A Conservative Finite Element Method for the Incompressible Euler Equations with Variable Density

Evan S. Gawlik* and François Gay-Balmaz†

Abstract

We construct a finite element discretization and time-stepping scheme for the incompressible Euler equations with variable density that exactly preserves total mass, total squared density, total energy, and pointwise incompressibility. The method uses Raviart-Thomas or Brezzi-Douglas-Marini finite elements to approximate the velocity and discontinuous polynomials to approximate the density and pressure. To achieve exact preservation of the aforementioned conserved quantities, we exploit a seldom-used weak formulation of the momentum equation and a second-order time-stepping scheme that is similar, but not identical, to the midpoint rule. We also describe and prove stability of an upwinded version of the method. We present numerical examples that demonstrate the order of convergence of the method.

1 Introduction

This paper considers the incompresible Euler equations with variable density on a bounded domain $\Omega \subset \mathbb{R}^n$, $n \in \{2, 3\}$. These equations seek a velocity field u, density ρ, and pressure p such that

$$\rho(\partial_t u + u \cdot \nabla u) = -\nabla p, \qquad \text{in } \Omega \times (0,T), \tag{1}$$
$$\partial_t \rho + \operatorname{div}(\rho u) = 0, \qquad \text{in } \Omega \times (0,T), \tag{2}$$
$$\operatorname{div} u = 0, \qquad \text{in } \Omega \times (0,T), \tag{3}$$
$$u \cdot n = 0, \qquad \text{on } \partial\Omega \times (0,T), \tag{4}$$
$$u(0) = u_0,\ \rho(0) = \rho_0, \qquad \text{in } \Omega, \tag{5}$$

where $u_0 : \Omega \to \mathbb{R}^n$ and $\rho_0 : \Omega \to \mathbb{R}$ are given, $\operatorname{div} u_0 = 0$, and $u_0 \cdot n = 0$ on $\partial\Omega$. The aim of this paper is to construct a finite element method for (1-3) that exactly conserves

$$\int_\Omega \rho \, dx, \quad \int_\Omega \rho^2 \, dx, \quad \int_\Omega \rho u \cdot u \, dx, \tag{6}$$

and the incompressibility constraint $\operatorname{div} u = 0$ at the (spatially and temporally) discrete level.

Several authors have considered related tasks in the past. Guermond and Quartapelle [16] have constructed a spatial discretization of the incompressible Navier-Stokes equations that preserves the three invariants (6) in the limit of vanishing viscosity. However, their temporal discretization does not preserve total (kinetic) energy nor total squared density. Moreover, the incompressibility constraint is only satisfied in a weak sense, and their construction relies on an evolution equation for $\sqrt{\rho}u$ – a quantity that lacks physical meaning. Another method that is closely related to ours

*Department of Mathematics, University of Hawaii at Manoa, egawlik@hawaii.edu
†CNRS - LMD, Ecole Normale Supérieure, francois.gay-balmaz@lmd.ens.fr

is the $H(\mathrm{div})$-conforming finite element spatial discretization of the incompressible Euler equations with constant density studied in [18, 25]. When $\rho \equiv 1$, our spatial discretization reduces to the one used there, and our temporal discretization reduces to the midpoint rule used in [25]. Lastly, the form of the momentum equation that we discretize in our method bears some resemblance to the one used in [23] for the compressible rotating shallow water equations, but with $\partial_t(\rho u)$ appearing rather than $\partial_t u$.

More broadly, conservative numerical methods for incompressible flow—with and without variable density—have received considerable attention over the last several decades. Energy conservation is widely considered to be important when performing simulations over long time spans, and many efforts have been devoted to achieving exact conservation of energy and other invariants for inviscid flow at the discrete level [2, 3, 9, 15, 24, 26, 27, 30, 32]; see the introduction of [26] for a more detailed history. Even schemes that do not preserve energy are, by and large, designed with energetic considerations in mind, since energy growth is tantamount to instability [22]. The role played by the incompressibility constraint $\operatorname{div} u = 0$ is equally important. Often it is enforced weakly [22, 30] or in a least squares sense [19, 28], but doing so can give rise to numerical artifacts. For example, finite element methods that violate pointwise incompressibility often violate pressure-robustness; the numerically computed velocity and pressure both change when a force $\nabla\phi$ is added to the right-hand side of the momentum equation, even though only the pressure should change [20]. In contrast, exactly divergence-free finite element methods are automatically pressure-robust [20]. Inexact enforcement of incompressibility can also interfere with the design of energy-conserving schemes, since proofs of energy conservation often rely on vector calculus identities that hold only for solenoidal vector fields. Finally, enforcing $\operatorname{div} u = 0$ weakly has the added disadvantage that $\operatorname{div} u$ may be far from zero on coarse meshes. For these reasons, there is growing interest in methods that enforce incompressibility pointwise [10, 12, 13, 18, 21, 31].

Despite the above considerations, we are not aware of any methods for the incompressible Euler equations with variable density that enforce incompressibility pointwise and conserve energy exactly at the spatially and temporally discrete level. Several schemes for (1-3) have been designed that enforce incompressibility approximately and adopt fractional step methods for the temporal discretization [1, 4, 8, 16, 17, 29]. These methods often update the velocity field by constructing a tentative, non-solenoidal velocity field, and then projecting onto a space of discretely divergence-free vector fields. Many of these methods are energy-stable, but not energy-conserving [16, 17, 29].

To motivate our numerical method, we first use the identity

$$\rho u \cdot \nabla u = \nabla(\rho u \cdot u) - u \times (\nabla \times (\rho u)) - (u \cdot \nabla \rho) u - \frac{1}{2}\rho \nabla (u \cdot u)$$

and equations (2-3) to write (1) in the form

$$\partial_t(\rho u) + \nabla(\rho u \cdot u) - u \times (\nabla \times (\rho u)) - \frac{1}{2}\rho \nabla (u \cdot u) = -\nabla p, \tag{7}$$

Next, we multiply (7), (2), and (3) by test functions v, σ, and q, respectively, and integrate by parts. Using the identity

$$\int_\Omega (\nabla(w \cdot u) - u \times (\nabla \times w)) \cdot v \, dx = \int_\Omega w \cdot (v \cdot \nabla u - u \cdot \nabla v) \, dx, \quad \text{if } \operatorname{div} u = 0, \tag{8}$$
$$\text{and } \; u \cdot n|_{\partial\Omega} = 0,$$

we arrive at the following observation. For every smooth vector field v satisfying $v \cdot n|_{\partial\Omega} = 0$ and every pair of smooth scalar fields σ and q, the solution (u, ρ, p) of (1-5) satisfies

$$\langle \partial_t(\rho u), v\rangle + a(\rho u, u, v) - \frac{1}{2} b(v, u \cdot u, \rho) = \langle p, \operatorname{div} v\rangle, \tag{9}$$

$$\langle \partial_t \rho, \sigma\rangle - b(u, \sigma, \rho) = 0, \tag{10}$$

$$\langle \operatorname{div} u, q\rangle = 0, \tag{11}$$

where $\langle u, v\rangle = \int_\Omega u \cdot v \, dx$ for vector fields u and v, $\langle f, g\rangle = \int_\Omega fg \, dx$ for scalar fields f and g, and

$$a(w, u, v) = \langle w, v \cdot \nabla u - u \cdot \nabla v\rangle,$$
$$b(w, f, g) = \langle w \cdot \nabla f, g\rangle.$$

Note that (9-11) can alternatively be derived from Hamilton's principle of least action; we refer the reader to [14] for details.

One advantage of the formulation (9-11) is that the conserved quantities (6) are straightforward consequences of two basic properties of the trilinear forms a and b. Namely, a is alternating in its last two arguments,

$$a(w, u, v) = -a(w, v, u), \tag{12}$$

and b is alternating in its last two arguments when its first argument is divergence-free:

$$b(w, f, g) = -b(w, g, f) \text{ if } \operatorname{div} w = 0 \text{ and } w \cdot n|_{\partial\Omega} = 0. \tag{13}$$

With these properties in mind, we take $\sigma = 1$ in the density equation (10) to deduce conservation of total mass:

$$\frac{d}{dt}\int_\Omega \rho \, dx = \langle \partial_t \rho, 1\rangle = b(u, 1, \rho) = 0.$$

If instead we take $\sigma = \rho$ in (10) and use (13), we deduce conservation of total squared density:

$$\frac{d}{dt}\frac{1}{2}\int_\Omega \rho^2 \, dx = \langle \partial_t \rho, \rho\rangle = b(u, \rho, \rho) = 0.$$

Finally, taking $v = u$ in the momentum equation (9) gives conservation of total (kinetic) energy:

$$\begin{aligned}\frac{d}{dt}\frac{1}{2}\int_\Omega \rho u \cdot u \, dx &= \langle \partial_t(\rho u), u\rangle - \frac{1}{2}\langle \partial_t \rho, u \cdot u\rangle \\ &= \langle p, \operatorname{div} u\rangle - a(\rho u, u, u) + \frac{1}{2} b(u, u \cdot u, \rho) - \frac{1}{2}\langle \partial_t \rho, u \cdot u\rangle \\ &= 0.\end{aligned}$$

Here, we have used the fact that $\operatorname{div} u = 0$, a is alternating in its last two arguments, and (10) holds. Our numerical method will inherit these conservation laws by using discretizations of a and b that satisfy analogues of (12) and (13).

2 Spatial Discretization

In this section, we propose a finite element spatial discretization of (9-11).

Let $\mathcal{T}_h$ be a triangulation of Ω. We denote by $\mathcal{E}_h$ the set of interior $(n-1)$-dimensional faces in $\mathcal{T}_h$ (edges in two dimensions). For each integer $s \geq 0$ and each simplex $K \in \mathcal{T}_h$, we denote by $P_s(K)$ the space of polynomials of degree at most s on K. We denote

$$H_0(\operatorname{div}, \Omega) = \{u \in L^2(\Omega)^n \mid \operatorname{div} u \in L^2(\Omega),\, u \cdot n = 0 \text{ on } \partial\Omega\}$$

and

$$\mathring{H}(\operatorname{div}, \Omega) = \{u \in H_0(\operatorname{div}, \Omega) \mid \operatorname{div} u = 0\}.$$

Our numerical method will make use of three approximation spaces: a space $U_h \subset H_0(\operatorname{div}, \Omega)$ for the velocity u, a space $F_h \subset L^2(\Omega)$ for the density ρ, and a space $Q_h \subset L^2_{\int=0}(\Omega) = \{q \in L^2(\Omega) \mid \int_\Omega q\, dx = 0\}$ for the pressure p. For the velocity, we use either the Raviart-Thomas space

$$RT_s(\mathcal{T}_h) = \{u \in H_0(\operatorname{div}, \Omega) \mid u|_K \in P_s(K)^n + xP_s(K),\, \forall K \in \mathcal{T}_h\}$$

or the Brezzi-Douglas-Marini space

$$BDM_{s+1}(\mathcal{T}_h) = \{u \in H_0(\operatorname{div}, \Omega) \mid u|_K \in P_{s+1}(K)^n,\, \forall K \in \mathcal{T}_h\},$$

where $s \geq 0$ is an integer. For the pressure, we use the zero-mean subspace of the discontinuous Galerkin space

$$DG_s(\mathcal{T}_h) = \{f \in L^2(\Omega) \mid f|_K \in P_s(K),\, \forall K \in \mathcal{T}_h\}.$$

For the density, we use $DG_m(\mathcal{T}_h)$, where $m \geq 0$ is an integer (not necessarily equal to s). In summary,

$$U_h \in \{RT_s(\mathcal{T}_h), BDM_{s+1}(\mathcal{T}_h)\}, \tag{14}$$

$$F_h = DG_m(\mathcal{T}_h), \tag{15}$$

$$Q_h = DG_s(\mathcal{T}_h) \cap L^2_{\int=0}(\Omega). \tag{16}$$

We also denote by W_h the (infinite-dimensional) space of vector fields on Ω that are piecewise smooth with respect to $\mathcal{T}_h$.

