INSTITUTE OF MATHEMATICS

THE CZECH ACADEMY OF SCIENCES

Stability and consistency of a finite difference scheme for compressible viscous isentropic flow in multi-dimension

Radim Hošek

Bangwei She

Preprint No. 8-2017

PRAHA 2017

Stability and consistency of a finite difference scheme for compressible viscous isentropic flow in multi-dimension

Radim Hošek, Bangwei She

Institute of Mathematics, Czech Academy of Sciences*

January 26, 2017

Abstract

Motivated by the work of Karper [27], we propose a numerical scheme to compressible Navier–Stokes system in multi-spatial dimensions, based on finite differences. The backward Euler method is applied for the time discretization, while a staggered grid, with continuity and momentum equations on different grids, is used in space. The existence of a solution to the implicit nonlinear scheme, strictly positivity of the numerical density, stability and consistency of the method are proved. The theoretical part is complemented by computational results that are performed in two spatial dimensions.

Key words: compressible Navier-Stokes, finite difference method, positivity preserving, energy stability, consistency

1 Introduction

The compressible Navier–Stokes system as a set of balance laws for mass and momentum, describes the flow of isentropic viscous gas, where the thermal effects are neglected. Let ϱ, $\mathbf{u}$ be the density and velocity field, the governing equations read

$$\partial_t \varrho + \mathrm{div}_x(\varrho \mathbf{u}) = 0, \tag{1}$$

$$\partial_t(\varrho \mathbf{u}) + \mathrm{div}_x(\varrho \mathbf{u} \otimes \mathbf{u}) + \nabla_x p(\varrho) = \mathrm{div}_x \mathbb{S} + \mathbf{f}. \tag{2}$$

Unlike the incompressible case, pressure in here is a function of density, assumed as

$$p(\varrho) = a\varrho^\gamma, \quad a > 0, \gamma > 1, \tag{3}$$

where the important features of the pressure are its convexity and asymptotic behaviour. Discussions about weakening this assumption can be found in [10]. For the consistency formulation, we need $\gamma > \frac{3}{2}$ for three-dimensional flow, which covers the case of a monatomic gas.

For the sake of easing the computation, the viscous stress tensor is assumed to take the form $\mathbb{S} = \mu \nabla_x \mathbf{u}$, $\mu > 0$ is the viscosity coefficient, and $\mathrm{div}_x \mathbb{S} = \mu \Delta_x \mathbf{u}$. We also omit the external forces, i.e. we set $\mathbf{f} \equiv 0$, bearing in mind that including them would not bring any insurmountable difficulties.

The system is complemented with initial conditions

$$\varrho|_{t=0} = \varrho_0 > 0, \qquad \mathbf{u}|_{t=0} = \mathbf{u}_0, \tag{4}$$

and homogeneous Dirichlet boundary condition for velocity

$$\mathbf{u}|_{\partial\Omega} = 0, \tag{5}$$

where $\Omega \subset \mathbb{R}^d$ is assumed to be a bounded Lipschitz domain, for space dimension $d = 2$ or 3. The time interval is $[0, T]$, without any assumptions on its size. More over, we expect the regularity $\varrho_0 \in L^\gamma(\Omega)$, $\mathbf{u}_0 \in W_0^{1,2}(\Omega)$.

The existence of strong solutions to (1–5) for *sufficiently smooth* initial data was proved in [37], however only for a possibly small time interval $[0, T^\star)$. Therefore, it was welcome, when the unconditional existence of *weak solution* was proved by Lions [32] and further developed in [18]. However, the existence result still requires $\gamma > \frac{3}{2}$, which does not cover the case of a diatomic gas. There are results on full system describing compressible flow, i.e. considering also the balance law of energy. Numerical schemes can be found in the framework of finite difference, finite

*The research of the authors leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ ERC Grant Agreement 320078. The Institute of Mathematics of the Academy of Sciences of the Czech Republic is supported by RVO:67985840.

volume, finite element, discontinuous Galerkin, gas kinetic BGK or mixtures of them. Representative examples are [31, 24, 3, 30, 6, 29, 35, 36, 38]. While considering the isentropic case, L^2-stable scheme has been studied in [1, 19], where upwind and pressure projection are the main technique. All speed asymptotic-preserving scheme can be found in [25] especially for low Mach number limit. Error estimates for the isentropic case was studied in [12, 23, 33]. The convergence of the compressible Navier-Stokes to its incompressible limit was numerically measured by a relative entropy at low Mach regime in [17]. Recently, Gallouët et al. proposed a MAC scheme similar to ours, for which they prove convergence results for (semi)stationary flows [20, 22], and error estimates for compressible Navier–Stokes [21].

Concerning the convergence of the numerical methods, to our best knowledge there is only one result in [27], where the scheme is based on finite element combined with discontinuous Galerkin method and uses also *upwind flux*. For linear problems, stability and consistency is enough to ensure convergence.In [27], Karper mimicked the proof of existence of weak solution for compressible Navier–Stokes system by Lions [32] and then showed for vanishing discretization parameter the convergence of the numerical solution, up to a subsequence, to a weak solution. This work had been further extended for smooth domains using non-fitted mesh [15] and to a heat conducting case [13, 14].

The scheme in [27] did not obtain a grateful acceptance, being labeled as *too academic* within the computational community. Therefore, an effort to prove convergence of a simpler numerical scheme motivated our result. In [28], Karper suggests a finite difference scheme for one dimensional compressible Navier–Stokes and shows its convergence. Moreover, it is suggested there to extend the result to multi-dimension, which we bring in this paper. Our result can be viewed as a starting point for two possible directions. One of them is continuation in the spirit of [27] in order to prove convergence of the (subsequence of) numerical solution to a weak solution. The other direction could be proving a convergence to measure-valued solution, which, in a suitable setting, coincides with a strong solution on its (possibly short) life span, see [11, 16].

In this paper we present the theoretical results of stability and consistency followed by numerical experiments. The paper is organized as follows. We explain the detailed scheme in Section 2. Then comes the proofs of positivity preserving of density, existence of the solution at any time level, energy stability and derivation of uniform estimates in Section 3, the consistency formulation in Section 4 and finally, numerical tests of the method in Section 5.

2 The numerical method

2.1 Time discretization

We discretize the time step equidistantly using Δt ($T = N_t \Delta t$) and define function only at these time instants $f^k := f(k\Delta t)$. The time derivative is approximated by the backward Euler method,

$$(\partial_h^t f)^n := \frac{f^n - f^{n-1}}{\Delta t}, \quad n = 1, 2, \ldots, N_t.$$

2.2 Spatial grids

2.2.1 Primary and dual grids

For convenience, the domain in our problem is set as $Q_T = I \times \Omega = [0, T] \times (0, L_x)^d$. A staggered grid is used in our spatial discretization. The domain Ω is uniformly discretized with mesh size $h = L_x/N_x$, i.e. $\overline{\Omega} := \bigcup \overline{Q_K}$ where the element Q_K is given by

$$Q_K = ((i-1)h; ih) \times ((j-1)h; jh) \times ((k-1)h; kh), \forall i, j, k \in \{1, \ldots, N_x\},$$

for example in three dimensions. The primary grid $\mathcal{T}$ is built by the centers K of these elements. Boundary of each element Q_K is created by faces F_σ, whose centers σ build the secondary grid $\mathcal{E}$, cf. Figure 1 which depicts the simpler two-dimensional case. Points $\sigma \in \mathcal{E}$ belonging to $\partial\Omega$ form $\mathcal{E}_{\text{ext}}$, while $\mathcal{E}_{\text{int}} = \mathcal{E} \setminus \mathcal{E}_{\text{ext}}$. We denote $\mathcal{E}(K)$ as the set of points that are at the center of the faces of element Q_K,

$$\mathcal{E}(K) := \left\{ \sigma = K \pm \frac{h}{2}\mathbf{e}_s, K \in \mathcal{T}, s = 1, \ldots, d \right\},$$

where $\mathbf{e}_s$ is a unit basis vector in one of the space directions (i.e. either $\mathbf{e}_1, \mathbf{e}_2$ or $\mathbf{e}_3$). Note that σ is linked with the direction of its normal vector $\mathbf{e}_s$, we denote it also as

$$\sigma_{,s\pm} = K \pm \frac{h}{2}\mathbf{e}_s.$$

On the other hand, any $\sigma \in \mathcal{E}_{\text{int}}$ adjacent to the elements K and $L \in \mathcal{N}(K)$ of the primary mesh, where $\mathcal{N}(K)$ is the collection neighbouring elements of K, we write $\sigma = K|L$ if $L = K + h\mathbf{e}_s$ for some $s = 1, \ldots, d$.

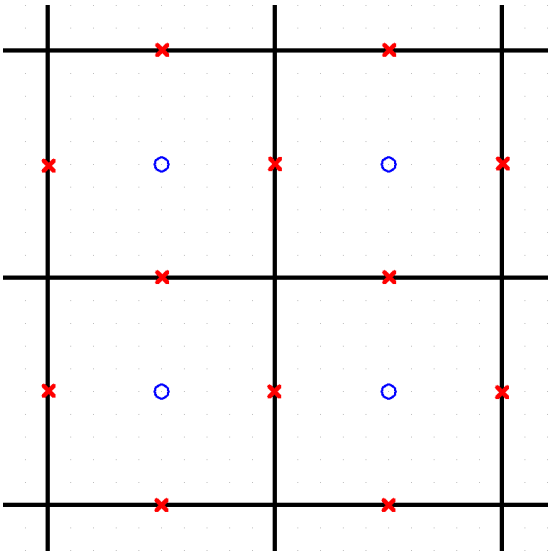


Figure 1: Space discretization: Blue circles o and red crosses × are the points of primary mesh and dual mesh, respectively.

2.2.2 Transferring quantities between grids

For any quantity f_h defined on the primary mesh $\mathcal{T}$ we denote its value at K as f_K. It can be interpolated to the dual one for $\sigma = K|L \in \mathcal{E}_{\text{int}}$ with

$$\{f\}_\sigma = \frac{1}{2}(f_K + f_L).$$

Mainly vector quantities are defined on dual grid. We define only the s-th component g^s_σ of vector quantity $\mathbf{g}_h$ on each face $\sigma \in \mathcal{E}$, if $\mathbf{e}_s$ is the normal vector to the face F_σ. Then the projection to the primary grid reads

$$\bar{\mathbf{g}}_K = \frac{1}{2}\sum_{s=1}^{d}\left(g^s_{\sigma,s+} + g^s_{\sigma,s-}\right)\mathbf{e}_s. \tag{6}$$

2.2.3 Extending discrete quantities

We will compute numerical solutions using decreasing discretization parameter and investigate the weak limit of the numerical solutions, considering these being L^p functions. For this purpose we interpret discrete quantities defined in primary mesh $\mathcal{T}$ as piecewise constant functions with respect to this mesh, defined by

$$f_h(\mathbf{x}) = f_K, \text{ for } \mathbf{x} \in Q_K.$$

We denote the space of piecewise constant functions with respect to the grid $\mathcal{T}$ by

$$X(\mathcal{T}) = \{f \in L^\infty(\Omega); f|_K \equiv f_K \in \mathbb{R}\}.$$

Discrete vector quantity defined component-wise on dual mesh (g^s_σ) can be also identified with piecewise constants, which is

$$g^s_h(\mathbf{x} - \frac{h}{2}\mathbf{e}_s) = g^s_{\sigma,s-}, \text{ for } \mathbf{x} \in Q_K. \tag{7}$$

for all $\sigma \in \mathcal{E}$. Note that the s-th component of $\mathbf{g}$ is constant in the neighbourhood Q_σ of σ, which is the center of the face F_σ. The space of such functions is denoted by $X(\mathcal{E})^d$, we also define

$$X(\mathcal{E}_{\text{int}})^d = \left\{\mathbf{g} \in X(\mathcal{E})^d; \mathbf{g}|_{\mathcal{E}_{\text{ext}}} = \mathbf{0}\right\}.$$

To indicate the mesh-dependence of these functions, here and hereafter we equip them with subscript h. This subscript will be omitted any time where values at particular points of the mesh are considered. Then, for all $f_h \in X(\mathcal{T})$, $\mathbf{g}_h \in X(\mathcal{E}_{\text{int}})^d$ we have

$$h^d \sum_{K\in\mathcal{T}} f_K = \int_\Omega f_h \,\mathrm{d}x, \qquad h^d \sum_{\sigma\in\mathcal{E}_{\text{int}}} g^s_\sigma \mathbf{e}_s = \int_\Omega \mathbf{g}_h \,\mathrm{d}x. \tag{8}$$

Besides the piecewise constant extension $\mathbf{g}_h \in X(\mathcal{E}_{\text{int}})^d$, we will need also an extension to the space of functions with piecewise constant first order derivatives, i.e. piecewise linears with respect to primary cells,

$$\widehat{\mathbf{g}}(\mathbf{x}) = \sum_{s=1}^{d} \left((g^s_{\sigma,s+} - g^s_{\sigma,s-}) \left(\frac{x_s}{h} - \left\lfloor \frac{x_s}{h} \right\rfloor \right) + g^s_{\sigma,s-} \right) \mathbf{e}_s, \qquad \text{for } \mathbf{x} = (x_1, \ldots, x_d) \in Q_K,$$

where $\lfloor \cdot \rfloor$ is the floor rounding operator. The values of discrete quantities outside Ω that we use, will be extrapolated according to the boundary conditions, see Section 2.5.

2.2.4 Projection of continuous quantities to the grid

We will also need to project smooth quantities to our grids. We define the projection operators $\Pi^P : L^1(\Omega) \to X(\mathcal{T})$ and $\Pi^D : W_0^{1,1}(\Omega; \mathbb{R}^d) \to X(\mathcal{E}_{\text{int}})^d$ with

$$(\Pi^P \phi)_K = \frac{1}{h^d} \int_{Q_K} \phi \, \mathrm{d}x, \qquad (\Pi^D \mathbf{v})_\sigma = \frac{\mathbf{e}_s}{h^{d-1}} \int_{F_\sigma} v^s \, \mathrm{d}S_x.$$

Note that $\mathbf{v} \in (W^{1,1}(\Omega))^d$ is the minimal requirement so that $\mathbf{v}$ has bounded traces and the projection Π^D is well defined. The zero trace at $\partial\Omega$ guarantees that $\Pi^D \mathbf{v}|_\sigma = \mathbf{0}$ for $\sigma \in \mathcal{E}_{\text{ext}}$.

The projection to the primary grid satisfies

$$\sum_{K \in \mathcal{T}} (\Pi^P \phi)_K = \int_\Omega \phi \, \mathrm{d}x, \tag{9}$$

and using Taylor expansion and (7), one can derive the following estimates,

$$\|\Pi^P \phi - \phi\|_{L^p(\Omega)} \le h \|\nabla \phi\|_{L^p(\Omega)}, \qquad \|\Pi^D \mathbf{v} - \mathbf{v}\|_{L^p(\Omega;\mathbb{R}^d)} \le h \|\nabla \mathbf{v}\|_{L^p(\Omega)}. \tag{10}$$

2.3 Standard Difference Operators

2.3.1 Definitions

In this paper we use two basic difference operators

$$(\partial^s_h f)_\sigma = \frac{f_L - f_K}{h}, \qquad \text{for } f_h \in X(\mathcal{T}), \tag{11}$$

$$(\partial^s_h g^s)_K = \frac{g^s_{\sigma,s+} - g^s_{\sigma,s-}}{h}, \qquad \text{for } \mathbf{g}_h \in X(\mathcal{E}_{\text{int}})^d. \tag{12}$$

A property worth noticing is that the discrete derivatives and therefore also all first order differential operators can be viewed as mappings between the grids. The mixed derivative is defined as

$$(\partial^r_h g^s)_{K + \frac{h}{2}\mathbf{e}_s \pm \frac{h}{2}\mathbf{e}_r} = \mp \frac{g^s_{K+\frac{h}{2}\mathbf{e}_s} - g^s_{K+\frac{h}{2}\mathbf{e}_s \pm h\mathbf{e}_r}}{h}, \qquad \text{for } \mathbf{g}_h \in X(\mathcal{E}_{\text{int}})^d \text{ and every } K \in \mathcal{T}. \tag{13}$$

Notice that (13) can cover (12) if $r = s$ and $K + \frac{h}{2}\mathbf{e}_s \pm \frac{h}{2}\mathbf{e}_r \in \Omega$.
We can naturally define the discrete divergence operator with

$$(\mathrm{div}_h \mathbf{g})_K = \sum_{s=1}^{d} (\partial^s_h g^s)_K,$$

and Laplace operators by

$$(\Delta_h f)_K = (\mathrm{div}_h \partial^s f)_K = \frac{1}{h^2} \sum_{L \in \mathcal{N}(K)} (f_L - f_K), \qquad (\Delta_h g^s)_\sigma = \frac{1}{h^2} \sum_{r=1}^{d} (g^s_{\sigma - \mathbf{e}_r} - 2g^s_\sigma + g^s_{\sigma + \mathbf{e}_r}).$$

2.3.2 Calculus for the discrete operators

From the definition of differential operators one deduces the following two properties that are a discrete counterpart of the integration by parts.

Lemma 2.1. *Let $f_h \in X(\mathcal{T}), \mathbf{g}_h \in X(\mathcal{E}_{\text{int}})^d, \mathbf{v}_h \in X(\mathcal{E}_{\text{int}})^d$. Then*

$$\sum_{K \in \mathcal{T}} (\mathrm{div}_h \mathbf{g})_K f_K = - \sum_{\sigma \in \mathcal{E}_{\text{int}}} g^s_\sigma (\partial^s_h f)_\sigma. \tag{14}$$

$$- \sum_{\sigma \in \mathcal{E}_{\text{int}}} (\Delta_h v^s)_\sigma g^s_\sigma = \sum_{K \in \mathcal{T}} \sum_{s=1}^{d} \left((\partial^s_h g^s)_K (\partial^s_h v^s)_K + \frac{1}{2} \sum_{\substack{r=1 \\ r \neq s}}^{d} \sum_{i=1}^{2} (\partial^r_h g^s)_{K + \frac{h}{2}\mathbf{e}_s + (-1)^i \frac{h}{2}\mathbf{e}_r} (\partial^r_h v^s)_{K + \frac{h}{2}\mathbf{e}_s + (-1)^i \frac{h}{2}\mathbf{e}_r} \right). \tag{15}$$

The proof can be found in the Appendix A. Taking $\mathbf{v}_h = \mathbf{g}_h$ in (15) we obtain

$$-\sum_{\sigma\in\mathcal{E}_{\rm int}}(\Delta_h g^s)_\sigma g^s_\sigma = \sum_{K\in\mathcal{T}}\sum_{s=1}^{d}\left(|\partial_h^s g^s|^2_K + \frac{1}{2}\sum_{\substack{r=1\\ r\neq s}}^{d}\left(|\partial_h^r g^s|^2_{K+\frac{h}{2}\mathbf{e}_s+\frac{h}{2}\mathbf{e}_r} + |\partial_h^r g^s|^2_{K+\frac{h}{2}\mathbf{e}_s-\frac{h}{2}\mathbf{e}_r}\right)\right) =: \sum_{K\in\mathcal{T}}\sum_{s=1}^{d}\sum_{r=1}^{d}|\widetilde{\partial_h^r g^s}|^2_K. \tag{16}$$

2.3.3 Inverse estimates

Inverse estimate is a typical powerful tool for obtaining compactness result for a sequence of numerical solutions. We introduce its analogue for our finite difference setting in the following two lemmas.

