Weierstraß-Institut
für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

A gradient-robust well-balanced scheme for the compressible isothermal Stokes problem

Mine Akbas [1], Thierry Gallouët [2], Almut Gaßmann [3], Alexander Linke [4], Christian Merdon[4]

submitted: 7th November 2019

[1] Düzce University
81620 Düzce
Turkey
E-Mail: mineakbas@duzce.edu.tr

[2] Aix-Marseille University
Site du Pharo - 58 bd Charles Livon
13284 Marseille Cedex 07
France
E-Mail: thierry.gallouet@univ-amu.fr

[3] Leibniz-Institut für Atmosphärenphysik e.V.
an der Universität Rostock
Schlossstraße 6
18225 Kühlungsborn
Germany
E-Mail: gassmann@iap-kborn.de

[4] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: alexander.linke@wias-berlin.de
christian.merdon@wias-berlin.de

No. 2641
Berlin 2019

2010 *Mathematics Subject Classification.* 76D07, 65N30, 65N12.

2008 *Physics and Astronomy Classification Scheme.* 47.10.ad, 47.11.Fg.

Key words and phrases. Compressible Stokes equations, finite element method, finite volume method, well-balanced scheme, upwind, convergence.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

A gradient-robust well-balanced scheme for the compressible isothermal Stokes problem

Mine Akbas , Thierry Gallouët , Almut Gaßmann , Alexander Linke , Christian Merdon

ABSTRACT. A novel notion for constructing a well-balanced scheme — a gradient-robust scheme — is introduced and a showcase application for a steady compressible, isothermal Stokes equations is presented. Gradient-robustness means that arbitrary gradient fields in the momentum balance are well-balanced by the discrete pressure gradient — if there is enough mass in the system to compensate the force. The scheme is asymptotic-preserving in the sense that it degenerates for low Mach numbers to a recent inf-sup stable and pressure-robust discretization for the incompressible Stokes equations. The convergence of the coupled FEM-FVM scheme for the nonlinear, isothermal Stokes equations is proved by compactness arguments. Numerical examples illustrate the numerical analysis, and show that the novel approach can lead to a dramatically increased accuracy in nearly-hydrostatic low Mach number flows. Numerical examples also suggest that a straight-forward extension to barotropic situations with nonlinear equations of state is feasible.

1. INTRODUCTION

In recent years, novel concepts and discretization approaches for the incompressible Navier-Stokes equations appeared around the so-called *pressure-robustness* property. Such discretizations allow for a priori error estimates of the discrete velocity that are independent of the pressure and the viscosity parameter that otherwise gives rise to a severe locking phenomenon [18, 25, 31, 2, 32, 31, 1, 19, 29, 21, 5, 29, 33, 42] demonstrated in several benchmark examples [18, 25, 24, 2, 15] even in coupled problems to simulate electrolyte flows [11]. Surprisingly, an astonishingly simple modification that manipulates locally the velocity test functions in the right-hand side (and the material derivative if present) in order to restore the orthogonality between discretely divergence-free functions and gradients renders any classical inf-sup stable finite element method a pressure-robust method [26, 2, 1, 20].

In this contribution we apply this modification to a provably convergent discretization of the compressible Stokes equations inspired by [12, 10]. The proposed modification does not compromise the convergence analysis, but improves the accuracy in nearly hydrostatic low Mach number flows. Thereby, a novel notion for a certain class of well-balanced schemes — *gradient-robust* schemes — for vector-valued partial differential equations like the compressible Euler, the compressible Navier–Stokes or the shallow-water equations is introduced.

The notion *gradient-robust* wants to emphasize that the accuracy of these schemes does not suffer from the appearance of dominant gradient fields in the momentum balance, leading to an accurate, implicitly defined discrete vorticity equation [18]. Indeed, several schemes for several different *vector PDEs* can be classified as *gradient-robust*, e.g., see [6, 28, 17]. In the meteorology community such schemes have been introduced by Cotter and Thuburn and their well-balanced property has been explained by *exact sequences* in the setting of finite element exterior calculus [6]. The proposed explanation for the well-balanced property of these schemes below is complementary, but sets a different focus. Emphasizing the importance of an $\mathbf{L}^2$ orthogonality of certain (discretely) divergence vector fields against (arbitrary) gradient fields for accuracy reason allows to build novel *gradient-robust* schemes.

In order to compare *gradient-robustness* with classical well-balanced schemes [38, 3], we regard the *momentum balance of isothermal hydrostatics*

$$(\rho\mathbf{u})_t + \nabla\cdot(\rho\mathbf{u}\otimes\mathbf{u}) + \nabla p = -\rho\nabla\phi, \tag{1.1}$$

for the compressible and the incompressible Euler equations on a bounded polyhedral Lipschitz domain $\mathcal{D}$. Here, the potential ϕ is assumed to depend on the space variable $\mathbf{x}$ only, i.e., we search for a steady density $\rho(\mathbf{x})$ and pressure solutions $p(\mathbf{x})$ fulfilling (1.1) with $\mathbf{u}\equiv\mathbf{0}$. The goal of the comparison is to understand the

DOI 10.20347/WIAS.PREPRINT.2641 Berlin 2019

necessary properties for a space discretization that discretely preserves incompressible or compressible *hydrostatics*.

In the incompressible case, it holds $\rho = \mathrm{const}$ and therefore one concludes for a hydrostatic balance

$$\nabla p = -\rho \nabla \phi = \nabla \left(-\rho\phi\right),$$

i.e., the hydrostatic pressure is given (up to a constant) by $p = -\rho\phi + \mathrm{const}$. A consistent discretization of incompressible hydrostatics requires to balance *arbitrary gradient fields* $\nabla(-\rho\phi)$ by the discrete pressure gradient, i.e., a consistent, pressure-robust discretization possesses an appropriately defined discrete Helmholtz projector $\mathbb{P}_h$ — an $\mathbf{L}^2$ projector onto discretely divergence-free vector fields — whose kernel contains *arbitrary (!) gradient fields* in $\mathbf{L}^2$, i.e., it holds

$$\mathbb{P}_h(-\rho\nabla\phi) = \mathbf{0}, \tag{1.2}$$

for all $\phi \in \mathbf{H}^1(\mathcal{D})$. Pressure-robust schemes for incompressible flows achieve this goal by exploiting the $\mathbf{L}^2$-orthogonality of vector-valued, *divergence-free* $H(\mathrm{div})$-conforming finite element test functions (with vanishing normal component at the boundary) against arbitrary gradient fields. Thus, pressure-robust schemes can be constructed on general unstructured grids. Note that most classical finite element, finite volume and Discontinuous Galerkin schemes are *not pressure-robust*. The kernel of their discrete Helmholtz projectors contains only a subspace of *discrete pressure gradients* [25, 18].

Assuming for the isothermal ($T = \mathrm{const}$) compressible case an ideal gas law $p = \rho\mathcal{R}T$, the hydrostatic balance is given by

$$\nabla p = -\rho\nabla\phi \qquad \Leftrightarrow \qquad \nabla\rho = \rho\nabla\left(-\frac{\phi}{\mathcal{R}T}\right),$$

which can be explicitely integrated, leading to

$$\rho = \rho_0 \exp\left(\frac{-\phi}{\mathcal{R}T}\right), \qquad p = \rho_0\mathcal{R}T\exp\left(\frac{-\phi}{\mathcal{R}T}\right).$$

Exploiting this *explicit solution*, one confirms the identity

$$\mathcal{R}T\exp\left(\frac{\phi}{\mathcal{R}T}\right)\nabla\left(\exp\left(\frac{-\phi}{\mathcal{R}T}\right)\right) = -\nabla\phi$$

and — as an example — the classical well-balanced scheme [38] is based on a space discretization of the right hand side term in the form

$$-\rho\nabla\phi = \rho_0\mathcal{R}T\,\nabla\left(\exp\left(\frac{-\phi}{\mathcal{R}T}\right)\right). \tag{1.3}$$

The key ideas of a *gradient-robust* well-balanced scheme for compressible flows, which is based on the notion of pressure-robustness for incompressible flows, rely on the following observations:

(a) A hydrostatic balance

$$\nabla p = -\rho\nabla\phi$$

is only possible if $\rho\nabla\phi$ is a gradient-field, which is only a-priori clear in the incompressible case $\rho = \mathrm{const}$; thus it seems to be plausible for variable ρ that a more accurate treatment of gradient forces may increase the overall accuracy of the scheme in a nearly-hydrostatic situation.

(b) Actually, we demonstrate in this contribution that it is possible to construct *gradient-robust schemes* on arbitrary unstructured grids, which allow a well-balanced property of the form

$$\nabla p = -\nabla\psi$$

for arbitrary gradient fields $\nabla\psi \in \mathbf{L}^2(\mathcal{D})$ — if there is enough mass in the system to compensate the gradient force $-\nabla\psi$.

(c) The velocity field of nearly-hydrostatic flows is of *low Mach number* type. Thus, its non-divergence-free part is small, i.e., of order $\mathcal{O}(\frac{1}{\mathrm{Ma}^2})$, and for an accurate treatment in nearly-hydrostatic, low Mach number flows it suffices to achieve that the *divergence-free part* of the velocity field vanishes in the hydrostatic limit case. Exactly this is achieved in this contribution for the barotropic compressible Stokes equations, since it is shown that the divergence-free part of the velocity fulfills the *incompressible Stokes equations* — where (incompressible) *pressure-robustness* can be exploited.

Although, this paper develops an appropriate space discretization for the rather simple compressible Stokes equations, it is a first step to develop and analyze *gradient-robust* schemes for the full compressible Navier–Stokes equations. Nevertheless, even this simple physical problem is highly relevant in atmospheric and oceanic modeling where the correct representation of the hydrostatic balance between pressure gradient and gravity is of vital importance in *stably stratified airflows over a topography*. It has been early recognized that discretization errors especially in terrain-following coordinates — leading to rather structured grids — can become large and deteriorate the accuracy of the numerical solutions. Such errors are especially severe, if a resting fluid is located over steep terrain [41]. Several attempts were made over the years to ameliorate the simulations. Proposed methods are the increase of the order of accuracy [7, 40, 27], the improvement of the lower boundary condition [13], the usage of cut cells or step mountain coordinates instead of the terrain following coordinates [35, 36, 34], the reduction of the steepness of the slopes [30], the damping of error induced noise [39], covariant formulations of the pressure gradient term [23, 22], curl-free pressure gradient formulations [37], and energy conserving schemes (discrete Poisson brackets) that hinder the spurious increase of kinetic energy in the perturbations [14]. These references are just to be thought to reflect the importance of the problem in applications.

The rest of the paper is structured as follows. Section 2 introduces the steady, compressible isothermal Stokes equations, which serve as a model problem. Section 3 explains our discretization, in particular the finite-volume scheme for the continuity equation and the finite element scheme for the momentum equation with the gradient-robust right-hand side modification. Section 4 motivates and discusses the new gradient-robustness property and links it to pressure-robustness in the incompressible setting or the well-balanced known from shallow water equations. Section 5 proves the existence of a discrete solution by standard compactness arguments, while Section 6 proves the convergence of a series of discrete solutions to a weak solution of the continuous system. In Section 7 the theoretical findings are validated by appropriate numerical benchmarks.

2. A model problem: the steady compressible isothermal Stokes equations

The isothermal compressible Stokes problem seeks for $(\mathbf{f},\mathbf{g}) \in \mathbf{L}^2(\Omega) \times \mathbf{L}^\infty(\Omega)$ some velocity field $\mathbf{u}$, pressure p and non-negative density $\varrho \geqslant 0$ with $\int_\Omega \varrho\, dx = M$ such that

$$\begin{aligned} -\nabla\cdot\boldsymbol{\sigma} + \nabla p &= \mathbf{f} + \varrho\mathbf{g}, \\ \operatorname{div}(\varrho\mathbf{u}) &= 0 \\ p &= \varphi(\rho) := c\rho, \end{aligned} \tag{2.1}$$

where friction is modeled as in linear elasticity by

$$\boldsymbol{\sigma} = 2\mu\boldsymbol{\epsilon}(\mathbf{u}) + \lambda(\nabla\cdot\mathbf{u})\boldsymbol{I}, \tag{2.2}$$

with $\boldsymbol{\epsilon}(\mathbf{u}) := \frac{1}{2}(\nabla\mathbf{u} + (\nabla\mathbf{u})^T)$, $\mu \in \mathbb{R}^+$, $\lambda \in \mathbb{R}$ with $\lambda \geqslant \underline{\lambda} > -2\mu$, compare e.g. with [9], and where the equation of state function $\varphi := c\rho$ with $c > 0$ is prescribed in addition to homogeneous Dirichlet velocity boundary conditions to close the system. Note, that the constant c may model (in a dimensionless setting) the squared inverse of the Mach number.

The compressible Stokes problem is thus a nonlinear problem and can be written in the following weak form [8]: search for $(\mathbf{u}, p, \varrho) \in \mathbf{H}_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega)$ with

$$\begin{aligned} a_1(\mathbf{u},\mathbf{v}) + a_2(\mathbf{u},\mathbf{v}) + b(p,\mathbf{v}) &= F(\mathbf{v}) + G(\varrho,\mathbf{v}) && \text{for all } \mathbf{v} \in \mathbf{H}_0^1(\Omega), \\ c(\varrho, u, \phi) &= 0 && \text{for all } \phi \in W^{1,\infty}(\Omega), \end{aligned} \tag{2.3}$$

where the multilinear forms used above read as

$$\begin{aligned} a_1(\mathbf{u},\mathbf{v}) &:= 2\mu \int_\Omega \boldsymbol{\epsilon}(\mathbf{u}) : \boldsymbol{\epsilon}(\mathbf{v})\, dx, & a_2(\mathbf{u},\mathbf{v}) &:= \lambda \int_\Omega \operatorname{div}(\mathbf{u})\operatorname{div}(\mathbf{v}), \\ b(p,\mathbf{u}) &:= -\int_\Omega p \operatorname{div}(\mathbf{u})\, dx, & c(\varrho,\mathbf{u},\phi) &:= \int_\Omega \varrho \mathbf{u} \cdot \nabla\phi\, dx, \\ F(\mathbf{v}) &:= \int_\Omega \mathbf{f} \cdot \mathbf{v}\, dx, & G(\varrho,\mathbf{v}) &:= \int_\Omega \varrho \mathbf{g} \cdot \mathbf{v}\, dx. \end{aligned}$$

