MOX-Report No. 74/2021

An efficient IMEX-DG solver for the compressible Navier-Stokes equations with a general equation of state

Orlando,G.; Barbante, P. F.; Bonaventura, L.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it http://mox.polimi.it

An efficient IMEX-DG solver for the compressible Navier-Stokes equations with a general equation of state

Giuseppe Orlando[(1)]
Paolo Francesco Barbante[(1)], Luca Bonaventura[(1)]

November 25, 2021

[(1)] MOX, Dipartimento di Matematica, Politecnico di Milano
Piazza Leonardo da Vinci 32, 20133 Milano, Italy
giuseppe.orlando@polimi.it, luca.bonaventura@polimi.it,
paolo.barbante@polimi.it

Keywords: Navier-Stokes equations, compressible flows, Discontinuous Galerkin methods, implicit methods, ESDIRK methods.

AMS Subject Classification: 65M08, 65N08, 65N12, 65Z05, 76R50

Abstract

We propose an efficient, accurate and robust IMEX solver for the compressible Navier-Stokes equation with general equation of state. The method, which is based on an $h-$adaptive Discontinuos Galerkin spatial discretization and on an Additive Runge Kutta IMEX method for time discretization, is tailored for low Mach number applications and allows to simulate low Mach regimes at a significantly reduced computational cost, while maintaining full second order accuracy also for higher Mach number regimes. The method has been implemented in the framework of the *deal.II* numerical library, whose adaptive mesh refinement capabilities are employed to enhance efficiency. Refinement indicators appropriate for real gas phenomena have been introduced. A number of numerical experiments on classical benchmarks for compressible flows and their extension to real gases demonstrate the properties of the proposed method.

1 Introduction

The efficient numerical solution of the compressible Navier-Stokes equations poses several major computational challenges. In particular, for flow regimes characterized by low Mach number and moderate Reynolds number values, severe time step restrictions may be required by standard explicit time discretization methods. The use of implicit and semi-implicit methods has a long tradition in low Mach number flows, see for example the seminal papers [11, 12, 36], as well as many other contributions in the literature on numerical weather prediction, see e.g. [6, 18, 20, 21, 29, 35, 37, 43, 42] and the reviews in [39, 8]. Other contributions have been proposed in the literature on more classical computational fluid dynamics, see e.g. [5, 4, 9, 10, 13, 31, 41]. Many of these contributions focus exclusively on the equations of motion of an ideal gas and their extension to real gases is not necessarily straightforward. Stability concerns are even more critical in these particular regimes for spatial discretizations based on the Discontinuous Galerkin (DG) method (see e.g. [19, 26] for a general presentation of this method), which is the spatial discretization used in many of the above referenced papers.

In this work, we seek to combine an accurate and flexible discontinuous DG space discretization with an implicit-explicit (IMEX) time discretization, see e.g. [27, 33], to obtain an efficient method for compressible flow of real gases at low to moderate Mach numbers. Our goal is to derive a method that can then be easily extended to handle multiphase compressible flows, where a number of coupling and forcing terms arise that cannot be dealt with efficiently by straightforward application of conventional solvers. In order to obtain a method that is robust in the low Mach number limit, following [11, 13], we couple implicitly the energy equation to the momentum equation, while treating the continuity equation in an explicit fashion.

Notice that a conceptually similar approach has been used in [29, 37] for the discretization employed in the IFS-FVM atmospheric model. In order to obtain a formulation that is efficient also in presence of non negligible viscous terms, we resort to an operator splitting approach, see e.g. [30]. More specifically, as commonly done in numerical models for atmospheric physics, we split the hyperbolic part of the problem, which is treated by an IMEX extension of the method proposed in [13], from the diffusive terms, which are treated implicitly. Second order accuracy can then be obtained by the Strang splitting approach [30, 40]. Notice that, with respect to the IMEX approach proposed for the Euler equations in [45], the technique presented here does not require to introduce reference solutions, does not introduce inconsistencies in the splitting with respect to a reference solution and only requires the solution of linear system of a size equal to that of the number of discrete degrees of freedom needed to describe a scalar variable, as in [13].

For the spatial discretization, we rely on the DG approach implemented in the numerical library *deal.II* [2], which is a very convenient environment to develop a reliable and easily accessible tool for large scale industrial applications. This software also provides $h-$refinement capabilities that are exploited by the proposed method. For the specific case of real gases, physically based refinement criteria have been developed and tested, which allow to track accurately convection phenomena also for general equations of state. The numerical experiments reported below show the ability of the proposed scheme and of its adaptive implementation to perform accurate simulations in different settings. The model equations and their non dimensional formulation are reviewed in Section 2. The time discretization approach is outlined and discussed in Section 3. The spatial discretization is presented in Section 4. Some implementation issues are described in Section 5, while the validation of the proposed method and its application to a number of significant benchmarks is reported in Section 6. Some conclusions and perspectives for future work are presented in Section 7.

2 The compressible Navier-Stokes equations

Let $\Omega \subset \mathbb{R}^d, 2 \leq d \leq 3$ be a connected open bounded set with a sufficiently smooth boundary $\partial\Omega$ and denote by $\mathbf{x}$ the spatial coordinates and by t the temporal coordinate. We consider the classical unsteady compressible

Navier-Stokes equations, written in flux form as:

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\mathbf{u}) &= 0 \\ \frac{\partial(\rho\mathbf{u})}{\partial t} + \nabla\cdot(\rho\mathbf{u}\otimes\mathbf{u}) + \nabla p &= \nabla\cdot\boldsymbol{\tau} + \rho\mathbf{f} \\ \frac{\partial(\rho E)}{\partial t} + \nabla\cdot[(\rho E + p)\mathbf{u}] &= \nabla\cdot(\boldsymbol{\tau}\mathbf{u} - \mathbf{q}) + \rho\mathbf{f}\cdot\mathbf{u} \end{aligned} \tag{1}$$

for $\mathbf{x} \in \Omega$, $t \in [0, T_f]$, supplied with suitable initial and boundary conditions. Here T_f is the final time, ρ is the density, $\mathbf{u}$ is the fluid velocity, p is the pressure, $\mathbf{q}$ denotes the heat flux and $\mathbf{f}$ represents volumetric forces. Notice that, at this stage, no more specific assumptions are made on the fluid. Possible choices of thermal and caloric equations of state will be specified in the following. ρE is the total energy, which can be rewritten as $\rho E = \rho e + \rho k$, where e is the internal energy and $k = \|\mathbf{u}\|^2/2$ is the kinetic energy. We also introduce the specific enthalpy $h = e + p/\rho$ and remark that one can also rewrite the energy flux as

$$(\rho E + p)\mathbf{u} = (e + k + \frac{p}{\rho})\rho\mathbf{u} = (h + k)\rho\mathbf{u}.$$

We assume that $\mathbf{q} = -\kappa\nabla T$, where T denotes the absolute temperature and κ the thermal conductivity. Furthermore, we assume that the linear stress constitutive equation holds and we neglect the bulk viscosity, so that

$$\boldsymbol{\tau} = \mu\left(\nabla\mathbf{u} + \nabla\mathbf{u}^T\right) - \frac{2\mu}{3}(\nabla\cdot\mathbf{u})\mathbf{I}.$$

The equations can then be rewritten as

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\mathbf{u}) &= 0 \\ \frac{\partial(\rho\mathbf{u})}{\partial t} + \nabla\cdot(\rho\mathbf{u}\otimes\mathbf{u}) + \nabla p &= \mu\,\nabla\cdot\left[\left(\nabla\mathbf{u} + \nabla\mathbf{u}^T\right) - \frac{2}{3}(\nabla\cdot\mathbf{u})\mathbf{I}\right] \\ &+ \rho\mathbf{f} \\ \frac{\partial(\rho E)}{\partial t} + \nabla\cdot[(h + k)\rho\mathbf{u}] &= \mu\,\nabla\cdot\left[\left(\nabla\mathbf{u} + \nabla\mathbf{u}^T\right)\mathbf{u} - \frac{2}{3}(\nabla\cdot\mathbf{u})\mathbf{u}\right] \\ &+ \kappa\Delta T + \rho\mathbf{f}\cdot\mathbf{u}. \end{aligned} \tag{2}$$

We now introduce reference scaling values $\mathcal{L}, \mathcal{T}, \mathcal{U}$ for the length, time and velocity, respectively, as well as reference values $\mathcal{P}, \mathcal{R}, \mathcal{TT}, \mathcal{E}, \mathcal{I}$ for pressure, density, temperature, total energy and internal energy, respectively. We assume unit Strouhal number $St = \mathcal{L}/\mathcal{UT} \approx 1$, that the enthalpy scales like $\mathcal{I} + \mathcal{P}/\mathcal{R}$ and that

$$\mathcal{E} = \mathcal{I} + \frac{\mathcal{P}}{\mathcal{R}} + \mathcal{U}^2.$$

The model equations can then be written in non-dimensional form as

$$\begin{aligned}
\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\mathbf{u}) &= 0 \\
\frac{\partial \rho\mathbf{u}}{\partial t} + \nabla\cdot(\rho\mathbf{u}\otimes\mathbf{u}) + \frac{\mathcal{P}}{\mathcal{R}\mathcal{U}^2}\nabla p &= \frac{\mu}{\mathcal{R}\mathcal{U}\mathcal{L}}\nabla\cdot\left[\left(\nabla\mathbf{u}+\nabla\mathbf{u}^T\right) - \frac{2}{3}(\nabla\cdot\mathbf{u})\mathbf{I}\right] \\
&+ \frac{\mathcal{T}}{\mathcal{U}}\rho\mathbf{f} \qquad (3)\\
\frac{\partial \rho E}{\partial t} + \nabla\cdot\left[\left(h\frac{\mathcal{I}+\mathcal{P}/\mathcal{R}}{\mathcal{E}} + k\frac{\mathcal{U}^2}{\mathcal{E}}\right)\rho\mathbf{u}\right] &= \frac{\mu\mathcal{U}}{\mathcal{R}\mathcal{E}\mathcal{L}}\nabla\cdot\left[\left(\nabla\mathbf{u}+\nabla\mathbf{u}^T\right)\mathbf{u} - \frac{2}{3}(\nabla\cdot\mathbf{u})\mathbf{u}\right] \\
&+ \frac{\kappa\mathcal{T}\mathcal{T}}{\mathcal{R}\mathcal{E}\mathcal{U}\mathcal{L}}\Delta T + \frac{\mathcal{L}}{\mathcal{E}}\rho\mathbf{f}\cdot\mathbf{u}.
\end{aligned}$$

We then define the Reynolds, Prandtl and Mach numbers as

$$Re = \frac{\mathcal{R}\mathcal{U}\mathcal{L}}{\mu} \quad \kappa = \frac{c_p\mu}{Pr} \quad Ma^2 = \frac{\mathcal{R}\mathcal{U}^2}{\mathcal{P}},$$

where c_p denotes the specific heat at constant pressure, so that, for low and moderate Mach numbers,

$$\begin{aligned}
\frac{\mathcal{U}^2}{\mathcal{E}} &= \frac{1}{\frac{\mathcal{I}}{\mathcal{U}^2}+1+\frac{1}{Ma^2}} = O(Ma^2) \\
\frac{\mathcal{I}+\mathcal{P}/\mathcal{R}}{\mathcal{E}} &= \frac{\frac{\mathcal{I}}{\mathcal{U}^2}+\frac{1}{Ma^2}}{\frac{\mathcal{I}}{\mathcal{U}^2}+1+\frac{1}{Ma^2}} = O(1). \qquad (4)
\end{aligned}$$

This justifies, in the above mentioned regimes, methods in which an implicit coupling between the pressure gradient and the energy flux is enforced. This strategy has been proposed in the seminal paper [11] and in the more recent works [13, 31]. We finally assume that the only acting volumetric force is gravity, so that $\mathbf{f} = -g\mathbf{k}$, where g denotes the acceleration of gravity and $\mathbf{k}$ the upward pointing unit vector in the standard Cartesian reference frame. It follows that

$$\begin{aligned}
\frac{\mathcal{T}}{\mathcal{U}}\rho g &= \frac{g\mathcal{T}\mathcal{U}}{\mathcal{U}^2}\rho = \frac{g\mathcal{L}}{\mathcal{U}^2}\rho = \frac{\rho}{Fr^2} \quad Fr^2 = \frac{\mathcal{U}^2}{g\mathcal{L}} \\
\frac{\mathcal{L}}{\mathcal{E}}\rho g &= \frac{g\rho\mathcal{L}}{\mathcal{I}+\frac{\mathcal{P}}{\mathcal{R}}+\mathcal{U}^2} = \frac{g\rho\mathcal{L}}{\mathcal{U}^2}\frac{1}{\frac{\mathcal{I}}{\mathcal{U}^2}+1+\frac{1}{Ma^2}} = \frac{\rho}{Fr^2}O(Ma^2). \qquad (5)
\end{aligned}$$

As a result, we will consider the non dimensional model equations

$$
\begin{aligned}
\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\mathbf{u}) &= 0 \\
\frac{\partial \rho\mathbf{u}}{\partial t} + \nabla\cdot(\rho\mathbf{u}\otimes\mathbf{u}) + \frac{1}{Ma^2}\nabla p &= \frac{1}{Re}\nabla\cdot\left[\left(\nabla\mathbf{u} + \nabla\mathbf{u}^T\right) - \frac{2}{3}(\nabla\cdot\mathbf{u})\mathbf{I}\right] \\
&- \frac{\rho}{Fr^2}\mathbf{k} \qquad (6)\\
\frac{\partial \rho E}{\partial t} + \nabla\cdot\left[\left(h + kMa^2\right)\rho\mathbf{u}\right] &= \frac{Ma^2}{Re}\nabla\cdot\left[\left(\nabla\mathbf{u} + \nabla\mathbf{u}^T\right)\mathbf{u} - \frac{2}{3}(\nabla\cdot\mathbf{u})\mathbf{u}\right] \\
&+ \frac{1}{PrRe}\Delta T - \rho\frac{Ma^2}{Fr^2}\mathbf{k}\cdot\mathbf{u},
\end{aligned}
$$

where we have taken $c_p\mathcal{TT}/\mathcal{E} \approx 1$, which can be justified at moderate values of the Mach number. Notice that these non dimensional equations are very similar to those derived in [31].

As previously remarked, the above equations must be complemented by an equation of state (EOS) for the compressible fluid. A classical choice is that of an ideal gas, whose EOS, in the non dimensional variables introduced above, is given by

$$
p = (\gamma - 1)\left(\rho E - \frac{1}{2}Ma^2\rho\mathbf{u}\cdot\mathbf{u}\right). \qquad (7)
$$

An example of non ideal gas model is given instead by van der Waals EOS, see e.g. [44]:

$$
p = \frac{\gamma - 1}{1 - \rho b}\left(\rho E - \frac{1}{2}Ma^2\rho\mathbf{u}\cdot\mathbf{u}\right) - \left(1 + \frac{1-\gamma}{1-\rho b}\right)a\rho^2 \qquad (8)
$$

where a and b are known as the van der Waals' constants. Notice that, for $a = b = 0$ the ideal gas law is retrieved. The caloric EOS for the ideal and van der Waals case reads

$$
e = \frac{R}{\gamma - 1}T - a\rho \qquad (9)
$$

where R is the specific gas constant. An important parameter to determine the regime in which real gas effects are relevant is the so-called compressibility factor $z = \frac{p}{\rho RT}$. When $z \approx 1$, the gas can be treated as an ideal one, while the ideal gas law is no longer valid for values of z very different from one. Another example is the Stiffened Gas equation of state (SG-EOS) [32], which is interesting for its convexity property and is given by

$$
p = (\gamma - 1)\left(\rho E - \frac{1}{2}Ma^2\rho\mathbf{u}\cdot\mathbf{u} - \rho q\right) - \gamma\pi \qquad (10)
$$

where q and π parameters that determine the gas characteristics. The caloric equation for the SG-EOS can be written as:

$$
e = \frac{(p + \gamma\pi)\,RT}{(\gamma - 1)\,(p + \pi)} + q. \qquad (11)
$$

3 The time discretization strategy

In the low Mach number limit, terms proportional to $1/Ma^2$ in equations (6) yield stiff components of the resulting semidiscretized ODE system. Therefore, following as remarked above [11, 13], it is appropriate to couple implicitly the energy equation to the momentum equation, while the continuity equation can be discretized in a fully explicit fashion. While this would be sufficient to yield an efficient time discretization approach for the purely hyperbolic system associated to (6) in absence of gravity, in regimes for which

$$Pr \approx O(1), \quad Fr << 1$$

thermal diffusivity and gravity terms would also have to be treated implicitly for the time discretization methods to be efficient. Straightforward application of any monolithic solver would then yield large algebraic systems with multiple couplings between discrete DOF associated to different physical variables. To avoid this, we resort to an operator splitting approach, see e.g. [30], as commonly done in numerical models for atmospheric physics, see e.g. the reviews in [8], [39]. More specifically, after spatial discretization, all diffusive terms on the right hand side of (6) are split from the hyperbolic part on the left hand side. The hyperbolic part is treated in a similar fashion to what outlined in [13], while the diffusive terms are treated implicitly. For simplicity, the gravity terms will be treated explictly in this first attempt and only a basic, first order splitting will be described, which can be easily improved to second order accuracy by the Strang splitting approach [30, 40].

For the time discretization, an IMplicit EXplicit (IMEX) Additive Runge Kutta method (ARK) [27] method will be used. These methods are useful for time dependent problems that can be formulated as $\mathbf{y}' = \mathbf{f}_S(\mathbf{y}, t) + \mathbf{f}_{NS}(\mathbf{y}, t)$, where the S and NS subscripts denote the stiff and non-stiff components of the system, to which the implicit and explicit companion methods are applied, respectively. If $\mathbf{v}^n \approx y(t^n)$, the generic $s-$ stage IMEX-ARK method can be defined as

$$\begin{aligned}\mathbf{v}^{(n,l)} = \mathbf{v}^n \quad &+ \quad \Delta t \sum_{m=1}^{s-1} \Big(a_{lm} \mathbf{f}_{NS}(\mathbf{v}^{(n,m)}, t + c_m \Delta t) \\ &+ \quad \tilde{a}_{lm} \mathbf{f}_S(\mathbf{v}^{(n,m)}, t + c_m \Delta t) \Big) + \Delta t \, \tilde{a}_{ll} \, \mathbf{f}_S(\mathbf{v}^{(n,l)}, t + c_l \Delta t).\end{aligned}$$

where $l = 1, \ldots, s$. After computation of the intermediate stages, $\mathbf{v}^{n+1}$ is computed as

$$\mathbf{v}^{n+1} = \mathbf{v}^n + \Delta t \sum_{l=1}^{s} b_l \left[\mathbf{f}_{NS}(\mathbf{v}^{(n,l)}, t + c_l \Delta t) + \mathbf{f}_S(\mathbf{v}^{(n,l)}, t + c_l \Delta t) \right].$$

Coefficients $a_{lm}, \tilde{a}_{lm}, c_l$ and b_l are determined so that the method is consistent of a given order. In particular, in addition to the order conditions

specific to each sub-method, the coefficients should respect coupling conditions. Here, we consider a variant of the IMEX method proposed in [20], whose coefficients are presented in the Butcher tableaux reported in Tables 1 and 2 for the explicit and implicit method, respectively, where $\gamma = 2 - \sqrt{2}$. The coefficients of the explicit method were proposed in [20], while the implicit method, also employed in the same paper, coincides indeed for the above choice of γ with the TR-BDF2 method proposed in [3, 24] and applied to the Euler equations in [43]. Notice that, even though we focus here on this specific second order method, the same strategy we outline is applicable to a generic DIRK method. In particular, higher order methods could be considered for coupling to high order spatial discretization, even though the effective overall accuracy would be limited by the splitting procedure if gravity and viscous terms are present.

0	0		
γ	γ	0	
1	$1-\alpha$	α	0
	$\frac{1}{2}-\frac{\gamma}{4}$	$\frac{1}{2}-\frac{\gamma}{4}$	$\frac{\gamma}{2}$

Table 1: *Butcher tableaux of the explicit ARK2 method*

0	0		
γ	$\frac{\gamma}{2}$	$\frac{\gamma}{2}$	
1	$\frac{1}{2\sqrt{2}}$	$\frac{1}{2\sqrt{2}}$	$1-\frac{1}{\sqrt{2}}$
	$\frac{1}{2}-\frac{\gamma}{4}$	$\frac{1}{2}-\frac{\gamma}{4}$	$\frac{\gamma}{2}$

Table 2: *Butcher tableaux of the implicit ARK2 method*

Notice that, as discussed in [20], the choice of the coefficients

$$\alpha = \frac{7-2\gamma}{6} \qquad 1-\alpha = \frac{2\gamma-1}{6}$$

in the third stage of the explicit part of the method is arbitrary. In [20], the above value of α was chosen with the aim of maximizing the stability region of the method. However, if a stability and absolute monotonicity analysis is carried out, as discussed in detail in Appendix A, it can be seen that different choices might be more advantageous, in order to improve the monotonicity of the method without compromising its stability. In particular, the value of $\alpha = 1/2$ appears to be a more appropriate choice, as also demonstrated by the numerical experiments reported in Section 6.

We now describe the application of this IMEX method and of the splitting approach outlined above to equations (6). Notice that, for simplicity, we first present the time semi-discretization only, while maintaining the continuous form of (6) with respect to the spatial variables. The detailed description of the algebraic problems resulting from the full space and time discretization according to the method outlined here will be presented in Section 4.