On an edge $e = K_1 \cap K_2 \in \mathcal{E}_h$, we denote the jump and average of a scalar function $f \in F_h$ by

$$[\![f]\!] = f_1 n_1 + f_2 n_2, \quad \{f\} = \frac{f_1 + f_2}{2},$$

where $f_i = f|_{K_i}$, n_1 is the the normal vector to e pointing from K_1 to K_2, and similarly for n_2. To define the jump and average of a vector field $u \in W_h$, we fix a choice of normal vector n for each edge $e \in \mathcal{E}_h$. Next, we determine K_1 and K_2 so that $e = K_1 \cap K_2$ and n points from K_1 to K_2, and we define

$$[\![u]\!] = u_1 - u_2, \quad \{u\} = \frac{u_1 + u_2}{2}.$$

We define trilinear forms $a_h : W_h \times U_h \times U_h \to \mathbb{R}$ and $b_h : U_h \times F_h \times F_h \to \mathbb{R}$ by

$$a_h(w, u, v) = \sum_{K \in \mathcal{T}_h} \int_K w \cdot (v \cdot \nabla u - u \cdot \nabla v)\, dx + \sum_{e \in \mathcal{E}_h} \int_e (n \times \{w\}) \cdot [\![u \times v]\!]\, ds,$$

$$b_h(u, f, g) = \sum_{K \in \mathcal{T}_h} \int_K (u \cdot \nabla f) g\, dx - \sum_{e \in \mathcal{E}_h} \int_e u \cdot [\![f]\!] \{g\}\, ds.$$

The trilinear forms a_h and b_h are well known: a_h has been used in discretizations the incompressible Euler equations with constant density [18, 25], and b_h is a standard discontinous Galerkin discretization of the scalar advection operator [6].

To understand these definitions, observe that if $u \in H_0(\mathrm{div}, \Omega)$, f is piecewise smooth with respect to $\mathcal{T}_h$, and g is smooth, then

$$\begin{aligned}
-\int_\Omega f \,\mathrm{div}(gu) &= -\sum_{K\in\mathcal{T}_h} \int_K f \,\mathrm{div}(gu)\,dx \\
&= \sum_{K\in\mathcal{T}_h} \left(\int_K (u\cdot\nabla f) g\,dx - \int_{\partial K} (u\cdot n) f g\,ds \right) \\
&= \sum_{K\in\mathcal{T}_h} \int_K (u\cdot\nabla f) g\,dx - \sum_{e\in\mathcal{E}_h} \int_e u\cdot[\![f]\!] g\,ds \\
&= \sum_{K\in\mathcal{T}_h} \int_K (u\cdot\nabla f) g\,dx - \sum_{e\in\mathcal{E}_h} \int_e u\cdot[\![f]\!]\{g\}\,ds,
\end{aligned}$$

where the last line holds because g is smooth. This shows that $b_h(u,\sigma,\rho)$ and $b_h(v, u\cdot u, \rho)$ are consistent discretizations of $-\int_\Omega \sigma\,\mathrm{div}(\rho u)\,dx$ and $-\int_\Omega (u\cdot u)\,\mathrm{div}(\rho v)\,dx = \int_\Omega \rho\nabla(u\cdot u)\cdot v\,dx$. Likewise, if w and u are smooth, $\mathrm{div}\, u = 0$, $u\cdot n|_{\partial\Omega} = 0$, and $v \in H_0(\mathrm{div},\Omega)\cap W_h$, then

$$\begin{aligned}
\int_\Omega (\nabla(w\cdot u) - u\times(\nabla\times w))\cdot v\,dx &= \sum_{K\in\mathcal{T}_h} \int_K (\nabla(w\cdot u) - u\times(\nabla\times w))\cdot v\,dx \\
&= \sum_{K\in\mathcal{T}_h} \left(\int_K w\cdot(v\cdot\nabla u - u\cdot\nabla v)\,dx + \int_{\partial K} (w\cdot v)(u\cdot n)\,ds \right) \\
&= \sum_{K\in\mathcal{T}_h} \left(\int_K w\cdot(v\cdot\nabla u - u\cdot\nabla v)\,dx + \int_{\partial K} (n\times w)\cdot(u\times v)\,ds + \int_{\partial K} (v\cdot n)(w\cdot u)\,ds \right) \\
&= \sum_{K\in\mathcal{T}_h} \int_K w\cdot(v\cdot\nabla u - u\cdot\nabla v)\,dx + \sum_{e\in\mathcal{E}_h} \int_e (n\times w)\cdot[\![u\times v]\!]\,ds + \sum_{e\in\mathcal{E}_h} \int_e v\cdot[\![w\cdot u]\!]\,ds \\
&= \sum_{K\in\mathcal{T}_h} \int_K w\cdot(v\cdot\nabla u - u\cdot\nabla v)\,dx + \sum_{e\in\mathcal{E}_h} \int_e (n\times\{w\})\cdot[\![u\times v]\!]\,ds,
\end{aligned}$$

where the last line holds because w and u are smooth and $v\cdot n$ is continuous across element interfaces. This shows that $a_h(\rho u, u, v)$ is a consistent discretization of $\int_\Omega (\nabla(\rho u\cdot u) - u\times(\nabla\times(\rho u)))\cdot v\,dx$.

The trilinear forms a_h and b_h possess several properties that play an essential role in the conservative nature of our numerical method. First, a_h is alternating in its last two arguments:

$$a_h(w,u,v) = -a_h(w,v,u), \quad \forall (w,u,v) \in W_h\times U_h\times U_h.$$

Second, using integration by parts, one checks that b_h is also alternating in its last two arguments if its first argument is divergence-free:

$$b_h(u,f,g) = -b_h(u,g,f), \quad \forall (u,f,g) \in (U_h\cap \mathring{H}(\mathrm{div},\Omega))\times F_h\times F_h. \tag{17}$$

As a consequence, we have

$$a_h(w,u,u) = 0, \qquad \forall (w,u)\in W_h\times U_h, \tag{18}$$

$$b_h(u,f,f) = 0, \qquad \forall (u,f)\in (U_h\cap\mathring{H}(\mathrm{div},\Omega))\times F_h. \tag{19}$$

We are now ready to define our semidiscrete numerical method. We first describe the simplest version of the method, which includes no upwinding and exactly preserves total mass, total squared density, total energy, and pointwise incompressibility. It seeks $u_h(t) \in U_h$, $\rho_h(t) \in F_h$, and $p_h(t) \in Q_h$ such that

$$\langle \partial_t(\rho_h u_h), v_h\rangle + a_h(\rho_h u_h, u_h, v_h) - \frac{1}{2} b_h(v_h, \overline{u_h \cdot u_h}, \rho_h) = \langle p_h, \operatorname{div} v_h\rangle, \qquad \forall v_h \in U_h, \tag{20}$$

$$\langle \partial_t \rho_h, \sigma_h\rangle - b_h(u_h, \sigma_h, \rho_h) = 0, \qquad \forall \sigma_h \in F_h, \tag{21}$$

$$\langle \operatorname{div} u_h, q_h\rangle = 0, \qquad \forall q_h \in Q_h. \tag{22}$$

Here, $\overline{f} \in F_h$ denotes the L^2-orthogonal projection of $f \in L^2(\Omega)$ onto F_h:

$$\langle \overline{f}, g_h\rangle = \langle f, g_h\rangle, \quad \forall g_h \in F_h.$$

Our use of $\overline{u_h \cdot u_h}$ rather than $u_h \cdot u_h$ in (20) is motivated by energetic considerations. As will be seen in Proposition 2.2, energy conservation is deduced by taking $v_h = u_h$ and $\sigma_h = \overline{u_h \cdot u_h}$ as test functions in (20-21) and combining the two equations. The naive alternative—taking $\sigma_h = u_h \cdot u_h$ in (21)—is generally not possible unless it happens that $u_h \cdot u_h \in F_h$.

Remark 2.1. If either

$$(m \geq 2s \text{ and } U_h = RT_s(\mathcal{T}_h)), \quad \text{or} \quad (m \geq 2s+2 \text{ and } U_h = BDM_{s+1}(\mathcal{T}_h)), \tag{23}$$

then $\overline{u_h \cdot u_h} = u_h \cdot u_h$. Indeed, u_h belongs to the divergence-free subspace of $U_h \in \{RT_s(\mathcal{T}_h), BDM_{s+1}(\mathcal{T}_h)\}$, which consists of piecewise polynomial vector fields of degree at most s (if $U_h = RT_s(\mathcal{T}_h)$) or $s+1$ (if $U_h = BDM_{s+1}(\mathcal{T}_h)$) [5, p. 116]. Thus, $u_h \cdot u_h|_K \in P_m(K)$ for every $K \in \mathcal{T}_h$ if (23) holds. In particular, $\overline{u_h \cdot u_h} = u_h \cdot u_h$ in the lowest-order version of this finite element method ($s = m = 0$ and $U_h = RT_0(\mathcal{T}_h)$). If the condition (23) is violated, the computation of $\overline{u_h \cdot u_h}$ is not a significant expense since it can be computed element by element. Note that an analogous projection of the squared fluid velocity appears in other conservative numerical methods for fluids; see, for instance, [23, p. 14].

