)) = 0.927452624. \tag{IV.40}$$

Density profiles are depicted on fig. IV.6. The red plain line represents the analytical solution. The blue dotted line represents the Sod's shock tube solution computed as a Riemann problem using both left and right initial states with the GoHy solver and the black dotted line represents the solution obtained with the present ILW method. Shock positions and density levels are in good agreement with the analytical solution for both methods. The contact continuity is even slightly better recovered with the ILW procedure than for the complete Riemann problem. Note that the MOOD procedure presented in section IV-1.3.1 is not used here.

IV-2 Extension of the ILW procedure to the 2D Euler system

The procedure designed for the 1D Euler system is now used with a high-order accurate dimensional splitting method on the 2D Euler system (IV.1-IV.3), as it is described in [50, 170, 35].



(a) 2^{nd}order (b) 3^{rd}order

(c) 4th order (d) 6th order

Figure IV.6 – Density profiles with initially 100 cells for the 2nd, 3rd, 4th and 6th-order GoHy schemes for the Sod piston problem.

IV-2.1 Formulation of the ILW procedure using directionnal splitting

Concerning the Lagrangian step, two subsystems will therefore be alternatively considered, depending on the sweep direction:

$$\left\{\begin{array}{lcl} \mathrm{D}_t^x\left(\rho_0\tau\right)-\partial_X u & = & 0,\\ \mathrm{D}_t^x\left(\rho_0 u\right)+\partial_X p & = & 0,\\ \mathrm{D}_t^x\left(\rho_0 v\right) & = & 0,\\ \mathrm{D}_t^x\left(\rho_0 e\right)+\partial_X\left(pu\right) & = & 0, \end{array}\right. \qquad \left\{\begin{array}{lcl} \mathrm{D}_t^y\left(\rho_0\tau\right)-\partial_Y v & = & 0,\\ \mathrm{D}_t^y\left(\rho_0 u\right) & = & 0,\\ \mathrm{D}_t^y\left(\rho_0 v\right)+\partial_Y p & = & 0,\\ \mathrm{D}_t^y\left(\rho_0 e\right)+\partial_Y\left(pv\right) & = & 0, \end{array}\right. \tag{IV.41}$$

where $\mathrm{D}_t^x=\partial_t+u\partial_x$ and $\mathrm{D}_t^y=\partial_t+v\partial_y$ denote the Lagrangian derivatives in $x-$ and $y-$directions respectively. Note that Lagrangian subsystems are simpler than Eulerian ones since convective terms, which are missing here, will be treated during the projection step. When replacing space derivatives by temporal ones this will lead to a simpler algebra in the sequel and a very close approach to the one proposed in section IV-1 for the 1D case.

Denoting $\boldsymbol{u}=(u,v)^t$, we recall that the considered boundary condition on Γ is given by $\boldsymbol{u}\cdot\boldsymbol{n}(t,s)=g(t,s)$, where s is the curvilinear coordinate along the boundary $\Gamma(t)$, and $\boldsymbol{n}(t,s)$ the normal at coordinate s and time t. The boundary is described by a set of points (or pearls) P_s distributed along Γ (see Figure IV.1). On each of these points, a problem similar to the one dealt with in the 1D case, is solved at high-order accuracy. From these data, values are then set in ghost-cells. To get close to the 1D case, velocity components are computed in a local basis $(\boldsymbol{t_s},\boldsymbol{n_s})$ where $\boldsymbol{t_s}$ and $\boldsymbol{n_s}$ are respectively tangent and normal vectors to Γ in P_s. Velocity components in this basis are denoted

$$\left\{\begin{array}{l} \widehat{u}=\boldsymbol{u}\cdot\boldsymbol{n_s}=u\cdot n_1+v\cdot n_2,\\ \widehat{v}=\boldsymbol{u}\cdot\boldsymbol{t_s}=u\cdot t_1+v\cdot t_2. \end{array}\right. \tag{IV.42}$$

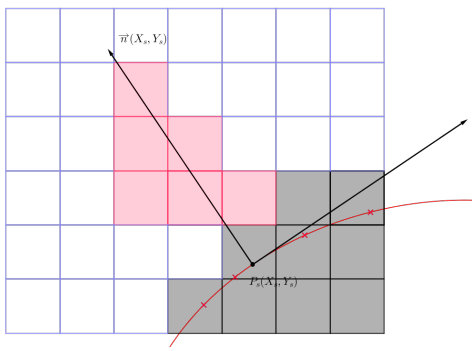


Figure IV.7 – Zoom on a point P_s on the discretized boundary with local coordinate system. The colored zone corresponds to a six points stencil for 3rd order reconstruction.

IV-2.1.1 Dimensional splitting technique

The DSM consists in alternatively applying the previous method in the x- and y-direction with appropriate weighted-time increments $\omega_k \Delta t$. To reach high-order accuracy in time, splitting sequences beyond the well-known 2nd-order Strang DSM must be used. Such weights, up to 8th-order, can be found in [50, 170, 35] and are reported in appendix, section A.2. During these sequences, prescribing time-dependent boundary conditions at intermediary time-steps can reveal quite tricky. The naive way yields only at most second order of accuracy. This is somehow similar to results found by Carpenter in [20]. To achieve this, the boundary condition is also rewritten as a 2D evolution system that is also split as explained now. Let us denote $\boldsymbol{n} = (n_1, n_2)^t$, $\boldsymbol{i_1} = (1,0)^t$, and $\boldsymbol{i_2} = (0,1)^t$. We introduce $\boldsymbol{g}(t,s) = g(t,s)\ \boldsymbol{n}(t,s)$ and in the sequel we also assume that $\mathrm{D}_t \boldsymbol{n} = 0$ and that $g(t,s)$ is known analytically. Letting $g_1 = \mathrm{D}_t g$ we therefore can write

$$\mathrm{D}_t \boldsymbol{g} = (g_1 n_1)\ \boldsymbol{i_1} + (g_1 n_2)\ \boldsymbol{i_2}. \tag{IV.43}$$

As for Euler equations, system (IV.43) is then split into the following two equations that will be alternatively solved according to the splitting sequence used for the inner scheme

$$\mathrm{D}_t \boldsymbol{g} = (g_1 n_1)\ \boldsymbol{i_1}, \quad \text{and} \quad \mathrm{D}_t \boldsymbol{g} = (g_1 n_2)\ \boldsymbol{i_2}. \tag{IV.44}$$

Assume that time weights ω_{2k-1} and ω_{2k} are respectively used for the x- and y-sweeps respectively and let us denote $t^{n+\omega_k}$ the fictitious time for the kth sweep (with $\omega_0 = 0$). We therefore get for any $l \geq 1$

$$\left\{ \begin{array}{lcl} \boldsymbol{g}^{n+\omega_{2l-1}} & = & \boldsymbol{g}^{n+\omega_{2l-2}} + \displaystyle\int_{t^n+\sum_{k=1}^{l-1}\omega_{2k-1}\Delta t}^{t^n+\sum_{k=1}^{l}\omega_{2k-1}\Delta t} (g_1 n_1)\ \boldsymbol{i_1} \mathrm{d}\theta, \\ \boldsymbol{g}^{n+\omega_{2l}} & = & \boldsymbol{g}^{n+\omega_{2l-1}} + \displaystyle\int_{t^n+\sum_{k=1}^{l-1}\omega_{2k}\Delta t}^{t^n+\sum_{k=1}^{l}\omega_{2k}\Delta t} (g_1 n_2)\ \boldsymbol{i_2} \mathrm{d}\theta, \end{array} \right.$$

which rewrites by induction, for any $l \geq 1$

$$\left\{ \begin{array}{lcl} \boldsymbol{g}^{n+\omega_{2l-1}} & = & \boldsymbol{g}^{n} + \displaystyle\int_{t^n}^{t^n+\sum_{k=1}^{l}\omega_{2k-1}\Delta t} (g_1 n_1)\ \boldsymbol{i_1} \mathrm{d}\theta + \int_{t^n}^{t^n+\sum_{k=1}^{l-1}\omega_{2k}\Delta t} (g_1 n_2)\ \boldsymbol{i_2} \mathrm{d}\theta, \\ \boldsymbol{g}^{n+\omega_{2l}} & = & \boldsymbol{g}^{n} + \displaystyle\int_{t^n}^{t^n+\sum_{k=1}^{l}\omega_{2k-1}\Delta t} (g_1 n_1)\ \boldsymbol{i_1} \mathrm{d}\theta + \int_{t^n}^{t^n+\sum_{k=1}^{l}\omega_{2k}\Delta t} (g_1 n_2)\ \boldsymbol{i_2} \mathrm{d}\theta. \end{array} \right.$$

Since $g_1 = \mathrm{D}_t g$ and $\mathrm{D}_t \boldsymbol{n} = 0$, exact integration therefore yields, for any $l \geq 1$

$$\left\{ \begin{array}{lcl} \boldsymbol{g}(t^{n+\omega_{2l-1}}) & = & \left(g(t^n + \Delta t \displaystyle\sum_{k=1}^{l} \omega_{2k-1})\ n_1 \right) \boldsymbol{i_1} + \left(g(t^n + \Delta t \displaystyle\sum_{k=1}^{l-1} \omega_{2k})\ n_2 \right) \boldsymbol{i_2}. \\ \boldsymbol{g}(t^{n+\omega_{2l}}) & = & \left(g(t^n + \Delta t \displaystyle\sum_{k=1}^{l} \omega_{2k-1})\ n_1 \right) \boldsymbol{i_1} + \left(g(t^n + \Delta t \displaystyle\sum_{k=1}^{l} \omega_{2k})\ n_2 \right) \boldsymbol{i_2}, \end{array} \right.$$

that is to say, performing the scalar product with $\boldsymbol{n}$,

$$\left\{\begin{array}{rcl} g(t^{n+\omega_{2l-1}}) & = & g\left(t^n+\Delta t\sum_{k=1}^{l}\omega_{2k-1}\right)\ n_1^2+g\left(t^n+\Delta t\sum_{k=1}^{l-1}\omega_{2k}\right)\ n_2^2, \\ g(t^{n+\omega_{2l}}) & = & g\left(t^n+\Delta t\sum_{k=1}^{l}\omega_{2k-1}\right)\ n_1^2+g\left(t^n+\Delta t\sum_{k=1}^{l}\omega_{2k}\right)\ n_2^2. \end{array}\right. \tag{IV.45}$$

These relations are used at the beginning of each dimensional sweep to prescribe boundary conditions.

IV-2.1.2 Methodology for a given sweep

We now consider a sweep in the $x-$direction so that only the first subsystem of (IV.41) is of interest – methods for other sweeps are strictly identical *modulo* slight modifications mentioned in section IV-2.1.1. In the following, we only use g and $\mathrm{D}_t^x g$ for building the non-linear problem. As in the 1D case, more material derivatives could be used but it would lead to a heavier algebra. To alleviate notations in 2D Taylor expansions we introduce

$$\psi_{i,j,k,l}=\frac{1}{k!}\left(\frac{(x_i+\frac{\Delta x}{2}-x_s)^{l+1}-(x_i-\frac{\Delta x}{2}-x_s)^{l+1}}{(l+1)\Delta x}\right)\left(\frac{(y_j+\frac{\Delta y}{2}-y_s)^{k-l+1}-(y_j-\frac{\Delta y}{2}-y_s)^{k-l+1}}{(k-l+1)\Delta y}\right).$$

Let us consider a mth-order scheme. Extending computations done in section IV-1 to the 2D case and performing the local change for velocity components, spatial Taylor expansions lead to

$$\left\{\begin{array}{rcl} \overline{\rho_0}_{i,j} & = & \sum_{k<m}\sum_{l\leq k}\binom{k}{l}\ \partial_X^l\partial_Y^{k-l}\rho_0\Big|_{x=x_s,y=y_s}\ \psi_{i,j,k,l}, \\ \overline{\rho_0\widehat{u}}_{i,j} & = & \sum_{k<m}\sum_{l\leq k}\binom{k}{l}\ \partial_X^l\partial_Y^{k-l}(\rho_0\widehat{u})\Big|_{x=x_s,y=y_s}\ \psi_{i,j,k,l}, \\ \overline{\rho_0\widehat{v}}_{i,j} & = & \sum_{k<m}\sum_{l\leq k}\binom{k}{l}\ \partial_X^l\partial_Y^{k-l}(\rho_0\widehat{v})\Big|_{x=x_s,y=y_s}\ \psi_{i,j,k,l}, \\ \overline{\rho_0\tau}_{i,j} & = & \sum_{k<m}\sum_{l\leq k}\binom{k}{l}\ \partial_X^l\partial_Y^{k-l}(\rho_0\tau)\Big|_{x=x_s,y=y_s}\ \psi_{i,j,k,l}, \\ \overline{\rho_0 e}_{i,j} & = & \sum_{k<m}\sum_{l\leq k}\binom{k}{l}\ \partial_X^l\partial_Y^{k-l}(\rho_0 e)\Big|_{x=x_s,y=y_s}\ \psi_{i,j,k,l}. \end{array}\right. \tag{IV.46}$$

The boundary condition and the equation of state provide the following informations in P_s

$$\left\{\begin{array}{rcl} \widehat{u}|_{P=P_s} & = & g(t) \\ \partial_X e\cdot n_1|_{P=P_s} & = & (\widehat{u}\partial_X\widehat{u}+\widehat{v}\partial_X\widehat{v})\cdot n_1-\left(\frac{\partial p}{\partial\epsilon}_{|\tau}\right)^{-1}\left(\rho_0\mathrm{D}_t g+\left(\frac{\partial p}{\partial\tau}_{|\epsilon}\right)\partial_X\tau\cdot n_1\right), \end{array}\right. \tag{IV.47}$$

Solving (IV.46) amounts to solve five subsystems:

— The first system is built using the first equation of (IV.46), considering $\frac{m(m+1)}{2}$ interior cells. It leads to a $\left(\frac{m(m+1)}{2}\right)^2$ linear system and allows to build ghost-cell values of $\overline{\rho_0}$.

— The second system is built using the second equation of (IV.46), considering $\frac{m(m+1)}{2} - 1$ interior cells together with the boundary condition on the normal velocity. It leads to a $\left(\frac{m(m+1)}{2} - 1)\right)^2$ linear system and allows to build ghost-cell values of $\overline{\rho_0 u}$.

— The third system is built using the third equation of (IV.46), considering $\frac{m(m+1)}{2}$ interior cells. It leads to a $\left(\frac{m(m+1)}{2}\right)^2$ linear system and allows to build ghost-cell values of $\overline{\rho_0 v}$.

— The fourth system is built using the fourth equation of (IV.46), considering $\frac{m(m+1)}{2}$ interior cells. It leads to a $\left(\frac{m(m+1)}{2}\right)^2$ linear system and allows to build ghost-cell values of $\overline{\rho_0 \tau}$.

— The fifth system is built using the last equation of (IV.46) considering $\frac{m(m+1)}{2} - 1$ interior cells together with (IV.47). In the special case where $n_1 = 0$, no information is provided by the boundary conditions, and thus it leads to a $\left(\frac{m(m+1)}{2}\right)^2$ linear system. If $n_1 \neq 0$, this system is linear for ϵ-affine EOS but may be non-linear for some EOS, thus requiring fixed-point algorithms to be solved and the size of the system is $\left(\frac{m(m+1)}{2} - 1)\right)^2$. It allows to build ghost-cell values of $\overline{\rho_0 e}$.

Lemma IV.7 (Linear system for ϵ-affine EOS)**.** *For any ϵ-affine EOS, the system to inverse at the boundary is linear.*

Proof. The proof is similar to the one in 1D. ■

The following flowchart summarizes the algorithm we propose in order to compute ghost-cell values for a given dimensional sweep in the 2D case.

— For each point/pearl P_s:

1. Do the local change of velocity components in the basis $(\boldsymbol{n_s}, \boldsymbol{t_s})$,
2. Build the stencil of interior points (see Figure IV.7),
3. Build and solve the five subsystems described above.

— Then, for each ghost-cell:

1. Find the nearest pearl P_{s_0},
2. Build ghost-cell values using Taylor expansions in the vicinity of P_{s_0},
3. Return to physical coordinates.

Remark IV.5. Due to spurious oscillations and linear instabilities of the 2D extrapolations (phenomena already noticed in [155]), rather than solving exactly all subsystems, it proves useful to use least square methods for $m \geq 2$, adding more points inside the stencil. In pratice the stencil is depicted in figure IV.8 and set as

$$S^{\beta}_{P_s} = \{P \in \Omega_f, \|P - P_s\|_2 < \beta, \langle P - P_s, \boldsymbol{n}\rangle \geq \frac{1}{\sqrt{2}} \frac{\|P - P_s\|^2}{\beta}\}$$

Commonly, β is set to $0.9(2m-1)\sqrt{\Delta X \Delta Y}$.

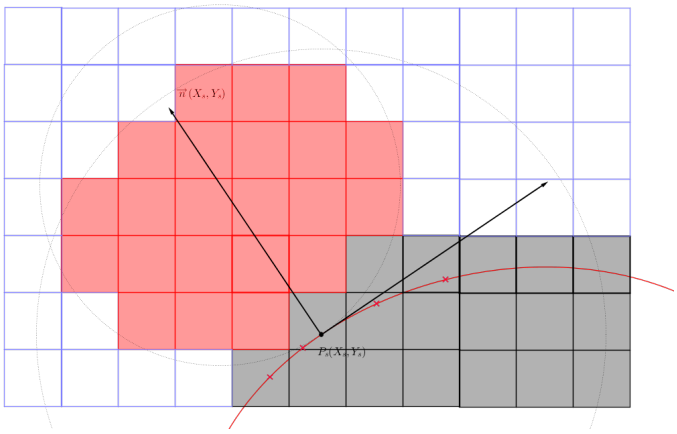


Figure IV.8 – Zoom on a point P_s on the discretized boundary with local coordinate system. The color zone corresponds to a least-squares stencil for 3rd order reconstruction.

IV-2.2 2D numerical validation

We assess in this part both the accuracy and the robustness of the method for the 2D Euler system. The study here is performed using the larger stencil based reconstruction applied to the GoHy schemes developed in [50, 170]. Similar results are obtained for the proposed staggered schemes introduced in [35] and detailed in chapter II, as well for smooth flows as for shock problems.

The discretization of the boundary Γ is always set such that the distance between two consecutive points does not exceed $C_\Gamma\sqrt{\Delta x \Delta y}$. In the following we set $C_\Gamma = 1$ which means that we have approximatively one pearl per cell. In practice, a large value of C_Γ leads to instabilities (boundaries are under-resolved). A smaller value of C_Γ is possible, increases accuracy but leads to heavier computations. The choice of this test-suite is made in order to ensure a large variety of test-cases, including continuous and isentropic flows, acoustic propagation around an obstacle, but also a large variety of shock impacting on infinite motionless obstacles with shapes that may or may not be Lipschitz continuous.

IV-2.2.1 2D isentropic vortex test-case [174]

We assess high-order accuracy on the 2D vortex test (see [174]) whose initial condition is reminded hereafter (with $r^2 = x^2 + y^2$)

$$\left\{\begin{array}{lcl} \rho_0(x,y) & = & \left(1 - \dfrac{(\gamma-1)\beta^2}{8\gamma\pi^2} e^{1-r^2}\right)^{\frac{1}{\gamma-1}}, \\ \boldsymbol{u}_0(x,y) & = & \dfrac{\beta}{2\pi} e^{\frac{1-r^2}{2}} \cdot (-y, x)^t, \\ p_0(x,y) & = & \rho_0(x,y)^\gamma, \end{array}\right. \tag{IV.48}$$

with $\gamma = 1.4$ and $\beta = 5$. Computations are performed on a disk of radius $R = 3.5$, centered at $(0,0)$ till $t = 1$ with a CFL number of 0.9 on the computational domain $\Omega = [-4,4]^2$. Boundary

conditions imposed at $R = 3.5$ are $\boldsymbol{u} \cdot \boldsymbol{n} = \boldsymbol{u}_0 \cdot \boldsymbol{n}$, with $\boldsymbol{u}_0$ defined in eq. (IV.48). Table IV.2 shows that the expected order of accuracy is reached. In the third column we also have reported numerical costs due to the ILW procedure, computed as the ratio between CPU time for ILW procedure and total CPU time as was done in [155]. These should of course be analysed cautiously since they strongly depend on the inner scheme and optimization of the boundary treatment (as well as the number of considered pearls on Γ). However, these figures show that the cost slope for the 1st-order ILW method is a bit less than one with respect to the number of cells per dimension. Second order ILW procedure cost slope is around 0.75 and for third order ILW procedure, it is about 0.66. One may guess the cost to follow the rule $\sim \frac{m+1}{2m}$.

N_x	GoHy-1			GoHy-2			GoHy-3		
50	4.96e-1	·	35%	5.33e-2	·	47%	9.93e-2	·	49%
100	2.52e-1	0.97	23%	1.40e-2	1.93	42%	2.04e-2	2.28	45%
200	1.20e-1	1.07	12%	4.50e-3	1.63	27%	3.46e-3	2.56	35%
400	5.66e-2	1.08	7%	1.28e-3	1.81	16%	6.43e-4	2.43	22%
800	2.74e-2	1.05	3.7%	3.23e-4	1.99	9.7%	9.31e-5	2.79	14%
1600	1.35e-2	1.03	1.9%	7.66e-5	2.08	6.2%	1.20e-5	2.95	9%
3200	6.70e-3	1.01	1.0%	1.90e-5	2.01	3.7%	1.51e-6	2.99	5%

Table IV.2 – l^1-error on density in both time and space, experimental order of convergence and cost in % of the ILW procedure for GoHy schemes on the 2D isentropic vortex at $t = 1.0$.

IV-2.2.2 Acoustic diffraction of a plane wave around a cylinder [15]

Next test-case is a challenging problem coming from the electromagnetic and aeroacoustic communities. Here we wish to assess the interest of increasing the order of accuracy of boundaries treatments. A plane acoustic wave is propagating in a barotropic gas and is scattered by a rigid and motionless cylinder. The main interest of this test lies in the fact that an analytical solution is available, in particular the pressure field on the cylinder.

The computational domain is $[-5, 5] \times [-5, 5]$ and the rigid wall boundary condition $\boldsymbol{u} \cdot \boldsymbol{n} = 0$ is applied on the rigid body which is a cylinder of radius $a = 0.5$ whose center is located at $(0, 0)$. Let ω be the frequency of the acoustic signal and $k = \frac{\omega}{c}$ the associated wave number, where c is the sound speed. The velocity potential of the incident wave is given by

$$\phi_0(t, x, y) = -\frac{\epsilon}{k} \cos(k(x - x_0) - \omega t)\chi_{\{x-ct<x_0\}}, \tag{IV.49}$$

from which we deduce the velocity, the pressure and the density according to

$$\boldsymbol{u} = \Re(\nabla\phi), \quad p = p_0 - \Re(\partial_t\phi), \quad \rho = \gamma p^{\frac{1}{\gamma}} \text{ with } \gamma = 1.4. \tag{IV.50}$$

For this test we took $\epsilon = 10^{-8}$ small enough so that Euler equations remain in the linear regime and approximate wave equations, $p_0 = \frac{1}{\gamma}$, $x_0 = 4.5$, and $\omega = 2\pi f$, $f \in 0.1\mathbb{N}$ (*ie* there exists $b \in \mathbb{N}$ such that $f = 0.1b$). Introducing $x = r\cos(\theta)$ and $y = r\sin(\theta)$, the harmonic solution for

the velocity potential is given by (see [15])

$$\phi(t,x,y) = -\frac{\epsilon}{k}\,\mathrm{e}^{-i(kx_0+\omega t)}\sum_{n=0}^{\infty} e_n i^n \left[J_n(kr) - \frac{J_n'(ka)}{H_n'(ka)}H_n(kr)\right]\cos(n\theta), \tag{IV.51}$$

where J_n is the first Bessel function, H_n the first Hankel function and $e_0 = 1, e_n = 2, n = 1, .., \infty$. From this potential, one gets harmonic velocity $\boldsymbol{u}$ and pressure p thanks to (IV.50).