For each time step, we first consider the discretization of the hyperbolic and forcing terms. For the first stage of the method, one simply has

$$\rho^{(n,1)} = \rho^n \quad \mathbf{u}^{(n,1)} = \mathbf{u}^n \quad E^{(n,1)} = E^n.$$

For the second stage, we can write formally

$$\begin{aligned} \rho^{(n,2)} &= \rho^n - a_{21}\Delta t\,\nabla\cdot(\rho^n\mathbf{u}^n) \\ \rho^{(n,2)}\mathbf{u}^{(n,2)} &+ \tilde{a}_{22}\frac{\Delta t}{Ma^2}\nabla p^{(n,2)} = \mathbf{m}^{(n,2)} \\ \rho^{(n,2)}E^{(n,2)} &+ \tilde{a}_{22}\Delta t\,\nabla\cdot\left(h^{(n,2)}\rho^{(n,2)}\mathbf{u}^{(n,2)}\right) = \mathbf{e}^{(n,2)}, \end{aligned} \tag{12}$$

where we have set

$$\begin{aligned} \mathbf{m}^{(n,2)} &= \rho^n\mathbf{u}^n \\ &- a_{21}\Delta t\,\nabla\cdot(\rho^n\mathbf{u}^n\otimes\mathbf{u}^n) - \tilde{a}_{21}\frac{\Delta t}{Ma^2}\nabla p^n - a_{21}\frac{\Delta t}{Fr^2}\rho^n\mathbf{k} \end{aligned} \tag{13}$$

$$\begin{aligned} \mathbf{e}^{(n,2)} &= \rho^n E^n - \tilde{a}_{21}\Delta t\,\nabla\cdot(h^n\rho^n\mathbf{u}^n) - a_{21}\Delta t Ma^2\,\nabla\cdot(k^n\rho^n\mathbf{u}^n) \\ &- a_{21}\frac{\Delta t Ma^2}{Fr^2}\rho^n\mathbf{k}\cdot\mathbf{u}^n. \end{aligned}$$

Notice that, substituting formally $\rho^{(n,2)}\mathbf{u}^{(n,2)}$ into the energy equation and taking into account the definitions $\rho E = \rho e + \rho k$ and $h = e + p/\rho$, the above system can be solved by computing the solution of

$$\begin{aligned} &\rho^{(n,2)}[e(p^{(n,2)},\rho^{(n,2)}) + k^{(n,2)}] \\ -\ &\tilde{a}_{22}^2\frac{\Delta t^2}{Ma^2}\nabla\cdot\left[\left(e(p^{(n,2)},\rho^{(n,2)}) + \frac{p^{(n,2)}}{\rho^{(n,2)}}\right)\nabla p^{(n,2)}\right] \\ +\ &\tilde{a}_{22}\Delta t\,\nabla\cdot\left[\left(e(p^{(n,2)},\rho^{(n,2)}) + \frac{p^{(n,2)}}{\rho^{(n,2)}}\right)\mathbf{m}^{(n,2)}\right] = \mathbf{e}^{(n,2)} \end{aligned} \tag{14}$$

in terms of $p^{(n,2)}$ according to the fixed point procedure described in [13]. More specifically, setting $\xi^{(0)} = p^{(n,2)}, k^{(n,2,0)} = k^{(n,1)}$ one solves for $l = 1,\ldots,M$ the equation

$$
\begin{aligned}
\rho^{(n,2)} e(\xi^{(l+1)}, \rho^{(n,2)}) &- \tilde{a}_{22}^2 \frac{\Delta t^2}{Ma^2} \nabla\cdot \left[\left(e(\xi^{(l)}, \rho^{(n,2)}) + \frac{\xi^{(l)}}{\rho^{(n,2)}} \right) \nabla \xi^{(l+1)} \right] \\
&= \mathbf{e}^{(n,2)} - \rho^{(n,2)} k^{(n,2,l)} \\
&- \tilde{a}_{22} \Delta t \nabla\cdot \left[\left(e(\xi^{(l)}, \rho^{(n,2)}) + \frac{\xi^{(l)}}{\rho^{(n,2)}} \right) \mathbf{m}^{(n,2)} \right]
\end{aligned} \tag{15}
$$

and updates the velocity as

$$
\mathbf{u}^{(n,2,l+1)} + \frac{\tilde{a}_{22} \Delta t}{\rho^{(n,2)} Ma^2} \nabla \xi^{(l+1)} = \mathbf{m}^{(n,2)}.
$$

In case of a non-ideal gas equation of state, $\rho^{(n,2)} e\left(\xi^{(l+1)}, \rho^{(n,2)}\right)$ contains a term that only depends on the density, as evident from Equation (8) and from Equation (10). Therefore, it has to be considered on the right-hand side. For the sake of clarity, we report the fixed point equation for the van der Waals EOS:

$$
\begin{aligned}
\rho^{(n,2)} e(\xi^{(l+1)}, \rho^{(n,2)}) &- \tilde{a}_{22}^2 \frac{\Delta t^2}{Ma^2} \nabla\cdot \left[\left(e(\xi^{(l)}, \rho^{(n,2)}) + \frac{\xi^{(l)}}{\rho^{(n,2)}} \right) \nabla \xi^{(l+1)} \right] \\
&= \mathbf{e}^{(n,2)} - \rho^{(n,2)} k^{(n,2,l)} - \left(1 + \frac{1-\gamma}{1 - \rho^{(n,2)} b} \right) a(\rho^{(n,2)})^2 \\
&- \tilde{a}_{22} \Delta t \nabla\cdot \left[\left(e(\xi^{(l)}, \rho^{(n,2)}) + \frac{\xi^{(l)}}{\rho^{(n,2)}} \right) \mathbf{m}^{(n,2)} \right].
\end{aligned}
$$

The same considerations as in [13] apply concerning the favourable properties of the weakly nonlinear system resulting from the discrete form of (15). Once the iterations have been completed, one sets $\mathbf{u}^{(n,2)} = \mathbf{u}^{(n,2,M+1)}$ and $E^{(n,2)}$ accordingly. For the third stage, one can write formally

$$
\begin{aligned}
\rho^{(n,3)} &= \rho^n - a_{31} \Delta t \nabla\cdot (\rho^n \mathbf{u}^n) - a_{32} \Delta t \nabla\cdot \left(\rho^{(n,2)} \mathbf{u}^{(n,2)} \right) \\
\rho^{(n,3)} \mathbf{u}^{(n,3)} &+ \tilde{a}_{33} \frac{\Delta t}{Ma^2} \nabla p^{(n,3)} = \mathbf{m}^{(n,3)} \\
\rho^{(n,3)} E^{(n,3)} &+ \tilde{a}_{33} \Delta t \nabla\cdot \left(h^{(n,3)} \rho^{(n,3)} \mathbf{u}^{(n,3)} \right) = \mathbf{e}^{(n,3)},
\end{aligned} \tag{16}
$$

where the right hand sides are defined as

$$\begin{aligned}
\mathbf{m}^{(n,3)} &= \rho^n \mathbf{u}^n \\
&- a_{31}\Delta t\,\nabla\cdot(\rho^n\mathbf{u}^n\otimes\mathbf{u}^n) - \tilde{a}_{31}\frac{\Delta t}{Ma^2}\nabla p^n - a_{31}\frac{\Delta t}{Fr^2}\rho^n\mathbf{k} \\
&- a_{32}\Delta t\,\nabla\cdot\left(\rho^{(n,2)}\mathbf{u}^{(n,2)}\otimes\mathbf{u}^{(n,2)}\right) - \tilde{a}_{32}\frac{\Delta t}{Ma^2}\nabla p^{(n,2)} - a_{32}\frac{\Delta t}{Fr^2}\rho^{(n,2)}\mathbf{k}
\end{aligned} \tag{17}$$

$$\begin{aligned}
\mathbf{e}^{(n,3)} &= \rho^n E^n - \tilde{a}_{31}\Delta t\,\nabla\cdot(h^n\rho^n\mathbf{u}^n) - a_{31}\Delta t Ma^2\,\nabla\cdot(k^n\rho^n\mathbf{u}^n) \\
&- a_{31}\Delta t\frac{Ma^2}{Fr^2}\rho^n\mathbf{k}\cdot\mathbf{u}^n \\
&- \tilde{a}_{32}\Delta t\,\nabla\cdot\left(h^{(n,2)}\rho^{(n,2)}\mathbf{u}^{(n,2)}\right) - a_{32}\Delta t Ma^2\,\nabla\cdot\left(k^{(n,2)}\rho^{(n,2)}\mathbf{u}^{(n,2)}\right) \\
&- a_{32}\Delta t\frac{Ma^2}{Fr^2}\rho^{(n,2)}\mathbf{k}\cdot\mathbf{u}^{(n,2)}.
\end{aligned}$$

Again, the solution of this stage is computed by substituting formally $\rho^{(n,3)}\mathbf{u}^{(n,3)}$ into the energy equation and computing the solution of

$$\begin{aligned}
&\rho^{(n,3)}[e(p^{(n,3)},\rho^{(n,3)}) + k^{(n,3)}] \\
-\ &\tilde{a}_{33}^2\frac{\Delta t^2}{Ma^2}\,\nabla\cdot\left[\left(e(p^{(n,3)},\rho^{(n,3)}) + \frac{p^{(n,3)}}{\rho^{(n,3)}}\right)\nabla p^{(n,3)}\right] \\
+\ &\tilde{a}_{33}\Delta t\,\nabla\cdot\left[\left(e(p^{(n,3)},\rho^{(n,3)}) + \frac{p^{(n,3)}}{\rho^{(n,3)}}\right)\mathbf{m}^{(n,3)}\right] = \mathbf{e}^{(n,3)}.
\end{aligned} \tag{18}$$

More specifically, setting $\xi^{(0)} = p^{(n,3)}, k^{(n,3,0)} = k^{(n,2)}$ one solves for $l = 1,\ldots,M$ the equation

$$\begin{aligned}
\rho^{(n,3)}e(\xi^{(l+1)},\rho^{(n,3)}) &- \frac{\tilde{a}_{33}^2\Delta t^2}{Ma^2}\,\nabla\cdot\left[\left(e(\xi^{(l)},\rho^{(n,3)}) + \frac{\xi^{(l)}}{\rho^{(n,3)}}\right)\nabla\xi^{(l+1)}\right] \\
&= \mathbf{e}^{(n,3)} - \rho^{(n,3)}k^{(n,3,l)} \\
&- \tilde{a}_{33}\Delta t\,\nabla\cdot\left[\left(e(\xi^{(l)},\rho^{(n,3)}) + \frac{\xi^{(l)}}{\rho^{(n,3)}}\right)\mathbf{m}^{(n,3)}\right]
\end{aligned} \tag{19}$$

and updates the velocity as

$$\mathbf{u}^{(n,3,l+1)} + \tilde{a}_{33}\frac{\Delta t}{\rho^{(n,3)}Ma^2}\nabla\xi^{(l+1)} = \mathbf{m}^{(n,3)}.$$

Once again, in case of a non-ideal gas equation of state, the expression of $\rho^{(n,3)}e\left(\xi^{(l+1)},\rho^{(n,3)}\right)$ is slightly different. For the sake of clarity, we report

also for this stage the fixed point equation for the van der Waals EOS:

$$\begin{aligned}
\rho^{(n,3)} e(\xi^{(l+1)}, \rho^{(n,3)}) &- \frac{\tilde{a}_{33}^2 \Delta t^2}{Ma^2} \nabla\cdot \left[\left(e(\xi^{(l)}, \rho^{(n,3)}) + \frac{\xi^{(l)}}{\rho^{(n,3)}} \right) \nabla \xi^{(l+1)} \right] \\
&= \mathbf{e}^{(n,3)} - \rho^{(n,3)} k^{(n,3,l)} \\
&- \left(1 + \frac{1-\gamma}{1-\rho^{(n,3)} b} \right) a(\rho^{(n,3)})^2 \\
&- \tilde{a}_{33} \Delta t \, \nabla\cdot \left[\left(e(\xi^{(l)}, \rho^{(n,3)}) + \frac{\xi^{(l)}}{\rho^{(n,3)}} \right) \mathbf{m}^{(n,3)} \right]
\end{aligned} \tag{20}$$

Let us consider now the diffusive part of the Navier-Stokes equations that, as already mentioned in Section 1, will be treated with an operator splitting technique. For the sake of clarity, we denote with $\sim$ the quantities computed in this part of the scheme; hence, we define

$$\tilde{\mathbf{u}}^{(n,1)} = \mathbf{u}^{(n,3)} \qquad \tilde{E}^{(n,1)} = E^{(n,3)}$$

and we proceed to the discretization of the viscous terms, which is carried out by the implicit part of the IMEX method previously described:

$$\rho^{n+1} \tilde{\mathbf{u}}^{(n,2)} - \tilde{a}_{22} \frac{\Delta t}{Re} \nabla\cdot \left[\left(\nabla \tilde{\mathbf{u}} + \nabla \tilde{\mathbf{u}}^T \right) - \frac{2}{3} (\nabla \cdot \tilde{\mathbf{u}}) \mathbf{I} \right]^{(n,2)} = \tilde{\mathbf{m}}^{(n,2)} \tag{21}$$

$$\begin{aligned}
\rho^{n+1} \tilde{E}^{(n,2)} &- \tilde{a}_{22} \frac{\Delta t Ma^2}{Re} \nabla\cdot \left[\left(\nabla \tilde{\mathbf{u}} + \nabla \tilde{\mathbf{u}}^T \right) \tilde{\mathbf{u}} - \frac{2}{3} (\nabla \cdot \tilde{\mathbf{u}}) \tilde{\mathbf{u}} \right]^{(n,2)} \\
&- \tilde{a}_{22} \frac{\Delta t}{PrRe} \Delta \tilde{T}^{(n,2)} = \tilde{\mathbf{e}}^{(n,2)},
\end{aligned}$$

where we have set

$$\tilde{\mathbf{m}}^{(n,2)} = \rho^{n+1} \tilde{\mathbf{u}}^{(n,1)} + \tilde{a}_{21} \frac{\Delta t}{Re} \nabla\cdot \left[\left(\nabla \tilde{\mathbf{u}} + \nabla \tilde{\mathbf{u}}^T \right) - \frac{2}{3} (\nabla \cdot \tilde{\mathbf{u}}) \mathbf{I} \right]^{(n,1)}$$

$$\begin{aligned}
\tilde{\mathbf{e}}^{(n,2)} &= \rho^{n+1} \tilde{E}^{(n,1)} \\
&+ \tilde{a}_{21} \frac{\Delta t Ma^2}{Re} \nabla\cdot \left[\left(\nabla \mathbf{u} + \nabla \mathbf{u}^T \right) \mathbf{u} - \frac{2}{3} (\nabla \cdot \mathbf{u}) \mathbf{u} \right]^{(n,1)} \\
&+ \tilde{a}_{21} \frac{\Delta t}{PrRe} \Delta \tilde{T}^{(n,1)}.
\end{aligned}$$

Notice that the momentum equation in (21) is decoupled from the energy equation and can be solved independently, so that in a subsequent step the equation for $\tilde{E}^{(n,2)}$ can be solved using temperature as an unknown. For the third stage, one can write formally

$$
\begin{aligned}
\rho^{n+1}\tilde{\mathbf{u}}^{(n,3)} &- \tilde{a}_{33}\frac{\Delta t}{Re}\nabla\cdot\left[\left(\nabla\tilde{\mathbf{u}}+\nabla\tilde{\mathbf{u}}^T\right)-\frac{2}{3}(\nabla\cdot\tilde{\mathbf{u}})\mathbf{I}\right]^{(n,3)} = \tilde{\mathbf{m}}^{(n,3)}\\
\rho^{n+1}\tilde{E}^{(n,3)} &- \tilde{a}_{33}\frac{\Delta t Ma^2}{Re}\nabla\cdot\left[\left(\nabla\tilde{\mathbf{u}}+\nabla\tilde{\mathbf{u}}^T\right)\tilde{\mathbf{u}}-\frac{2}{3}(\nabla\cdot\tilde{\mathbf{u}})\tilde{\mathbf{u}}\right]^{(n,3)}\\
&- \tilde{a}_{33}\frac{\Delta t}{PrRe}\Delta\tilde{T}^{(n,3)} = \tilde{\mathbf{e}}^{(n,3)},
\end{aligned}
$$

where the right hand sides are defined as

$$
\begin{aligned}
\tilde{\mathbf{m}}^{(n,3)} &= \rho^{n+1}\tilde{\mathbf{u}}^{(n,1)} + \tilde{a}_{31}\frac{\Delta t}{Re}\nabla\cdot\left[\left(\nabla\tilde{\mathbf{u}}+\nabla\tilde{\mathbf{u}}^T\right)-\frac{2}{3}(\nabla\cdot\tilde{\mathbf{u}})\mathbf{I}\right]^{(n,1)}\\
&+ \tilde{a}_{32}\frac{\Delta t}{Re}\nabla\cdot\left[\left(\nabla\tilde{\mathbf{u}}+\nabla\tilde{\mathbf{u}}^T\right)-\frac{2}{3}(\nabla\cdot\tilde{\mathbf{u}})\mathbf{I}\right]^{(n,2)} \qquad (22)
\end{aligned}
$$

$$
\begin{aligned}
\tilde{\mathbf{e}}^{(n,3)} &= \rho^{n+1}\tilde{E}^{(n,1)} \qquad (23)\\
&+ \tilde{a}_{31}\frac{\Delta t Ma^2}{Re}\nabla\cdot\left[\left(\nabla\tilde{\mathbf{u}}+\nabla\tilde{\mathbf{u}}^T\right)\tilde{\mathbf{u}}-\frac{2}{3}(\nabla\cdot\tilde{\mathbf{u}})\tilde{\mathbf{u}}\right]^{(n,1)}\\
&+ \tilde{a}_{32}\frac{\Delta t Ma^2}{Re}\nabla\cdot\left[\left(\nabla\tilde{\mathbf{u}}+\nabla\tilde{\mathbf{u}}^T\right)\tilde{\mathbf{u}}-\frac{2}{3}(\nabla\cdot\tilde{\mathbf{u}})\tilde{\mathbf{u}}\right]^{(n,2)}\\
&+ \tilde{a}_{31}\frac{\Delta t}{PrRe}\Delta\tilde{T}^{(n,1)} + \tilde{a}_{32}\frac{\Delta t}{PrRe}\Delta\tilde{T}^{(n,2)}.
\end{aligned}
$$

Again, the momentum equation in (22) is decoupled from the energy equation and can be solved independently, so that in a subsequent step the equation for $E^{(n,3)}$ can be solved using temperature as an unkown. Finally, one sets

$$\mathbf{u}^{n+1} = \tilde{\mathbf{u}}^{(n,3)} \qquad E^{n+1} = \tilde{E}^{(n,3)}$$

and the computation of the $n-$th time step is completed.