As with most finite element methods for advection-dominated problems, solutions of (20-22) can often exhibit unphysical oscillations if upwinding is not incorporated into the discretization [11]. If upwinding is desired, we propose the following generalization of (20-22): Seek $u_h(t) \in U_h$, $\rho_h(t) \in F_h$, and $p_h(t) \in Q_h$ such that

$$\sum_{e \in \mathcal{E}_h} \int_e \left(\alpha_e(u_h)(n \times [\![\rho_h u_h]\!]) \cdot [\![u_h \times v_h]\!] + \frac{\beta_e(u_h)}{2|u_h \cdot n|^2}(u_h \cdot n)(v_h \cdot n)[\![\overline{u_h \cdot u_h}]\!] \cdot [\![\rho_h]\!] \right) ds$$

$$+ \langle \partial_t(\rho_h u_h), v_h\rangle + a_h(\rho_h u_h, u_h, v_h) - \frac{1}{2} b_h(v_h, \overline{u_h \cdot u_h}, \rho_h) - \langle p_h, \operatorname{div} v_h\rangle = 0, \quad \forall v_h \in U_h, \tag{24}$$

$$\langle \partial_t \rho_h, \sigma_h\rangle - b_h(u_h, \sigma_h, \rho_h) + \sum_{e \in \mathcal{E}_h} \int_e \beta_e(u_h) [\![\sigma_h]\!] \cdot [\![\rho_h]\!] \, ds = 0, \quad \forall \sigma_h \in F_h, \tag{25}$$

$$\langle \operatorname{div} u_h, q_h\rangle = 0, \quad \forall q_h \in Q_h. \tag{26}$$

Here, $\{\alpha_e\}_{e \in \mathcal{E}_h}$ and $\{\beta_e\}_{e \in \mathcal{E}_h}$ are nonnegative parameters which may depend on u_h. Standard choices for α_e and β_e are [6, 25]

$$\alpha_e(u_h) = c_1 \frac{u_h \cdot n}{|u_h \cdot n|}, \quad \beta_e(u_h) = c_2 |u_h \cdot n|, \quad c_1, c_2 \in \left[0, \tfrac{1}{2}\right],$$

where the choice $c_1 = c_2 = \frac{1}{2}$ corresponds to full upwinding. To understand why, observe that if $\beta_e(u_h) = \frac{1}{2}|u_h \cdot n|$, then

$$(u_h\{\rho_h\} + \beta_e(u_h)[\![\rho_h]\!]) \cdot n_1 = \begin{cases} \rho_{h1} u_h \cdot n_1 & \text{if } u_h \cdot n_1 > 0, \\ \rho_{h2} u_h \cdot n_1 & \text{if } u_h \cdot n_1 < 0, \end{cases}$$

so the upwinding in (25) has the effect of replacing $\{\rho_h\}$ with its upwind value (either ρ_{h1} or ρ_{h2}) in the expression (17) for $b_h(u_h, \sigma_h, \rho_h)$. The effect of choosing $\alpha_e(u_h) = \frac{1}{2}\frac{u_h \cdot n}{|u_h \cdot n|}$ is similar. Note that an additional term involving β_e has been added to the momentum equation (24) to counteract any energy imbalance introduced by the terms involving β_e in the density equation (25). When ρ_h is constant, the terms involving β_e vanish, and we recover the momentum upwinding strategy proposed in [25].

2.1 Properties of the Spatial Discretization

Although (26) imposes incompressibility weakly, our choice of finite element spaces (14-16) ensures that the velocity field determined by (24-26) is divergence-free pointwise.

Proposition 2.1. *For every t, $\operatorname{div} u_h \equiv 0$.*

Proof. Since $U_h \in \{RT_s(\mathcal{T}_h), BDM_{s+1}(\mathcal{T}_h)\}$ and $\int_\Omega \operatorname{div} u_h \, dx = \int_{\partial\Omega} u_h \cdot n \, ds = 0$, we have $\operatorname{div} u_h \in DG_s(\mathcal{T}_h) \cap L^2_{\int=0}(\Omega) = Q_h$, so we may take $q_h = \operatorname{div} u_h$ in (26). □

The following proposition shows that the numerical method (24-26) exactly preserves total mass, total energy, and, if $\beta_e = 0$, total squared density.

Proposition 2.2. *For every t, we have*

$$\frac{d}{dt}\int_\Omega \rho_h \, dx = 0, \tag{27}$$

$$\frac{d}{dt}\int_\Omega \rho_h^2 \, dx \le 0, \ \textit{with equality if } \beta_e = 0, \ \forall e \in \mathcal{E}_h, \tag{28}$$

$$\frac{d}{dt}\int_\Omega \rho_h u_h \cdot u_h \, dx = 0. \tag{29}$$

Proof. Taking $\sigma_h \equiv 1$ in (25) gives

$$\frac{d}{dt}\int_\Omega \rho_h \, dx = \langle \partial_t \rho_h, 1\rangle = b_h(u_h, 1, \rho_h) - \sum_{e\in\mathcal{E}_h}\int_e \beta_e(u_h)[\![1]\!]\cdot[\![\rho_h]\!]\, ds = 0.$$

Taking $\sigma_h = \rho_h$ in (25) and invoking (19) gives

$$\frac{d}{dt}\frac{1}{2}\int_\Omega \rho_h^2 \, dx = \langle \partial_t \rho_h, \rho_h\rangle = b_h(u_h, \rho_h, \rho_h) - \sum_{e\in\mathcal{E}_h}\int_e \beta_e(u_h)[\![\rho_h]\!]\cdot[\![\rho_h]\!]\, ds \le 0.$$

To prove (29), we take $v_h = u_h$ in (24) and invoke the definition of the L^2-projection to write

$$\begin{aligned}
&\frac{d}{dt}\frac{1}{2}\int_\Omega \rho_h u_h \cdot u_h \, dx \\
&\quad = \langle \partial_t(\rho_h u_h), u_h\rangle - \frac{1}{2}\langle \partial_t \rho_h, u_h \cdot u_h\rangle \\
&\quad = \langle p_h, \operatorname{div} u_h\rangle - a_h(\rho_h u_h, u_h, u_h) + \frac{1}{2} b_h(u_h, \overline{u_h \cdot u_h}, \rho_h) - \sum_{e\in\mathcal{E}_h}\int_e \frac{1}{2}\beta_e(u_h)[\![\overline{u_h\cdot u_h}]\!]\cdot[\![\rho_h]\!]\, ds \\
&\qquad - \frac{1}{2}\langle \partial_t \rho_h, \overline{u_h \cdot u_h}\rangle.
\end{aligned}$$

The first two terms above vanish by (26) and (18). The last three terms vanish according to the density evolution equation (25). It follows that (29) holds. □

2.2 Inhomogeneous Boundary Conditions

Let us discuss how to handle more general boundary conditions than (4). Suppose that the boundary condition $u \cdot n = 0$ is replaced by $u \cdot n = g$, where $g : \partial\Omega \times (0,T) \to \mathbb{R}$ is a given function satisfying $\int_{\partial\Omega} g\, ds = 0$ for every t. Assume that the sets $\Gamma_-(t) = \{x \in \partial\Omega \mid g(x,t) < 0\}$ and $\Gamma_+(t) = \{x \in \partial\Omega \mid g(x,t) > 0\}$ do not vary with t. Then we must prescribe the values of ρ on Γ_-, leading to boundary conditions

$$\begin{aligned}
u \cdot n &= g, \text{ on } \partial\Omega, \\
\rho &= \theta, \text{ on } \Gamma_-,
\end{aligned}$$

where $\theta : \Gamma_- \times (0,T) \to \mathbb{R}$ is a given positive function. To handle these boundary conditions, we note that the identity (8) becomes

$$\int_\Omega (\nabla(w\cdot u) - u \times (\nabla \times w)) \cdot v \, dx = \int_\Omega w \cdot (v \cdot \nabla u - u \cdot \nabla v)\, dx + \int_{\partial\Omega} g w \cdot v,$$

if $\operatorname{div} u = 0$ and $u \cdot n|_{\partial\Omega} = g$. By accounting for the extra term above, and by accounting for boundary terms in the density equation, we arrive at the following spatial discretization:

$$\begin{aligned}
\langle \partial_t(\rho_h u_h), v_h\rangle + a_h(\rho_h u_h, u_h, v_h) - \frac{1}{2} b_h(v_h, \overline{u_h\cdot u_h}, \rho_h) + \int_{\partial\Omega} g\rho_h u_h \cdot v_h \, ds &= \langle p_h, \operatorname{div} v_h\rangle, && \forall v_h \in U_h, \\
\langle \partial_t \rho_h, \sigma_h\rangle - b_h(u_h, \sigma_h, \rho_h) + \int_{\Gamma_-} g\theta\sigma_h + \int_{\Gamma_+} g\rho_h\sigma_h &= 0, && \forall \sigma_h \in F_h, \\
\langle \operatorname{div} u_h, q_h\rangle &= 0, && \forall q_h \in Q_h.
\end{aligned}$$

Here, $\rho_h(t)$ and $p_h(t)$ are sought within F_h, and Q_h, respectively, whereas $u_h(t)$ is sought within $U_h + w(t)$, where $w(t) : \Omega \to \mathbb{R}^n$ is a vector field satisfying $w \cdot n = g$ on $\partial\Omega$ (or a suitable interpolant thereof).

3 Temporal Discretization

We now propose a temporal discretization of (24-26). To reduce notational clutter, we supress the subscript h when referring to functions that belong to finite element spaces in this section. We also suppress the subscript h on a_h and b_h.

Our temporal discretization seeks $u_1, u_2, \ldots \in U_h$, $\rho_1, \rho_2, \ldots \in F_h$, and $p_1, p_2, \ldots \in Q_h$ such that for every k and every $(v, \sigma, q) \in U_h \times F_h \times Q_h$,

$$\sum_{e \in \mathcal{E}_h} \int_e \alpha_e(u_{k+1/2})(n \times [\![(\rho u)_{k+1/2}]\!]) \cdot [\![u_{k+1/2} \times v]\!]\, ds$$

$$+ \sum_{e \in \mathcal{E}_h} \int_e \frac{\beta_e(u_{k+1/2})}{2|u_{k+1/2} \cdot n|^2} (u_{k+1/2} \cdot n)(v \cdot n) [\![\overline{u_k \cdot u_{k+1}}]\!] \cdot [\![\rho_{k+1/2}]\!]\, ds \tag{30}$$

$$+ \left\langle \frac{\rho_{k+1} u_{k+1} - \rho_k u_k}{\Delta t}, v \right\rangle + a((\rho u)_{k+1/2}, u_{k+1/2}, v) - \frac{1}{2} b\left(v, \overline{u_k \cdot u_{k+1}}, \rho_{k+1/2}\right) - \langle p_{k+1}, \operatorname{div} v \rangle = 0,$$

$$\left\langle \frac{\rho_{k+1} - \rho_k}{\Delta t}, \sigma \right\rangle - b(u_{k+1/2}, \sigma, \rho_{k+1/2}) + \sum_{e \in \mathcal{E}_h} \int_e \beta_e(u_{k+1/2}) [\![\sigma]\!] \cdot [\![\rho_{k+1/2}]\!]\, ds = 0, \tag{31}$$

$$\langle \operatorname{div} u_{k+1}, q \rangle = 0, \tag{32}$$

where $\Delta t > 0$ is a time step, $u_{k+1/2} = \frac{u_k + u_{k+1}}{2}$, $\rho_{k+1/2} = \frac{\rho_k + \rho_{k+1}}{2}$, and

$$(\rho u)_{k+1/2} = \frac{\rho_k u_k + \rho_{k+1} u_{k+1}}{2}. \tag{33}$$

In the case where $\alpha_e = \beta_e = 0$ for every $e \in \mathcal{E}_h$, the scheme reduces to

$$\left\langle \frac{\rho_{k+1} u_{k+1} - \rho_k u_k}{\Delta t}, v \right\rangle + a\left((\rho u)_{k+1/2}, \frac{u_k + u_{k+1}}{2}, v\right)$$
$$- \frac{1}{2} b\left(v, \overline{u_k \cdot u_{k+1}}, \frac{\rho_k + \rho_{k+1}}{2}\right) - \langle p_{k+1}, \operatorname{div} v \rangle = 0, \qquad \forall v \in U_h \tag{34}$$

$$\left\langle \frac{\rho_{k+1} - \rho_k}{\Delta t}, \sigma \right\rangle - b\left(\frac{u_k + u_{k+1}}{2}, \sigma, \frac{\rho_k + \rho_{k+1}}{2}\right) = 0, \qquad \forall \sigma \in F_h \tag{35}$$

$$\langle \operatorname{div} u_{k+1}, q \rangle = 0, \qquad \forall q \in Q_h. \tag{36}$$

Remark 3.1. Just as in Remark 2.1, we have $\overline{u_k \cdot u_{k+1}} = u_k \cdot u_{k+1}$ if (23) holds, since then $u_k \cdot u_{k+1}|_K \in P_m(K)$ for every $K \in \mathcal{T}_h$.