Lemma 2.2. *Let $f_h \in X(\mathcal{T})$ and $\mathbf{g}_h \in X(\mathcal{E}_{\rm int})^d$. Then we have*

$$\|\partial_h^s f\|_{L^p(\Omega)} \le c(p)h^{-1}\|f\|_{L^p(\Omega)}, \qquad \|{\rm div}_h\mathbf{g}\|_{L^p(\Omega)} \le c(p)h^{-1}\|\mathbf{g}\|_{L^p(\Omega)},$$

with a positive constant $c(p)$, independent of h.

Proof. We observe, by virtue of the generalized triangle inequality, that

$$h^d\sum_{\sigma\in\mathcal{E}_{\rm int}}|(\partial_h^s f)_\sigma|^p = h^{d-p}\sum_{\sigma\in\mathcal{E}_{\rm int}}|f_L - f_K|^p \le c(p)h^{d-p}\sum_{K\in\mathcal{T}}|f_K|^p,$$
$$h^d\sum_{K\in\mathcal{T}}|({\rm div}_h\mathbf{g})_K|^p = h^{d-p}\sum_{K\in\mathcal{T}}|g^s_{\sigma,s+} - g^s_{\sigma,s-}|^p \le c(p)h^{d-p}\sum_{\sigma\in\mathcal{E}_{\rm int}}|g^s_\sigma|^p.$$

Using (8) concludes the proof. □

Lemma 2.3. *Let $p > q \ge 1$ and $f_h \in X(\mathcal{T}), \mathbf{g}_h \in X(\mathcal{E}_{\rm int})^d$. Then we have the estimate*

$$\|f\|_{L^p(\Omega)} \le c(p,q)h^{d(\frac{1}{p}-\frac{1}{q})}\|f\|_{L^q(\Omega)}, \qquad \|\mathbf{g}\|_{L^p(\Omega)} \le c(p,q)h^{d(\frac{1}{p}-\frac{1}{q})}\|\mathbf{g}\|_{L^q(\Omega)}.$$

Proof. We show the proof for $f \in X(\mathcal{T})$ only, leaving the other part to the kind reader. By definition, $\|f\|^p_{L^p(K)} = h^d|f_K|^p$, which implies $\|f\|_{L^p(K)} = h^{d(\frac{1}{p}-\frac{1}{q})}\|f\|_{L^q(K)}$. Then from the inequality

$$\sqrt[m]{S^m+1} \le S+1, \quad S\ge 0, m\ge 1, \qquad \text{setting } S = \frac{a^q}{b^q}, m = \frac{p}{q},$$

we deduce $\sqrt[p]{A^p+B^p} \le \sqrt[q]{A^q+B^q}$ and using induction also $\sqrt[p]{\sum_i a_i^p} \le \sqrt[q]{\sum_i a_i^q}$, which implies

$$\|f\|_{L^p(\Omega)} = \sqrt[p]{\sum_{K\in\mathcal{T}}\|f\|^p_{L^p(K)}} \le c(p,q)h^{d(\frac{1}{p}-\frac{1}{q})}\sqrt[p]{\sum_{K\in\mathcal{T}}\|f\|^p_{L^q(K)}} \le c(p,q)h^{d(\frac{1}{p}-\frac{1}{q})}\sqrt[q]{\sum_{K\in\mathcal{T}}\|f\|^q_{L^q(K)}} = c(p,q)h^{d(\frac{1}{p}-\frac{1}{q})}\|f\|_{L^q(\Omega)}.$$

□

Remark 1. *Analogously one would show that for any quantity f that is piecewise constant in time with respect to Δt-equidistant discretization of $[0,T]$ and any $p > q \ge 1$ it holds that*

$$\|f\|_{L^p(0,T)} \lesssim (\Delta t)^{(\frac{1}{p}-\frac{1}{q})}\|f\|_{L^q(0,T)}, \tag{17}$$

2.4 Upwind discretization and upwind derivative

The 'upwinding' or 'upstreaming' is a method vastly used in finite volume schemes for discretizing flow quantities. For its locally conservative properties (see [9, Section 1.1]), it appears useful in wider set of methods. First, we set $f^+ = \max\{0, f\}, \quad f^- = \min\{0, f\}$. Then, we can write $f = f^+ + f^-, \quad f^+ = \frac{1}{2}(f+|f|)$ and $f^- = \frac{1}{2}(f-|f|)$.

Let $\mathbf{u}_h \in X(\mathcal{E}_{\rm int})^d$ and $\sigma = K|L, L = K + h\mathbf{e}_s, s = 1,\dots,d$. Then we define the upwind flux of the quantity $f \in X(\mathcal{T})$ with respect to velocity $\mathbf{u}$ by

$${\rm Up}[f,\mathbf{u}]_\sigma = f_K(u^s_\sigma)^+ + f_L(u^s_\sigma)^-,$$

and the *upwind discrete derivative* and the upwind divergence with

$$\partial_s^{\rm Up}[f,\mathbf{u}]_K = \frac{{\rm Up}[f,\mathbf{u}]_{\sigma,s+} - {\rm Up}[f,\mathbf{u}]_{\sigma,s-}}{h}, \qquad {\rm div}_{\rm Up}[f,\mathbf{u}]_K = \sum_{s=1}^{d}\partial_s^{\rm Up}[f,\mathbf{u}]_K.$$

The following lemma is then a simple corollary of Lemma 2.1.

Lemma 2.4. *Let $f_h \in X(\mathcal{T}), \mathbf{v}_h = [v^1, \cdots, v^d] \in X(\mathcal{E}_{\rm int})^d$, then* $\sum_{K\in\mathcal{T}} \mathrm{div}_{\rm Up}[f,\mathbf{v}]_K = 0.$

The next lemma shows the difference between upwinding and averaging. It can be obtained by direct calculation.

Lemma 2.5. *Let $f \in X(\mathcal{T}), \mathbf{v} = [v^1, v^2, v^3] \in X(\mathcal{E}_{\rm int})^d$. Then,*

$$\mathrm{Up}[f,\mathbf{v}]_\sigma = \{f\}_\sigma (v^s)_\sigma - \frac{h}{2}|v^s_\sigma|(\partial^s_h f)_\sigma.$$

2.5 The Method

We introduce the following implicit scheme,

$$\partial^t_h \varrho^n_K + \mathrm{div}_{\rm Up}[\varrho^n, \mathbf{u}^n]_K - h^\alpha(\Delta_h \varrho^n)_K = 0, \tag{18}$$

$$\partial^t_h(\{\varrho\bar{u}\}_\sigma)^n + \{\mathrm{div}_{\rm Up}[\varrho^n\bar{\mathbf{u}}^n, \mathbf{u}^n]\}_\sigma + \big(\partial^s_h p(\varrho^n)\big)_\sigma \mathbf{e}_s - \mu(\Delta_h \mathbf{u}^n)_\sigma - h^\alpha \sum_{r=1}^d \big\{\partial^r_h\big(\{\bar{\mathbf{u}}^n\}\partial^r_h\varrho^n\big)\big\}_\sigma = 0, \tag{19}$$

for all $K \in \mathcal{T}, \sigma \in \mathcal{E}_{\rm int}$ and $n = \{1, \ldots, N_t\}$, with initial values

$$\varrho^0_K = \Pi^P \varrho_0, \qquad \bar{\mathbf{u}}^0_K = \Pi^P \mathbf{u}_0. \tag{20}$$

and boundary conditions

$$\mathbf{u}^n_\sigma = 0, \qquad (\mathbf{n}\cdot\nabla_h\rho^n)_\sigma = 0, \qquad \text{for } \sigma \in \mathcal{E}_{\rm ext}, n = 0, \ldots N_t, \tag{21}$$

To be more specific, the boundary conditions are implemented as $\rho_{\sigma-\frac{h}{2}\mathbf{e}_s} = \rho_{\sigma+\frac{h}{2}\mathbf{e}_s}$ and $\mathbf{u}_{\sigma+\frac{h}{2}\mathbf{e}_r-\frac{h}{2}\mathbf{e}_s} = -\mathbf{u}_{\sigma+\frac{h}{2}\mathbf{e}_r+\frac{h}{2}\mathbf{e}_s}$ for any $\sigma \in \mathcal{E}_{\rm ext}$ and $r \neq s$.

The way of projecting the initial velocity is motivated by the fact that nothing like (9) holds true for Π^D and also that we do not need the initial velocity on the faces $\sigma \in \mathcal{E}$.

Remark 2. *There is no boundary condition for density on the continuous level. However we need to equip the scheme with the no flux boundary condition for the density due to the additional artificial diffusion term in the scheme, which regularizes the continuity equation.*

3 Existence, stability and energy estimates

We start with showing the stability of the numerical method and deriving energy estimates. Prior to that we introduce two auxiliary results.

3.1 Renormalized continuity equation

Under certain regularity assumptions, density and velocity that satisfy continuity equation are known to satisfy its *renormalized* form (see DiPerna, Lions [5] or [10, Proposition 4.2]). Here we introduce its discrete counterpart.

Lemma 3.1. *Let $(\varrho_h, \mathbf{u}_h)$ satisfy the discrete continuity equation (18). Then for any $B \in C^2(\mathbb{R})$, $(\varrho_h, \mathbf{u}_h)$ satisfy the discrete renormalized equation,*

$$h^d \sum_{K\in\mathcal{T}} \Big(\partial^t_h B(\varrho^n_K) + (B'(\varrho^n_K)\varrho^n_K - B(\varrho^n_K))(\mathrm{div}_h \mathbf{u}^n)_K + \mathcal{P}_K\Big) = 0, \tag{22}$$

where

$$\mathcal{P}_K = \Delta t \frac{B''(\varrho^\eta_K)}{2}|\varrho^n_K|^2 + \frac{1}{2}\sum_{s=1}^d \Big((h^\alpha + hu^{s\ +}_{\sigma,s-})B''(\varrho^{n,\star}_{\sigma,s-})|(\partial^s_h\varrho)_{\sigma,s-}|^2 + (h^\alpha - hu^{s\ -}_{\sigma,s+})B''(\varrho^{n,\star}_{\sigma,s+})|(\partial^s_h\varrho)_{\sigma,s+}|^2\Big). \tag{23}$$

The intermediate values $\varrho^\eta_K, \varrho^{n,\star}_{\sigma,s\pm}$ are from the Lagrangian remainders of Taylor expansions.

Proof. We multiply (18) with $B'(\varrho^n_K)$ and handle the uprising terms.

Step 1. Using the Taylor expansion for the discrete time derivative of $B(\varrho^n_K)$ we get

$$\partial^t_h B(\varrho^n_K) = \frac{B(\varrho^n_K) - B(\varrho^{n-1}_K)}{\Delta t} = B'(\varrho^n_K)\partial^t_h\varrho^n_K - \frac{\Delta t}{2}B''(\varrho^\eta_K)|\partial^t_h\varrho^n_K|^2,$$

i.e. the time derivative term yields the first terms in both (22) and (23).

Step 2. We omit the time index n which is constant along the whole rest of the proof.

As $\mathrm{div}_{\mathrm{Up}}[\varrho,\mathbf{u}]_K = \sum_{s=1}^d (\partial_h^s \mathrm{Up}[\varrho,\mathbf{u}])_K$, we will prove it for one component only, leaving the summation over s as the very last step of the proof. Using the notation $\sigma, s- = J|K$ and $\sigma, s+ = K|L$, we can write

$$\begin{aligned} B'(\varrho_K)(\partial_h^s \mathrm{Up}[\varrho,\mathbf{u}])_K &= \frac{B'(\varrho_K)}{h}\Big(\varrho_K u^{s\,+}_{\sigma,s+} + \varrho_L u^{s\,-}_{\sigma,s+} - \varrho_J u^{s\,+}_{\sigma,s-} - \varrho_K u^{s\,-}_{\sigma,s-}\Big) \\ &= \frac{B'(\varrho_K)}{h}\Big(\varrho_K (u^s_{\sigma,s+} - u^s_{\sigma,s-}) + u^{s\,-}_{\sigma,s+}(\varrho_L - \varrho_K) + u^{s\,+}_{\sigma,s-}(\varrho_K - \varrho_J)\Big). \end{aligned} \tag{24}$$

Taylor expansion gives

$$\begin{aligned} B(\varrho_L) - B(\varrho_K) &= B'(\varrho_K)(\varrho_L - \varrho_K) + \frac{1}{2}B''(\varrho^\star_{\sigma,s+})(\varrho_L - \varrho_K)^2 \\ B(\varrho_K) - B(\varrho_J) &= B'(\varrho_K)(\varrho_K - \varrho_J) - \frac{1}{2}B''(\varrho^\star_{\sigma,s-})(\varrho_K - \varrho_J)^2, \end{aligned} \tag{25}$$

which, having used the definition of discrete derivative, yields

$$\begin{aligned} \frac{1}{h}B'(\varrho_K)(\varrho_L - \varrho_K) &= (\partial_h^s B(\varrho))_{\sigma,s+} - \frac{1}{2h}B''(\varrho^\star_{\sigma,s+})(\varrho_L - \varrho_K)^2, \\ \frac{1}{h}B'(\varrho_K)(\varrho_K - \varrho_J) &= (\partial_h^s B(\varrho))_{\sigma,s-} + \frac{1}{2h}B''(\varrho^\star_{\sigma,s-})(\varrho_K - \varrho_J)^2. \end{aligned} \tag{26}$$

Substitution from (26) into (24) yields

$$\begin{aligned} B'(\varrho_K)\partial_h^s \mathrm{Up}[\varrho,\mathbf{u}]_K &= B'(\varrho_K)\varrho_K(\partial_h^s u^s)_K + u^{s\,-}_{\sigma,s+}(\partial_h^s B(\varrho))_{\sigma,s+} + u^{s\,+}_{\sigma,s-}(\partial_h^s B(\varrho))_{\sigma,s-} \\ &\quad - \frac{1}{2h}u^{s\,-}_{\sigma,s+}B''(\varrho^\star_{\sigma,s+})(\varrho_L - \varrho_K)^2 + \frac{1}{2h}u^{s\,+}_{\sigma,s-}B''(\varrho^\star_{\sigma,s-})(\varrho_K - \varrho_J)^2. \end{aligned} \tag{27}$$

The last two terms are a contribution to $\mathcal{P}$, while the first three are rewritten as

$$\begin{aligned} &B'(\varrho_K)\varrho_K(\partial_h^s u^s)_K + u^{s\,-}_{\sigma,s+}(\partial_h^s B(\varrho))_{\sigma,s+} + u^{s\,+}_{\sigma,s-}(\partial_h^s B(\varrho))_{\sigma,s-} \\ &\quad = \big(B'(\varrho_K)\varrho_K - B(\varrho_K)\big)(\partial_h^s u^s)_K \\ &\qquad + \frac{B(\varrho_K)}{h}\left(\underbrace{u^s_{\sigma,s+} - u^{s\,-}_{\sigma,s+}}_{u^{s\,+}_{\sigma,s+}} + \underbrace{u^{s\,+}_{\sigma,s-} - u^s_{\sigma,s-}}_{-u^{s\,-}_{\sigma,s-}}\right) + \frac{B(\varrho_L)}{h}u^{s\,-}_{\sigma,s+} - \frac{B(\varrho_J)}{h}u^{s\,+}_{\sigma,s-} \\ &\quad = \big(B'(\varrho_K)\varrho_K - B(\varrho_K)\big)(\partial_h^s u^s)_K + \partial_h^{\mathrm{Up}}[B(\varrho),\mathbf{u}]_K. \end{aligned} \tag{28}$$

Let us substitute (28) to (27), sum over s and over $K \in \mathcal{T}$. Thanks to Lemma 2.4 , we obtain (22).

Step 3. To conclude the proof we show that the artificial diffusion term will contribute to (23) only. By virtue of (25), we get

$$\begin{aligned} &- h^\alpha B'(\varrho_K)(\Delta_h \varrho)_K = -h^{\alpha-2}B'(\varrho_K)\left((\varrho_L - \varrho_K) - (\varrho_K - \varrho_J)\right) \\ &\quad = -h^\alpha(\Delta_h B(\varrho))_K + \frac{1}{2}h^{\alpha-2}B''(\varrho^\star_{\sigma,s+})(\varrho_L - \varrho_K)^2 + \frac{1}{2}h^{\alpha-2}B''(\varrho^\star_{\sigma,s-})(\varrho_K - \varrho_J)^2. \end{aligned} \tag{29}$$

Summing (29) over s and over $K \in \mathcal{T}$, the first term on the right-hand side vanishes due to Neumann boundary condition of the density, while the other two terms contribute to the pollution term (23). □

Note that $\mathcal{P}_\mathcal{K} \geq 0$ provided B is convex.

Remark 3. *One can weaken the assumptions on B in Lemma 3.1 and allow jumps of its second derivatives, paying the price that all $B''(\xi), \xi \in (a,b)$ in (23) are replaced by some $B_2(\xi) \in \mathrm{co}\{B''_-(z), B''_+(z)\}$, which are the one-sided second derivatives of B at ξ. Anyway, $P_K \geq 0$ as long as B is convex. The proof of such assertion remains the same as in Lemma 3.1, with one exception. Instead of the standard Taylor's Theorem one just uses its generalized version, see [26].*

3.2 Positivity of density

We show that the discrete density is positive. Motivated by Karper [27], we present a complete proof of the following lemma. The lemma plays a role of an induction step, where the initial step is $0 < \varrho_K^0 = h^{-d}\int_K \varrho_0 \,\mathrm{d}x$ for all $K \in \mathcal{T}$, since $\varrho_0 > 0$ by assumption.

Lemma 3.2. *Suppose that $\varrho_h^n \in X(\mathcal{T})$ and $\mathbf{u}_h^n \in X(\mathcal{E}_{\mathrm{int}})^d$ satisfy (18), where $\varrho_h^{n-1} > 0$ in Ω_h. Then*

$$\varrho_h^n > 0, \ \text{in } \Omega_h.$$

Proof. The proof is stated in two steps, and the first being its nonnegativity. We use the renormalized continuity equation (22) with the one-parametric family of functions

$$B_\eta(z) = \begin{cases} (-z)^\eta & \text{for } z < 0, \\ 0 & \text{for } z \geq 0, \end{cases}$$

for $\eta > 1$. Notice that every B_η satisfies the weakened assumptions of Lemma 3.1 in the sense of Remark 3, i.e. $B_\eta \in C^1(\mathbb{R})$ and B_η'' is a continuous function, with an exception in the form of a jump discontinuity at 0, but since B_η is convex, we have $P_K > 0$. Moreover, $\eta \to 1^+$ yields $B_\eta(z) \to B(z) = \max\{-z, 0\}$ and

$$B_\eta'(z)z - B_\eta(z) = (\eta - 1)(-z)^\eta \to 0, \qquad \text{as } \eta \to 1^+, \quad \text{for } z < 0, \tag{30}$$

while for $z \geq 0$ the convergence is satisfied trivially. Since by assumption $\varrho_K^0 > 0$, it remains to show the induction step. Then (22) together with $P_K > 0$ and $B_\eta(\varrho_K^{n-1}) = 0$ for all $K \in \mathcal{T}$ (since we assume $\varrho_K^{n-1} > 0$) yields

$$\sum_{K\in\mathcal{T}} B_\eta(\varrho_K^n) \leq -\Delta t \sum_{K\in\mathcal{T}} \left(B_\eta'(\varrho_K^n)\varrho_K^n - B_\eta(\varrho_K^n)\right)(\mathrm{div}_h \mathbf{u}^n)_K. \tag{31}$$

Sending $\eta \to 1^+$ in (31), one gets by virtue of (30) that

$$\sum_{K\in\mathcal{T}} \max\{-\varrho_K^n, 0\} \leq 0,$$

from which we conclude $\varrho_K^n \geq 0$ for any $K \in \mathcal{T}$.