To ensure a harmonic regime in a neighbourhood of the cylinder without generating interferences with the computational domain boundaries, the final time is $t = 8.4$. We give on figure IV.9 pressure variations $|p - p_0|$ around the cylinder for 1st, 2nd and 3rd-order ILW methods and schemes for two space discretizations ($\Delta x = \Delta y = \frac{1}{20}$ and $\frac{1}{40}$) and three signal frequencies ($f = 0.5$, 1 and 2). As expected, it shows that high-order accurate methods lead to better results. But since interior schemes are also of different orders, it is hard to see benefits given by ILW methods of increasing accuracy here. We therefore give on figure IV.10 pressure variations $|p - p_0|$ around the cylinder for 1st, 2nd and 3rd-order ILW methods, but with the same 3rd-order GoHy-3 interior scheme in all cases. Results indeed show the benefits of formally 3rd-order accurate ILW reconstruction procedures.

IV-2.2.3 Reflected shock wave

This test-case is a theoretical one. It consists of a shock wave impacting an oblique wall. The computation domain is $\Omega = [-0.5 : 0.5]^2$. An oblique wall is parametrized by the starting point $(-0.2, 0)$ forming an angle θ with the horizontal plane, with $\theta = 11.99°$. The shock wave is reflected with an angle β. Denote M the Mach number, $M = \frac{u}{c}$, then β satisfies the following identity

$$\tan(\theta) = 2\cot(\beta)\frac{M^2\sin^2(\beta) - 1}{M^2(\gamma + \cos(2\beta)) + 2}. \tag{IV.52}$$

Initial data are

$$\begin{cases} \rho &= 1, \\ u &= 2.9\chi_{x<-0.3}, \\ p &= \frac{1}{\gamma}, \\ \gamma &= \frac{7}{5}, \end{cases} \tag{IV.53}$$

which gives a static speed sound $c = 1$, and so a Mach number $M = \frac{u}{c} = 2.9$.

Using eq. (IV.52), one finds that for such parameters, the angle formed by the oblique shock β is 30°. In fig. IV.11, the density profile is depicted at time $t = 1$, and the expected angle of the reflected shock is depicted by the white line. The expected angle is reached by the first, second and third order proposed effective schemes. Moreover, the fluid perfectly slips along the boundary without any boundary effects.

IV-2.2.4 Double Mach Reflection [171]

The ILW procedure is again applied on solid wall boundaries that may be curved or unaligned with the grid. For inviscid flows this leads to the boundary condition $\boldsymbol{u} \cdot \boldsymbol{n} = 0$. The first shock example considered here is the double Mach reflection problem [171, 155]. A solid wall is set at $(0, 0)$ forming a $30°$ angle with the x-axis and a horizontally moving Mach 10 shock, initially located at $x = 0$, is propagating in a perfect gas ($\gamma = 1.4$) at rest. Ahead of the shock, the gas has a density of 1.4 and a pressure of 1. The computational domain $[-1, 3] \times [0, 2]$ is discretized with a constant space step $\Delta x = \Delta y = \frac{1}{200}$. The choice of such a coarse mesh is done to easily point out differences between the different orders of accuracy.

Results, depicted in Figure IV.12, are very close to those found in the literature [171, 155] and the jet propagates along the wall without any numerical friction. For this test we have used the MOOD procedure (see section IV-1.3.1) to decrease the order of accuracy wherever we encountered stability issues. In practice this only happens near the wall in the immediate vicinity of the Mach stem propagating perpendicularly to it.

IV-2.2.5 Mach shock on a cylinder – Whitham test-case [23]

We now consider the Whitham test-case which consists in a planar shock propagating in a perfect gas ($\gamma = 1.4$) which interacts with a rigid and motionless circular cylinder (see [23] and included references). At $t = 0$, a 2.81 Mach shock coming from the left is located at $x = 0$. Ahead of the shock, the gas has a density of 1 and a pressure of 5 10^4. The cylinder's center, whose radius is 5.10^{-3}, is located at $(6.10^{-3}, 0)$. The computational domain $[-10.10^{-3}, 70.10^{-3}] \times [-40.10^{-3}, 40.10^{-3}]$ is discretized with a constant space step $\Delta x = \Delta y = 4.10^{-4}$.

Here again a MOOD method is used on the boundary to improve robustness. Combined with high-order accuracy this leads to a better restitution of the flow structure behind the cylinder as it can be seen in Figure IV.13 where 1st, 2nd and 3rd-order results at $t = 3.10{-}5$ and $= 6.10{-}5$ are reported. The bow shock is well captured and less diffused as the order of accuracy is increased. The MOOD procedure applies essentially on the shock front.

IV-2.2.6 Mach shock on a prism – Schardin test-case [23]

We now consider the Schardin test-case which consists in a planar shock propagating in a perfect gas ($\gamma = 1.4$) which interacts with a rigid and motionless prism (see [23] and included references). At $t = 0$, a 1.3 Mach shock coming from the left is located at $x = 0$. Ahead of the shock, the gas has a density of 1 and a pressure of 5.10^4. The prism's tip is located at $(1.5.10^{-2}, 0)$ and the edge length is set to 20.10^{-3}. The computational domain $[-10.10^{-3}, 70.10^{-3}] \times [-40.10^{-3}, 40.10^{-3}]$ is discretized with a constant space step $\Delta x = \Delta y = 4.10^{-4}$.

A MOOD method is used on the boundary to improve robustness. Combined with high-order accuracy this leads to a better restitution of the flow structure behind the prism as it can be seen in Figure IV.13 where 1st, 2nd and 3rd-order results at $t = 1.5.10^{-4}$ are reported. The bow shock

is well captured and less diffused as the order of accuracy is increased. The MOOD procedure applies essentially on the shock front. The expected structure of the flow is recovered, especially the presence of vortices behind the prism.

IV-2.2.7 Mach shock on a NACA0018 profile [88]

We now consider a classical aerodynamics test-case which consists in a planar shock propagating in a perfect gas ($\gamma = 1.4$) which interacts with a rigid and motionless NACA0018 airfoil with a $30°$ angle of attack (see [88] and included references). At $t = 0$, a 1.5 Mach shock coming from the left is located at $x = 0.55$. Ahead of the shock, the gas has a density of 1.4 and a pressure of 1. The airfoil's head is located at $(0.6, 1)$ and the chord length is set to 1. The computational domain $[-0.2, 1.8] \times [0, 2]$ is discretized with 100, 200 and 400 cells in each direction.

Figure IV.15 shows the obtained results for the first, second and third order schemes on a 400×400 grid at time $t = 0.64$. These results are in good agreement with the results provided in [88] concerning the shock structure. As the order is increased, the shock front is sharper but also more oscillatory, and flow structures near both tip and head of the airfoil are better recovered.

Imposing free stream velocity u_∞ and density ρ_∞ with the post-shock values, both lift C_l and drag C_d coefficients are computed using

$$\begin{pmatrix} C_d \\ C_l \end{pmatrix} = -\frac{2}{\rho_\infty u_\infty^2 L} \int_\Gamma (p - p_0)\boldsymbol{n}\mathrm{d}S. \tag{IV.54}$$

where L is the chord of the airfoil, set here to 1. The computed lift and drag coefficients are depicted in fig. IV.16 as a function of time for different grid sizes. For both schemes, the convergence error in the drag coefficient appears to be linear while more than quadratic convergence seems to be reached for the lift coefficient.

How to adapt the method to the staggered schemes

To tackle the procedure for the discretization of boundary conditions in the case of staggered schemes, two key ingredients are required. The first one is that the Taylor expansion of the total energy variable is replaced by the Taylor expansion of the internal energy. The second one is that Taylor expansions are performed on variables which are located on two (resp. three) different grids in 1D (resp. 2D). Lemma III.2 details how to build boundary conditions at intermediate Runge–Kutta time-steps.

Considering again the acoustic diffraction test-case presented in section IV-2.2.2, comparisons are drawn between the results obtained with the GoHy-3 scheme and the third order staggered scheme (STAG-3). Results are displayed in fig. IV.17. Pressure variations are very close for all frequencies f for both schemes.

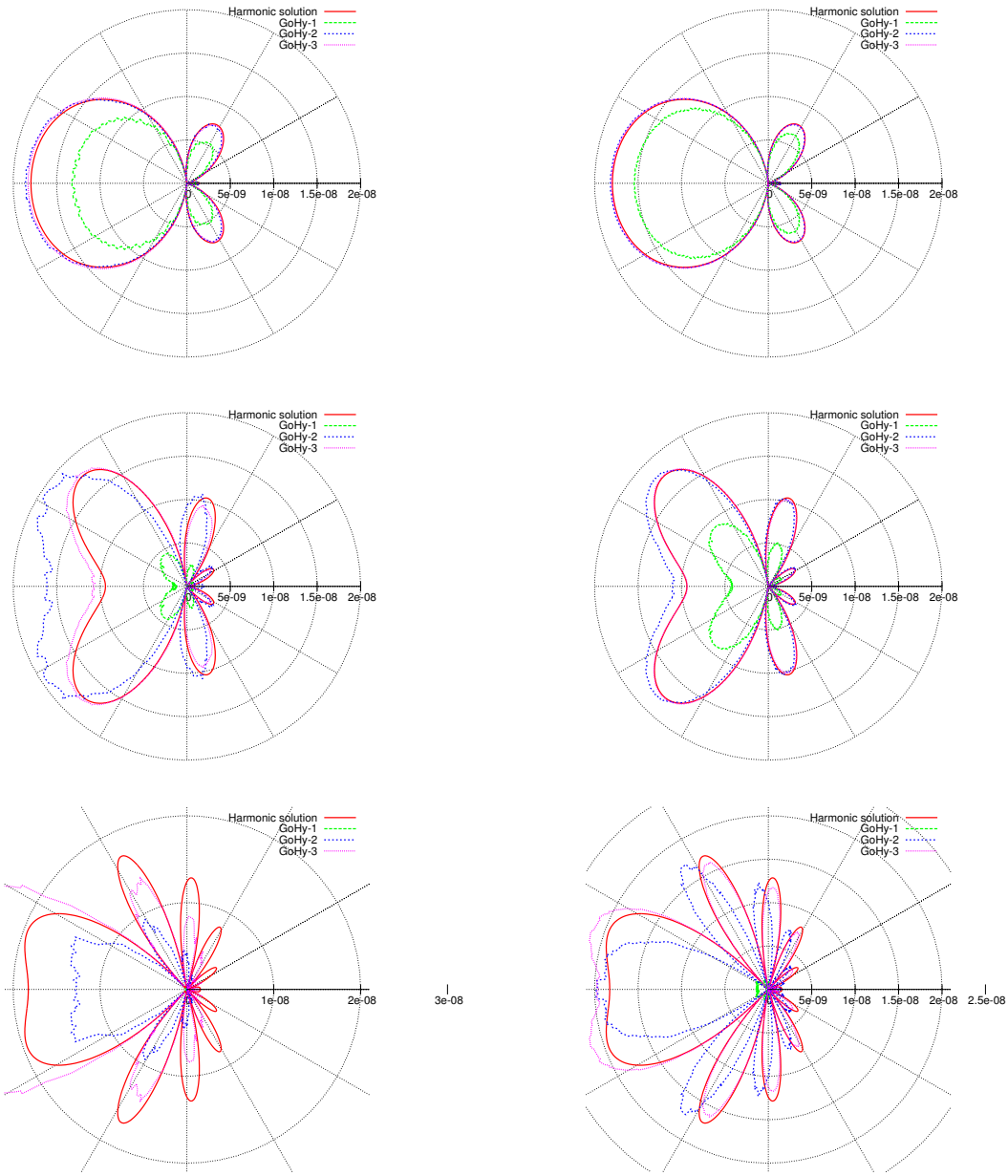


Figure IV.9 – Pressure variations $|p-p_0|$ around the cylinder as a function of θ for $f = 0.5$ (top), $f = 1.0$ (middle), $f = 2.0$ (bottom) for 1st, 2nd and 3rd-order accurate schemes with $\Delta x = \Delta y = \frac{1}{20}$ (left) and $\Delta x = \Delta y = \frac{1}{40}$ (right).

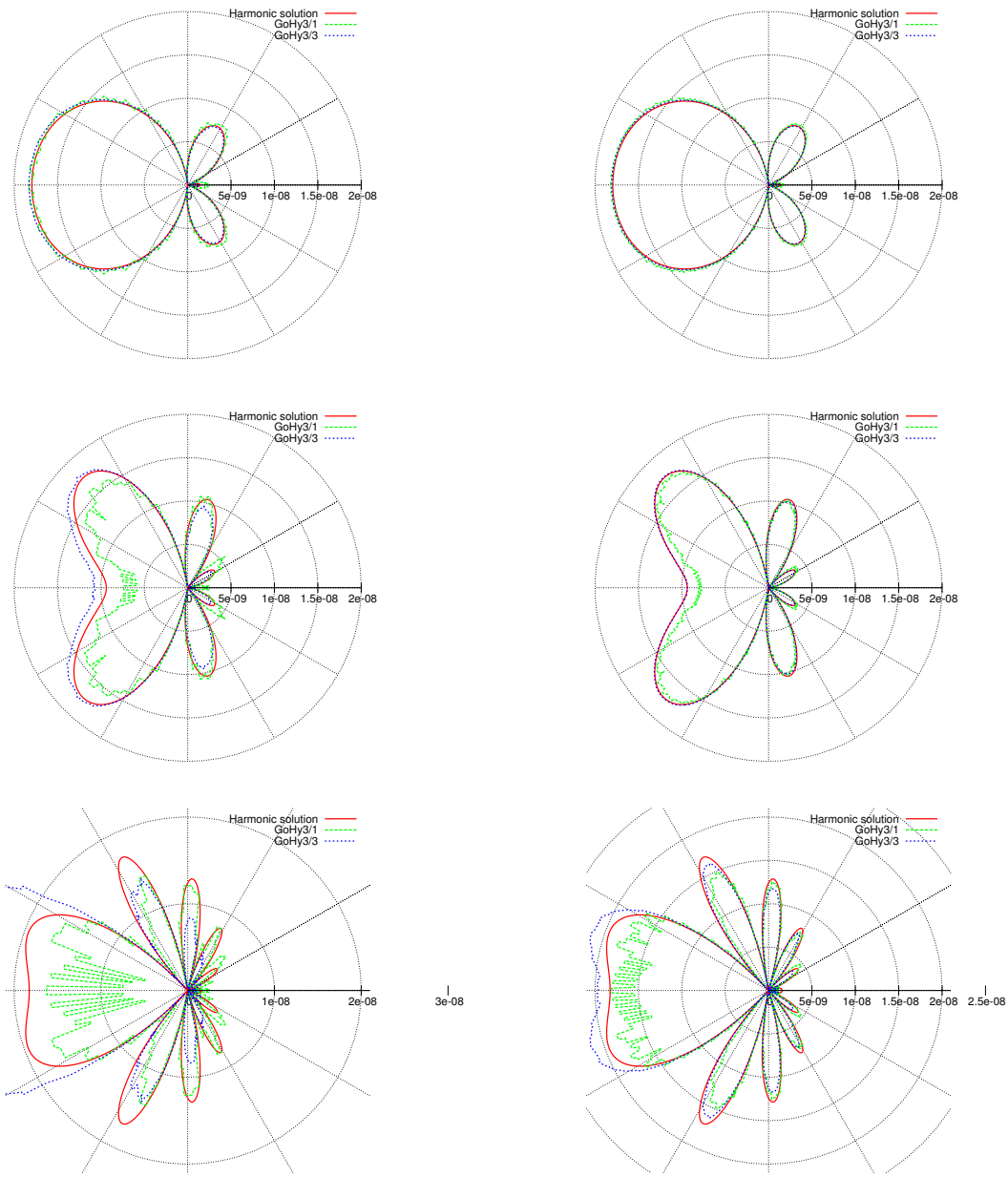
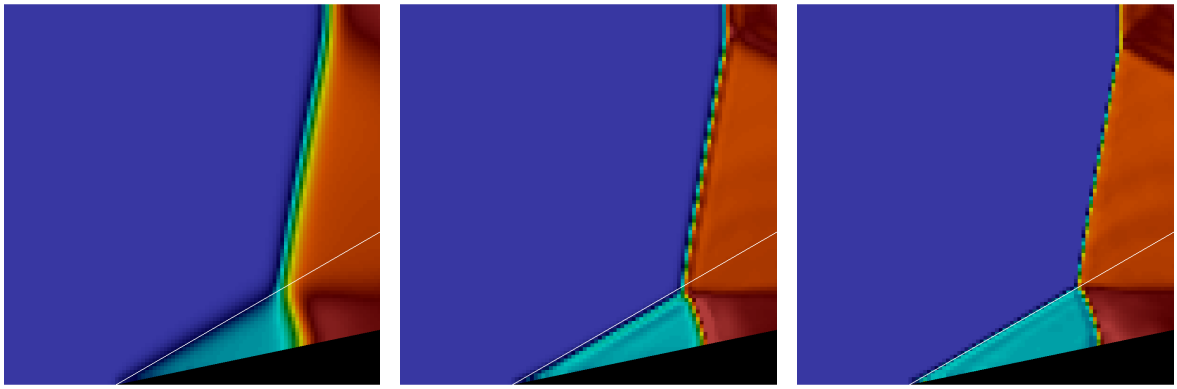
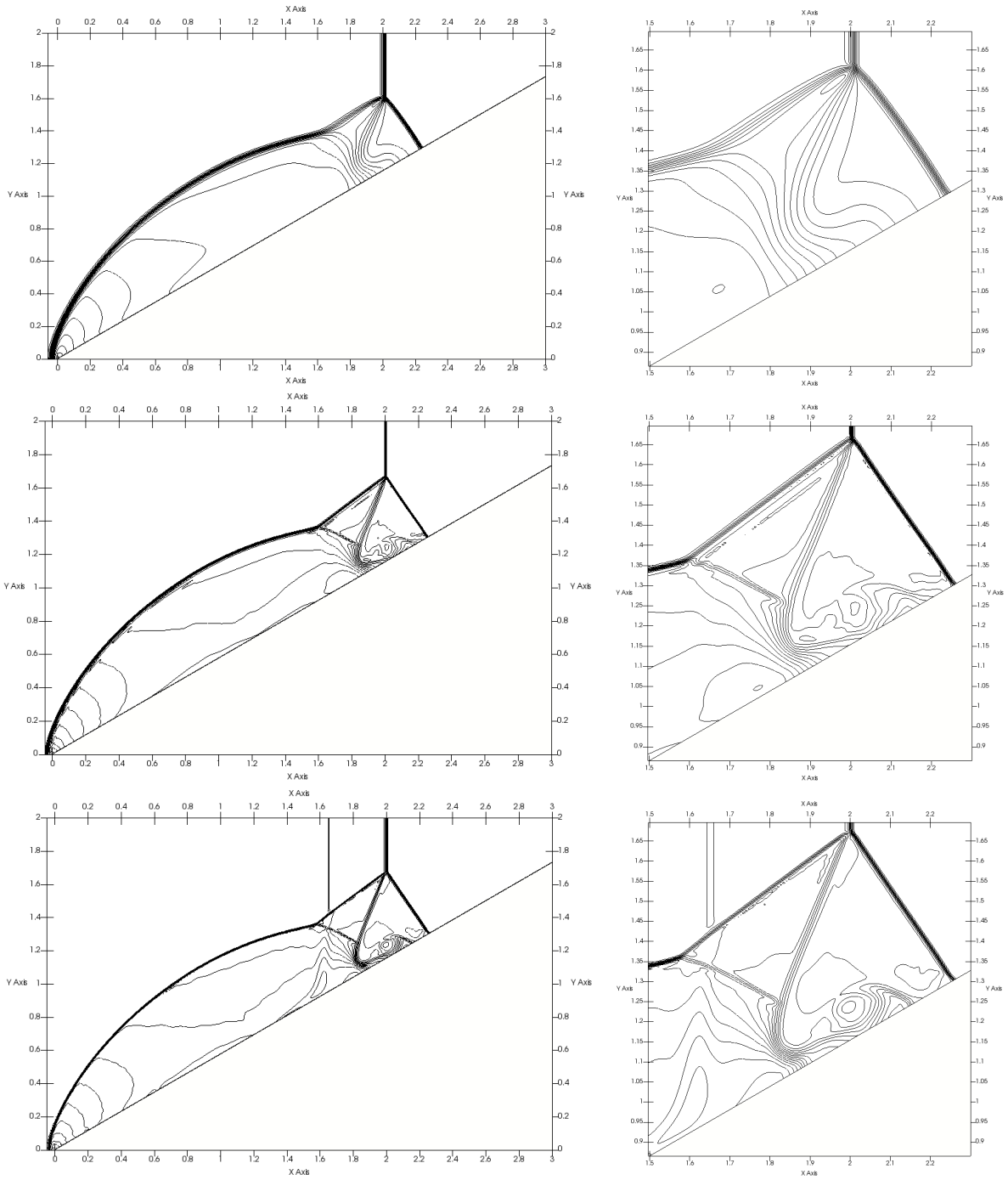


Figure IV.10 – Pressure variations $|p - p_0|$ around the cylinder as a function of θ for $f = 0.5$ (top), $f = 1.0$ (middle), $f = 2.0$ (bottom) for the GoHy-3 interior scheme and 1st, 2nd and 3rd-order accurate ILW methods with $\Delta x = \Delta y = \frac{1}{20}$ (left) and $\Delta x = \Delta y = \frac{1}{40}$ (right).