Qualitative properties of the velocity solution $\mathbf{u}$ can be investigated by introducing the spaces

$$\begin{aligned} \mathbf{V}^0 &= \{\mathbf{v} \in \mathbf{H}_0^1(\Omega) : \nabla \cdot \mathbf{v} = 0\} \\ \mathbf{V}^\perp &= \{\mathbf{v} \in \mathbf{H}_0^1(\Omega) : (\boldsymbol{\epsilon}(\mathbf{v}), \boldsymbol{\epsilon}(\mathbf{w})) = 0 \quad \text{for all } \mathbf{w} \in \mathbf{V}^0\} \end{aligned} \tag{2.4}$$

and with the help of the orthogonal splitting — in the scalar product $(\boldsymbol{\epsilon}(\bullet), \boldsymbol{\epsilon}(\bullet))$ —

$$\mathbf{u} = \mathbf{u}^0 + \mathbf{u}^\perp$$

with $\mathbf{u}^0 \in \mathbf{V}^0$ and $\mathbf{u}^\perp \in \mathbf{V}^\perp$. Testing the equation with an arbitrary $\mathbf{v}^0 \in \mathbf{V}^0$ one recognizes that it holds

$$2\mu(\boldsymbol{\epsilon}(\mathbf{u}^0), \boldsymbol{\epsilon}(\mathbf{v}^0)) = (\mathbf{f} + \rho\mathbf{g}, \mathbf{v}^0) \tag{2.5}$$

for all $\mathbf{v}^0 \in \mathbf{V}^0$. Thus, for fixed ρ or for $\mathbf{g} = \mathbf{0}$ the *divergence-free part* $\mathbf{u}^0$ of the solution $\mathbf{u}$ fulfills a *linear incompressible Stokes equations*. Moreover, introducing the space

$$\mathbf{L}^2_\sigma = \{\mathbf{v} \in \mathbf{L}^2(\Omega) : (\mathbf{v}, \nabla\phi) = 0 \text{ for all } \phi \in H^1(\Omega)\}, \tag{2.6}$$

one obtains the orthogonal decomposition

$$\mathbf{L}^2(\Omega) = \mathbf{L}^2_\sigma \oplus_{\mathbf{L}^2} \{\nabla\phi : \phi \in H^1(\Omega)\} \tag{2.7}$$

and $\mathbf{L}^2_\sigma$ represents the space of *weakly divergence-free* $\mathbf{L}^2$ vector fields with vanishing normal component at the boundary [18]. Exploiting the Helmholtz–Hodge decomposition (2.7), one can introduce the $\mathbf{L}^2$-orthogonal Helmholtz–Hodge projector $\mathbb{P} : \mathbf{L}^2(\Omega) \to \mathbf{L}^2_\sigma$ of a vector field $\mathbf{f} \in \mathbf{L}^2(\Omega)$ by

$$(\mathbb{P}(\mathbf{f}), \mathbf{w}) = (\mathbf{f}, \mathbf{w}) \qquad \text{for all } \mathbf{w} \in \mathbf{L}^2_\sigma,$$

see [18]. Then, due to the orthogonal decomposition (2.7) one obtains:

Lemma 2.1. For all $\phi \in H^1(\Omega)$ it holds

$$\mathbb{P}(\nabla\phi) = \mathbf{0}.$$

Using the concept of the Helmholtz–Hodge projector, one can refine (2.5) to observe

$$\mu(\boldsymbol{\epsilon}(\mathbf{u}^0), \boldsymbol{\epsilon}(\mathbf{v}^0)) = (\mathbb{P}(\mathbf{f} + \rho\mathbf{g}), \mathbf{v}^0) \tag{2.8}$$

for all $\mathbf{v}^0 \in \mathbf{V}^0$, i.e., the divergence-free part $\mathbf{u}^0$ does not depend on the entire data $\mathbf{f} + \rho\mathbf{g}$, but only on its divergence-free part $\mathbb{P}(\mathbf{f} + \rho\mathbf{g})$.

3. Well-balanced Bernardi–Raugel finite element - finite volume method

The proposed discretization is based on the finite element-finite volume scheme of [12]. Here, the continuity equation is discretized by some finite volume technique that ensures the non-negativity and mass constraints of the piecewise-constant discrete density ϱ_h.

For the velocity the classical $\mathbf{H}^1$-conforming Bernardi–Raugel finite element method is employed — instead of the *nonconforming Crouzeix–Raviart* element used in [12, 10]. This has several advantages: First, the conforming method is cheaper in terms of the number of degrees of freedom. Second, it easily allows for the use of the stress tensor $\boldsymbol{\sigma}$, whereas the Crouzeix–Raviart element does not fulfill a *discrete Korn inequality*. Third, the conforming setting makes some of the compactness arguments easier, in order to prove convergence to a weak solution of this nonlinear problem, without resorting to additional stability terms.

However, the main important difference to the scheme [12] is a modified discretization of the right-hand side $\mathbf{f} + \rho\mathbf{g}$ that delivers more accurate results in nearly hydrostatic situations. The modification is inspired by certain *pressure-robust schemes* for the incompressible Stokes equations, see e.g. [18, 25]. Fundamental is an appropriate discrete equivalent of (2.8), where the discretely divergence-free part $\mathbf{u}_h^0$ of the discrete solution $\mathbf{u}_h$ does only depend on the *continuous* Helmholtz–Hodge projector $\mathbb{P}(\mathbf{f} + \rho_h\mathbf{g})$.

3.1. **Notation.** Consider a shape-regular triangulation $\mathcal{T}$ with nodes $\mathcal{N}$ and faces $\mathcal{F}$. The subset $\mathcal{F}(\Omega)$ denotes the interior faces of the triangulation. The set $P_k(T)$ consists of all scalar-valued polynomials of total degree k on the simplex $T \in \mathcal{T}$. Moreover, the set of piecewise polynomials is denoted by

$$P_k(\mathcal{T}) := \{v_h \in L^2(\Omega) : v_h|_T \in P_k(T) \text{ for all } T \in \mathcal{T}\}.$$

Vector-valued quantities or functions are addressed by bold letters.

3.2. **Finite Element Method and a divergence-free reconstruction operator.** The numerical discretization employs the Bernardi–Raugel finite element spaces

$$\mathbf{V}_h := \big(\mathbf{P}_1(\mathcal{T}) \cap \mathbf{H}_0^1(\Omega)\big) \oplus \mathcal{B}(\mathcal{F}(\Omega)) \quad \text{and} \quad Q_h := P_0(\mathcal{T}) \cap L_0^2(\Omega),$$

where $\mathcal{B}(\mathcal{F}(\Omega))$ denotes the normal-weighted face bubbles, i.e.

$$\mathcal{B}(\mathcal{F}(\Omega)) := \{b_F \mathbf{n}_F : F \in \mathcal{F}(\Omega)\}.$$

For $d = 2$, b_F is the standard quadratic face bubble on the face $F \in \mathcal{F}$. For $d = 3$, the corresponding standard face bubble is cubic. The L^2 projection in the discrete pressure Q_h will be denoted in the following by π_0.

Then, the *discrete divergence* operator $\operatorname{div}_h : \mathbf{V}_h \to Q_h$ of the Bernardi–Raugel element is denoted by

$$\operatorname{div}_h(\mathbf{v}_h) := \pi_0(\operatorname{div} \mathbf{v}_h). \tag{3.1}$$

Note that the Bernardi–Raugel element is *discretely inf-sup stable* on shape-regular meshes [16]. The space of *discretely divergence-free* vector fields will be denoted as

$$\mathbf{V}_h^0 = \{\mathbf{v}_h \in \mathbf{V}_h : \operatorname{div}_h \mathbf{v}_h = 0\}. \tag{3.2}$$

Due to the general theory of mixed finite element spaces [16], the space of discretely divergence-free Bernardi–Raugel functions has optimal approximation properties versus $\mathbf{V}^0$. More precisely, it holds for all $\mathbf{v}^0 \in \mathbf{V}^0$ that

$$\inf_{\mathbf{v}_h^0 \in \mathbf{V}_h^0} \|\nabla(\mathbf{v}^0 - \mathbf{v}_h^0)\| \leqslant (1 + C_F) \inf_{\mathbf{v}_h \in \mathbf{V}_h} \|\nabla(\mathbf{v}^0 - \mathbf{v}_h)\|, \tag{3.3}$$

where C_F denotes the (uniformly bounded) stability constant of the corresponding *Fortin operator* [16] of the Bernardi–Raugel element.

The gradient-robust modification of the Bernardi–Raugel finite element method employs a reconstruction operator Π in the right-hand side functionals, which maps *discretely divergence-free functions* onto *weakly divergence-free* ones in the sense of $\mathbf{L}^2_\sigma$ [25]. For the Bernardi–Raugel finite element method, this can be ensured by standard interpolators into either the Raviart–Thomas RT_0 or the Brezzi–Douglas–Marini BDM_1 finite element spaces [4]. Here, we employ the Brezzi–Douglas–Marini standard BDM_1 standard interpolator defined by

$$\int_F q_h(\Pi \mathbf{v}_h - \mathbf{v}_h) \cdot \mathbf{n}_F \, ds = 0 \quad \text{for all } q_h \in P_1(F) \text{ and } F \in \mathcal{F}.$$

Also note, that $\Pi(\mathbf{v}_h) = \mathbf{v}_h$ whenever $\mathbf{v}_h \in \mathbf{P}_1(\mathcal{T}) \cap \mathbf{H}^1_0$, hence only face bubbles are modified by the reconstruction operator (and their reconstruction equals their RT_0 standard interpolation).

Remark 3.1. In order to give an impression how the proposed space discretization can actually be implemented, we describe the discretization variant with the Raviart–Thomas standard interpolator in detail, although we will not use this slightly less accurate variant in our numerical experiments.

The Raviart–Thomas standard interpolator can be elementwise defined in an explicit way by

$$\Pi^{\mathrm{RT}_0}(\mathbf{v}_h)_{|T} = \mathbf{a}_T + \frac{c_T}{d}\left(\mathbf{x} - \mathbf{x}_T\right), \tag{3.4}$$

where $\mathbf{x}_T$ denotes the barycenter of the element T, c_T denotes the elementwise divergence computable by

$$c_T := \frac{1}{|T|} \sum_{F \in \mathcal{F}(T)} \int_F \mathbf{v}_h \cdot \mathbf{n}_F \, dS \tag{3.5}$$

and $\mathbf{a}_T$ denotes the *average velocity* computable by

$$\mathbf{a}_T := \frac{1}{|T|} \sum_{F \in \mathcal{F}(T)} \left(\int_F \mathbf{v}_h \cdot \mathbf{n}_F \, dS \right) (\mathbf{x}_F - \mathbf{x}_T),$$

where $\mathbf{x}_F$ denotes a face barycenter of the face F.

The following lemma collects some more important properties.

Lemma 3.2 (Properties of Π)**.** It holds, for all $\mathbf{v} \in \mathbf{H}^1_0(\Omega)$,

$$\mathrm{div}(\Pi \mathbf{v}) = \mathrm{div}_h \mathbf{v}, \tag{3.6}$$

$$\|\mathbf{u} - \Pi \mathbf{v}\|_{L^2(T)} \leqslant h_T \|\nabla \mathbf{v}\|_{L^2(T)} \quad \text{for all } T \in \mathcal{T}, \tag{3.7}$$

$$\int_\Omega \nabla \phi \cdot (\Pi \mathbf{v}) \, dx = -\int_\Omega \phi \, \mathrm{div}_h \mathbf{v} \, dx \quad \text{for all } \phi \in H^1(\Omega). \tag{3.8}$$

Proof. The properties (3.6) and (3.7) follow from the properties of the standard interpolation into the spaces BDM_1 (and RT_0), see e.g. [4]. Property (3.8) follows from an integration by parts and property (3.6). □

Remark 3.3. The cornerstone of the novel gradient-robust scheme is given by the following statement: for all *discretely-divergence-free* Bernardi–Raugel vector fields $\mathbf{v}_h^0 \in \mathbf{V}_h^0$ and all $\phi \in H^1(\Omega)$ it holds

$$\int_\Omega \nabla \phi \cdot (\Pi \mathbf{v}_h^0) \, dx = -\int_\Omega \phi \, \mathrm{div}_h \mathbf{v}_h^0 \, dx = 0,$$

i.e., the reconstruction operator Π enables to repair the $\mathbf{L}^2$ orthogonality of *discretely divergence-free* vector fields and arbitrary gradient fields $\nabla \phi$.

3.3. **Coupling to finite volume upwind discretization of continuity equation.** This finite element scheme is coupled to a finite volume discretization for the continuity equation. Altogether, our discretization seeks some $(\mathbf{u}_h, p_h, \varrho_h) \in \mathbf{V}_h \times Q_h \times Q_h$ such that

$$
\begin{aligned}
a_1(\mathbf{u}_h, \mathbf{v}_h) + a_2(\Pi\mathbf{u}_h, \Pi\mathbf{v}_h) + b(p_h, \mathbf{v}_h) &= F(\Pi\mathbf{v}_h) + G(\varrho_h, \Pi\mathbf{v}_h) \qquad \text{for all } \mathbf{v}_h \in \mathbf{V}_h,\\
\mathrm{div}_{\text{upw}}(\varrho_h\mathbf{u}_h) &= 0,\\
p_h &= \varphi(\varrho_h).
\end{aligned} \tag{3.9}
$$

The upwind discretization $\mathrm{div}_{\text{upw}}(\varrho_h\mathbf{u}_h) \in P_0(\mathcal{T})$ of $\mathrm{div}(\varrho_h\mathbf{u}_h)$ is defined on all $T \in \mathcal{T}$ by

$$
\begin{aligned}
\mathrm{div}_{\text{upw}}(\varrho_h\mathbf{u}_h)|_T &:= \frac{1}{|T|} \sum_{F\in\mathcal{F}(T)\cap\mathcal{F}(L)} u^+_{T,F}\varrho_h|_T - u^-_{T,F}\varrho_h|_L\\
&= \frac{1}{|T|} \sum_{F\in\mathcal{F}(T)\cap\mathcal{F}(L)} \varrho_F^{\text{upw}} u_{T,F},
\end{aligned}
$$

where $u_{T,F} = \int_F \mathbf{u}_h \cdot \mathbf{n}_T \, ds$ is the integral over the face F in outer normal direction of the simplex T and $u^+_{K,F} \geqslant 0$ and $u^-_{T,F} \geqslant 0$ is the positive and negative part, respectively. Hence, $\varrho_F^{\text{upw}} := \varrho_h|_T$ if $u_{T,F} > 0$ and $\varrho_F^{\text{upw}} := \varrho_h|_L$ else for $F = \partial T \cap \partial L$.

The introduction of the upwind divergence leads to a (singular) matrix

$$
\mathrm{div}_{\text{upw}}(\varrho_h\mathbf{u}_h) = 0 \quad \Leftrightarrow \quad D\varrho_h = 0 \quad \text{where} \quad D_{jk} := \mathrm{div}_{\text{upw}}(\chi_j\mathbf{u}_h)|_{T_k} \tag{3.10}
$$

where χ_j is the characteristic function of $T_j \in \mathcal{T}$.

Lemma 3.4 (Properties of D)**.** It holds

(1) D is weakly diagonal-dominant,i.e.

$$
D_{jj} \geqslant 0 \quad \text{and} \quad \sum_k D_{jk} = 0 \quad \text{for all } j = 1, \dots, \mathcal{T},
$$

(2) $D^T\mathbf{1} = 0,$

(3) $D\mathbf{1} = \mathrm{div}_{\text{upw}}(\mathbf{u}_h) = \pi_0\mathrm{div}(\mathbf{u}_h).$

Proof. The proof of (1) and (2) is common for finite volume discretizations and follows straightforwardly from the relation $u^{\pm}_{T,F} = u^{\mp}_{L,F}$ for $F \in \mathcal{F}(T) \cap \mathcal{F}(L)$. For the proof of (3) recall $u_{L,F} = u^+_{L,F} - u^-_{L,F}$. □

Since (3.9) is a nonlinear problem, it has to be solved iteratively and one has to choose a reasonable solution ϱ_h with $\mathrm{div}_{\text{upw}}(\varrho_h\mathbf{u}_h) = 0$ that satisfies the non-negativity and mass constraints. Consider a given approximation $\mathbf{u}_h$ and ϱ_h^{n-1} (from a previous fixpoint iterate or an initial solution). To compute a unique update ϱ_h^n of the discrete density that preserves the non-negativity and the integral mean of ϱ_h^{n-1}, we suggest to employ the backward Euler method. Given the (diagonal) P_0 mass matrix M, i.e. $M_{jj} := |T_j|$ and some time step τ, this leads to the linear problem

$$
(M + \tau D)\varrho_h^n = M\varrho_h^{n-1}. \tag{3.11}
$$

Here, ϱ_h^n has to be understood as a column vector with the elementwise constant values of $\varrho_h^n \in P_0(\mathcal{T})$.