Notice that the temporal discretization proposed above is similar, but not identical, to the midpoint rule. The key difference is that $\overline{u_k \cdot u_{k+1}}$ appears rather than $\overline{u_{k+1/2} \cdot u_{k+1/2}}$. It is well known that in the case of constant density, the midpoint rule can be used to construct energy-preserving schemes if the nonlinear term in the momentum equation is discretized in skew-symmetric form [25]. Here, in the variable density setting, the midpoint rule does not preserve energy, but replacing $\overline{u_{k+1/2} \cdot u_{k+1/2}}$ with $\overline{u_k \cdot u_{k+1}}$ as we have done above rectifies this. A proof is given below in Proposition 3.2.

3.1 Properties of the Temporal Discretization

By our choice of finite element spaces, we again have exact incompressibility of the discrete solution.

Proposition 3.1. *For every k, $\operatorname{div} u_k \equiv 0$.*

Proof. Take $q = \operatorname{div} u_{k+1}$ in (32). □

The next proposition shows that the fully discrete scheme (30-32) enjoys the same conservative properties as the semidiscrete scheme (24-26). In particular, it is unconditionally stable for any $\alpha_e, \beta_e \geq 0$, and it exactly preserves total mass, total energy, and, if $\beta_e = 0$ for every $e \in \mathcal{E}_h$, total squared density.

Proposition 3.2. *For every k, we have*

$$\int_\Omega \rho_{k+1}\,dx = \int_\Omega \rho_k\,dx, \tag{37}$$

$$\int_\Omega \rho_{k+1}^2\,dx \le \int_\Omega \rho_k^2\,dx, \ \textit{with equality if } \beta_e = 0,\ \forall e \in \mathcal{E}_h, \tag{38}$$

$$\int_\Omega \rho_{k+1}u_{k+1}\cdot u_{k+1}\,dx = \int_\Omega \rho_k u_k \cdot u_k\,dx. \tag{39}$$

Proof. Taking $\sigma \equiv 1$ in (31) gives

$$\int_\Omega \frac{\rho_{k+1}-\rho_k}{\Delta t}\,dx = \left\langle \frac{\rho_{k+1}-\rho_k}{\Delta t}, 1\right\rangle = b\left(\frac{u_k+u_{k+1}}{2}, 1, \frac{\rho_k+\rho_{k+1}}{2}\right) = 0.$$

Taking $\sigma = \frac{\rho_k+\rho_{k+1}}{2} = \rho_{k+1/2}$ in (31) and invoking (19) gives

$$\begin{aligned}
\frac{1}{2}\int_\Omega \frac{\rho_{k+1}^2-\rho_k^2}{\Delta t}\,dx &= \left\langle \frac{\rho_{k+1}-\rho_k}{\Delta t}, \frac{\rho_k+\rho_{k+1}}{2}\right\rangle \\
&= b\left(\frac{u_k+u_{k+1}}{2}, \frac{\rho_k+\rho_{k+1}}{2}, \frac{\rho_k+\rho_{k+1}}{2}\right) - \sum_{e\in\mathcal{E}_h}\int_e \beta_e(u_{k+1/2})[\![\rho_{k+1/2}]\!]\cdot[\![\rho_{k+1/2}]\!]\,ds \\
&= -\sum_{e\in\mathcal{E}_h}\int_e \beta_e(u_{k+1/2})[\![\rho_{k+1/2}]\!]\cdot[\![\rho_{k+1/2}]\!]\,ds \\
&\le 0.
\end{aligned}$$

To prove (39), we first observe the identity

$$\frac{1}{2}\int_\Omega \frac{\rho_{k+1}u_{k+1}\cdot u_{k+1}-\rho_k u_k\cdot u_k}{\Delta t}\,dx = \left\langle \frac{\rho_{k+1}u_{k+1}-\rho_k u_k}{\Delta t}, \frac{u_k+u_{k+1}}{2}\right\rangle - \frac{1}{2}\left\langle \frac{\rho_{k+1}-\rho_k}{\Delta t}, u_k\cdot u_{k+1}\right\rangle.$$

Next, we invoke (30) and the definition of the L^2-projection to write

$$\begin{aligned}
&\left\langle \frac{\rho_{k+1}u_{k+1}-\rho_k u_k}{\Delta t}, \frac{u_k+u_{k+1}}{2}\right\rangle - \frac{1}{2}\left\langle \frac{\rho_{k+1}-\rho_k}{\Delta t}, u_k\cdot u_{k+1}\right\rangle \\
&= \left\langle p_{k+1}, \operatorname{div}\left(\frac{u_k+u_{k+1}}{2}\right)\right\rangle - a\left((\rho u)_{k+1/2}, \frac{u_k+u_{k+1}}{2}, \frac{u_k+u_{k+1}}{2}\right) \\
&\quad + \frac{1}{2}b\left(\frac{u_k+u_{k+1}}{2}, \overline{u_k\cdot u_{k+1}}, \frac{\rho_k+\rho_{k+1}}{2}\right) - \frac{1}{2}\left\langle \frac{\rho_{k+1}-\rho_k}{\Delta t}, \overline{u_k\cdot u_{k+1}}\right\rangle \\
&\quad - \frac{1}{2}\sum_{e\in\mathcal{E}_h}\beta_e(u_{k+1/2})[\![\overline{u_k\cdot u_{k+1}}]\!]\cdot[\![\rho_{k+1/2}]\!]\,ds.
\end{aligned}$$

The first two terms above vanish by (32) and (18). The last three terms vanish according to the discrete density evolution equation (31). It follows that (39) holds. □

Remark 3.2. The definition (33) plays no role in the proof of Proposition 3.2. Thus, the conclusions (37-39) remain valid if (33) is replaced, for instance, by $(\rho u)_{k+1/2} = \left(\frac{\rho_k+\rho_{k+1}}{2}\right)\left(\frac{u_k+u_{k+1}}{2}\right)$, $(\rho u)_{k+1/2} = \rho_k u_k$ or $(\rho u)_{k+1/2} = \rho_{k+1}u_{k+1}$.

s	h^{-1}	$\|u_h - u\|_{L^2(\Omega)}$	Rate	$\|\rho_h - \rho\|_{L^2(\Omega)}$	Rate	$\|p_h - p\|_{L^2(\Omega)}$	Rate
0	1	$6.29 \cdot 10^{-1}$		$3.80 \cdot 10^{-1}$		$8.51 \cdot 10^{-1}$	
	2	$3.18 \cdot 10^{-1}$	0.98	$2.22 \cdot 10^{-1}$	0.77	$3.74 \cdot 10^{-1}$	1.18
	4	$1.58 \cdot 10^{-1}$	1.01	$1.11 \cdot 10^{-1}$	1.01	$1.81 \cdot 10^{-1}$	1.05
	8	$7.72 \cdot 10^{-2}$	1.03	$5.08 \cdot 10^{-2}$	1.12	$8.85 \cdot 10^{-2}$	1.04
1	1	$2.11 \cdot 10^{-1}$		$1.21 \cdot 10^{-1}$		$2.66 \cdot 10^{-1}$	
	2	$1.18 \cdot 10^{-1}$	0.84	$4.56 \cdot 10^{-2}$	1.40	$1.66 \cdot 10^{-1}$	0.68
	4	$3.02 \cdot 10^{-2}$	1.97	$1.64 \cdot 10^{-2}$	1.47	$4.20 \cdot 10^{-2}$	1.98
	8	$1.23 \cdot 10^{-2}$	1.29	$7.45 \cdot 10^{-3}$	1.14	$1.74 \cdot 10^{-2}$	1.27
2	1	$3.55 \cdot 10^{-2}$		$2.30 \cdot 10^{-2}$		$7.57 \cdot 10^{-2}$	
	2	$3.99 \cdot 10^{-3}$	3.15	$3.90 \cdot 10^{-3}$	2.56	$7.81 \cdot 10^{-3}$	3.28
	4	$5.54 \cdot 10^{-4}$	2.85	$4.36 \cdot 10^{-4}$	3.16	$1.01 \cdot 10^{-3}$	2.95
	8	$9.53 \cdot 10^{-5}$	2.54	$5.25 \cdot 10^{-5}$	3.05	$1.45 \cdot 10^{-4}$	2.80

Table 1: L^2-errors in the velocity, density, and pressure at time $T = 0.5$ without upwinding.

s	h^{-1}	$\|u_h - u\|_{L^2(\Omega)}$	Rate	$\|\rho_h - \rho\|_{L^2(\Omega)}$	Rate	$\|p_h - p\|_{L^2(\Omega)}$	Rate
0	1	$5.97 \cdot 10^{-1}$		$3.82 \cdot 10^{-1}$		$1.00 \cdot 10^{0}$	
	2	$3.00 \cdot 10^{-1}$	0.99	$2.04 \cdot 10^{-1}$	0.91	$4.21 \cdot 10^{-1}$	1.25
	4	$1.60 \cdot 10^{-1}$	0.90	$1.06 \cdot 10^{-1}$	0.94	$2.09 \cdot 10^{-1}$	1.01
	8	$8.20 \cdot 10^{-2}$	0.97	$5.32 \cdot 10^{-2}$	0.99	$1.02 \cdot 10^{-1}$	1.04
1	1	$1.89 \cdot 10^{-1}$		$8.06 \cdot 10^{-2}$		$2.83 \cdot 10^{-1}$	
	2	$5.05 \cdot 10^{-2}$	1.90	$2.12 \cdot 10^{-2}$	1.93	$8.28 \cdot 10^{-2}$	1.77
	4	$1.28 \cdot 10^{-2}$	1.98	$5.95 \cdot 10^{-3}$	1.83	$2.15 \cdot 10^{-2}$	1.95
	8	$3.28 \cdot 10^{-3}$	1.96	$1.57 \cdot 10^{-3}$	1.92	$5.41 \cdot 10^{-3}$	1.99
2	1	$2.28 \cdot 10^{-2}$		$1.48 \cdot 10^{-2}$		$7.35 \cdot 10^{-2}$	
	2	$3.06 \cdot 10^{-3}$	2.89	$2.83 \cdot 10^{-3}$	2.39	$7.46 \cdot 10^{-3}$	3.30
	4	$3.77 \cdot 10^{-4}$	3.02	$3.74 \cdot 10^{-4}$	2.92	$9.28 \cdot 10^{-4}$	3.01
	8	$4.84 \cdot 10^{-5}$	2.96	$5.14 \cdot 10^{-5}$	2.86	$1.17 \cdot 10^{-4}$	2.99

Table 2: L^2-errors in the velocity, density, and pressure at time $T = 0.5$ with upwinding.