Next we show that the density is strictly positive. Choose $K \in \mathcal{T}$ such that $\varrho_K^n \leq \varrho_L^n$ for all $L \in \mathcal{T}$. Then we have

$$\begin{aligned} \varrho_K^n - \varrho_K^{n-1} &= -\Delta t \mathrm{div}_{\mathrm{Up}}[\varrho^n, \mathbf{u}^n]_K + \Delta t h^\alpha (\Delta_h \varrho^n) \\ &\geq -\frac{\Delta t}{h}\sum_{s=1}^d \left(\varrho_K^n u_{\sigma_{s,+}}^s - \varrho_K^n u_{\sigma_{s,-}}^s + (\varrho_{K+h\mathbf{e}_s}^n - \varrho_K^n)u_{\sigma_{s,+}}^{s\,-} + (\varrho_K^n - \varrho_{K-h\mathbf{e}_s}^n)u_{\sigma_{s,-}}^{s\,+}\right) \\ &\geq -\Delta t \varrho_K^n (\mathrm{div}_h \mathbf{u}^n)_K \geq -\Delta t \varrho_K^n |(\mathrm{div}_h \mathbf{u}^n)_K|, \end{aligned} \tag{32}$$

where we have used the minimality of ϱ_K^n to estimate the last term on the first row and last two terms on the second row from below with 0. Then, from (32) we get

$$\varrho_L^n \geq \varrho_K^n \geq \frac{1}{1 + \Delta t |(\mathrm{div}_h \mathbf{u}^n)_K|}\varrho_K^{n-1} > 0, \qquad \text{for any } L \in \mathcal{T},$$

which concludes the proof. □

3.3 Energy estimates

For the upcoming energy estimates we will need to handle the convective term, where we use the following identity.

Lemma 3.3. *For the convective term from (19), the following identity holds,*

$$h^d \sum_{K\in\mathcal{T}} \mathrm{div}_{\mathrm{Up}}[\varrho^n \bar{\mathbf{u}}^n, \mathbf{u}^n]_K \cdot \bar{\mathbf{u}}_K = -h^d \sum_{\sigma\in\mathcal{E}_{\mathrm{int}}} \mathrm{Up}[\varrho^n, \mathbf{u}^n]_\sigma \left(\partial_h^s \frac{|\bar{\mathbf{u}}^n|^2}{2}\right)_\sigma + \mathcal{N}, \tag{33}$$

where $\mathcal{N}$, the numerical diffusion term reads

$$\mathcal{N} = \frac{h^{d+1}}{4} \sum_{\sigma\in\mathcal{E}_{\mathrm{int}}} |\mathrm{Up}[\varrho^n, \mathbf{u}^n]_\sigma||(\partial_h^s \bar{\mathbf{u}}^n)_\sigma|^2.$$

Proof. We omit the time index n for the sake of brevity. Applying Lemma 2.1, the left hand side $\mathcal{L}$ of (33) equals

$$\mathcal{L} = -h^d \sum_{\sigma\in\mathcal{E}_{\rm int}} \mathrm{Up}[\varrho\bar{\mathbf{u}},\mathbf{u}]_\sigma \cdot (\partial_h^s \bar{\mathbf{u}})_\sigma := h^{d-1} \sum_{\sigma\in\mathcal{E}_{\rm int}} \mathcal{L}_\sigma.$$

Considering $\sigma = K|L$, we can write

$$\begin{aligned}
\mathcal{L}_\sigma &= -\big(\varrho_K \bar{\mathbf{u}}_K u_\sigma^{s\,+} + \varrho_L \bar{\mathbf{u}}_L u_\sigma^{s\,-}\big)\cdot(\bar{\mathbf{u}}_L - \bar{\mathbf{u}}_K)\\
&= \varrho_K u_\sigma^{s\,+}\left(\frac{|\bar{\mathbf{u}}|_K^2}{2} + \frac{|\bar{\mathbf{u}}|_K^2}{2} - \bar{\mathbf{u}}_K\cdot\bar{\mathbf{u}}_L + \frac{|\bar{\mathbf{u}}_L|^2}{2} - \frac{|\bar{\mathbf{u}}_L|^2}{2}\right) + \varrho_L u_\sigma^{s\,-}\left(\frac{|\bar{\mathbf{u}}_K|^2}{2} - \frac{|\bar{\mathbf{u}}_K|^2}{2} + \bar{\mathbf{u}}_K\cdot\bar{\mathbf{u}}_L - \frac{|\bar{\mathbf{u}}_L|^2}{2} - \frac{|\bar{\mathbf{u}}_L|^2}{2}\right)\\
&= (\varrho_K u_\sigma^{s\,+} + \varrho_L u_\sigma^{s\,-})\left(\frac{|\bar{\mathbf{u}}_K|^2}{2} - \frac{|\bar{\mathbf{u}}_L|^2}{2}\right) + (\varrho_K u_\sigma^{s\,+} - \varrho_L u_\sigma^{s\,-})\left|\frac{\bar{\mathbf{u}}_K - \bar{\mathbf{u}}_L}{2}\right|^2\\
&= -h\mathrm{Up}[\varrho,\mathbf{u}]_\sigma\left(\partial_h^s\frac{|\bar{\mathbf{u}}|^2}{2}\right)_\sigma + \frac{h^2}{4}|\mathrm{Up}[\varrho,\mathbf{u}]_\sigma||(\partial_h^s\bar{\mathbf{u}})_\sigma|^2.
\end{aligned}$$

Summation over σ concludes the proof. □

Now we can deduce the following energy estimates on the numerical solution.

Theorem 3.4. *Let $(\varrho_h,\mathbf{u}_h)$ be the numerical solution obtained through the scheme (18–20). For any time step $m = 1,\dots,N_t$ the following stability estimate holds,*

$$h^d\sum_{K\in\mathcal{T}}\left(\varrho_K^m\frac{|\bar{\mathbf{u}}_K^m|^2}{2} + \frac{1}{\gamma-1}p(\varrho_K^m)\right) + \Delta t h^d\mu\sum_{n=1}^m\sum_{K\in\mathcal{T}}\sum_{r=1}^d\sum_{s=1}^d|\widetilde{\partial_h^r u^{s,n}}|_K^2 + \sum_{j=1}^4 N_j^m \le h^d\sum_{K\in\mathcal{T}}\left(\varrho_K^0\frac{|\bar{\mathbf{u}}_K^0|^2}{2} + \frac{1}{\gamma-1}p(\varrho_K^0)\right), \tag{34}$$

where

$$\begin{aligned}
N_1^m &= \Delta t h^d\sum_{n=1}^m\sum_{K\in\mathcal{T}}\sum_{s=1}^d\frac{1}{2}\Big((h^\alpha + h^2(u_{\sigma,s-}^{s,n})^+)p''(\varrho_{\sigma,s-}^{n,\star})|(\partial_h^s\varrho^n)_{\sigma,s-}|^2 + (h^\alpha - h^2(u_{\sigma,s+}^{s,n})^-)p''(\varrho_{\sigma,s+}^{n,\star})|(\partial_h^s\varrho^n)_{\sigma,s+}|^2\Big),\\
N_2^m &= (\Delta t)^2h^d\sum_{n=1}^m\sum_{K\in\mathcal{T}}\frac{p''(\varrho_K^\eta)}{2}|(\partial_h^t\varrho_K)^n|^2,\\
N_3^m &= (\Delta t)^2h^d\sum_{n=1}^m\sum_{K\in\mathcal{T}}\frac{\varrho_K^{n-1}}{2}|(\partial_h^t\bar{\mathbf{u}}_K)^n|^2,\\
N_4^m &= \Delta t h^{d+1}\frac{1}{4}\sum_{n=1}^m\sum_{\sigma\in\mathcal{E}_{\rm int}}|\mathrm{Up}[\varrho^n,\mathbf{u}^n]_\sigma||(\partial_h^s\bar{\mathbf{u}}^n)_\sigma|^2.
\end{aligned}$$

Proof. We take the scalar product of the discrete momentum equation (19) and $h^d(u^s)_\sigma^n\mathbf{e}_s$, sum over $\sigma\in\mathcal{E}_{\rm int}$ and handle term by term.

Time difference term. We use the notation $\sigma = K|L$ and the definition of projection to primary grid (6) to get

$$\frac{h^d}{2}\sum_{\sigma\in\mathcal{E}_{\rm int}}\partial_h^t(\varrho_K\bar{\mathbf{u}}_K + \varrho_L\bar{\mathbf{u}}_L)^n u_\sigma^{s,n}\cdot\mathbf{e}_s = h^d\sum_{K\in\mathcal{T}}\partial_h^t(\varrho_K\bar{\mathbf{u}}_K)^n\cdot\bar{\mathbf{u}}_K^n. \tag{35}$$

Convective term. Using the projection into primary grid (6), Lemma 3.3, summation by parts (14) and the continuity equation (18), we can write

$$\begin{aligned}
h^d\sum_{\sigma\in\mathcal{E}_{\rm int}}\frac{\mathrm{div}_{\rm Up}[\varrho^n\bar{\mathbf{u}}^n,\mathbf{u}^n]_K + \mathrm{div}_{\rm Up}[\varrho^n\bar{\mathbf{u}}^n,\mathbf{u}^n]_L}{2}\cdot(u^s)_\sigma^n\mathbf{e}_s &= h^d\sum_{K\in\mathcal{T}}\mathrm{div}_{\rm Up}[\varrho^n\bar{\mathbf{u}}^n,\mathbf{u}^n]_K\cdot\bar{\mathbf{u}}_K^n\\
&= -h^d\sum_{\sigma\in\mathcal{E}_{\rm int}}\mathrm{Up}[\varrho^n,\mathbf{u}^n]_\sigma\left(\partial_h^s\frac{|\bar{\mathbf{u}}^n|^2}{2}\right)_\sigma + \mathcal{N} = h^d\sum_{K\in\mathcal{T}}(\mathrm{div}_{\rm Up}[\varrho^n,\mathbf{u}^n])_K\frac{|\bar{\mathbf{u}}_K^n|^2}{2} + \mathcal{N}\\
&= -h^d\sum_{K\in\mathcal{T}}(\partial_h^t\varrho_K)^n\frac{|\bar{\mathbf{u}}_K^n|^2}{2} + h^{d+\alpha}\sum_{K\in\mathcal{T}}(\Delta_h\varrho^n)_K\frac{|\bar{\mathbf{u}}_K^n|^2}{2} + \mathcal{N}.
\end{aligned} \tag{36}$$

Pressure term. Using (14), one gets

$$h^d\sum_{\sigma\in\mathcal{E}_{\rm int}}(\partial_h^s p(\varrho^n))_\sigma\mathbf{e}_s\cdot(u^s)_\sigma^n = -h^d\sum_{K\in\mathcal{T}}p(\varrho_K^n)(\mathrm{div}_h\mathbf{u}^n)_K.$$

Then, we apply Lemma 3.1 with $B(z) = \frac{1}{\gamma-1}p(z)$ to deduce

$$-h^d \sum_{K\in\mathcal{T}} p(\varrho_K^n)(\mathrm{div}_h \mathbf{u}^n)_K = \frac{h^d}{\gamma-1}\sum_{K\in\mathcal{T}}(\partial_h^t p(\varrho_K))^n + h^d \sum_{K\in\mathcal{T}}(\mathcal{P}_K)^n. \tag{37}$$

Viscosity term. Direct application of (16) gives

$$-h^d\mu \sum_{\sigma\in\mathcal{E}_{\mathrm{int}}} (\Delta_h \mathbf{u}^n)_\sigma \cdot (u^s)_\sigma^n \mathbf{e}_s = \mu h^d \sum_{K\in\mathcal{T}}\sum_{s=1}^{d}\sum_{r=1}^{d} |\widetilde{\partial_h^s u^s}|_K^2. \tag{38}$$

Additional term. Using (6) and summation by parts (14), we can write

$$-h^{d+\alpha}\sum_{\sigma\in\mathcal{E}_{\mathrm{int}}}\sum_{r=1}^{d}\left\{\partial_h^r\left(\{\bar{\mathbf{u}}^n\}\partial_h^r\varrho^n\right)\right\}_\sigma\cdot(u^s)_\sigma^n\mathbf{e}_s = -h^{d+\alpha}\sum_{K\in\mathcal{T}}\sum_{r=1}^{d}\partial_h^r\left(\{\bar{\mathbf{u}}^n\}\partial_h^r\varrho^n\right)_K\cdot\bar{\mathbf{u}}_K^n = h^{d+\alpha}\sum_{\sigma\in\mathcal{E}_{\mathrm{int}}}\{\bar{\mathbf{u}}^n\}_\sigma(\partial_h^s\varrho^n)_\sigma\cdot(\partial_h^s\bar{\mathbf{u}}^n)_\sigma,$$

Then, employing

$$\{\bar{\mathbf{u}}^n\}_\sigma\cdot(\partial_h^s\bar{\mathbf{u}}^n)_\sigma = \frac{1}{2h}(\bar{\mathbf{u}}_L^n+\bar{\mathbf{u}}_K^n)\cdot(\bar{\mathbf{u}}_L^n-\bar{\mathbf{u}}_K^n) = \frac{|\bar{\mathbf{u}}_L^n|^2-|\bar{\mathbf{u}}_K^n|^2}{2h} = \left(\partial_h^s\frac{|\bar{\mathbf{u}}^n|^2}{2}\right)_\sigma,$$

to the chain of equalities above and using (14) with the no-flux boundary condition for density (21), we obtain

$$-h^{d+\alpha}\sum_{\sigma\in\mathcal{E}_{\mathrm{int}}}\sum_{r=1}^{d}\left\{\partial_h^r\left(\{\bar{\mathbf{u}}^n\}\partial_h^r\varrho^n\right)\right\}_\sigma\cdot(u^s)_\sigma^n\mathbf{e}_s = h^{d+\alpha}\sum_{\sigma\in\mathcal{E}_{\mathrm{int}}}\left(\partial_h^s\frac{|\bar{\mathbf{u}}^n|^2}{2}\right)_\sigma(\partial_h^s\varrho^n)_\sigma = -h^{d+\alpha}\sum_{K\in\mathcal{T}}(\Delta_h\varrho^n)_K\frac{|\bar{\mathbf{u}}_K^n|^2}{2}. \tag{39}$$

Final step. We observe the identity

$$\partial_h^t(\varrho_K\bar{\mathbf{u}}_K)^n\bar{\mathbf{u}}_K^n - \partial_h^t\varrho_K^n\left(\frac{|\bar{\mathbf{u}}_K^n|^2}{2}\right) = \partial_h^t\left(\varrho_K\frac{|\bar{\mathbf{u}}_K|^2}{2}\right)^n + \varrho_K^{n-1}\frac{|\bar{\mathbf{u}}_K^n-\bar{\mathbf{u}}_K^{n-1}|^2}{2}. \tag{40}$$

Finally, we collect the right-hand sides of (35–39), employ (40), multiply by Δt and sum over time to obtain the desired result. Notice that the artificial diffusion terms get canceled out. □

3.4 Existence of the numerical solution

As the numerical scheme (18-19) is implicit and nonlinear, the existence of its solution (i.e. of the quantities in the next step) is not a priori known. We prove it in the upcoming section using Schaeffer's fixed point theorem, see e.g. [8, Theorem 9.2.4]. Note that nothing about the uniqueness of the solution is claimed.

Theorem 3.5 (Schaeffer's fixed point theorem)**.** *Let $\mathcal{S} : Z \to Z$ be a continuous mapping defined on a finite-dimensional space Z and let the set*

$$\{z\in Z, z=\kappa\mathcal{S}(z), \kappa\in[0,1]\},$$

be nonempty and bounded. Then there exists $z \in Z$ such that

$$z = \mathcal{S}(z).$$

Before stating the existence theorem, we prove an auxiliary lemma concerning the viscosity term.

Lemma 3.6. *Let $\mathcal{L} : X(\mathcal{E}_{\mathrm{int}})^d \to X(\mathcal{E}_{\mathrm{int}})^d$ be a linear mapping given by*

$$(\mathcal{L}(\mathbf{v}))_\sigma := (\Delta_h\mathbf{v})_\sigma. \tag{41}$$

Then its inverse operator $\mathcal{L}^{-1}$ is bounded with constant depending on the discretization parameter h.

Proof. Note that for fixed h $X(\mathcal{E}_{\mathrm{int}})^d$ is a finite-dimensional space and thus all norms are equivalent. Therefore, we aim at proving

$$\|\mathcal{L}(\mathbf{v})\|_\infty \geq c(h) > 0, \qquad \text{for all } \mathbf{v} = (v^1, v^2, v^3) \in X(\mathcal{E}_{\mathrm{int}})^d, \|\mathbf{v}\|_\infty = 1. \tag{42}$$

From $\|\mathbf{v}\|_\infty = 1$ we have that $|v_\sigma^s| = 1$ for some $\sigma \in \mathcal{E}_{\mathrm{int}}$. Without loss of generality we may assume that $v_\sigma^s = -1$. And as $v_{\sigma'}^s = 0$ when $\sigma' \in \mathcal{E}_{\mathrm{ext}}$, there exist $K_1, K_2 \in \mathcal{T}$ such that

$$(\partial^s v^s)_{K_2} \geq \frac{h}{N_x} = \frac{h^2}{L_x}, \qquad \text{and} \quad (\partial^s v^s)_{K_1} \leq -\frac{h^2}{L_x},$$

where K_1, K_2 differ only in the s-component and $hN_x = L_x$. Therefore, using the same argument as before, there exists $\widetilde{\sigma} \in \mathcal{E}_{\rm int}$ such that

$$\|(\Delta_h \mathbf{v})\|_\infty \geq |(\partial_h^s \partial_h^s \mathbf{v})_{\widetilde{\sigma}}| \geq \frac{(\partial^s v^s)_{K_2} - (\partial^s v^s)_{K_1}}{L_x} \geq \frac{2h^2}{L_x^2},$$

which is (42) with $c(h) = \frac{2h^2}{L_x^2}$. □

Theorem 3.7. *Let $p(\varrho) = a\varrho^\gamma$ and $\varrho_h^{n-1} \in X(\mathcal{T})$, $\mathbf{u}_h^{n-1} \in X(\mathcal{E}_{\rm int})^d$ be given; $\varrho_K^{n-1} > 0$ for all $K \in \mathcal{T}$. Then the numerical scheme (18-19) admits a solution*

$$\varrho_h^n \in X(\mathcal{T}), \varrho_K^n > 0 \text{ for all } K \in \mathcal{T}, \mathbf{u}_h^n \in X(\mathcal{E}_{\rm int})^d.$$

Moreover, is satisfies the discrete conservation of mass

$$\sum_{K \in \mathcal{T}} \varrho_K^n = \sum_{K \in \mathcal{T}} \varrho_K^{n-1}. \tag{43}$$

Proof. We show the existence in two steps. We treat the continuity equation first.

Step 1. We claim, that for ϱ_h^{n-1} given, the continuity scheme (18) provides a unique solution depending continuously on the parameter $\mathbf{u}_h^n \in X(\mathcal{E}_{\rm int})^d$.

In fact, for all $K \in \mathcal{T}$ (18) builds a system of N_e linear equations with N_e unknowns, where N_e denotes the number of points in the primary mesh, where ϱ_K^{n-1} represents the (known) right-hand side and $\mathbf{u}_h^n$ is a parameter.