Lemma 3.5 (Preservation of non-negativity and integral mean)**.** It holds

(1) $\varrho_h^n \geqslant 0$ if $\varrho_h^{n-1} \geqslant 0,$

(2) $M(\varrho_h^n - \varrho_h^{n-1}) \cdot \mathbf{1} = 0.$

Proof. Since M is a positive diagonal matrix, $M+\tau D$ is diagonal-dominant and hence an M-matrix. This implies that the inverse $(M+\tau D)^{-1}$ is totally positive and hence preserves the non-negativity of ϱ_h^{n-1}. The second property follows from $D^T\mathbf{1}=0$ used in the identity

$$(M\varrho_h^n)\cdot\mathbf{1}=\varrho_h^n\cdot((M+\tau D)^T\mathbf{1})=((M+\tau D)\varrho_h^n)\cdot\mathbf{1}=(M\varrho_h^{n-1})\cdot\mathbf{1}.$$

This concludes the proof. □

The pseudo time-stepping (3.11) is embedded into the iterative algorithm in Section 5.1.

4. On gradient-robustness and well-balanced schemes

In analogy to (2.4), consider the discrete space

$$\mathbf{V}_h^\perp := \{\mathbf{v}_h\in\mathbf{V}_h : (\boldsymbol{\epsilon}(\mathbf{v}_h),\boldsymbol{\epsilon}(\mathbf{w}_h^0))=0 \text{ for all } \mathbf{w}_h^0\in\mathbf{V}_h^0\}$$

which allows for the orthogonal splitting $\mathbf{V}_h := \mathbf{V}_h^0\oplus\mathbf{V}_h^\perp$ in the discrete scalar product $(\boldsymbol{\epsilon}(\bullet),\boldsymbol{\epsilon}(\bullet))$. The main structural property of the gradient-robust scheme (3.9) is now derived by:

Theorem 4.1. Exploiting the splitting $\mathbf{u}_h=\mathbf{u}_h^0+\mathbf{u}_h^\perp$ with $\mathbf{u}_h^0\in\mathbf{V}_h^0$ and $\mathbf{u}^\perp\in\mathbf{V}_h^\perp$, the *discretely divergence-free part* $\mathbf{u}_h^0$ fulfills a *pressure-robust* discretization of the *incompressible Stokes* problem in the form: for all $\mathbf{v}_h^0\in\mathbf{V}_h^0$ it holds

$$a_1(\mathbf{u}_h^0,\mathbf{v}_h^0)=F(\Pi\mathbf{v}_h^0)+G(\varrho_h,\Pi\mathbf{v}_h^0)=(\mathbb{P}(\mathbf{f}+\rho_h\mathbf{g}),\Pi\mathbf{v}_h^0)).$$

Remark 4.2. Theorem 4.1 is the discrete equivalent to the continuous relation (2.8). We emphasize the appearance of the *continuous* Helmholtz–Hodge projector $\mathbb{P}(\mathbf{f}+\rho_h\mathbf{g})$. Actually, it is again Theorem 4.1 that makes the scheme *asymptotic-preserving* in the low Mach number limit, where the non-divergence-free part $\mathbf{u}_h^\perp$ of the discrete velocity solution $\mathbf{u}_h$ should vanish in the limit.

Consider the compressible Stokes problem (2.3) with the right-hand sides

$$\mathbf{f}:=\nabla q \quad\text{and}\quad \mathbf{g}=0$$

for some $q\in H^1(\Omega)$. For this setting, one can observe that the solution $(\mathbf{u},p)=(\mathbf{0},q+C)$ of the incompressible Stokes problem also solves the compressible problem if there is enough mass in the system. Indeed, if it exists a (global) constant C, such that $\varrho:=q/c+C$ satisfies the mass constraint $\int_\Omega\varrho\,dx=M$ and is non-negative $\varrho\geqslant 0$, then the solution of the incompressible Stokes problem also is a solution of the compressible problem.

Vice versa, assume that $\varrho\geqslant 0$ satisfies the mass constraint and $\nabla(\varphi(\varrho))=\nabla q$. Then, it is clear that $(\mathbf{u},p,\varrho)=(\mathbf{0},\varphi(\varrho),\varrho)$ solves the compressible Stokes problem and $(\mathbf{u},p)=(\mathbf{0},\varphi(\varrho))$ solves the incompressible Stokes problem. This proves the following lemma.

Lemma 4.3. The compressible Stokes problem with right-hand sides

$$\mathbf{f}:=\nabla q \quad\text{and}\quad \mathbf{g}=0$$

has a hydrostatic solution $\mathbf{u}=\mathbf{0}$, if and only if it exists a (global) constant C, such that $\varrho:=q/c+C$ satisfies the mass constraint $\int_\Omega\varrho\,dx=M$ and is non-negative, i.e. $\varrho\geqslant 0$. The pair $(\mathbf{u},p):=(\mathbf{0},q)$ also solves the incompressible Stokes problem.

Definition 4.4 (Well-balanced property)**.** A discretization of the compressible Stokes problem is called *well-balanced* if it computes hydrostatic solutions $\mathbf{u}=\mathbf{0}$ correctly if the right-hand side is balanced by the gradient of some admissible pressure-density pair.

5. Existence of discrete solutions

The discussion in Subsection 3.3 and in Section 4 motivates the following pseudo-time stepping algorithm and the choice of its initial value. Subsection 5.2 proves that this algorithm has a fixed point, which is a discrete solution of (3.9).

5.1. **An iterative algorithm with well-balanced initial solution.** The previous discussion motivates to choose the initial solution by a solve of the incompressible Stokes equations and a rescaling of its pressure. In case of a well-balanced situation as in Lemma 4.3, this then already gives a discrete solution of the compressible system. Otherwise, one enters a suitable fixed point iteration.

Input.

- some triangulation $\mathcal{T}$,
- stepsize $\tau > 0$.

Initial Step.

- Set $\varrho_{-1} \equiv M/|\Omega|$.
- Solve the incompressible Stokes system, i.e., find $\mathbf{u}_0 \in \mathbf{V}_h$ and $p_0 \in Q_h$ such that

$$\begin{aligned} a_1(\mathbf{u}_0, \mathbf{v}_h) + b(p_0, \mathbf{v}_h) &= F(\Pi \mathbf{v}_h) + G(\varrho_{-1}, \Pi \mathbf{v}_h) && \text{for all } \mathbf{v} \in \mathbf{V}_h, \\ b(q_h, \mathbf{u}_0) &= 0 && \text{for all } q_h \in Q_h. \end{aligned}$$

- Set $\varrho_0 := p_0/c + C$, where $C \in \mathbb{R}$ is chosen such that ϱ_0 satisfies $\int_\Omega \varrho_0 \, dx = M$ and $\varrho_0 \geqslant 0$. If this is not possible, start with $\varrho_0 = \varrho_{-1}$ and $\mathbf{u}_h = 0$.

Loop (start with $n = 1$).

- Update matrix D according to (3.10) (with $\mathbf{u}_h = \mathbf{u}_h^{n-1}$) and find $\varrho_h^n \in Q_h$ such that

$$(M + \tau D)\varrho_h^n = M\varrho_h^{n-1}. \tag{5.1}$$

- Update the pressure according to the equation of state, i.e.

$$p_h^n := \varphi(\varrho_h^n) = c\varrho_h^n. \tag{5.2}$$

- Find $\mathbf{u}_h^n \in \mathbf{V}_h$ that satisfies the momentum equation

$$a_1(\mathbf{u}_h^n, \mathbf{v}_h) + a_2(\Pi \mathbf{u}_h^n, \Pi \mathbf{v}_h) = F(\Pi \mathbf{v}_h) + G(\varrho_h^n, \Pi \mathbf{v}_h) - b(p_h^n, \mathbf{v}_h) \quad \text{for all } \mathbf{v}_h \in \mathbf{V}_h. \tag{5.3}$$

- Compute residuals of the stationary momentum equation and the continuity equation, i.e.

$$\text{res} := \|a_1(\mathbf{u}_h^n, \bullet) + a_2(\Pi \mathbf{u}_h^n, \Pi \bullet) - F(\Pi \bullet) + G(\varrho_h^n, \Pi \bullet) - b(p_h^n, \bullet)\|_{l^2} + |\mathrm{div}_{\mathsf{upw}}(\varrho_h^n \mathbf{u}_n)|$$

- Stop if res $<$ tol, otherwise increase n by one and restart loop.

Remark 5.1. Note, that one only can prove that there exists some discrete solution (see Subsection 5.2), but it is not guaranteed that the algorithm converges. In our numerical benchmarks, we could enforce convergence by choosing small enough time steps τ.

5.2. **Existence of a fixed point.** Note, that there is no uniqueness result for the continuous compressible Stokes system, but one can show existence of a (discrete) solution for the (discretized) compressible Stokes problem. To do so we mainly follow the argumentation in [12]. There the existence of a weak solution with $\varrho \in L^2$ and $p = \varphi(\varrho) := c\varrho$ is proven. The main argument concerns the proof of the a priori stability estimate

$$\min\{2\mu + \lambda, \mu\}\|\nabla \mathbf{u}_h\|_{L^2} + \|p_h\|_{L^2(\Omega)} + \|\varrho_h\|_{L^2(\Omega)} \lesssim 1$$

which is needed in the convergence proof via some Brouwer fixed point argument. The crucial point is the vanishing term

$$\int_\Omega p\mathrm{div}(\mathbf{u}) = c\int_\Omega \varrho\mathrm{div}(\mathbf{u}) = -c\int_\Omega \nabla(\log\varrho)\cdot(\varrho\mathbf{u}) = c\int_\Omega \log\varrho\,\mathrm{div}(\varrho\mathbf{u}) = 0 \tag{5.4}$$

for $\varrho \in C^1(\overline{\Omega})$, which can also be generalized to $\varrho \in L^2(\Omega)$, see [8, Appendix A] for details.

A similar stability estimate holds for the discrete scheme which requires the following Lemma.

Lemma 5.2. For any convex and twice continuously differentiable function $\phi : [0,\infty) \to \mathbb{R}^+$, it holds

$$\int_\Omega \phi'(\varrho_h)\mathrm{div}_{\mathsf{upw}}(\varrho_h\mathbf{u}_h)\,dx - \int_\Omega (\varrho_h\phi'(\varrho_h) - \phi(\varrho_h))\mathrm{div}(\mathbf{u}_h)\,dx$$
$$= \frac{1}{2}\sum_{F_{KL}\in\mathcal{F}(\Omega)} \phi''(\varrho_{KL})\frac{|u_{K,F_{KL}}|}{|F_{KL}|}\|[[\varrho_h]]\|^2_{L^2(F_{KL})} \geqslant 0,$$

where the quantities $\varrho_{KL} \in (\varrho_h|_K, \varrho_h|_L)$ denote intermediate values on every face $F_{KL} \in \mathcal{F}(\Omega)$ according to remainders in corresponding Taylor expansions.

Proof. By convexity of ϕ and Taylor expansion, it holds

$$\phi'(x)(x-y) - \phi(x) + \phi(y) = \frac{1}{2}\phi''(s)(x-y)^2 \geqslant 0 \quad \text{for some } s \in (x,y). \tag{5.5}$$

The integrals in the assertion can be rewritten into

$$\int_\Omega \phi'(\varrho_h)\mathrm{div}_{\mathsf{upw}}(\varrho_h\mathbf{u}_h)\,dx = \sum_{T\in\mathcal{T}}\sum_{F\in\mathcal{F}(T)} \phi'(\varrho_h|_T)\varrho_F^{\mathsf{upw}}u_{T,F}$$
$$\int_\Omega (\varrho_h\phi'(\varrho_h) - \phi(\varrho_h))\mathrm{div}(\mathbf{u}_h)\,dx = \sum_{T\in\mathcal{T}}\sum_{F\in\mathcal{F}(T)} (\varrho_h|_T\phi'(\varrho_h|_T) - \phi(\varrho_h|_T))u_{T,F}.$$

Hence, their difference reads

$$\begin{aligned}
&\int_\Omega \phi'(\varrho_h)\mathrm{div}_{\mathsf{upw}}(\varrho_h\mathbf{u}_h)\,dx - \int_\Omega (\varrho_h\phi'(\varrho_h) - \phi(\varrho_h))\mathrm{div}(\mathbf{u}_h)\,dx\\
&= \sum_{T\in\mathcal{T}}\sum_{F\in\mathcal{F}(T)} \left(\phi'(\varrho_h|_T)\varrho_F^{\mathsf{upw}} - \varrho_h|_T\phi'(\varrho_h|_T) + \phi(\varrho_h|_T)\right)u_{T,F}\\
&= \sum_{F_{KL}\in\mathcal{F}(\Omega)} \left(\phi'(\varrho_h|_K)(\varrho_{F_{KL}}^{\mathsf{upw}} - \varrho_h|_K) + \phi'(\varrho_h|_L)(\varrho_h|_L - \varrho_{F_{KL}}^{\mathsf{upw}}) + \phi(\varrho_h|_K) - \phi(\varrho_h|_L)\right)u_{K,F_{KL}}\\
&=: \sum_{F_{KL}\in\mathcal{F}(\Omega)} u_{K,F_{KL}}\theta_{KL}
\end{aligned}$$

where the last sum collects the flux jumps θ_{KL} over all interior faces $F_{KL} \in \mathcal{F}(\Omega)$ (on boundary faces it holds $u_{K,F_{KL}} = 0$). It remains to show that each summand is non-negative. The first case assumes

$u_{K,F_{KL}} > 0$ and hence $\varrho^{\text{upw}}_{F_{KL}} = \varrho_h|_K$. Then, one obtains for the jump term

$$\theta_{KL} = \phi'(\varrho_h|_L)(\varrho_h|_L - \varrho_h|_K) + \phi(\varrho_h|_K) - \phi(\varrho_h|_L) = \phi''(\varrho_{KL})(\varrho_h|_K - \varrho_h|_L)^2 \geqslant 0$$

due to (5.5) where s is renamed to ϱ_{KL}. In the other case $u_{K,F_{KL}} < 0$ it holds $\varrho^{\text{upw}}_{F_{KL}} = \varrho_h|_L$ and hence

$$\theta_{KL} = \phi'(\varrho_h|_K)(\varrho_h|_L - \varrho_h|_K) + \phi(\varrho_h|_K) - \phi(\varrho_h|_L) = -\phi''(\varrho_{KL})(\varrho_h|_K - \varrho_h|_L)^2 \leqslant 0$$

again by (5.5) (multiplied by -1). Hence

$$\begin{aligned}\int_\Omega \phi'(\varrho_h)\mathrm{div}_{\text{upw}}(\varrho_h \mathbf{v}_h)\,dx - \int_\Omega (\varrho_h\phi'(\varrho_h) - \phi(\varrho_h))\mathrm{div}(\mathbf{v}_h)\,dx \\ &= \sum_{F_{KL}\in\mathcal{F}(\Omega)} |u_{K,F_{KL}}|\phi''(\varrho_{KL})(\varrho_h|_K - \varrho_h|_L)^2 \\ &= \frac{1}{2}\sum_{F_{KL}\in\mathcal{F}(\Omega)} \phi''(\varrho_{KL})\frac{|u_{K,F_{KL}}|}{|F_{KL}|}\|[[\varrho_h]]\|^2_{L^2(F_{KL})} \geqslant 0.\end{aligned}$$

This concludes the proof. □

Lemma 5.3 (Discrete stability estimate)**.** For any solution $(\mathbf{u}_h, p_h, \varrho_h)$ of the discrete scheme, it holds

$$\min\{2\mu+\lambda,\mu\}\|\nabla \mathbf{u}_h\|_{L^2} \lesssim \|\mathbf{f}\|_{L^2} + \|\varrho_h\|_{L^2(\Omega)}\|\mathbf{g}\|_{L^\infty}, \tag{5.6}$$

$$\sum_{F_{KL}\in\mathcal{F}(\Omega)} \varrho_{KL}^{-1}\frac{|u_{K,F_{KL}}|}{|F_{KL}|}\|[[\varrho_h]]\|^2_{L^2(F_{KL})} \lesssim c^{-1}\min\{2\mu+\lambda,\mu\}^{-1}\left(\|\mathbf{f}\|_{L^2} + \|\varrho_h\|_{L^2(\Omega)}\|\mathbf{g}\|_{L^\infty}\right), \tag{5.7}$$

$$\text{If } \mathbf{g} \equiv \mathbf{0} \implies \|p_h\|_{L^2(\Omega)} = c\|\varrho_h\|_{L^2(\Omega)} \lesssim \|\mathbf{f}\|_{L^2} + c. \tag{5.8}$$

The hidden constants in $\lesssim$ depend neither on the mesh width h, nor the viscosity parameters μ and λ nor on c.