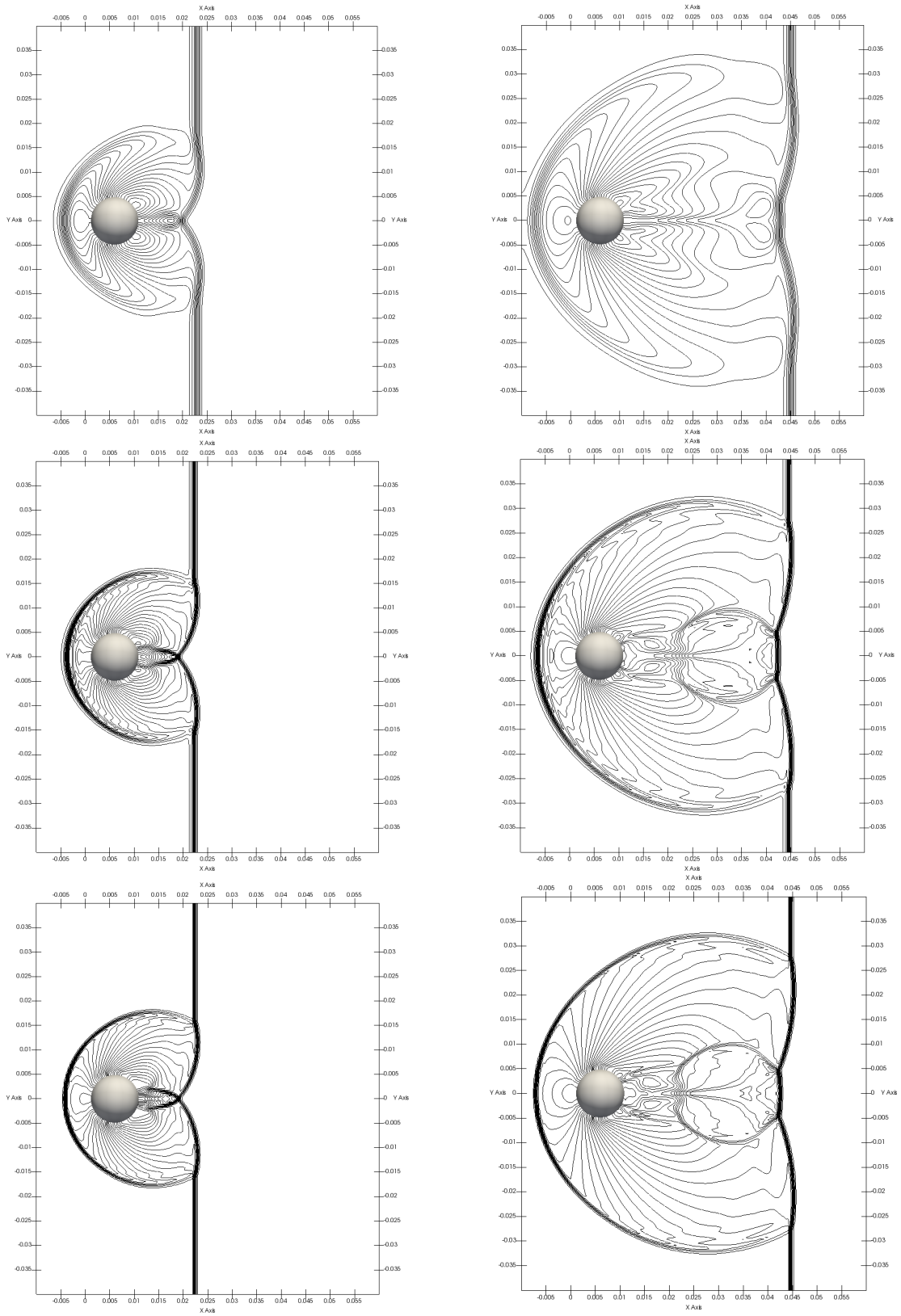
(a) 1st order (b) 2nd order (c) 3rd order

Figure IV.11 – Density colors of a reflected shock wave on a wedge at CFL=0.5 with 100 cells in each direction. The expected angle of the oblique shock, depicted by the white line, is recovered by the schemes.



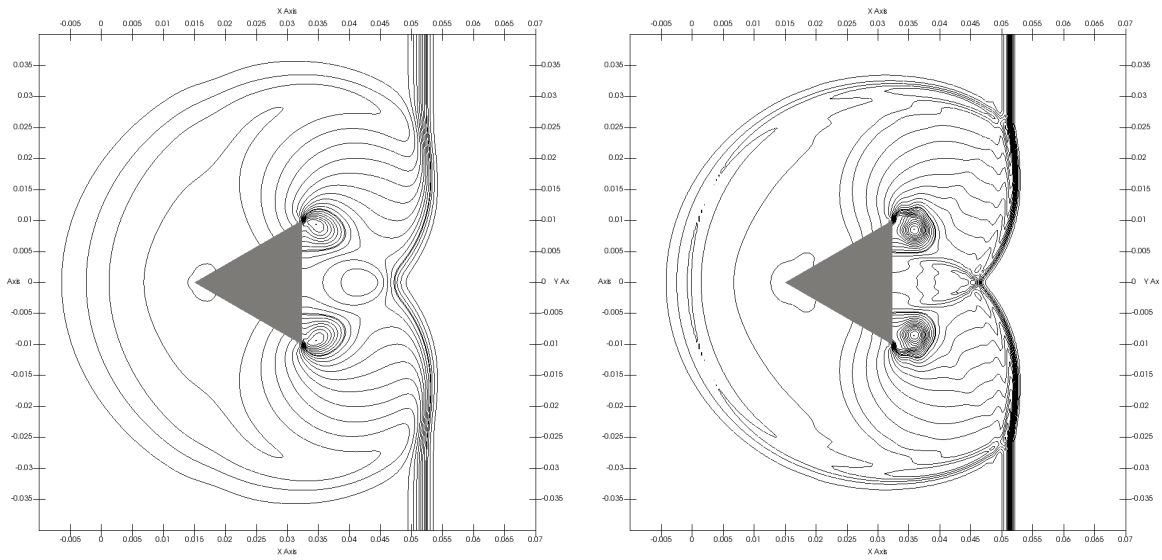
(a) Whole domain (b) Zoom near the double Mach stem

Figure IV.12 – Density contours of double Mach reflection for 1st (top), 2nd (middle) and 3rd-order (bottom) ILW-GoHy schemes with $\Delta x = \Delta y = \frac{1}{200}$; 30 contours from 1.731 to 20.92 as in [155].

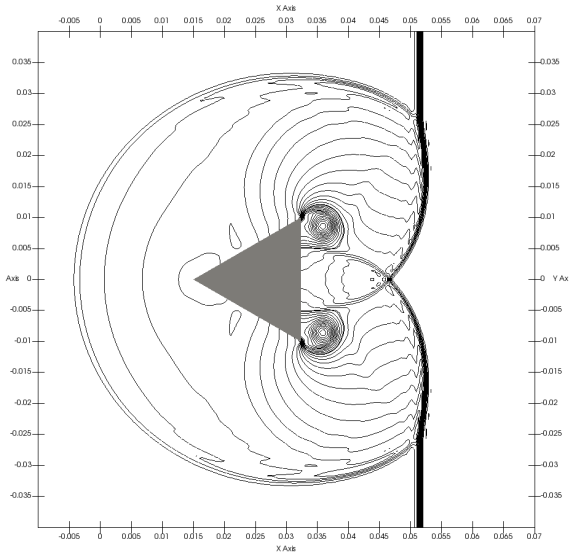


(a) T = 30 μs (b) T = 60 μs

Figure IV.13 – Density contours of Mach 2.81 flow past a cylinder for 1st (top), 2nd (middle) and 3rd-order (bottom) ILW-GoHy schemes with $\Delta x = \Delta y = 4.10^{-4}$ at $t = 3.10^{-5}$ (left) and $= 6.10^{-5}$ (right); 30 contours from 0.3 to 8.

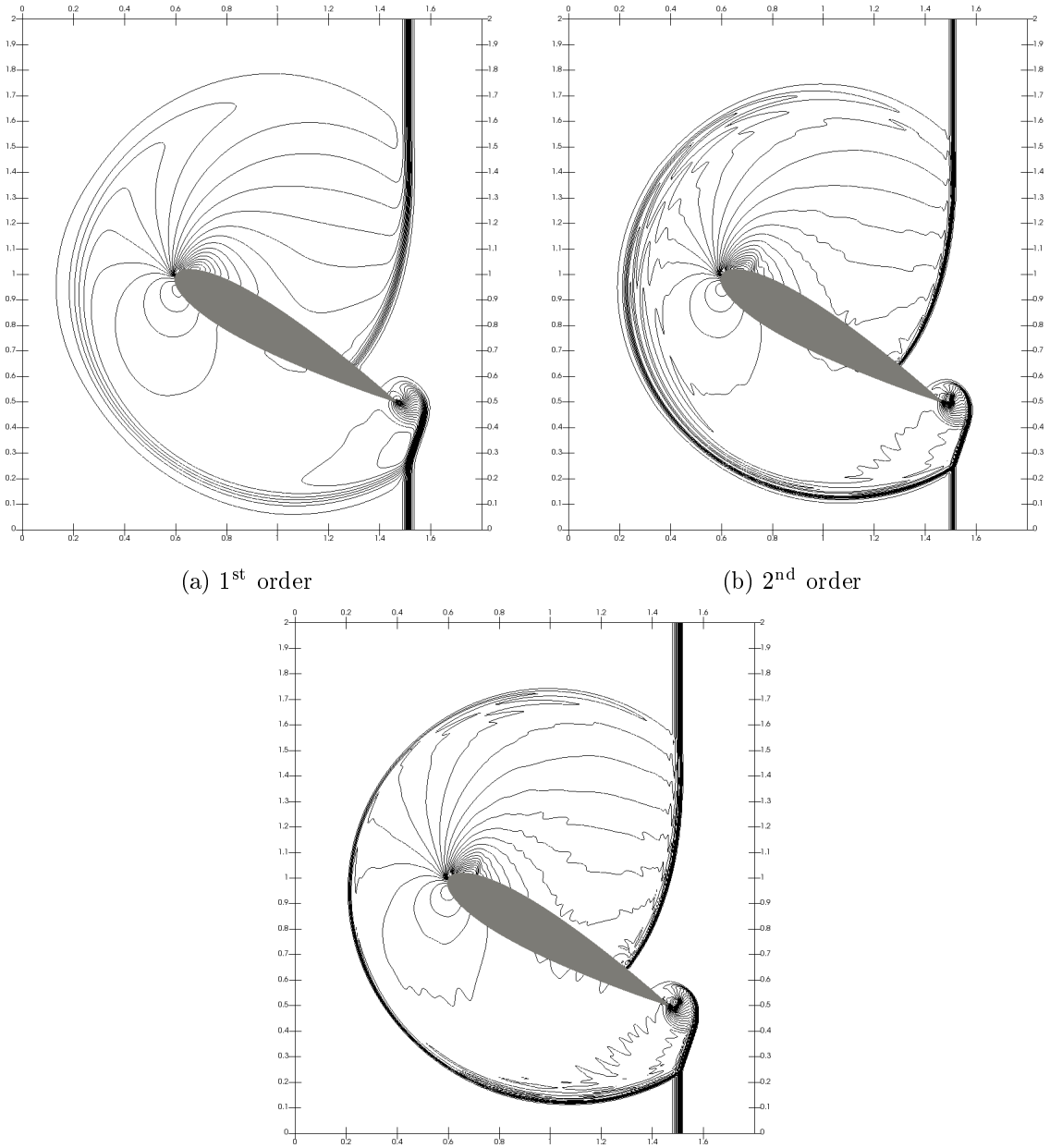


(a) 1st order (b) 2nd order



(c) 3rd order

Figure IV.14 – Density contours of Mach 1.3 flow past a prism for 1st (top, left), 2nd (top, right) and 3rd-order (bottom) ILW-GoHy schemes with $\Delta x = \Delta y = 4.10^{-4}$ at $t = 1.5.10{-}4$, CFL=0.5; 30 contours from 0.5 to 1.8

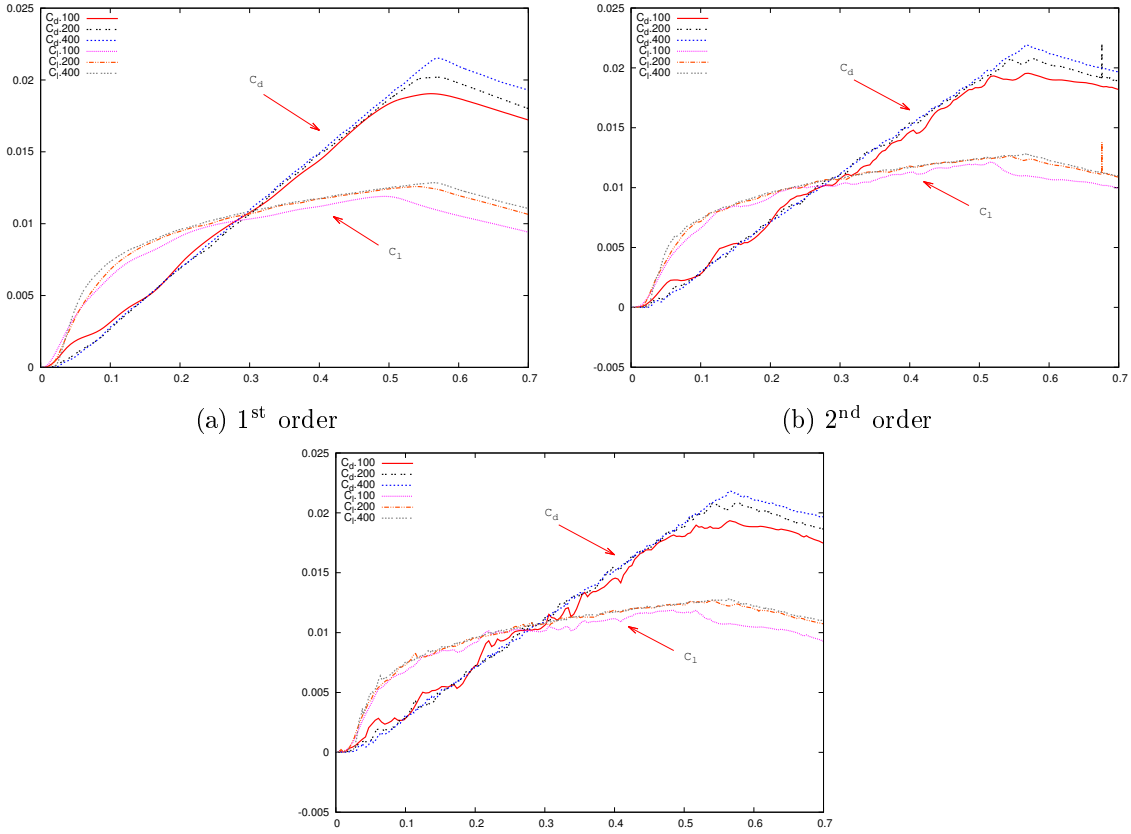


(a) 1st order

(b) 2nd order

(c) 3rd order

Figure IV.15 – Pressure contours of a Mach shock on a NACA0018 for 1st (top, left), 2nd (top, right) and 3rd-order (bottom) ILW-GoHy schemes with 400 cells in each direction, CFL=0.5; 35 contours from 0.0 to 3.5



(a) 1st order

(b) 2nd order

(c) 3rd order

Figure IV.16 – Lift and drag coefficients as a function of time for the Mach shock on the NACA0018 profile considering 100, 200 and 400 cells in each direction for 1st (top, left), 2nd (top, right) and 3rd-order (bottom) ILW-GoHy schemes.

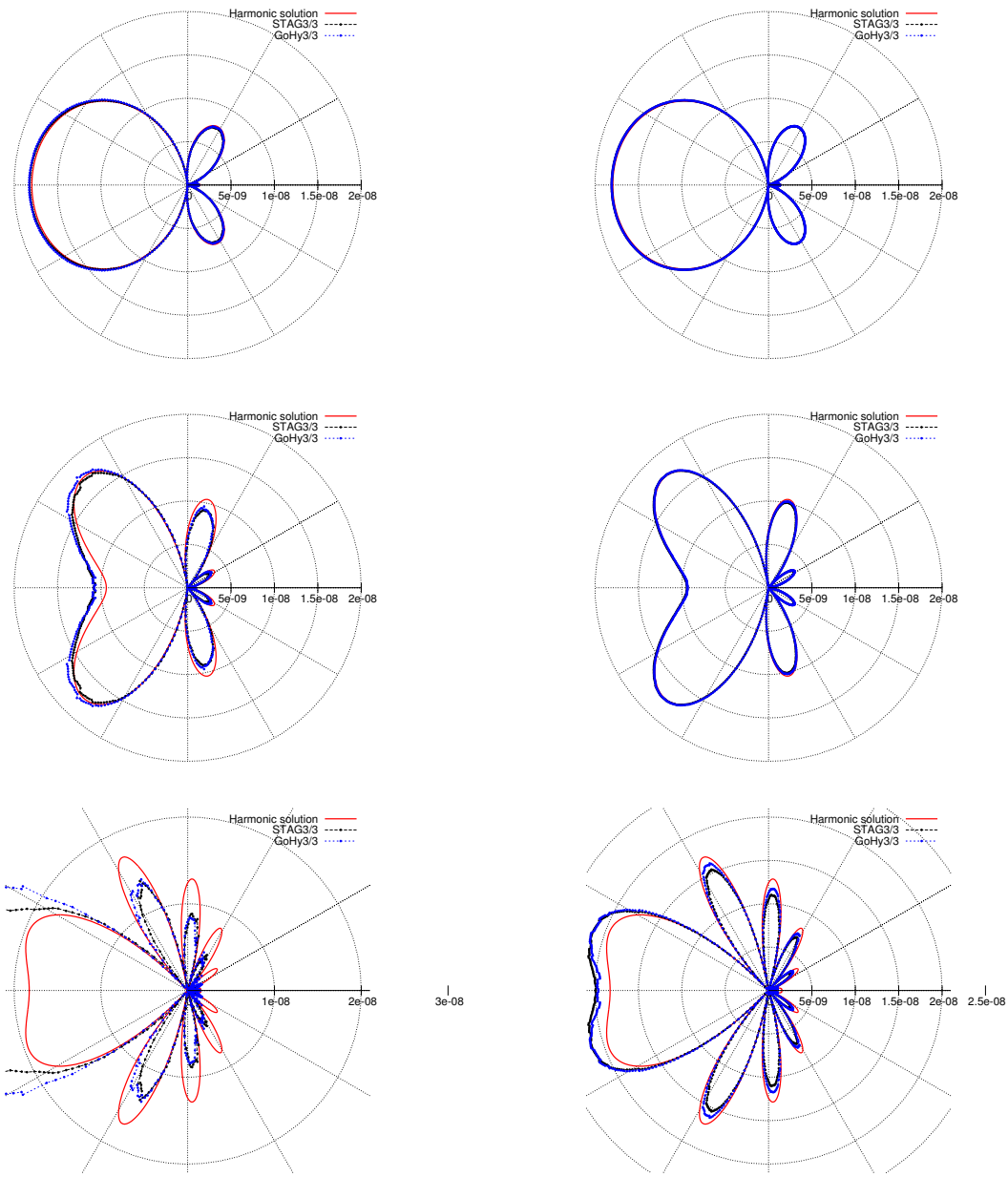


Figure IV.17 – Pressure variations $|p - p_0|$ around the cylinder as a function of θ for $f = 0.5$ (top), $f = 1.0$ (middle), $f = 2.0$ (bottom) for the third order cell-centered scheme (GoHy-3, blue) and for the third order staggered scheme (STAG-3, black) with $\Delta x = \Delta y = \frac{1}{20}$ (left) and $\Delta x = \Delta y = \frac{1}{40}$ (right).

Chapter V

Extension to fluid-rigid body interaction

Partant de la procédure de Lax–Wendroff inverse établie pour les équations d'Euler présentée dans le chapitre IV, un algorithme de couplage fluide-corps rigide est déduit. Après une courte introduction concernant les caractéristiques physiques et mathématiques du mouvement de corps rigide, un schéma semi-discret permettant de calculer à l'ordre élevé en espace les forces et moments exercés sur la frontière du corps rigide est proposé. Deux procédures d'intégrations en temps sont ensuite développées. La première est basée, tout comme les schémas hydrodynamiques présentés dans le chapitre II, sur une intégration en temps de type Runge–Kutta. La seconde est basée sur une approche de type Cauchy–Kowalevski comme dans [50, 170]. Ce choix d'intégration en temps permet de faire correspondre sur la même échelle en temps les deux solveurs. Enfin l'extension 2D de ces schémas est ensuite faite via splitting directionnel comme pour les schémas hydrodynamiques utilisés. La procédure de Lax–Wendroff inverse donne une définition naturelle des forces et moments de pression exercés sur la frontière du corps rigide. Ainsi le couplage est immédiat et d'autant plus facile à implémenter. Des résultats numériques sont proposés à la fin du chapitre afin d'illustrer la stabilité et la robustesse du couplage utilisé.

In this chapter, we propose a simple and straightforward way for coupling rigid body and compressible fluid dynamics. Considering rigid body dynamics, a semi-discrete scheme is first proposed for 1D motion, then for 2D motion using directional splitting method. The computations of forces and torques is done considering a regular discretization of the boundary. Such a discretization enables for a high-order accurate way of computing the forces and torques integrals along the boundary. Two fully discrete version are then proposed. Mostly those versions strongly rely on the hydrodynamics schemes used. Indeed, using a one-step cell-centered schemes [50, 170], a one-step scheme is proposed for the integration of forces and torques exerted on the rigid body boundary. As a contrary, using the staggered schemes introduced in [35] and extensively detailed in chapter II. The coupling between fluid and solid is then straightforward using the ILW procedure developed in chapter IV.

The outline of the chapter is the following. First, an overview of rigid body motion and dynamics is proposed in section V-1. Then, starting from a semi-discrete high-order accurate in space scheme, two time integration are proposed in section V-2. The first one is based on Runge–Kutta sequences, whereas the second is based on Cauchy–Kovalevskaya time-integration. The extension to 2D relies on directional splitting method. The choice has been made for both schemes to match the time-integration used for the hydrodynamics ones. This is done to avoid any loss of accuracy due to the time-coupling. Last the coupling between the fluid and rigid body solvers is done using the Inverse Lax–Wendroff procedure designed in chapter II. The procedure gives naturally definition of the pressure forces and torques exerted on the rigid body boundary. Thus, the coupling method is straightforward and quite easy to implement. Numerical examples are proposed then in 1D and 2D to assert the viability of the coupling.

V-1 Rigid body motion and dynamics 192
V-1.1 Description of a rigid body 193
V-1.2 Immersed rigid body dynamics 194
V-2 High-order Lagrangian schemes for rigid body dynamics 195
V-2.1 High-order schemes for rigid body dynamics in 1D 195
V-2.2 High-order schemes for rigid body dynamics in 2D 196
V-3 Fluid - Rigid body coupling 203
V-3.1 Description of the algorithm 203
V-3.2 Numerical results 204

V-1 Rigid body motion and dynamics

We consider that the motion of the boundary is no longer prescribed analytically. Instead we consider the boundary Γ to be the boundary of a rigid body whose mass is finite. Its motion is then induced by the forces exerted by the fluid on the boundary. One may refer to [89, 54] for further informations concerning rigid body motion and dynamics.

Space dimension	Number of variables
d = 1	1
d = 2	3
d = 3	6

Table V.1 – Number of variables for rigid body motion as a function of given space dimensions

V-1.1 Description of a rigid body

In physics, a rigid body is considered as a body where no deformation can be induced in it. Consider two points (or particles) belonging to the rigid body, denoted by the greek subscript α and β. Then, for any α and β, rigid body constraint writes

$$\|\boldsymbol{x}_\alpha - \boldsymbol{x}_\beta\| = \text{constant}, \tag{V.1}$$

meaning that the distance separating two abstract points α and β in a rigid body is always constant.

V-1.1.1 Invariant of rigid body motion

Using only eq. (V.1), one can prove that for any space dimension d, the rigid body motion can be reduced to solving $d + (\frac{1}{2}d(d-1))$ equations [14]. It implies in particular that the rigid body motion is described by a set of $d + (\frac{1}{2}d(d-1))$ variables.