4 Numerical Examples

4.1 Convergence Tests

To test the performance of the numerical method (30-32), we used it to approximate the solution of (9-11) on the square $\Omega = (-1, 1) \times (-1, 1)$ with initial conditions

$$u(x, y, 0) = (-\cos(\pi x/2)\sin(\pi y/2),\ \sin(\pi x/2)\cos(\pi y/2)),$$
$$\rho(x, y, 0) = 2 + \sin(xy).$$

We used the finite element spaces $U_h = RT_s(\mathcal{T}_h)$, $F_h = DG_s(\mathcal{T}_h)$, and $Q_h = DG_s(\mathcal{T}_h) \cap L^2_{\int=0}(\Omega)$ with $s \in \{0, 1, 2\}$ on a uniform triangulation $\mathcal{T}_h$ with maximum element diameter $h = 2^{-j}$, $j = 0, 1, 2, 3$. We used a small time step $\Delta t = 0.00625$ to ensure that temporal discretization errors were negligible, and we measured the L^2-error in the computed solution (u_h, ρ_h, p_h) at time $T = 0.5$. The "exact" solution was obtained with $s = 2$, $h = 2^{-5}$, and $\Delta t = 0.00625$. Tables 1-2 show the results for two choices of the parameters α_e, β_e: $\alpha_e = \beta_e = 0$ (no upwinding) and $\alpha_e(u) = \frac{u \cdot n}{2|u \cdot n|}$, $\beta_e(u) = \frac{1}{2}|u \cdot n|$ (upwinding). When using upwinding, the errors converged at the optimal rates 1, 2, and 3, respectively, for polynomial degrees $s = 0, 1, 2$. In the absence of upwinding, the convergence rates for $s = 1$ were lower (closer to first order than to second order). These observations are consistent with those of [18, 25], where the case of constant density is considered.

Δt^{-1}	$\|u_h - u\|_{L^2(\Omega)}$	Rate	$\|\rho_h - \rho\|_{L^2(\Omega)}$	Rate	$\|p_h - p\|_{L^2(\Omega)}$	Rate
2	$7.64 \cdot 10^{-3}$		$2.90 \cdot 10^{-2}$		$4.47 \cdot 10^{-2}$	
4	$2.27 \cdot 10^{-3}$	1.75	$8.35 \cdot 10^{-3}$	1.80	$1.37 \cdot 10^{-2}$	1.70
8	$6.03 \cdot 10^{-4}$	1.91	$2.19 \cdot 10^{-3}$	1.93	$3.63 \cdot 10^{-3}$	1.92
16	$1.77 \cdot 10^{-4}$	1.76	$5.52 \cdot 10^{-4}$	1.99	$9.23 \cdot 10^{-4}$	1.97

Table 3: Convergence with respect to Δt of the L^2-errors in the velocity, density, and pressure at time $T = 0.5$ without upwinding.

Δt^{-1}	$\|u_h - u\|_{L^2(\Omega)}$	Rate	$\|\rho_h - \rho\|_{L^2(\Omega)}$	Rate	$\|p_h - p\|_{L^2(\Omega)}$	Rate
2	$7.64 \cdot 10^{-3}$		$2.90 \cdot 10^{-2}$		$4.47 \cdot 10^{-2}$	
4	$2.27 \cdot 10^{-3}$	1.75	$8.35 \cdot 10^{-3}$	1.80	$1.37 \cdot 10^{-2}$	1.70
8	$5.99 \cdot 10^{-4}$	1.92	$2.19 \cdot 10^{-3}$	1.93	$3.62 \cdot 10^{-3}$	1.92
16	$1.58 \cdot 10^{-4}$	1.92	$5.52 \cdot 10^{-4}$	1.99	$9.19 \cdot 10^{-4}$	1.98

Table 4: Convergence with respect to Δt of the L^2-errors in the velocity, density, and pressure at time $T = 0.5$ with upwinding.

Figure 1 plots the squared density errors $|1 - F(t)/F(0)|$, $F(t) = \int_\Omega \rho_h(t)^2 \, dx$, for the simulations in Tables 1-2 that used $h = \frac{1}{4}$. As expected, squared density decayed monotonically with upwinding and remained constant (up to roundoff errors) without upwinding. Total mass and energy errors (not plotted) remained below 10^{-13} in these experiments, both with and without upwinding.

To test the convergence of the method with respect to Δt, we repeated the above experiment with $s = 2$, $h = 2^{-4}$, and $\Delta t = 2^{-j}$, $j = 1, 2, 3, 4$. The results in Tables 3-4 suggest that the method converges at a second-order rate with respect to Δt, both with and without upwinding.

4.2 Rayleigh-Taylor Instability

As a second test, we simulated the Rayleigh-Taylor instability on a rectangle $\Omega = (-1/2, 1/2) \times (-2, 2)$ with initial conditions

$$u(x, y, 0) = (0, 0),$$
$$\rho(x, y, 0) = 2 + \tanh\left(\frac{y + 0.1\cos(2\pi x)}{0.1}\right).$$

For this test, we added a gravitational forcing term $\langle (0, -g)\rho_{k+1/2}, v\rangle$, $g = 10$, to right-hand side of the momentum equation (30). We used the upwind scheme with a time step $\Delta t = 0.01$ and finite element spaces $U_h = RT_0(\mathcal{T}_h)$, $F_h = DG_1(\mathcal{T}_h)$, and $Q_h = DG_0(\mathcal{T}_h) \cap L^2_{\int=0}(\Omega)$ on a uniform triangulation $\mathcal{T}_h$ with maximum element diameter $h = 2^{-j}$, $j = 4, 5, 6$. Plots of the density at various times t are shown in Figures 2-4. The simulations at different resolutions agree with one another qualitatively. In the absence of upwinding, the simulation is corrupted by unphysical oscillations; see Figure 5.

For comparison, we simulated the same Rayleigh-Taylor instability with the spatial discretization introduced in [16], adapted to the inviscid setting. This method seeks $u_h(t) \in \widetilde{U}_h$, $\rho_h(t) \in \widetilde{F}_h$, and $p_h(t) \in \widetilde{Q}_h$ such that

$$\langle \sigma_h \partial_t(\sigma_h u_h), v_h\rangle + \langle \rho_h u_h \cdot \nabla u_h, v_h\rangle + \frac{1}{2}\langle u_h \operatorname{div}(\rho_h u_h), v_h\rangle = \langle f, v_h\rangle - \langle \nabla p_h, v_h\rangle, \quad \forall v_h \in \widetilde{U}_h, \quad (40)$$

$$\langle \partial_t \rho_h, \theta_h\rangle + \langle u_h \cdot \nabla \rho_h, \theta_h\rangle + \frac{1}{2}\langle \rho_h \operatorname{div} u_h, \theta_h\rangle = 0, \quad \forall \theta_h \in \widetilde{F}_h, \quad (41)$$

$$\langle \operatorname{div} u_h, q_h\rangle = 0, \quad \forall q_h \in \widetilde{Q}_h, \quad (42)$$

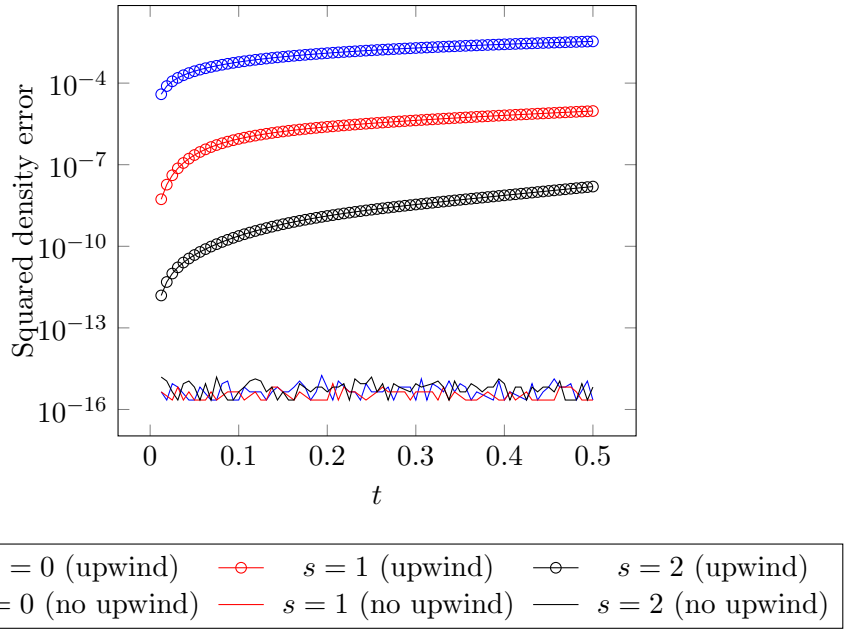

Figure 1: Squared density errors $|1 - F(t)/F(0)|$, $F(t) = \int_\Omega \rho_h(t)^2\,dx$, for the simulations in Tables 1-2 that used $h = \frac{1}{4}$. Total mass and energy errors (not plotted) remained below 10^{-13} in these experiments, both with and without upwinding.

where $\sigma_h = \sqrt{\rho_h}$, $f = (0, -g)\rho_h$, and $\widetilde{U}_h$, $\widetilde{F}_h$, and $\widetilde{Q}_h$ are finite-dimensional subspaces of $\{u \in H^1(\Omega) \mid u \cdot n = 0 \text{ on } \partial\Omega\}$, $H^1(\Omega)$, and $H^1(\Omega) \cap L^2_{\int=0}(\Omega)$, respectively. We chose

$$\begin{aligned}
\widetilde{U}_h &= \{u \in CG_2(\mathcal{T}_h)^n \mid u \cdot n = 0 \text{ on } \partial\Omega\},\\
\widetilde{F}_h &= CG_1(\mathcal{T}_h),\\
\widetilde{Q}_h &= CG_1(\mathcal{T}_h) \cap L^2_{\int=0}(\Omega),
\end{aligned}$$

where $CG_s(\mathcal{T}_h) = \{q \in C^0(\overline{\Omega}) \mid q|_K \in P_s(K), \forall K \in \mathcal{T}_h\}$. We discretized (40-42) in time with the midpoint rule and added streamline-upwind Petrov-Galerkin (SUPG) stabilization terms [7] to (40-41) to mitigate spurious oscillations. Note that Guermond and Quartapelle [16] propose a more sophisticated temporal discretization (a fractional step method), but we adopted the midpoint rule here for simplicity. Figure 6 shows the results obtained with $h = 2^{-5}$ and $\Delta t = 0.01$. The two methods under comparison produce qualitatively similar results for $t < 1$, and begin to deviate somewhat as t increases; compare Figures 3 and 6.