Proof. Testing the momentum equation with $\mathbf{u}_h$ yields

$$2\mu\|\epsilon(\mathbf{u}_h)\|^2_{L^2} + \lambda\|\mathrm{div}_h\mathbf{u}_h\|^2_{L^2} - \int_\Omega p_h\mathrm{div}\mathbf{u}_h\,dx = \int \mathbf{f}\cdot\Pi\mathbf{u}_h\,dx + \int \varrho_h\mathbf{g}\cdot\Pi\mathbf{u}_h\,dx. \tag{5.9}$$

The approximation properties of Π yield

$$\int \mathbf{f}\cdot\Pi\mathbf{u}_h\,dx \leqslant (C_F + hC_\Pi)\|\mathbf{f}\|_{L^2}\|\nabla\mathbf{u}_h\|_{L^2}.$$

The integral with $\mathbf{g}$ is estimated similarly by

$$\int \varrho_h\mathbf{g}\cdot\Pi\mathbf{u}_h\,dx \leqslant (C_F + hC_\Pi)\|\varrho_h\|_{L^2(\Omega)}\|\mathbf{g}\|_{L^\infty}\|\nabla\mathbf{u}_h\|_{L^2}$$

and it remains to handle the integral on the left-hand side. Lemma 5.2 shows, for $\phi(s) := s\log(s)$ and with $\mathrm{div}_{\text{upw}}(\varrho_h\mathbf{u}_h) = 0$ due to (3.9), that

$$\begin{aligned}\int_\Omega p_h\mathrm{div}(\mathbf{u}_h)\,dx &= c\int_\Omega \varrho_h\mathrm{div}(\mathbf{u}_h)\,dx \\ &= c\int_\Omega (1+\log(\varrho_h))\mathrm{div}_{\text{upw}}(\varrho_h\mathbf{u}_h)\,dx - c\sum_{F_{KL}\in\mathcal{F}(\Omega)}\frac{\phi''(\varrho_{KL})}{2|F_{KL}|}|u_{K,F_{KL}}|\|[[\varrho_h]]\|^2_{L^2(F_{KL})} \\ &= -c\sum_{F_{KL}\in\mathcal{F}(\Omega)}\frac{\varrho_{KL}^{-1}}{2|F_{KL}|}|u_{K,F_{KL}}|\|[[\varrho_h]]\|^2_{L^2(F_{KL})} \leqslant 0.\end{aligned} \tag{5.10}$$

Assume that $\lambda \geqslant 0$. Then, it holds

$$\lambda \|\mathrm{div}_h \mathbf{u}\|_{L^2}^2 \geqslant 0$$

and hence

$$\mu \|\nabla \mathbf{u}_h\|_{L^2}^2 \leqslant 2\mu \|\epsilon(\mathbf{u}_h)\|_{L^2}^2 \lesssim \left(\|\mathbf{f}\|_{L^2} + \|\varrho_h\|_{L^2(\Omega)} \|\mathbf{g}\|_{L^\infty}\right) \|\nabla \mathbf{u}_h\|_{L^2}.$$

Division by $\|\nabla \mathbf{u}_h\|_{L^2}$ concludes the proof in this case.

In the case $0 > \lambda > -2\mu$, elementary vector calculus identities yield

$$\min\{\mu, 2\mu + \lambda\} \|\nabla \mathbf{u}_h\|_{L^2}^2 \leqslant \mu \|\mathrm{rot}\mathbf{u}\|_{L^2}^2 + (2\mu + \lambda) \|\mathrm{div}\mathbf{u}\|_{L^2}^2 = 2\mu \|\epsilon(\mathbf{u}_h)\|_{L^2}^2 + \lambda \|\mathrm{div}\mathbf{u}\|_{L^2}^2.$$

Moreover, it holds

$$0 \geqslant \lambda \|\mathrm{div}_h \mathbf{u}\|_{L^2}^2 \geqslant \lambda \|\mathrm{div}\mathbf{u}\|_{L^2}^2$$

and hence

$$\min\{\mu, 2\mu + \lambda\} \|\nabla \mathbf{u}_h\|_{L^2}^2 \leqslant 2\mu \|\epsilon(\mathbf{u}_h)\|_{L^2}^2 + \lambda \|\mathrm{div}_h \mathbf{u}\|_{L^2}^2 \lesssim \left(\|\mathbf{f}\|_{L^2} + \|\varrho_h\|_{L^2(\Omega)} \|\mathbf{g}\|_{L^\infty}\right) \|\nabla \mathbf{u}_h\|_{L^2}.$$

Division by $\|\nabla \mathbf{u}_h\|_{L^2}$ concludes the proof of (5.6). The proof of (5.7) follows from a combination of (5.9) and (5.10) together with the already proven estimate (5.6), i.e.

$$\begin{aligned}\sum_{F_{KL} \in \mathcal{F}(\Omega)} \varrho_{KL}^{-1} \frac{|u_{K,F_{KL}}|}{|F_{KL}|} \|[[\varrho_h]]\|_{L^2(F_{KL})}^2 &\lesssim c^{-1} \left(\|\mathbf{f}\|_{L^2} + \|\varrho_h\|_{L^2(\Omega)} \|\mathbf{g}\|_{L^\infty}\right) \|\nabla \mathbf{u}_h\|_{L^2} \\ &\lesssim c^{-1} \min\{2\mu + \lambda, \mu\}^{-1} \left(\|\mathbf{f}\|_{L^2} + \|\varrho_h\|_{L^2(\Omega)} \|\mathbf{g}\|_{L^\infty}\right)^2.\end{aligned}$$

For the proof of (5.8), consider a test function $\mathbf{v}_h$ with $\mathrm{div}_h(\mathbf{v}_h) = p_h - \overline{p}_h$, where $\overline{p}_h := |\Omega|^{-1} \int_\Omega p_h \, dx$, and $\|\nabla \mathbf{v}_h\|_{L^2} \lesssim \|p_h - \overline{p}_h\|_{L^2}$, using discrete inf-sup stability. Inserting this test function in the momentum equation and using also bounds from (5.6), it follows the estimate

$$\begin{aligned}\|p_h - \overline{p}_h\|_{L^2}^2 &= \lambda \int_\Omega \mathrm{div}_h(\mathbf{u}_h) \mathrm{div}_h(\mathbf{v}_h) \, dx + 2\mu \int_\Omega \epsilon(\mathbf{u}_h) : \epsilon(\mathbf{v}_h) \, dx - \int \mathbf{f} \cdot \Pi \mathbf{v}_h \, dx \\ &\lesssim \|\mathbf{f}\|_{L^2} \|p_h - \overline{p}_h\|_{L^2}\end{aligned} \tag{5.11}$$

where the constant C also depends on $\|\mathbf{f}\|_{L^2}$ (note that we assumed here that $\mathbf{g} \equiv 0$). For the following estimate we also need a bound on $\|\overline{p}_h\|_{L^2}$ that can be obtained due to

$$\overline{p}_h = \frac{1}{|\Omega|} \int_\Omega p_h \, dx = \frac{c}{|\Omega|} \int_\Omega \varrho_h \, dx = \frac{cM}{|\Omega|}. \tag{5.12}$$

A Pythagoras argument and the combination of (5.11) and (5.12) results in

$$\|p_h\|_{L^2}^2 = \|p_h - \overline{p}_h\|_{L^2}^2 + \|\overline{p}_h\|_{L^2}^2 \lesssim \|\mathbf{f}\|_{L^2}^2 + \frac{(cM)^2}{|\Omega|} \lesssim \|\mathbf{f}\|_{L^2}^2 + c^2.$$

This concludes the proof of (5.8). □

Lemma 5.4 (Existence of a discrete solution)**.** On every (fixed) shape-regular mesh in the sense of Section 3, the discrete nonlinear equation system (3.9) has at least one solution.

Proof. The existence of a discrete weak solution $(\mathbf{u}_h, p_h, \varrho_h) \in \mathbf{V}_h \times Q_h \times Q_h$ is proved by the Brouwer fixed-point theorem. We start the algorithm presented in Subsection 5.1 with $\varrho_{-1} \equiv M/|\Omega|$ and $\mathbf{u}_{-1} = \mathbf{0}$, $p_{-1} = \varphi(\varrho_{-1})$. Obviously, the discrete start value $(\mathbf{u}_{-1}, p_{-1}, \varrho_{-1}) \in \mathbf{V}_h \times Q_h \times Q_h$ lies in a finite-dimensional product space of convex sets with finite diameter that is itself convex. Now a mapping

$$\begin{aligned} f : \mathbf{V}_h \times Q_h \times Q_h &\to \mathbf{V}_h \times Q_h \times Q_h, \\ (\mathbf{u}_n, p_n, \varrho_n) &\mapsto f(\mathbf{u}_n, p_n, \varrho_n) := (\mathbf{u}_{n+1}, p_{n+1}, \varrho_{n+1}) \end{aligned}$$

is constructed by composition of (5.1), (5.2) and (5.3) where an (arbitrary) stepsize $\tau > 0$ is fixed. Then, the mapping defined by (5.1) is linear and continuous, since $(M + \tau D)$ is an invertible matrix. The mapping defined by (5.2) is again continuous. Finally, the mapping defined by (5.3) is linear and continuous, since the assumptions on the viscosities μ and λ assure that the discrete bilinear form $a_1(\mathbf{u}_h, \mathbf{v}_h) + a_2(\Pi\mathbf{u}_h, \Pi\mathbf{v}_h)$ is coercive. Therefore, the composed mapping f constructed in the algorithm in Subsection 5.1 is continuous.

Due to discrete mass conservation in (5.1) (see Lemma 3.5), it follows

$$\|\rho_{n+1}\|_{L^1} = \|\rho_n\|_{L^1} = \ldots = \|\rho_{-1}\|_{L^1} = M.$$

Due to this and the equivalence of all norms in finite dimensions, it holds $\|p_n\|_{L^2} = c\|\rho_n\|_{L^2} \leqslant C(h)$ for all $n \geqslant -1$. Hence, all (p_n, ϱ_n) for $n \geqslant -1$ lie in the same convex set. Finally, similar to the proof of (5.6), a discrete bound can be proved for $\|\nabla\mathbf{u}_{n+1}\|_{L^2}$. The only difference is that one cannot assume that $\mathbf{u}_{n+1}$ fulfills the discrete mass conservation in (3.9) with $\mathrm{div}_{\mathsf{upw}}(\varrho_{n+1}\mathbf{u}_{n+1}) = 0$. Therefore, the term $(p_{n+1}, \mathrm{div}(\mathbf{u}_{n+1}))$ has to be estimated in a different way. However, since the grid is fixed, the argument above yields

$$|(p_{n+1}, \mathrm{div}(\mathbf{u}_{n+1}))| \leqslant \|p_{n+1}\|_{L^2}\|\nabla\mathbf{u}_{n+1}\|_{L^2} \leqslant C(h)\|\nabla\mathbf{u}_{n+1}\|_{L^2},$$

and one derives a similar estimate like in (5.6). Therefore, f is a continuous function that maps a convex set into itself. According to the Brouwer fixed point theorem, this mapping has a fixed-point that is a solution of (3.9). □

6. Convergence of the scheme

This section proves convergence of the discrete solutions to a weak solution of (2.3).

Theorem 6.1. We assume that it holds $\mathbf{g} = \mathbf{0}$. Consider a sequence of shape-regular triangulations $(\mathcal{T}_k)_{k\in\mathbb{N}})$ with mesh width $h_k \to 0$. Let $(\mathbf{u}_k, p_k, \varrho_k)$ denote the corresponding discrete solutions of (3.9) on the meshes $\mathcal{T}_k$. Then, up extraction of a subsequence, it holds

(i) the sequence $(\mathbf{u}_k)_{k\in\mathbb{N}}$ converges weakly/strongly in $\mathbf{H}_0^1(\Omega)/\mathbf{L}^2(\Omega)$ to a limit $\mathbf{u} \in \mathbf{H}_0^1(\Omega)$,

(ii) the sequence $(p_k)_{k\in\mathbb{N}} = (c\varrho_k)_{k\in\mathbb{N}}$ converges weakly in $L^2(\Omega)$ to a limit $p = c\varrho \in L^2(\Omega)$,

(iii) the limit $(\mathbf{u}, p, \varrho)$ is a solution of (2.3).

Proof. The weak convergence and the existence of the limit $(\mathbf{u}, p, \varrho) \in \mathbf{H}_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega)$ follows from Lemma 5.3 and standard arguments from linear functional analysis. Hence, it remains to prove (iii).

Step 1. $(\mathbf{u}, p)$ satisfy the momentum equation, i.e., the first equation of (2.3).

Consider an arbitrary test function $\mathbf{v} \in \mathbf{C}_0^\infty$, which is dense in $\mathbf{H}_0^1(\Omega)$, and define $\mathbf{v}_k \in \mathbf{V}_k$ on $\mathcal{T}_k$ as its best approximation in the $\mathbf{H}_0^1$ norm, i.e., it holds for all $\mathbf{w}_h \in \mathbf{V}_h$: $(\nabla\mathbf{v}_h, \nabla\mathbf{w}_h) = (\nabla\mathbf{v}, \nabla\mathbf{w}_h)$. From standard arguments follows the strong convergence

$$\mathbf{v}_h \to \mathbf{v} \text{ in } \mathbf{H}_0^1(\Omega). \tag{6.1}$$

This strong convergence and the weak convergence of $(\mathbf{u}_k)$ to $\mathbf{u}$ in $\mathbf{H}_0^1(\Omega)$ allows to conclude

$$\int_\Omega \epsilon(\mathbf{u}_k) : \epsilon(\mathbf{v}_k)\, dx \;\to\; \int_\Omega \epsilon(\mathbf{u}) : \epsilon(\mathbf{v})\, dx.$$

Similarly, (6.1) and the weak convergence of (p_k) to p in $L^2(\Omega)$ yield

$$\int_\Omega p_k \mathrm{div}(\mathbf{v}_k)\, dx \;\to\; \int_\Omega p\mathrm{div}(\mathbf{v})\, dx.$$

Since also $\|\mathrm{div}_h(\mathbf{v}) - \mathrm{div}(\mathbf{v})\|_{L^2(\Omega)} \to 0$, it follows

$$\int_\Omega \mathrm{div}_h(\mathbf{u}_k)\mathrm{div}_h(\mathbf{v}_k)\,dx = \int_\Omega \mathrm{div}(\mathbf{u}_k)\mathrm{div}(\mathbf{v})\,dx + \int_\Omega \mathrm{div}(\mathbf{u}_k)\left(\mathrm{div}_h(\mathbf{v}_k) - \mathrm{div}(\mathbf{v})\right)\,dx$$
$$\to \int_\Omega \mathrm{div}(\mathbf{u})\mathrm{div}(\mathbf{v})\,dx.$$

It remains to show convergence of the right-hand side integrals, which follows again by (6.1) and the weak convergence of (ϱ_k) to ϱ in $L^2(\Omega)$, i.e.

$$\int_\Omega \mathbf{f}\cdot\mathbf{v}_k\,dx \;\to\; \int_\Omega \mathbf{f}\cdot\mathbf{v}\,dx.$$

The combination of all convergence results concludes the proof of Step 1.

Step 2. $(\mathbf{u},\varrho)$ satisfy the continuity equation, i.e. the second equation of (2.3).