The associated homogeneous problem

$$\varrho_K^n + \Delta t \mathrm{div}_{\rm Up}[\varrho^n, \mathbf{u}^n]_K - \Delta t\, h^\alpha (\Delta_h \varrho^n)_K = 0, \tag{44}$$

admits a unique solution and hence the trivial one. It is easy to verify that $\varrho_h^n \equiv 0$ indeed solves (44). To show uniqueness one uses the same procedure as in the proof of Lemma 3.2 to get

$$\sum_{K \in \mathcal{T}} \max\{-\varrho_K^n, 0\} \leq 0,$$

and hence $\varrho_K^n = 0$ for all $K \in \mathcal{T}$.

Therefore, for given ϱ_h^{n-1}, the continuity scheme (18) supplies us with a unique solution $\varrho_h^n = \varrho_h^n(\mathbf{u}_h^n)$, where the mapping

$$\mathbf{u}_h^n \mapsto \varrho_h^n(\mathbf{u}_h^n),$$

is continuous in $X(\mathcal{E}_{\rm int})^d$. Moreover, Lemma 3.2 gives $\varrho_h^n > 0$. The discrete conservation of mass (43) can be obtained by simple summation of the discrete continuity equation (18) over all $K \in \mathcal{T}$.

Step 2. We rewrite the momentum scheme (19) as follows,

$$\mu(\Delta_h \mathbf{u}^n)_\sigma \mathbf{e}_r = \kappa \mathcal{F}_\sigma(\mathbf{u}_h^n), \qquad \sigma \in \mathcal{E}_{\rm int}, \tag{45}$$

where

$$\mathcal{F}_\sigma(\mathbf{u}_h^n) := -\frac{\{\varrho^n[\mathbf{u}_h^n]\bar{\mathbf{u}}^n\}_\sigma - \{\varrho^{n-1}\bar{\mathbf{u}}^{n-1}\}_\sigma}{\Delta t} - \{\mathrm{div}_{\rm Up}[\varrho^n[\mathbf{u}_h^n]\bar{\mathbf{u}}^n, \mathbf{u}^n]\}_\sigma - (\partial_h^s p(\varrho^n[\mathbf{u}_h^n]))_\sigma + h^\alpha \sum_{r=1}^d \{\partial_h^r(\{\bar{\mathbf{u}}^n\}\partial_h^r \varrho^n[\mathbf{u}_h^n])\}_\sigma.$$

Note that $\bar{\mathbf{u}}^{n-1}$ was determined in the previous step and $\bar{\mathbf{u}}^0$ is given by the initial conditions (20). We define $\mathcal{F} := (\mathcal{F}_\sigma)_{\{\sigma \in \mathcal{E}_{\rm int}\}}$ together with $\mathbf{u}_{\sigma'}^n = 0$ for $\sigma' \in \mathcal{E}_{\rm ext}$.

We are searching for $\mathbf{u}_h^n$ being a fixed point of the mapping $\mathcal{F} \circ \mathcal{L}^{-1}$, with $\mathcal{L}$ defined by (41). We verify the assumptions of the Schaeffer's fixed point theorem (Theorem 3.5). As $\mathcal{F}$ is clearly continuous and $\mathcal{L}^{-1}$ is linear and bounded, their composition is continuous in the finite dimensional space $X(\mathcal{E}_{\rm int})^d$. Any possible solution $\mathbf{u}_{h,\kappa}^n$ of (45) is indeed a solution of the momentum scheme with the diffusion constant enlarged to $\frac{\mu}{\kappa}$, i.e. the energy estimate (34), with μ replaced by μ/κ, implies that $\mathcal{F}(\mathbf{u}_{h,\kappa}^n)$ is bounded in $X(\mathcal{E}_{\rm int})^d$, independently of κ. The boundedness of $\mathcal{L}^{-1}$ further implies that also

$$\left\{\mathbf{u}_{h,\kappa}^n \in X(\mathcal{E}_{\rm int})^d, \quad \mathbf{u}_{h,\kappa}^n = \kappa \mathcal{F} \circ \mathcal{L}^{-1}(\mathbf{u}_{h,\kappa}^n)\right\}, \tag{46}$$

is bounded independently of κ. Note that the set in (46) is nonempty, as zero obviously solves (45) with $\kappa = 0$.

□

3.5 Uniform bounds

The convergence proof requires some compactness results which are usually gained through uniform bounds of approximate quantities. In sequel, the notation $A \lesssim B$ means $A \leq cB$, where $c > 0$ is a constant that does not depend on the discretization parameter h. $A \approx B$ means $A \lesssim B$ and $B \lesssim A$.

The energy estimate (34) allows us to establish the following uniform bounds.

Proposition 3.8. *Let $(\varrho_h, \mathbf{u}_h)$ be a numerical solution obtained through the scheme (18)–(20) and let the total initial energy D be defined by*

$$D = \int_\Omega \frac{1}{2}\varrho_0 \mathbf{u}_0^2 + \frac{1}{\gamma - 1} p(\varrho_0)\,\mathrm{d}x. \tag{47}$$

Then

$$\|\varrho_h\|_{L^\infty(L^\gamma(\Omega))} \lesssim D, \qquad \|p(\varrho_h)\|_{L^\infty(L^1(\Omega))} \lesssim D, \qquad \|\sqrt{\varrho_h}\bar{\mathbf{u}}_h\|_{L^\infty(L^2(\Omega))} \lesssim D. \tag{48}$$

Note that the constant D depends solely on the initial data $(\varrho_0, \mathbf{u}_0)$.

Proof. Due to the convexity of $p(z)$, all the terms in the left hand side of (34) are non-negative, hence every single one can be estimated by the right-hand side of (34).

Further, it is the definition of initial conditions to the numerical scheme (20), property (9) and the inequality $\|\varrho_0\|_{L^1(\Omega)} \leq \|\varrho_0\|_{L^\gamma(\Omega)}$ that guarantee

$$\frac{h^d}{\gamma - 1} \sum_{K \in \mathcal{T}} p(\varrho_K^0) \leq \frac{1}{\gamma - 1} \int_\Omega p(\varrho_0)\,\mathrm{d}x.$$

Then we apply the Jensen's inequality on each cell twice to get also

$$h^d \sum_K \varrho_K^0 |\bar{\mathbf{u}}_K^0|^2 \leq \sum_{K \in \mathcal{T}} \varrho_K \left(\int_{Q_K} |\mathbf{u}_0|^2\,\mathrm{d}x \right) \leq \sum_{K \in \mathcal{T}} h^{-d} \int_{Q_K} \int_{Q_K} \varrho_0 |\mathbf{u}_0|^2\,\mathrm{d}x\,\mathrm{d}y = \sum_{K \in \mathcal{T}} \int_{Q_K} \varrho_0 |\mathbf{u}_0|^2\,\mathrm{d}x = \int_\Omega \varrho_0 |\mathbf{u}_0|^2\,\mathrm{d}x.$$

□

The reader can observe a slight abuse of notation, concerning the Bochner spaces. Since we have not defined the extension of our discrete quantities to integrable functions in time, keep in mind that the equiintegrability of some v_h in a Bochner space $L^q(0, T; X)$ should be understood as

$$\left(\Delta t \sum_{n=1}^{N_t} (\|v^n\|_X)^q \right)^{\frac{1}{q}} \leq c,$$

which corresponds to the standard Bochner norm for the piecewise constant extension in time.

Using Hölder inequality one deduces from (48) also

$$\|\varrho_h \bar{\mathbf{u}}_h\|_{L^\infty\left(L^{\frac{2\gamma}{\gamma+1}}(\Omega;\mathbb{R}^d)\right)} \lesssim D. \tag{49}$$

3.6 Discrete inequality of the Sobolev type and velocity estimates

Similarly to the continuous case, we would like to obtain information about better (equi)integrability of $\mathbf{u}_h$. For this purpose we introduce a version of the discrete Sobolev embedding theorem. Both the claim and its proof are inspired by an analogous assertion from [2, Lemma 1]. Prior to that, we introduce the following auxiliary algebraic inequality.

Lemma 3.9. *For any $a, b \in \mathbb{R}$ and any $p > 2$ the following inequality holds,*

$$\left| |a|^{p-1}a - |b|^{p-1}b \right| \leq \frac{p}{2} \left(|a|^{p-1} + |b|^{p-1} \right) |a - b|. \tag{50}$$

Proof. Without loss of generality we can assume that $a \geq b$, then it holds, that

$$|a|^{p-1}a - |b|^{p-1}b \geq 0. \tag{51}$$

It can be shown through discussing the signs of a and b:

1. Let $a \geq 0, b \geq 0$. Then the left-hand side of (51) equals $a^p - b^p \geq b^{p-1}(a - b) \geq 0$.
2. Let $a \geq 0, b < 0$. Then $a^p - |b|^{p-1}b \geq 0$.

3. Let $a<0, b<0$. Then $(-b)|b|^{p-1}-(-a)|a|^{p-1} \geq |b|^{p-1}(a-b) \geq 0$.

Therefore, it remains to show that $|a|^{p-1}a - |b|^{p-1}b \leq \frac{p}{2}(|a|^{p-1}+|b|^{p-1})(a-b)$. We will use Taylor expansion of the function $f(x) := |x|^{p-1}x$, notice that $f'(x) = p|x|^{p-1}$ and $f''(x) = p(p-1)|x|^{p-3}x$ is increasing.

Then

$$\begin{aligned} |x|^{p-1}x &= |a|^{p-1}a + p|a|^{p-1}(x-a) + \frac{1}{2}f''(\zeta_a)(x-a)^2, \\ |x|^{p-1}x &= |b|^{p-1}b + p|b|^{p-1}(x-b) + \frac{1}{2}f''(\zeta_b)(x-b)^2. \end{aligned} \tag{52}$$

We take $x = \frac{1}{2}(a+b)$ and subtract the equations in (52) to obtain

$$|a|^{p-1}a - |b|^{p-1}b = p(|a|^{p-1}+|b|^{p-1})\frac{a-b}{2} + (f''(\zeta_b) - f''(\zeta_a))\frac{(a-b)^2}{4}. \tag{53}$$

As $\zeta_a \geq \zeta_b$ and f'' is increasing, the last term on the right-hand side of (53) is negative which, together with (51), recovers (50). □

Now we can prove the Sobolev-type inequality for discrete quantities.

Proposition 3.10. *Let* $\mathbf{w} = (w^1, \cdots, w^d) \in X(\mathcal{E}_{\text{int}})^d$*, then the following inequality holds*

$$h^d \left(\sum_{\sigma \in \mathcal{E}_{\text{int}}} |w^s_\sigma|^q \right)^{\frac{2}{q}} \lesssim h^d \sum_{r=1}^{d}\sum_{s=1}^{d}\sum_{K\in\mathcal{T}} |\widetilde{\partial^r_h w^s}|^2 =: \|\widetilde{\nabla_h \mathbf{w}_h}\|_2^2 \qquad \text{for } \begin{cases} d=2, & q\in[1,\infty), \\ d=3, & q\in[1,6]. \end{cases}$$

Proof. We start with $d=3$, $\mathbf{v} \in X(\mathcal{E}_{\text{int}})^d$, $\mathbf{v}|_{\partial\Omega} = 0$, whose relation to $\mathbf{w}$ will be specified later. Any component v^s of $\mathbf{v}$ can be expressed, by virtue of the definition (13), as

$$v^s_\sigma(x) = h \sum_{K\in\mathcal{T}} (\partial^r v^s)_K \chi^{r,s}_K(x), \tag{54}$$

for any $r = 1, \ldots, d$, where the characteristic function $\chi^{r,s}_K$ equals one at $x \in \sigma$, for which K participates on creating the value v_σ and zero otherwise. In particular, if $r = s$, we define

$$\chi^{s,s}_K(x) = \begin{cases} 1 & \text{if } x \in Q_\sigma : (\sigma - K)\cdot \mathbf{e}_s \geq 0 \ \wedge\ (\sigma-K)\cdot\mathbf{e}_p = 0, \forall p \in \{1,\ldots,d\}\setminus\{s\}, \\ 0 & \text{otherwise,} \end{cases} \tag{55}$$

and for $r \neq s$

$$\chi^{r,s}_K(x) = \begin{cases} 1 & \text{if } x \in Q_\sigma : (\sigma - K)\cdot \mathbf{e}_s = \frac{1}{2} \ \wedge\ (\sigma-K)\cdot\mathbf{e}_r \geq 0 \ \wedge\ (\sigma-K)\cdot\mathbf{e}_p = 0, p \in \{1,\ldots,d\}\setminus\{r,s\}, \\ 0 & \text{otherwise.} \end{cases} \tag{56}$$

We comment on the definitions (55–56), that since every x belongs to three distinct cubes Q_σ, we pick always the one, whose face F_σ has the normal vector $\mathbf{e}_s$, where s is indicated by the second item at the upper index of $\chi^{\cdot,s}_K$ and was fixed at the beginning of the proof.

Integrating (54) over Ω and estimating the characteristic functions $\chi^{s,s}_K$ from above yields

$$\int_\Omega |v^s_\sigma| \,\mathrm{d}x \leq h \sum_{K\in\mathcal{T}} |\partial^s v^s|_K h^{d-1} \leq h^d \sum_{K\in\mathcal{T}} |\partial^s v^s|_K. \tag{57}$$

Further, denoting $\dot{K} = K + \frac{h}{2}\mathbf{e}_s - \frac{h}{2}\mathbf{e}_{r_1}$ and $\ddot{K} = K + \frac{h}{2}\mathbf{e}_s - \frac{h}{2}\mathbf{e}_{r_2}$, we can express

$$|v^s_\sigma(x)|^2 = \sum_{K\in\mathcal{T}} (\partial^{r_1} v_s)_{\dot{K}} h\chi^{r_1,s}_K(x) \sum_{K\in\mathcal{T}} (\partial^{r_2} v_s)_{\ddot{K}} h\chi^{r_2,s}_K(x) \leq \left(\sum_{K\in\mathcal{T}} |(\partial^{r_1} v_s)_{\dot{K}}| h\overline{\chi}^{r_1,s}_K(x)\right)\left(\sum_{K\in\mathcal{T}} |(\partial^{r_2} v_s)_{\ddot{K}}| h\overline{\chi}^{r_2,s}_K(x)\right)$$

with three mutually distinct indices r_1, r_2, s, where

$$\overline{\chi}^{r_i,s}_K(x) = \begin{cases} 1 & \text{if } x \in Q_\sigma : (\sigma - K)\cdot \mathbf{e}_s = \frac{h}{2} \ \wedge\ (\sigma-K)\cdot\mathbf{e}_p = 0, p \in \{1,\ldots,d\}\setminus\{r_i,s\}, \\ 0 & \text{otherwise,} \end{cases} \tag{58}$$

which is a dominating function to $\chi^{r_i,s}_K$, independent of r_i. In particular, $\overline{\chi}^{r_1,s}_K(x), \overline{\chi}^{r_2,s}_K(x)$ depend only on (x_{r_2}, x_s), (x_{r_1}, x_s), respectively. Thus we can compute

$$
\begin{aligned}
&\int_{\mathbb{R}}\int_{\mathbb{R}}(v_\sigma^s)^2\,\mathrm{d}x_{r_1}\,\mathrm{d}x_{r_2}\\
&\quad\le \sum_{K\in\mathcal{T}} h|(\partial_h^{r_1}v^s)_{\dot K}|\int_{\mathbb{R}}\overline{\chi}_K^{r_1,s}(x_{r_2},x_s)\,\mathrm{d}x_{r_2}\sum_{K\in\mathcal{T}} h|(\partial_h^{r_2}v^s)_{\ddot K}|\int_{\mathbb{R}}\overline{\chi}_K^{r_2,s}(x_{r_1},x_s)\,\mathrm{d}x_{r_1}\\
&\quad= h^4\sum_{K\in\mathcal{T}}|(\partial_h^{r_1}v^s)_{\dot K}|\sum_{K\in\mathcal{T}}|(\partial_h^{r_2}v^s)_{\ddot K}|(\mathbf{1}_{K-\frac{h}{2}\mathbf{e}_s}(x_s))^2,
\end{aligned}
\tag{59}
$$

where we used the fact that after integrating with respect to x_{r_1}, x_{r_2}, the functions $\overline{\chi}_K^{r_1,s}(x), \overline{\chi}_K^{r_2,s}(x)$ leave their projections to the line x_s, which are in both cases equal to $\mathbf{1}_{K-\frac{h}{2}\mathbf{e}_s}(x_s)$. Integrating (59) over the remaining variable x_s, we get

$$
\int_{\mathbb{R}^d}|v_\sigma^s|^2\,\mathrm{d}x=\int_\Omega|v_\sigma^s|^2\le h^6\sum_{K\in\mathcal{T}}|(\partial_h^{r_1}v^s)_{\dot K}||(\partial_h^{r_1}v^s)_{\ddot K}|.
\tag{60}
$$

Having all the ingredients, we enter the main part of the proof. We start with the standard interpolation inequality and substitute from (57) and (60) to obtain

$$
\|v_h^s\|_{\frac{3}{2}}^{\frac{3}{2}}\le\|v_h^s\|_1^{\frac{1}{2}}\|v_h^s\|_2\le h^{\frac{9}{2}}\left(\sum_{K\in\mathcal{T}}|\partial_h^s v^s|_K\sum_{K\in\mathcal{T}}|\partial_h^{r_1}v^s|_{\dot K}\sum_{K\in\mathcal{T}}|\partial_h^{r_2}v^s|_{\ddot K}\right)^{\frac{1}{2}}.
\tag{61}
$$

Using the AG-inequality $ABC\le\frac{1}{3^3}(A+B+C)^3$, (61) becomes

$$
\|v_h^s\|_{\frac{3}{2}}^{\frac{3}{2}}\le 3^{-\frac{3}{2}}h^{\frac{9}{2}}\left(\sum_{K\in\mathcal{T}}|\partial_h^s v^s|_K+\sum_{K\in\mathcal{T}}|\partial_h^{r_1}v^s|_{\dot K}+\sum_{K\in\mathcal{T}}|\partial_h^{r_2}v^s|_{\ddot K}\right)^{\frac{3}{2}}\lesssim h^{\frac{9}{2}}\left(\sum_{K\in\mathcal{T}}\sum_{r=1}^{d}|\partial_h^r v^s|_K\right)^{\frac{3}{2}}.
\tag{62}
$$

Now we set $v_h^s=|w_h^s|^3w_h^s$ and apply Lemma 3.9 to (62), to get

$$
\|w_h^s\|_6^6\le\left(\frac{2}{3}h^d\sum_{r=1}^{d}\sum_{K\in\mathcal{T}}|\widetilde{\partial_h^r w^s}|_K\{|w^s|^3\}_K^{\star r}\right)^{\frac{3}{2}}.
\tag{63}
$$

where $\{v^s\}_K^{\star r}$ is rather unusual interpolation. In particular,

$$
\{v^s\}_K^{\star r}:=\begin{cases}\frac{1}{2}\left(v_{K+\frac{h}{2}\mathbf{e}_s}+v_{K+\frac{h}{2}\mathbf{e}_s-\mathbf{e}_r}\right)+\frac{1}{2}\left(v_{K+\frac{h}{2}\mathbf{e}_s}+v_{K+\frac{h}{2}\mathbf{e}_s+\mathbf{e}_r}\right) & \text{for } r\neq s,\\ v_{K+\frac{h}{2}\mathbf{e}_s}+v_{K-\frac{h}{2}\mathbf{e}_s} & \text{for } r=s.\end{cases}
$$