4 The spatial discretization strategy

We consider a decomposition of the domain Ω into a family of hexahedra $\mathcal{T}_h$ (quadrilaterals in the two-dimensional case) and denote each element by K. The skeleton $\mathcal{E}$ denotes the set of all element faces and $\mathcal{E} = \mathcal{E}^I \cup \mathcal{E}^B$, where $\mathcal{E}^I$ is the subset of interior faces and $\mathcal{E}^B$ is the subset of boundary faces. Suitable jump and average operators can then be defined as customary for finite element discretizations. A face $\Gamma \in \mathcal{E}^I$ shares two elements that we denote by K^+ with outward unit normal $\mathbf{n}^+$ and K^- with outward unit normal $\mathbf{n}^-$, whereas for a face $\Gamma \in \mathcal{E}^B$ we denote by $\mathbf{n}$ the outward unit normal. For a scalar function φ the jump is defined as

$$[[\varphi]] = \varphi^+\mathbf{n}^+ + \varphi^-\mathbf{n}^- \quad \text{if } \Gamma \in \mathcal{E}^I \qquad [[\varphi]] = \varphi\mathbf{n} \quad \text{if } \Gamma \in \mathcal{E}^B.$$

The average is defined as

$$\{\{\varphi\}\} = \frac{1}{2}\left(\varphi^+ + \varphi^-\right) \quad \text{if } \Gamma \in \mathcal{E}^I \qquad \{\{\varphi\}\} = \varphi \quad \text{if } \Gamma \in \mathcal{E}^B.$$

Similar definitions apply for a vector function $\boldsymbol{\varphi}$:

$$[[\boldsymbol{\varphi}]] = \boldsymbol{\varphi}^+ \cdot \mathbf{n}^+ + \boldsymbol{\varphi}^- \cdot \mathbf{n}^- \quad \text{if } \Gamma \in \mathcal{E}^I \qquad [[\boldsymbol{\varphi}]] = \boldsymbol{\varphi} \cdot \mathbf{n} \quad \text{if } \Gamma \in \mathcal{E}^B$$

$$\{\{\boldsymbol{\varphi}\}\} = \frac{1}{2}\left(\boldsymbol{\varphi}^+ + \boldsymbol{\varphi}^-\right) \quad \text{if } \Gamma \in \mathcal{E}^I \qquad \{\{\boldsymbol{\varphi}\}\} = \boldsymbol{\varphi} \quad \text{if } \Gamma \in \mathcal{E}^B.$$

For vector functions, it is also useful to define a tensor jump as:

$$\langle\langle\boldsymbol{\varphi}\rangle\rangle = \boldsymbol{\varphi}^+ \otimes \mathbf{n}^+ + \boldsymbol{\varphi}^- \otimes \mathbf{n}^- \quad \text{if } \Gamma \in \mathcal{E}^I \qquad \langle\langle\boldsymbol{\varphi}\rangle\rangle = \boldsymbol{\varphi} \otimes \mathbf{n} \quad \text{if } \Gamma \in \mathcal{E}^B.$$

We also introduce the following finite element spaces

$$Q_k = \left\{v \in L^2(\Omega) : v|_K \in \mathbb{Q}_k \quad \forall K \in \mathcal{T}_h\right\} \quad \mathbf{V}_k = [Q_k]^d,$$

where $\mathbb{Q}_k$ is the space of polynomials of degree k in each coordinate direction. We then denote by $\boldsymbol{\varphi}_i(\mathbf{x})$ the basis functions for the space $\mathbf{V}_k$ and by $\psi_i(\mathbf{x})$ the basis functions for the space Q_k, the finite element spaces chosen for the discretization of the velocity and of the pressure (as well as the density), respectively.

$$\mathbf{u} \approx \sum_{j=1}^{\dim(\mathbf{V}_h)} u_j(t)\boldsymbol{\varphi}_j(\mathbf{x}) \qquad p \approx \sum_{j=1}^{\dim(Q_h)} p_j(t)\psi_j(\mathbf{x}).$$

Given these definitions, the weak formulation for the momentum equation of the second stage (12) reads as follows:

$$\begin{aligned}
& \sum_K \int_K \rho^{(n,2)}\mathbf{u}^{(n,2)} \cdot \mathbf{v} d\Omega \\
- & \sum_K \int_K \tilde{a}_{22}\frac{\Delta t}{Ma^2} p^{(n,2)} \nabla \cdot \mathbf{v} d\Omega + \sum_{\Gamma\in\mathcal{E}} \int_\Gamma \tilde{a}_{22}\frac{\Delta t}{Ma^2}\left\{\left\{p^{(n,2)}\right\}\right\}[[\mathbf{v}]]\, d\Sigma \\
= & \sum_K \int_K \rho^n \mathbf{u}^n \cdot \mathbf{v} d\Omega - \sum_K \int_K a_{21}\frac{\Delta t}{Fr^2}\rho^n \mathbf{k} \cdot \mathbf{v} d\Omega \\
+ & \sum_K \int_K a_{21}\Delta t\left(\rho^n \mathbf{u}^n \otimes \mathbf{u}^n\right) : \nabla \mathbf{v} d\Omega + \sum_K \int_K \tilde{a}_{21}\frac{\Delta t}{Ma^2} p^n \nabla \cdot \mathbf{v} d\Omega \\
- & \sum_{\Gamma\in\mathcal{E}} \int_\Gamma a_{21}\Delta t\left\{\{\rho^n \mathbf{u}^n \otimes \mathbf{u}^n\}\right\} : \langle\langle\mathbf{v}\rangle\rangle\, d\Sigma \\
- & \sum_{\Gamma\in\mathcal{E}} \int_\Gamma \tilde{a}_{21}\frac{\Delta t}{Ma^2}\left\{\{p^n\}\right\}[[\mathbf{v}]]\, d\Sigma \\
- & \sum_{\Gamma\in\mathcal{E}} \int_\Gamma a_{21}\Delta t\frac{\lambda^{(n,1)}}{2}\langle\langle\rho^n \mathbf{u}^n\rangle\rangle : \langle\langle\mathbf{v}\rangle\rangle\, d\Sigma
\end{aligned} \tag{24}$$

where

$$\lambda^{(n,1)} = \max\left(\left|\mathbf{u}^{n^+}\cdot\mathbf{n}^+\right|, \left|\mathbf{u}^{n^-}\cdot\mathbf{n}^-\right|\right)$$

In view of the implicit coupling between the momentum and the energy equations, we need to derive the algebraic formulation of (24) in order to formally substitute the degrees of freedom of the velocity into the algebraic formulation of the energy equation. We take $\mathbf{v} = \boldsymbol{\varphi}_i$, $i = 1...\dim(\mathbf{V}_h)$ and we exploit the representation introduced above to obtain

$$\begin{aligned}
& \sum_K \int_K \rho^{(n,2)} \sum_{j=1}^{\dim(\mathbf{V}_h)} u_j^{(n,2)} \boldsymbol{\varphi}_j\cdot\boldsymbol{\varphi}_i d\Omega - \sum_K \int_K \tilde{a}_{22}\frac{\Delta t}{Ma^2} \sum_{j=1}^{\dim(Q_h)} p_j^{(n,2)}\psi_j \nabla\cdot\boldsymbol{\varphi}_i d\Omega \\
+ & \sum_{\Gamma\in\mathcal{E}} \int_\Gamma \tilde{a}_{22}\frac{\Delta t}{Ma^2} \sum_{j=1}^{\dim(Q_h)} p_j^{(n,2)} \{\{\psi_j\}\} [[\boldsymbol{\varphi}_i]] \, d\Sigma \\
= & \sum_K \int_K \rho^n \mathbf{u}^n\cdot\boldsymbol{\varphi}_i d\Omega - \sum_K \int_K a_{21}\frac{\Delta t}{Fr^2}\rho^n \mathbf{k}\cdot\boldsymbol{\varphi}_i d\Omega \\
+ & \sum_K \int_K a_{21}\Delta t \left(\rho^n\mathbf{u}^n\otimes\mathbf{u}^n\right) : \nabla\boldsymbol{\varphi}_i d\Omega + \sum_K \int_K \tilde{a}_{21}\frac{\Delta t}{Ma^2} p^n \nabla\cdot\boldsymbol{\varphi}_i d\Omega \\
- & \sum_{\Gamma\in\mathcal{E}} \int_\Gamma a_{21}\Delta t \{\{\rho^n\mathbf{u}^n\otimes\mathbf{u}^n\}\} : \langle\langle\boldsymbol{\varphi}_i\rangle\rangle \, d\Sigma \\
- & \sum_{\Gamma\in\mathcal{E}} \int_\Gamma \tilde{a}_{21}\frac{\Delta t}{Ma^2} \{\{p^n\}\} [[\boldsymbol{\varphi}_i]] \, d\Sigma \\
- & \sum_{\Gamma\in\mathcal{E}} \int_\Gamma a_{21}\Delta t \frac{\lambda^{(n,1)}}{2} \langle\langle\rho^n\mathbf{u}^n\rangle\rangle : \langle\langle\boldsymbol{\varphi}_i\rangle\rangle \, d\Sigma
\end{aligned} \tag{25}$$

which can be written in compact form as

$$\mathbf{A}^{(n,2)}\mathbf{U}^{(n,2)} + \mathbf{B}^{(n,2)}\mathbf{P}^{(n,2)} = \mathbf{F}^{(n,2)} \tag{26}$$

where we have set

$$A_{ij}^{(n,2)} = \sum_K \int_K \rho^{(n,2)} \boldsymbol{\varphi}_j\cdot\boldsymbol{\varphi}_i d\Omega \tag{27}$$

$$B_{ij}^{(n,2)} = \sum_K \int_K -\tilde{a}_{22}\frac{\Delta t}{Ma^2}\nabla\cdot\boldsymbol{\varphi}_i\psi_j d\Omega + \sum_{\Gamma\in\mathcal{E}} \int_\Gamma \tilde{a}_{22}\frac{\Delta t}{Ma^2} \{\{\psi_j\}\} [[\boldsymbol{\varphi}_i]] \, d\Sigma \tag{28}$$

with $\mathbf{U}^{(n,2)}$ denoting the vector of the degrees of freedom associated to the velocity field and $\mathbf{P}^{(n,2)}$ denoting the vector of the degrees of freedom associated to the pressure. Consider now the weak formulation for the energy equation of the second stage (12)

$$
\begin{aligned}
& \sum_K \int_K \rho^{(n,2)} E^{(n,2)} w d\Omega - \sum_K \int_K \tilde{a}_{22} \Delta t h^{(n,2)} \rho^{(n,2)} \mathbf{u}^{(n,2)} \cdot \nabla w d\Omega \\
+ & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma \tilde{a}_{22} \Delta t \left\{ \left\{ h^{(n,2)} \rho^{(n,2)} \mathbf{u}^{(n,2)} \right\} \right\} \cdot [[w]] \, d\Sigma \\
= & \sum_K \int_K \rho^n E^n w d\Omega - \sum_K \int_K a_{21} \frac{\Delta t Ma^2}{Fr^2} \rho^n \mathbf{k} \cdot \mathbf{u}^n w d\Omega \\
+ & \sum_K \int_K a_{21} \Delta t Ma^2 \left(k^n \rho^n \mathbf{u}^n \right) \cdot \nabla w d\Omega + \sum_K \int_K \tilde{a}_{21} \Delta t \left(h^n \rho^n \mathbf{u}^n \right) \cdot \nabla w d\Omega \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma a_{21} \Delta t Ma^2 \left\{ \{ k^n \rho^n \mathbf{u}^n \} \right\} \cdot [[w]] \, d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}^I} \int_\Gamma \tilde{a}_{21} \Delta t \left\{ \{ h^n \rho^n \mathbf{u}^n \} \right\} \cdot [[w]] \, d\Sigma
\end{aligned} \tag{29}
$$

Take $w = \psi_i$ and consider the expansion for $\mathbf{u}^{(n,2)}$ in (29) to get

$$
\begin{aligned}
& \sum_K \int_K \rho^{(n,2)} E^{(n,2)} \psi_i d\Omega - \sum_K \int_K \tilde{a}_{22} \Delta t h^{(n,2)} \rho^{(n,2)} \sum_{j=1}^{\dim(\mathbf{V}_h)} u_j^{(n,2)} \boldsymbol{\varphi}_j \cdot \nabla \psi_i d\Omega \\
+ & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma \tilde{a}_{22} \Delta t \sum_{j=1}^{\dim(\mathbf{V}_h)} u_j^{(n,2)} \left\{ \left\{ h^{(n,2)} \rho^{(n,2)} \boldsymbol{\varphi}_j \right\} \right\} \cdot [[\psi_i]] \, d\Sigma \\
= & \sum_K \int_K \rho^n E^n \psi_i d\Omega - \sum_K \int_K a_{21} \frac{\Delta t Ma^2}{Fr^2} \rho^n \mathbf{k} \cdot \mathbf{u}^n \psi_i d\Omega \\
+ & \sum_K \int_K a_{21} \Delta t Ma^2 \left(k^n \rho^n \mathbf{u}^n \right) \cdot \nabla \psi_i d\Omega + \sum_K \int_K \tilde{a}_{21} \Delta t \left(h^n \rho^n \mathbf{u}^n \right) \cdot \nabla \psi_i d\Omega \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma a_{21} \Delta t Ma^2 \left\{ \{ k^n \rho^n \mathbf{u}^n \} \right\} \cdot [[\psi_i]] \, d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma \tilde{a}_{21} \Delta t \left\{ \{ h^n \rho^n \mathbf{u}^n \} \right\} \cdot [[\psi_i]] \, d\Sigma
\end{aligned} \tag{30}
$$

which can be expressed in compact form as

$$
\mathbf{C}^{(n,2)} \mathbf{U}^{(n,2)} = \mathbf{G}^{(n,2)} \tag{31}
$$

where we have set

$$C_{ij}^{(n,2)} = \sum_K \int_K -\tilde{a}_{22}\Delta t h^{(n,2)}\rho^{(n,2)}\boldsymbol{\varphi}_j \cdot \nabla\psi_i d\Omega + \sum_{\Gamma\in\mathcal{E}} \int_\Gamma \tilde{a}_{22}\Delta t \left\{\left\{h^{(n,2)}\rho^{(n,2)}\boldsymbol{\varphi}_j\right\}\right\} \cdot [[\psi_i]]\, d\Sigma \tag{32}$$

Formally we can then derive $\mathbf{U}^{(n,2)} = (\mathbf{A}^{(n,2)})^{-1}\left(\mathbf{F}^{(n,2)} - \mathbf{B}^{(n,2)}\mathbf{P}^{(n,2)}\right)$ and obtain the following relation

$$\mathbf{C}^{(n,2)}(\mathbf{A}^{(n,2)})^{-1}\left(\mathbf{F}^{(n,2)} - \mathbf{B}^{(n,2)}\mathbf{P}^{(n,2)}\right) = \mathbf{G}^{(n,2)} \tag{33}$$

Taking into account that

$$\rho^{(n,2)}E^{(n,2)} = \rho^{(n,2)}e^{(n,2)}(p^{(n,2)}) + Ma^2\rho^{(n,2)}k^{(n,2)},$$

we finally obtain

$$\mathbf{C}^{(n,2)}(\mathbf{A}^{(n,2)})^{-1}\left(\mathbf{F}^{(n,2)} - \mathbf{B}^{(n,2)}\mathbf{P}^{(n,2)}\right) = -\mathbf{D}^{(n,2)}\mathbf{P}^{(n,2)} + \tilde{\mathbf{G}}^{(n,2)} \tag{34}$$

where we have set

$$D_{ij}^{(n,2)} = \sum_K \int_K \rho^{(n,2)}e^{(n,2)}(\psi_j)\psi_i d\Omega \tag{35}$$

and $\tilde{\mathbf{G}}^{(n,2)}$ takes into account all the other terms (the one at previous stage and the kinetic energy). The above system can be solved in terms of $\mathbf{P}^{(n,2)}$ according to the fixed point procedure described in [13]. More specifically, setting $\mathbf{P}^{(n,2,0)} = \mathbf{P}^{(n,1)}, k^{(n,2,0)} = k^{(n,1)}$, for $l = 1, \ldots, M$ one solves the equation

$$\left(\mathbf{D}^{(n,2,l)} - \mathbf{C}^{(n,2,l)}(\mathbf{A}^{(n,2)})^{-1}\mathbf{B}^{(n,2)}\right)\mathbf{P}^{(n,2,l+1)} = \tilde{\mathbf{G}}^{(n,2,l)} - \mathbf{C}^{(n,2,l)}(\mathbf{A}^{(n,2)})^{-1}\mathbf{F}^{(n,2,l)}$$

and updates the velocity solving

$$\mathbf{A}^{(n,2)}\mathbf{U}^{(n,2,l+1)} = \mathbf{F}^{(n,2,l)} - \mathbf{B}^{(n,2)}\mathbf{P}^{(n,2,l+1)}.$$

For the third stage, we proceed in a similar manner. We start with the weak formulation of the momentum equation in (16):

$$
\begin{aligned}
& \sum_K \int_K \rho^{(n,3)} \mathbf{u}^{(n,3)} \cdot \mathbf{v} d\Omega \\
- & \sum_K \int_K \tilde{a}_{33} \frac{\Delta t}{Ma^2} p^{(n,3)} \nabla \cdot \mathbf{v} d\Omega + \sum_{\Gamma \in \mathcal{E}} \int_\Gamma \tilde{a}_{33} \frac{\Delta t}{Ma^2} \left\{ \left\{ p^{(n,3)} \right\} \right\} [[\mathbf{v}]] \, d\Sigma \\
= & \sum_K \int_K \rho^n \mathbf{u}^n \cdot \mathbf{v} d\Omega - \sum_K \int_K a_{31} \frac{\Delta t}{Fr^2} \rho^n \mathbf{k} \cdot \mathbf{v} d\Omega - \sum_K \int_K a_{32} \frac{\Delta t}{Fr^2} \rho^{(n,2)} \mathbf{k} \cdot \mathbf{v} d\Omega \\
+ & \sum_K \int_K a_{31} \Delta t \left(\rho^n \mathbf{u}^n \otimes \mathbf{u}^n \right) : \nabla \mathbf{v} d\Omega + \sum_K \int_K \tilde{a}_{31} \frac{\Delta t}{Ma^2} p^n \nabla \cdot \mathbf{v} d\Omega \\
+ & \sum_K \int_K a_{32} \Delta t \left(\rho^{(n,2)} \mathbf{u}^{(n,2)} \otimes \mathbf{u}^{(n,2)} \right) : \nabla \mathbf{v} d\Omega + \sum_K \int_K \tilde{a}_{32} \frac{\Delta t}{Ma^2} p^{(n,2)} \nabla \cdot \mathbf{v} d\Omega \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma a_{31} \Delta t \left\{ \{ \rho^n \mathbf{u}^n \otimes \mathbf{u}^n \} \right\} : \langle \langle \mathbf{v} \rangle \rangle \, d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma \tilde{a}_{31} \frac{\Delta t}{Ma^2} \left\{ \{ p^n \} \right\} [[\mathbf{v}]] \, d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma a_{32} \Delta t \left\{ \left\{ \rho^{(n,2)} \mathbf{u}^{(n,2)} \otimes \mathbf{u}^{(n,2)} \right\} \right\} : \langle \langle \mathbf{v} \rangle \rangle \, d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma \tilde{a}_{32} \frac{\Delta t}{Ma^2} \left\{ \left\{ p^{(n,2)} \right\} \right\} [[\mathbf{v}]] \, d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma a_{31} \Delta t \frac{\lambda^{(n,1)}}{2} \langle \langle \rho^n \mathbf{u}^n \rangle \rangle : \langle \langle \mathbf{v} \rangle \rangle \, d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma a_{32} \Delta t \frac{\lambda^{(n,2)}}{2} \left\langle \left\langle \rho^{(n,2)} \mathbf{u}^{(n,2)} \right\rangle \right\rangle : \langle \langle \mathbf{v} \rangle \rangle \, d\Sigma
\end{aligned} \tag{36}
$$

where

$$
\lambda^{(n,2)} = \max \left(\left| \mathbf{u}^{(n,2)^+} \cdot \mathbf{n}^+ \right|, \left| \mathbf{u}^{(n,2)^-} \cdot \mathbf{n}^- \right| \right)
$$

Now, taking $\mathbf{v} = \boldsymbol{\varphi}_i$ and exploiting the representation of $\mathbf{u}^{(n,3)}$ and $p^{(n,3)}$, we end up with the following relation

$$
\begin{aligned}
& \sum_K \int_K \rho^{(n,3)} \sum_{j=1}^{\dim(\mathbf{V}_h)} u_j^{(n,3)} \boldsymbol{\varphi}_j \cdot \boldsymbol{\varphi}_i d\Omega - \sum_K \int_K \tilde{a}_{33} \frac{\Delta t}{Ma^2} \sum_{j=1}^{\dim(Q_h)} p_j^{(n,3)} \psi_j \nabla \cdot \boldsymbol{\varphi}_i d\Omega \\
+ & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma \tilde{a}_{33} \frac{\Delta t}{Ma^2} \sum_{j=1}^{\dim(Q_h)} p_j^{(n,3)} \left\{\{\psi_j\}\right\} \left[[\boldsymbol{\varphi}_i]\right] d\Sigma \\
= & \sum_K \int_K \rho^n \mathbf{u}^n \cdot \boldsymbol{\varphi}_i d\Omega - \sum_K \int_K a_{31} \frac{\Delta t}{Fr^2} \rho^n \mathbf{k} \cdot \boldsymbol{\varphi}_i d\Omega - \sum_K \int_K a_{32} \frac{\Delta t}{Fr^2} \rho^{(n,2)} \mathbf{k} \cdot \boldsymbol{\varphi}_i d\Omega \\
+ & \sum_K \int_K a_{31} \Delta t \left(\rho^n \mathbf{u}^n \otimes \mathbf{u}^n\right) : \nabla \boldsymbol{\varphi}_i d\Omega + \sum_K \int_K \tilde{a}_{31} \frac{\Delta t}{Ma^2} p^n \nabla \cdot \boldsymbol{\varphi}_i d\Omega \\
+ & \sum_K \int_K a_{32} \Delta t \left(\rho^{(n,2)} \mathbf{u}^{(n,2)} \otimes \mathbf{u}^{(n,2)}\right) : \nabla \boldsymbol{\varphi}_i d\Omega + \sum_K \int_K \tilde{a}_{32} \frac{\Delta t}{Ma^2} p^{(n,2)} \nabla \cdot \boldsymbol{\varphi}_i d\Omega \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma a_{31} \Delta t \left\{\{\rho^n \mathbf{u}^n \otimes \mathbf{u}^n\}\right\} : \left\langle\langle \boldsymbol{\varphi}_i \rangle\right\rangle d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma \tilde{a}_{31} \frac{\Delta t}{Ma^2} \left\{\{p^n\}\right\} \left[[\boldsymbol{\varphi}_i]\right] d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma a_{32} \Delta t \left\{\left\{\rho^{(n,2)} \mathbf{u}^{(n,2)} \otimes \mathbf{u}^{(n,2)}\right\}\right\} : \left\langle\langle \boldsymbol{\varphi}_i \rangle\right\rangle d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma \tilde{a}_{32} \frac{\Delta t}{Ma^2} \left\{\left\{p^{(n,2)}\right\}\right\} \left[[\boldsymbol{\varphi}_i]\right] d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma a_{31} \Delta t \frac{\lambda^{(n,1)}}{2} \left\langle\langle \rho^n \mathbf{u}^n \right\rangle\rangle : \left\langle\langle \boldsymbol{\varphi}_i \right\rangle\rangle d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma a_{32} \Delta t \frac{\lambda^{(n,2)}}{2} \left\langle\left\langle \rho^{(n,2)} \mathbf{u}^{(n,2)} \right\rangle\right\rangle : \left\langle\langle \boldsymbol{\varphi}_i \right\rangle\rangle d\Sigma
\end{aligned} \tag{37}
$$