We define on every element $T \in \mathcal{T}$

$$\mathbf{q}_h|_T := I_h^{\mathrm{RT}_0}(\varrho_{\mathsf{upw}}\mathbf{u}_h|_T), \tag{6.2}$$

which is $\mathbf{H}(\mathrm{div},\Omega)$-conforming and divergence-free. For an arbitrary scalar P_1 function ψ_h, it holds

$$0 = (\psi_h, \mathrm{div}\,\mathbf{q}_h) = -(\mathbf{q}_h, \nabla\psi_h) = -(\mathbf{q}_h - \varrho_h\mathbf{u}_h, \nabla\psi_h) - (\varrho_h\mathbf{u}_h, \nabla\psi_h).$$

We estimate now the term

$$|(\mathbf{q}_h - \varrho_h\mathbf{u}_h, \nabla\psi_h)| = \left|\sum_T \nabla\psi_h|_T \cdot \int_T (\mathbf{q}_h - \varrho_h\mathbf{u}_h)\,dx\right|.$$

For the term under the integral we get

$$\int_T (\mathbf{q}_h - \varrho_h\mathbf{u}_h)\,dx = \int_T (\mathbf{q}_h - \varrho_h I_h^{\mathrm{RT}_0}\mathbf{u}_h)\,dx + \int_T \varrho_h(I_h^{\mathrm{RT}_0}\mathbf{u}_h - \mathbf{u}_h)\,dx.$$

Thus one obtains by the triangle inequality

$$\begin{aligned}
|(\mathbf{q}_h - \varrho_h\mathbf{u}_h, \nabla\psi_h)| &\leqslant \sum_T \left|\nabla\psi_h|_T \cdot \left(\int_T (\mathbf{q}_h - \varrho_h I_h^{\mathrm{RT}_0}\mathbf{u}_h)\,dx + \int_T \varrho_h(I_h^{\mathrm{RT}_0}\mathbf{u}_h - \mathbf{u}_h)\,dx\right)\right| \\
&\leqslant C\sum_T \left(\|\mathbf{q}_h - \varrho_h I_h^{\mathrm{RT}_0}\mathbf{u}_h\|_{L^1(T)} + \|\varrho_h(I_h^{\mathrm{RT}_0}\mathbf{u}_h - \mathbf{u}_h)\|_{L^1(T)}\right) \\
&\leqslant C\sum_T \|\mathbf{q}_h - \varrho_h I_h^{\mathrm{RT}_0}\mathbf{u}_h\|_{L^1(T)} + C\|\varrho_h\|_{L^2}\,\|I_h^{\mathrm{RT}_0}\mathbf{u}_h - \mathbf{u}_h\|_{L^2}.
\end{aligned} \tag{6.3}$$

The term $\|\varrho_h\|_{L^2}\,\|I_h^{\mathrm{RT}_0}\mathbf{u}_h - \mathbf{u}_h\|_{L^2}$ converges to 0, according to the interpolation properties of $I_h^{\mathrm{RT}_0}$ and the stability estimate for $\|\varrho_h\|_{L^2}$. It remains to estimate $\sum_T \|\mathbf{q}_h - \varrho_h I_h^{\mathrm{RT}_0}\mathbf{u}_h\|_{L^1(T)}$. Interpolation properties of $I_h^{\mathrm{RT}_0}$ yield

$$\begin{aligned}
\sum_T \|\mathbf{q}_h - \varrho_h I_h^{\mathrm{RT}_0}\mathbf{u}_h\|_{L^1(T)} &\lesssim \sum_T h_T \sum_{F\in\mathcal{F}(T)} \left|(\varrho_F^{\mathsf{upw}} - \varrho_h|_T)\int_F \mathbf{u}_h\cdot\mathbf{n}_F\,ds\right| \\
&\lesssim \sum_{F\in\mathcal{F}(\Omega)} h_F|[[\varrho_h]]_F|\,|\mathbf{u}_F| := A.
\end{aligned}$$

It holds $A \to 0$ which can be proven as follows. A Cauchy inequality shows

$$A \lesssim \left(\sum_{F\in\mathcal{F}(\Omega)} |\mathbf{u}_F|\varrho_{KL}^{-1}[[\varrho_h]]_F^2\right)^{1/2} \left(\sum_{F\in\mathcal{F}(\Omega)} h_F^2|\mathbf{u}_F|\varrho_{KL}\right)^{1/2}. \tag{6.4}$$

The left sum is bounded by (5.7) and it remains to show that the second sum converges to zero. A Hölder inequality, a trace inequality and a inverse inequality on some neighboring simplex T_F of F show

$$|\mathbf{u}_F| \lesssim \|\mathbf{u}\|_{L^2(T_F)}^{1/2} \|\nabla \mathbf{u}\|_{L^2(T_F)}^{1/2} \|1\|_{L^2(F)} \lesssim h_F^{d/2-1} \|\mathbf{u}\|_{L^2(T_F)}. \tag{6.5}$$

Then, another Cauchy inequality and some overlap arguments yield

$$\begin{aligned}\left(\sum_{F\in\mathcal{F}(\Omega)} h_F^2 |u_F| \varrho_{KL}\right)^{1/2} &\leqslant \left(\sum_{F\in\mathcal{F}(\Omega)} \|\mathbf{u}\|_{L^2(T_F)}^2\right)^{1/4} \left(\sum_{F\in\mathcal{F}(\Omega)} h_F^{d+2} \varrho_{KL}^2\right)^{1/4} \\ &\lesssim \|\nabla \mathbf{u}\|_{L^2(\Omega)}^{1/2} \left(\sum_{F\in\mathcal{F}(\Omega)} h_F^{d+2} \varrho_{KL}^2\right)^{1/4}\end{aligned}$$

To show that the last sum converges to zero, we use that ϱ_{KL}^2 is smaller than $\varrho_h|_{T_F}^2$ for some neighboring simplex T_F of T and hence

$$\left(\sum_{F\in\mathcal{F}(\Omega)} h_F^{d+2} \varrho_{KL}^2\right)^{1/4} \lesssim \left(\sum_{F\in\mathcal{F}(\Omega)} h_F^2 |T_F| \varrho_h|_{T_F}^2\right)^{1/4} \leqslant h^{1/2} \|\varrho_h\|_{L^2}^{1/2}.$$

According to (5.8) $\|\varrho_h\|_{L^2}$ is bounded and hence, one arrives at

$$A \lesssim h^{1/2}$$

which concludes the proof of Step 2.

Step 3. (p, ϱ) **satisfy the equation of state, i.e.** $p = \varphi(\varrho) = c\varrho$**.**

Consider any function $\varphi \in C_C^\infty(\Omega)$ and its piecewise-constant approximation $\varphi_h := \pi_0 \varphi$ which converges strongly to φ. Then its holds

$$\begin{aligned}\int_\Omega p_h \varphi_h \, dx &\to \int_\Omega p\varphi \, dx \\ \int_\Omega c\varrho_h \varphi_h \, dx &\to \int_\Omega c\varrho\varphi \, dx.\end{aligned}$$

Since the integrals on the right-hand side are equal for any h, also their limit integrals have to be equal, i.e.

$$\int_\Omega p\varphi \, dx = \int_\Omega c\varrho\varphi \, dx \quad \text{for all } \varphi \in C_C^\infty(\Omega)$$

Hence, it follows $p = c\varrho$. □

7. Numerical Experiments

This section reports on some two-dimensional numerical experiments assessing accuracy and asymptotic convergence rates of the novel scheme, which especially illustrates the increased robustness with respect to gradients in the momentum balance. Some experiments also show that the scheme might also converge in barotropic cases where $p = \varphi(\varrho) := c\varrho^\gamma$ with $\gamma > 1$. In this case however the proof of Step 3 in the convergence proof (without using additional stability terms in the scheme) is non-trivial and open.

The loop in the algorithm was stopped in all experiments until the tolerance criterion was satisfied with tol $:= 10^{-11}$. The time step in the evolution of the density was chosen small enough and usually $\tau \approx \nu/c$. The term 'ndof' refers to the number of degrees of freedom and is approximately $2\mathcal{T} + \mathcal{N} + \mathcal{E}$, i.e. the size of the ansatz spaces for density, pressure and velocity.

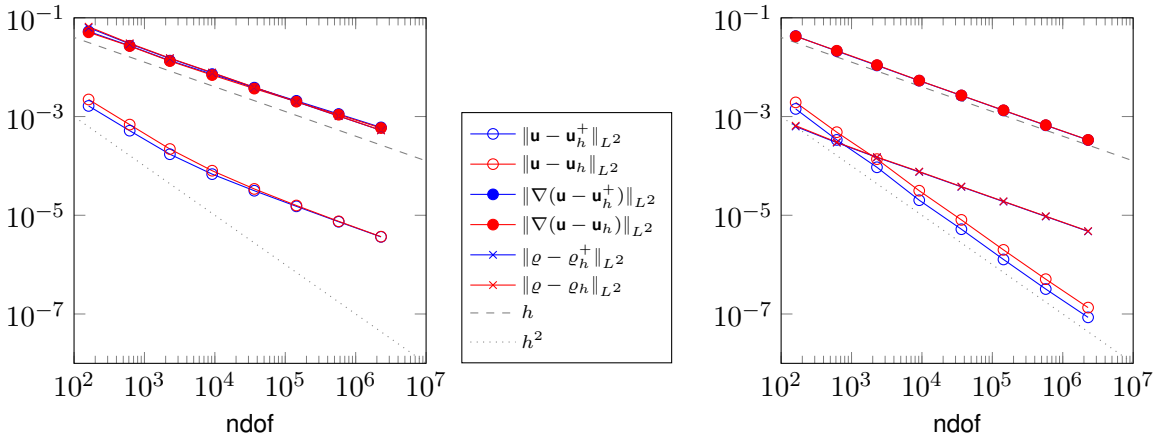

FIGURE 7.1. Convergence histories for the modified method and classical method for $\gamma = 1$ and $\mu = 1$ and $c = 1$ (left), $c = 100$ (right).

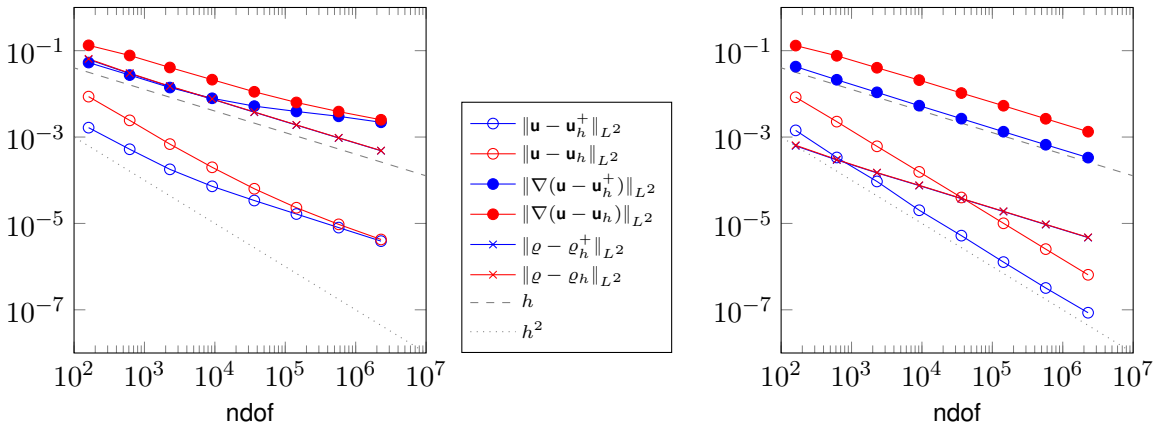

FIGURE 7.2. Convergence histories for the modified method and classical method for $\gamma = 1$ and $\mu = 10^{-1}$ and $c = 1$ (left), $c = 100$ (right).

Throughout this section, $(\mathbf{u}_h^+, \varrho_h^+)$ denotes the solution of the 'modified scheme' (3.9) and $(\mathbf{u}_h, \varrho_h)$ denotes the solution of the 'classical scheme' (3.9) where $\Pi = 1$.

7.1. **Manufactured solutions to study error convergence and locking.** This example on the unit square $\Omega := (0,1)^2$ studies the convergence rates of our discretization scheme and examines the exact solution

$$\mathbf{u} := \mathrm{curl}(x^2(x-1)^2y^2(y-1)^2)/\varrho, \qquad p = \varphi(\varrho) := c\varrho^\gamma$$

for different choices of γ, μ and $\lambda = -2\mu/3$. Assuming a linear density $\varrho := 1 + (y - 1/2)/c$, the first test case considers the isothermal configuration $\gamma = 1$ and afterwards a barotropic configuration with $\gamma = 1.4$ is presented. In all cases $\int_\Omega \varrho\, dx = 1$ holds independent of c. The right-hand side functions are chosen such that $(\mathbf{u}, p, \rho)$ is a solution of the compressible Stokes system with

$$\mathbf{f} := -2\mu\epsilon(\mathbf{u}) - \frac{\mu}{3}\nabla(\mathrm{div}\mathbf{u}) + \nabla p, \qquad \mathbf{g} := 0.$$

The experiments want to answer the question, whether the same locking behavior from the incompressible Stokes problem can be observed in the compressible setting for the unmodified scheme.

$\gamma = 1 \qquad \mu = 1 \qquad c = 1$

ndof	161	617	2297	9152	36326	143945	573386	2290184
$\|\mathbf{u} - \mathbf{u}_h^+\|_{L^2}$	$1.64 \cdot 10^{-3}$	$5.17 \cdot 10^{-4}$	$1.72 \cdot 10^{-4}$	$6.81 \cdot 10^{-5}$	$3.14 \cdot 10^{-5}$	$1.52 \cdot 10^{-5}$	$7.39 \cdot 10^{-6}$	$3.65 \cdot 10^{-6}$
rate		1.67	1.58	1.34	1.12	1.05	1.04	1.02
$\|\mathbf{u} - \mathbf{u}_h\|_{L^2}$	$2.22 \cdot 10^{-3}$	$6.87 \cdot 10^{-4}$	$2.20 \cdot 10^{-4}$	$7.99 \cdot 10^{-5}$	$3.41 \cdot 10^{-5}$	$1.58 \cdot 10^{-5}$	$7.53 \cdot 10^{-6}$	$3.68 \cdot 10^{-6}$
rate		1.69	1.64	1.46	1.23	1.11	1.07	1.03
$\|\nabla(\mathbf{u} - \mathbf{u}_h^+)\|_{L^2}$	$5.31 \cdot 10^{-2}$	$2.72 \cdot 10^{-2}$	$1.36 \cdot 10^{-2}$	$7.24 \cdot 10^{-3}$	$3.85 \cdot 10^{-3}$	$2.08 \cdot 10^{-3}$	$1.13 \cdot 10^{-3}$	$6.03 \cdot 10^{-4}$
rate		0.96	1.00	0.91	0.91	0.89	0.89	0.90
$\|\nabla(\mathbf{u} - \mathbf{u}_h)\|_{L^2}$	$5.09 \cdot 10^{-2}$	$2.68 \cdot 10^{-2}$	$1.33 \cdot 10^{-2}$	$6.92 \cdot 10^{-3}$	$3.68 \cdot 10^{-3}$	$2.00 \cdot 10^{-3}$	$1.09 \cdot 10^{-3}$	$5.85 \cdot 10^{-4}$
rate		0.92	1.01	0.94	0.91	0.88	0.88	0.90
$\|\varrho - \varrho_h^+\|_{L^2}$	$6.33 \cdot 10^{-2}$	$2.97 \cdot 10^{-2}$	$1.50 \cdot 10^{-2}$	$7.68 \cdot 10^{-3}$	$3.88 \cdot 10^{-3}$	$1.99 \cdot 10^{-3}$	$1.02 \cdot 10^{-3}$	$5.25 \cdot 10^{-4}$
rate		1.09	0.99	0.96	0.98	0.96	0.96	0.96
$\|\varrho - \varrho_h\|_{L^2}$	$6.62 \cdot 10^{-2}$	$3.05 \cdot 10^{-2}$	$1.53 \cdot 10^{-2}$	$7.80 \cdot 10^{-3}$	$3.93 \cdot 10^{-3}$	$2.01 \cdot 10^{-3}$	$1.03 \cdot 10^{-3}$	$5.27 \cdot 10^{-4}$
rate		1.12	1.00	0.97	0.99	0.97	0.97	0.96