In particular rigid body motion can be described as

$$\mathrm{D}_t\boldsymbol{x}_\alpha = \mathrm{D}_t\boldsymbol{x}_0 + \underline{\boldsymbol{Q}}(t)\boldsymbol{x}_\alpha, \tag{V.2}$$

where $\boldsymbol{x_0}$ is in the rigid body, $\underline{\boldsymbol{Q}}$ is antisymmetric, meaning that $\underline{\boldsymbol{Q}}(t) = -\underline{\boldsymbol{Q}}(t)^t$. In the following, only one and two space dimensions problems are considered. For one space dimension, eq. (V.2) is reduced to

$$\mathrm{D}_t x_\alpha = \mathrm{D}_t x_0, \tag{V.3}$$

since the only antisymmetric matrix in one space dimension is 0. Physically, it implies that the only possible motion for a rigid body in 1D is a translation. However, in two space dimensions, eq. (V.2) leads to

$$\mathrm{D}_t \begin{pmatrix} x_\alpha \\ y_\alpha \end{pmatrix} = \mathrm{D}_t \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + \begin{pmatrix} 0 & -q \\ q & 0 \end{pmatrix} \begin{pmatrix} x_\alpha \\ y_\alpha \end{pmatrix}, \tag{V.4}$$

which leads to a translation and a rotation. More often than not, eq. (V.4) is written under the more convenient form

$$\mathrm{D}_t \begin{pmatrix} x_\alpha \\ y_\alpha \end{pmatrix} = \mathrm{D}_t \begin{pmatrix} x_s \\ y_s \end{pmatrix} + \begin{pmatrix} 0 & -\omega \\ \omega & 0 \end{pmatrix} \begin{pmatrix} x_\alpha - x_s \\ y_\alpha - y_s \end{pmatrix}. \tag{V.5}$$

where the point $\boldsymbol{x}_s$ is called the center of mass and is only translated. In addition to the

description of the rigid body motion, some quantities must be defined to study the rigid body dynamics.

V-1.1.2 Definition of physical quantities

Consider a rigid body whose motion is prescribed by eq. (V.5), which is described by a bounded domain Ω_s of $\mathbb{R}^2$. Given a positive bounded function ρ_s which described the material density of the rigid body, then one defines the solid mass M_s, the gravity center $\boldsymbol{x}_s$ and the moment of inertia J_s as

$$\left\{\begin{array}{lcl} M_s & = & \displaystyle\int_{\Omega_s} \rho_s(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \\ \boldsymbol{x}_s & = & \displaystyle\frac{1}{M_s}\int_{\Omega_s} \rho_s(\boldsymbol{x})\boldsymbol{x}\mathrm{d}\boldsymbol{x} \\ J_s & = & \displaystyle\int_{\Omega_s} \rho_s(\boldsymbol{x})\|\boldsymbol{x}-\boldsymbol{x}_s\|^2\mathrm{d}\boldsymbol{x} \end{array}\right. \tag{V.6}$$

And at last, let $\boldsymbol{u}_s = \mathrm{D}_t\boldsymbol{x}_s$, one defines the kinetic energy of the rigid body as

$$\mathcal{E}_s = \frac{1}{2}M_s\|\boldsymbol{u}_s\|^2 + \frac{1}{2}J_s\omega^2. \tag{V.7}$$

V-1.2 Immersed rigid body dynamics

Using the previously defined quantities, one writes the system of equations describing the rigid body dynamics, without any external forces, as

$$\left\{\begin{array}{lcl} M_s\mathrm{D}_t\boldsymbol{u}_s & = & \displaystyle\int_{\partial\Omega_s} \underline{\boldsymbol{\sigma}}\cdot\boldsymbol{n}\mathrm{d}S, \\ J_s\mathrm{D}_t\omega & = & \displaystyle\int_{\partial\Omega_s} \underline{\boldsymbol{\sigma}}\cdot\boldsymbol{n}\cdot\begin{pmatrix}-y+y_s\\ x-x_s\end{pmatrix}\mathrm{d}S, \\ \mathrm{D}_t\boldsymbol{x} & = & \boldsymbol{u}_s+\omega\begin{pmatrix}-y+y_s\\ x-x_s\end{pmatrix}, \end{array}\right. \tag{V.8}$$

where σ is the stress tensor. Considering that the rigid body is immersed in an inviscid fluid, then $\underline{\boldsymbol{\sigma}} = -p\underline{\boldsymbol{I}}$. For a viscous one, it leads to $\underline{\boldsymbol{\sigma}} = -p\underline{\boldsymbol{I}} + \underline{\boldsymbol{\Upsilon}}$. For inviscid fluid, it writes

$$\left\{\begin{array}{lcl} M_s\mathrm{D}_t\boldsymbol{u}_s & = & \displaystyle-\int_{\partial\Omega_s} p\boldsymbol{n}\mathrm{d}S, \\ J_s\mathrm{D}_t\omega & = & \displaystyle-\int_{\partial\Omega_s} p\boldsymbol{n}\cdot\begin{pmatrix}-y+y_s\\ x-x_s\end{pmatrix}\mathrm{d}S, \\ \mathrm{D}_t\boldsymbol{x} & = & \boldsymbol{u}_s+\omega\begin{pmatrix}-y+y_s\\ x-x_s\end{pmatrix}. \end{array}\right. \tag{V.9}$$

In the following, the emphasis is laid on solving system (V.9).

V-2 High-order Lagrangian schemes for rigid body dynamics

First, system (V.9) is considered in one dimensional space. A semi-discrete scheme is proposed to approximate its solution. Two different discretizations are then proposed. The first one is based on a Runge–Kutta type integration in time, which is particularly adapted to schemes presented in chapter II. The second one, based on a Cauchy–Kovalevskaya integration in time, as the GoHy schemes used in chapter IV is then proposed. The extension to two space dimensions of these schemes is then proposed using a directionnal splitting method. First, the case of the rigid homogeneous cylinder is detailed, and then it is extended to any kind of geometry and mass repartition.

V-2.1 High-order schemes for rigid body dynamics in 1D

In one dimensional, we consider a rigid body occupying the domain $\Omega_s = [x_l, x_r]$. Then system (V.9) leads to the simplified 1D system

$$\left\{ \begin{array}{lcl} M_s \mathrm{D}_t u_s & = & -\left(p(x_r) - p(x_l)\right), \\ \mathrm{D}_t x & = & u_s, \end{array} \right. \tag{V.10}$$

where $p(x_r)$ and $p(x_l)$ are respectively the pressure applied at $x = x_r$ and at $x = x_l$. The semi-discrete scheme therefore writes

$$\left\{ \begin{array}{lcl} \mathrm{D}_t u_s & = & -\dfrac{p_r - p_l}{M_s}, \\ \mathrm{D}_t x_l & = & u_s, \\ \mathrm{D}_t x_r & = & u_s. \end{array} \right. \tag{V.11}$$

The pressure values p_r and p_l are respectively the pressure applied on the right and the left boundaries of rigid body. In practice, they are given using the Inverse Lax–Wendroff method proposed in chapter IV. Two approaches to realize the time integration of eq. (V.11) are proposed. The first one is based on a Runge–Kutta approach, the second one using a Cauchy–Kovalevskaya approach.

V-2.1.1 Runge–Kutta based approach

Using notations of chapter II for Runge–Kutta sequences, the fully discrete scheme writes

$$\left\{ \begin{array}{lcl} u_s^{n+\alpha_m} & = & u_s^n - \dfrac{\Delta t}{M_s} \displaystyle\sum_{l=0}^{m-1} a_{m,l} \left(p_r^{n+\alpha_l} - p_l^{n+\alpha_l}\right), \\ x_l^{n+\alpha_m} & = & x_l^n - \Delta t \displaystyle\sum_{l=0}^{m-1} a_{m,l} u_s^{n+\alpha_m}, \\ x_r^{n+\alpha_m} & = & x_r^n - \Delta t \displaystyle\sum_{l=0}^{m-1} a_{m,l} u_s^{n+\alpha_m}, \end{array} \right. \tag{V.12}$$

where the pressure $p_r^{n+\alpha_m}$ and $p_l^{n+\alpha_m}$ are given in practice by the Inverse Lax–Wendroff procedure using the values inside the fluid domain and the velocity at the boundary.

$$\left\{\begin{array}{rcl} u_s^{n+1} & = & u_s^n - \dfrac{\Delta t}{M_s}\displaystyle\sum_{l=0}^{s-1}\theta_l\left(p_r^{n+\alpha_l} - p_l^{n+\alpha_l}\right), \\ x_l^{n+1} & = & x_l^n - \Delta t\displaystyle\sum_{l=0}^{s-1}\theta_l u_s^{n+\alpha_m}, \\ x_r^{n+1} & = & x_r^n - \Delta t\displaystyle\sum_{l=0}^{s-1}\theta_l u_s^{n+\alpha_m}. \end{array}\right. \tag{V.13}$$

V-2.1.2 Cauchy–Kovalevskaya based approach

The Cauchy–Kovalevskaya based approach is identical to the one used in [50, 170]. It relies on using the information provided by the EOS and also by the fluid system of equations. In particular, one uses that

$$\rho_0 \mathrm{D}_t p + (\rho c)^2 \partial_x u = 0, \tag{V.14}$$

where c is the speed of sound. It yields without expliciting the time derivatives that

$$\left\{\begin{array}{rcl} u_s^{n+1} & = & u_s^n - \dfrac{\Delta t}{M_s}\displaystyle\sum_{k\geq 0}\left(\mathrm{D}_t^k p_r^n - \mathrm{D}_t^k p_l^n\right)\frac{\Delta t^k}{k!}, \\ x_l^{n+1} & = & x_l^n - \Delta t\displaystyle\sum_{k\geq 0}\mathrm{D}_t^k u_s^n \frac{\Delta t^k}{k!}, \\ x_r^{n+1} & = & x_r^n - \Delta t\displaystyle\sum_{k\geq 0}\mathrm{D}_t^k u_s^n \frac{\Delta t^k}{k!}. \end{array}\right. \tag{V.15}$$

V-2.2 High-order schemes for rigid body dynamics in 2D

In order to study rigid body dynamics in 2D, a choice of space discretization must first be made. Indeed, contrarily to the 1D case, the rigid body is no longer described by only two points. We consider a rigid body which is described by a closed bounded domain $\Omega_s \subset \Omega \subset \mathbb{R}^2$. We denote by $\Gamma = \partial\Omega_s$. As the external forces are exerted on the boundary Γ, it is all but natural to lay the emphasis on the discretization of Γ, then to devise a semi-discrete scheme and last to consider the fully discrete scheme for rigid body motion.

V-2.2.1 Rigid body space discretization

The choice has been made to consider a discretization of Γ instead of Ω_s since the forces exerted on the rigid body are exerted on the boundary Γ. Γ is parametrized by a function $\boldsymbol{\gamma} : [0:1] \longrightarrow \mathbb{R}^2$.

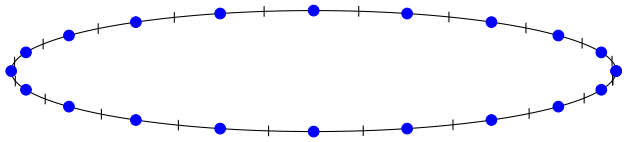


Figure V.1 – Regular curvilinear discretization of an ellipse with $\Gamma : s \to (5\cos(2\pi s), \sin(2\pi s))^t$ using 20 pearls (blue dots)

In the following the curvilinear abscissa is denoted s. It writes

$$\Gamma = \{\boldsymbol{x}, \exists s \in [0,1], \boldsymbol{\gamma}(s) = \boldsymbol{x}\}.$$

We consider a discretization with N elements $\Gamma_{i+\frac{1}{2}}$ such that

$$\begin{cases} s_0 & = & 0, & \\ s_N & = & 1, & \\ s_{i+1} - s_i & = & \Delta s, & \forall i \in \{0, ..., N-1\}, \\ \Gamma_{i+\frac{1}{2}} & = & \{\boldsymbol{x}, \exists s \in [s_i, s_{i+1}], \boldsymbol{\gamma}(s) = \boldsymbol{x}\} & \forall i \in \{0, ..., N-1\}. \end{cases} \tag{V.16}$$

Denote in particular that the two points of abscissa s_0 and s_N are identical. One trivially gets that

$$\bigcup_{i=0}^{N-1} \Gamma_{i+\frac{1}{2}} = \Gamma$$

We define also the staggered curvilinear abscissae as

$$s_{i+\frac{1}{2}} = \frac{s_{i+1} - s_i}{2}, \quad \forall i \in \{0, ..., N-1\}.$$

The pearls $\boldsymbol{P}_{i+\frac{1}{2}}$ are located as

$$\boldsymbol{P}_{i+\frac{1}{2}} = \boldsymbol{\gamma}(s_{i+\frac{1}{2}})$$

This discretization, which is depicted for an ellipse in fig. V.1, is particularly appropriate to compute integrals of the form

$$\begin{aligned} \int_\Gamma \phi(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} &= \sum_{i=0}^{N-1} \int_{\Gamma_{i+\frac{1}{2}}} \phi(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ &= \sum_{i=0}^{N-1} \int_{s_i}^{s_{i+1}} \phi(\boldsymbol{\gamma}(s)) \|\boldsymbol{\gamma}\prime(s)\| ds \\ &= \Delta s \sum_{i=0}^{N-1} \frac{1}{\Delta s} \int_{s_i}^{s_{i+1}} \phi(\boldsymbol{\gamma}(s)) \|\boldsymbol{\gamma}\prime(s)\| ds \end{aligned} \tag{V.17}$$

The following lemma gives an accuracy result on the spatial discretization concerning the computation of such an integral. It is a corollary of a result proved by Kurganov and Rauch in [**kurganov2009order**] about spectral accuracy of low order quadrature formulae for periodic function. It is proved here for smooth functions on a closed curve using the interpolation coefficients $\widehat{C}_k$ which are central in this work.

Lemma V.1. *Assume that Γ is a closed curve. Let $\boldsymbol{\gamma}$ and ϕ smooth enough and $m > 0$. Assume the following approximation $\phi^{\gamma}_{i+\frac{1}{2}} = \phi(\boldsymbol{\gamma}(s_{i+\frac{1}{2}}))\|\boldsymbol{\gamma}\prime(s_{i+\frac{1}{2}})\|$, then*

$$\int_{\Gamma} \phi(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \Delta s \sum_{i=0}^{N-1} \phi^{\gamma}_{i+\frac{1}{2}} + \mathscr{O}(\Delta s^m).$$

Remark V.1. Lemma V.1 implies in particular that trapezoidal rule yields immediately spectral accuracy for the integral computation on a closed curve.

Proof. Denoting $\overline{\phi}^{\gamma}_{i+\frac{1}{2}} = \dfrac{1}{\Delta s}\displaystyle\int_{s_i}^{s_{i+1}} \phi(\boldsymbol{\gamma}(s))\|\boldsymbol{\gamma}\prime(s)\|ds$, one has in particular from chapter II, for $r > 0$ that

$$\overline{\phi}^{\gamma}_{i+\frac{1}{2}} = \sum_{k=-r}^{r} \widehat{C}_k \phi^{\gamma}_{i+k+\frac{1}{2}} + \mathscr{O}(\Delta s^{2r+1}) \tag{V.18}$$

where the coefficients $\widehat{C}_k$ are available in table II.2, and $\phi^{\gamma}_{i+\frac{1}{2}}$ is defined with periodic boundary conditions as

$$\phi^{\gamma}_{i+\frac{1}{2}} = \begin{cases} \phi(\boldsymbol{\gamma}(s_{i+\frac{1}{2}}))\|\boldsymbol{\gamma}\prime(s_{i+\frac{1}{2}})\|), & i \in \{0, ..., N-1\}, \\ \phi(\boldsymbol{\gamma}(s_{i+N+\frac{1}{2}}))\|\boldsymbol{\gamma}\prime(s_{i+N+\frac{1}{2}})\|, & i \leq -1, \\ \phi(\boldsymbol{\gamma}(s_{i-N+\frac{1}{2}}))\|\boldsymbol{\gamma}\prime(s_{i-N+\frac{1}{2}})\|, & i \geq N. \end{cases} \tag{V.19}$$

Then, for a given $r > 0$ one has

$$\begin{aligned} \int_{\Gamma} \phi(\boldsymbol{x})\mathrm{d}\boldsymbol{x} &= \Delta s \sum_{i=0}^{N-1} \overline{\phi}^{\gamma}_{i+\frac{1}{2}} \\ &= \Delta s \sum_{i=0}^{N-1} \left(\sum_{k=-r}^{r} \widehat{C}_k \phi^{\gamma}_{i+k+\frac{1}{2}} + \mathscr{O}(\Delta s^{2r+1}) \right) \\ &= \Delta s \left(\sum_{k=-r}^{r} \widehat{C}_k \right) \sum_{i=0}^{N-1} \left(\phi^{\gamma}_{i+\frac{1}{2}} + \mathscr{O}(\Delta s^{2r+1}) \right) \end{aligned} \tag{V.20}$$

Taking r such that $2r + 1 > m$, using that $\displaystyle\sum_{k=-r}^{r} \widehat{C}_k = 1$ and definition of $\phi^{\gamma}_{i+\frac{1}{2}}$ in eq. (V.19), it yields

$$\int_{\Gamma} \phi(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \Delta s \sum_{i=0}^{N-1} \left(\phi^{\gamma}_{i+\frac{1}{2}} + \mathscr{O}(\Delta s^m) \right)$$

Using that N is inversely proportional to Δs, it leads to

$$= \Delta s \sum_{i=0}^{N-1} \phi^{\gamma}_{i+k+\frac{1}{2}} + \mathscr{O}(\Delta s^m)$$

Hence the result. ■

Lemma V.1 is useful to compute the line integral of torques and forces exerted on the rigid body boundary Γ. As for rigid body dynamics, we first explain how we design the semi-discrete scheme for irrotational rigid bodies and then extend the semi-discrete scheme to the general case of rigid body dynamics including both translation and rotation.

V-2.2.2 Irrotational rigid body semi-discrete scheme

Consider the system of equations (V.9) with $J_s \to \infty$. It yields an irrotational field of velocity inside the rigid body with, thus, $\omega = 0$. The system writes

$$\left\{ \begin{array}{lcl} M_s \mathrm{D}_t \boldsymbol{u}_s & = & -\displaystyle\int_{\partial\Omega_s} p\boldsymbol{n}\mathrm{d}S, \\ \mathrm{D}_t \boldsymbol{x} & = & \boldsymbol{u}_s. \end{array} \right. \tag{V.21}$$

The only possible motion for the rigid body is therefore a translation. As the interior fluid schemes in 2D are based on directional splitting, the choice has been made to apply the same strategy to eq. (V.21). Denoting $\boldsymbol{u}_s = (u_s, v_s)^t$, $\boldsymbol{n} = (n_1, n_2)^t$, $\boldsymbol{x} = (x, y)^t$, it leads to

$$\left\{ \begin{array}{lcl} M_s \mathrm{D}_t^x u_s & = & -\displaystyle\int_{\Gamma} p n_1 \mathrm{d}S, \\ M_s \mathrm{D}_t^x v_s & = & 0, \\ \mathrm{D}_t^x x & = & u_s, \\ \mathrm{D}_t^x y & = & 0. \end{array} \right. \qquad \left\{ \begin{array}{lcl} M_s \mathrm{D}_t^y u_s & = & 0, \\ M_s \mathrm{D}_t^y v_s & = & -\displaystyle\int_{\Gamma} p n_2 \mathrm{d}S, \\ \mathrm{D}_t^y x & = & 0, \\ \mathrm{D}_t^y y & = & v_s. \end{array} \right. \tag{V.22}$$

Considering the first system (in the x-direction) of eq. (V.22), its semi-discrete form using lemma V.1 writes

$$\left\{ \begin{array}{lcl} M_s \mathrm{D}_t^x u_s & = & -\Delta s \displaystyle\sum_{i=0}^{N-1} (p n_1)^{\gamma}_{i+\frac{1}{2}}, \\ M_s \mathrm{D}_t^x v_s & = & 0, \\ \mathrm{D}_t^x x_{i+\frac{1}{2}} & = & u_s, \\ \mathrm{D}_t^x y_{i+\frac{1}{2}} & = & 0, \end{array} \right. \tag{V.23}$$

where $\boldsymbol{P}_{i+\frac{1}{2}} = (x_{i+\frac{1}{2}}, y_{i+\frac{1}{2}})^t$. Considering that Γ is known analytically and that the pressure at point $\boldsymbol{P}_{i+\frac{1}{2}}$ can be computed with a mth order of accuracy, then the semi-discrete form is also of order m in space.

V-2.2.3 General rigid body semi-discrete scheme

Consider the system of equations (V.9) without any assumption on M_s or J_s. We introduce the notations $\boldsymbol{T}$ and $\boldsymbol{N}$ for the non-normalized tangent and normal. Meaning in particular that one has

$$\boldsymbol{N} = \boldsymbol{n}\|\gamma\|.$$

The equation on $\boldsymbol{N}$ is immediately obtained using the laws of rigid body motion. Indeed,

$$\boldsymbol{T} = \begin{pmatrix} \partial_s x \\ \partial_s y \end{pmatrix}, \quad \boldsymbol{N} = \pm \begin{pmatrix} \partial_s y \\ \partial_s x \end{pmatrix},$$

and thus one gets that

$$\mathrm{D}_t \boldsymbol{T} = \begin{pmatrix} \mathrm{D}_t \partial_s x \\ \mathrm{D}_t \partial_s y \end{pmatrix} = \partial_s \left(\boldsymbol{u}_s + \omega \begin{pmatrix} -y + y_s \\ x - x_s \end{pmatrix} \right) = \begin{pmatrix} 0 & -\omega \\ \omega & 0 \end{pmatrix} \boldsymbol{T}, \tag{V.24}$$

and similarly the non-normalized vector satisfies

$$\mathrm{D}_t \boldsymbol{N} = \begin{pmatrix} 0 & -\omega \\ \omega & 0 \end{pmatrix} \boldsymbol{N}, \tag{V.25}$$

Using directional splitting method, the semi-discrete scheme for (V.9) and (V.25) writes

$$\left\{ \begin{array}{lcl} M_s \mathrm{D}_t^x u_s & = & -\displaystyle\int_\Gamma p n_1 \mathrm{d}S, \\ M_s \mathrm{D}_t^x v_s & = & 0, \\ J_s \mathrm{D}_t^x \omega & = & \displaystyle\int_\Gamma p n_1 (y - y_g) \mathrm{d}S, \\ \mathrm{D}_t^x x & = & u_s - \omega(y - y_g), \\ \mathrm{D}_t^x y & = & 0, \\ \mathrm{D}_t^x N_1 & = & 0, \\ \mathrm{D}_t^x N_2 & = & \omega N_1. \end{array} \right. \qquad \left\{ \begin{array}{lcl} M_s \mathrm{D}_t^y u_s & = & 0, \\ M_s \mathrm{D}_t^y v_s & = & -\displaystyle\int_\Gamma p n_2 \mathrm{d}S, \\ J_s \mathrm{D}_t^y \omega & = & -\displaystyle\int_\Gamma p n_2 (x - x_g) \mathrm{d}S, \\ \mathrm{D}_t^y x & = & 0, \\ \mathrm{D}_t^y y & = & v_s + \omega(x - x_g), \\ \mathrm{D}_t^y N_1 & = & -\omega N_2, \\ \mathrm{D}_t^y N_2 & = & 0. \end{array} \right. \tag{V.26}$$

Considering first system (in the x-direction) of eq. (V.26), its semi-discrete form using lemma V.1 writes

$$\begin{cases} M_s \mathrm{D}_t^x u_s &= -\Delta s \sum_{i=0}^{N-1} (pn_1)^\gamma_{i+\frac{1}{2}}, \\ M_s \mathrm{D}_t^x v_s &= 0, \\ J_s \mathrm{D}_t^x \omega &= -\Delta s \sum_{i=0}^{N-1} (pn_1(y-y_g))^\gamma_{i+\frac{1}{2}}, \\ \mathrm{D}_t^x x_{i+\frac{1}{2}} &= u_s - \omega(y-y_g), \\ \mathrm{D}_t^x y_{i+\frac{1}{2}} &= 0, \\ \mathrm{D}_t^x N_{1,i+\frac{1}{2}} &= 0, \\ \mathrm{D}_t^x N_{2,i+\frac{1}{2}} &= \omega N_{1,i+\frac{1}{2}}. \end{cases} \tag{V.27}$$

The main differences with the case of irrotational motion is obviously that the rigid body is rotating due to the torques exerted at the boundary, which implies also that the normals are rotating as well. Hence the equation on both N_1 and N_2. In practice, one rewrites eq. (V.27) substituting the term ϕ^γ with respectively terms of the form (ϕN_1) for the $x-$direction and of the form (ϕN_2) for the y-direction.

$$\begin{cases} M_s \mathrm{D}_t^x u_s &= -\Delta s \sum_{i=0}^{N-1} (pN_1)_{i+\frac{1}{2}}, \\ M_s \mathrm{D}_t^x v_s &= 0, \\ J_s \mathrm{D}_t^x \omega &= -\Delta s \sum_{i=0}^{N-1} (pN_1(y-y_g))_{i+\frac{1}{2}}, \\ \mathrm{D}_t^x x_{i+\frac{1}{2}} &= u_s - \omega(y_{i+\frac{1}{2}}-y_g), \\ \mathrm{D}_t^x y_{i+\frac{1}{2}} &= 0, \\ \mathrm{D}_t^x N_{1,i+\frac{1}{2}} &= 0, \\ \mathrm{D}_t^x N_{2,i+\frac{1}{2}} &= \omega N_{1,i+\frac{1}{2}}. \end{cases} \tag{V.28}$$

Two integrations in time are know proposed. The first one is based on Runge–Kutta time integration and the second one on a Cauchy–Kovalevskaya one using repetitively time-derivatives of system (V.28) as well as information provided by the fluid part.