However, all we care about is its estimate $\sum_{K\in\mathcal{T}}\{|v^s|\}_K^\star\le 2\sum_{\sigma\in\mathcal{E}_{\mathrm{int}}}|v^s|$. With that and Cauchy-Schwarz inequality, (63) remains

$$
\|w_h^s\|_6^6\lesssim\left\|\sum_{r=1}^{d}|\widetilde{\partial_h^r w^s}|\right\|_2^{\frac{3}{2}}\left\||w_h^s|^3\right\|_2^{\frac{3}{2}},
$$

i.e., after summation over all components

$$
\|w_h^s\|_6^{6-\frac{9}{2}}\lesssim\|\widetilde{\nabla_h w_h^s}\|_2^{\frac{3}{2}}.
$$

The proof for $d=2$ follows the same step and is a bit simpler. We have

$$
|v_\sigma^s(x)|^2\le\left(\sum_{K\in\mathcal{T}}(\partial_h^r v^s)_{\dot K}h\chi_K^{r,s}(x)\right)\left(\sum_{K\in\mathcal{T}}(\partial_h^s v^s)_K h\chi_K^{s,s}(x)\right),
\tag{64}
$$

where $r=s\in\{1,2\}, r\neq s$ and $\dot K=K+\frac{h}{2}\mathbf{e}_s-\frac{h}{2}\mathbf{e}_r$. We recall the definition of $\overline{\chi_K^{r,s}}$ (58) and introduce $\overline{\chi}_K^{s,s}$, a dominating function to $\chi_K^{s,s}$, with

$$
\overline{\chi}_K^{s,s}(x)=\begin{cases}1 & \text{if } x\in Q_\sigma:(\sigma-K)\cdot\mathbf{e}_p=0, p\in\{1,\dots,d\}\setminus\{r_i,s\},\\ 0 & \text{otherwise.}\end{cases}
$$

Similarly as before, $\overline{\chi}_K^{r,s}(x)=\overline{\chi}_K^{r,s}(x_s)$ and $\overline{\chi}_K^{s,s}(x)=\overline{\chi}_K^{s,s}(x_r)$. Therefore, the integration of (64) yields

$$
\int_\Omega|v^s(x)|^2\,\mathrm{d}x\le h^4\sum_{K\in\mathcal{T}}(\partial_h^r v^s)_{\dot K}\sum_{K\in\mathcal{T}}(\partial_h^s v^s)_K.
\tag{65}
$$

Then we set $\mathbf{v} = |\mathbf{w}|^{\lambda-1}\mathbf{w}$, with $\mathbf{w} \in X(\mathcal{E}_{\rm int})^2$ and $\lambda > 2$. Substituting into (65) and applying Lemma 3.9 one gets

$$\|w^s\|_{2\lambda}^{\lambda} \lesssim \left(h^2 \sum_{K\in\mathcal{T}} \{|w^s|^{\lambda-1}\|_K^{\star r}(\partial_h^r w^s)_{\dot{K}} \right)^{\frac{1}{2}} \left(h^2 \sum_{K\in\mathcal{T}} \{|w^s|^{\lambda-1}\}\|_K^{\star s}(\partial_h^s w^s)_K \right)^{\frac{1}{2}} \lesssim \||w^s|^{\lambda-1}\|_p \|\partial_h^r w^s\|_{p'}^{\frac{1}{2}} \|\partial_h^s w^s\|_{p'}^{\frac{1}{2}}, \quad (66)$$

where we applied Hölder's inequality in the last step. Now we fix p with $2\lambda = p(\lambda-1)$ (and therefore $p'(\lambda+1) = 2\lambda$) and divide both sides of (66) with the norm of w^s and apply the Young inequality to get

$$\|w^s\|_{2\lambda} \lesssim \sum_{r=1}^{2} \|\partial_h^r w^s\|_{p'}. \quad (67)$$

The final step is the chain of inequalities build on (67) and standard Lebesgue embeddings

$$\|w^s\|_q \lesssim \|w^s\|_{2\lambda} \lesssim \|\widetilde{\nabla_h w^s}\|_{\frac{2\lambda}{\lambda+1}} \lesssim \|\widetilde{\nabla_h w^s}\|_2,$$

as $p' = \frac{2\lambda}{\lambda+1} < 2$ for any admissible λ.

□

Remark 4. *To prove discrete Sobolev inequality we use the cross derivatives of the velocity, which are, in the finite difference scheme, employed in a rather awkward way. It is interesting that in the three-dimensional case, thanks to the interpolation (61), we do not need to use all 3×3 derivatives, but only 3×2, as we could alternatively use the same derivative twice in the inequality in (61).*

Due to the positivity of the density we can deduce from (34) that

$$\|\widetilde{\nabla_h \mathbf{u}_h}\|_{L^2(L^2(\Omega))} \lesssim D. \quad (68)$$

and using Proposition 3.10 we get also that

$$\|\mathbf{u}_h\|_{L^2(L^q(\Omega))} \lesssim D, \qquad \|\bar{\mathbf{u}}_h\|_{L^2(L^q(\Omega))} \lesssim D, \quad (69)$$

with $q \in [1, 6]$ for $d = 3$ and $q \in [1, \infty)$ for $d = 2$.

4 Consistency of the numerical method

One step towards the convergence to a weak solution is the consistency of numerical solutions, i.e. verifying that the numerical solution satisfies the weak formulation of the problem up to a residual term $\mathcal{R}(\varrho_h, \mathbf{u}_h)$ which satisfies

$$\mathcal{R}(\varrho_h, \mathbf{u}_h) \to 0, \quad \text{as } h \to 0.$$

In this section we formulate the results both for $d = 2, d = 3$. The difference in these cases occurs only in the inverse estimates and the discrete Sobolev inequality (Proposition 3.10) and its consequences, mainly the velocity integrability (69).

We want to emphasize that our result on consistency is not the only possibility. Our goal was to enable as large set of admissible values for γ as possible. Stronger assumptions on the integrability properties of test functions is the price to pay.

4.1 Preliminary material for proving consistency

First, we show some useful estimates on projections and artificial diffusion terms in order to shorten the proofs of consistency. First let us recall the estimates (10).

Lemma 4.1. *Let $\phi \in W^{1,p}(\Omega)$. Then*

$$\|\partial_h \Pi^P \phi\|_{L^p(\Omega)} \lesssim \|\nabla\phi\|_{L^p(\Omega)}, \quad \|\partial_h \Pi^P \Pi^D \mathbf{v}\|_{L^p(\Omega)} \lesssim \|\nabla \mathbf{v}\|_{L^p(\Omega)}, \quad (70)$$

$$\|\Pi^P \Pi^D \mathbf{v} - \mathbf{v}\|_{L^p(\Omega)} \lesssim h\|\nabla \mathbf{v}\|_{L^p(\Omega)}, \quad (71)$$

$$\|\Pi^P \nabla_h \Pi^P \Pi^D \mathbf{v} - \nabla \mathbf{v}\|_{L^p(\Omega)} \lesssim h\|\nabla_x^2 \mathbf{v}\|_{L^p(\Omega)}. \quad (72)$$

Proof. Estimates (70) are the direct consequences of the mean value theorem, with its double application in the latter case,

$$|(\partial^s \Pi^P \phi)_\sigma| = h^{-1}|\phi(\xi_L) - \phi(\xi_K)| \lesssim |\nabla_h \phi|, \quad \text{with some } \xi_K \in Q_K, \xi_L \in Q_L,$$
$$|(\partial^s \Pi^P \Pi^D \mathbf{v})_\sigma| = h^{-1}|\mathbf{v}(\widetilde{\xi_L}) - \mathbf{v}(\widetilde{\xi_K})| \lesssim |\nabla_h \mathbf{v}|, \quad \text{with some } \widetilde{\xi_K} \in Q_K, \widetilde{\xi_L} \in Q_L.$$

Similarly, to get (71) we can write

$$|(\Pi^P \Pi^D \mathbf{v} - \mathbf{v}| \le |\Pi^P \Pi^D \mathbf{v} - \Pi^D \mathbf{v}| + |\Pi^D \mathbf{v} - \mathbf{v}| \lesssim h|\nabla_x \widehat{\Pi^D \mathbf{v}}| + h|\nabla_x \mathbf{v}| \lesssim h|\nabla_x \mathbf{v}|.$$

To prove (72) we show using Taylor expansion that

$$|(\Pi^P \partial_h^s \Pi^P \Pi^D v^r)(x) - \partial^s v^r(x)| \lesssim h|\nabla_x^2 v^r|, \tag{73}$$

where $x \in K$. Let us denote $L = K + h\mathbf{e}_s, J = K - h\mathbf{e}_s$, then

$$(\Pi^P \partial_h^s \Pi^P \Pi^D v^r)(x) = \frac{1}{2h}\left((\Pi^P \Pi^D v^r)_L - (\Pi^P \Pi^D v^r)_J\right). \tag{74}$$

Expressing the Taylor expansion of v^r at each cell K gives

$$v^r(x) = v^r(x_K) + \nabla_x v^r(x_K)(x - x_K) + \frac{1}{2}(x - x_K)^T \nabla_x^2 v^r(\xi(x))(x - x_K),$$

where x_K is its center. Further, as the affine function with zero mean belong to the kernel of the combined projection $Pi^{PD} := \Pi^P \Pi^D$, we have

$$(\Pi^P \Pi^D v^r)_K = v^r(x_K) + \frac{1}{4h^2}\left(\int_{F_{\sigma,r+}} \nabla_x^2 v^r(\xi(x))(x - x_K)^2 \,\mathrm{d}S_x + \int_{F_{\sigma,r-}} \nabla_x^2 v^r(\xi(x))(x - x_K)^2 \,\mathrm{d}S_x\right). \tag{75}$$

Combining (74) and (75), we can write

$$\left|\frac{1}{2h}\left((\Pi^P \Pi^D v^r)_L - (\Pi^P \Pi^D v^r)_J\right) - \partial^s v^r(x)\right| \lesssim \left|\frac{1}{2h}\left(v^r(x_L) - v^r(x_J)\right) - \partial^s v^r(x)\right| + h|\nabla^2 v^r|. \tag{76}$$

Further we use the Mean Value Theorem to express

$$\partial^s v^r(x) = \partial^s v^r(x_K) + \nabla_x \partial^s v^r(\xi_K)(x - x_K), \tag{77}$$

for $x \in K$. The combination of (74, 76, 77) finally yields (73), which proves (72).

□

We introduce the following lemma that will simplify the treatment of the artificial viscosity term.

Lemma 4.2. *Let ϱ_h be obtained through the scheme (18–19) with $\gamma > 1$. Then it holds that*

$$h^\alpha \|\partial_h^s \varrho_h\|_{L^2(0,T,L^2(\Omega))} \lesssim h^\beta c(D),$$

with $\beta = \frac{\alpha}{2} + \min\{0, d(\frac{1}{4} - \frac{1}{\gamma})\}$ and D is defined by (47).

Proof. First let $\gamma \ge 2$. We use the renormalized equation (22) with $B(z) = z^2$. Thanks to the fact that $\mathcal{P}_K \ge 0$, we obtain

$$h^\alpha \int_0^T \int_\Omega (\partial_h \varrho_h)^2 \le \int_\Omega \varrho_0^2 \,\mathrm{d}x - \int_\Omega \varrho^2(T) \,\mathrm{d}x + \int_0^T \int_\Omega |\varrho_h|^2 |\mathrm{div}_h \mathbf{u}_h| \,\mathrm{d}x \lesssim D^2 + \|\varrho_h\|^2_{L^\infty(0,T;L^4(\Omega))} \|\mathrm{div}_h \mathbf{u}_h\|_{L^2(0,T;L^2(\Omega))}, \tag{78}$$

where we used the Hölder inequality and energy estimate (48). Applying the inverse estimate to the latter term in (78), one gets

$$\|\varrho_h\|^2_{L^4(\Omega)} \lesssim h^{\min\left\{0, 2d\left(\frac{1}{4} - \frac{1}{\gamma}\right)\right\}} \|\varrho_h\|^2_{L^\gamma} \lesssim D^2 h^{\min\{0, \frac{d(\gamma-4)}{2\gamma}\}}. \tag{79}$$

Combining (78)–(79)) together with the energy estimates (48) and (68), one gets

$$h^\alpha \|\partial_h^s \varrho_h\|_{L^2(L^2)} = h^{\frac{\alpha}{2}} \|h^{\frac{\alpha}{2}} \partial_h^s \varrho_h\|_{L^2(0,T,L^2(\Omega))} \le h^{\alpha/2} D^{1/2} + h^{\frac{\alpha}{2} + \min\{0, d(\frac{1}{4} - \frac{1}{\gamma})\}} D^{\frac{3}{2}}.$$

For $\gamma \in (1, 2)$, one just uses one more inverse estimate to get $\|\varrho_h\|_\gamma \lesssim h^{d(\frac{1}{2} - \frac{1}{\gamma})} \|\varrho_h\|_\gamma$, but this term will be dominated by $h^\beta c(D)$ for low values of h, anyway. □

Let us write out explicitly the assumptions on α and γ that ensure $\beta > 0$ in Lemma 4.2.

$$\beta > 0 \quad \text{if we have} \quad d = 2: \begin{cases} \gamma \in (1,4), & \alpha > \frac{4}{\gamma} - 1, \\ \gamma \geq 4, & \alpha > 0, \end{cases} \quad \text{or} \quad d = 3: \begin{cases} \gamma \in (1,4), & \alpha > \frac{6}{\gamma} - \frac{3}{2}, \\ \gamma \geq 4, & \alpha > 0. \end{cases} \tag{80}$$

The two following lemmas find their use in the proof of consistency of the momentum scheme.

Lemma 4.3. *For any $f_h \in X(\mathcal{T})$, $\mathbf{g}_h \in X(\mathcal{E}_{\rm int})^d, \mathbf{v} \in W^{2,q}(\Omega)$ we have*

$$\int_\Omega f_h \mathrm{div}_x \mathbf{v} \,\mathrm{d}x = \int_\Omega f_h \mathrm{div}_h(\Pi_h^D \mathbf{v}) \,\mathrm{d}x, \tag{81}$$

Proof. The proof of both identities is based on the Divergence theorem and decomposition of the domain Ω to cells Q_K, where f_h and $(\nabla_h \mathbf{g})$ are constant. The chain of equalities

$$\begin{aligned} \int_\Omega f_h \mathrm{div}_x \mathbf{v} \,\mathrm{d}x &= \sum_{K\in\mathcal{T}} f_K \int_{Q_K} \mathrm{div}_x \mathbf{v} \,\mathrm{d}x = \sum_{K\in\mathcal{T}} f_K \int_{\partial Q_K} \mathbf{v}\cdot\mathbf{n} \,\mathrm{d}S_x \\ &= h^2 \sum_{K\in\mathcal{T}} f_K \sum_{s=1}^d \Big((\Pi^D \mathbf{v})_{\sigma,s+} - (\Pi^D \mathbf{v})_{\sigma,s-} \Big) = h^d \sum_{K\in\mathcal{T}} f_K (\mathrm{div}_h \Pi_h^D \mathbf{v})_K = \int_\Omega f_h \mathrm{div}_h(\Pi_h^D \mathbf{v}) \,\mathrm{d}x, \end{aligned}$$

recovers (81). □

Next, let us define the extension for $(\partial_h^r g^s)_{K+\frac{h}{2}\mathbf{e}_s \pm \frac{h}{2}\mathbf{e}_r}$ for $r \neq s$ and $\mathbf{g} \in X(\mathcal{E}_{\rm int})^d$ to be piecewise constant in its neighbourhood. In particular we define

$$(q^{r,s})(x) = (q^{r,s})_{K+\frac{h}{2}\mathbf{e}_s \pm \frac{h}{2}\mathbf{e}_r}, \qquad \text{when } x - \frac{h}{2}\mathbf{e}_s \mp \frac{h}{2}\mathbf{e}_r \in Q_K \ \wedge x \in \Omega, \tag{82}$$

with

$$q^{r,s} = \partial_h^r g_h^s \qquad \text{or} \qquad q^{r,s} = \partial_h^r g_h^s \partial_h^r v_h^s, \tag{83}$$

where $\mathbf{g} \in X(\mathcal{E}_{\rm int})^d$ and $\mathbf{v} \in X(\mathcal{E})^d$.

As a consequence of (82) we have also

$$h^d \sum_{K\in\mathcal{T}} \left(\frac{1}{2}(q^{r,s})_{K+\frac{h}{2}\mathbf{e}_s+\frac{h}{2}\mathbf{e}_r} + \frac{1}{2}(q^{r,s})_{K+\frac{h}{2}\mathbf{e}_s-\frac{h}{2}\mathbf{e}_r} \right) = \int_\Omega (q^{r,s}) \,\mathrm{d}x,$$

and thus also

$$h^d \sum_{K\in\mathcal{T}} \sum_{s=1}^d \left((q^{s,s})_K + \sum_{\substack{r=1 \\ r\neq s}}^d \left(\frac{1}{2}(q^{r,s})_{K+\frac{h}{2}\mathbf{e}_s+\frac{h}{2}\mathbf{e}_r} + \frac{1}{2}(q^{r,s})_{K+\frac{h}{2}\mathbf{e}_s-\frac{h}{2}\mathbf{e}_r} \right) \right) = \int_\Omega \sum_{r=1}^d \sum_{s=1}^d q^{r,s} \,\mathrm{d}x, \tag{84}$$

where $q^{r,s}$ satisfies (83). The core of the argument is that all nonzero $q_K^{r,s}$ are covered twice with one-half, beside the border ones, whose intersection with Ω is of the size $h^d/2$.

The extension (82) might be viewed as another mesh, and that is the reason why Gallouet at al. define it at the beginning in [20]. We prefer to state it here at the only place where we use it. Notice that for $r = s$ we have $\partial_h^s g^s \in X(\mathcal{T})$, for which the extension is defined in Section 2.2.3.

Lemma 4.4. *Let $\mathbf{g} \in X(\mathcal{E}_{\rm int})^d$ and $\mathbf{v} \in W_0^{1,1}(\Omega)$. Then it holds that*

$$\begin{aligned} &\sum_{K\in\mathcal{T}} \sum_{s=1}^d \left((\partial_h^s g^s)_K (\partial_h^s \Pi^D \mathbf{v})_K + \frac{1}{2} \sum_{\substack{r=1 \\ r\neq s}}^d \sum_{i=1}^2 (\partial_h^r g^s)_{K+\frac{h}{2}\mathbf{e}_s+(-1)^i\frac{h}{2}\mathbf{e}_r} (\partial_h^r \Pi^D \mathbf{v})_{K+\frac{h}{2}\mathbf{e}_s+(-1)^i\frac{h}{2}\mathbf{e}_r} \right) \\ &\quad = \int_\Omega \sum_{s=1}^d \sum_{r=1}^d \widetilde{\partial_h^r g_h^s \partial_x^r v^s}(x) \,\mathrm{d}x + R = \int_\Omega \widetilde{\nabla_h \mathbf{g}_h} : \nabla_x \mathbf{v} \,\mathrm{d}x + R, \end{aligned} \tag{85}$$

where $|R| \leq h \|\widetilde{\nabla_h \mathbf{g}_h}\|_2 \|\nabla_x^2 \mathbf{v}\|_2$.

Proof. Let $\dot{K} := K + \frac{h}{2}\mathbf{e}_s + (-1)^i \frac{h}{2}\mathbf{e}_r$ for $i \in \{1,2\}$, no matter whether $r \neq s$ or not. If we extend $\mathbf{v}$ with zero outside Ω, we can express

$$(\partial_h^r(\Pi^D \mathbf{v})^s)_{\dot{K}} = \frac{1}{h}\left[\frac{1}{h^{d-1}}\int_{F_{\sigma+}} v^s \,\mathrm{d}S_x - \frac{1}{h^{d-1}}\int_{F_{\sigma-}} v^s \,\mathrm{d}S_x\right], \tag{86}$$

where $\sigma\pm := \dot{K} \pm \frac{h}{2}\mathbf{e}_r$.