In Figures 7-9, we compare the conservation properties of our method (both with and without upwinding) to (40-42). All of the methods under consideration preserve total mass (not plotted) to machine precision. Without upwinding, our method preserves total squared density to machine precision, and with upwinding it dissipates total squared density slightly more slowly than (40-42) does. Our method preserves total energy to machine precision both with and without upwinding, and the L^2-norm of $\operatorname{div} u_h$ drifts slightly but remains below 10^{-13}. In contrast, with (40-42), the energy drifts somewhat and $\|\operatorname{div} u_h\|_{L^2(\Omega)}$ drifts considerably.

It is worth noting that the methods under comparison above have similar computational complexity. On the triangulations with $h = 2^{-4}$, $h = 2^{-5}$, and $h = 2^{-6}$, our method uses 22,607, 90,722, and 360,768 degrees of freedom, respectively, whereas (40-42) uses 20,864, 82,884, and 329,604 degrees of freedom, respectively.

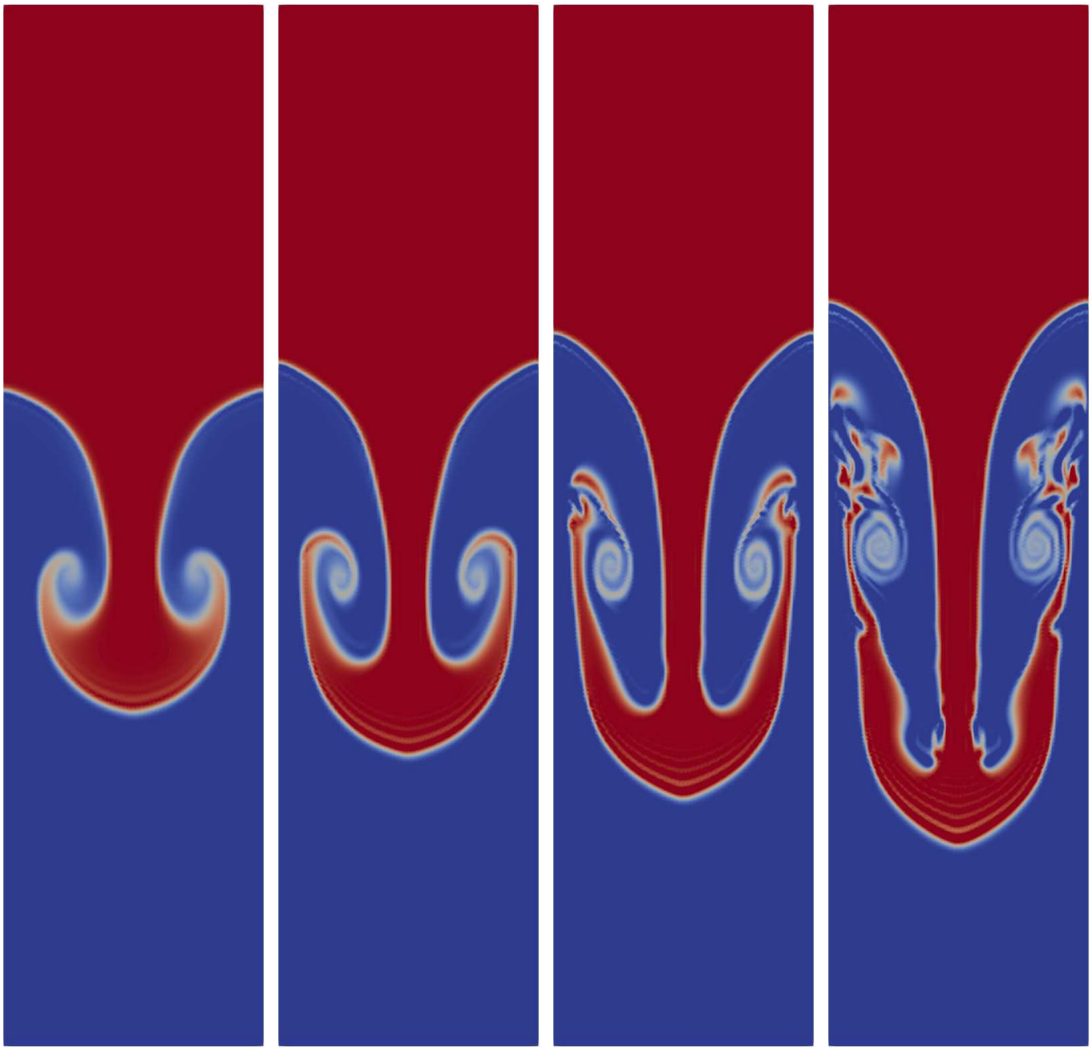

Figure 2: Density contours at $t = 0.8, 0.95, 1.1, 1.25$ in the Rayleigh-Taylor instability simulation with $h = 2^{-6}$.

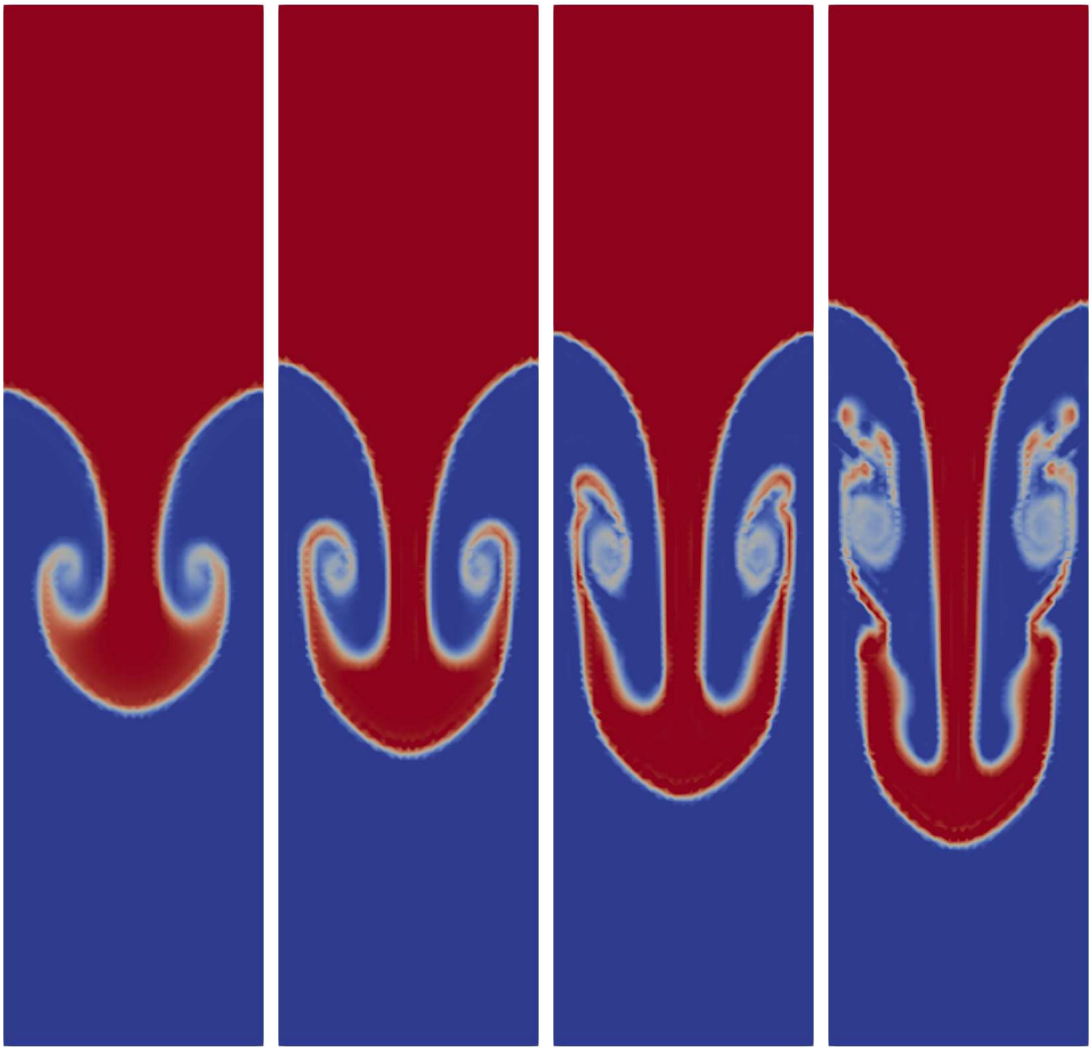

Figure 3: Density contours at $t = 0.8, 0.95, 1.1, 1.25$ in the Rayleigh-Taylor instability simulation with $h = 2^{-5}$.

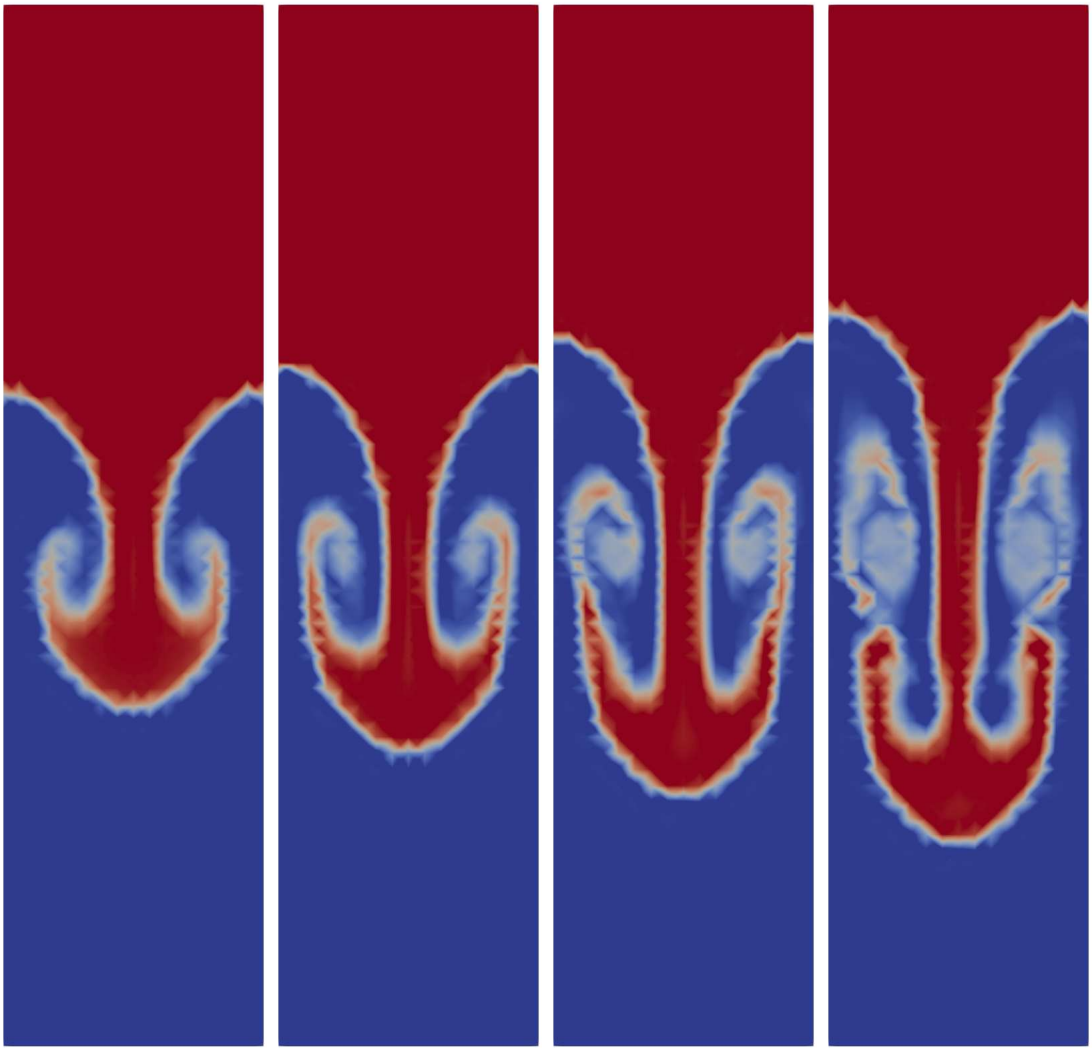

Figure 4: Density contours at $t = 0.8, 0.95, 1.1, 1.25$ in the Rayleigh-Taylor instability simulation with $h = 2^{-4}$.