which can be written in compact form as

$$
\mathbf{A}^{(n,3)} \mathbf{U}^{(n,3)} + \mathbf{B}^{(n,3)} \mathbf{P}^{(n,3)} = \mathbf{F}^{(n,3)} \tag{38}
$$

where we have set

$$
A_{ij}^{(n,3)} = \sum_K \int_K \rho^{(n,3)} \boldsymbol{\varphi}_j \cdot \boldsymbol{\varphi}_i d\Omega \tag{39}
$$

$$
\begin{aligned}
B_{ij}^{(n,3)} = & \sum_K \int_K -\tilde{a}_{33} \frac{\Delta t}{Ma^2} \nabla \cdot \boldsymbol{\varphi}_i \psi_j d\Omega \\
& + \sum_{\Gamma \in \mathcal{E}} \int_\Gamma \tilde{a}_{33} \frac{\Delta t}{Ma^2} \left\{\{\psi_j\}\right\} \left[[\boldsymbol{\varphi}_i]\right] d\Sigma
\end{aligned} \tag{40}
$$

and $\mathbf{U}^{(n,3)}$ denotes the vector of the degrees of freedom associated to the velocity field, whereas $\mathbf{P}^{(n,3)}$ denotes the vector of the degrees of freedom associated to the pressure. Consider now the weak formulation for the energy equation in (16)

$$
\begin{aligned}
& \sum_K \int_K \rho^{(n,3)} E^{(n,3)} w d\Omega - \sum_K \int_K \tilde{a}_{33} \Delta t h^{(n,3)} \rho^{(n,3)} \mathbf{u}^{(n,3)} \cdot \nabla w d\Omega \\
+ & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma \tilde{a}_{33} \Delta t \left\{ \left\{ h^{(n,3)} \rho^{(n,3)} \mathbf{u}^{(n,3)} \right\} \right\} \cdot [[w]] \, d\Sigma \\
= & \sum_K \int_K \rho^n E^n w d\Omega - \sum_K \int_K a_{31} \frac{\Delta t M a^2}{Fr^2} \rho^n \mathbf{k} \cdot \mathbf{u}^n w d\Omega - \sum_K \int_K a_{32} \frac{\Delta t M a^2}{Fr^2} \rho^{(n,2)} \mathbf{k} \cdot \mathbf{u}^{(n,2)} w d\Omega \\
+ & \sum_K \int_K a_{31} \Delta t M a^2 \left(k^n \rho^n \mathbf{u}^n \right) \cdot \nabla w d\Omega + \sum_K \int_K \tilde{a}_{31} \Delta t \left(h^n \rho^n \mathbf{u}^n \right) \cdot \nabla w d\Omega \\
+ & \sum_K \int_K a_{32} \Delta t M a^2 \left(k^{(n,2)} \rho^{(n,2)} \mathbf{u}^{(n,2)} \right) \cdot \nabla w d\Omega + \sum_K \int_K \tilde{a}_{32} \Delta t \left(h^{(n,2)} \rho^{(n,2)} \mathbf{u}^{(n,2)} \right) \cdot \nabla w d\Omega \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma a_{31} \Delta t M a^2 \left\{ \{ k^n \rho^n \mathbf{u}^n \} \right\} \cdot [[w]] \, d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma \tilde{a}_{31} \Delta t \left\{ \{ h^n \rho^n \mathbf{u}^n \} \right\} \cdot [[w]] \, d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma a_{32} \Delta t M a^2 \left\{ \left\{ k^{(n,2)} \rho^{(n,2)} \mathbf{u}^{(n,2)} \right\} \right\} \cdot [[w]] \, d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma \tilde{a}_{32} \Delta t \left\{ \left\{ h^{(n,2)} \rho^{(n,2)} \mathbf{u}^{(n,2)} \right\} \right\} \cdot [[w]] \, d\Sigma \qquad (41)
\end{aligned}
$$

Take now $w = \psi_i$ and consider the expansion for $\mathbf{u}^{(n,3)}$

$$
\begin{aligned}
& \sum_K \int_K \rho^{(n,3)} E^{(n,3)} \psi_i d\Omega - \sum_K \int_K \tilde{a}_{33} \Delta t h^{(n,3)} \rho^{(n,3)} \sum_{j=1}^{\dim(\mathbf{V}_h)} u_j^{(n,3)} \boldsymbol{\varphi}_j \cdot \nabla \psi_i d\Omega \\
+ & \sum_{\Gamma \in \mathcal{E}^I} \int_\Gamma \tilde{a}_{33} \Delta t \sum_{j=1}^{\dim(\mathbf{V}_h)} u_j^{(n,3)} \left\{ \left\{ h^{(n,3)} \rho^{(n,3)} \boldsymbol{\varphi}_j \right\} \right\} \cdot [[\psi_i]] \, d\Sigma \\
= & \sum_K \int_K \rho^n E^n \psi_i d\Omega - \sum_K \int_K a_{31} \frac{\Delta t M a^2}{Fr^2} \rho^n \mathbf{k} \cdot \mathbf{u}^n \psi_i d\Omega - \sum_K \int_K a_{32} \frac{\Delta t M a^2}{Fr^2} \rho^{(n,2)} \mathbf{k} \cdot \mathbf{u}^{(n,2)} \psi_i d\Omega \\
+ & \sum_K \int_K a_{31} \Delta t M a^2 \left(k^n \rho^n \mathbf{u}^n \right) \cdot \nabla \psi_i d\Omega + \sum_K \int_K \tilde{a}_{31} \Delta t \left(h^n \rho^n \mathbf{u}^n \right) \cdot \nabla \psi_i d\Omega \\
+ & \sum_K \int_K a_{32} \Delta t M a^2 \left(k^{(n,2)} \rho^{(n,2)} \mathbf{u}^{(n,2)} \right) \cdot \nabla \psi_i d\Omega + \sum_K \int_K \tilde{a}_{32} \Delta t \left(h^{(n,2)} \rho^{(n,2)} \mathbf{u}^{(n,2)} \right) \cdot \nabla \psi_i d\Omega \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma a_{31} \Delta t M a^2 \left\{ \{ k^n \rho^n \mathbf{u}^n \} \right\} \cdot [[\psi_i]] \, d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma \tilde{a}_{31} \Delta t \left\{ \{ h^n \rho^n \mathbf{u}^n \} \right\} \cdot [[\psi_i]] \, d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma a_{32} \Delta t M a^2 \left\{ \left\{ k^{(n,2)} \rho^{(n,2)} \mathbf{u}^{(n,2)} \right\} \right\} \cdot [[\psi_i]] \, d\Sigma \\
- & \sum_{\Gamma \in \mathcal{E}} \int_\Gamma \tilde{a}_{32} \Delta t \left\{ \left\{ h^{(n,2)} \rho^{(n,2)} \mathbf{u}^{(n,2)} \right\} \right\} \cdot [[\psi_i]] \, d\Sigma
\end{aligned} \tag{42}
$$

which can be expressed in compact form as

$$
\mathbf{C}^{(n,3)} \mathbf{U}^{(n,3)} = \mathbf{G}^{(n,3)} \tag{43}
$$

where

$$
\begin{aligned}
C_{ij}^{(n,3)} = & \sum_K \int_K -\tilde{a}_{33} \Delta t h^{(n,3)} \rho^{(n,3)} \boldsymbol{\varphi}_j \cdot \nabla \psi_i d\Omega \\
& + \sum_{\Gamma \in \mathcal{E}} \int_\Gamma \tilde{a}_{33} \Delta t \left\{ \left\{ h^{(n,3)} \rho^{(n,3)} \boldsymbol{\varphi}_j \right\} \right\} \cdot [[\psi_i]] \, d\Sigma
\end{aligned} \tag{44}
$$

Formally, one can derive $\mathbf{U}^{(n,3)} = (\mathbf{A}^{(n,3)})^{-1} \left(\mathbf{F}^{(n,3)} - \mathbf{B}^{(n,3)} \mathbf{P}^{(n,3)} \right)$ and obtain the following relation

$$
\mathbf{C}^{(n,3)} (\mathbf{A}^{(n,3)})^{-1} \left(\mathbf{F}^{(n,3)} - \mathbf{B}^{(n,3)} \mathbf{P}^{(n,3)} \right) = \mathbf{G}^{(n,3)} \tag{45}
$$

Taking into account that

$$
\rho^{(n,3)} E^{(n,3)} = \rho^{(n,3)} e^{(n,3)} + Ma^2 \rho^{(n,3)} k^{(n,3)},
$$

we obtain

$$\mathbf{C}^{(n,3)}(\mathbf{A}^{(n,3)})^{-1}\left(\mathbf{F}^{(n,3)}-\mathbf{B}^{(n,3)}\mathbf{P}^{(n,3)}\right)=-\mathbf{D}^{(n,3)}\mathbf{P}^{(n,3)}+\tilde{\mathbf{G}}^{(n,3)} \tag{46}$$

where

$$D_{ij}^{(n,3)}=\sum_K\int_K\rho^{(n,3)}e^{(n,3)}(\psi_j)\psi_i d\Omega \tag{47}$$

and $\tilde{\mathbf{G}}^{(n,3)}$ takes into account all the other terms (the one at previous stage and the kinetic energy). Again, the above system is solved by a fixed point procedure. More specifically, setting $\mathbf{P}^{(n,3,0)}=\mathbf{P}^{(n,3)}, k^{(n,3,0)}=k^{(n,2)}$ for $l=1,\dots,M$ one solves the equation

$$\left(\mathbf{D}^{(n,3,l)}-\mathbf{C}^{(n,3,l)}(\mathbf{A}^{(n,3)})^{-1}\mathbf{B}^{(n,3)}\right)\mathbf{P}^{(n,3,l+1)}=\tilde{\mathbf{G}}^{(n,3,l)}-\mathbf{C}^{(n,3,l)}(\mathbf{A}^{(n,3)})^{-1}\mathbf{F}^{(n,2,l)}$$

and then updates the velocity solving

$$\mathbf{A}^{(n,3)}\mathbf{U}^{(n,3,l+1)}=\mathbf{F}^{(n,3,l)}-\mathbf{B}^{(n,3)}\mathbf{P}^{(n,3,l+1)}$$

Once the iterations have been completed, one sets $\mathbf{u}^{(n,3)}=\mathbf{u}^{(n,3,M+1)}$ and $E^{(n,3)}$ accordingly. One sets then

$$\rho^{n+1}=\rho^{(n,3)}\qquad\tilde{\mathbf{u}}^{(n,1)}=\mathbf{u}^{(n,3)}\qquad\tilde{E}^{(n,1)}=E^{(n,3)}$$

and proceeds to the implicit discretization of the viscous terms, which is carried out by the implicit part of the IMEX method described above. We would like to stress that the method outlined above does not require to introduce reference solutions, does not introduce inconsistencies in the splitting and only requires the solution of linear systems of a size equal to that of the number of discrete degrees of freedom needed to describe a scalar variable, as in [13]. This contrasts with other low Mach approaches based on IMEX methods, such as e.g. the technique proposed for the Euler equations in [45].

5 Implementation issues

As stated in Section 1, the proposed method has been implemented using the numerical library *deal.II*, which is based on a matrix-free approach. As a consequence, no global sparse matrix is built and only the action of the linear operators defined above on a vector is actually computed. The implementation follows the operator splitting strategy previously described.

Therefore, we have built two *ad-hoc* structures, one for the hyperbolic part and one for the diffusive part. Another feature of the library employed during the numerical simulations is the mesh adaptation capability, as we will see in the presentation of the results. The preconditioned conjugate gradient method implemented in the function *SolverCG* of the library was employed to solve the linear systems for the density and for the update of the velocity in the fixed point iterations, as well as to solve the linear systems associated to the matrices $\mathbf{A}^{(n,2)}$ and $\mathbf{A}^{(n,3)}$, while the GMRES solver implemented in the function *SolverGMRES* of the same library was used for the remaining linear systems.

On the other hand, in the diffusive part, a Symmetric Interior Penalty (SIP) approach has been adopted for the space discretization [1]. Moreover, following [15], we set for each face Γ of a cell K

$$\sigma_{\Gamma,K} = (k+1)^2 \frac{\text{diam}\,(\Gamma)}{\text{diam}\,(K)}$$

and we define the penalization constant of the SIP method as

$$\overline{C} = \frac{1}{2}\left(\sigma_{\Gamma,K^+} + \sigma_{\Gamma,K^-}\right)$$

if $\Gamma \in \mathcal{E}^I$ and $\overline{C} = \sigma_{\Gamma,K}$ if $\Gamma \in \mathcal{E}^B$. All the linear systems for this part are solved using the preconditioned conjugated gradient mentioned above. A geometric multigrid preconditioner has been employed for the solution of the symmetric linear systems using the procedure described in [25], whereas a Jacobi preconditioner has been used for the non symmetric ones.

6 Numerical tests

The numerical scheme outlined in the previous Sections has been validated in a number of benchmarks. We set $\mathcal{H} = \min\{\text{diam}(K) | K \in \mathcal{T}_h\}$ and we define two Courant numbers, one based on the speed of sound denoted by C and one based on the local velocity of the flow, the so-called advective Courant number, denoted by C_u:

$$C = \frac{1}{Ma} kc\Delta t/\mathcal{H}, \qquad C_u = ku\Delta t/\mathcal{H} \tag{48}$$

where c is the magnitude of the speed of sound and u is the magnitude of the flow velocity. In most of the tests, the ideal gas law with $\gamma = 1.4$ as the specific heat ratio is employed. In general, we consider $k = 1$ for all the simulations, unless differently stated.

6.1 Isentropic vortex

As a first benchmark, we consider for an ideal gas the two dimensional inviscid isentropic vortex also studied in [41, 45]. For this test, an analytic solution is available, that can be used to assess the convergence properties of the scheme. The initial conditions are given as a perturbation of a reference state

$$\rho(\mathbf{x},0)=\rho_\infty+\delta\rho \qquad \mathbf{u}(\mathbf{x},0)=\mathbf{u}_\infty+\delta\mathbf{u} \qquad p(\mathbf{x},0)=p_\infty+\delta p$$

and the exact solution is a propagation of the initial condition at the background velocity

$$\rho(\mathbf{x},t)=\rho(\mathbf{x}-\mathbf{u}_\infty t,0) \qquad \mathbf{u}(\mathbf{x},t)=\mathbf{u}(\mathbf{x}-\mathbf{u}_\infty t,0) \qquad p(\mathbf{x},t)=p(\mathbf{x}-\mathbf{u}_\infty t,0)$$

The typical perturbation is defined as

$$\widetilde{\delta T}=\frac{1-\gamma}{8\gamma\pi^2}\beta^2 e^{1-r^2}$$

with $r=\sqrt{(x-x_0)^2+(y-y_0)^2}$ denoting the radial coordinate and β being the vortex strength. As explained in [45], however, in order to emphasize the role of the Mach number Ma, we define

$$\delta T=\frac{1-\gamma}{8\gamma\pi^2}Ma^2\beta^2 e^{1-r^2}$$

and we set

$$\rho(\mathbf{x},0)=(1+\delta T)^{\frac{1}{\gamma-1}} \qquad p(\mathbf{x},0)=Ma^2\,(1+\delta T)^{\frac{\gamma}{\gamma-1}}\,. \tag{49}$$

For what concerns the velocity the typical perturbation is defined as

$$\widetilde{\delta\mathbf{u}}=\beta\begin{pmatrix}-y\\ x\end{pmatrix}\frac{e^{\frac{1}{2}\left(1-r^2\right)}}{2\pi} \tag{50}$$

and also in this case we rescale it using Ma

$$\delta\mathbf{u}=\beta Ma\begin{pmatrix}-y\\ x\end{pmatrix}\frac{e^{\frac{1}{2}\left(1-r^2\right)}}{2\pi}.$$

We apply the same reasoning also to the background velocity and therefore we define $\mathbf{u}_\infty=Ma\tilde{\mathbf{u}}_\infty$ with $\tilde{\mathbf{u}}_\infty=[10,10]^T$. To avoid any kind of issue from the boundary conditions, we choose a sufficiently large domain $\Omega=(-10,10)^2$ and periodic boundary conditions and, eventually, we set $\rho_\infty=1$, $p_\infty=1$, $x_0=y_0=0$, $\beta=10$, the final time $T_f=1$ and $Ma=0.1$. Notice that we refrain from investigating the properties of the method in the very

low Mach number regime for this test, since this entails an almost constant solution. The numerical experiments have been carried out on Cartesian meshes of square elements with N_{el} elements in each coordinate direction, choosing for each spatial resolution time steps so that the Courant numbers remained constant (hyperbolic scaling).

We first consider the original IMEX-ARK scheme with $\alpha = \frac{7-2\gamma}{6}$ for the explicit part. In Tables 3-5, the L^2 errors for density, velocity and pressure are reported at various resolutions for the $k = 1$ case, while the corresponding errors in the $k = 2$ case are given in Tables 6-8. We observe that, in general, convergence rates of at least $k + \frac{1}{2}$ are observed for $k = 1$ while for $k = 2$ the convergence rate seem to degrade at the finest resolution as soon as the Courant number grows.

N_{el}	L^2 rel. error ρ	L^2 rate ρ	L^2 rel. error $\mathbf{u}$	L^2 rate $\mathbf{u}$	L^2 rel. error p	L^2 rate p
10	$1.99 \cdot 10^{-3}$		$1.19 \cdot 10^{-2}$		$2.79 \cdot 10^{-3}$	
20	$7.87 \cdot 10^{-4}$	1.34	$3.87 \cdot 10^{-3}$	1.62	$1.11 \cdot 10^{-3}$	1.33
40	$2.56 \cdot 10^{-4}$	1.62	$1.08 \cdot 10^{-3}$	1.84	$3.62 \cdot 10^{-4}$	1.62
80	$7.22 \cdot 10^{-5}$	1.83	$2.73 \cdot 10^{-4}$	1.98	$1.01 \cdot 10^{-4}$	1.84

Table 3: Convergence test for the inviscid isentropic vortex at $C \approx 0.01$, $C_u \approx 0.01$ with $k = 1$ and $\alpha = \frac{7-2\gamma}{6}$ for the explicit part. Relative errors for the density, the velocity and the pressure in L^2 norm. N_{el} denotes the number of elements along each direction.

N_{el}	L^2 rel. error ρ	L^2 rate ρ	L^2 rel. error $\mathbf{u}$	L^2 rate $\mathbf{u}$	L^2 rel. error p	L^2 rate p
10	$1.99 \cdot 10^{-3}$		$1.20 \cdot 10^{-2}$		$2.78 \cdot 10^{-3}$	
20	$7.92 \cdot 10^{-4}$	1.33	$3.93 \cdot 10^{-3}$	1.61	$1.11 \cdot 10^{-3}$	1.32
40	$2.60 \cdot 10^{-4}$	1.61	$1.12 \cdot 10^{-3}$	1.81	$3.64 \cdot 10^{-4}$	1.61
80	$7.43 \cdot 10^{-5}$	1.81	$2.91 \cdot 10^{-4}$	1.94	$1.03 \cdot 10^{-4}$	1.82

Table 4: Convergence test for the inviscid isentropic vortex at $C \approx 0.05$, $C_u \approx 0.05$ with $k = 1$ and $\alpha = \frac{7-2\gamma}{6}$ for the explicit part. Relative errors for the density, the velocity and the pressure in L^2 norm. N_{el} denotes the number of elements along each direction.

N_{el}	L^2 rel. error ρ	L^2 rate ρ	L^2 rel. error $\mathbf{u}$	L^2 rate $\mathbf{u}$	L^2 rel. error p	L^2 rate p
10	$2.10 \cdot 10^{-3}$		$1.20 \cdot 10^{-2}$		$2.73 \cdot 10^{-3}$	
20	$8.00 \cdot 10^{-4}$	1.39	$4.07 \cdot 10^{-3}$	1.56	$1.09 \cdot 10^{-3}$	1.32
40	$2.67 \cdot 10^{-4}$	1.58	$1.21 \cdot 10^{-3}$	1.75	$3.64 \cdot 10^{-4}$	1.58
80	$7.95 \cdot 10^{-5}$	1.75	$3.44 \cdot 10^{-4}$	1.81	$1.06 \cdot 10^{-4}$	1.78

Table 5: Convergence test for the inviscid isentropic vortex at $C \approx 0.15$, $C_u \approx 0.14$ with $k = 1$ and $\alpha = \frac{7-2\gamma}{6}$ for the explicit part. Relative errors for the density, the velocity and the pressure in L^2 norm. N_{el} denotes the number of elements along each direction.

N_{el}	L^2 rel. error ρ	L^2 rate ρ	L^2 rel. error $\mathbf{u}$	L^2 rate $\mathbf{u}$	L^2 rel. error p	L^2 rate p
10	$6.37 \cdot 10^{-4}$		$2.61 \cdot 10^{-3}$		$9.09 \cdot 10^{-4}$	
20	$1.18 \cdot 10^{-4}$	2.43	$3.59 \cdot 10^{-4}$	2.86	$1.64 \cdot 10^{-4}$	2.47
40	$1.83 \cdot 10^{-5}$	2.69	$4.39 \cdot 10^{-5}$	3.03	$2.55 \cdot 10^{-5}$	2.69
80	$3.08 \cdot 10^{-6}$	2.57	$6.96 \cdot 10^{-6}$	2.66	$4.21 \cdot 10^{-6}$	2.60

Table 6: Convergence test for the inviscid isentropic vortex at $C \approx 0.01$, $C_u \approx 0.01$ with $k = 2$ and $\alpha = \frac{7-2\gamma}{6}$ for the explicit part. Relative errors for the density, the velocity and the pressure in L^2 norm. N_{el} denotes the number of elements along each direction.