V-2.2.4 Runge–Kutta based approach

We use the notation introduced in chapter II. The fully discrete scheme in the x-direction writes for the intermediary time-step

$$\left\{\begin{array}{lcl}
u_s^{n+\alpha_m} & = & u_s^n - \dfrac{\Delta t}{M_s}\Delta s \displaystyle\sum_{l=0}^{m-1} a_{m,l} \sum_{i=0}^{N-1} (p^{n+\alpha_l} N_1^n)_{i+\frac{1}{2}}, \\
v_s^{n+\alpha_m} & = & v_s^n, \\
\omega^{n+\alpha_m} & = & \omega^n - \dfrac{\Delta t}{J_s}\Delta s \displaystyle\sum_{l=0}^{m-1} a_{m,l} \sum_{i=0}^{N-1} (p^{n+\alpha_l} N_1^n (y^n - y_g^n))_{i+\frac{1}{2}}, \\
x_{i+\frac{1}{2}}^{n+\alpha_m} & = & x_{i+\frac{1}{2}}^n + \Delta t \left(\displaystyle\sum_{l=0}^{m-1} a_{m,l} (u_s^{n+\alpha_l} - \omega^{n+\alpha_l} (y_{i+\frac{1}{2}}^n - y_g^n)) \right), \\
y_{i+\frac{1}{2}}^{n+\alpha_m} & = & y_{i+\frac{1}{2}}^n, \\
N_{1,i+\frac{1}{2}}^{n+\alpha_m} & = & N_{1,i+\frac{1}{2}}^n, \\
N_{2,i+\frac{1}{2}}^{n+\alpha_m} & = & N_{2,i+\frac{1}{2}}^n + \Delta t \left(\displaystyle\sum_{l=0}^{m-1} a_{m,l} \omega^{n+\alpha_l} N_{1,i+\frac{1}{2}}^n \right),
\end{array}\right. \tag{V.29}$$

and for the final time-step as

$$\left\{\begin{array}{lcl}
u_s^{n+1} & = & u_s^n - \dfrac{\Delta t}{M_s}\Delta s \displaystyle\sum_{l=0}^{s-1} \theta_l \sum_{i=0}^{N-1} (p^{n+\alpha_l} N_1^n)_{i+\frac{1}{2}}, \\
v_s^{n+1} & = & v_s^n, \\
\omega^{n+1} & = & \omega^n - \dfrac{\Delta t}{J_s}\Delta s \displaystyle\sum_{l=0}^{s-1} \theta_l \sum_{i=0}^{N-1} (p^{n+\alpha_l} N_1^n (y^n - y_g^n))_{i+\frac{1}{2}}, \\
x_{i+\frac{1}{2}}^{n+1} & = & x_{i+\frac{1}{2}}^n + \Delta t \left(\displaystyle\sum_{l=0}^{s-1} \theta_l (u_s^{n+\alpha_l} - \omega^{n+\alpha_l} (y_{i+\frac{1}{2}}^n - y_g^n)) \right), \\
y_{i+\frac{1}{2}}^{n+1} & = & y_{i+\frac{1}{2}}^n, \\
N_{1,i+\frac{1}{2}}^{n+1} & = & N_{1,i+\frac{1}{2}}^n, \\
N_{2,i+\frac{1}{2}}^{n+1} & = & N_{2,i+\frac{1}{2}}^n + \Delta t \left(\displaystyle\sum_{l=0}^{s-1} \theta_l \omega^{n+\alpha_l} N_{1,i+\frac{1}{2}}^n \right),
\end{array}\right. \tag{V.30}$$

V-2.2.5 Cauchy–Kovalevskaya based approach

The Cauchy–Kovalevskaya based approach is identical to the one used in [50, 170]. It relies on informations provided by the EOS, by the fluid system of equations but also by the rigid body system of equations. Concerning fluid and EOS, the equation

$$\rho_0 \mathrm{D}_t p + (\rho c)^2 \partial_x u = 0, \tag{V.31}$$

is derivated in time repetitively to transform time derivatives of p into space derivatives. Moreover, one uses that

$$\left\{\begin{array}{lcl}
\mathrm{D}_t^x y_{i+\frac{1}{2}} & = & 0, \\
\mathrm{D}_t^x N_{1,i+\frac{1}{2}} & = & 0.
\end{array}\right. \tag{V.32}$$

Then starting from eq. (V.28) and integrating in time yield

$$\begin{cases} u_s^{n+1} &= u_s^n - \frac{1}{M_s}\Delta s \int_{t^n}^{t^{n+1}} \sum_{i=0}^{N-1} (pN_1)_{i+\frac{1}{2}}(\theta)\mathrm{d}\theta, \\ v_s^{n+1} &= v_s^n, \\ \omega^{n+1} &= \omega^n - \frac{1}{J_s}\Delta s \int_{t^n}^{t^{n+1}} \sum_{i=0}^{N-1} (pN_1(y-y_g))_{i+\frac{1}{2}}(\theta)\mathrm{d}\theta, \\ x_{i+\frac{1}{2}}^{n+1} &= x_{i+\frac{1}{2}}^n + \int_{t^n}^{t^{n+1}} \left(u_s - \omega(y_{i+\frac{1}{2}} - y_g)\right)(\theta)\mathrm{d}\theta, \\ y_{i+\frac{1}{2}}^{n+1} &= y_{i+\frac{1}{2}}^n, \\ N_{1,i+\frac{1}{2}}^{n+1} &= N_{1,i+\frac{1}{2}}^n, \\ N_{2,i+\frac{1}{2}}^{n+1} &= N_{2,i+\frac{1}{2}}^n + \int_{t^n}^{t^{n+1}} \omega N_{1,i+\frac{1}{2}}(\theta)\mathrm{d}\theta. \end{cases} \tag{V.33}$$

Performing Taylor expansion in the θ variable and using eq. (V.32) lead to

$$\begin{cases} u_s^{n+1} &= u_s^n - \frac{\Delta t}{M_s}\Delta s \sum_{i=0}^{N-1} \left(\sum_k \mathrm{D}_t^{x,k} p_{i+\frac{1}{2}}^n \frac{\Delta t^k}{(k+1)!}\right) N_{1,i+\frac{1}{2}}^n, \\ v_s^{n+1} &= v_s^n, \\ \omega^{n+1} &= \omega^n - \frac{\Delta t}{J_s}\Delta s \sum_{i=0}^{N-1} \left(\sum_k \mathrm{D}_t^{x,k} p_{i+\frac{1}{2}}^n \frac{\Delta t^k}{(k+1)!}\right) N_{1,i+\frac{1}{2}}^n (y_{i+\frac{1}{2}}^n - y_g^n)), \\ x_{i+\frac{1}{2}}^{n+1} &= x_{i+\frac{1}{2}}^n + \Delta t \sum_k \left(\mathrm{D}_t^{x,k} u_s^n - \mathrm{D}_t^{x,k}\omega^n (y_{i+\frac{1}{2}}^n - y_g^n)\right) \frac{\Delta t^k}{(k+1)!}, \\ y_{i+\frac{1}{2}}^{n+1} &= y_{i+\frac{1}{2}}^n, \\ N_{1,i+\frac{1}{2}}^{n+1} &= N_{1,i+\frac{1}{2}}^n, \\ N_{2,i+\frac{1}{2}}^{n+1} &= N_{2,i+\frac{1}{2}}^n + \Delta t \sum_k \left(\mathrm{D}_t^{x,k}\omega^n \frac{\Delta t^k}{(k+1)!}\right) N_{1,i+\frac{1}{2}}^n. \end{cases} \tag{V.34}$$

V-3 Fluid - Rigid body coupling

After detailing the two proposed numerical schemes for the integration of forces and torques exerted on the boundary, we propose a simple and straightforward scheme to couple the fluid and the rigid body solvers.

V-3.1 Description of the algorithm

Since the inverse Lax–Wendroff procedure has been developed in a Lagrange-remap formalism and since the rigid body motion is described in a Lagrangian formalism, there is no further work to be done. The fluid-rigid body coupling is depicted in Figure V.2. It follows a simple flow chart, where the space and time coupling is realized using our Inverse–Lax Wendroff boundary treatment. At time $t = t^n$, one knows the value of $\boldsymbol{\mathcal{U}}^+$ which are the values inside the fluid domain and also the rigid body state among which is the normal velocity. Using the normal velocity known at the boundary, one applies the ILW procedure, and deduces values inside $\boldsymbol{\mathcal{U}}^-$

as well as the integral of forces and torques exerted on the rigid body boundary.

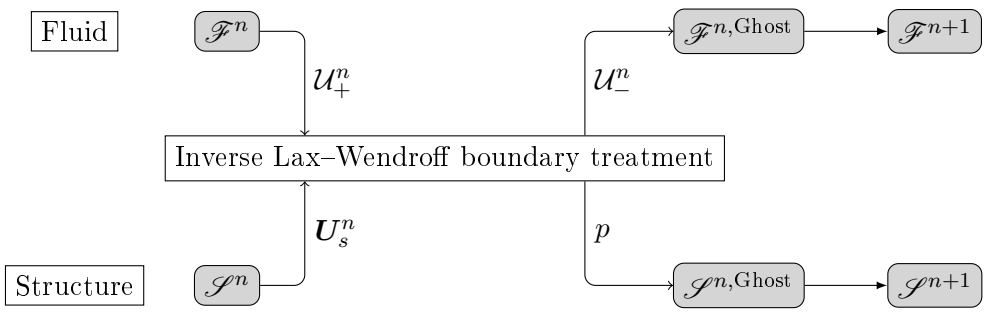


Figure V.2 – Using the Inverse Lax–Wendroff procedure as a time and space coupling for rigid body interaction.

The rigid body motion solver also adds a constraint on the time step Δt. In addition to the classical CFL condition, in practice the time-step is asked to satisfy the constraint for one dimensional problem

$$\Delta t < \frac{\Delta x}{|u_s|},$$

and for two dimension problems

$$\Delta t < \frac{1}{\max_k \omega_k} \min\left(\frac{\Delta X}{\max_i |u_{i+\frac{1}{2}}|}), \frac{\Delta Y}{\max_i |v_{i+\frac{1}{2}}|}) \right),$$

where $(u_{i+\frac{1}{2}}, v_{i+\frac{1}{2}})$ is the velocity of the pearl $i + \frac{1}{2}$.

V-3.2 Numerical results

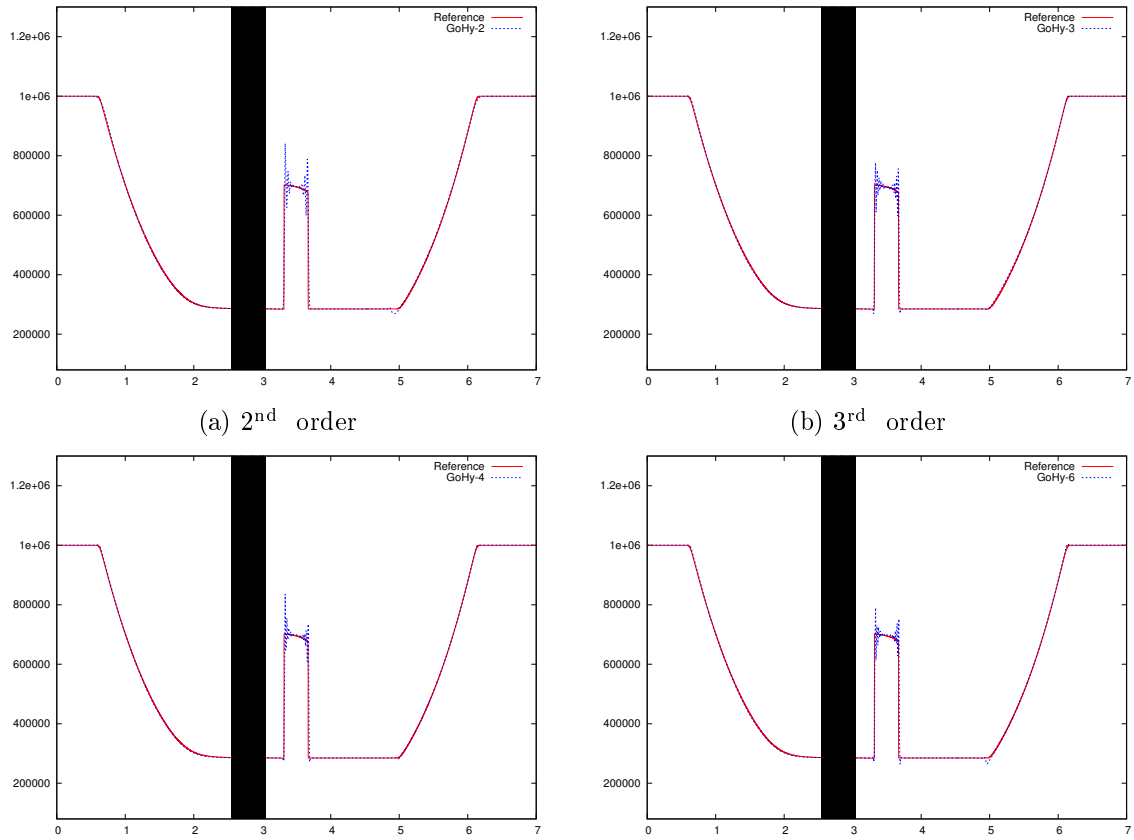
A test-suite is proposed to assess both accuracy and robustness of the fluid-rigid body schemes. We begin with a 1D case problem consisting of a piston whose motion is triggered by a pressure differential [120]. Then, the ability of the 2D schemes to handle strong shocks is assessed. The first test concerns the lift-off of a cylinder proposed in [53, 6, 88, 120]. The problem is then extended to more complex geometries with first an ellipse and then a rhombus.

V-3.2.1 Pressure motion driven piston in 1D [120]

This test-case has been proposed in [120] to study the coupling between fluid and rigid body in 1D. The computational domain is $[0 : 7]$. Initially a rigid body of length 0.5m and of mass 1.0kg is centered at $x = 2$m. The fluid initial state is

$$\begin{cases} \rho_0(x) = 10\chi_{\{x<2,x>5\}} + 1\chi_{\{2<x<5\}}, \\ u_0(x) = 0, \\ p_0(x) = 10^6\chi_{\{x<2,x>5\}} + 10^5\chi_{\{2<x<5\}}, \\ \quad \gamma \quad = 1.4. \end{cases} \tag{V.35}$$

The movement of the rigid body (in black in the figure) is triggered by the pressure differential between the left and right sides of the piston. In return, it induces propagation waves in the fluid regions. Fluid states as well as the piston position are depicted in fig. V.3 at time $t = 3$ ms.



(a) 2nd order (b) 3rd order

(c) 4th order (d) 6th order

Figure V.3 – Pressure profiles at time t=3 ms with 800 cells for the pressure motion driven piston in 1D for second, third, fourth and sixth order ILW-GoHy schemes.

V-3.2.2 Lift-Off of a cylinder [6, 88, 120]

The lift-off of a cylinder has been proposed in [53] to study the coupling between a fluid solver and a rigid body motion one. It is a challenging problem coupling both a fluid and a moving rigid body. The computational domain is $[0.0 : 1.0] \times [0.0 : 0.2]$. A disk of radius 5 cm and of density $\rho = 7.6$ kg.m^{-2} lies at the bottom of a channel. Initially the center of the disk is at point $(15.10^{-2}, 5.10^{-2})$. A Mach 3 shock enters the domain, and due to the asymmetry of the problem lifts the disk. Equivalent initial datas are presented in [6, 88, 120]:

$$\begin{cases} p_0 = & 1.0\chi_{\{x>0.08\}} + \frac{31}{3}\chi_{\{x<0.08\}}, \\ u_0 = & 2.6293688\chi_{\{x<0.08\}}, \\ v_0 = & 0, \\ \rho_0 = & 1.0\chi_{\{x>0.08\}} + 3.8571429\chi_{\{x<0.08\}}, \\ \gamma = & 1.4. \end{cases} \tag{V.36}$$

Figure V.8 shows the pressure contours at $t = 0.14$ and $t = 0.255$ for a grid size $\Delta x = \Delta y = 6.25 \times 10^{-4}$ using the third order scheme GoHy-3. Figure V.9 shows density contours at $t = 0.255$ for the same grid size and the same scheme. A MOOD method is used on the boundary. General profiles are in accordance with results found in the literature. We also compare in table V.2 the final position of the cylinder of [88] and the final position obtained for the reflection method presented in [6] for different grid sizes and order. Final positions are in good agreements with those found in the literature, especially with [6]. As presented in [120], the presence of strong vortices are denoted under the cylinder which does not disappear as the mesh is refined. We assume that a highly dissipative scheme prevents such vortices from appearing. Here, high-order accuracy and reduced dissipation allow such mechanisms to appear and develop.

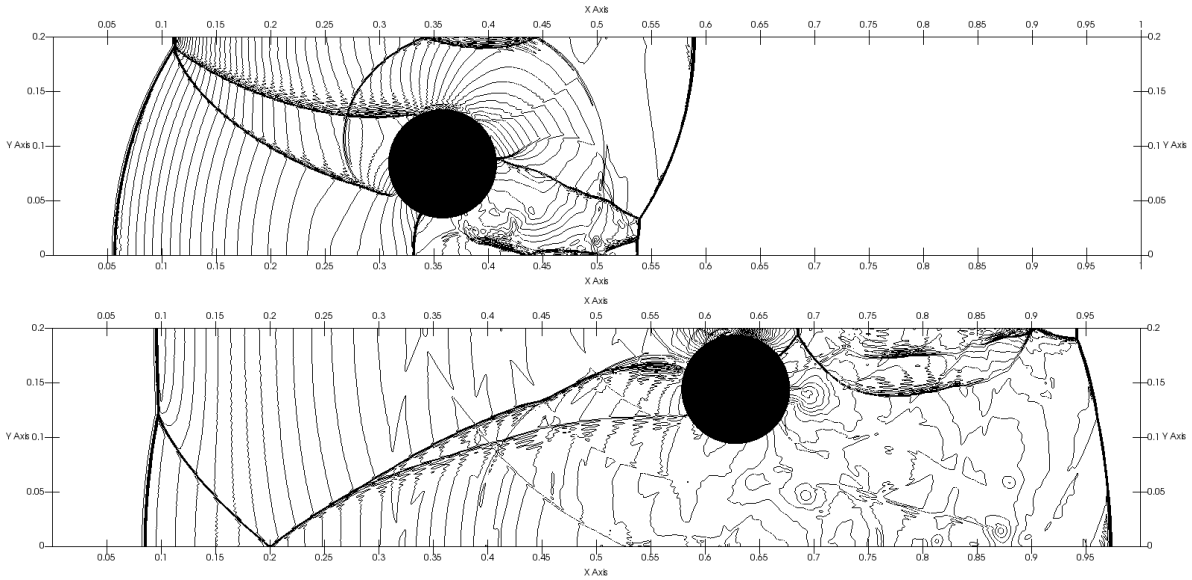


Figure V.4 – 60 contours of fluid pressure from 0 to 28 at times t=0.14 (top) and t=0.255 (bottom) for the third order scheme, $\Delta x = \Delta y = 6.25 \times 10^{-4}$.

$\Delta x = \Delta y$	Hu and al. [88]	Arienti and al. [6]	GoHy-1	GoHy-2	GoHy-3
2.5×10^{-3}	$(0.659, 0.132)$	$(0.624, 0.143)$	$(0.623, 0.126)$	$(0.628, 0.136)$	$(0.627, 0.136)$
1.25×10^{-3}	$(0.649, 0.145)$	$(0.626, 0.145)$	$(0.621, 0.131)$	$(0.626, 0.141)$	$(0.625, 0.140)$
6.25×10^{-4}	$(0.641, 0.147)$	$(0.627, 0.145)^\star$	$(0.623, 0.136)$	$(0.628, 0.144)$	$(0.628, 0.144)$

Table V.2 – Comparisons of the position of the cylinder's center at $t = 0.255$. $\star$ denotes results for $\Delta x = \Delta y = 10^{-3}$.

Integration of forces and torques exerted on the cylinder depends on the number of points used to discretize the cylinder. Here, it is noticed that if one takes greater value of C_Γ, the position is changed only at the fourth digit. We present in Table V.3, absolute errors made on conservation of mass and total energy which seem to converge with a slope of $0.7 - 0.8$ for the first order scheme, and near unity for the second and third order ones.