Similarly as in the proof of Lemma 4.1, we use Taylor theorem to express

$$v^s(x) = v^s(\sigma) + \nabla_x v^s(\sigma)(x-\sigma) + \frac{1}{2}\nabla_x^2 v^s(x-\sigma), \tag{87}$$

for $x \in F_\sigma$ Substituting (87) into (86) yields

$$(\partial^r(\Pi^D \mathbf{v})^s)_{\dot{K}} \leq \frac{1}{h}(v^s(\sigma+) - v^s(\sigma-)) + h(|\nabla_x^2 v^s(\sigma+)| + |\nabla_x^2 v^s(\sigma-)|),$$

as the affine function with zero mean belongs to the kernel of the projection Π^D. Then we use the mean value theorem twice to get for $x \in Q_{\dot{K}}$ (we apologize for an abuse of notation)

$$\begin{aligned}
&(\partial^r(\Pi^D \mathbf{v})^s)_{\dot{K}} - \partial_x^r v^s(x) \\
&\quad\leq \frac{1}{h}(v^s(\sigma+) - v^s(\sigma-)) - \partial_x^r v^s(x) + h(|\nabla_x^2 v^s(\sigma+)| + |\nabla_x^2 v^s(\sigma-)|) \\
&\quad= \partial_x^r v^s(\xi) - \partial_x^r v^s(x) + h(|\nabla_x^2 v^s(\sigma+)| + |\nabla_x^2 v^s(\sigma-)|) \\
&\quad\leq h\sqrt{2}\nabla_x \partial_x^r v^s(\xi') + + h(|\nabla_x^2 v^s(\sigma+)| + |\nabla_x^2 v^s(\sigma-|).
\end{aligned}$$

Now we have for any $\dot{K}$

$$(\partial_h^r g^s)_{\dot{K}}(\partial^r(\Pi^D \mathbf{v})^s)_{\dot{K}} \leq (\partial_h^r g^s)_{\dot{K}} h^{-d}\int_{Q_{\dot{K}}} \partial_x^r v^s \,\mathrm{d}x + (\partial_h^r g^s)_{\dot{K}} h^{1-d}\int_{Q_{\dot{K}}} |\nabla_x^2 v^s(x)| \,\mathrm{d}x. \tag{88}$$

Finally we apply (84) to (88) and Cauchy-Schwarz inequality to obtain (85). □

4.2 Consistency of the continuity scheme

The weak formulation of the continuity method reads as follows.

Theorem 4.5 (Consistency formulation for the continuity)**.** *Let $\varrho_h, \hat{\mathbf{u}}_h$ be piecewise constant and piecewise affine representations, respectively in space and piecewise constant in time, of the solution to the numerical scheme (18–19), with the following parameters: $\gamma > \frac{2d}{d+2}, \alpha > \max\left\{\frac{d(4-\gamma)}{2\gamma}, 0\right\}$, i.e.*

$$\begin{aligned}
d = 2: &\qquad \gamma > 1, \qquad \alpha > \max\left\{\frac{4}{\gamma} - 1, 0\right\}, \\
d = 3: &\qquad \gamma > \frac{6}{5}, \qquad \alpha > \max\left\{\frac{6}{\gamma} - \frac{3}{2}, 0\right\}.
\end{aligned} \tag{89}$$

Then for any $\phi \in C^2(\Omega)$ it holds that

$$\int_\Omega \partial_h^t \varrho_h^n \phi \,\mathrm{d}x - \int_\Omega \varrho_h^n \hat{\mathbf{u}}_h^n \cdot \nabla_x \phi \,\mathrm{d}x = h^{\theta_1}\langle \mathbf{r}_h, \nabla_x \phi\rangle + h^{\theta_2}\langle \mathbb{Q}_h, \nabla_x^2 \phi\rangle,$$

where $\theta_1, \theta_2 > 0$ and $\|\mathbf{r}_h\|_{L^1(0,T,L^{p'}(\Omega))} \lesssim 1$, $\|\mathbb{Q}_h\|_{L^1(0,T;L^{q'}(\Omega))} \lesssim 1$ for $p' = \frac{p}{p-1}$ and $q' = \frac{q}{q-1}$ satisfying

$$d = 2: \quad \begin{cases} p \geq 2 \\ q > \frac{2\gamma}{3\gamma-2} \\ q \geq 1 \end{cases} \qquad \text{or} \qquad d = 3: \quad \begin{cases} p \geq 2 \\ p > \frac{6\gamma}{5\gamma-6} \\ q > \frac{6\gamma}{7\gamma-6} \\ q \geq 1 \end{cases}.$$

Proof. We multiply (18) with $h^d(\Pi^P \phi)_K$ and sum over $K \in \mathcal{T}$. Then we handle the product term by term as following.

Time derivative. We use (9) to get

$$h^d \sum_{K\in\mathcal{T}} (\partial_h^t \varrho_K)^n (\Pi^P \phi) = \sum_{K\in\mathcal{T}} (\partial_h^t \varrho_K)^n \int_K \phi(x) \,\mathrm{d}x = \int_\Omega (\partial_h^t \varrho_h)^n \phi \,\mathrm{d}x.$$

Convective term. Using the definition of the projection and standard integration by parts we get

$$\begin{aligned}
&h^d \sum_{K\in\mathcal{T}} \mathrm{div}_{\mathrm{Up}}[\varrho^n, \mathbf{u}^n]_K (\Pi^P \phi)_K \\
&= \int_\Omega \mathrm{div}_{\mathrm{Up}}[\varrho_h^n, \mathbf{u}_h^n]\phi \,\mathrm{d}x \\
&= -\sum_{s=1}^d \int_\Omega \mathrm{Up}[\varrho_h^n, \mathbf{u}_h^n] \frac{\phi(. + \frac{h}{2}\mathbf{e}_s) - \phi(. - \frac{h}{2}\mathbf{e}_s)}{h} \,\mathrm{d}x \\
&= -\sum_{s=1}^d \int_\Omega \{\varrho_h^n\} u_h^{s,n} \partial_h^s \phi \,\mathrm{d}x + \sum_{s=1}^d \int_\Omega \frac{h}{2} |u^{s,n}| (\partial_h^s \varrho_h^n) \partial_h^s \phi \,\mathrm{d}x =: I_1 + R_1,
\end{aligned}$$

where the equality on the last row follows from the application of Lemma 2.5. Further

$$\begin{aligned}
I_1 &= -\int_\Omega \{\varrho_h^n\} \mathbf{u}_h^n \cdot \nabla_h \phi \,\mathrm{d}x = -\int_\Omega \varrho_h^n \mathbf{u}_h^n \cdot \nabla_h \phi \,\mathrm{d}x - \int_\Omega \left(\frac{\varrho_h^n(x + \frac{h}{2}\mathbf{e}_s) - \varrho_h^n(x)}{2} - \frac{\varrho_h^n(x) - \varrho_h^n(x - \frac{h}{2}\mathbf{e}_s)}{2} \right) \mathbf{u}_h^n \cdot \nabla_h \phi \,\mathrm{d}x \\
&=: I_2 + R_2.
\end{aligned}$$

Then, using standard integration by parts together with $\mathbf{u}_h|_{\partial\Omega} = 0$, the identities

$$\partial_h^s v_h^s|_K = \partial_h^s \hat{v}_h^s|_K \equiv \partial^s \hat{v}_h^s|_K,$$

for any $\mathbf{v} = (v^1, v^2, v^3) \in X(\mathcal{E}_{\mathrm{int}})^d$ and $\mathrm{div}_x \hat{\mathbf{u}}$ being constant on each cell, we get

$$\begin{aligned}
I_2 &= \int_\Omega \mathrm{div}_h(\varrho_h^n \mathbf{u}_h^n)\phi \,\mathrm{d}x = \sum_{K\in\mathcal{T}} \varrho_K^n \int_K \mathrm{div}_h \mathbf{u}_h^n \phi = \sum_{K\in\mathcal{T}} \varrho_K^n \int_{Q_K} \phi \mathrm{div}_x \hat{\mathbf{u}}_h^n \,\mathrm{d}x \\
&= \sum_{K\in\mathcal{T}} \int_{Q_K} \phi \mathrm{div}_x(\varrho_h^n \hat{\mathbf{u}}_h^n) = -\int_\Omega \varrho_h^n \hat{\mathbf{u}}_h^n \cdot \nabla_x \phi \,\mathrm{d}x.
\end{aligned}$$

We need to show that the residual terms R_1, R_2 contribute to $\mathbf{r}_h, \mathbb{Q}_h$. To see that, we perform summation by parts to R_1, R_2 to obtain

$$|R_1| + |R_2| \lesssim h \left| \int_\Omega \partial_h(\mathbf{u}_h^n \nabla_x \phi) \varrho_h^n \,\mathrm{d}x \right| \leq h \int_\Omega |\nabla_h \mathbf{u}_h^n||\nabla_x \phi||\varrho_h^n| \,\mathrm{d}x + h \int_\Omega |\mathbf{u}_h^n||\partial_h \nabla_x \phi||\varrho_h^n| \,\mathrm{d}x =: R_1' + R_2'.$$

Using Hölder inequality with exponents p_1, p_2, p, where $\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p} = 1$, and using inverse estimates we can estimate

$$|R_1'| = h \int_\Omega |\nabla_h \mathbf{u}_h^n||\nabla_x \phi||\varrho_h^n| \,\mathrm{d}x \lesssim h \|\nabla_h \mathbf{u}_h^n\|_{p_1} \|\varrho_h^n\|_{p_2} \|\nabla_x \phi\|_p, \lesssim h^{\theta_1} \|\nabla_h \mathbf{u}_h^n\|_2 \|\varrho_h^n\|_\gamma \|\nabla_x \phi\|_p, \tag{90}$$

where $\theta_1 > 0$ as long as $p > \frac{2d\gamma}{\gamma(2+d)-2d}$, which implies the restriction on γ such that $\gamma > \frac{2d}{2+d}$, see also Remark 5.

Similarly we deduce

$$|R_2'| \lesssim h^\theta \|\mathbf{u}_h^n\|_{q_1} \|\varrho_h^n\|_\gamma \|\nabla_x^2 \phi\|_q, \tag{91}$$

where $\theta > 0$ if $q \geq 1$ and $q > \frac{dq_1\gamma}{(q_1+dq_1-d)\gamma - dq_1}, \gamma > \frac{dq_1}{q_1+dq_1-d}, q_1 \geq 1$. More specifically, the lower bounds read

$$d = 3: \ q > \frac{6\gamma}{7\gamma - 6} \text{ with } q_1 = 6, \quad \text{or} \quad d = 2: \ q = q(q_1) > \frac{2\gamma}{3\gamma - 2} \text{ with } q_1 \text{ arbitrarily large.}$$

We recall also the basic constraint $\gamma \geq 1$ which is crucial for stability of the method.

Then, summing over time one gets

$$\Delta t \sum_{n=1}^{N_t} (|R_1| + |R_2|) \lesssim h^{\theta_1} c(D) \|\nabla_x \phi\|_p + h^\theta c(D) \|\nabla_x^2 \phi\|_q,$$

after using the energy estimates (48), (69) and (68).

Artificial viscosity term. We perform integration by parts (14) to get

$$h^{d+\alpha}\sum_{K\in\mathcal{T}}(\Delta_h\varrho^n)_K(\Pi^P\phi)_K = h^{d+\alpha}\sum_{\sigma\in\mathcal{E}_{\rm int}}(\nabla_h\varrho^n)_\sigma(\partial_h^s\Pi^P\phi)_\sigma,$$

which can be further estimated using Hölder inequality to obtain

$$h^{d+\alpha}\sum_{K\in\mathcal{T}}(\Delta_h\varrho^n)_K(\Pi^P\phi)_K \le h^\alpha\left(\int_\Omega(\partial_h\varrho_h^n)^2\,\mathrm{d}x\right)^{\frac12}\left(\int_\Omega(\partial_h\Pi^P\phi)^2\,\mathrm{d}x\right)^{\frac12} \lesssim h^\alpha\|\partial_h\varrho_h^n\|_2\|\nabla\phi\|_2, \tag{92}$$

where we used Lemma 4.1 in the last inequality. Then the summation over time and Lemma 4.2 supply the estimate $h^\beta c(D)$ as well as the lower bound on α, see (80). Moreover, $p\ge 2$ is required.

The existence of $\mathbf{r}_h, \mathbb{Q}_h$ with properties stated in the Theorem is a consequence of appropriate boundedness of terms on the right-hand sides of (90), (91), (92), the Riesz representation theorem and $\theta_2 = \min\{\theta,\beta\}$. □

Remark 5. *In the above computation, we can formally apply the inverse estimate to smooth functions as well. For instance in (90), since* $\frac{1}{p_1}+\frac{1}{p_2}+\frac{1}{p}=1$, *we have*

$$0<\theta_1 = 1+d\left(\frac{1}{p_1}-\frac12\right)+d\left(\frac{1}{p_2}-\frac{1}{\gamma}\right) = 1+d\left(1-\frac12-\frac{1}{\gamma}-\frac1p\right) = d\left(\frac{\gamma(2+d)-2d}{2d\gamma}-\frac1p\right),$$

which indicates

$$p>\frac{2d\gamma}{\gamma(2+d)-2d},\quad \gamma>\frac{2d}{2+d}.$$

4.3 Consistency of the momentum scheme

Theorem 4.6 (Consistency formulation for the momentum)**.** *Let* $(\varrho_h^n,\mathbf{u}_h^n)$ *be piecewise constant representations of the solution to numerical scheme (18–19) with* $\Delta t\approx h$ *and the following parameters*

$$\gamma>\frac{d}{2}, \alpha>\max\left\{\frac{d(4-\gamma)}{2\gamma},0\right\}. \tag{93}$$

Then for any $\mathbf{v}\in C^2(\Omega)^3$, *it holds that*

$$\begin{aligned}&\int_\Omega\partial_h^t(\varrho_h\bar{\mathbf{u}}_h)^n\cdot\mathbf{v}\,\mathrm{d}x-\int_\Omega\varrho_h^n\bar{\mathbf{u}}_h^n\otimes\bar{\mathbf{u}}_h^n:\nabla_x\mathbf{v}\,\mathrm{d}x-\int_\Omega p(\varrho_h^n)\mathrm{div}_x\mathbf{v}\,\mathrm{d}x+\mu\int_\Omega(\nabla_h\mathbf{u}_h^n):\nabla_x\mathbf{v}\,\mathrm{d}x\\ &\quad=h^{\theta_1}\langle\mathbf{r_h},\nabla_x\mathbf{v}\rangle+h^{\theta_2}\langle\mathbb{Q}_h,\nabla_x^2\mathbf{v}\rangle,\end{aligned} \tag{94}$$

with $\|\mathbf{r_h}\|_{L^1(0,T;L^{p'}(\Omega)}\lesssim 1$ *and* $\|\mathbb{Q}_h\|_{L^1(0,T;L^{q'}(\Omega)}\lesssim 1$, *where* $p'=\frac{p}{p-1}$ *and* $q'=\frac{q}{q-1}$ *which satisfy:*

$$d=2:\quad\begin{cases}p\ge 3,\\ p>\frac{2\gamma}{\gamma-1},\\ q>\frac{2\gamma}{\gamma-1},\end{cases}\qquad or\qquad d=3:\quad\begin{cases}p>\frac{6\gamma}{2\gamma-3},\\ q>\frac{6\gamma}{4\gamma-3}.\end{cases} \tag{95}$$

Proof. We multiply momentum scheme (19) by $h^d\Pi^D\mathbf{v}$ and handle term by term. We would like to point out, that the values of exponents θ_i may vary throughout the proof. To find the proper values of θ_i for (94) should be obtained as the minima of $\theta_i, i=1,2$ throughout their occurrences in the proof.

Time difference term. Using the transition between grids (6) one gets

$$h^d\sum_{\sigma\in\mathcal{E}_{\rm int}}\partial_h^t\{\varrho\bar{\mathbf{u}}\}_\sigma^n\cdot\Pi^D\mathbf{v} = h^d\partial_h^t\left(\sum_{K\in\mathcal{T}}(\varrho\bar{\mathbf{u}})_K\cdot(\Pi^P\Pi^D\mathbf{v})_K\right)^n = \int_\Omega\partial_h^t(\varrho_h\bar{\mathbf{u}}_h)^n\cdot v + R_1+R_2,$$

where

$$\begin{aligned}R_1 &= h^d\sum_{K\in\mathcal{T}}\sqrt{\varrho_K^{n-1}}\sqrt{\varrho_K^{n-1}}\frac{\bar{\mathbf{u}}_K^n-\bar{\mathbf{u}}_K^{n-1}}{\Delta t}\int_{Q_K}(\Pi^P\Pi^D\mathbf{v}-\mathbf{v})\,\mathrm{d}x\\ &\le\|\varrho_h^{n-1}\|_\gamma^{\frac12}h\|\nabla_x\mathbf{v}\|_{\frac{2\gamma}{\gamma-1}}\left((\Delta t)\int_\Omega\varrho_h^{n-1}\left(\frac{\bar{\mathbf{u}}_h^n-\bar{\mathbf{u}}^{n-1}}{\Delta t}\right)^2\mathrm{d}x\right)^{\frac12}(\Delta t)^{-\frac12}\\ &=: h^{\theta_1}\|\varrho_h^{n-1}\|_\gamma^{\frac12}\|\nabla_x\mathbf{v}\|_{\frac{2\gamma}{\gamma-1}}\mathbf{U}_h^n,\quad \theta_1=\frac12.\end{aligned}$$

By virtue of (34) we have $\Delta t\sum_{n=1}^{N_t}(\mathbf{U}_h^n)^2\lesssim c(D)$, which implies, together with (48), that $\Delta t\sum_{n=1}^{N_t}|R_1|\lesssim h^{\frac{1}{2}}\|\nabla_x\mathbf{v}\|_{\frac{2\gamma}{\gamma-1}}c(D)$. The other residual term reads

$$R_2=h^d\sum_{K\in\mathcal{T}}\bar{\mathbf{u}}_K^n\frac{\varrho_K^n-\varrho_K^{n-1}}{\Delta t}(\Pi^P\Pi^D\mathbf{v}-\mathbf{v}). \tag{96}$$

From energy estimates (34) we have

$$\Delta t^2h^d\sum_{n=1}^{N_t}\sum_{K\in\mathcal{T}}(\varrho_K^\eta)^{\gamma-2}\left(\frac{\varrho_K^n-\varrho_K^{n-1}}{\Delta t}\right)^2\lesssim c(D). \tag{97}$$

Using the properties of Legendre remainder points of strictly convex functions, formulated e.g. in [7, Lemma 2.1], (97) implies also

$$(\Delta t)^\gamma h^d\sum_{n=1}^{N_t}\sum_{K\in\mathcal{T}}\left(\frac{\varrho_K^n-\varrho_K^{n-1}}{\Delta t}\right)^\gamma\lesssim c(D,\gamma). \tag{98}$$

Thus, applying Hölder inequality and estimate (71) to (96), one gets

$$|R_2|\le\|\mathbf{u}^n\|_6h\|\nabla_x\mathbf{v}\|_{\frac{6\gamma}{5\gamma-6}}(\Delta t)^{-\frac{\gamma-1}{\gamma}}\left((\Delta t)^{\gamma-1}h^d\sum_{K\in\mathcal{T}}\left(\frac{\varrho_K^n-\varrho_K^{n-1}}{\Delta t}\right)^\gamma\right)^{\frac{1}{\gamma}}\le h^{\frac{1}{\gamma}}\|\mathbf{u}_h^n\|_6\|\nabla_x\mathbf{v}\|_{\frac{6\gamma}{5\gamma-6}}H_h^n,$$

where (98) yields $\Delta t\sum_{n=1}^{N_t}(H_h^n)^\gamma\lesssim c(D)$. This together with (69) implies

$$\Delta t\sum_{n=1}^{N_t}|R_2|\lesssim h^{\frac{1}{\gamma}}c(D,\gamma)\|\nabla_x\mathbf{v}\|_{\frac{6\gamma}{5\gamma-6}},\quad\text{for } d=3,$$

For $d=2$ we can estimate analogously

$$|R_2|\le h^{\frac{1}{\gamma}}\|\mathbf{u}^n\|_{p_1}\|\nabla_x\mathbf{v}\|_pH_h^n,\qquad\text{with }\frac{p_1\gamma}{(p_1-1)\gamma-p_1},$$

and thus we get a lower bound $p=p(p_1)>\frac{\gamma}{\gamma-1}$, as p_1 can be arbitrarily large.