N_{el}	L^2 rel. error ρ	L^2 rate ρ	L^2 rel. error $\mathbf{u}$	L^2 rate $\mathbf{u}$	L^2 rel. error p	L^2 rate p
10	$6.34 \cdot 10^{-4}$		$2.64 \cdot 10^{-3}$		$9.08 \cdot 10^{-4}$	
20	$1.19 \cdot 10^{-4}$	2.41	$3.80 \cdot 10^{-4}$	2.80	$1.65 \cdot 10^{-4}$	2.46
40	$1.95 \cdot 10^{-5}$	2.61	$5.82 \cdot 10^{-5}$	2.71	$2.63 \cdot 10^{-5}$	2.65
80	$4.31 \cdot 10^{-6}$	2.18	$1.81 \cdot 10^{-5}$	1.69	$5.06 \cdot 10^{-6}$	2.38

Table 7: Convergence test for the inviscid isentropic vortex at $C \approx 0.05$, $C_u \approx 0.05$ with $k = 2$ and $\alpha = \frac{7-2\gamma}{6}$ for the explicit part. Relative errors for the density, the velocity and the pressure in L^2 norm. N_{el} denotes the number of elements along each direction.

N_{el}	L^2 rel. error ρ	L^2 rate ρ	L^2 rel. error $\mathbf{u}$	L^2 rate $\mathbf{u}$	L^2 rel. error p	L^2 rate p
10	$6.25 \cdot 10^{-4}$		$2.74 \cdot 10^{-3}$		$8.94 \cdot 10^{-4}$	
20	$1.24 \cdot 10^{-4}$	2.33	$4.63 \cdot 10^{-4}$	2.57	$1.65 \cdot 10^{-4}$	2.44
40	$2.55 \cdot 10^{-5}$	2.28	$1.17 \cdot 10^{-4}$	1.98	$3.03 \cdot 10^{-5}$	2.45
80	$9.36 \cdot 10^{-6}$	1.45	$5.14 \cdot 10^{-5}$	1.19	$9.14 \cdot 10^{-6}$	1.73

Table 8: Convergence test for the inviscid isentropic vortex at $C \approx 0.15$, $C_u \approx 0.14$ with $k = 2$ and $\alpha = \frac{7-2\gamma}{6}$ for the explicit part. Relative errors for the density, the velocity and the pressure in L^2 norm. N_{el} denotes the number of elements along each direction.

Analogous results are shown in Tables 9-14 for the modified scheme with $\alpha = 0.5$, chosen, as discussed in Appendix A, in order to increase the region of absolute monotonicity without affecting too much stability. It can be seen that, for $k = 1$, slightly lower errors are obtained, especially for the density, while the behaviour for $k = 2$ is similar to that of the original scheme.

N_{el}	L^2 rel. error ρ	L^2 rate ρ	L^2 rel. error $\mathbf{u}$	L^2 rate $\mathbf{u}$	L^2 rel. error p	L^2 rate p
10	$2.02 \cdot 10^{-3}$		$1.19 \cdot 10^{-2}$		$2.80 \cdot 10^{-3}$	
20	$7.82 \cdot 10^{-4}$	1.37	$3.81 \cdot 10^{-3}$	1.64	$1.11 \cdot 10^{-3}$	1.33
40	$2.51 \cdot 10^{-4}$	1.64	$1.04 \cdot 10^{-3}$	1.87	$3.58 \cdot 10^{-4}$	1.63
80	$6.96 \cdot 10^{-5}$	1.85	$2.50 \cdot 10^{-4}$	2.06	$9.89 \cdot 10^{-5}$	1.86

Table 9: Convergence test for the inviscid isentropic vortex at $C \approx 0.01$, $C_u \approx 0.01$ with $k = 1$ and $\alpha = 0.5$ for the explicit part. Relative errors for the density, the velocity and the pressure in L^2 norm. N_{el} denotes the number of elements along each direction.

N_{el}	L^2 rel. error ρ	L^2 rate ρ	L^2 rel. error $\mathbf{u}$	L^2 rate $\mathbf{u}$	L^2 rel. error p	L^2 rate p
10	$2.08 \cdot 10^{-3}$		$1.19 \cdot 10^{-2}$		$2.81 \cdot 10^{-3}$	
20	$7.70 \cdot 10^{-4}$	1.43	$3.66 \cdot 10^{-3}$	1.70	$1.10 \cdot 10^{-3}$	1.35
40	$2.41 \cdot 10^{-4}$	1.68	$9.44 \cdot 10^{-4}$	1.95	$3.49 \cdot 10^{-4}$	1.66
80	$6.52 \cdot 10^{-5}$	1.89	$2.08 \cdot 10^{-4}$	2.18	$9.45 \cdot 10^{-5}$	1.88

Table 10: Convergence test for the inviscid isentropic vortex at $C \approx 0.05$, $C_u \approx 0.05$ with $k = 1$ and $\alpha = 0.5$ for the explicit part. Relative errors for the density, the velocity and the pressure in L^2 norm. N_{el} denotes the number of elements along each direction.

N_{el}	L^2 rel. error ρ	L^2 rate ρ	L^2 rel. error $\mathbf{u}$	L^2 rate $\mathbf{u}$	L^2 rel. error p	L^2 rate p
10	$2.34 \cdot 10^{-3}$		$1.17 \cdot 10^{-2}$		$2.88 \cdot 10^{-3}$	
20	$7.49 \cdot 10^{-4}$	1.64	$3.29 \cdot 10^{-3}$	1.83	$1.09 \cdot 10^{-3}$	1.40
40	$2.26 \cdot 10^{-4}$	1.73	$7.68 \cdot 10^{-4}$	2.10	$3.32 \cdot 10^{-4}$	1.72
80	$6.76 \cdot 10^{-5}$	1.74	$2.18 \cdot 10^{-4}$	1.82	$9.10 \cdot 10^{-5}$	1.87

Table 11: Convergence test for the inviscid isentropic vortex at $C \approx 0.15$, $C_u \approx 0.14$ with $k = 1$ and $\alpha = 0.5$ for the explicit part. Relative errors for the density, the velocity and the pressure in L^2 norm. N_{el} denotes the number of elements along each direction.

N_{el}	L^2 rel. error ρ	L^2 rate ρ	L^2 rel. error $\mathbf{u}$	L^2 rate $\mathbf{u}$	L^2 rel. error p	L^2 rate p
10	$6.41 \cdot 10^{-4}$		$2.59 \cdot 10^{-3}$		$9.07 \cdot 10^{-4}$	
20	$1.18 \cdot 10^{-4}$	2.44	$3.41 \cdot 10^{-4}$	2.93	$1.65 \cdot 10^{-4}$	2.43
40	$1.84 \cdot 10^{-5}$	2.68	$4.46 \cdot 10^{-5}$	2.94	$2.54 \cdot 10^{-5}$	2.50
80	$3.72 \cdot 10^{-6}$	2.31	$1.33 \cdot 10^{-5}$	1.75	$4.61 \cdot 10^{-6}$	2.46

Table 12: Convergence test for the inviscid isentropic vortex at $C \approx 0.01$, $C_u \approx 0.01$ with $k = 2$ and $\alpha = 0.5$ for the explicit part. Relative errors for the density, the velocity and the pressure in L^2 norm. N_{el} denotes the number of elements along each direction.

N_{el}	L^2 rel. error ρ	L^2 rate ρ	L^2 rel. error $\mathbf{u}$	L^2 rate $\mathbf{u}$	L^2 rel. error p	L^2 rate p
10	$6.53 \cdot 10^{-4}$		$2.56 \cdot 10^{-3}$		$9.05 \cdot 10^{-4}$	
20	$1.24 \cdot 10^{-4}$	2.40	$3.49 \cdot 10^{-4}$	2.87	$1.68 \cdot 10^{-4}$	2.43
40	$2.53 \cdot 10^{-5}$	2.29	$1.03 \cdot 10^{-4}$	1.76	$2.96 \cdot 10^{-5}$	2.50
80	$9.76 \cdot 10^{-6}$	1.37	$5.01 \cdot 10^{-5}$	1.00	$9.51 \cdot 10^{-6}$	1.64

Table 13: Convergence test for the inviscid isentropic vortex at $C \approx 0.05$, $C_u \approx 0.05$ with $k = 2$ and $\alpha = 0.5$ for the explicit part. Relative errors for the density, the velocity and the pressure in L^2 norm. N_{el} denotes the number of elements along each direction.

N_{el}	L^2 rel. error ρ	L^2 rate ρ	L^2 rel. error $\mathbf{u}$	L^2 rate $\mathbf{u}$	L^2 rel. error p	L^2 rate p
10	$7.10 \cdot 10^{-4}$		$2.60 \cdot 10^{-3}$		$9.17 \cdot 10^{-4}$	
20	$1.68 \cdot 10^{-4}$	2.08	$6.10 \cdot 10^{-4}$	2.09	$1.98 \cdot 10^{-4}$	2.21
40	$5.91 \cdot 10^{-5}$	1.51	$3.04 \cdot 10^{-4}$	1.00	$5.76 \cdot 10^{-5}$	1.78
80	$2.85 \cdot 10^{-5}$	1.05	$1.52 \cdot 10^{-4}$	1.00	$2.66 \cdot 10^{-5}$	1.11

Table 14: Convergence test for the inviscid isentropic vortex at $C \approx 0.15$, $C_u \approx 0.14$ with $k = 2$ and $\alpha = 0.5$ for the explicit part. Relative errors for the density, the velocity and the pressure in L^2 norm. N_{el} denotes the number of elements along each direction.

In further numerical experiments, we have observed that the lack of absolute monotonicity strongly effectively affects the computation of density and, as a consequence, the stability of the whole numerical scheme. For Courant number around $C \approx 0.2$ the original method becomes unstable, while the modified scheme with $\alpha = 0.5$ is still able to recover the expected convergence rates at least in the $k = 1$ case, as evident from Table 15, while again in the $k = 2$ reported in Table 16 we observe a degradation of the convergence rates. In order to be able to run at slightly longer time steps we have then chosen to use the $\alpha = 0.5$ value for the IMEX scheme for the rest of the numerical simulations carried out in this paper. We notice also that, for both schemes, the results compare well with the analogous results presented in [41] and with those obtained in [45] with a higher order IMEX method.

N_{el}	L^2 rel. error ρ	L^2 rate ρ	L^2 rel. error $\mathbf{u}$	L^2 rate $\mathbf{u}$	L^2 rel. error p	L^2 rate p
10	$2.71 \cdot 10^{-3}$		$1.16 \cdot 10^{-2}$		$2.96 \cdot 10^{-3}$	
20	$7.74 \cdot 10^{-4}$	1.81	$2.95 \cdot 10^{-3}$	1.98	$1.09 \cdot 10^{-3}$	1.44
40	$2.34 \cdot 10^{-4}$	1.73	$7.71 \cdot 10^{-4}$	1.94	$3.28 \cdot 10^{-4}$	1.73
80	$8.91 \cdot 10^{-5}$	1.39	$3.74 \cdot 10^{-4}$	1.04	$1.01 \cdot 10^{-4}$	1.70

Table 15: Convergence test for the inviscid isentropic vortex at $C \approx 0.2$, $C_u \approx 0.2$ with $k = 1$ and $\alpha = 0.5$ for the explicit part. Relative errors for the density, the velocity and the pressure in L^2 norm. N_{el} denotes the number of elements along each direction.

N_{el}	L^2 rel. error ρ	L^2 rate ρ	L^2 rel. error $\mathbf{u}$	L^2 rate $\mathbf{u}$	L^2 rel. error p	L^2 rate p
10	$8.18 \cdot 10^{-4}$		$2.92 \cdot 10^{-3}$		$9.58 \cdot 10^{-4}$	
20	$2.42 \cdot 10^{-4}$	1.76	$1.02 \cdot 10^{-3}$	1.52	$2.57 \cdot 10^{-4}$	1.90
40	$9.88 \cdot 10^{-5}$	1.29	$5.20 \cdot 10^{-4}$	0.97	$9.36 \cdot 10^{-5}$	1.46
80	$4.79 \cdot 10^{-5}$	1.04	$2.58 \cdot 10^{-4}$	1.01	$4.47 \cdot 10^{-5}$	1.07

Table 16: Convergence test for the inviscid isentropic vortex at $C \approx 0.2$, $C_u \approx 0.2$ with $k = 2$ and $\alpha = 0.5$ for the explicit part. Relative errors for the density, the velocity and the pressure in L^2 norm. N_{el} denotes the number of elements along each direction.

We have also tested in this case the $h-$adaptive version of the method, in order to validate the implementation also in case of non-conforming meshes. The local refinement criterion is based on the gradient of the density. More specifically, we define for each element K the quantity

$$\eta_K = \|\nabla \rho\|_{\infty,K}$$

that acts as local refinement indicator, where $\|\cdot\|_{\infty,K}$ denotes the $L^\infty-$norm over the element K. Table 17 shows the relative errors for all the quantities on a sequence of adaptive simulations keeping the maximum Courant numbers fixed. Figure 1 shows instead the density and the adapted mesh at $t = T_f$, from which it can be seen that the refinement criterion is able to track the vortex correctly.

N_{el}	L^2 rel. error ρ	L^2 rel. error $\mathbf{u}$	L^2 rel. error p
271	$2.19 \cdot 10^{-2}$	$1.20 \cdot 10^{-2}$	$2.97 \cdot 10^{-3}$
586	$6.39 \cdot 10^{-4}$	$3.09 \cdot 10^{-3}$	$9.08 \cdot 10^{-4}$
1999	$1.80 \cdot 10^{-4}$	$7.93 \cdot 10^{-4}$	$2.56 \cdot 10^{-4}$
7678	$5.16 \cdot 10^{-5}$	$1.84 \cdot 10^{-4}$	$7.42 \cdot 10^{-5}$

Table 17: Adaptive simulations of the inviscid isentropic vortex at different resolutions with a maximum $C \approx 0.1$, $C_u \approx 0.1$, relative errors for the density, the velocity and the pressure in L^2 norm with $k = 1$.

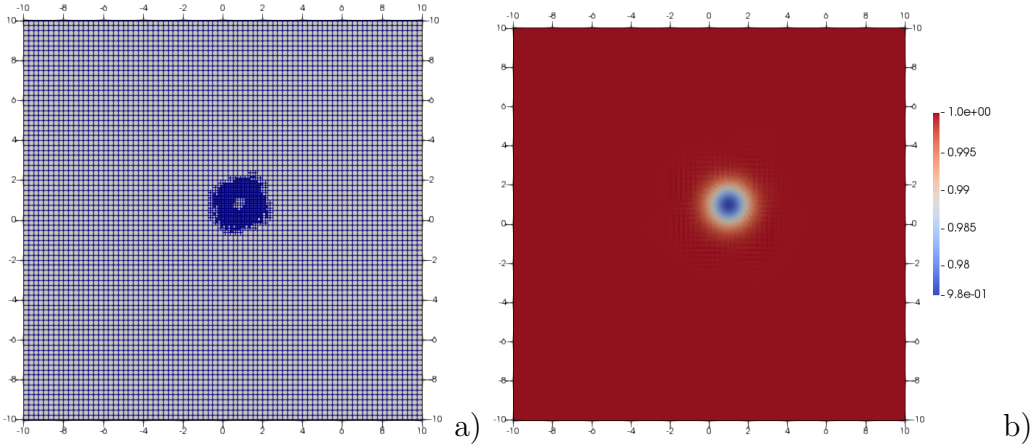


Figure 1: Adaptive simulation of the inviscid isentropic vortex benchmark: a) computational mesh at $t = T_f$, b) contour plot of the density at $t = T_f$.

6.2 Sod shock tube problem

Even though the proposed method is particularly well suited for low Mach number flows, we have also tested its behaviour also in a situation in which shock waves occur. For this purpose, we have first considered the classical Sod shock tube problem for an ideal gas proposed in [38] and also discussed in [13]. It consists of a right-moving shock wave, an intermediate contact discontinuity and a left-moving rarefaction fan. In this higher Mach number regime it is more appropriate to use the Local Lax-Friedrichs flux (LLF), defined by setting

$$\lambda^{(n,1)} = \max\left(\left\|\mathbf{u}^{(n,1)^+}\right\| + \frac{1}{Ma}c^{(n,1)^+}, \left\|\mathbf{u}^{(n,1)^-}\right\| + \frac{1}{Ma}c^{(n,1)^-}\right)$$
$$\lambda^{(n,2)} = \max\left(\left\|\mathbf{u}^{(n,2)^+}\right\| + \frac{1}{Ma}c^{(n,2)^+}, \left\|\mathbf{u}^{(n,2)^-}\right\| + \frac{1}{Ma}c^{(n,2)^-}\right)$$

where $c = \sqrt{\gamma \frac{p}{\rho}}$ is the speed of sound.

The presence of different discontinuities requires the use of a monotonic scheme to avoid undershoots and overshoots. It is well known that using Q_0 finite elements in combination with LLF and explicit time integration that complies with the monotonicity constraints discussed in [16, 23, 22] guarantees the monotonicity of the solution. Hence, a way to obtain monotonic results is to project the numerical solution onto the Q_0 subspace for each element in which a suitable jump indicator exceeds a certain threshold. Similar projections onto low order components of the solution are also used in several monotonization approaches, see e.g. [14]. However, since in the proposed scheme only the density is treated in a full explicit fashion, in order

to avoid an excessive complication in the structure of the resulting method we choose to apply this Q_0 projection strategy only for the density variable, without introducing monotonization for the velocity and the pressure. While we are aware that this is not sufficient to guarantee full monotonicity, the derivation of a fully monotonic IMEX scheme goes beyond the scope of this work and we do not investigate this issue further here. Therefore, the results in this Section are to be interpreted merely as a first stress test of the proposed scheme at higher Mach number values.

We use a smoothness indicator based on the jump of the density across two faces. More in detail, we define for each element K the quantity

$$\eta_K = \sum_{\Gamma \in \mathcal{E}_K} \left\| \rho^+ - \rho^- \right\|^2_{2,\Gamma}$$

where $\mathcal{E}_K$ denotes the set of all faces belonging to cell K and $\|\cdot\|$ represents the standard L^2 norm on Γ. Table 18 highlights the setting of the problem in terms of initial conditions and position of the initial discontinuity. We consider a 2D domain $(-0.5, 0.5) \times (0, 0.1)$ in order to test the ability of the method to capture one-dimensional waves also on multi-dimensional grids. The other component of the velocity is initialized to 0 and periodic boundary conditions are imposed in the transverse direction y. The mesh is composed by 500×50 elements and the time step is chosen in such a way that the maximum Courant number $C \approx 0.07$, while the maximum advective Courant number C_u is around 0.06. Figure 2 shows the results for the density, the velocity and the pressure at $t = 0.2$ compared with the exact solution. One can easily notice that the discontinuities are located at the right position and that the post-discontinuity values are correct.

ρ_L	u_L	p_L	ρ_R	u_R	p_R	x_d
1	0	1	0.125	0	0.1	0

Table 18: Initial left and right states for Sod shock tube problem. x_d denotes the position of the initial discontinuity

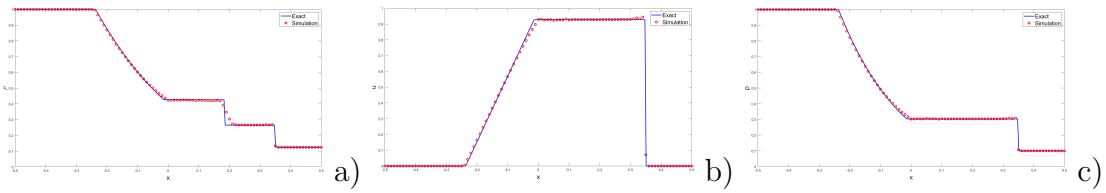


Figure 2: Sod shock tube problem at $t = 0.2$, comparison with exact solution, a) density, b) velocity, c) pressure

We have also considered the same problem in the case of the van der Waals EOS, taking $a = b = 0.5$ as proposed in [13]. All the other parameters are the same as in the previous case. Figure 3 shows the results for the density, the velocity and the pressure at $t = 0.2$. Again, a good agreement between the numerical results and the exact solution is observed for all the quantities, also in the case of a real gas equation of state.

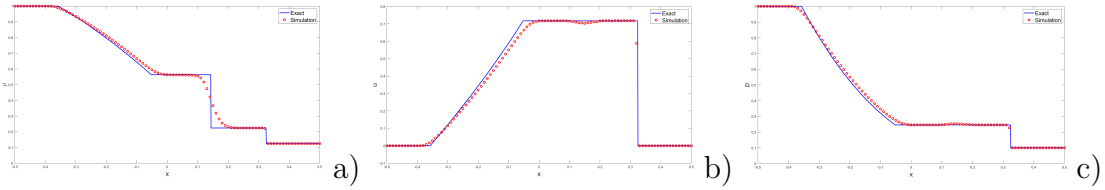


Figure 3: Sod shock tube problem at $t = 0.2$, comparison with exact solution, a) density, b) velocity, c) pressure

6.3 2D Lid-driven cavity

We consider now the classical 2D Lid-driven cavity test case. The computational domain is the box $\Omega = (0, 1) \times (0, 1)$ which is initialized with a density $\rho = 1$ and a velocity $\mathbf{u} = \mathbf{0}$. The flow is driven by the upper boundary, whose velocity is set to $\mathbf{u} = (1, 0)^T$, while on the other three boundaries a no-slip condition is imposed. We set $Re = 100$ and $Ma^2 = 10^{-5}$. The advantage of the proposed scheme is that the allowed time step is more than 100 times larger compared to that of a fully explicit scheme. Indeed, the time-step chosen is such that the maximum advective Courant number C_u is around 0.12, while the maximum Courant number C is around 49. The streamlines are shown in Figure 4 and highlight the formation of the main recirculation pattern. A comparison of the horizontal component of the velocity along the vertical middle line and of the vertical component of the velocity along the horizontal middle line with the reference solutions in [17, 41] is also presented.