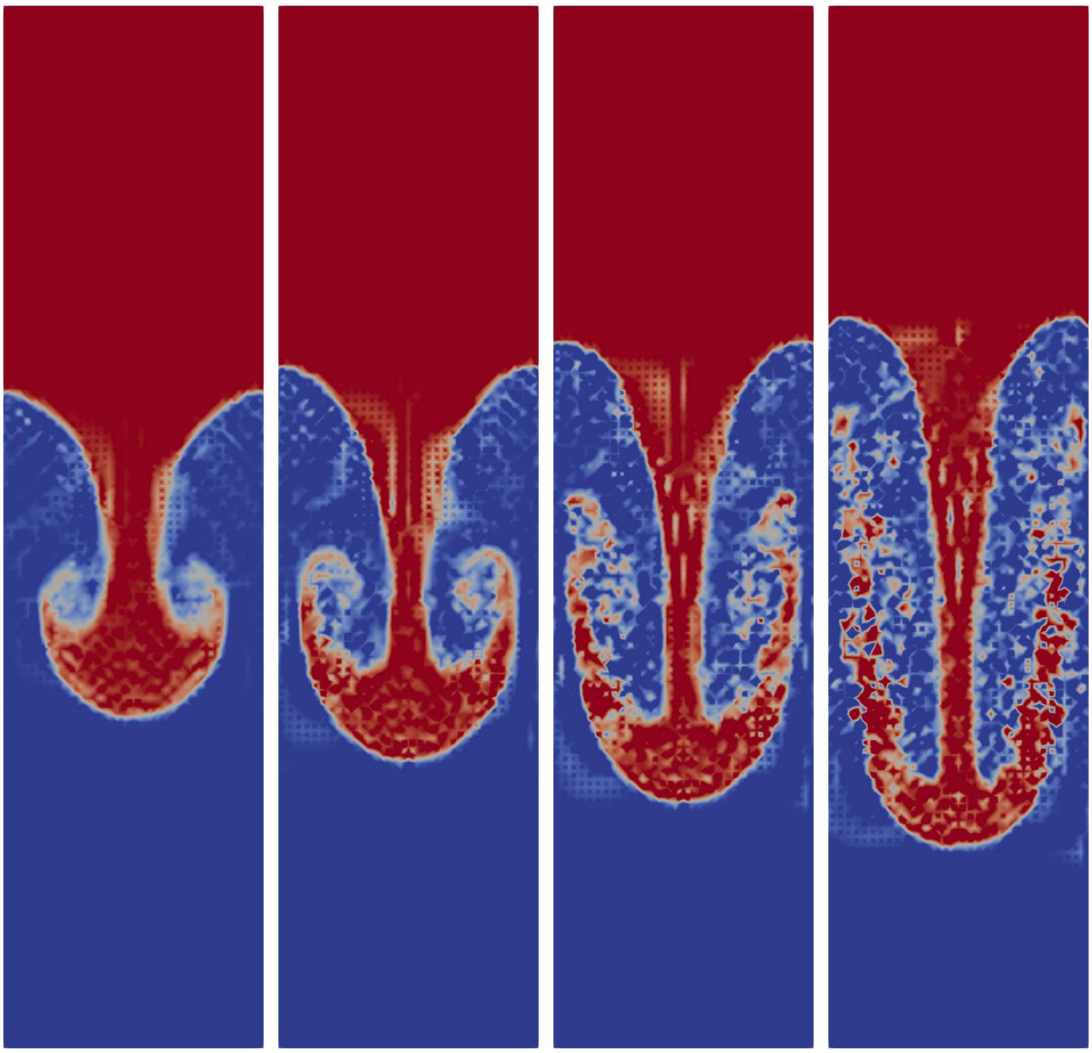

Figure 5: Density contours at $t = 0.8, 0.95, 1.1, 1.25$ in the Rayleigh-Taylor instability simulation with $h = 2^{-5}$ and no upwinding.

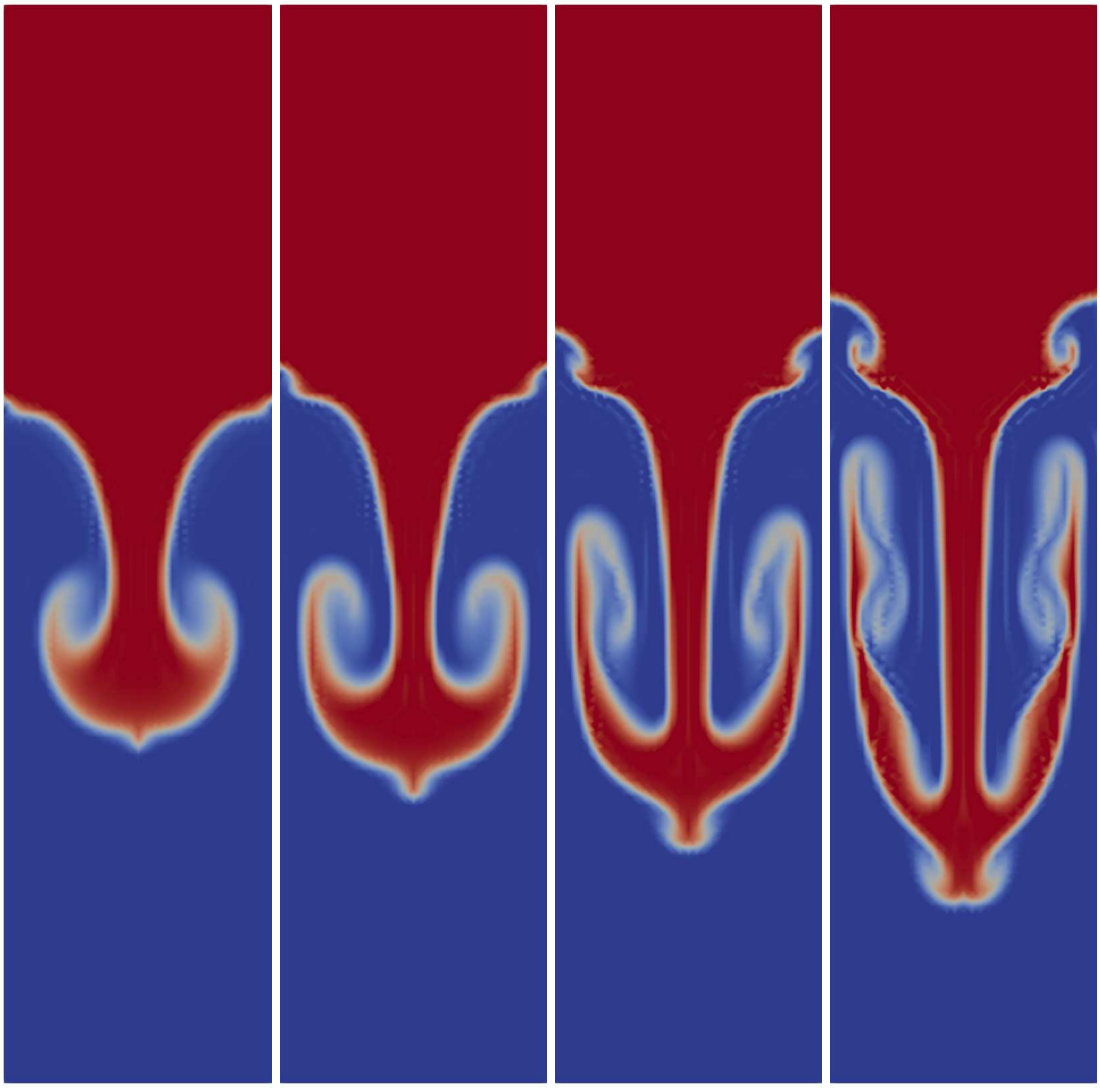

Figure 6: Density contours at $t = 0.8, 0.95, 1.1, 1.25$ in the Rayleigh-Taylor instability simulation, obtained using (40-42) with $h = 2^{-5}$.

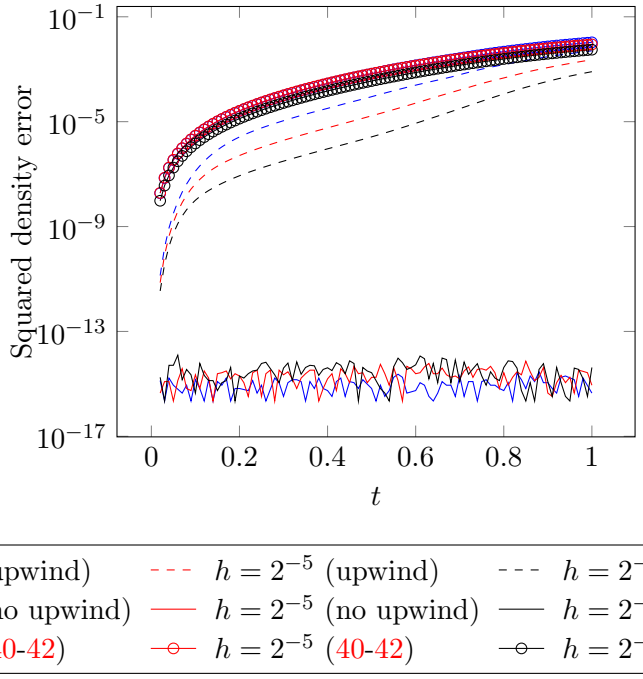

Figure 7: Squared density errors $|1 - F(t)/F(0)|$, $F(t) = \int_\Omega \rho_h(t)^2\,dx$, in the Rayleigh-Taylor instability simulation.