In both choices of d we have $\theta_2=\frac{1}{\gamma}$. It is possible, but not effective to lower the integrability exponent of $\nabla_x\mathbf{v}$ by inverse estimates, since this constraint on integrability is not active.

Notice that we used the relation $\Delta t\approx h$ in this part of the proof.

Convective term. We use the transition between grids, summation by parts (14) and Lemma 2.5 to obtain

$$\begin{aligned}
&\Delta th^d\sum_{n=1}^{N_t}\sum_{\sigma\in\mathcal{E}_{\rm int}}\{\mathrm{div}_{\rm Up}[\varrho^n\bar{\mathbf{u}}^n,\mathbf{u}^n]\}_\sigma\cdot(\Pi^D\mathbf{v})_\sigma=-\Delta th^d\sum_{n=1}^{N_t}\sum_{\sigma\in\mathcal{E}_{\rm int}}\mathrm{Up}[\varrho^n\bar{\mathbf{u}}^n,\mathbf{u}^n]\cdot(\partial_h^s\Pi^P\Pi^D\mathbf{v})_\sigma\\
&=-\Delta th^d\sum_{n=1}^{N_t}\sum_{\sigma\in\mathcal{E}_{\rm int}}u_\sigma^{s,n}\{\varrho_h^n\bar{\mathbf{u}}_h^n\}_\sigma\cdot(\partial_h^s\Pi^P\Pi^D\mathbf{v})_\sigma+h^{d+1}\sum_{\sigma\in\mathcal{E}_{\rm int}}|u_\sigma^{s,n}|\partial_h^s(\varrho^n\bar{\mathbf{u}}^n)_\sigma\cdot(\partial_h^s\Pi^P\Pi^D\mathbf{v})_\sigma\\
&=-h^d\Delta t\sum_{n=1}^{N_t}\sum_{K\in\mathcal{T}}\varrho^n\bar{\mathbf{u}}^n\otimes\bar{\mathbf{u}}^n:\{\nabla_h\Pi^P\Pi^D\mathbf{v}\}_K+R_3=-\Delta t\sum_{n=1}^{N_t}\sum_{K\in\mathcal{T}}\int_{Q_K}\varrho_K^n\bar{\mathbf{u}}_K^n\otimes\bar{\mathbf{u}}_K^n:\nabla_x\mathbf{v}+R_3+R_4.
\end{aligned}$$

We need to estimate the residual terms. Before starting that, we perform summation by parts (14) to R_3 and we split the discrete derivative of the product,

$$\begin{aligned}
R_3=&-h^{d+1}\sum_{K\in\mathcal{T}}(\varrho^n\bar{\mathbf{u}}^n)_K\cdot\sum_{s=1}^d(\partial_h^s(|u^{s,n}|\partial_h^s\Pi^P\Pi^D\mathbf{v}))_K\\
=&-h^{d+1}\sum_{K\in\mathcal{T}}(\varrho^n\bar{\mathbf{u}}^n)_K\cdot\sum_{s=1}^d(\partial_h^su^{s,n})_K(\partial_h^s\Pi^P\Pi^D\mathbf{v})_{\sigma,s-}-h^{d+1}\sum_{K\in\mathcal{T}}(\varrho^n\bar{\mathbf{u}}^n)_K\cdot\sum_{s=1}^d(\partial_h^s\partial_h^s\Pi^P\Pi^D\mathbf{v})_Ku_{\sigma,s+}^{s,n}\\
=:&R_{3,1}+R_{3,2}.
\end{aligned}$$

Then we use Hölder inequality using $p,p_1,p_2:\frac{1}{p}+\frac{1}{p_1}+\frac{1}{p_2}=1$, the inequality

$$|\partial_h^s|g||\le|\partial_h^sg|,$$

relation (70) and inverse estimate (Lemma 2.3) twice to get

$$|R_{3,1}| \le h^{d+1} \sum_{K\in\mathcal{T}} |\varrho_K^n \bar{\mathbf{u}}_K^n| \sum_{s=1}^{d} |(\partial_h^s u^{s,n})_K| |(\partial_h^s \Pi^P \Pi^D \mathbf{v})_{\sigma,s-}| \le h \|\nabla_h \mathbf{u}\|_{p_1} \|\varrho_h^n \bar{\mathbf{u}}_h^n\|_{p_2} \|\partial_h \Pi^P \Pi^D \mathbf{v}\|_p$$
$$\lesssim h^{\theta_1} \|\nabla_h \mathbf{u}_h^n\|_2 \|\varrho_h^n \bar{\mathbf{u}}_h^n\|_{\frac{2\gamma}{\gamma+1}} \|\nabla_x \mathbf{v}\|_p,$$

where the exponent θ_1 remains positive as long as $p > \frac{2d\gamma}{2\gamma-d}, \gamma > \frac{d}{2}$, i.e.,

$$d=2: \ p > \frac{2\gamma}{\gamma-1}, \gamma > 1, \quad \text{or} \quad d=3: \ p > \frac{6\gamma}{2\gamma-3}, \gamma > \frac{3}{2}.$$

Similarly, for $R_{3,2}$ we use the same tools and Mean Value Theorem to obtain

$$|R_{3,2}| \le h^{d+1} \sum_{K\in\mathcal{T}} |\varrho_K^n \bar{\mathbf{u}}_K^n| \sum_{s=1}^{d} |u_{\sigma,s+}^{s,n}| |(\partial_h^s \partial_h^s \Pi^P \Pi^D \mathbf{v})_K| \lesssim h \|\mathbf{u}_h^n\|_{q_1} \|\varrho_h^n \bar{\mathbf{u}}_h^n\|_{q_2} \|\partial_h |\nabla_x \mathbf{v}|\|_q$$
$$\lesssim h^{\theta_2} \|\mathbf{u}_h^n\|_{q_1} \|\varrho_h^n \bar{\mathbf{u}}_h^n\|_{\frac{2\gamma}{\gamma+1}} \|\nabla_x^2 \mathbf{v}\|_q,$$

where θ_2 is positive as long as

$$d=2: \ q = q(q_1) > \frac{2\gamma}{2\gamma-1}, \quad \text{or} \quad d=3: \ q > \frac{6\gamma}{4\gamma-3},$$

since q_1 can be arbitrarily large for $d=2$ and $q_1 = 6$ for $d=3$.

Applying summation over time, uniform estimates (49, 69) and the assumptions on test function $\mathbf{v}$ one gets that $\Delta t \sum_{n=1}^{N_t} |R_{3,1}| + |R_{3,2}| = c(D) \left(h^{\theta_1} \|\nabla_x \mathbf{v}\|_p + h^{\theta_2} \|\nabla_x^2 \mathbf{v}\|_q \right)$.

Pressure term. By virtue of summation by parts (14) and Lemma 4.3 we write the following chain of equalities:

$$h^d \sum_{\sigma\in\mathcal{E}_{\rm int}} (\partial_h^s p(\varrho^n))_\sigma \mathbf{e}_s \cdot (\Pi_h^D \mathbf{v})_\sigma = -h^d \sum_{K\in\mathcal{T}} p(\varrho_K^n) \mathrm{div}_h (\Pi_h^D \mathbf{v})_K = -\int_\Omega p(\varrho_h^n) \mathrm{div}_x \mathbf{v} \, \mathrm{d}x.$$

Viscosity term. We apply summation by parts (Lemma 2.1) and Lemma 4.4 to get

$$h^d \Delta t \sum_{\sigma\in\mathcal{E}_{\rm int}} (\Delta_h \mathbf{u}_h^n)_\sigma \cdot (\Pi^D \mathbf{v})_\sigma^s = h^d \Delta t \int_\Omega \widetilde{\nabla_h \mathbf{u}_h^n} : \nabla_x \mathbf{v} + R_4,$$

where $\Delta t \sum_{n=1}^{N_t} |R_4| \lesssim h \|\widetilde{\nabla_h \mathbf{u}_h}\|_{2,2} \|\nabla_x^2 \mathbf{v}\|_2 \lesssim C(D) h^\theta \|\nabla_x^2 \mathbf{v}\|_q$, with $\theta_2 = 1 + \min\left\{0, d(\frac{1}{2} - \frac{1}{q})\right\}$, thus $\theta_2 > 0$ as long as $q > \frac{2d}{2+d}$, i.e.,

$$d=2: \ q > 1, \quad \text{or} \quad d=3: \ q > \frac{6}{5}.$$

Artificial viscosity term. Finally we treat the last term using summation by parts and transition between grids to get

$$R_5 := h^{d+\alpha} \sum_{\sigma\in\mathcal{E}_{\rm int}} \sum_{r=1}^{d} \{\partial_h^r (\{\bar{\mathbf{u}}^n\} \partial_h^r \varrho^n)\}_\sigma \cdot (\Pi^D \mathbf{v})_\sigma = h^{d+\alpha} \sum_{K\in\mathcal{T}} \sum_{r=1}^{d} \sum_{s=1}^{d} \partial_h^r (\{\bar{u}^{s,n}\} \partial_h^r \varrho^n)_K (\Pi^P \Pi^D v^s)_K$$
$$= h^{d+\alpha} \sum_{\sigma\in\mathcal{E}_{\rm int}} \sum_{r=1}^{d} \{\bar{u}^{r,n}\}_\sigma (\partial_h^s \varrho^n)_\sigma (\partial_h^s \Pi^P \Pi^D v^r)_\sigma,$$

where in the last inequality we interchanged the role of r and s in order to get the standard summation over σ, which is associated with s. Applying the Hölder inequality we get

$$|R_5| \le h^\alpha \|\bar{\mathbf{u}}_h^n\|_6 \|\partial_h^s \varrho_h^n\|_2 \|\nabla_x \mathbf{v}\|_3.$$

Summation over time together with applying the uniform bounds gives

$$\Delta t \sum_{n=1}^{N_t} |R_5| \le h^\alpha \|\mathbf{u}_h\|_{2,6} \|\nabla_x \mathbf{v}\|_3 \|\partial_h^s \varrho_h\|_{2,2} \le h^{\theta_1} c(D) \|\nabla \mathbf{v}\|_3,$$

where $\theta_1 = \beta > 0$ is ensured by the assumptions on the lower bounds of α, see (80). Moreover, $p \ge 3$ is required. □

5 Numerical experiments

In this section we perform two numerical experiments for the scheme in two dimensional space, one with Dirichlet boundary condition and the other is periodic type. Our computational domain is always $\Omega = [0,1]^2$, and some constants are chosen as $\mu = 0.01,\ a = 1.0, \gamma = 1.4.\ \alpha = 1.86$ is chosen to satisfy the restriction (80).

Implementation – fix point iteration for the implicit scheme We solve the implicit nonlinear scheme by fix-point iteration. Given the data $\left(\varrho_h^n, \mathbf{u}_h^n\right)$ at time t^n, let $\left(\varrho_h^{n,0}, \mathbf{u}_h^{n,0}\right) = \left(\varrho_h^n, \mathbf{u}_h^n\right)$, then for $\ell = 0, 1, \cdots$, we linearize the nonlinear system and solve

$$\frac{\varrho_K^{n,\ell+1} - \varrho_K^{n,0}}{\Delta t} + \mathrm{div}_{\mathrm{Up}}[\varrho^{n,\ell}, \mathbf{u}^{n,\ell}]_K - h^\alpha(\Delta_h \varrho^{n,\ell})_K = 0,$$

$$\frac{\{\varrho^{n,\ell+1}\bar{\mathbf{u}}^{n,\ell+1}\}_\sigma - \{\varrho^{n,0}\bar{\mathbf{u}}^{n,0}\}_\sigma}{\Delta t} + \{\mathrm{div}_{\mathrm{Up}}[\varrho^{n,\ell}\bar{\mathbf{u}}^{n,\ell}, \mathbf{u}^{n,\ell}]\}_\sigma + \left(\partial_h^s p(\varrho^{n,\ell})\right)_\sigma e_s$$
$$- \mu(\Delta_h \mathbf{u}^{n,\ell+1})_\sigma - h^\alpha \sum_{r=1}^{d} \{\partial_h^r(\hat{u}^{n,\ell}\partial_h^r \varrho^{n,\ell+1})\}_\sigma = 0,$$

until $\|w_h^{n,\ell} - w_h^{n,\ell+1}\| < \xi\|w_h^{n,\ell}\|$, for $w_h \in \{\varrho_h, \mathbf{u}_h\}$, where ξ is a very small positive parameter, e.g. $\xi = 1.0e-6$. Then the solution at next time step t^{n+1} is obtain by $w_h^{n+1} = w_h^{n,\ell+1}$. As we solve the above iterative steps explicitly, a CFL condition is required for preserving the stability $\Delta t = \mathrm{CFL}\frac{h_{\min}}{|\mathbf{u}|_{\max}}$ with $\mathrm{CFL} = 0.6$.

5.1 Cavity flow

In this experiment we simulate the two dimensional cavity flow supplied with Dirichlet data $\mathbf{u} = (16x^2(1-x)^2, 0)^T$ on the top boundary, and zero otherwise. Starting with the initial values $\mathbf{u} = \mathbf{0}$, and $\varrho = 1$ we show in Figure 2 the evolution of the contour mapping for density and velocity components till time $T = 1$ with mesh parameter $h = 1/128$. In order to present the Experimental Order of Convergence(EOC), we calculate the errors in relative norms for different mesh sizes till $t = 0.1$ while the reference solution is computed at the fine mesh $h = 1/512$. From Table 1 we observe first order convergence.

Table 1: Convergence results of cavity flow

h	$\|e_{\nabla\mathbf{u}}\|_{l^2(L^2)}$	EOC	$\|e_{\mathbf{u}}\|_{l^2(L^2)}$	EOC	$\|e_\varrho\|_{l^1(L^1)}$	EOC	$\|e_\varrho\|_{l^\infty(L^\gamma)}$	EOC
1/32	9.22e-03	–	2.84e-01	–	6.08e-05	–	1.79e-03	–
1/64	4.46e-03	1.05	1.37e-01	1.05	2.79e-05	1.12	9.15e-04	0.97
1/128	2.06e-03	1.11	7.14e-02	0.94	1.45e-05	0.95	4.79e-04	0.93
1/256	9.03e-04	1.19	3.09e-02	1.21	5.98e-06	1.27	2.11e-04	1.18

5.2 Gresho-vortex

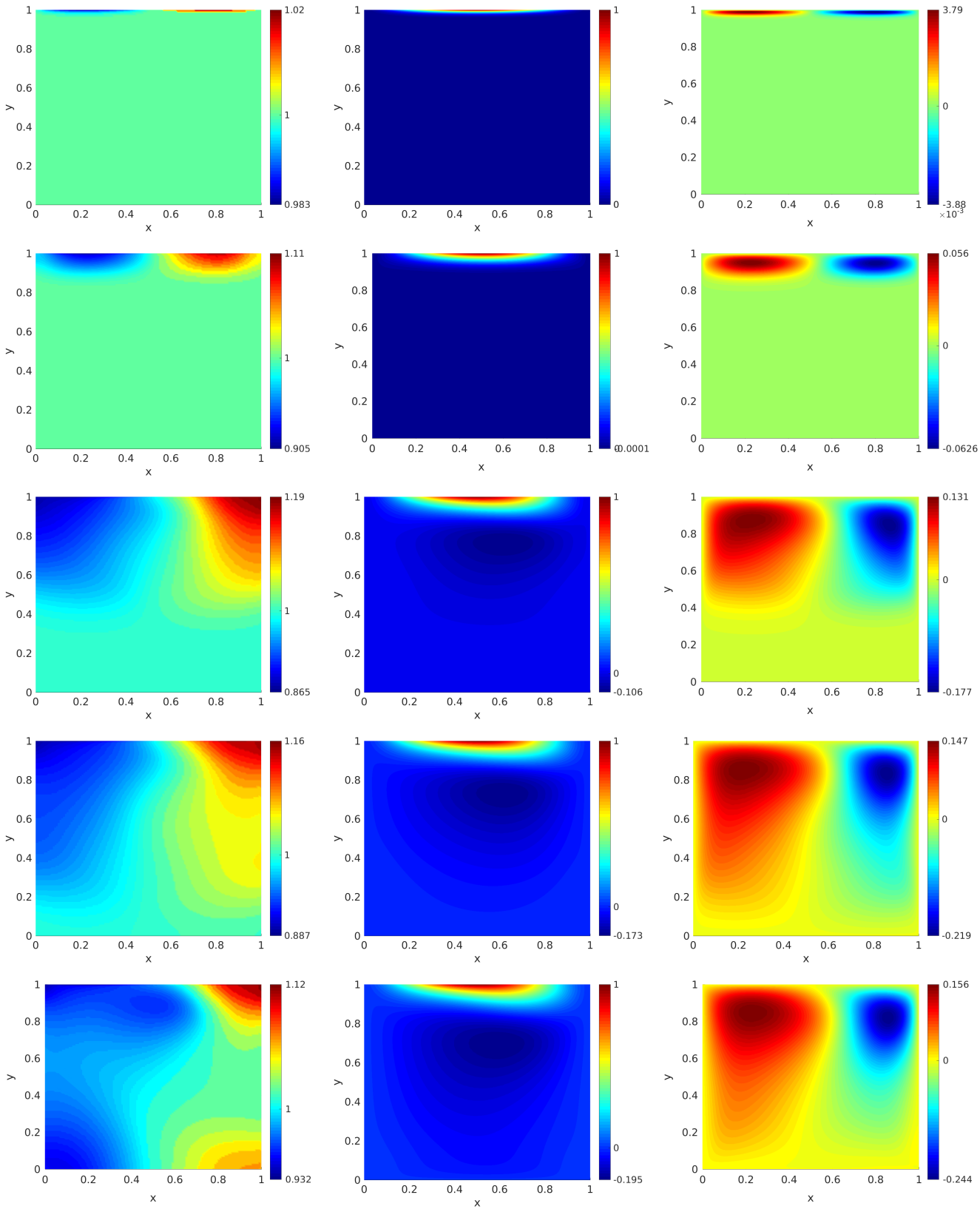
This experiment is an example of rotating vortex, that has been studied in [4, 17, 34] and reference therein for the isentropic flow. Initially, a vortex of radius $R = 0.2$ is prescribed at location $(x_0, y_0) = (0.5, 0.5)$ with the velocity field given by

$$\begin{cases} u_1(0,x,y) = u_r(r) * (y - 0.5)/r, \\ u_2(0,x,y) = u_r(r) * (0.5 - x)/r. \end{cases}$$

where $r = \sqrt{(x-0.5)^2 + (y-0.5)^2}$ and the radial velocity of the vortex u_r is

$$u_r(r) = \sqrt{\gamma}\begin{cases} 2r/R & \text{if } 0 \le r < R/2, \\ 2(1 - r/R) & \text{if } R/2 \le r < R, \\ 0 & \text{if } r \ge R. \end{cases}$$

By setting the periodic boundary condition, we show in Figure 3 the evolution of the flow with mesh parameter $h = 1/128$, from which we see obvious diffusion effects. Analogous to the settings of the previous cavity test, EOC Table 2 indicates similarly first order convergence in the related norms.