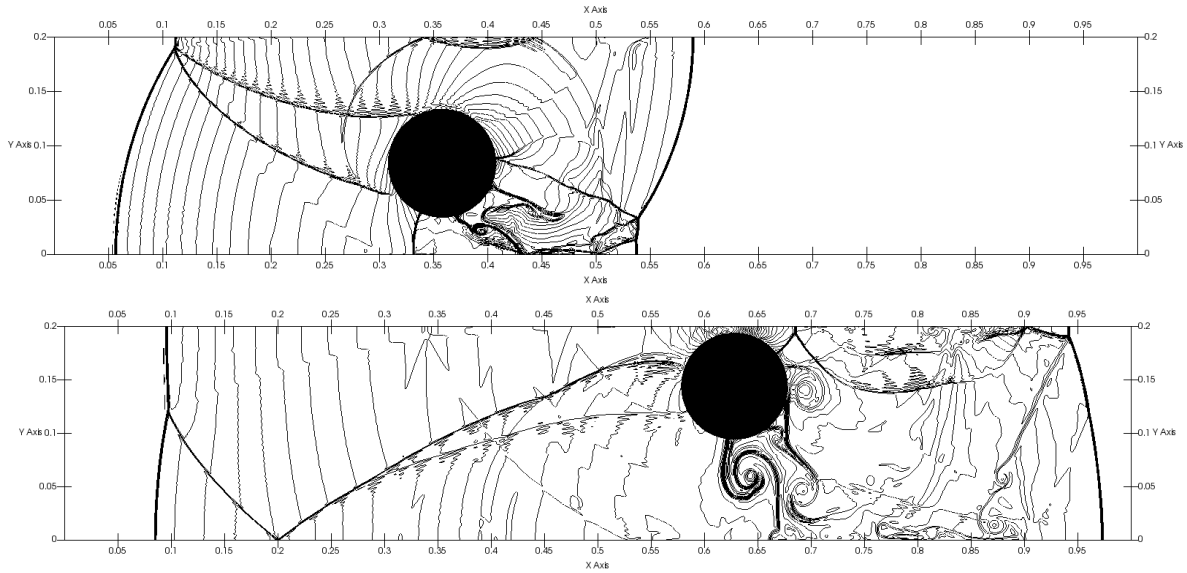


Figure V.5 – 60 contours of fluid density from 0 to 12 at times t=0.14 (top) and t=0.255 (bottom) for the third order scheme, $\Delta x = \Delta y = 6.25 \times 10^{-4}$.

$\Delta x = \Delta y$	GoHy-1		GoHy-2		GoHy-3	
	$\lvert\Delta m\rvert$	$\lvert\Delta e\rvert$	$\lvert\Delta m\rvert$	$\lvert\Delta e\rvert$	$\lvert\Delta m\rvert$	$\lvert\Delta e\rvert$
2.5×10^{-3}	1.55e-2	4.24e-2	8.07e-3	1.71e-2	1.1e-2	2.5e-2
1.25×10^{-3}	9.41e-3	2.62e-2	4.12e-3	8.89e-3	5.58e-3	1.29e-2
6.25×10^{-4}	5.36e-3	1.54e-2	2.16e-3	4.58e-3	2.81e-3	6.47e-3

Table V.3 – Conservation on mass and total energy at $t = 0.255$ for the lift-off cylinder test-case.

V-3.2.3 Lift-Off of an ellipse

This test-case is very similar to the previous one. The initial data are unchanged. However the form of the rigid body is changed. Indeed, for the cylinder test-case and in absence of any viscous forces, the rigid body motion is irrotational. In this test-case, we consider an ellipse lying at the bottom of the channel. The ellipse is defined by a semi-major axe in the x-direction of length 7 cm and a semi-minor axe of length 4 cm. Its density is set to $\rho = 9.0$ kg.m^{-2}. Initially, the ellipse lies at the bottom of a channel, and its center is at point $(17.10^{-2}, 4.10^{-2})$. A Mach 3 shock enters the domain, and due to the asymmetry of the problem lifts the ellipse.

V-3.2.4 Lift-Off of a rhombus

This test-case is very similar to the previous ones. The initial data are unchanged. However the form of the rigid body is changed. In this test-case, we consider a rhombus which as undefined normals at each of its angles. The rhombus is defined by the following equation

$$\left\{ \begin{array}{ll} |\frac{x^\dagger}{b}| + |\frac{y^\dagger}{h}| = & 1 \\ \begin{pmatrix} x^\dagger \\ y^\dagger \end{pmatrix} = & \begin{pmatrix} \cos(\theta & \sin(\theta) \\ -\sin(\theta & \cos(\theta) \end{pmatrix} \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} \end{array} \right. \tag{V.37}$$

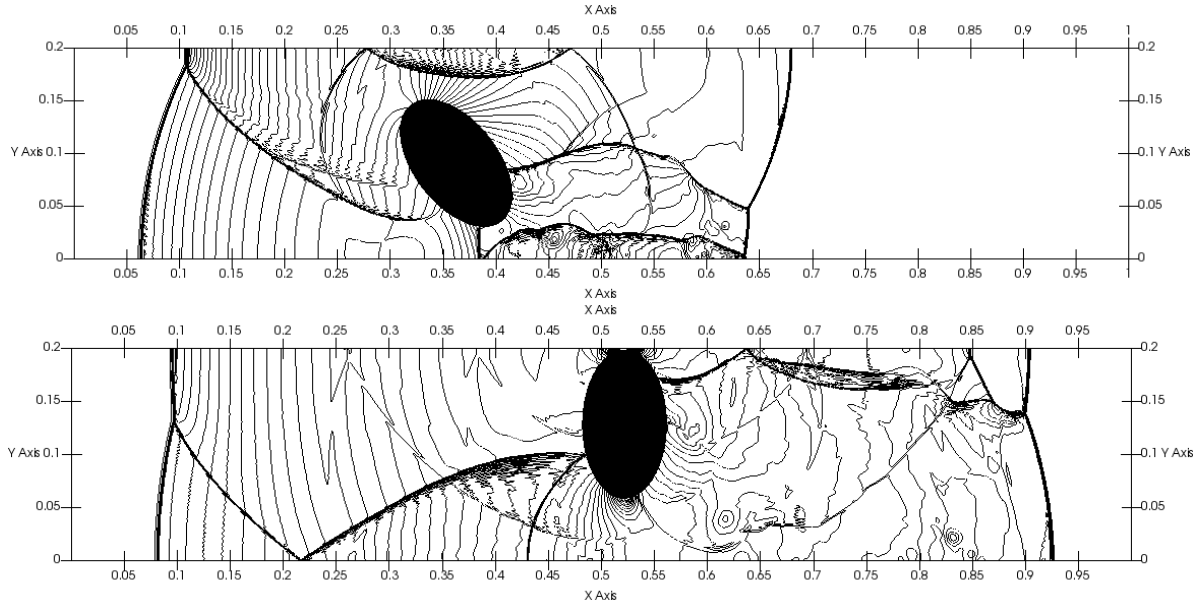


Figure V.6 – 60 contours of fluid pressure from 0 to 28 at times t=0.14 (top) and t=0.255 (bottom) for the third order scheme, $\Delta x = \Delta y = 6.25 \times 10^{-4}$.

and the parameters $x_0 = 15.10^{-2}$, $y_0 = 5.10^{-2}$, $\theta = -\frac{\pi}{10}$, $b = 3.10^{-2}$, $h = 5.10^{-2}$. Its density is set to $\rho = 9.0$ kg.m^{-2}. Initially, the rhombus is motionless. A Mach 3 shock enters the domain, and due to the asymmetry of the problem lifts the rhombus.

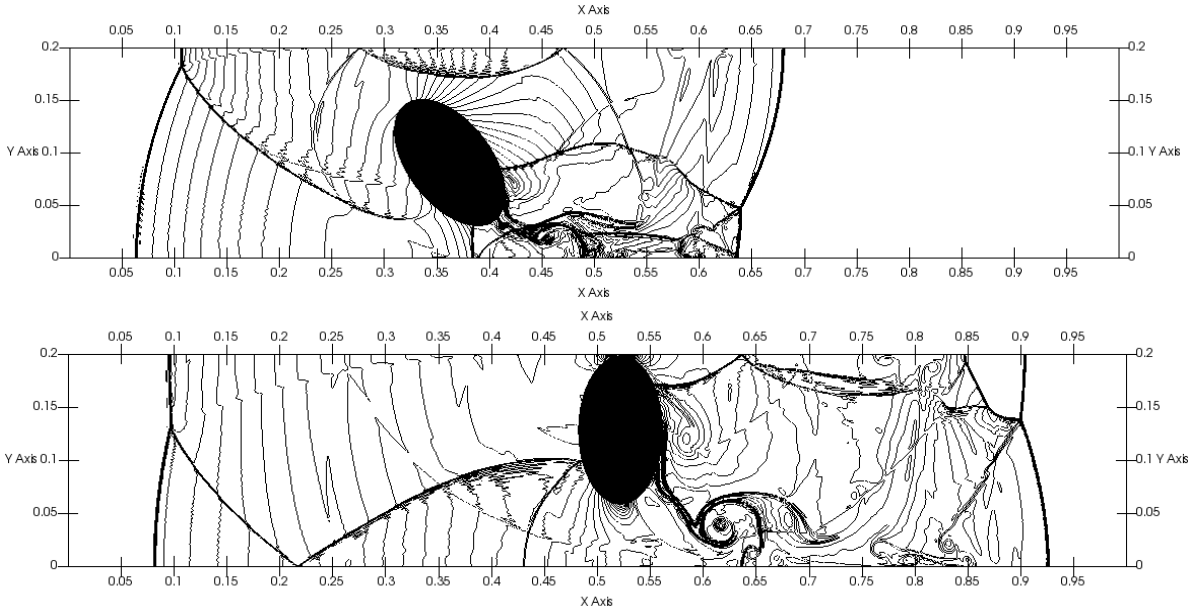


Figure V.7 – 60 contours of fluid density from 0 to 12 at times t=0.14 (top) and t=0.255 (bottom) for the third order scheme, $\Delta x = \Delta y = 6.25 \times 10^{-4}$.

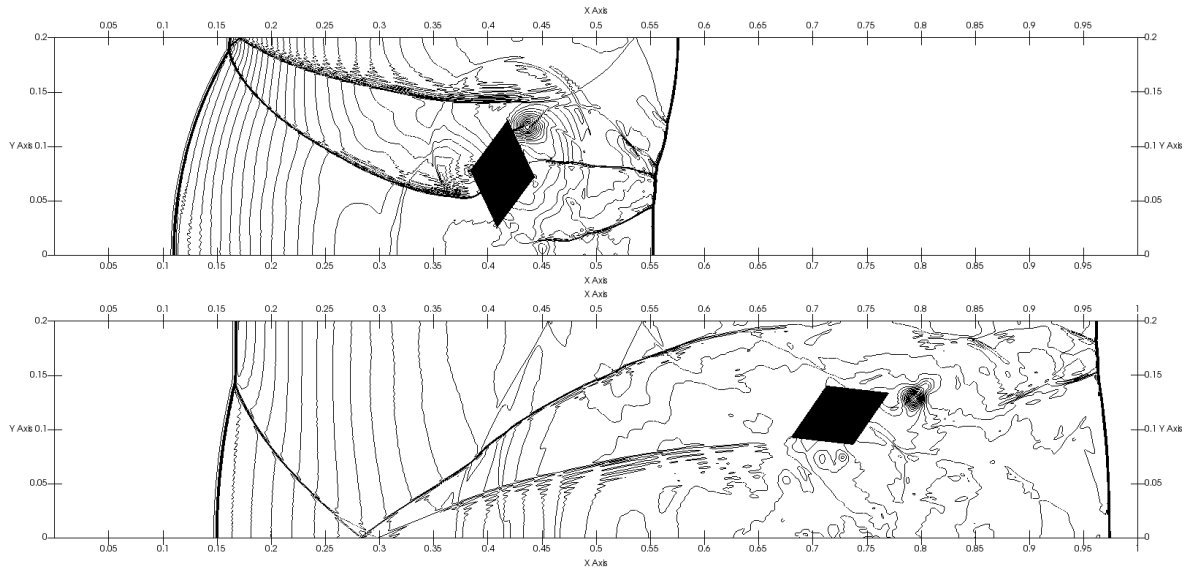


Figure V.8 – 60 contours of fluid pressure from 0 to 28 at times t=0.14 (top) and t=0.255 (bottom) for the third order scheme, $\Delta x = \Delta y = 6.25 \times 10^{-4}$.

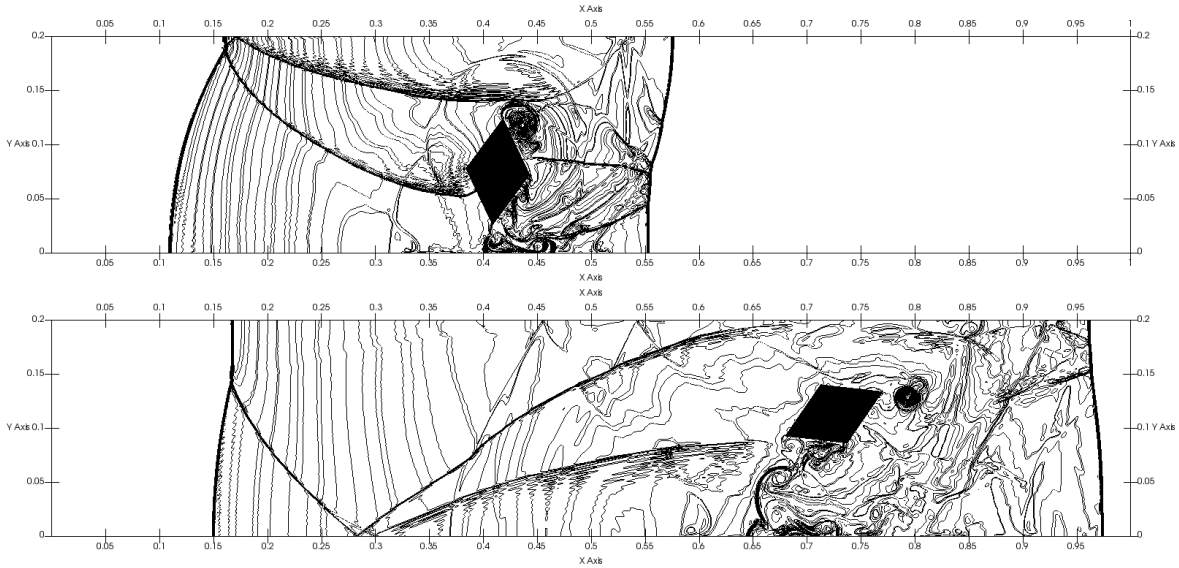


Figure V.9 – 60 contours of fluid density from 0 to 12 at times t=0.14 (top) and t=0.255 (bottom) for the third order scheme, $\Delta x = \Delta y = 6.25 \times 10^{-4}$.

Conclusions and perspectives

Dans ce manuscrit, nous nous sommes intéressés à des questions de simulation numériques pour l'interaction fluide-structure. Le modèle considéré a été celui de l'interaction entre un fluide compressible et une structure indéformable. Pour ce faire, une méthode de type Lax–Wendroff inverse a été mise au point pour réaliser un schéma de couplage fluide-structure explicite et stable. Ce couplage permet de faire communiquer un solveur de type Volumes Finis pour le fluide compressible avec un solveur pour la dynamique des corps rigides.

Dans un tout premier temps, des résultats principaux liés aux systèmes hyperboliques de lois de conservation ont été présentés. Puis, l'accent a été mis sur les méthodes de résolution des équations d'Euler pour un fluide compressible, ainsi que les différents couplages en espace comme en temps trouvés dans la littérature. Du fait des grandes disparités physiques entre les matériaux, la méthode des cellules mixes a été mise de côté, tout comme les méthodes épousant les contours du corps rigide (body-fitted) car non-adaptées aux schémas numériques sur grilles cartésiennes. Nous avons fait le choix de nous intéresser plus précisément aux méthodes de domaine fictif. Le dévolu a été jeté sur la méthode de frontières immergés (Immersed boundaries) en calculant les cellules fantômes par la procédure de Lax–Wendroff inverse. En effet, bien que n'assurant pas la conservation de la masse, de la quantité de mouvement et de l'énergie totale, elle permet une reconstruction à l'ordre très élevé des valeurs fantômes et assure ainsi un schéma final lui aussi d'ordre très élevé. Enfin, le choix a été fait de considérer un couplage explicite en temps afin d'éviter de devoir résoudre un problème non plus local, mais global autour de la frontière.

Ensuite, nous avons présenté la famille de schémas sur grilles cartésiennes décalées, potentiellement utilisée pour former le solveur fluide compressible. Cette famille de schéma a été démontrée comme étant conservative en masse, quantité de mouvement et énergie totale, ainsi que faiblement consistante avec les équations d'Euler. Le passage en multidimensionnel se fait par l'utilisation de séquences de splitting directionnel d'ordre élevé. Puis, l'extension de cette famille de schémas pour les équations de Navier–Stokes compressibles a été réalisée, impliquant une distribution particulière sur les grilles décalées des termes visqueux non-diagonaux. Des résultats numériques sont venus illustrer tout autant la précision que la robustesse de cette famille de schémas.

Puis, notre étude s'est portée sur la discrétisation des conditions aux bords, sur la précision ainsi que sur la stabilité qui en découlent. Afin de pouvoir se référer à des résultats théoriques, le problème a d'abord été traité dans le cas des systèmes hyperboliques linéaires. La procédure de

calcul des cellules fantômes a été développée dans le cas de l'équation de l'advection en 1D. Elle a été ensuite étendue au cas du système des équations des ondes, en considérant deux conditions aux bords différentes. Une première forçant la vitesse au bord, tandis que la seconde forçant une relation entre vitesse et pression. L'extension générique pour un système hyperbolique linéaire de lois de conservation a ensuite été détaillée. Bien que permettant de construire une méthode d'ordre très élevé, la procédure de Lax–Wendroff inverse n'assure pas pour autant la stabilité du schéma final obtenu. Cela a été mis en évidence par des expériences numériques sur le système des équations des ondes. Outre une étude de type GKS sur un schéma donné, il a été proposé de définir un critère de stabilité permettant en pratique de grandement simplifier les calculs nécessaires pour déterminer la stabilité d'un schéma. Ce critère s'est avéré, dans de nombreux cas, en parfaite concordance avec l'analyse GKS. Ce travail a mis en évidence la nécessité de s'intéresser tout particulièrement à la stabilité du schéma final obtenu et a permis de très largement simplifier l'étude faite ensuite dans le cas du système des équations d'Euler.

Pour la discrétisation des conditions aux bords imposées en vitesse pour les équations d'Euler, une méthode est déduite de l'analyse linéaire pour construire des cellules fantômes stables et d'ordre très élevé. Plus particulièrement, considérant des schémas intérieurs de type Lagrange-projection sur grilles cartésiennes, deux méthodes sont isolées pour l'imposition des valeurs fantômes. La première consiste à faire l'hypothèse d'isentropie spatiale locale proche de la frontière, tandis que la seconde consiste à élargir le stencil pour effectuer la reconstruction des valeurs fantômes. Des résultats théoriques permettent de caractériser les conditions d'existence et d'unicité de la reconstruction proposée par ces deux méthodes. Dans le but de traiter le cas de chocs forts impactant la frontière, une procédure de type MOOD a été développée. Enfin, l'extension au cas 2D a été faite. L'extrapolation polynomiale 2D étant fortement oscillante et ayant tendance à être instable, une procédure de type moindre carré a été introduite afin de lisser un tel comportement. Des résultats numériques sont venus illustrer tout autant la précision que la robustesse de la méthode proposée.

Enfin, le couplage entre un fluide compressible et une structure indéformable a été réalisé à partir de la procédure de Lax–Wendroff inverse développée précédemment. Un schéma semi-discret permettant de calculer à l'ordre élevé en espace les forces et moments exercés sur la frontière du corps rigide a été proposé. Deux procédures d'intégrations en temps ont ensuite été développées, une de type Runge–Kutta et une seconde de type Cauchy–Kowalevski. Ce choix d'intégration en temps a permis de faire correspondre sur la même échelle en temps les solveurs fluide et corps rigide. Enfin l'extension 2D de ces schémas a été faite via splitting directionnel. La procédure de Lax–Wendroff inverse nous a permis de définir naturellement les forces et moments de pression exercés sur la frontière du corps rigide. Ainsi le couplage fut immédiat et d'autant plus facile à implémenter. Quelques résultats numériques ont été proposés afin de mettre en évidence la stabilité et la robustesse du couplage utilisé.

Plusieurs perspectives sont désormais possibles. Dans un premier temps, il apparaît important d'étendre la méthode à trois dimensions d'espace. Cela permettrait d'approcher des situations plus réalistes. La méthode proposée devrait s'appliquer directement, sans modifications préalables, au

3D, à la condition de pouvoir répartir les "perles" de la méthode de Lax–Wendroff inverse sur la surface du solide. Cela ne devrait pas occasionner un surcoût prohibitif de la procédure par rapport au coût des solveurs fluide et structure.

Ensuite, il paraît intéressant de pouvoir considérer que la structure n'est plus simplement un corps rigide, donc indéformable, mais qu'elle suit d'autres lois de comportements (élasticité linéaire, hypoélasticité, plasticité, fracturation, ...). En maillant ainsi la structure, il paraîtrait dès lors tout à fait naturel de faire correspondre sommets du maillage sur la frontière et les "perles" utilisées lors de la procédure de Lax–Wendroff inverse. Considérant que l'analyse linéaire a déjà été faite, nous pouvons dès à présent nous assurer que le couplage ne devrait pas souffrir en terme de stabilité d'un tel traitement à la condition d'en faire aussi l'étude pour la partie structure.

De même, le modèle fluide pourrait être complexifié en prenant en compte une viscosité de type Navier–Stokes compressible. Bien que le solveur fluide ait déjà été proposée dans ce travail, l'analyse linéaire de stabilité n'a pas encore été effectuée et le couplage fluide visqueux et corps rigides n'en est encore qu'à ses prémices. De plus, il serait tout aussi possible d'utiliser la méthode proposée afin de réaliser un couplage entre deux fluides non-miscibles aux propriétés différentes, ou encore de considérer des conditions aux bords plus complexes.

Pour conclure, dans un contexte HPC, le code développé pour cette thèse est déjà entièrement parallélisé via MPI/OpenMP. Les principales procédures sont locales, hormis le calcul des résultantes des forces et des moments. En effet, le solveur corps rigide nécessite de nombreuses synchronisations afin de calculer les résultantes des forces et moments à sa surface, ce qui rend certainement le code non optimal. Réduire le nombre de communications globales, dans un contexte HPC, apparaît comme vital pour assurer un correct passage à l'échelle. Enfin l'insertion de la procédure au sein d'une plateforme AMR multi-physique [91] présenterait aussi son intérêt afin d'améliorer encore davantage la précision et le temps de calcul. Cela permettrait de pouvoir simuler des cas d'écoulements plus complexes.

In this manuscript, numerical simulation of fluid-structure interaction was of most interest to us, considering a compressible fluid interacting with a rigid body. In order to realize the coupling between the two, the inverse Lax–Wendroff procedure has been developed for stability and explicit time-coupling purposes. This coupling is done in a stable way for a compressible hydrodynamics solver and a rigid body dynamics one.