$\gamma = 1 \qquad \mu = 1 \qquad c = 100$

ndof	161	617	2297	9152	36326	143945	573386	2290184
$\|\mathbf{u} - \mathbf{u}_h^+\|_{L^2}$	$1.43 \cdot 10^{-3}$	$3.37 \cdot 10^{-4}$	$9.38 \cdot 10^{-5}$	$2.02 \cdot 10^{-5}$	$5.21 \cdot 10^{-6}$	$1.27 \cdot 10^{-6}$	$3.20 \cdot 10^{-7}$	$8.55 \cdot 10^{-8}$
rate		2.08	1.84	2.22	1.95	2.03	1.99	1.90
$\|\mathbf{u} - \mathbf{u}_h\|_{L^2}$	$1.94 \cdot 10^{-3}$	$4.82 \cdot 10^{-4}$	$1.37 \cdot 10^{-4}$	$3.14 \cdot 10^{-5}$	$8.06 \cdot 10^{-6}$	$1.99 \cdot 10^{-6}$	$5.07 \cdot 10^{-7}$	$1.35 \cdot 10^{-7}$
rate		2.01	1.81	2.13	1.96	2.02	1.97	1.91
$\|\nabla(\mathbf{u} - \mathbf{u}_h^+)\|_{L^2}$	$4.27 \cdot 10^{-2}$	$2.12 \cdot 10^{-2}$	$1.08 \cdot 10^{-2}$	$5.33 \cdot 10^{-3}$	$2.67 \cdot 10^{-3}$	$1.33 \cdot 10^{-3}$	$6.64 \cdot 10^{-4}$	$3.34 \cdot 10^{-4}$
rate		1.01	0.97	1.03	1.00	1.00	1.00	0.99
$\|\nabla(\mathbf{u} - \mathbf{u}_h)\|_{L^2}$	$4.23 \cdot 10^{-2}$	$2.15 \cdot 10^{-2}$	$1.10 \cdot 10^{-2}$	$5.36 \cdot 10^{-3}$	$2.68 \cdot 10^{-3}$	$1.34 \cdot 10^{-3}$	$6.70 \cdot 10^{-4}$	$3.36 \cdot 10^{-4}$
rate		0.97	0.96	1.04	1.00	1.00	1.00	0.99
$\|\varrho - \varrho_h^+\|_{L^2}$	$6.28 \cdot 10^{-4}$	$2.96 \cdot 10^{-4}$	$1.49 \cdot 10^{-4}$	$7.55 \cdot 10^{-5}$	$3.77 \cdot 10^{-5}$	$1.89 \cdot 10^{-5}$	$9.44 \cdot 10^{-6}$	$4.72 \cdot 10^{-6}$
rate		1.09	0.99	0.98	1.00	0.99	1.00	1.00
$\|\varrho - \varrho_h\|_{L^2}$	$6.53 \cdot 10^{-4}$	$3.02 \cdot 10^{-4}$	$1.51 \cdot 10^{-4}$	$7.68 \cdot 10^{-5}$	$3.82 \cdot 10^{-5}$	$1.92 \cdot 10^{-5}$	$9.56 \cdot 10^{-6}$	$4.78 \cdot 10^{-6}$
rate		1.11	1.00	0.98	1.01	1.00	1.00	1.00

TABLE 7.1. Errors of the modified gradient-robust scheme $(\mathbf{u}_h^+, \rho^+)$ and the classical scheme $(\mathbf{u}_h, \rho)$ for $\gamma = 1$ and $\mu = 1$.

$\gamma = 1 \qquad \mu = 10^{-2} \qquad c = 1$

ndof	161	617	2297	9152	36326	143945	573386	2290184
$\|\mathbf{u} - \mathbf{u}_h^+\|_{L^2}$	$1.65 \cdot 10^{-3}$	$5.22 \cdot 10^{-4}$	$1.80 \cdot 10^{-4}$	$7.31 \cdot 10^{-5}$	$3.47 \cdot 10^{-5}$	$1.71 \cdot 10^{-5}$	$8.47 \cdot 10^{-6}$	$4.19 \cdot 10^{-6}$
rate		1.66	1.54	1.30	1.08	1.02	1.02	1.02
$\|\mathbf{u} - \mathbf{u}_h\|_{L^2}$	$8.44 \cdot 10^{-2}$	$2.30 \cdot 10^{-2}$	$6.05 \cdot 10^{-3}$	$1.57 \cdot 10^{-3}$	$4.10 \cdot 10^{-4}$	$1.08 \cdot 10^{-4}$	$2.98 \cdot 10^{-5}$	$9.04 \cdot 10^{-6}$
rate		1.88	1.92	1.95	1.94	1.92	1.86	1.72
$\|\nabla(\mathbf{u} - \mathbf{u}_h^+)\|_{L^2}$	$5.30 \cdot 10^{-2}$	$2.72 \cdot 10^{-2}$	$1.41 \cdot 10^{-2}$	$8.09 \cdot 10^{-3}$	$5.75 \cdot 10^{-3}$	$4.97 \cdot 10^{-3}$	$4.67 \cdot 10^{-3}$	$4.34 \cdot 10^{-3}$
rate		0.96	0.95	0.80	0.49	0.21	0.09	0.11
$\|\nabla(\mathbf{u} - \mathbf{u}_h)\|_{L^2}$	$1.25 \cdot 10^{0}$	$7.31 \cdot 10^{-1}$	$3.87 \cdot 10^{-1}$	$2.02 \cdot 10^{-1}$	$1.02 \cdot 10^{-1}$	$5.18 \cdot 10^{-2}$	$2.61 \cdot 10^{-2}$	$1.35 \cdot 10^{-2}$
rate		0.77	0.92	0.94	0.99	0.97	0.99	0.94
$\|\varrho - \varrho_h^+\|_{L^2}$	$6.25 \cdot 10^{-2}$	$2.94 \cdot 10^{-2}$	$1.48 \cdot 10^{-2}$	$7.52 \cdot 10^{-3}$	$3.75 \cdot 10^{-3}$	$1.88 \cdot 10^{-3}$	$9.40 \cdot 10^{-4}$	$4.70 \cdot 10^{-4}$
rate		1.09	0.99	0.98	1.00	0.99	1.00	1.00
$\|\varrho - \varrho_h\|_{L^2}$	$6.45 \cdot 10^{-2}$	$3.01 \cdot 10^{-2}$	$1.50 \cdot 10^{-2}$	$7.64 \cdot 10^{-3}$	$3.80 \cdot 10^{-3}$	$1.91 \cdot 10^{-3}$	$9.53 \cdot 10^{-4}$	$4.77 \cdot 10^{-4}$
rate		1.10	1.00	0.98	1.01	0.99	1.00	1.00

$\gamma = 1 \qquad \mu = 10^{-2} \qquad c = 100$

ndof	161	617	2297	9152	36326	143945	573386	2290184
$\|\mathbf{u} - \mathbf{u}_h^+\|_{L^2}$	$1.43 \cdot 10^{-3}$	$3.37 \cdot 10^{-4}$	$9.38 \cdot 10^{-5}$	$2.02 \cdot 10^{-5}$	$5.22 \cdot 10^{-6}$	$1.27 \cdot 10^{-6}$	$3.20 \cdot 10^{-7}$	$8.56 \cdot 10^{-8}$
rate		2.08	1.84	2.22	1.95	2.03	1.99	1.90
$\|\mathbf{u} - \mathbf{u}_h\|_{L^2}$	$8.28 \cdot 10^{-2}$	$2.26 \cdot 10^{-2}$	$5.94 \cdot 10^{-3}$	$1.53 \cdot 10^{-3}$	$3.91 \cdot 10^{-4}$	$9.87 \cdot 10^{-5}$	$2.47 \cdot 10^{-5}$	$6.21 \cdot 10^{-6}$
rate		1.88	1.92	1.96	1.97	1.98	2.00	1.99
$\|\nabla(\mathbf{u} - \mathbf{u}_h^+)\|_{L^2}$	$4.27 \cdot 10^{-2}$	$2.12 \cdot 10^{-2}$	$1.08 \cdot 10^{-2}$	$5.33 \cdot 10^{-3}$	$2.67 \cdot 10^{-3}$	$1.33 \cdot 10^{-3}$	$6.64 \cdot 10^{-4}$	$3.34 \cdot 10^{-4}$
rate		1.01	0.97	1.03	1.00	1.00	1.00	0.99
$\|\nabla(\mathbf{u} - \mathbf{u}_h)\|_{L^2}$	$1.25 \cdot 10^{0}$	$7.31 \cdot 10^{-1}$	$3.88 \cdot 10^{-1}$	$2.03 \cdot 10^{-1}$	$1.02 \cdot 10^{-1}$	$5.17 \cdot 10^{-2}$	$2.58 \cdot 10^{-2}$	$1.29 \cdot 10^{-2}$
rate		0.77	0.92	0.94	1.00	0.97	1.00	0.99
$\|\varrho - \varrho_h^+\|_{L^2}$	$6.25 \cdot 10^{-4}$	$2.94 \cdot 10^{-4}$	$1.48 \cdot 10^{-4}$	$7.52 \cdot 10^{-5}$	$3.75 \cdot 10^{-5}$	$1.88 \cdot 10^{-5}$	$9.40 \cdot 10^{-6}$	$4.70 \cdot 10^{-6}$
rate		1.09	0.99	0.98	1.00	0.99	1.00	1.00
$\|\varrho - \varrho_h\|_{L^2}$	$6.45 \cdot 10^{-4}$	$3.01 \cdot 10^{-4}$	$1.50 \cdot 10^{-4}$	$7.64 \cdot 10^{-5}$	$3.80 \cdot 10^{-5}$	$1.91 \cdot 10^{-5}$	$9.53 \cdot 10^{-6}$	$4.76 \cdot 10^{-6}$
rate		1.10	1.00	0.98	1.01	0.99	1.00	1.00

TABLE 7.2. Errors of the modified gradient-robust scheme $(\mathbf{u}_h^+, \rho^+)$ and the classical scheme $(\mathbf{u}_h, \rho)$ for $\gamma = 1$ and $\mu = 10^{-2}$.

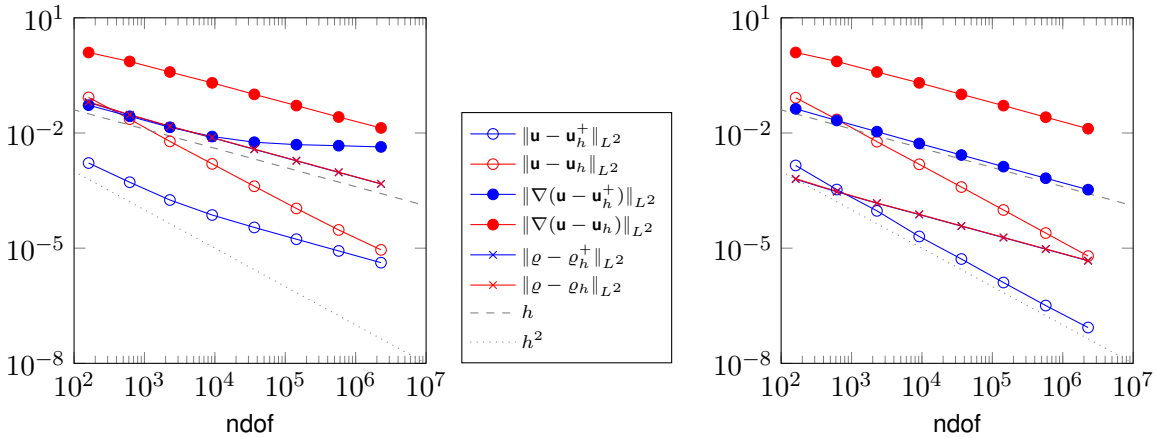

FIGURE 7.3. Convergence histories for the modified method and classical method for $\gamma = 1$ and $\mu = 10^{-2}$ and $c = 1$ (left), $c = 100$ (right).

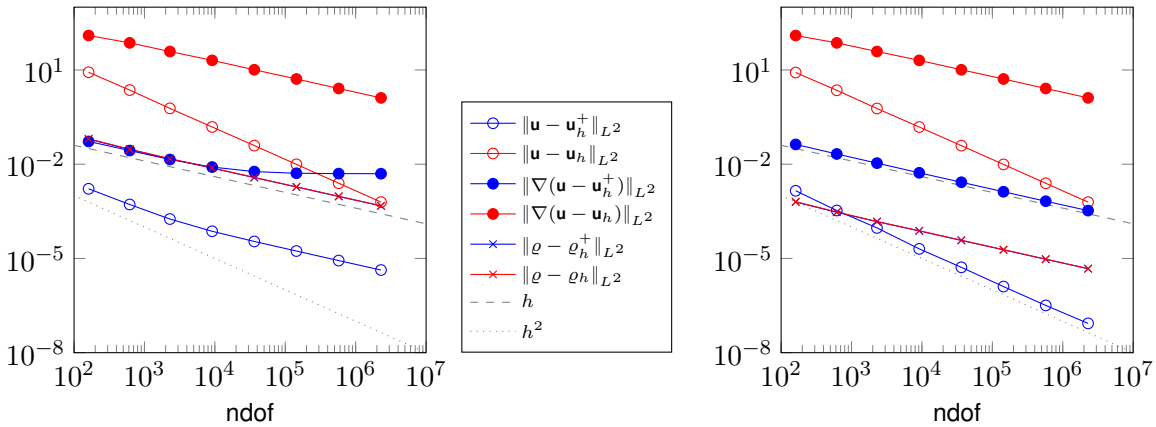

FIGURE 7.4. Convergence histories for the modified method and classical method for $\gamma = 1$ and $\mu = 10^{-4}$ and $c = 1$ (left), $c = 100$ (right).

Table 7.1 displays the calculated errors for $\gamma = 1$, $\mu = 1$ and $c = 1$ or $c = 100$. Figure 7.1 shows the corresponding convergence histories and convergence rates. In the compressible case ($c = 1$), the classical and the modified method both give very similar results. Interestingly, the convergence rate of the L^2 velocity error drops asymptotically to the suboptimal rate 1 with respect to the mesh size $h \approx \text{ndof}^{-1/2}$. In the nearly incompressible case ($c = 100$) however, all convergence rates are optimal as one would expect from the linear incompressible Stokes problem. A similar trend can be observed for $\mu = 10^{-1}$ in Figure 7.2. Here, also the convergence rate of $\|\nabla(\mathbf{u} - \mathbf{u}_h^+)\|_{L^2}$ is asymptotically significantly suboptimal (but above 0.33). The unmodified method begins to show a similar behavior a bit later, possibly due to the pressure-dependent consistency error that dominates in the beginning but is reduced with optimal order.