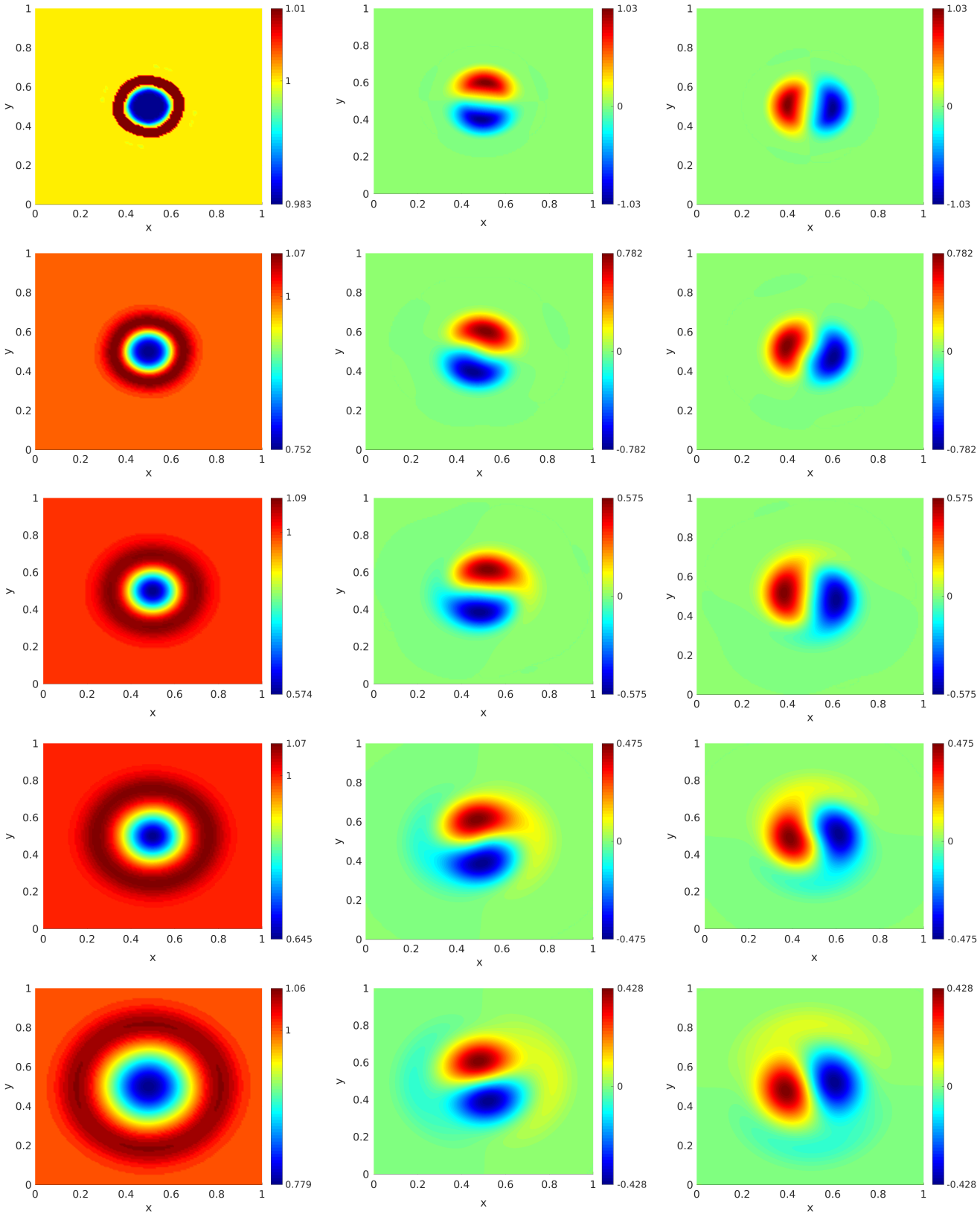


(a) density ϱ (b) velocity component u_1 (c) velocity component u_2

Figure 2: Time evolution of cavity flow, from top to bottom are $t = 0.01, 0.1, 0.5, 0.75, 1$, from left to right are densities and velocity components

Table 2: Convergence results of Gresho vortex test

h	$\|e_{\nabla \mathbf{u}}\|_{l^2(L^2)}$	EOC	$\|e_{\mathbf{u}}\|_{l^2(L^2)}$	EOC	$\|e_{\varrho}\|_{l^1(L^1)}$	EOC	$\|e_{\varrho}\|_{l^\infty(L^\gamma)}$	EOC
1/32	1.10e-02	–	3.74e-01	–	4.40e-04	–	1.35e-02	–
1/64	5.57e-03	0.98	1.88e-01	1.00	2.22e-04	0.99	6.72e-03	1.00
1/128	2.69e-03	1.05	8.71e-02	1.11	1.02e-04	1.12	3.10e-03	1.12
1/256	1.15e-03	1.22	3.37e-02	1.37	3.86e-05	1.40	1.16e-03	1.42



(a) density ϱ (b) velocity component u_1 (c) velocity component u_2

Figure 3: Time evolution of gresho vortex, from top to bottom are $t = 0.01, 0.05, 0.1, 0.15, 0.2$, from left to right are densities and velocity components

A Appendix: Proof of Lemma 2.1

Before proving Lemma 2.1, let us introduce a simplified version in one dimension.

Lemma A.1. *Let the computational domain Ω degenerate to one dimensional interval $I = [a, b]$ and be equally divided into M intervals of the size $h = \frac{b-a}{M}$. Assume that the functions f, ψ, v are discretized at the interval centres, while g, ϕ are located at the division points. Moreover, we assume homogeneous Dirichlet boundary conditions for g, v, e.g.*

$$g_{1/2} = 0, \quad g_{M+1/2} = 0, \quad v_0 = -v_1, \quad v_{M+1} = -v_M.$$

Then the following equalities hold

$$\sum_{i=1}^{M} f_i \frac{g_{i+1/2} - g_{i-1/2}}{h} = -\sum_{i=1}^{M-1} \frac{f_{i+1} - f_i}{h} g_{i+1/2} \tag{99a}$$

$$\sum_{i=1}^{M-1} \frac{\phi_{i+3/2} - 2\phi_{i+1/2} + \phi_{i-1/2}}{h^2} g_{i+1/2} = -\sum_{i=1}^{M} \frac{\phi_{i+1/2} - \phi_{i-1/2}}{h} \frac{g_{i+1/2} - g_{i-1/2}}{h}. \tag{99b}$$

$$-\sum_{i=1}^{M} \frac{\psi_{i+1} - 2\psi_i + \psi_{i-1}}{h^2} v_i = \frac{1}{2}\sum_{i=1}^{M} \frac{\psi_{i+1} - \psi_i}{h} \frac{v_{i+1} - v_i}{h} + \frac{1}{2}\sum_{i=1}^{M} \frac{\psi_i - \psi_{i-1}}{h} \frac{v_i - v_{i-1}}{h} \tag{99c}$$

Proof. By using the boundary conditions $g_{1/2} = g_{M+1/2} = 0$ we directly obtain (99a)

$$\begin{aligned}
&\sum_{i=1}^{M} f_i \frac{g_{i+1/2} - g_{i-1/2}}{h} = \frac{1}{h}\Big(\sum_{i=1}^{M} f_i g_{i+1/2} - \sum_{i=1}^{M} f_i g_{i-1/2}\Big) = \frac{1}{h}\Big(\sum_{i=1}^{M} f_i g_{i+1/2} - \sum_{j=0}^{M-1} f_{j+1} g_{j+1/2}\Big) \\
&= \frac{1}{h}\Big(\sum_{i=1}^{M-1} f_i g_{i+1/2} - \sum_{j=1}^{M-1} f_{j+1} g_{j+1/2} + f_M g_{M+1/2} - f_1 g_{1/2}\Big) = -\sum_{i=1}^{M-1} \frac{f_{i+1} - f_i}{h} g_{i+1/2},
\end{aligned}$$

and (99b)

$$\begin{aligned}
&\sum_{i=1}^{M-1} \frac{\phi_{i+3/2} - 2\phi_{i+1/2} + \phi_{i-1/2}}{h^2} g_{i+1/2} = \frac{1}{h^2}\left(\sum_{j=2}^{M} (\phi_{j+1/2} - \phi_{j-1/2}) g_{j-1/2} - \sum_{i=1}^{M-1} (\phi_{i+1/2} - \phi_{i-1/2}) g_{i+1/2}\right) \\
&= -\sum_{i=1}^{M} \frac{\phi_{i+1/2} - \phi_{i-1/2}}{h} \frac{g_{i+1/2} - g_{i-1/2}}{h} - (\phi_{3/2} - \phi_{1/2}) g_{1/2} - (\phi_{M+1/2} - \phi_{M-1/2}) g_{M+1/2} \\
&= -\sum_{i=1}^{M} \frac{\phi_{i+1/2} - \phi_{i-1/2}}{h} \frac{g_{i+1/2} - g_{i-1/2}}{h}.
\end{aligned}$$

Applying the Dirichlet boundary condition for v we can show (99c)

$$\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{M} \frac{\psi_{i+1} - \psi_i}{h} \frac{v_{i+1} - v_i}{h} + \frac{1}{2}\sum_{i=1}^{M} \frac{\psi_i - \psi_{i-1}}{h} \frac{v_i - v_{i-1}}{h} \\
&= \frac{1}{2h^2}\Big(\sum_{j=2}^{M+1} (\psi_j - \psi_{j-1}) v_j - \sum_{i=1}^{M} (\psi_{i+1} - \psi_i) v_i + \sum_{i=1}^{M} (\psi_i - \psi_{i-1}) v_i - \sum_{j=0}^{M-1} (\psi_{j+1} - \psi_j) v_j\Big) \\
&= \frac{1}{2h^2}\Big(-2\sum_{i=1}^{M} (\psi_{i+1} - 2\psi_i + \psi_{i-1}) v_i + (\psi_{M+1} - \psi_M)(v_{M+1} + v_M) - (\psi_1 - \psi_0)(v_1 + v_0)\Big) \\
&= -\frac{1}{h^2}\sum_{i=1}^{M} (\psi_{i+1} - 2\psi_i + \psi_{i-1}) v_i.
\end{aligned}$$

□

Lemma 2.1 is to show for $f \in X(\mathcal{T}), \mathbf{g} \in X(\mathcal{E}_{\text{int}})^d$ the following equalities.

$$\sum_{K \in \mathcal{T}} (\text{div}_h \mathbf{g})_K f_K = -\sum_{\sigma \in \mathcal{E}_{\text{int}}} g^s_\sigma (\partial^s_h f)_\sigma.$$

$$-\sum_{\sigma \in \mathcal{E}_{\text{int}}} (\Delta_h v^s)_\sigma g^s_\sigma = \sum_{K \in \mathcal{T}} \left((\partial^s_h g^s)_K (\partial^s_h v^s)_K + \frac{1}{2}\sum_{r=1, r\neq s}^{d} \sum_{i=1}^{2} (\partial^r_h g^s)_{K + \frac{h}{2}\mathbf{e}_s + (-1)^i \frac{h}{2}\mathbf{e}_r} (\partial^r_h v^s)_{K + \frac{h}{2}\mathbf{e}_s + (-1)^i \frac{h}{2}\mathbf{e}_r} \right).$$

Proof. It is obvious to obtain the first equality by using (99a) for $s = 1, \ldots, d$ and summing them up. The second equality can be done with same strategy by applying (99b) for the first term on the right hand side and (99c) for the latter term on the right hand side. Summing them up concludes the proof. □

References

[1] G. Ansanay-Alex, F. Babik, J. C. Latché, and D. Vola. An L2-stable approximation of the Navier—Stokes convection operator for low-order non-conforming finite elements. *Int. J. Numer. Meth. Fluids*, 66(5):555–580, 2011.

[2] Y. Coudière, T. Gallouët, and R. Herbin. Discrete Sobolev inequalities and Lp error estimates for finite volume solutions of convection diffusion equations. *ESAIM: M2AN*, 35:767–778, 7 2001.

[3] P. I. Crumpton, J. A. Mackenzie, and K. W. Morton. Cell vertex algorithms for the compressible Navier-Stokes equations. *J. Comput. Phys.*, 109(1):1–15, 1993.

[4] P. Degond and M. Tang. All speed scheme for the low mach number limit of the isentropic Euler equations. *Commun. Comput. Phys.*, 10:1–31, 7 2011.

[5] R. J. DiPerna and P. L. Lions. Ordinary differential equations, transport theory and Sobolev spaces. *Inventiones mathematicae*, 98(3):511–547, 1989.

[6] V. Dolejší and M. Feistauer. *Discontinuous Galerkin method*, volume 48 of *Springer Series in Computational Mathematics*. Springer, Cham, 2015. Analysis and applications to compressible flow.

[7] P. Drábek and R. Hošek. Properties of solution diagrams for bistable equations. *Electron. J. Differ. Equ.*, 2015(156):1–19, 2015.

[8] L. C. Evans. *Partial differential equations.* Providence, RI: American Mathematical Society, 1998.

[9] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In *Handbook of numerical analysis. Vol. 7: Solution of equations in $\mathbb{R}^n$ (Part 3). Techniques of scientific computing (Part 3)*, pages 713–1020. Amsterdam: North-Holland/ Elsevier, 2000.

[10] E. Feireisl. *Dynamics of viscous compressible fluids.* Oxford: Oxford University Press, 2004.

[11] E. Feireisl, P. Gwiazda, A. Świerczewska-Gwiazda, and E. Wiedemann. Dissipative measure-valued solutions to the compressible Navier–Stokes system. *Calculus of Variations and Partial Differential Equations*, 55(6):141, 2016.

[12] E. Feireisl, R. Hošek, D. Maltese, and A. Novotný. Error estimates for a numerical method for the compressible Navier–Stokes system on sufficiently smooth domains. *ESAIM: M2AN*, 51(1):279–319, 2017.

[13] E. Feireisl, R. Hošek, and M. Michálek. A convergent numerical method for the full Navier–Stokes–Fourier system in smooth physical domains. *SIAM J. Numer. Anal.*, 54(5):3062–3082, 2016.

[14] E. Feireisl, T. Karper, and M. Michálek. Convergence of a numerical method for the compressible Navier–Stokes system on general domains. *Numer. Math.*, 134(4):667–704, 2016.

[15] E. Feireisl, T. Karper, and A. Novotný. A convergent numerical method for the Navier–Stokes–Fourier system. *IMA J. Numer. Anal.*, 36(4):1477, 2015.

[16] E. Feireisl and M. Lukáčová-Medvid'ová. Convergence of a mixed finite element finite volume scheme for the isentropic Navier–Stokes system via dissipative measure-valued solutions. *preprint available at ArXiv*, Aug. 2016.

[17] E. Feireisl, M. Lukáčová-Medvid'ová, Š. Nečasová, A. Novotný, and B. She. Error estimate for a numerical approximation to the compressible barotropic Navier-Stokes equations. Preprint, 2016.

[18] E. Feireisl, A. Novotný, and H. Petzeltová. On the existence of globally defined weak solutions to the Navier-Stokes equations. *J. Math. Fluid Mech.*, 3(4):358–392, 2001.

[19] T. Gallouët, L. Gastaldo, R. Herbin, and J.-C. Latché. An unconditionally stable pressure correction scheme for the compressible barotropic Navier-Stokes equations. *ESAIM: M2AN*, 42:303–331, 3 2008.

[20] T. Gallouët, R. Herbin, J.-C. Latché, and D. Maltese. Convergence of the MAC scheme for the compressible stationary Navier-Stokes equations. *ArXiv e-prints*, July 2016.

[21] T. Gallouët, R. Herbin, D. Maltese, and A. Novotný. Implicit MAC scheme for compressible Navier-Stokes equations: unconditional error estimates. *Preprint*, 2016.

[22] T. Gallouët, R. Herbin, D. Maltese, and A. Novotný. Convergence of the marker-and-cell scheme for the semi-stationary compressible Stokes problem. *Mathematics and Computers in Simulation*, 2016. available on line.

[23] T. Gallouët, R. Herbin, D. Maltese, and A. Novotný. Error estimates for a numerical approximation to the compressible barotropic Navier–Stokes equations. *IMA J. Numer. Anal.*, 36(2):543–592, 2016.

[24] F. Grasso and C. Meola. Euler and Navier-Stokes equations for compressible flows: finite-volume methods. In *Handbook of computational fluid mechanics*, pages 159–282. Academic Press, San Diego, CA, 1996.

[25] J. Haack, S. Jin, and J. Liu. An all-speed asymptotic-preserving method for the isentropic Euler and Navier-Stokes equations. *Commun. Comput. Phys.*, 12:955–980, 10 2012.

[26] R. Hošek. Expressing the remainder of Taylor polynomial when the function lacks smoothness. *Preprint*, 2017.

[27] T. K. Karper. A convergent FEM-DG method for the compressible Navier-Stokes equations. *Numer. Math.*, 125(3):441–510, 2013.

[28] T. K. Karper. Convergent finite differences for 1D viscous isentropic flow in Eulerian coordinates. *Discrete Contin. Dyn. Syst., Ser. S*, 7(5):993–1023, 2014.

[29] C. M. Klaij, J. J. W. van der Vegt, and H. van der Ven. Space-time discontinuous Galerkin method for the compressible Navier-Stokes equations. *J. Comput. Phys.*, 217(2):589–611, 2006.

[30] M. Kouhi and E. Oñate. An implicit stabilized finite element method for the compressible Navier-Stokes equations using finite calculus. *Comput. Mech.*, 56(1):113–129, 2015.

[31] M. Kupiainen and B. Sjögreen. A Cartesian embedded boundary method for the compressible Navier-Stokes equations. *J. Sci. Comput.*, 41(1):94–117, 2009.

[32] P.-L. Lions. *Mathematical topics in fluid mechanics. Vol. 2: Compressible models.* Oxford: Clarendon Press, 1998.

[33] B. Liu. The analysis of a finite element method with streamline diffusion for the compressible Navier-Stokes equations. *SIAM J. Numer. Anal.*, 38(1):1–16 (electronic), 2000.

[34] S. Noelle, G. Bispen, K. R. Arun, M. Lukáčová-Medvid'ová, and C.-D. Munz. A weakly asymptotic preserving low Mach number scheme for the Euler equations of gas dynamics. *SIAM J. Sci. Comput.*, 36(6):B989–B1024, 2014.

[35] J. S. Park and C. Kim. Higher-order multi-dimensional limiting strategy for discontinuous Galerkin methods in compressible inviscid and viscous flows. *Comput. & Fluids*, 96:377–396, 2014.

[36] F. Renac, S. Gérald, C. Marmignon, and F. Coquel. Fast time implicit-explicit discontinuous Galerkin method for the compressible Navier-Stokes equations. *J. Comput. Phys.*, 251:272–291, 2013.

[37] A. Valli. An existence theorem for compressible viscous fluids. *Ann. Mat. Pura Appl. (4)*, 130:197–213, 1982.

[38] K. Xu, C. Kim, L. Martinelli, and A. Jameson. BGK-based schemes for the simulation of compressible flow. *Int. J. Comput. Fluid Dyn.*, 7(3):213–235, 1996.