Firstly, an overview of main theoretical results concerning hyperbolic systems of conservation laws has been made. The emphasis was then laid on numerical methods for the resolution of compressible Euler equations as well as for space and time coupling used for fluid-structure interaction found in the literature. Due to tremendous materials physical discontinuities, the mixed-cells method was discarded. Methods based on body-fitted meshes were also discarded as they were irrelevant for hydrodynamics solver on Cartesian grids. The choice has been made to focus on fictitious domain methods, and more precisely on the immersed boundary methods. The selected method for the space coupling was to build high-order accurate ghost-cells values using the inverse Lax–Wendroff procedure. Although, this method does not ensure conservation of mass, momentum and total energy, contrarily to the embedded boundary methods, it yields high-order accuracy which is of most use for high-order hydrodynamics solver. Last, an explicit coupling has been chosen, rather than implicit or semi-implicit ones, in order to solve a local problem instead of a global one.

Secondly, as a possible choice for the hydrodynamics solver, a scheme based on staggered Cartesian grids has been detailed. The scheme was proven to be conservative in mass, momentum and total energy and also weakly consistent with the Euler equations. The key for both conservation and weak consistency is the internal energy corrector that has been proposed. For multiple space dimensions, the scheme was used with a high-order directional splitting method. Then, the extension of the scheme for the resolution of the compressible Navier–Stokes equations was made. It relies on a peculiar distribution of non-diagonal viscous terms on a grid staggered in both directions. Numerical results have illustrated both the accuracy and the robustness of the scheme.

Afterwards, numerical boundary treatment was considered, with a special focus on both high-order accuracy and stability. In order to use theoretical results, especially concerning linear stability for initial boundary value problems, the problem was dealt with for linear hyperbolic systems of conservation laws. The ghost-values computation procedure, called in the manuscript "reconstruction operator", was first developed for the special case of linear advection problems in 1D. Then, it was extended to the wave equations system considering two different but well-posed boundary conditions. The first boundary condition imposed only the velocity at the boundary, whereas the second linked both velocity and pressure at the boundary. The extension was then realized for generic linear hyperbolic systems. Although giving high-order accuracy for ghost values in the fictitious domain, the inverse Lax–Wendroff procedure does not ensure the stability of the effective scheme. It was pointed out by numerical experiments performed for the wave equations. Besides a GKS stability analysis done for a given scheme and reconstruction operator, a new stability criterion was proposed in order to ease greatly stability characterization for the

discretization of the initial value boundary problem. Numerical experiments assess the pratical relevancy of such a criterion. Our findings highlighted the need to focus particularly on linear stability of the effective scheme before tackling the case of non-linear problems. It alleviated greatly the study that was then performed for the Euler equations.

For the extension of numerical boundary treatment to compressible Euler equations, the boundary conditions was considered to be imposed as a slip boundary condition, enforcing the normal velocity. A method has been deduced from the linear analysis of the inverse Lax–Wendroff procedure to obtain high-order and stable effective schemes. More precisely, considering Lagrange-remap interior schemes based on Cartesian grids, the non-inversibility of the Jacobian matrix pointed out the need for another equation. Two methods were developed to build the reconstruction operator. The first one consisted in considering that the flow near the boundary was spatially isentropic. Whereas the second one consisted in enlarging the stencil used to build the reconstruction operator. Theoretical results to characterize conditions for existence and uniqueness of the reconstruction operator were proved for both methods. In order to deal with strong incoming or outgoing shocks, a MOOD procedure was developed. Then the extension to two space dimensions problems was done. A special procedure of least-square was also developed in order to prevent 2D extrapolation instabilities from occurring. Numerical experiments have been performed to illustrate both accuracy and robustness of the method.

Last, the coupling between a compressible fluid and a rigid body was made, starting from the previously introduced inverse Lax–Wendroff procedure for Lagrange-remap schemes. A semi-discrete scheme was derived, computing with high-order accuracy in space the resultants of forces and torques exerted on the rigid body boundary. Two time-integrations were proposed: A Runge–Kutta one, and a Cauchy–Kovalevskaya one. These time integration choices result from the hydrodynamics solver choices, and was done in order to maintain both solvers on the same time-scale, easing the coupling. Then, the two space dimensions extension was performed using directional splitting method. The inverse Lax–Wendroff procedure yielded natural definitions for pressure forces and torques exerted on the rigid body boundary. Thereby the coupling was straightforward and easy to implement. Some numerical results have been presented to emphasize the stability and robustness of the coupling.

New perspectives seem now to be reachable. Firstly, extending the method to three space dimensions should be quite straightforward and of very high interest. It would allow to get closer to more realistic situations. The proposed method can be applied straightforwardly provided one can map the inverse Lax–Wendroff pearls on the surface of a rigid body. Going from 2D to 3D should not induce large prohibitive numerical costs due to the procedure.

Then, considering a deformable structure instead of a rigid body one should be of great interest for the CEA needs. Many deformations models are available in the literature such as linear elasticity, hypo-elasticity, plasticity and fracturation. Once again, be given a set of pearls describing the structures boundaries, the ILW procedure should be applicable straightforwardly. The structure being described by a mesh, it seems all but natural to consider that the vertices on the boundary of the mesh are exactly the pearls used in the ILW procedure. The space and time coupling

should still hold for such a more complex multi-physics problem, provided the linear stability is also performed for the structure part.

Identically, the fluid model could be made more complex. The ILW procedure was designed whether for internal energy affine equations of state or for equations of state such that the square of the sound speed is Lipschitz continuous but without any viscous components. Same analysis and works could be performed considering the fluid to follow the compressible Navier–Stokes equations instead of the compressible Euler ones. Although a compressible Navier–Stokes solver was proposed in this manuscript, the linear analysis for initial boundary values problem was not performed, and the viscous fluid rigid body coupling is still in its early stages. Moreover, the method could also be applied to realize a coupling between two immiscible fluids with different constitutive laws or to consider more complex boundary conditions than just slip boundary conditions.

In conclusion, in a HPC context, the code that was implemented during this PhD is already running in parallel using MPI/OpenMP. Since every procedure is local, the parallel computing is straightforward for the fluid part and for the discretization of boundary conditions. However the rigid body solver requires many synchronizations between the processes to get the values of forces and torques resultants and then to compute the displacement. Reducing the number of global communications, in a HPC context, is of the essence to enforce correct scalability of the method. As a last word, implementing such procedures inside the multiphysics AMR platform [91] would be of special interest to improve even more accuracy and computational cost, and so to run even more complex simulations.

Appendix A

Butcher tables and weights for directional splitting methods

L'annexe comprend l'ensemble des tableaux de coefficients de grande taille afin de fournir au lecteur la possibilité de reproduire les méthodes utilisées et décrites dans le manuscrit.

A.1 Butcher table for usual Runge–Kutta sequences

We remind here briefly the Butcher table for a given explicit Runge–Kutta sequence.

$$
\begin{array}{c|ccccc}
\alpha_1 & a_{1,0} & 0 & 0 & 0 & \cdots \\
\alpha_2 & a_{2,0} & a_{2,1} & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \ddots & \cdots & \cdots \\
\alpha_s & a_{s,0} & \cdots & \cdots & a_{s,s-1} & 0 \\
\hline
1 & \theta_0 & \theta_1 & \cdots & \theta_{s-1} & \theta_s
\end{array}
$$

$$
\begin{array}{c|cc}
\alpha & \alpha & 0 \\
\hline
1 & 1-\frac{1}{2\alpha} & \frac{1}{2\alpha}
\end{array}
$$

Table A.1 – Generic second order Runge–Kutta sequence

$$
\begin{array}{c|ccc}
0 & & & \\
1 & 1 & & \\
\frac{1}{2} & \frac{1}{4} & \frac{1}{4} & \\
\hline
1 & \frac{1}{6} & \frac{1}{6} & \frac{2}{3}
\end{array}
$$

Table A.2 – Third order TVD Runge–Kutta sequence [70]

$$
\begin{array}{c|cccc}
0 & & & & \\
\frac{1}{2} & \frac{1}{2} & & & \\
\frac{1}{2} & 0 & \frac{1}{2} & & \\
1 & 0 & 0 & 1 & \\
\hline
 & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}
\end{array}
$$

Table A.3 – Original Kutta sequence [99]

0				
$\frac{1}{3}$	$\frac{1}{3}$			
$\frac{2}{3}$	-$\frac{1}{3}$	1		
1	1	-1	1	
	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Table A.4 – The $\frac{3}{8}$-Kutta sequence [99]

0						
$\frac{1}{5}$	$\frac{1}{5}$					
$\frac{3}{10}$	$\frac{3}{40}$	$\frac{9}{40}$				
$\frac{4}{5}$	$\frac{44}{45}$	$-\frac{56}{15}$	$\frac{32}{9}$			
$\frac{8}{9}$	$\frac{19372}{6561}$	$\frac{-25360}{2187}$	$\frac{64448}{6561}$	$\frac{-212}{729}$		
1	$\frac{9017}{3168}$	$\frac{-355}{33}$	$\frac{46732}{5247}$	$\frac{49}{176}$	$\frac{-5103}{18656}$	
	$\frac{35}{384}$	0	$\frac{500}{1113}$	$\frac{125}{192}$	$\frac{-2187}{6784}$	$\frac{11}{84}$

Table A.5 – Dormand–Prince Runge–Kutta sequence [49]

A.2 Directional splitting weights sequences

ω_1	1
ω_2	1

Table A.6 – First order Godunov splitting weights ω_k

ω_1	0.5
ω_2	1
ω_3	0.5

Table A.7 – Second order Strang splitting weights ω_k

ω_1	0.26833009578175993
ω_2	0.91966152301739986
ω_3	-0.18799161879915978
ω_4	-0.18799161879915978
ω_5	0.91966152301739986
ω_6	0.26833009578175993

Table A.8 – Third order directional splitting weights ω_k

ω_1	0.5
ω_2	-0.05032120814910445
ω_3	-0.27516060407455222
ω_4	0.55032120814910445
ω_5	0.55032120814910445
ω_6	0.55032120814910445
ω_7	-0.27516060407455222
ω_8	-0.05032120814910445
ω_9	0.5

Table A.9 – Fourth order directional splitting weights ω_k

ω_1	0.3922568052387787
ω_2	0.7845136104775573
ω_3	0.5100434119184577
ω_4	0.2355732133593581
ω_5	-0.4710533854097564
ω_6	-1.1776799841788710
ω_7	0.0687531682525201
ω_8	1.3151863206839112
ω_9	0.0687531682525201
ω_{10}	-1.1776799841788710
ω_{11}	-0.4710533854097564
ω_{12}	0.2355732133593581
ω_{13}	0.5100434119184577
ω_{14}	0.7845136104775573
ω_{15}	0.3922568052387787

Table A.10 – Sixth order directional splitting weights ω_k [175]

ω_1	0.3145153251052165
ω_2	0.629030650210433
ω_3	0.9991900571895715
ω_4	1.36934946416871
ω_5	0.152381158138440
ω_6	-1.06458714789183
ω_7	0.299385475870660
ω_8	1.66335809963315
ω_9	-0.007805591481625
ω_{10}	-1.67896928259640
ω_{11}	-1.619218660405435
ω_{12}	-1.55946803821447
ω_{13}	-0.6238386128980215
ω_{14}	0.311790812418427
ω_{15}	0.98539084848119350
ω_{16}	1.6589908845439600
ω_{17}	0.98539084848119350
ω_{18}	0.311790812418427
ω_{19}	-0.6238386128980215
ω_{20}	-1.55946803821447
ω_{21}	-1.619218660405435
ω_{22}	-1.67896928259640
ω_{23}	-0.007805591481625
ω_{24}	1.66335809963315
ω_{25}	0.299385475870660
ω_{26}	-1.06458714789183
ω_{27}	0.152381158138440
ω_{28}	1.36934946416871
ω_{29}	0.9991900571895715
ω_{30}	0.629030650210433
ω_{31}	0.3145153251052165

Table A.11 – Eighth order directional splitting weights ω_k [175]

References

[1] R. Abgrall and S. Karni. "A comment on the computation of non-conservative products". In: *Journal of Computational Physics* 229.8 (2010), pp. 2759–2763.

[2] G. Allaire. *Numerical analysis and optimization: an introduction to mathematical modelling and numerical simulation.* Oxford University Press, 2007.

[3] G. Allaire, S. M. Kaber, and K. Trabelsi. *Numerical linear algebra.* Vol. 55. Springer, 2008.

[4] P. Angot, C.-H. Bruneau, and P. Fabrie. "A penalization method to take into account obstacles in incompressible viscous flows". In: *Numerische Mathematik* 81.4 (1999), pp. 497–520.

[5] A. Arakawa and V. R. Lamb. "Computational design of the basic dynamical processes of the UCLA general circulation model". In: *Methods in computational physics* 17 (1977), pp. 173–265.

[6] M. Arienti, P. Hung, E. Morano, and J. E. Shepherd. "A level set approach to Eulerian–Lagrangian coupling". In: *Journal of Computational Physics* 185.1 (2003), pp. 213–251.

[7] E. Arquis and J. Caltagirone. "Sur les conditions hydrodynamiques au voisinage d'une interface milieu fluide-milieu poreux: application à la convection naturelle". In: *CR Acad. Sci. Paris II* 299 (1984), pp. 1–4.

[8] K. Attenborough, S. Taherzadeh, H. E. Bass, X. Di, R. Raspet, G. Becker, A. Güdesen, A. Chrestman, G. A. Daigle, A. L'Espérance, et al. "Benchmark cases for outdoor sound propagation models". In: *The Journal of the Acoustical Society of America* 97.1 (1995), pp. 173–191.

[9] A. L. Bauer, D. E. Burton, E. Caramana, R. Loubère, M. J. Shashkov, and P. Whalen. "The internal consistency, stability, and accuracy of the discrete, compatible formulation of Lagrangian hydrodynamics". In: *Journal of Computational Physics* 218.2 (2006), pp. 572–593.

[10] M. Belliard, M. Chandesris, J. Dumas, Y. Gorsse, D. Jamet, C. Josserand, and B. Mathieu. "An analysis and an affordable regularization technique for the spurious force oscillations in the context of direct-forcing immersed boundary methods". In: *Computers & Mathematics with Applications* 71.5 (2016), pp. 1089–1113.

[11] D. J. Benson. "Computational methods in Lagrangian and Eulerian hydrocodes". In: *Computer methods in Applied mechanics and Engineering* 99.2-3 (1992), pp. 235–394.

[12] M. J. Berger, C. Helzel, and R. J. LeVeque. "H-box methods for the approximation of hyperbolic conservation laws on irregular grids". In: *SIAM Journal on Numerical Analysis* 41.3 (2003), pp. 893–918.

[13] A. Bhagatwala and S. K. Lele. "A modified artificial viscosity approach for compressible turbulence simulations". In: *Journal of Computational Physics* 228.14 (2009), pp. 4965–4969.

[14] K. Bisshopp. "Note on rigid body motion". In: *Journal of Mechanisms* 6.3 (1971), pp. 259–266.

[15] J. J. Bowman, T. B. Senior, and P. L. Uslenghi. "Electromagnetic and acoustic scattering by simple shapes (Revised edition)". In: *New York, Hemisphere Publishing Corp., 1987, 747 p.* 1 (1987).

[16] J. C. Butcher. "Coefficients for the study of Runge-Kutta integration processes". In: *Journal of the Australian Mathematical Society* 3.02 (1963), pp. 185–201.

[17] J. C. Butcher. "On Runge-Kutta processes of high order". In: *Journal of the Australian Mathematical Society* 4.02 (1964), pp. 179–194.

[18] E. J. Caramana, M. J. Shashkov, and P. P. Whalen. "Formulations of artificial viscosity for multi-dimensional shock wave computations". In: *Journal of Computational Physics* 144.1 (1998), pp. 70–97.

[19] E. Caramana, D. Burton, M. Shashkov, and P. Whalen. "The construction of compatible hydrodynamics algorithms utilizing conservation of total energy". In: *Journal of Computational Physics* 146.1 (1998), pp. 227–262.

[20] M. H. Carpenter, D. Gottlieb, S. Abarbanel, and W.-S. Don. "The theoretical accuracy of Runge-Kutta time discretizations for the initial boundary value problem: a study of the boundary error". In: *SIAM Journal on Scientific Computing* 16.6 (1995), pp. 1241–1252.

[21] G. Carré, S. Del Pino, B. Després, and E. Labourasse. "A cell-centered Lagrangian hydrodynamics scheme on general unstructured meshes in arbitrary dimension". In: *Journal of Computational Physics* 228.14 (2009), pp. 5160–5183.

[22] J. G. Charney, R. Fjörtoft, and J. v. Neumann. "Numerical integration of the barotropic vorticity equation". In: *Tellus* 2.4 (1950), pp. 237–254.

[23] A. Chaudhuri, A. Hadjadhj, and A. Chinnayya. "On the use of immersed boundary methods for shock/obstacle interactions". In: *J. Comput. Physics* 230 (2011), pp. 1731–1748.

[24] S. Clain, S. Diot, and R. Loubere. "A high-order finite volume method for systems of conservation laws—Multi-dimensional Optimal Order Detection (MOOD)". In: *Journal of computational Physics* 230.10 (2011), pp. 4028–4050.

[25] B. Cockburn, G. E. Karniadakis, and C.-W. Shu. *The development of discontinuous Galerkin methods.* Springer, 2000.

[26] P. Colella, D. T. Graves, B. J. Keen, and D. Modiano. "A Cartesian grid embedded boundary method for hyperbolic conservation laws". In: *Journal of Computational Physics* 211.1 (2006), pp. 347–366.

[27] A. W. Cook and W. H. Cabot. "Hyperviscosity for Shock-turbulence Interactions". In: *J. Comput. Phys.* 203 (2005), pp. 379–385.

[28] A. W. Cook and W. H. Cabot. "A high-wavenumber viscosity for high-resolution numerical methods". In: *Journal of Computational Physics* 195.2 (2004), pp. 594–601.

[29] J.-F. Coulombel. "Fully discrete hyperbolic initial boundary value problems with nonzero initial data". In: *arXiv preprint arXiv:1412.0851* (2014).

[30] J.-F. Coulombel. "The Leray-Gårding method for finite difference schemes". In: *arXiv preprint arXiv:1505.06060* (2015).

[31] J.-F. Coulombel and A. Gloria. "Semigroup stability of finite difference schemes for multidimensional hyperbolic initial-boundary value problems". In: *Mathematics of computation* 80.273 (2011), pp. 165–203.

[32] R. Courant, K. Friedrichs, and H. Lewy. "Über die partiellen Differenzengleichungen der mathematischen Physik". In: *Mathematische annalen* 100.1 (1928), pp. 32–74.

[33] J. Crank and P. Nicolson. "A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type". In: *Mathematical Proceedings of the Cambridge Philosophical Society.* Vol. 43. 01. Cambridge Univ Press. 1947, pp. 50–67.

[34] G. Dakin, B. Desprès, and S. Jaouen. "Inverse Lax–Wendroff boundary treatment for compressible hydrodynamics Lagrange-remap schemes on Cartesian grids". In: *Journal of computational physics* (submitted).

[35] G. Dakin and H. Jourdren. "High-order accurate Lagrange-remap hydrodynamic schemes on staggered Cartesian grids". In: *Comptes Rendus Mathematique* (2016).

[36] G. Dal Maso, P. G. Lefloch, and F. Murat. "Definition and weak stability of nonconservative products". In: *Journal de mathématiques pures et appliquées* 74.6 (1995), pp. 483–548.

[37] R. B. DeBar. *Fundamentals of the KRAKEN code.* Tech. rep. Lawrence Livermore National Lab., CA (USA), 1974.

[38] R. B. DeBar. *Method in two-D Eulerian hydrodynamics.* Tech. rep. Lawrence Livermore National Lab., CA (USA), 1974.

[39] S. Del Pino, B. Desprès, P. Havé, H. Jourdren, and P. Piserchia. "3D Finite Volume simulation of acoustic waves in the earth atmosphere". In: *Computers & Fluids* 38.4 (2009), pp. 765–777.

[40] S. Del Pino and H. Jourdren. "Arbitrary high-order schemes for the linear advection and wave equations : application to hydrodynamics and aeroacoustics". In: *C. R. Acad. Sci. Paris, Ser. I* 342 (2006), pp. 441–446.

[41] I. Demirdžić and S. Muzaferija. "Numerical method for coupled fluid flow, heat transfer and stress analysis using unstructured moving meshes with cells of arbitrary topology". In: *Computer methods in applied mechanics and engineering* 125.1 (1995), pp. 235–255.

[42] B. Després. "Lagrangian systems of conservation laws". In: *Numerische Mathematik* 89.1 (2001), pp. 99–134.

[43] B. Després. "Finite volume transport schemes". In: *Numerische Mathematik* 108.4 (2008), pp. 529–556.

[44] B. Després. "Uniform asymptotic stability of Strang's explicit compact schemes for linear advection". In: *SIAM Journal on Numerical Analysis* 47.5 (2009), pp. 3956–3976.

[45] B. Després. *Lois de conservations eulériennes, lagrangiennes et méthodes numériques.* Vol. 68. Springer Science & Business Media, 2010.

[46] B. Després. "Weak consistency of the cell-centered Lagrangian GLACE scheme on general meshes in any dimension". In: *Computer Methods in Applied Mechanics and Engineering* 199.41 (2010), pp. 2669–2679.

[47] B. Després and F. Dubois. *Systèmes hyperboliques de lois de conservation: Application à la dynamique des gaz.* Editions Ecole Polytechnique, 2005.

[48] B. Després and C. Mazeran. "Lagrangian gas dynamics in two dimensions and Lagrangian systems". In: *Archive for Rational Mechanics and Analysis* 178.3 (2005), pp. 327–372.

[49] J. R. Dormand and P. J. Prince. "A family of embedded Runge-Kutta formulae". In: *Journal of computational and applied mathematics* 6.1 (1980), pp. 19–26.

[50] F. Duboc, C. Enaux, S. Jaouen, H. Jourdren, and M. Wolff. "High-order dimensionally split Lagrange-remap schemes for compressible hydrodynamics". In: *C. R. Acad. Sci. Paris, Ser. I* 348 (2010), pp. 105–110.

[51] T. Engels, D. Kolomenskiy, K. Schneider, and J. Sesterhenn. "Numerical simulation of fluid–structure interaction with the volume penalization method". In: *Journal of Computational Physics* 281 (2015), pp. 96–115.

[52] E. Fadlun, R. Verzicco, P. Orlandi, and J. Mohd-Yusof. "Combined immersed-boundary finite-difference methods for three-dimensional complex flow simulations". In: *Journal of computational physics* 161.1 (2000), pp. 35–60.

[53] J. Falcovitz, G. Alfandary, and G. Hanoch. "A two-dimensional conservation laws scheme for compressible flows with moving boundaries". In: *Journal of Computational Physics* 138.1 (1997), pp. 83–102.

[54] R. Featherstone. *Rigid body dynamics algorithms.* Springer, 2014.

[55] E. Fehlberg. "Low-order classical Runge-Kutta formulas with stepsize control and their application to some heat transfer problems". In: (1969).

[56] M. A. Fernández. "Coupling schemes for incompressible fluid-structure interaction: implicit, semi-implicit and explicit". In: *SeMA Journal* 55.1 (2011), pp. 59–108.

[57] H. Forrer and M. Berger. "Flow simulations on Cartesian grids involving complex moving geometries". In: *Hyperbolic problems: theory, numerics, applications*. Springer, 1999, pp. 315–324.