To study the locking behavior, Figures 7.3 and 7.4 show the convergence histories of the calculated errors for $\mu = 10^{-2}$ and $\mu = 10^{-4}$, respectively, with $c \in \{1, 100\}$. The results for $\mu = 10^{-2}$ are also printed in Table 7.2. The first important observation is that the classical scheme $(\mathbf{u}_h, \varrho_h)$ indeed shows locking and produces errors that are several magnitudes larger than the errors of the modified scheme $(\mathbf{u}_h^+, \varrho_h^+)$. The factor on coarse meshes is approximately $1/\mu$ as expected by the theory. However, for the case $c = 1$ on finer meshes the velocity error convergence rates of the modified scheme deteriorates earlier than the ones of the classical scheme. Nevertheless, the error of the modified scheme on the finest mesh is still

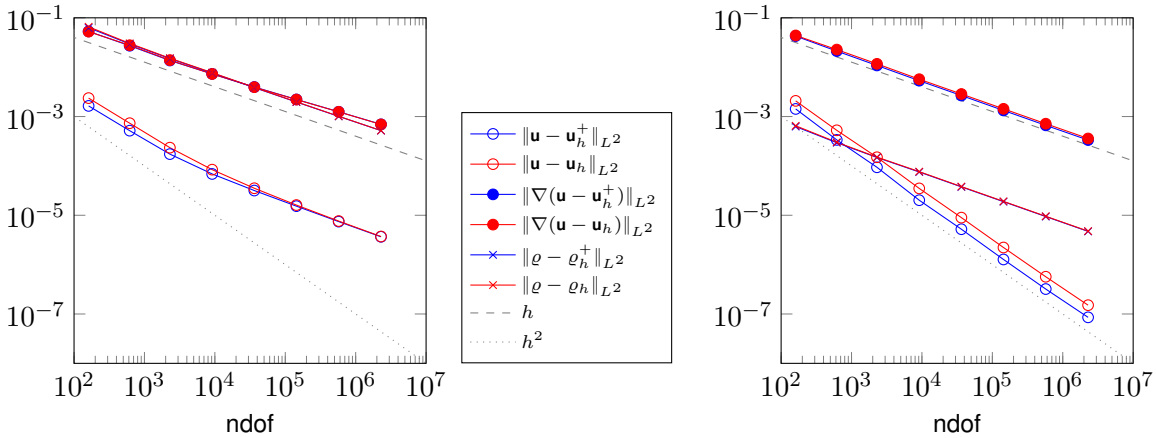

FIGURE 7.5. Convergence histories for the modified method and classical method for $\gamma = 1.4$ and $\mu = 1$ and $c = 1$ (left), $c = 100$ (right).

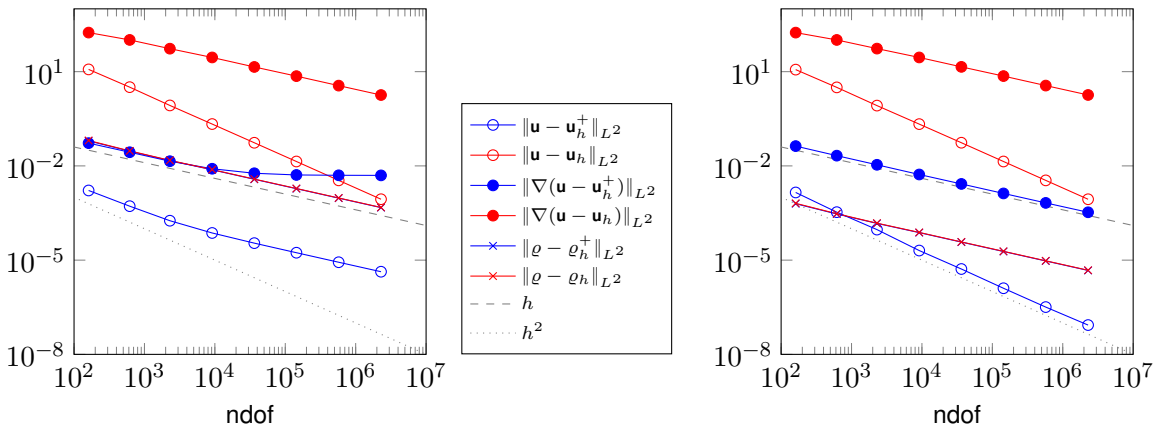

FIGURE 7.6. Convergence histories for the modified method and classical method for $\gamma = 1.4$ and $\mu = 10^{-4}$ and $c = 1$ (left), $c = 100$ (right).

much smaller than the error of the classical scheme and it is expected that the classical scheme also shows suboptimal convergence once it arrives at the same error level similar to the case $\mu = 10^{-1}$. Note, that for the nearly incompressible case $c = 100$, all convergence rates are again optimal and the gap between $(\mathbf{u}_h, \varrho_h)$ and $(\mathbf{u}_h^+, \varrho_h^+)$ due to the locking is as large as in the other case and, more importantly, persists even on the finest mesh. This is the known locking behavior from the incompressible Stokes setting.

Figures 7.5 and 7.6 show some results for $\gamma = 1.4$ and $\mu = 1$ and $\mu = 10^{-4}$, respectively. The convergence histories of the error are very similar to the isothermal case $\gamma = 1$, quantitatively and qualitatively concerning the locking behavior and the suboptimal convergence rates for small mu and small c.

Remark 7.1. The experiments convey that the suboptimal convergence rates on finer meshes have to do with the discretization of the continuity equation and the compressibility of the fluid. If $\operatorname{div}(\mathbf{u}) \neq 0$, the upwind discretization introduces an error that does not allow any guaranteed convergence rates for the error of the velocity gradient. However, due to an Aubin–Nitsche argument, the linear convergence of the L^2 error of the velocity is still granted and was observed in all experiments.

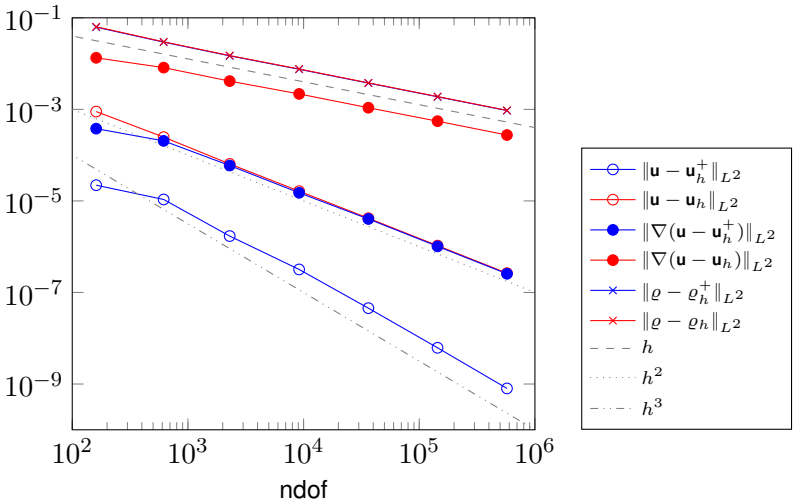

FIGURE 7.7. Convergence histories for the modified gradient-robust scheme $(\mathbf{u}_h^+, \rho^+)$ and the classical scheme $(\mathbf{u}_h, \rho)$ for $c = 1$ and $\gamma = 1$ on unstructured meshes in Section 7.2.

NDOF	$\|\mathbf{u}-\mathbf{u}_h^+\|_{L^2}$	$\|\nabla(\mathbf{u}-\mathbf{u}_h^+)\|_{L^2}$	$\|\rho-\rho_h^+\|_{L^2}$	$\|\mathbf{u}-\mathbf{u}_h\|_{L^2}$	$\|\nabla(\mathbf{u}-\mathbf{u}_h)\|_{L^2}$	$\|\rho-\rho_h\|_{L^2}$
161	$2.2053 \cdot 10^{-5}$	$3.7917 \cdot 10^{-4}$	$6.2504 \cdot 10^{-2}$	$8.9555 \cdot 10^{-4}$	$1.3380 \cdot 10^{-2}$	$6.3856 \cdot 10^{-2}$
617	$1.0756 \cdot 10^{-5}$	$2.0488 \cdot 10^{-4}$	$2.9448 \cdot 10^{-2}$	$2.4783 \cdot 10^{-4}$	$8.1682 \cdot 10^{-3}$	$2.9954 \cdot 10^{-2}$
2297	$1.7069 \cdot 10^{-6}$	$5.9351 \cdot 10^{-5}$	$1.4785 \cdot 10^{-2}$	$6.4319 \cdot 10^{-5}$	$4.1439 \cdot 10^{-3}$	$1.5004 \cdot 10^{-2}$
9152	$3.1842 \cdot 10^{-7}$	$1.5058 \cdot 10^{-5}$	$7.5158 \cdot 10^{-3}$	$1.6349 \cdot 10^{-5}$	$2.1742 \cdot 10^{-3}$	$7.6230 \cdot 10^{-3}$
36326	$4.5439 \cdot 10^{-8}$	$4.0321 \cdot 10^{-6}$	$3.7484 \cdot 10^{-3}$	$4.1878 \cdot 10^{-6}$	$1.0860 \cdot 10^{-3}$	$3.7978 \cdot 10^{-3}$
143945	$6.1867 \cdot 10^{-9}$	$1.0165 \cdot 10^{-6}$	$1.8826 \cdot 10^{-3}$	$1.0555 \cdot 10^{-6}$	$5.5162 \cdot 10^{-4}$	$1.9064 \cdot 10^{-3}$
573386	$8.0138 \cdot 10^{-10}$	$2.5576 \cdot 10^{-7}$	$9.4006 \cdot 10^{-4}$	$2.6432 \cdot 10^{-7}$	$2.7507 \cdot 10^{-4}$	$9.5180 \cdot 10^{-4}$

TABLE 7.3. Errors of the modified gradient-robust scheme $(\mathbf{u}_h^+, \rho^+)$ and the classical scheme $(\mathbf{u}_h, \rho)$ for $c = 1$ and $\gamma = 1$ on unstructured meshes in Section 7.2.

7.2. **Incompressibility limit.** This example on the unit square $\Omega := (0,1)^2$ examines the exact solution

$$\mathbf{u} := 0, \qquad p = \varphi(\varrho) := c\varrho^\gamma, \qquad \varrho(x,y) := 1.0 + (y - 1/2)/c$$

for $\mu = 1$ and $\lambda = -2/3$. These functions satisfy the compressible Stokes system with the right-hand side functions

$$\mathbf{f} = 0 \quad \text{and} \quad \mathbf{g} = \gamma\varrho^{\gamma-2}\begin{pmatrix}0\\1\end{pmatrix}.$$

Note that the constant c behaves like the squared inverse of the Mach number and for $c \to \infty$ the compressible System converges to the incompressible Stokes system, i.e., the density converges to the constant function 1.0.

Table 7.3 compares the error of the solutions of the classical Bernardi–Raugel scheme ($\Pi = 1$) with the gradient-robust scheme (where Π is chosen as described above) for $c = 1$ and $\gamma = 1$. One can clearly see, that the velocity errors of the novel scheme are improved by about two orders of magnitudes and also show some superconvergence behavior as depicted in Figure 7.7, i.e., the L^2 velocity gradient error converges quadratically.

For the following discussion we fix two meshes, one is the unstructured mesh with 489 triangles used before and the other one is a structured mesh with 450 triangles, see Figure 7.8. Table 7.4 compares the velocity

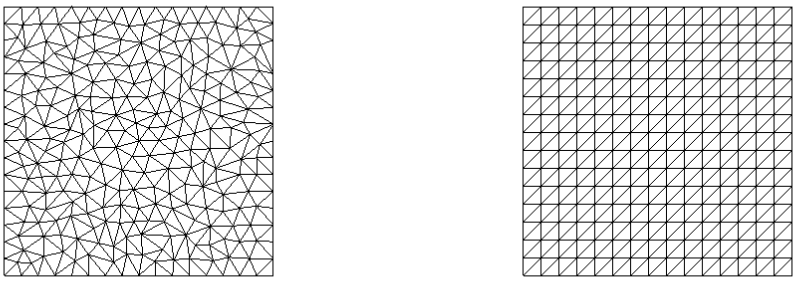

FIGURE 7.8. Unstructured grid (G1) with 489 triangles (left) and structured mesh grid (G2) with 450 triangles (right) used in the examples from Section 7.2 with varying c.

c	$\|\nabla(\mathbf{u}-\mathbf{u}_h^+)\|_{L^2}$ (G1)	$\|\nabla(\mathbf{u}-\mathbf{u}_h)\|_{L^2}$ (G1)	$\|\nabla(\mathbf{u}-\mathbf{u}_h^+)\|_{L^2}$ (G2)	$\|\nabla(\mathbf{u}-\mathbf{u}_h)\|_{L^2}$ (G2)
1	$1.0219 \cdot 10^{-4}$	$8.6144 \cdot 10^{-3}$	$2.9412 \cdot 10^{-13}$	$1.3473 \cdot 10^{-2}$
10	$1.0236 \cdot 10^{-5}$	$8.2877 \cdot 10^{-3}$	$9.3288 \cdot 10^{-14}$	$1.3001 \cdot 10^{-2}$
100	$1.0236 \cdot 10^{-6}$	$8.2862 \cdot 10^{-3}$	$7.8271 \cdot 10^{-13}$	$1.2996 \cdot 10^{-2}$
1000	$1.0236 \cdot 10^{-7}$	$8.2864 \cdot 10^{-3}$	$8.3695 \cdot 10^{-12}$	$1.2996 \cdot 10^{-2}$
10000	$1.0236 \cdot 10^{-8}$	$8.2864 \cdot 10^{-3}$	$6.8088 \cdot 10^{-11}$	$1.2996 \cdot 10^{-2}$

TABLE 7.4. Errors $\|\nabla(\mathbf{u}-\mathbf{u}_h)\|_{L^2}$ of the classical and gradient-robust scheme computed on the two fixed grids from Figure 7.8 for $\gamma = 2$ and different choices of c in the example from Section 7.2.

c	$\|\nabla(\mathbf{u}-\mathbf{u}_h^+)\|_{L^2}$ (G1)	$\|\nabla(\mathbf{u}-\mathbf{u}_h)\|_{L^2}$ (G1)	$\|\nabla(\mathbf{u}-\mathbf{u}_h^+)\|_{L^2}$ (G2)	$\|\nabla(\mathbf{u}-\mathbf{u}_h)\|_{L^2}$ (G2)
1	$7.6708 \cdot 10^{-5}$	$5.7730 \cdot 10^{-3}$	$1.8710 \cdot 10^{-6}$	$9.0664 \cdot 10^{-3}$
10	$7.1800 \cdot 10^{-6}$	$5.7996 \cdot 10^{-3}$	$1.3956 \cdot 10^{-8}$	$9.0971 \cdot 10^{-3}$
100	$7.1664 \cdot 10^{-7}$	$5.8004 \cdot 10^{-3}$	$1.3977 \cdot 10^{-10}$	$9.0974 \cdot 10^{-3}$
1000	$7.1653 \cdot 10^{-8}$	$5.8005 \cdot 10^{-3}$	$1.4607 \cdot 10^{-11}$	$9.0974 \cdot 10^{-3}$
10000	$7.1665 \cdot 10^{-9}$	$5.8005 \cdot 10^{-3}$	$1.0489 \cdot 10^{-10}$	$9.0974 \cdot 10^{-3}$

TABLE 7.5. Errors $\|\nabla(\mathbf{u}-\mathbf{u}_h)\|_{L^2}$ of the classical and gradient-robust scheme computed on the two fixed grids from Figure 7.8 for $\gamma = 1.4$ and different choices of c in the example from Section 7.2.

c	$\|\nabla(\mathbf{u}-\mathbf{u}_h^+)\|_{L^2}$ (G1)	$\|\nabla(\mathbf{u}-\mathbf{u}_h)\|_{L^2}$ (G1)	$\|\nabla(\mathbf{u}-\mathbf{u}_h^+)\|_{L^2}$ (G2)	$\|\nabla(\mathbf{u}-\mathbf{u}_h)\|_{L^2}$ (G2)
1	$5.9299 \cdot 10^{-5}$	$4.1439 \cdot 10^{-3}$	$2.5549 \cdot 10^{-6}$	$6.4992 \cdot 10^{-3}$
10	$5.1372 \cdot 10^{-6}$	$4.1432 \cdot 10^{-3}$	$1.6635 \cdot 10^{-8}$	$6.4981 \cdot 10^{-3}$
100	$5.1194 \cdot 10^{-7}$	$4.1432 \cdot 10^{-3}$	$1.6628 \cdot 10^{-10}$	$6.4982 \cdot 10^{-3}$
1000	$5.1181 \cdot 10^{-8}$	$4.1432 \cdot 10^{-3}$	$1.5016 \cdot 10^{-11}$	$6.4982 \cdot 10^{-3}$
10000	$5.1275 \cdot 10^{-9}$	$4.1432 \cdot 10^{-3}$	$1.6017 \cdot 10^{-10}$	$6.4982 \cdot 10^{-3}$

TABLE 7.6. Errors $\|\nabla(\mathbf{u}-\mathbf{u}_h)\|_{L^2}$ of the classical and gradient-robust scheme computed on the two fixed grids from Figure 7.8 for $\gamma = 1$ and different choices of c in the example from Section 7.2.

error on these two meshes for different choices of c. There are two interesting observations. First, the velocity errors of the gradient-robust scheme converge to zero for $c \to \infty$, while the errors of the classical scheme stagnates. Second, the velocity of the gradient-robust scheme is exact on structured meshes for every c, while the classical scheme is not.