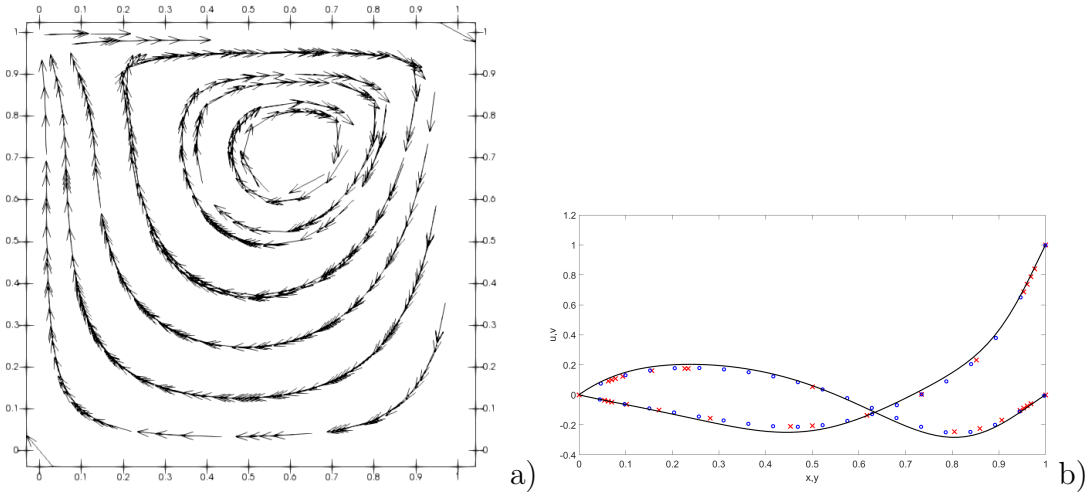


Figure 4: Computational results for the 2D lid-driven cavity, a) streamlines, b) comparison with the solutions in [17] and in [41]. Blue dots denote the results in [17], red crosses the results in [41] and the black line our numerical results.

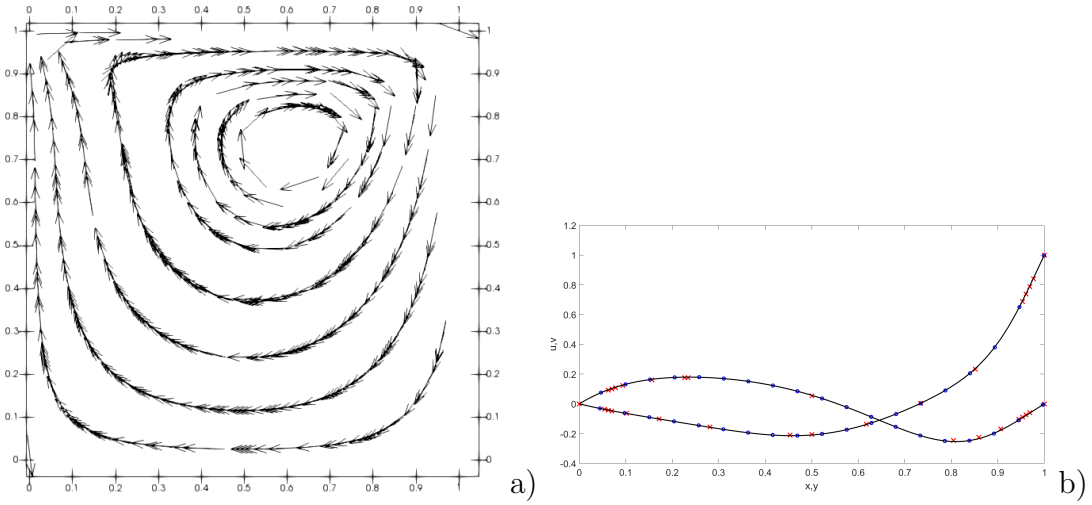


Figure 5: Computational results for the 2D lid-driven cavity with $k = 2$, a) streamlines, b) comparison with the solutions in [17] and in [41]. Blue dots denote the results in [17], red crosses the results in [41] and black line our numerical results.

We note a reasonable agreement between the different solutions, even though there is a still visible discrepancy between our results and the reference ones. Since the solution in [41] is obtained using third degree polynomials, in order to further improve the results, we consider also the case

$k = 2$. For this higher order approximation we note that our results fit very well both the reference solutions.

We have also tested the $h-$adaptive version of the same algorithm, using a refinement criterion based on the vorticity. More specifically, we define

$$\eta_K = \text{diam}\,(K)\,\|\nabla \times \mathbf{u}\|_{2,K}$$

as local indicator. We start from a uniform Cartesian with 16 elements along each direction. We allowed refinement for 5% of the elements with largest indicator values and coarsening for 30% of the elements with the smallest indicator values. The minimum element diameter allowed is $\mathcal{H} = \frac{1}{64}$, while the maximum element diameter is $\mathcal{H} = \frac{1}{16}$. Figure 6 reports the computational mesh at steady state and the computed streamlines. One can easily notice that the local refinement criterion is able to detect and to automatically enhance the resolution in the zones where vortices appear, as well as along the top boundary of the domain. For a more quantitative view, in Figure 7, we compare again the components of the velocity along the middle lines and, moreover, the absolute difference between the velocities of the fixed mesh and adaptive simulations is plotted over the whole domain, showing that no substantial loss of accuracy has occurred with a reduction of around 25% of the required computational time.

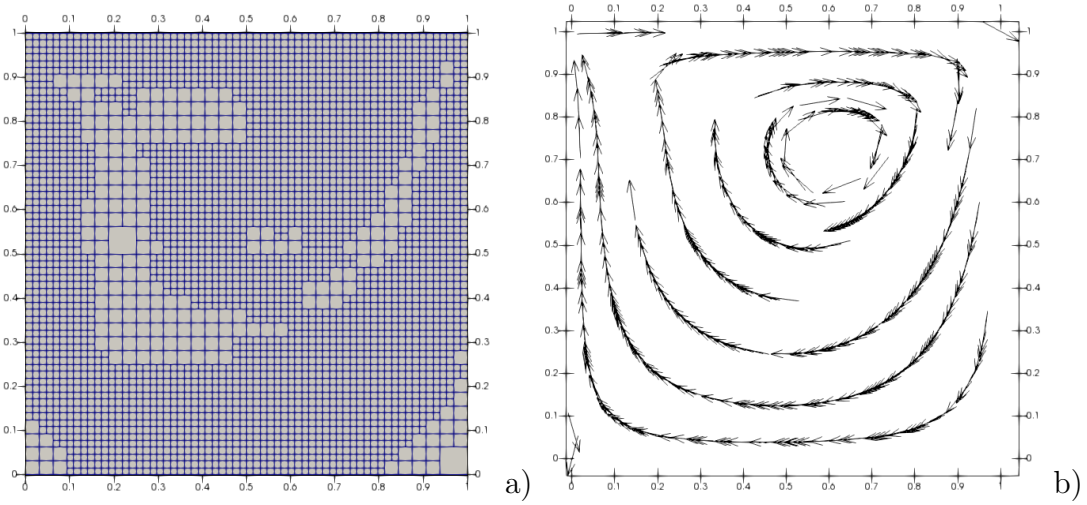
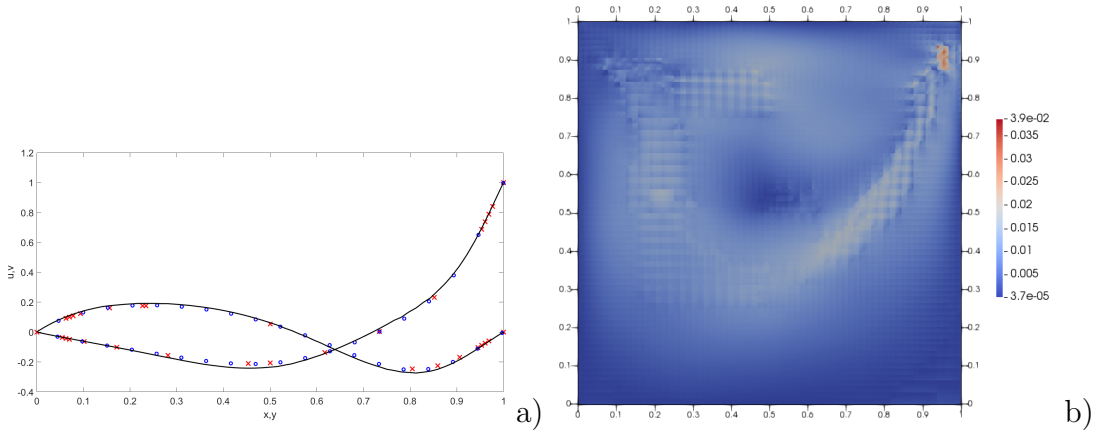


Figure 6: Adaptive simulation for the 2D lid-driven cavity, a) mesh at steady state, b) streamlines.



a) b)

Figure 7: Adaptive simulation for the 2D lid-driven cavity, a) comparison with the solutions in [17] and in [41]. Blue dots denote the results in [17], red crosses the results in [41] and black line our numerical results, b) difference for velocity magnitude between the fixed grid simulation and the adaptive simulation (interpolated to the fixed grid).

6.4 Cold bubble

In this Section, we consider a test case proposed in [34], in which gravity effects are also taken into account. This problem will also serve as a first benchmark for the application of the method to the case of real gases equation of state. The computational domain is the rectangle $(0, 1000) \times (0, 2000)$ and the initial condition is represented by a thermal anomaly introduced in an isentropic background atmosphere with constant potential temperature $\theta_0 = 303$. We recall that the potential temperature is defined for an ideal gas as

$$\theta = T \left(\frac{p_0}{p} \right)^{\frac{\gamma-1}{\gamma}}$$

where p_0 is a reference pressure. The perturbation potential temperature θ' defines the initial datum, given by

$$\theta' = \begin{cases} A & \text{if } r \leq r_0 \\ A \exp\left(-\frac{(r-r_0)^2}{\sigma^2} \right) & \text{if } r > r_0 \end{cases}$$

with $r^2 = (x - x_0)^2 + (y - y_0)^2$ and $x_0 = 500$, $y_0 = 1250$, $r_0 = 50$, $\sigma = 100$ and $A = -15$. Moreover, we set $Re = 1000$, $Fr = \frac{1}{9.8}$ and $Ma^2 = 10^{-5}$. Notice that we have employed a larger value of the Reynolds number with respect to the original configuration in [34]. Concerning the boundary conditions, wall boundary conditions are imposed at all the boundaries.

Therefore, we set $\rho^- = \rho^+, \mathbf{u}^- = \mathbf{u}^+ - 2\left(\mathbf{u}^+ \cdot \mathbf{n}^+\right)\mathbf{n}^+$ and $\rho^+ E^+ = \rho^- E^-$, which reduces to

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

It is worth to notice that the computation of p^- necessary to impose the wall boundary conditions clearly depends on the specific equation of state. In order to enhance the computational efficiency, we employ again mesh $h-$adaptivity tool, as refinement indicator the gradient of the potential temperature, since this quantity allows to identify the cold bubble. More specifically, we set

$$\eta_K = \|\nabla\theta\|_{\infty,K}$$

as local indicator and we allow to refine when η_K exceeds a certain threshold and to coarsen below another threshold. The initial computational grid is composed by 50×100 elements and we allowed up to two local refinements only, so as to keep under control the advective Courant number. The time step is taken equal to 0.08, yielding to a maximum Courant number $C \approx 5.6$ and a maximum advective Courant number $C_u \approx 0.24$.

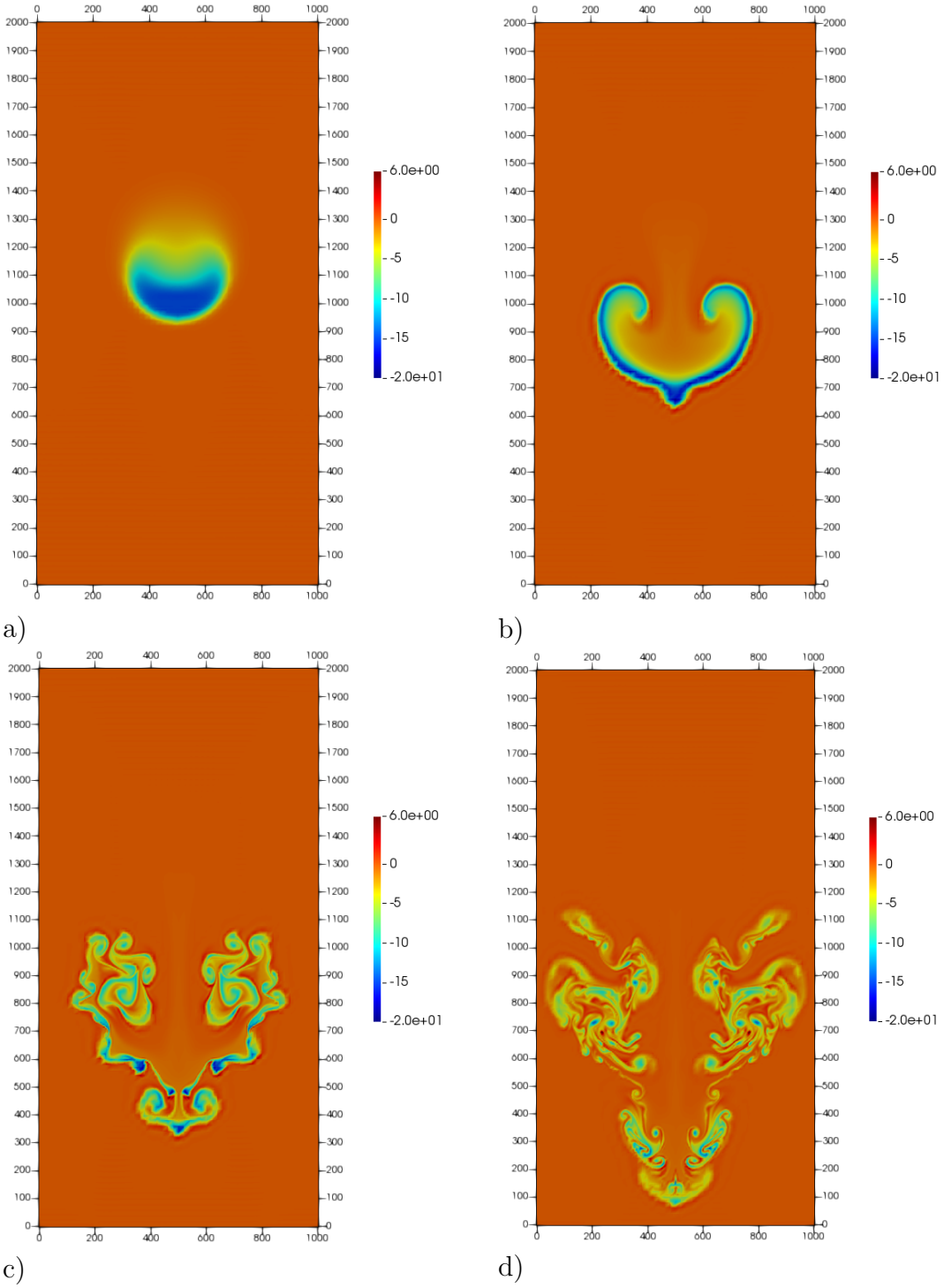


Figure 8: Cold bubble in an isentropic atmosphere, deviation from the basic-state of the potential temperature for adaptive simulation at: a) $t = 50$, b) $t = 100$, c) $t = 150$, d) $t = 200$.

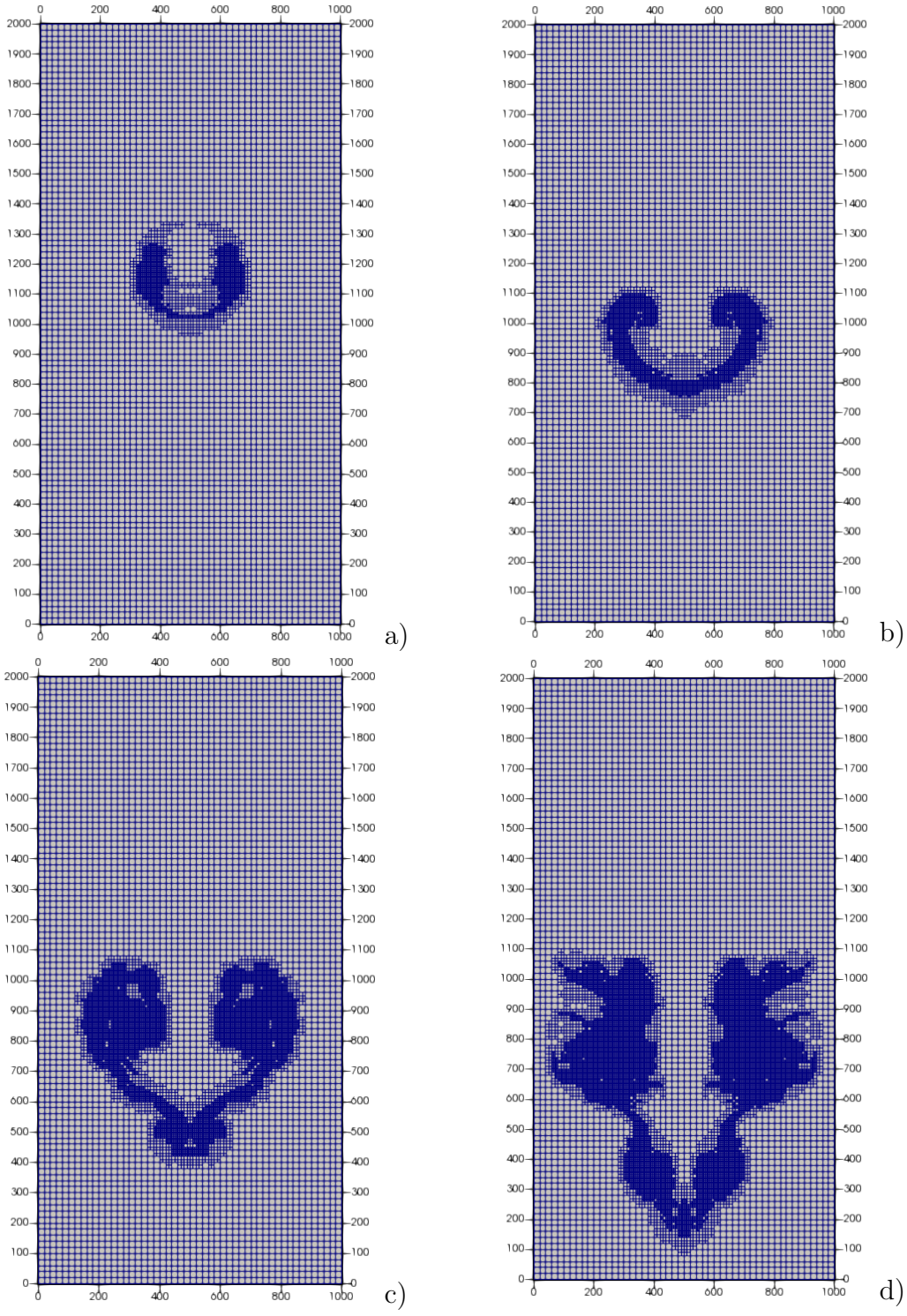


Figure 9: Cold bubble in an isentropic atmosphere, computational grid for adaptive simulation at: a) $t = 50$, b) $t = 100$, c) $t = 150$, d) $t = 200$.

Figure 8 shows the deviation of the potential temperature from θ_0 at $t = 50$, $t = 100$, $t = 150$ and $t = 200$. The qualitative behaviour is the one expected, since the bubble falls and deforms, gradually developing Kelvin-Helmholtz type instabilities. Three fixed-point iterations on average for each stage were required. It is worth to point out that a reduction in computational time of around 30% with respect to a full resolution grid has been obtained. We also show in Figure 9 the evolution of the mesh during the simulation; as one can easily notice, the refinement criterion is able to track the bubble and the final mesh consists of 19181 elements instead of the 80000 elements of the full resolution mesh.

6.5 Non ideal gas

In this Section, we investigate the ability of the proposed scheme to deal with different equations of state. We first consider the van der Waals equation (8) with $R = 287.05$, $a = 11000$ and $b = 0.0005$, leading to an average compressibility factor $z \approx 0.83$. We want to employ again the mesh adaptivity procedure in order to enhance the computational efficiency. However, for non-ideal gas equations of state, the definition of a potential temperature is not trivial. It can be proven, after some calculations reported in Appendix B, that the quantity $\beta = \log(T) - 2(\gamma - 1)\,\mathrm{atanh}(2\rho b - 1)$ is constant in an isentropic process for the van der Waals EOS. Therefore, we consider it as the counterpart of the potential temperature and the local refinement indicator for each element is defined by

$$\eta_K = \|\nabla\beta\|_{\infty,K}\,.$$

The same remeshing procedure described in the previous Section is then employed also in this case. Figure 10 shows the contour plots of the temperature at $t = 50$, $t = 100$ and $t = 150$; the qualitative behaviour is in agreement with the ideal gas case, even though the Kevin Helmholtz instability seems to appear earlier. We also report in Figure 11 the evolution of the computational grid: one can easily notice that, also in this case, the criterion is able to automatically detect the bubble.