[58] T. Gallouët, R. Herbin, and J.-C. Latché. "Kinetic energy control in explicit finite volume discretizations of the incompressible and compressible Navier-Stokes equations". In: *International Journal on Finite Volumes* 7.2 (2010), pp. 1–6.

[59] R. Glowinski, T.-W. Pan, T. I. Hesla, D. D. Joseph, and J. Periaux. "A distributed Lagrange multiplier/fictitious domain method for the simulation of flow around moving rigid bodies: application to particulate flow". In: *Computer methods in applied mechanics and engineering* 184.2 (2000), pp. 241–267.

[60] R. Glowinski, T.-W. Pan, T. I. Hesla, and D. D. Joseph. "A distributed Lagrange multiplier/fictitious domain method for particulate flows". In: *International Journal of Multiphase Flow* 25.5 (1999), pp. 755–794.

[61] E. Godlewski and P.-A. Raviart. *Numerical approximation of hyperbolic systems of conservation laws*. Vol. 118. Springer Science & Business Media, 2013.

[62] S. K. Godunov. "A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics". In: *Matematicheskii Sbornik* 89.3 (1959), pp. 271–306.

[63] S. K. Godunov and V. S. Ryaben'kii. "Spectral stability criteria for boundary-value problems for non-self-adjoint difference equations". In: *Russian Mathematical Surveys* 18.3 (1963), pp. 1–12.

[64] M. Goldberg. "On a boundary extrapolation theorem by Kreiss". In: *Mathematics of Computation* 31.138 (1977), pp. 469–477.

[65] M. Goldberg and E. Tadmor. "Scheme-independent stability criteria for difference approximations of hyperbolic initial-boundary value problems. I". In: *Mathematics of Computation* 32.144 (1978), pp. 1097–1107.

[66] M. Goldberg and E. Tadmor. "Scheme-independent stability criteria for difference approximations of hyperbolic initial-boundary value problems. II". In: *mathematics of computation* 36.154 (1981), pp. 603–626.

[67] M. Goldberg and E. Tadmor. "Simple stability criteria for difference approximations of hyperbolic initial-boundary value problems". In: *Nonlinear Hyperbolic Equations—Theory, Computation Methods, and Applications*. Springer, 1989, pp. 179–185.

[68] Y. Gorsse, A. Iollo, H. Telib, and L. Weynans. "A simple second order cartesian scheme for compressible Euler flows". In: *Journal of Computational Physics* 231.23 (2012), pp. 7780–7794.

[69] S. Gottlieb, D. I. Ketcheson, and C.-W. Shu. "High order strong stability preserving time discretizations". In: *Journal of Scientific Computing* 38.3 (2009), pp. 251–289.

[70] S. Gottlieb and C.-W. Shu. "Total variation diminishing Runge-Kutta schemes". In: *Mathematics of computation of the American Mathematical Society* 67.221 (1998), pp. 73–85.

[71] S. Gottlieb, C.-W. Shu, and E. Tadmor. "Strong stability-preserving high-order time discretization methods". In: *SIAM review* 43.1 (2001), pp. 89–112.

[72] B. E. Griffith, R. D. Hornung, D. M. McQueen, and C. S. Peskin. "An adaptive, formally second order accurate version of the immersed boundary method". In: *Journal of Computational Physics* 223.1 (2007), pp. 10–49.

[73] B. E. Griffith and C. S. Peskin. "On the order of accuracy of the immersed boundary method: Higher order convergence rates for sufficiently smooth problems". In: *Journal of Computational Physics* 208.1 (2005), pp. 75–105.

[74] J.-L. Guermond and B. Popov. "Viscous regularization of the Euler equations and entropy principles". In: *SIAM Journal on Applied Mathematics* 74.2 (2014), pp. 284–305.

[75] B. Gustafsson. "The Godunov-Ryabenkii condition: The beginning of a new stability theory". In: *Godunov Methods*. Springer, 2001, pp. 425–443.

[76] B. Gustafsson, H.-O. Kreiss, and A. Sundström. "Stability theory of difference approximations for mixed initial boundary value problems. II". In: *Mathematics of Computation* (1972), pp. 649–686.

[77] A. Harten, B. Engquist, S. Osher, and S. R. Chakravarthy. "Uniformly high order accurate essentially non-oscillatory schemes, III". In: *Journal of computational physics* 71.2 (1987), pp. 231–303.

[78] D. Hartmann, M. Meinke, and W. Schröder. "A strictly conservative Cartesian cut-cell method for compressible viscous flows on adaptive grids". In: *Computer Methods in Applied Mechanics and Engineering* 200.9 (2011), pp. 1038–1052.

[79] P. He and R. Qiao. "A full-Eulerian solid level set method for simulation of fluid–structure interactions". In: *Microfluidics and nanofluidics* 11.5 (2011), pp. 557–567.

[80] R. Herbin, J.-C. Latché, and T. T. Nguyen. "Explicit staggered schemes for the compressible Euler equations". In: *ESAIM: Proceedings*. Vol. 40. EDP Sciences. 2013, pp. 83–102.

[81] R. Herbin, J.-C. Latché, and T. T. Nguyen. "Consistent segregated staggered schemes with explicit steps for the isentropic and full Euler equations". In: (2017).

[82] R. Herbin and J.-C. Latché. "Kinetic energy control in the MAC discretization of the compressible Navier-Stokes equations". In: *International Journal of Finites Volumes* 7 (2010).

[83] R. Herbin, J.-C. Latché, and T. T. Nguyen. "Consistent explicit staggered schemes for compressible flows Part II: the Euler equation". In: (2013).

[84] O. Heuzé, S. Jaouen, and H. Jourdren. "Dissipative issue of high-order shock capturing schemes with non-convex equations of state". In: *J. Comput. Physics* 228 (2009), pp. 833–860.

[85] W. B. Holzapfel. "Equations of state and thermophysical properties of solids under pressure". In: *High-Pressure Crystallography*. Springer, 2004, pp. 217–236.

[86] H. H. Hu, D. D. Joseph, and M. J. Crochet. "Direct simulation of fluid particle motions". In: *Theoretical and Computational Fluid Dynamics* 3.5 (1992), pp. 285–306.

[87] H. H. Hu, N. A. Patankar, and M. Zhu. "Direct numerical simulations of fluid–solid systems using the arbitrary Lagrangian–Eulerian technique". In: *Journal of Computational Physics* 169.2 (2001), pp. 427–462.

[88] X. Hu, B. Khoo, N. A. Adams, and F. Huang. "A conservative interface method for compressible flows". In: *Journal of Computational Physics* 219.2 (2006), pp. 553–578.

[89] R. N. Jazar. "Rigid-Body Dynamics". In: *Advanced Dynamics*. John Wiley Sons, Inc., 2011, pp. 1072–1188.

[90] G. Jiang and C. Shu. "Efficient Implementation of Weighted ENO Schemes". In: *J. Comput. Physics* 126 (1996), pp. 202–228.

[91] H. Jourdren. "HERA: a Hydrodynamic AMR Platform for Multi-Physics Simulations". In: *LNCSE* Springer, 41 (2005), pp. 283–294.

[92] R. Käppeli and S. Mishra. "Well-balanced schemes for the Euler equations with gravitation". In: *Journal of Computational Physics* 259 (2014), pp. 199–219.

[93] S. Kawai, S. K. Shankar, and S. K. Lele. "Assessment of localized artificial diffusivity scheme for large-eddy simulation of compressible turbulent flows". In: *Journal of Computational Physics* 229.5 (2010), pp. 1739–1762.

[94] M. Kenamond, M. Bement, and M. Shashkov. "Compatible, total energy conserving and symmetry preserving arbitrary Lagrangian–Eulerian hydrodynamics in 2D rz–Cylindrical coordinates". In: *Journal of Computational Physics* 268 (2014), pp. 154–185.

[95] R. E. Kidder. *The Theory of Homogeneous Isentropic Compression and its Application to Laser Fusion*. Springer. Vol. 3B. 1974, pp. 449–464.

[96] H.-O. Kreiss. "Stability theory for difference approximations of mixed initial boundary value problems. I". In: *Mathematics of Computation* (1968), pp. 703–714.

[97] H.-O. Kreiss. "Initial boundary value problems for hyperbolic systems". In: *Communications on Pure and Applied Mathematics* 23.3 (1970), pp. 277–298.

[98] L. Krivodonova and M. Berger. "High-order accurate implementation of solid wall boundary conditions in curved geometries". In: *Journal of computational physics* 211.2 (2006), pp. 492–512.

[99] W. Kutta. "Beitrag zur näherungweisen Integration totaler Differentialgleichungen". In: (1901).

[100] M.-C. Lai and C. S. Peskin. "An immersed boundary method with formal second-order accuracy and reduced numerical viscosity". In: *Journal of Computational Physics* 160.2 (2000), pp. 705–719.

[101] R. Landshoff. *A numerical method for treating fluid flow in the presence of shocks*. Tech. rep. DTIC Document, 1955.

[102] P. D. Lax. "Hyperbolic systems of conservation laws II". In: *Communications on Pure and Applied Mathematics* 10.4 (1957), pp. 537–566.

[103] P. D. Lax. *Hyperbolic systems of conservation laws and the mathematical theory of shock waves*. Vol. 11. SIAM, 1973.

[104] P. D. Lax and X.-D. Liu. "Solution of two-dimensional Riemann problems of gas dynamics by positive schemes". In: *SIAM Journal on Scientific Computing* 19.2 (1998), pp. 319–340.

[105] P. D. Lax and R. D. Richtmyer. "Survey of the stability of linear finite difference equations". In: *Communications on pure and applied mathematics* 9.2 (1956), pp. 267–293.

[106] P. D. Lax and B. Wendroff. "Systems of conservation laws". In: *Communications on Pure and Applied mathematics* 13.2 (1960), pp. 217–237.

[107] P. Le Tallec and J. Mouro. "Fluid structure interaction with large structural displacements". In: *Computer Methods in Applied Mechanics and Engineering* 190.24 (2001), pp. 3039–3067.

[108] R. Liska and B. Wendroff. "Comparison of several difference schemes on 1D and 2D test problems for the Euler equations". In: *SIAM Journal on Scientific Computing* 25.3 (2003), pp. 995–1017.

[109] A. Llor, A. Claisse, and C. Fochesato. "Energy preservation and entropy in Lagrangian space-and time-staggered hydrodynamic schemes". In: *Journal of Computational Physics* 309 (2016), pp. 324–349.

[110] P.-H. Maire. "A high-order cell-centered Lagrangian scheme for two-dimensional compressible fluid flows on unstructured meshes". In: *Journal of Computational Physics* 228.7 (2009), pp. 2391–2425.

[111] A. E. Mattsson and W. J. Rider. "Artificial viscosity: back to the basics". In: *International Journal for Numerical Methods in Fluids* 77.7 (2015), pp. 400–417.

[112] R. I. McLachlan. "On the numerical integration of ordinary differential equations by symmetric composition methods". In: *SIAM Journal on Scientific Computing* 16.1 (1995), pp. 151–168.

[113] R. I. McLachlan and P. Atela. "The accuracy of symplectic integrators". In: *Nonlinearity* 5.2 (1992), p. 541.

[114] R. I. McLachlan and G. R. W. Quispel. "Splitting methods". In: *Acta Numerica* 11 (2002), pp. 341–434.

[115] M. Meyer, A. Devesa, S. Hickel, X. Hu, and N. Adams. "A conservative immersed interface method for large-eddy simulation of incompressible flows". In: *Journal of Computational Physics* 229.18 (2010), pp. 6300–6317.

[116] C. Michler, S. Hulshoff, E. Van Brummelen, and R. De Borst. "A monolithic approach to fluid–structure interaction". In: *Computers & fluids* 33.5 (2004), pp. 839–848.

[117] R. Mittal and G. Iaccarino. "Immersed boundary methods". In: *Annu. Rev. Fluid Mech.* 37 (2005), pp. 239–261.

[118] J. Mohd-Yusof. "Combined immersed-boundary/B-spline methods for simulations of ow in complex geometries". In: *Annual Research Briefs. NASA Ames Research Center= Stanford University Center of Turbulence Research: Stanford* (1997), pp. 317–327.

[119] L. Monasse. "Analyse d'une méthode de couplage entre un fluide compressible et une structure déformable". PhD thesis. Université Paris-Est, 2011.

[120] L. Monasse, V. Daru, C. Mariotti, S. Piperno, and C. Tenaud. "A conservative coupling algorithm between a compressible flow and a rigid body using an Embedded Boundary method". In: *Journal of Computational Physics* 231.7 (2012), pp. 2977–2994.

[121] B. Muralidharan and S. Menon. "A high-order adaptive Cartesian cut-cell method for simulation of compressible viscous flow over immersed bodies". In: *Journal of Computational Physics* 321 (2016), pp. 342–368.

[122] S. Nagarajan, S. K. Lele, and J. H. Ferziger. "A robust high-order compact method for large eddy simulation". In: *Journal of Computational Physics* 191.2 (2003), pp. 392–419.

[123] H. Nessyahu and E. Tadmor. "Non-oscillatory central differencing for hyperbolic conservation laws". In: *Journal of computational physics* 87.2 (1990), pp. 408–463.

[124] J. von Neumann and R. D. Richtmyer. "A method for numerical calculation of hydrodynamic shocks". In: *J. Appl. Phys.* 21 (1950), pp. 232–237.

[125] W. Noh. "Fundamental methods of hydrodynamics". In: *Methods of Computational Physics* 3 (1964), pp. 117–179.

[126] W. Noh. "Numerical methods in hydrodynamic calculations". In: *Lawrence Livermore Laboratory Report* Technical Report UCRL-52112 (1976).

[127] W. F. Noh. "Errors for calculations of strong shocks using an artificial viscosity and an artificial heat flux". In: *Journal of Computational Physics* 72.1 (1987), pp. 78–120.

[128] P. Olsson. "Summation by parts, projections, and stability. I". In: *Mathematics of Computation* 64.211 (1995), pp. 1035–1065.

[129] P. Olsson. "Summation by parts, projections, and stability. II". In: *Mathematics of Computation* 64.212 (1995), pp. 1473–1493.

[130] S. Osher. "Systems of difference equations with general homogeneous boundary conditions". In: *Transactions of the American Mathematical Society* (1969), pp. 177–201.

[131] N. A. Patankar, P. Singh, D. D. Joseph, R. Glowinski, and T.-W. Pan. "A new formulation of the distributed Lagrange multiplier/fictitious domain method for particulate flows". In: *International Journal of Multiphase Flow* 26.9 (2000), pp. 1509–1524.

[132] R. B. Pember, J. B. Bell, P. Colella, W. Y. Curtchfield, and M. L. Welcome. "An adaptive Cartesian grid method for unsteady compressible flow in irregular regions". In: *Journal of computational Physics* 120.2 (1995), pp. 278–304.

[133] C. S. Peskin. "Flow patterns around heart valves: a numerical method". In: *Journal of computational physics* 10.2 (1972), pp. 252–271.

[134] C. S. Peskin. "The immersed boundary method". In: *Acta numerica* 11 (2002), pp. 479–517.

[135] Y. P. Popov and A. A. Samarskii. "Completely conservative difference schemes". In: *Zhurnal Vysshei Matematiki Matematicheskoi Fisiki 9* 4 (1969), pp. 953–958.

[136] M. A. Puscas, L. Monasse, A. Ern, C. Tenaud, C. Mariotti, and V. Daru. "A time semi-implicit scheme for the energy-balanced coupling of a shocked fluid flow with a deformable structure". In: *Journal of Computational Physics* 296 (2015), pp. 241–262.

[137] R. D. Richtmyer. "Proposed numerical method for calculation of shocks". In: *Los Alamos Report* 671 (1948).

[138] R. D. Richtmyer and K. W. Morton. *Difference methods for Initial Value problems.* Wiley-Interscience. 1967.

[139] C. W. Schulz-Rinne, J. P. Collins, and H. M. Glaz. "Numerical solution of the Riemann problem for two-dimensional gas dynamics". In: *SIAM Journal on Scientific Computing* 14.6 (1993), pp. 1394–1414.

[140] L. I. Sedov. "Propagation of strong shock waves". In: *Journal of Applied Mathematics and Mechanics* 10 (1946), pp. 241–250.

[141] D. Serre. *Systems of Conservation Laws 1: Hyperbolicity, entropies, shock waves.* Cambridge University Press, 1999.

[142] S. Shanmuganathan, D. Youngs, J. Griffond, B. Thornber, and R. Williams. "Accuracy of high-order density-based compressible methods in low Mach vortical flows". In: *International Journal for Numerical Methods in Fluids* 74.5 (2014), pp. 335–358.

[143] C.-W. Shu. "High-order finite difference and finite volume WENO schemes and discontinuous Galerkin methods for CFD". In: *International Journal of Computational Fluid Dynamics* 17.2 (2003), pp. 107–118.

[144] C.-W. Shu and S. Osher. "Efficient Implementation of Essentially Non-oscillatory Shock-capturing Schemes, II". In: *J. Comput. Phys.* 83 (1989), pp. 32–78.

[145] P. Singh, D. Joseph, T. Hesla, R. Glowinski, and T.-W. Pan. "A distributed Lagrange multiplier/fictitious domain method for viscoelastic particulate flows". In: *Journal of Non-Newtonian Fluid Mechanics* 91.2 (2000), pp. 165–188.

[146] G. A. Sod. "A Survey of Several Finite Difference Methods for Systems of Nonlinear Hyperbolic Conservation Laws". In: *J. Comput. Physics* 27 (1978), pp. 1–31.

[147] F. Sotiropoulos and X. Yang. "Immersed boundary methods for simulating fluid–structure interaction". In: *Progress in Aerospace Sciences* 65 (2014), pp. 1–21.

[148] G. Strang. "Trigonometric polynomials and difference methods of maximum accuracy". In: *Journal of Mathematics and Physics* 41.1 (1962), pp. 147–154.

[149] G. Strang. "On the construction and comparison of difference schemes". In: *SIAM Journal on Numerical Analysis* 5.3 (1968), pp. 506–517.

[150] J. C. Strikwerda. "Initial boundary value problems for the method of lines". In: *Journal of Computational Physics* 34.1 (1980), pp. 94–107.

[151] J. Strutt and L. Rayleigh. "Investigation of the character of the equilibrium of an incompressible heavy fluid of variable density". In: *Proc. London Math. Soc* 14.1 (1883), p. 8.

[152] A. Suresh and H. Huynh. "Accurate Monotonicity-Preserving Schemes with Runge-Kutta Stepping". In: *J. Comput. Physics* 136 (1997), pp. 83–99.

[153] W. G. Sutcliffe. "BBC Hydrodynamics". In: *Lawrence Livermore Laboratory Report* Technical Report UCID-17013 (1974).

[154] A. Szepessy. "Convergence of a streamline diffusion finite element method for scalar conservation laws with boundary conditions". In: *RAIRO-Modélisation mathématique et analyse numérique* 25.6 (1991), pp. 749–782.

[155] S. Tan and C.-W. Shu. "Inverse Lax-Wendroff procedure for numerical boundary conditions of conservation laws". In: *Journal of Computational Physics* 229.21 (2010), pp. 8144–8166.

[156] S. Tan and C.-W. Shu. "A high order moving boundary treatment for compressible inviscid flows". In: *Journal of Computational Physics* 230.15 (2011), pp. 6023–6036.

[157] S. Tan and C.-W. Shu. "Inverse Lax–Wendroff Procedure for Numerical Boundary Conditions of Hyperbolic Equations: Survey and New Developments". In: *Advances in Applied Mathematics, Modeling, and Computational Science.* Springer, 2013, pp. 41–63.

[158] S. Tan, C. Wang, C.-W. Shu, and J. Ning. "Efficient implementation of high order inverse lax–wendroff boundary treatment for conservation laws". In: *Journal of Computational Physics* 231.6 (2012), pp. 2510–2527.

[159] G. Taylor. "The instability of liquid surfaces when accelerated in a direction perpendicular to their planes. I". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences.* Vol. 201. 1065. The Royal Society. 1950, pp. 192–196.

[160] G. Taylor and A. Green. "Mechanism of the production of small eddies from large ones". In: *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 158.895 (1937), pp. 499–521.

[161] K. Thompson. "Time-dependent boundary conditions for hyperbolic systems, I". In: *J. Comput. Phys.* 68.1 (1987), pp. 1–24.

[162] K. Thompson. "Time-dependent boundary conditions for hyperbolic systems, II". In: *J. Comput. Phys.* 89.2 (1990), pp. 439–461.

[163] B. Thornber, D. Drikakis, and D. Youngs. "Large-eddy simulation of multi-component compressible turbulent flows using high resolution methods". In: *Computers & Fluids* 37.7 (2008), pp. 867–876.

[164] V. A. Titarev and E. F. Toro. "ADER: Arbitrary high order Godunov approach". In: *Journal of Scientific Computing* 17.1-4 (2002), pp. 609–618.

[165] J. G. Trulio and K. R. Trigger. "Numerical solution of the one dimensional Lagrangian hydrodynamics equations". In: *Lawrence Radiation Laboratory Report* Technical Report UCRL-6267 (1961).

[166] P. Tsoutsanis, I. W. Kokkinakis, L. Könözsy, D. Drikakis, R. J. Williams, and D. L. Youngs. "Comparison of structured-and unstructured-grid, compressible and incompressible methods using the vortex pairing problem". In: *Computer Methods in Applied Mechanics and Engineering* (2015).

[167] J. Verner. "Jim Verner's Refuge for Runge-Kutta Pairs". In: *http://people.math.sfu.ca/ jverner/* (2013).

[168] F. Vilar and C.-W. Shu. "Development and stability analysis of the inverse lax-wendroff boundary treatment for central compact schemes". In: *ESAIM: Mathematical Modelling and Numerical Analysis* (2014), Published–online.

[169] M. L. Wilkins. "Use of artificial viscosity in multidimensional fluid dynamic calculations". In: *Journal of computational physics* 36.3 (1980), pp. 281–303.

[170] M. Wolff. "Mathematical and numerical analysis of the resistive magnetohydrodynamics system with self-generated magnetic field terms". PhD thesis. Université de Strasbourg, 2011.

[171] P. Woodward and P. Colella. "The Numerical Simulation of Two-Dimensional Fluid Flow with Strong Shocks". In: *J. Comput. Physics* 54 (1984), pp. 115–173.

[172] L. Wu. "The semigroup stability of the difference approximations for initial-boundary value problems". In: *Mathematics of computation* 64.209 (1995), pp. 71–88.

[173] J. Yang and F. Stern. "A simple and efficient direct forcing immersed boundary framework for fluid–structure interactions". In: *Journal of Computational Physics* 231.15 (2012), pp. 5029–5061.

[174] H. Yee, N. Sandham, and M. Djomehri. "Low dissipative high-order shock-capturing methods using characteristics-based filters". In: *Journal of Computational Physics* 150 (1999), pp. 199–238.

[175] H. Yoshida. "Construction of higher order symplectic integrators". In: *Phys. Letters A* 150 (1990), pp. 262–267.

[176] D. L. Youngs. ""The Lagrangian Remap Method"". In: *Implicit Large Eddy Simulation: computing turbulent flow dynamics* Cambridge University Press (2007).

[177] X. Zeng and C. Farhat. "A systematic approach for constructing higher-order immersed boundary and ghost fluid methods for fluid–structure interaction problems". In: *Journal of Computational Physics* 231.7 (2012), pp. 2892–2923.