NDOF	$\|\mathbf{u}-\mathbf{u}_h^+\|_{L^2}$	$\|\nabla(\mathbf{u}-\mathbf{u}_h^+)\|_{L^2}$	$\|\rho-\rho_h^+\|_{L^2}$	$\|\mathbf{u}-\mathbf{u}_h\|_{L^2}$	$\|\nabla(\mathbf{u}-\mathbf{u}_h)\|_{L^2}$	$\|\rho-\rho_h\|_{L^2}$
161	$7.5658\cdot10^{-14}$	$1.5323\cdot10^{-12}$	$6.2501\cdot10^{-2}$	$1.2578\cdot10^{-3}$	$1.8979\cdot10^{-2}$	$6.4038\cdot10^{-2}$
617	$1.3088\cdot10^{-16}$	$5.2005\cdot10^{-15}$	$2.9447\cdot10^{-2}$	$3.4992\cdot10^{-4}$	$1.1486\cdot10^{-2}$	$2.9970\cdot10^{-2}$
2297	$8.0186\cdot10^{-17}$	$5.5293\cdot10^{-15}$	$1.4785\cdot10^{-2}$	$9.0438\cdot10^{-5}$	$5.7741\cdot10^{-3}$	$1.5006\cdot10^{-2}$
9152	$8.2615\cdot10^{-17}$	$1.1330\cdot10^{-14}$	$7.5158\cdot10^{-3}$	$2.2972\cdot10^{-5}$	$3.0452\cdot10^{-3}$	$7.6233\cdot10^{-3}$
36326	$8.4962\cdot10^{-17}$	$2.2574\cdot10^{-14}$	$3.7484\cdot10^{-3}$	$5.8270\cdot10^{-6}$	$1.5099\cdot10^{-3}$	$3.7979\cdot10^{-3}$

TABLE 7.7. Errors of the modified gradient-robust scheme $(\mathbf{u}_h^+,\rho^+)$ and the classical scheme $(\mathbf{u}_h,\rho)$ for $c=1$ and $\gamma=1.4$ on unstructured grids with right-hand sides (7.1).

NDOF	$\|\mathbf{u}-\mathbf{u}_h^+\|_{L^2}$	$\|\nabla(\mathbf{u}-\mathbf{u}_h^+)\|_{L^2}$	$\|\rho-\rho_h^+\|_{L^2}$	$\|\mathbf{u}-\mathbf{u}_h\|_{L^2}$	$\|\nabla(\mathbf{u}-\mathbf{u}_h)\|_{L^2}$	$\|\rho-\rho_h\|_{L^2}$
161	$6.9935\cdot10^{-17}$	$1.2646\cdot10^{-15}$	$6.2500\cdot10^{-2}$	$8.9980\cdot10^{-4}$	$1.3467\cdot10^{-2}$	$6.3956\cdot10^{-2}$
617	$6.6351\cdot10^{-17}$	$2.4263\cdot10^{-15}$	$2.9446\cdot10^{-2}$	$2.4970\cdot10^{-4}$	$8.1662\cdot10^{-3}$	$2.9961\cdot10^{-2}$
2297	$7.3217\cdot10^{-17}$	$5.0692\cdot10^{-15}$	$1.4785\cdot10^{-2}$	$6.4488\cdot10^{-5}$	$4.1437\cdot10^{-3}$	$1.5005\cdot10^{-2}$
9152	$7.3142\cdot10^{-17}$	$1.0417\cdot10^{-14}$	$7.5158\cdot10^{-3}$	$1.6409\cdot10^{-5}$	$2.1739\cdot10^{-3}$	$7.6232\cdot10^{-3}$
36326	$7.6058\cdot10^{-17}$	$2.0748\cdot10^{-14}$	$3.7484\cdot10^{-3}$	$4.1974\cdot10^{-6}$	$1.0860\cdot10^{-3}$	$3.7978\cdot10^{-3}$

TABLE 7.8. Errors of the modified gradient-robust scheme $(\mathbf{u}_h^+,\rho^+)$ and the classical scheme $(\mathbf{u}_h,\rho)$ for $c=1$ and $\gamma=1$ on unstructured grids with right-hand sides (7.1).

Table 7.5 and 7.6 repeat this experiment for $\gamma=1.4$ and $\gamma=1$, respectively. Here the results are similar as for the case with $\gamma=2$ in the sense that the gradient-robust scheme is more accurate than the classical scheme. However, the gradient-robust variant is not exact on structured meshes in these cases which most likely is due to the non-constant vector **g**.

7.3. **Well-balanced property.** We repeat the experiment from the previous section, but this time we assume the right-hand sides

$$\mathbf{f}=\gamma\varrho^{\gamma-1}\begin{pmatrix}0\\1\end{pmatrix}\quad\text{and}\quad\mathbf{g}=0. \tag{7.1}$$

Table 7.8 displays the errors for $\gamma=1$ and $c=1$ for the classical and the gradient-robust scheme. Surprisingly, the novel gradient-robust scheme computes the exact velocity even on unstructured meshes. Also note, that the gradient-robust scheme converges after the first iteration, since the initial value based on the (rescaled) discrete pressure from incompressible Stokes problem is already the correct discrete density. Table 7.7 leads to the same conclusions for for $\gamma=1.4$.

REFERENCES

1. Naveed Ahmed, Alexander Linke, and Christian Merdon, *On really locking-free mixed finite element methods for the transient incompressible Stokes equations*, SIAM J. Numer. Anal. **56** (2018), no. 1, 185–209. MR 3743746
2. ———, *Towards pressure-robust mixed methods for the incompressible Navier-Stokes equations*, Comput. Methods Appl. Math. **18** (2018), no. 3, 353–372. MR 3824769
3. N. Botta, R. Klein, S. Langenberg, and S. Lützenkirchen, *Well balanced finite volume methods for nearly hydrostatic flows*, J. Comput. Phys. **196** (2004), no. 2, 539–565. MR 2054350
4. F. Brezzi and M. Fortin, *Mixed and hybrid finite elements*, Springer Series in Computational Mathematics, vol. 15, Springer, 1991.
5. Bernardo Cockburn, Guido Kanschat, and Dominik Schötzau, *A note on discontinuous Galerkin divergence-free solutions of the Navier-Stokes equations*, J. Sci. Comput. **31** (2007), no. 1-2, 61–73. MR 2304270
6. C. J. Cotter and J. Thuburn, *A finite element exterior calculus framework for the rotating shallow-water equations*, J. Comput. Phys. **257** (2014), no. part B, 1506–1526. MR 3133448
7. Darren Engwirda, Maxwell Kelley, and John Marshall, *High-order accurate finite-volume formulations for the pressure gradient force in layered ocean models*, Ocean Modelling **116** (2017), 1 – 15.

8. R. Eymard, T. Gallouët, R. Herbin, and J. C. Latché, *A convergent finite element-finite volume scheme for the compressible Stokes problem. II. The isentropic case*, Math. Comp. **79** (2010), no. 270, 649–675. MR 2600538
9. Eduard Feireisl, *Mathematical models of incompressible fluids as singular limits of complete fluid systems*, Milan J. Math. **78** (2010), no. 2, 523–560. MR 2781851
10. Eduard Feireisl, Mária Lukáčová-Medviďová, Šárka Nečasová, Antonín Novotný, and Bangwei She, *Asymptotic preserving error estimates for numerical solutions of compressible Navier-Stokes equations in the low Mach number regime*, Multiscale Model. Simul. **16** (2018), no. 1, 150–183. MR 3749377
11. J. Fuhrmann, C. Guhlke, A. Linke, C. Merdon, and R. Müller, *Models and numerical methods for electrolyte flows*, WIAS preprint No. 2525 (2018), 1–19.
12. T. Gallouët, R. Herbin, and J.-C. Latché, *A convergent finite element-finite volume scheme for the compressible Stokes problem. Part I: The isothermal case*, Math. Comp. **78** (2009), no. 267, 1333–1352.
13. A. Gassman, *Formulation of the lm's dynamical lower boundary condition*, COSMO Newsletter **4** (2004), 155–158.
14. Almut Gassmann, *A global hexagonal c-grid non-hydrostatic dynamical core (icon-iap) designed for energetic consistency*, Quarterly Journal of the Royal Meteorological Society **139** (2013), no. 670, 152–175.
15. Nicolas R. Gauger, Alexander Linke, and Philipp W. Schroeder, *On high-order pressure-robust space discretisations, their advantages for incompressible high Reynolds number generalised Beltrami flows and beyond*, SMAI Journal of Computational Mathematics **5** (2019), 89–129.
16. V. Girault and P.-A. Raviart, *Finite element methods for Navier-Stokes equations: Theory and algorithms*, volume 5 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1980.
17. Johnny Guzmán, Chi-Wang Shu, and Filánder A. Sequeira, $\mathrm{H}(\mathrm{div})$ *conforming and DG methods for incompressible Euler's equations*, IMA J. Numer. Anal. **37** (2017), no. 4, 1733–1771. MR 3712173
18. Volker John, Alexander Linke, Christian Merdon, Michael Neilan, and Leo G. Rebholz, *On the divergence constraint in mixed finite element methods for incompressible flows*, SIAM Rev. **59** (2017), no. 3, 492–544. MR 3683678
19. Philip L. Lederer, Christoph Lehrenfeld, and Joachim Schöberl, *Hybrid discontinuous Galerkin methods with relaxed $H(\mathrm{div})$-conformity for incompressible flows. Part I*, SIAM J. Numer. Anal. **56** (2018), no. 4, 2070–2094. MR 3826676
20. Philip L. Lederer, Alexander Linke, Christian Merdon, and Joachim Schöberl, *Divergence-free reconstruction operators for pressure-robust Stokes discretizations with continuous pressure finite elements*, SIAM J. Numer. Anal. **55** (2017), no. 3, 1291–1314. MR 3656505
21. Christoph Lehrenfeld and Joachim Schöberl, *High order exactly divergence-free hybrid discontinuous Galerkin methods for unsteady incompressible flows*, Comput. Methods Appl. Mech. Engrg. **307** (2016), 339–361. MR 3511719
22. Jin-Xi LI, Yi-Yuan LI, and Bin WANG, *Pressure gradient errors in a covariant method of implementing the ÏČ-coordinate: idealized experiments and geometric analysis*, Atmospheric and Oceanic Science Letters **9** (2016), no. 4, 270–276.
23. Shian-Jiann Lin, *A finite-volume integration method for computing pressure gradient force in general vertical coordinates*, Quarterly Journal of the Royal Meteorological Society **123** (1997), no. 542, 1749–1762.
24. A. Linke and C. Merdon, *On velocity errors due to irrotational forces in the Navier-Stokes momentum balance*, J. Comput. Phys. **313** (2016), 654–661. MR 3481034
25. ———, *Pressure-robustness and discrete Helmholtz projectors in mixed finite element methods for the incompressible Navier–Stokes equations*, Comput. Methods Appl. Mech. Engrg. **311** (2016), 304–326.
26. Alexander Linke, *On the role of the Helmholtz decomposition in mixed methods for incompressible flows and a new variational crime*, Comput. Methods Appl. Mech. Engrg. **268** (2014), 782–800. MR 3133522
27. Y. Mahrer, *An improved numerical approximation of the horizontal gradients in a terrain-following coordinate system*, Monthly Weather Review **112** (1984), no. 5, 918–922.
28. Andrea Natale and Colin J. Cotter, *A variational $H(\mathrm{div})$ finite-element discretization approach for perfect incompressible fluids*, IMA J. Numer. Anal. **38** (2018), no. 3, 1388–1419. MR 3861250
29. Sander Rhebergen and Garth N. Wells, *A hybridizable discontinuous Galerkin method for the Navier-Stokes equations with pointwise divergence-free velocity field*, J. Sci. Comput. **76** (2018), no. 3, 1484–1501. MR 3833698
30. Christoph Schärr, Daniel Leuenberger, Oliver Fuhrer, Daniel Lüthi, and Claude Girard, *A new terrain-following vertical coordinate formulation for atmospheric prediction models*, Monthly Weather Review **130** (2002), no. 10, 2459–2480.
31. Philipp W. Schroeder and Gert Lube, *Pressure-robust analysis of divergence-free and conforming FEM for evolutionary incompressible Navier-Stokes flows*, J. Numer. Math. **25** (2017), no. 4, 249–276. MR 3767413
32. ———, *Divergence-free $H(\mathrm{div})$-FEM for time-dependent incompressible flows with applications to high Reynolds number vortex dynamics*, J. Sci. Comput. **75** (2018), no. 2, 830–858. MR 3780790
33. L. R. Scott and M. Vogelius, *Conforming finite element methods for incompressible and nearly incompressible continua*, Large-scale computations in fluid mechanics, Part 2 (La Jolla, Calif., 1983), Lectures in Appl. Math., vol. 22, Amer. Math. Soc., Providence, RI, 1985, pp. 221–244. MR 818790
34. James Shaw and Hilary Weller, *Comparison of terrain-following and cut-cell grids using a nonhydrostatic model*, Monthly Weather Review **144** (2016), no. 6, 2085–2099.

35. J. Steppeler, J. Li, I. M. Navon, F. Fang, and Z. Xiao, *Medium range forecasts using cut-cells: a sensitivity study*, Meteorology and Atmospheric Physics (2019), 1–9.
36. Jürgen Steppeler, Heinz-Werner Bitzer, Maud Minotte, and Luca Bonaventura, *Nonhydrostatic atmospheric modeling using a z-coordinate representation*, Monthly Weather Review **130** (2002), no. 8, 2143–2149.
37. Hilary Weller and Ava Shahrokhi, *Curl-free pressure gradients over orography in a solution of the fully compressible Euler equations with implicit treatment of acoustic and gravity waves*, Monthly Weather Review **142** (2014), no. 12, 4439–4457.
38. Yulong Xing and Chi-Wang Shu, *High order well-balanced WENO scheme for the gas dynamics equations under gravitational fields*, J. Sci. Comput. **54** (2013), no. 2-3, 645–662. MR 3011375
39. Günther Zängl, *An improved method for computing horizontal diffusion in a sigma-coordinate model and its application to simulations over mountainous topography*, Monthly Weather Review **130** (2002), no. 5, 1423–1432.
40. _____, *Extending the numerical stability limit of terrain-following coordinate models over steep slopes*, Monthly Weather Review **140** (2012), no. 11, 3722–3733.
41. Günther Zängl, Leonhard Gantner, Gisela Hartjenstein, and Heike Noppel, *Numerical errors above steep topography: A model intercomparison*, Meteorologische Zeitschrift **13** (2004), no. 2, 69–76.
42. Shangyou Zhang, *A new family of stable mixed finite elements for the 3D Stokes equations*, Math. Comp. **74** (2005), no. 250, 543–554. MR 2114637