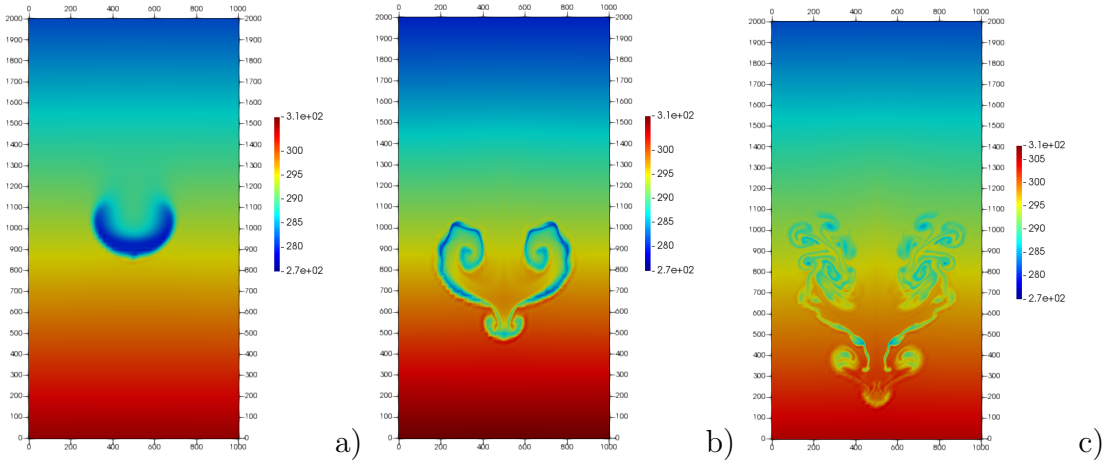


Figure 10: Cold bubble test case, contour plot of the temperature for adaptive simulation using van der Waals equation of state at: a) $t = 50$, b) $t = 100$, c) $t = 150$.

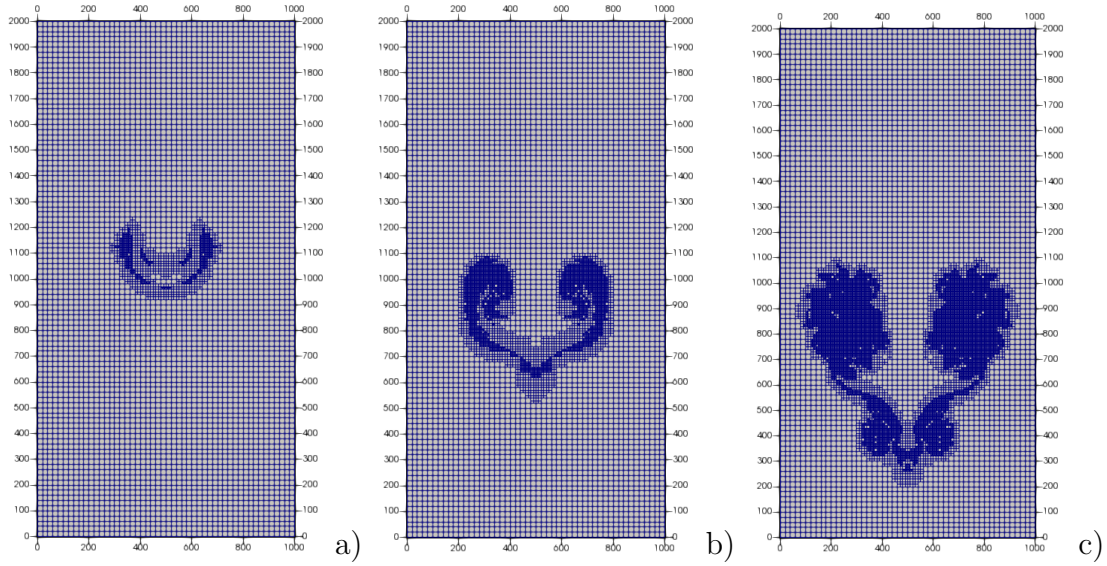


Figure 11: Cold bubble test case, evolution of the mesh for adaptive simulation using van der Waals equation of state at: a) $t = 50$, b) $t = 100$, c) $t = 150$.

Finally, let us consider now the Stiffened Gas equation of state (SG-EOS) (10) with $q = 0$ and $\pi = 18000$ in order to obtain again an average compressibility factor $z \approx 0.83$. In this case it can be shown (see Appendix B for further details) that the quantity which remains constant in an isentropic process is $\delta = \frac{p+\pi}{\rho^\gamma}$ and again we use its gradient as local refinement indicator.

Figure 12 shows the contour plots of the temperature at $t = 50$, $t = 100$ and $t = 150$, while Figure 13 reports the evolution of the computational grid at the same instants. The qualitative behaviour in this case is more similar to the one obtained in Section 6.4 and we note again that the refinement criterion is able to automatically track the bubble. Both the simulations also in this case required an average of 3 fixed-point iterations for each stage.

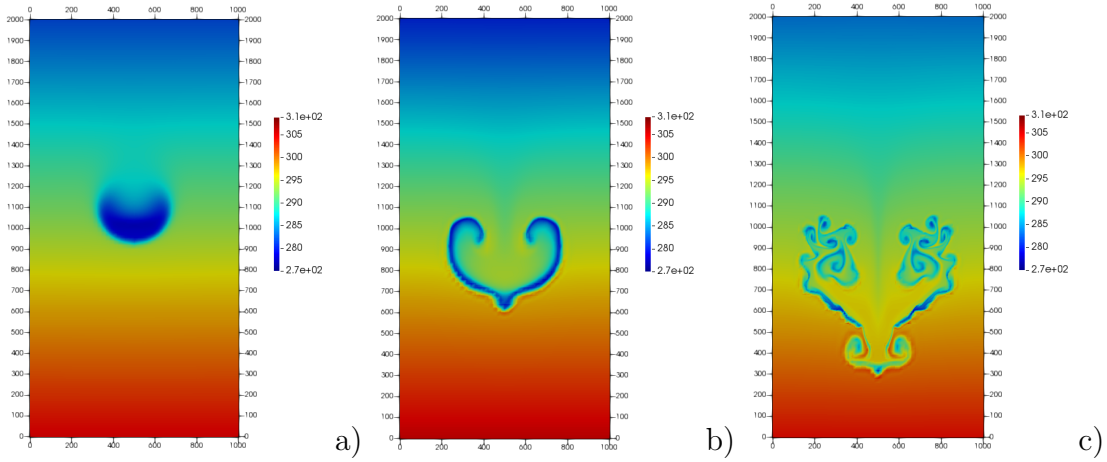


Figure 12: Cold bubble test case, contour plot of the temperature for adaptive simulation using SG-EOS at: a) $t = 50$, b) $t = 100$, c) $t = 150$.

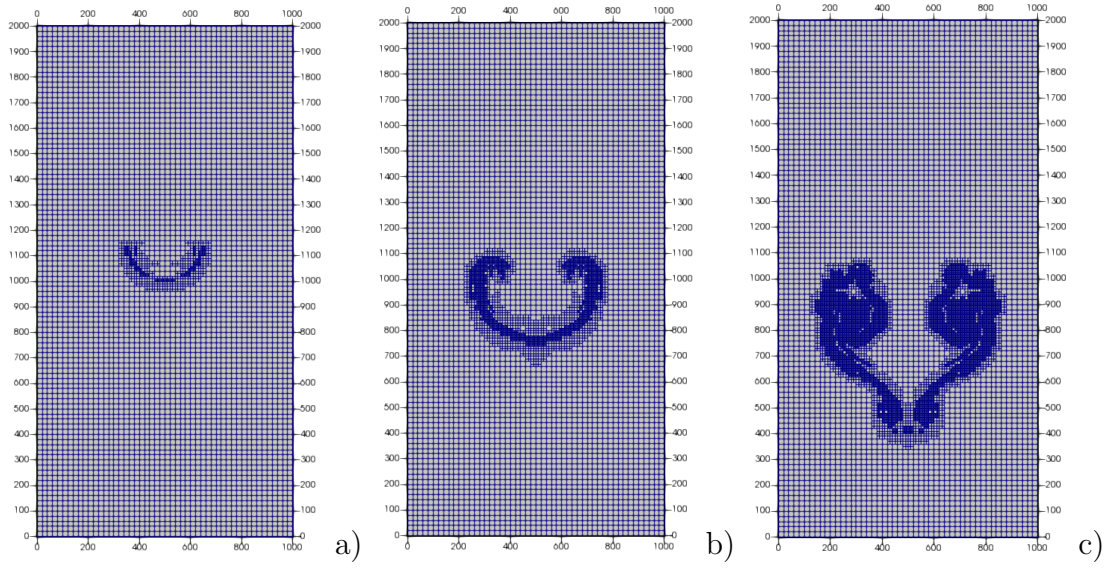


Figure 13: Cold bubble test case, evolution of the computational grid for adaptive simulation using SG-EOS at: a) $t = 50$, b) $t = 100$, c) $t = 150$.

6.6 Warm bubble

Up to now we have checked the method in absence of heat conduction. Let us consider now the test case proposed in [10] of a rising warm bubble. The domain is the square box $\Omega = (-0.5, 1.5) \times (-0.5, 1.5)$ with periodic boundary conditions on the lateral boundaries and wall boundary conditions on the top and on the bottom of the domain. The initial temperature corresponds to a Gaussian profile

$$T(\mathbf{x}, 0) = \begin{cases} 386.48 & \text{if } r > r_b \\ \dfrac{p_0}{R\left(1 - 0.1 e^{\frac{r^2}{\sigma^2}}\right)} & \text{if } r \leq r_b \end{cases}$$

where $r = |\mathbf{x} - \mathbf{x}_b|$ is the distance from the center $\mathbf{x}_b = (0.5, 0.35)^T$, $r_b = 0.25$ is the radius and $\sigma = 2$. We set $p_0 = 10^5$ and $R = 287$. Moreover, following [10], we consider:

$$Re = 804.9 \qquad Pr = 0.71 \qquad Fr \approx 0.004 \qquad Ma \approx 0.01.$$

The grid is composed by 120 elements along each direction and the time step is such that the maximum Courant number $C \approx 118$ and the maximum value of advective Courant number C_u is around 0.03. Figures 14,15 and 16 show the results at $t \in \{10, 15, 20\}$ both in terms of contours and plots along the same specific cuts along $x-$axis chosen in [10]. All the results are in good agreement with the reference ones and, also in this case, we are able to recover the Kelvin-Helmholtz instability.

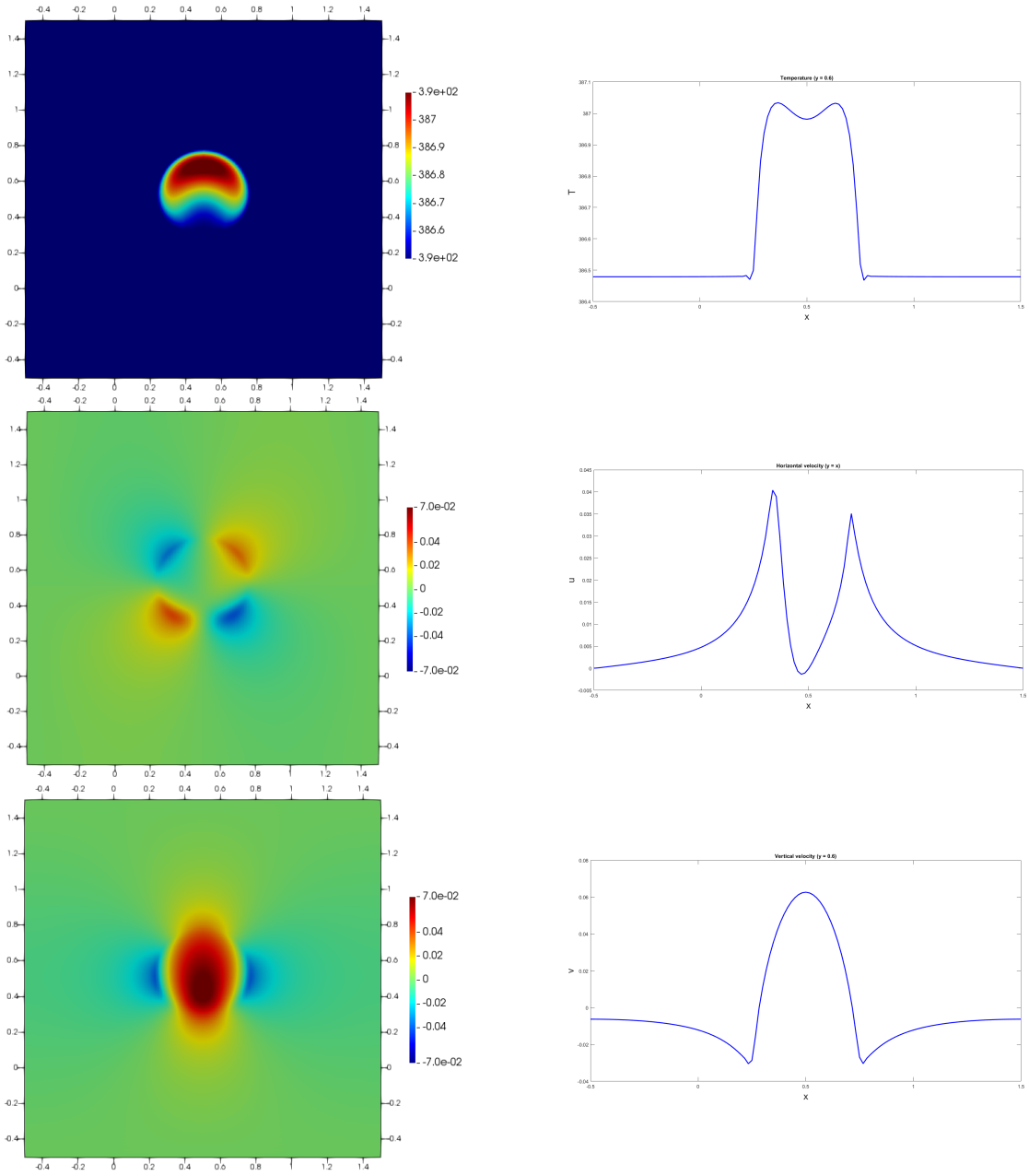


Figure 14: Warm bubble test case, results at $t = 10$. From bottom to top: temperature, horizontal velocity and vertical velocity.

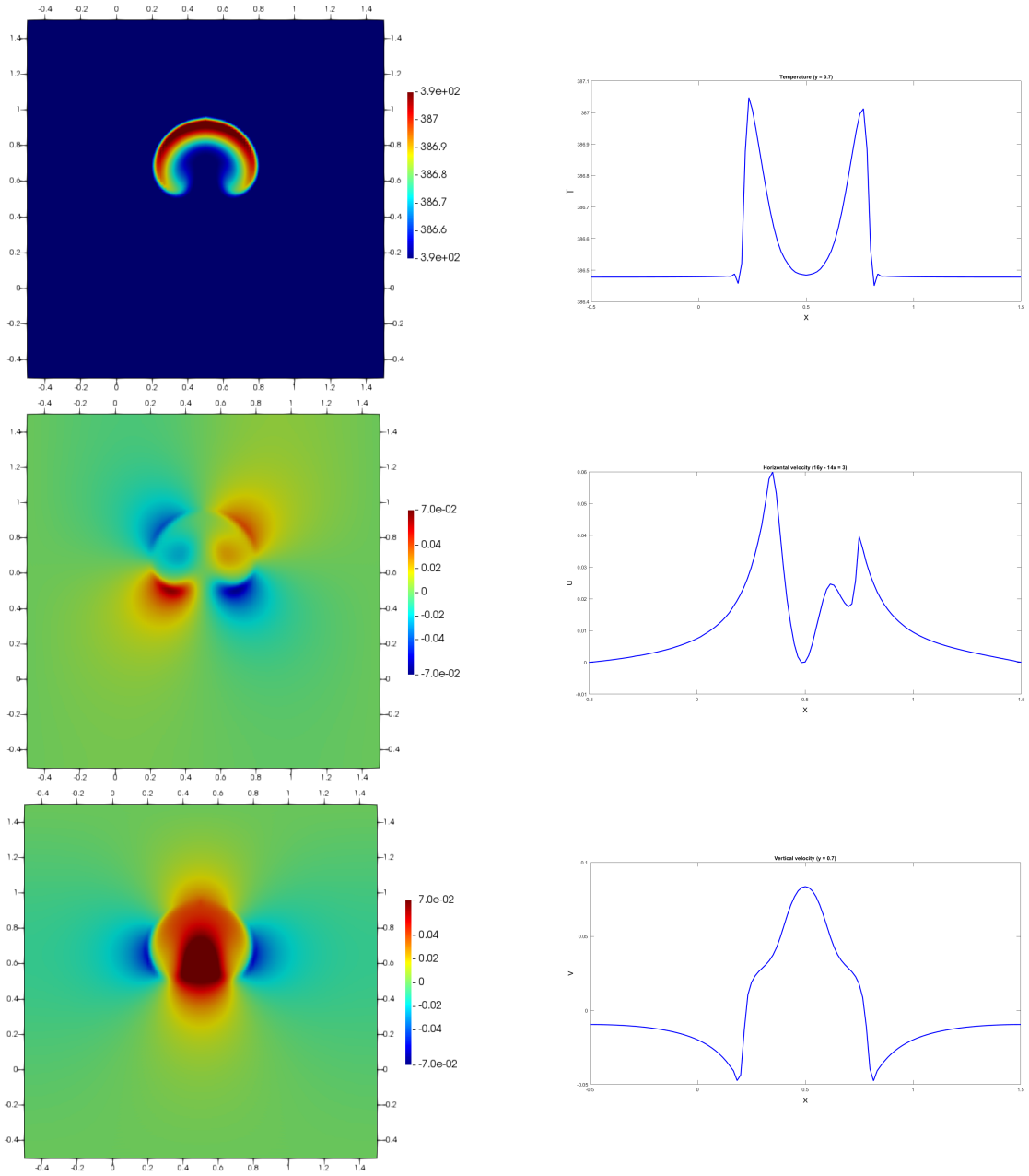


Figure 15: Warm bubble test case, results at $t = 15$. From bottom to top: temperature, horizontal velocity and vertical velocity.

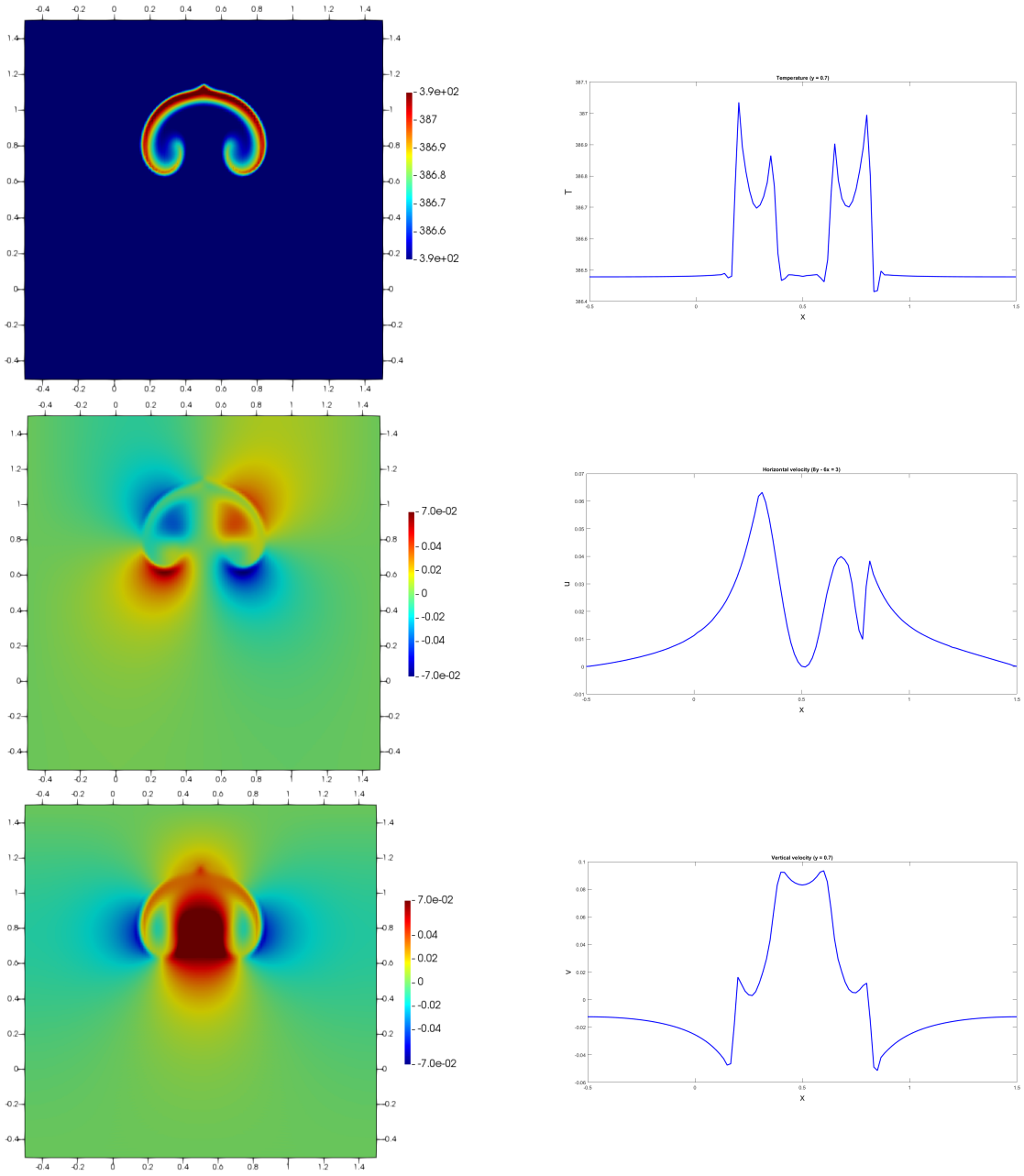


Figure 16: Warm bubble test case, results at $t = 20$. From bottom to top: temperature, horizontal velocity and vertical velocity.