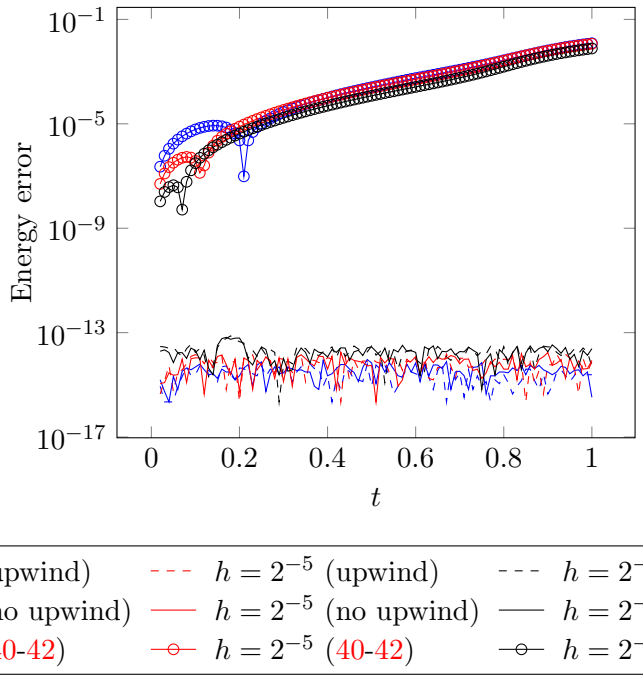

Figure 8: Energy errors $|1 - E(t)/E(0)|$, $E(t) = \int_\Omega \left(\frac{1}{2}\rho_h(t)u_h(t)\cdot u_h(t) + \rho_h(t)gy\right)\,dx$, in the Rayleigh-Taylor instability simulation. (The curves labelled (40-42) appear nonsmooth because the sign of $1 - E(t)/E(0)$ changes from negative to positive near $t = 0$.)

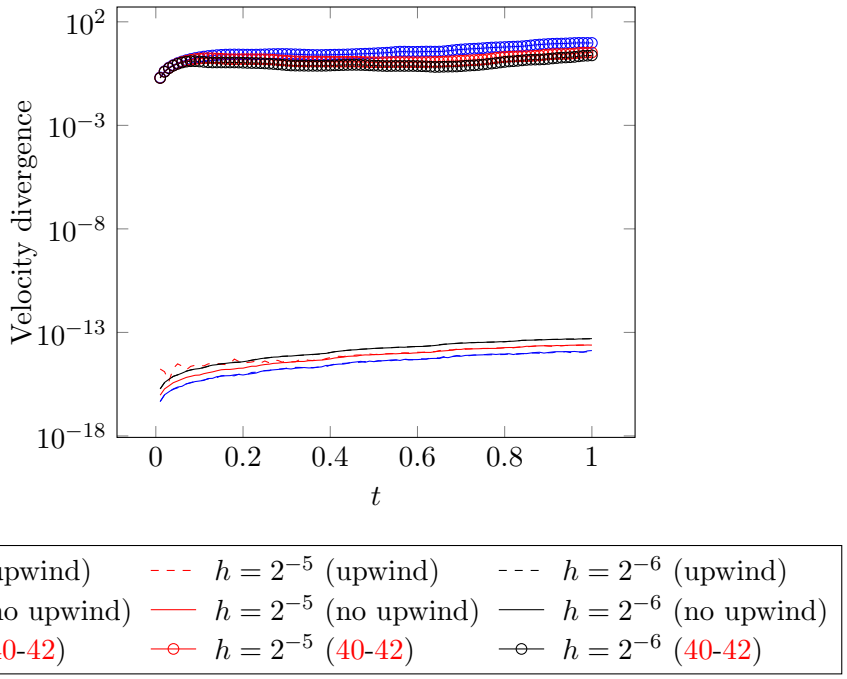

Figure 9: L^2-norm of the divergence of the velocity field in the Rayleigh-Taylor instability simulation.

5 Conclusion

We have constructed a numerical method for the incompressible Euler equations with variable density that exactly preserves total mass, total squared density, total energy, and pointwise incompressibility at the spatially and temporally discrete levels. The method achieves second-order accuracy in time, and allows for the use of high-order finite elements to achieve high-order accuracy in space. An upwinded version of the method was also described and proved to be stable. Numerical tests illustrated its convergence order and its performance on a simulation of the Rayleigh-Taylor instability.

6 Acknowledgements

EG was partially supported by NSF grant DMS-1703719. FGB was partially supported by the ANR project GEOMFLUID, ANR-14-CE23-0002-01.

References

[1] A. S. Almgren, J. B. Bell, P. Colella, L. H. Howell, and M. L. Welcome. "A conservative adaptive projection method for the variable density incompressible Navier–Stokes equations". In: *Journal of Computational Physics* 142.1 (1998), pp. 1–46.

[2] A. Arakawa. "Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow. Part I". In: *Journal of Computational Physics* 1.1 (1966), pp. 119–143.

[3] W. Bauer and C. J. Cotter. "Energy–enstrophy conserving compatible finite element schemes for the rotating shallow water equations with slip boundary conditions". In: *Journal of Computational Physics* 373 (2018), pp. 171–187.

[4] J. B. Bell and D. L. Marcus. "A second-order projection method for variable-density flows". In: *Journal of Computational Physics* 101.2 (1992), pp. 334–348.

[5] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Vol. 15. Springer-Verlag, 1991.

[6] F. Brezzi, L. D. Marini, and E. Süli. "Discontinuous Galerkin methods for first-order hyperbolic problems". In: *Mathematical Models and Methods in Applied Sciences* 14.12 (2004), pp. 1893–1903.

[7] A. N. Brooks and T. J. Hughes. "Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations". In: *Computer Methods in Applied Mechanics and Engineering* 32.1-3 (1982), pp. 199–259.

[8] C. Calgaro, E. Creusé, and T. Goudon. "An hybrid finite volume–finite element method for variable density incompressible flows". In: *Journal of Computational Physics* 227.9 (2008), pp. 4671–4696.

[9] S. Charnyi, T. Heister, M. A. Olshanskii, and L. G. Rebholz. "On conservation laws of Navier–Stokes Galerkin discretizations". In: *Journal of Computational Physics* 337 (2017), pp. 289–308.

[10] B. Cockburn, G. Kanschat, and D. Schötzau. "A note on discontinuous Galerkin divergence-free solutions of the Navier–Stokes equations". In: *Journal of Scientific Computing* 31.1-2 (2007), pp. 61–73.

[11] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements*. Vol. 159. Springer Science & Business Media, 2004.

[12] J. A. Evans and T. J. Hughes. "Isogeometric divergence-conforming B-splines for the unsteady Navier–Stokes equations". In: *Journal of Computational Physics* 241 (2013), pp. 141–167.

[13] R. S. Falk and M. Neilan. "Stokes complexes and the construction of stable finite elements with pointwise mass conservation". In: *SIAM Journal on Numerical Analysis* 51.2 (2013), pp. 1308–1326.

[14] E. S. Gawlik and F. Gay-Balmaz. "A variational finite element discretization of compressible flow". In: *https://arxiv.org/abs/1910.05648* (2019).

[15] E. S. Gawlik, P. Mullen, D. Pavlov, J. E. Marsden, and M. Desbrun. "Geometric, variational discretization of continuum theories". In: *Physica D: Nonlinear Phenomena* 240.21 (2011), pp. 1724–1760.

[16] J.-L. Guermond and L. Quartapelle. "A projection FEM for variable density incompressible flows". In: *Journal of Computational Physics* 165.1 (2000), pp. 167–188.

[17] J.-L. Guermond and A. Salgado. "A splitting method for incompressible flows with variable density based on a pressure Poisson equation". In: *Journal of Computational Physics* 228.8 (2009), pp. 2834–2846.

[18] J. Guzmán, C.-W. Shu, and F. A. Sequeira. "H(div) conforming and DG methods for incompressible Euler's equations". In: *IMA Journal of Numerical Analysis* 37.4 (2016), pp. 1733–1771.

[19] J. J. Heys, E. Lee, T. A. Manteuffel, and S. F. McCormick. "On mass-conserving least-squares methods". In: *SIAM Journal on Scientific Computing* 28.5 (2006), pp. 1675–1693.

[20] V. John, A. Linke, C. Merdon, M. Neilan, and L. G. Rebholz. "On the divergence constraint in mixed finite element methods for incompressible flows". In: *SIAM Review* 59.3 (2017), pp. 492–544.

[21] J. Kreeft and M. Gerritsma. "Mixed mimetic spectral element method for Stokes flow: A pointwise divergence-free solution". In: *Journal of Computational Physics* 240 (2013), pp. 284–309.

[22] R. J. Labeur and G. N. Wells. "Energy stable and momentum conserving hybrid finite element method for the incompressible Navier–Stokes equations". In: *SIAM Journal on Scientific Computing* 34.2 (2012), A889–A913.

[23] D. Lee, A. Palha, and M. Gerritsma. "Discrete conservation properties for shallow water flows using mixed mimetic spectral elements". In: *Journal of Computational Physics* 357 (2018), pp. 282–304.

[24] Y. Morinishi, T. S. Lund, O. V. Vasilyev, and P. Moin. "Fully conservative higher order finite difference schemes for incompressible flow". In: *Journal of Computational Physics* 143.1 (1998), pp. 90–124.

[25] A. Natale and C. J. Cotter. "A variational finite-element discretization approach for perfect incompressible fluids". In: *IMA Journal of Numerical Analysis* 38.3 (2017), pp. 1388–1419.

[26] A. Palha and M. Gerritsma. "A mass, energy, enstrophy and vorticity conserving (MEEVC) mimetic spectral element discretization for the 2D incompressible Navier–Stokes equations". In: *Journal of Computational Physics* 328 (2017), pp. 200–220.

[27] D. Pavlov et al. "Structure-preserving discretization of incompressible fluids". In: *Physica D: Nonlinear Phenomena* 240.6 (2011), pp. 443–458.

[28] M. M. Proot and M. I. Gerritsma. "Mass-and momentum conservation of the least-squares spectral element method for the Stokes problem". In: *Journal of Scientific Computing* 27.1-3 (2006), pp. 389–401.

[29] J.-H. Pyo and J. Shen. "Gauge–Uzawa methods for incompressible flows with variable density". In: *Journal of Computational Physics* 221.1 (2007), pp. 181–197.

[30] L. G. Rebholz. "An energy-and helicity-conserving finite element scheme for the Navier–Stokes equations". In: *SIAM Journal on Numerical Analysis* 45.4 (2007), pp. 1622–1638.

[31] S. Rhebergen and G. N. Wells. "A hybridizable discontinuous Galerkin method for the Navier–Stokes equations with pointwise divergence-free velocity field". In: *Journal of Scientific Computing* 76.3 (2018), pp. 1484–1501.

[32] B. Sanderse. "Energy-conserving Runge–Kutta methods for the incompressible Navier–Stokes equations". In: *Journal of Computational Physics* 233 (2013), pp. 100–131.