7 Conclusions and future perspectives

We have proposed an efficient, $h-$adaptive IMEX-DG solver for the compressible Naiver-Stokes equations with general equation of state. The solver combines ideas from the discretization approaches in [11, 13, 20, 29] and proposes an improvement in the choice of the free parameter employed by the explicit part of the IMEX scheme described in [20]. The resulting method achieves full second order accuracy also including viscous terms and implements an $h-$adaptive approach in the framework of the numerical library *deal.II*. A number of physically based adaptation criteria have been proposed, which allow to exploit the full efficiency of the adaptive technique also for real gas simulations. A number of numerical experiments validate the proposed method and show its potential for low Mach number problems. In future work, we plan to extend the scheme to multiphase flows and to demonstrate its potential for application to atmospheric flows.

Acknowledgements

We thank M. Tavelli for providing the original data of the cavity flow simulation discussed in Section 6. We also gratefully acknowledge several useful discussions with A. Della Rocca on numerical methods related to those presented here.

A Stability and monotonicity of the explicit time discretization

In this Appendix we study the stability and monotonicity of the explicit part of the IMEX scheme applied in the paper. We recall that the Butcher tableaux for the explicit part of the method is given by

$$\begin{array}{c|ccc} 0 & 0 & & \\ \gamma & \gamma & 0 & \\ 1 & 1-\alpha & \alpha & 0 \\ \hline & \frac{1}{2}-\frac{\gamma}{4} & \frac{1}{2}-\frac{\gamma}{4} & \frac{\gamma}{2} \end{array}$$

In [20], the choice $\alpha = \frac{7-2\gamma}{6}$ was made to maximize the stability region of the resulting scheme, but this coefficient is indeed a free parameter and can also be chosen in different ways, as long as stability is not compromised. In order to identify possible alternative choices, we perform an analysis using the concepts introduced in [28], [23], [16] (see also the review in [22]). A

similar analysis for the implicit part of the IMEX scheme was carried out in [7], to which we refer for a summary of the related theoretical results. We then define

$$A = \begin{bmatrix} 0 & 0 & 0 \\ \gamma & 0 & 0 \\ 1-\alpha & \alpha & 0 \end{bmatrix} \qquad b^T = \begin{bmatrix} \frac{1}{2} - \frac{\gamma}{4} & \frac{1}{2} - \frac{\gamma}{4} & \frac{\gamma}{2} \end{bmatrix}$$

with $\gamma = 2 - \sqrt{2}$. We define for $\xi \in \mathbb{R}$ the quantities

$$\begin{aligned} A(\xi) &= A\left(I - \xi A\right)^{-1} & b^T(\xi) &= b^T\left(I - \xi A\right)^{-1} \\ e(\xi) &= \left(I - \xi A\right)^{-1} e & \varphi(\xi) &= 1 + \xi b^T\left(I - \xi A\right)^{-1} e \end{aligned} \tag{51}$$

where I is the 3×3 identity matrix and e is a vector whose all components are equal to 1. Therefore, for the specific scheme, we obtain

$$A(\xi) = \begin{bmatrix} 0 & 0 & 0 \\ \gamma & 0 & 0 \\ 1 + \alpha\left(\gamma\xi - 1\right) & \alpha & 0 \end{bmatrix}$$

$$b^T(\xi) = \begin{bmatrix} \frac{1}{4}\left[2 + \gamma\left(-1 + \xi\left(4 - \gamma + 2\alpha\left(\gamma\xi - 1\right)\right)\right)\right] \\ \frac{1}{4}\left[2 + \gamma\left(2\alpha\xi - 1\right)\right] \\ \frac{\gamma}{2} \end{bmatrix}$$

$$e(\xi) = \begin{bmatrix} 1 \\ 1 + \gamma\xi \\ 1 + \xi + \alpha\gamma\xi^2 \end{bmatrix}$$

$$\varphi(\xi) = 1 + \xi + \frac{\xi^2}{2} + \left(3 - 2\sqrt{2}\right)\alpha\xi^3$$

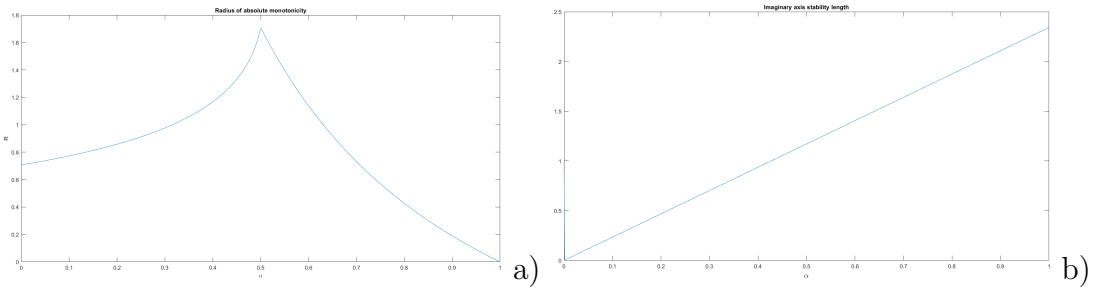
A method with tableaux $\left(A, b^T\right)$ is absolutely monotone at $\xi \in \mathbb{R}$ if $A(\xi) \geq 0$, $b^T(\xi) \geq 0$, $e(\xi) \geq 0$ and $\varphi(\xi) \geq 0$ elementwise; moreover the radius of absolute monotonicity is defined for all ξ in $-r \leq \xi \leq 0$ as

$$R(a, b) = \sup\left[r | r \geq 0, A(\xi) \geq 0, b^T(\xi) \geq 0, e(\xi) \geq 0, \varphi(\xi) \geq 0\right].$$

Figure 17 shows the behaviour of the radius of absolute monotonicity as α varies, along with the behaviour of the stability region along the imaginary axis. As already mentioned before, $\alpha = \frac{7-2\gamma}{6}$ was chosen originally to maximize the stability region, but in this case $R = \frac{2\sqrt{2}-3}{2+\sqrt{2}} \approx 0.05$, so that the region of absolute monotonicity is quite small. After some manipulations, it can be shown that the region of absolute stability is given by

$$S = \left\{z \in \mathbb{C} : \left|1 + z + \alpha\gamma z^2\right| < 1\right\}.$$

The alternative value $\alpha = 0.5$ maximizes the region of absolute monotonicity without compromising too much the stability. The impact of this alternative choice on numerical results is discussed in Section 6.



a) b)

Figure 17: Analysis of the explicit part of IMEX scheme: a) Radius of absolute monotonicity as function of α, b) Size of stability region along the imaginary axis as α varies.

B Isentropic processes for real gases

The potential temperature is an important thermodynamic quantity in density driven flows and can be easily derived for the ideal gas equation of state. In case of real gases, however, the derivation of quantities with similar properties is less straightforward. In this Appendix we analyze isentropic processes for the non-ideal equations of state considered in the work, for the purpose of deriving analogous quantities that are conserved under isentropic transformations. Let us recall the first law of thermodynamics

$$de = Tds - pdv = Tds + \frac{p}{\rho^2} d\rho \tag{52}$$

where s denotes the specific enthalpy. If we divide by T the previous relationship, we obtain

$$\frac{1}{T} de = ds + \frac{p}{\rho^2 T} d\rho \tag{53}$$

which in an isentropic process reduces to

$$\frac{1}{T} de - \frac{p}{\rho^2 T} d\rho = 0. \tag{54}$$

Thanks to (9) we then obtain :

$$\frac{R}{(\gamma - 1)\, T} dT - \frac{a\rho^2 + p}{\rho^2 T} d\rho = 0. \tag{55}$$

The EOS can be rewritten as [44]

$$T = \frac{\left(p + a\rho^2\right)(1 - \rho b)}{\rho R}. \tag{56}$$

If we substitute (56) into (55), we obtain

$$\frac{R}{(\gamma-1)\,T}dT - \frac{R}{\rho\,(1-\rho b)}d\rho \tag{57}$$

which can then be integrated to yield

$$\frac{R}{\gamma-1}\log(T) - 2R\,\mathrm{atanh}\,(2\rho b-1) = const \tag{58}$$

or, equivalently,

$$\log(T) - 2\,(\gamma-1)\,\mathrm{atanh}\,(2\rho b-1) = const. \tag{59}$$

The same computations can be repeated for the SG-EOS. Thanks to (11), we obtain for an isentropic process

$$\frac{R}{(\gamma-1)\,T}dT - \frac{\pi}{\rho^2 T}d\rho - \frac{p}{\rho^2 T}d\rho = 0. \tag{60}$$

Since $T = \frac{p+\pi}{\rho R}$, we obtain

$$\frac{R}{(\gamma-1)\,T}dT - \frac{R}{\rho}d\rho = 0, \tag{61}$$

which by integration yields

$$\frac{R}{\gamma-1}\log(T) - R\log(\rho) = const. \tag{62}$$

Thanks to the properties of the logarithm, we can rewrite

$$\log\left(\frac{T}{\rho^{\gamma-1}}\right) = const \tag{63}$$

and therefore

$$\frac{T}{\rho^{\gamma-1}} = const \tag{64}$$

which is equivalent to

$$\frac{p+\pi}{\rho^{\gamma}} = const. \tag{65}$$

References

[1] D.N. Arnold. An interior penalty finite element method with discontinuous elements. *SIAMNUM*, 19:742–760, 1982.

[2] W. Bangerth, R. Hartmann, and G. Kanschat. deal II: a general-purpose object-oriented finite element library. *ACM Transactions on Mathematical Software (TOMS)*, 33:24–51, 2007.

[3] R.E. Bank, W.M. Coughran, W. Fichtner, E.H. Grosse, Rose D.J., and R.K. Smith. Transient Simulation of Silicon Devices and Circuits. *IEEE Transactions on Electron Devices.*, 32:1992–2007, 1985.

[4] F. Bassi, L. Botti, A. Colombo, A. Ghidoni, and F. Massa. Linearly implicit Rosenbrock-type Runge–Kutta schemes applied to the Discontinuous Galerkin solution of compressible and incompressible unsteady flows. *Computers & Fluids*, 118, 2015.

[5] F. Bassi, A. Crivellini, D.A. Di Pietro, and S. Rebay. An implicit high-order discontinuous Galerkin method for steady and unsteady incompressible flows. *Computers & Fluids*, 36:1529–1546, 2007.

[6] L. Bonaventura. A semi-implicit, semi-Lagrangian scheme using the height coordinate for a nonhydrostatic and fully elastic model of atmospheric flows. *Journal of Computational Physics*, 158:186–213, 2000.

[7] L. Bonaventura and A. Della Rocca. Unconditionally strong stability preserving extensions of the TR-BDF2 method. *Journal of Scientific Computing*, 70:859–895, 2017.

[8] L. Bonaventura, R. Redler, and R. Budich. *Earth System Modelling 2: Algorithms, Code Infrastructure and Optimisation.* Springer Verlag, New York, 2012.

[9] W. Boscheri and L. Pareschi. High order pressure-based semi-implicit IMEX schemes for the 3D Navier-Stokes equations at all Mach numbers. *Journal of Computational Physics*, 434:110206, 2021.

[10] S. Busto, M. Tavelli, W. Boscheri, and M. Dumbser. Efficient high order accurate staggered semi-implicit discontinuous Galerkin methods for natural convection problems. *Computers & Fluids*, 198:104399, 2020.

[11] V. Casulli and D. Greenspan. Pressure method for the numerical solution of transient, compressible fluid flows. *International Journal for Numerical Methods in Fluids*, 4:1001–1012, 1984.

[12] M.J.P. Cullen. A test of a semi-implicit integration technique for a fully compressible non-hydrostatic model. *Quarterly Journal of the Royal Meteorological Society*, 116:1253–1258, 1990.

[13] M. Dumbser and V. Casulli. A conservative, weakly nonlinear semi-implicit finite volume scheme for the compressible navier- stokes equations with general equation of state. *Applied Mathematics and Computation*, 272:479–497, 2016.

[14] M. Dumbser, O. Zanotti, R. Loubère, and S. Diot. A posteriori subcell limiting of the discontinuous galerkin finite element method for hyperbolic conservation laws. *Journal of Computational Physics*, 278:47–75, 2014.

[15] N. Fehn, M. Kronbichler, C. Lehrenfeld, G. Lube, and P. Schroeder. High-order DG solvers for under-resolved turbulent incompressible flows: A comparison of l^2 and $h(div)$ methods. *International Journal of Numerical Methods in Fluids*, 91:533–556, 2019.

[16] L. Ferracina and M.N. Spijker. Stepsize restrictions for the Total-Variation-Diminishing property in general Runge-Kutta methods. *SIAM Journal of Numerical Analysis*, 42(3):1073–1093, 2004.

[17] U. Ghia, K.N. Ghia, and C.T. Shin. High-re solutions for incompressible flow using the navier-stokes equations and a multigrid method. *Journal of Computational Physics*, 48(3):387–411, 1982.

[18] F.X. Giraldo. Semi-implicit time-integrators for a scalable spectral element atmospheric model. *Quarterly Journal of the Royal Meteorological Society*, 131:2431–2454, 2005.

[19] F.X. Giraldo. *An Introduction to Element-Based Galerkin Methods on Tensor-Product Bases.* Springer Nature, 2020.

[20] F.X. Giraldo, J.F. Kelly, and E.M. Constantinescu. Implicit-explicit formulations of a three-dimensional nonhydrostatic unified model of the atmosphere (NUMA). *SIAM Journal of Scientific Computing*, 35:1162–1194, 2013.

[21] F.X. Giraldo, M. Restelli, and M. Läuter. Semi-implicit formulations of the Navier–Stokes equations: Application to nonhydrostatic atmospheric modeling. *SIAM Journal on Scientific Computing*, 32:3394–3425, 2010.

[22] S. Gottlieb, C.W. Shu, and E. Tadmor. Strong stability-preserving high-order time discretization methods. *SIAM Review*, 43(1):89–112, 2001.

[23] I. Higueras. On strong stability preserving time discretization methods. *Journal of Scientific Computing*, 21:193–223, 2004.

[24] M.E. Hosea and L.F: Shampine. Analysis and implementation of TR-BDF2. *Applied Numerical Mathematics*, 20:21–37, 1996.

[25] B. Janssen and G. Kanschat. Adaptive multilevel methods with local smoothing for H^1- and H^{curl}-conforming high order finite element methods. *SIAM Journal on Scientific Computing*, 33(4):2095–2114, 2011.

[26] G.E. Karniadakis and S. Sherwin. *Spectral hp−Element Methods for Computational Fluid Dynamics.* Oxford University Press, 2005.

[27] C.A. Kennedy and M.H. Carpenter. Additive Runge-Kutta schemes for convection-diffusion-reaction equations. *Applied Numerical Mathematics*, 44:139–181, 2003.

[28] J.F.B.M. Kraaijevanger. Contractivity of Runge-Kutta methods. *BIT*, 31:482–528, 1991.

[29] C. Kühnlein, W. Deconinck, R. Klein, S. Malardel, Z.P. Piotrowski, P.K. Smolarkiewicz, J. Szmelter, and N.P. Wedi. FVM 1.0: A nonhydrostatic finite-volume dynamical core formulation for IFS. *Geoscientific Model Development*, 12:651–676, 2019.

[30] R.J. LeVeque. *Finite volume methods for hyperbolic problems*. Cambridge University Press, 2002.

[31] C.D. Munz, S. Roller, R. Klein, and K.J. Geratz. The extension of incompressible flow solvers to the weakly compressible regime. *Computers and Fluids*, 32:173–196, 2003.

[32] O. Métayer and R. Saurel. The noble-abel stiffened-gas equation of state. *Physics of Fluids*, 28:046102, 04 2016.

[33] L. Pareschi and G. Russo. Implicit–explicit Runge–Kutta schemes and applications to hyperbolic systems with relaxation. *Journal of Scientific computing*, 25:129–155, 2005.

[34] M. Restelli. *Semi-Lagrangian and Semi-Implicit Discontinuous Galerkin Methods for Atmospheric Modeling Applications*. PhD thesis, Politecnico di Milano, 3 2007.

[35] M. Restelli and F.X. Giraldo. A conservative Discontinuous Galerkin semi-implicit formulation for the Navier-Stokes equations in nonhydrostatic mesoscale modeling. *SIAM Journal of Scientific Computing*, 31:2231–2257, 2009.

[36] A. Robert. A semi-Lagrangian and semi-implicit numerical integration scheme for the primitive meteorological equations. *Journal of the Meteorological Society of Japan*, 60:319–325, 1982.

[37] P.K. Smolarkiewicz, C. Kühnlein, and N.P. Wedi. Semi-implicit integrations of perturbation equations for all-scale atmospheric dynamics. *Journal of Computational Physics*, 376:145–159, 2019.

[38] G.A. Sod. A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws. *Journal of Computational Physics*, 27(1):1–31, 1978.

[39] J. Steppeler, R. Hess, G. Doms, U. Schättler, and L. Bonaventura. Review of numerical methods for nonhydrostatic weather prediction models. *Meteorology and Atmospheric Physics*, 82:287–301, 2003.

[40] G. Strang. On the construction and comparison of difference schemes. *SIAM Journal on Numerical Analysis*, 5:506–517, 1968.

[41] M. Tavelli and M. Dumbser. A pressure-based semi-implicit space–time discontinuous Galerkin method on staggered unstructured meshes for

the solution of the compressible Navier–Stokes equations at all Mach numbers. *Journal of Computational Physics*, 341:341–376, 2017.

[42] G. Tumolo. A mass conservative TR-BDF2 semi-implicit semi-Lagrangian DG discretization of the shallow water equations on general structured meshes of quadrilaterals. *Communications in Applied and Industrial Mathematics*, 7:165–190, 2016.

[43] G. Tumolo and L. Bonaventura. A semi-implicit, semi-Lagrangian discontinuous Galerkin framework for adaptive numerical weather prediction: SISL-DG framework for adaptive NWP. *Quarterly Journal of the Royal Meteorological Society*, 141:2582–2601, 2015.

[44] J. Vidal. Thermodynamics. application to chemical engineering and petroleum industry., Dec 1997.

[45] J. Zeifang, J. Schütz, K. Kaiser, A. Beck, M. Lukáčová-Medvid'ová, and S. Noelle. A novel full-euler low mach number imex splitting. *Communications in Computational Physics*, 27(1):292–320, 2019.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

73/2021 Marcinno, F.; Zingaro, A.; Fumagalli, I.; Dede', L.; Vergara, C.
A computational study of blood flow dynamics in the pulmonary arteries

70/2021 Beirao da Veiga, L.; Canuto, C.; Nochetto, R.H.; Vacca, G.; Verani, M.
Adaptive VEM: Stabilization-Free A Posteriori Error Analysis

71/2021 Franco, N.; Manzoni, A.; Zunino, P.
A Deep Learning approach to Reduced Order Modelling of parameter dependent Partial Differential Equations

72/2021 Fresca, S.; Manzoni, A.
POD-DL-ROM: enhancing deep learning-based reduced order models for nonlinear parametrized PDEs by proper orthogonal decomposition

69/2021 Antonietti, P.F.; Caldana, M.; Dede', L.
Accelerating Algebraic Multigrid Methods via Artificial Neural Networks

65/2021 Mazzieri, I.; Muhr, M.; Stupazzini, M.; Wohlmuth, B.
Elasto-acoustic modelling and simulation for the seismic response of structures: The case of the Tahtali dam in the 2020 Izmir earthquake

67/2021 Salvador, M.; Regazzoni, F.; Pagani, S.; Dede', L.; Trayanova, N.; Quarteroni, A.
The role of mechano-electric feedbacks and hemodynamic coupling in scar-related ventricular tachycardia

68/2021 Regazzoni, F.; Salvador, M.; Dede', L.; Quarteroni, A.
A machine learning method for real-time numerical simulations of cardiac electromechanics

66/2021 Antonietti, P.F.; Botti, M.; Mazzieri, I.
On mathematical and numerical modelling of multiphysics wave propagation with polygonal Discontinuous Galerkin methods

63/2021 Rosafalco, L.; Torzoni, M.; Manzoni, A.; Mariani, S.; Corigliano, A.
Online structural health monitoring by model order reduction and deep learning algorithms