Published in: **Numer. Math.** (2009) 113(1):123-161, DOI 10.1007/s00211-009-0224-8.
The original publication is available at `www.springerlink.com`

Stefan Vater · **Rupert Klein**

Stability of a Cartesian Grid Projection Method for Zero Froude Number Shallow Water Flows

Received: 30 May 2007 / Revised: 24 November 2008 / Published online: 21 April 2009

Abstract In this paper a Godunov-type projection method for computing approximate solutions of the zero Froude number (incompressible) shallow water equations is presented. It is second-order accurate and locally conserves height (mass) and momentum. To enforce the underlying divergence constraint on the velocity field, the predicted numerical fluxes, computed with a standard second order method for hyperbolic conservation laws and applied to an auxiliary system, are corrected in two steps. First, a MAC-type projection adjusts the advective velocity divergence. In a second projection step, additional momentum flux corrections are computed to obtain new time level cell-centered velocities, which satisfy another discrete version of the divergence constraint.

The scheme features an exact and stable second projection. It is obtained by a Petrov-Galerkin finite element ansatz with piecewise bilinear trial functions for the unknown height and piecewise constant test functions. The key innovation compared to existing finite volume projection methods is a correction of the in-cell slopes of the momentum by the second projection. The stability of the projection is proved using a generalized theory for mixed finite elements. In order to do so, the validity of three different inf-sup conditions has to be shown.

The results of preliminary numerical test cases demonstrate the method's applicability. On fixed grids the accuracy is improved by a factor four compared to a previous version of the scheme.

Keywords incompressible flows · projection method · numerical stability · mixed finite elements · inf-sup-condition · shallow water equations

Mathematics Subject Classification (2000) 65M12 · 76M12 · 76M10 · 35L65

S. Vater (✉) · R. Klein
Numerische Mathematik/Scientific Computing, FB Mathematik und Informatik,
Institut für Mathematik, Freie Universität (FU) Berlin, Arnimallee 6, 14195 Berlin, Germany
E-mail: vater@math.fu-berlin.de

R. Klein
E-mail: rupert.klein@math.fu-berlin.de

1 Introduction

Starting with the fundamental work of Chorin [11] and Temam [35], the use of projection methods for the numerical solution of the incompressible flow equations has a long tradition (see e.g. [39,6,7,28,2] and references therein). In these methods, solutions are first advanced in time ignoring the solenoidal constraint of the velocity field. In a second step, the velocity field is corrected using a suitable approximation of the pressure to enforce compliance with the divergence constraint.

The stability of the projection step in exact projection methods for the incompressible Euler or shallow water equations has been an unsolved issue in the past. Difficulties arise in this context from a decoupling of the velocity and the pressure variables, which, in turn, is a consequence of using discrete gradient approximations with kernel dimension larger than one. Examples of such methods are given by [6], [7] and [28]. To resolve this problem, approximate projection methods were introduced in [3], which use the same discrete divergence and gradient operators as in exact projection methods, but a modified version of the discrete Laplacian. This approach results in velocity fields that satisfy the underlying divergence constraint only up to the order of accuracy of the gradient and divergence discretizations. In the present paper we propose an alternative approach that utilizes discretizations of the differential operators, which guarantee exact projections while avoiding the velocity-pressure decoupling. The resulting discretization of the pressure Poisson equation was first described by Süli [34] for the solution of Poisson's equation, and can be derived by a Petrov-Galerkin finite element ansatz with piecewise bilinear trial functions for the pressure and piecewise constant test functions.

The divergence constraint on the velocity field, which arises in the zero Mach number limit of the Euler equations (see [20,29,21], and also the review by [30]), leads to a saddle point problem, in which the velocity is coupled with the gradient of the pressure. The fundamental theory of (discretizations of) such problems goes back to Babuška [5] and Brezzi [10], who analyzed finite element schemes for elliptic partial differential equations with additional side constraints. This theory provides the so-called "inf-sup conditions" for existence and uniqueness of solutions and stable discretizations of such problems.

To the best of the authors knowledge, stability estimates of the Babuška-Brezzi-type have not been derived for projection methods applied to inviscid flow problems so far. This is different in the viscous case (cf. [18] for an overview). However, in contrast to the inviscid case in the incompressible Navier-Stokes equations the Laplacian of the velocity interacts with the pressure gradient, which leads to a saddle point problem of the Stokes type involving higher spatial derivatives compared to the inviscid case. Consequently, the stability proofs for methods solving the Navier-Stokes equations cannot be easily transferred.

The presented method is a non-incremental pressure-correction method for the incompressible (zero Froude number) shallow water equations. To represent advection of mass and momentum, the scheme relies on second order conservative finite volume Godunov-type methods in its predictor step. It is shown that the projection step, which corrects the cell-centered momentum to satisfy the underlying divergence constraint, is stable in the sense of mixed finite element methods. This is the main result of this paper and is summarized in Theorem 8. The discretization features both, a compact Poisson stencil, and an exact projection. The key to achieving both of these properties at the same time lies in the fact that we let part of the in-cell slopes, which are normally determined by standard slope limiting procedures, be assigned in the projection step.

After introducing the governing equations and the consequences of the zero Froude number limit in the remainder of the introduction, we describe the construction of the numerical method in Section 2. The stability of the projection step is investigated using the theory of generalized mixed finite elements in Section 3. To demonstrate the applicability of the scheme, some basic numerical test cases are presented in Section 4. The major conclusions of this work are reported in the last Section.

1.1 Governing Equations

The shallow water equations are a set of partial differential equations, which describe the depth averaged flow with velocity $\boldsymbol{v}(\boldsymbol{x},t)$ under a free surface $h(\boldsymbol{x},t)$. In their non-dimensional form and without any source terms (such as bottom topography) they are given by the two equations

$$\begin{aligned} \mathsf{Sr}\,\frac{\partial h}{\partial t} &+ \nabla\cdot(h\boldsymbol{v}) &&= 0 \\ \mathsf{Sr}\,\frac{\partial (h\boldsymbol{v})}{\partial t} &+ \nabla\cdot\left(h\boldsymbol{v}\circ\boldsymbol{v} + \frac{1}{2\,\mathsf{Fr}^2}h^2\boldsymbol{I}\right) &&= 0\;, \end{aligned} \tag{1}$$

which express conservation of height h and momentum $h\boldsymbol{v}$. The "$\circ$" represents the dyadic product of two vectors. Here, two dimensionless characteristic quantities have been introduced, namely

$$\mathsf{Sr} := \frac{\ell'_{\text{ref}}}{t'_{\text{ref}}\,v'_{\text{ref}}} \quad \text{and} \quad \mathsf{Fr} := \frac{v'_{\text{ref}}}{\sqrt{g'\,h'_{\text{ref}}}}\;,$$

which are known as the *Strouhal* and the *Froude number*, respectively. The first one describes the ratio between the advection timescale $\ell'_{\text{ref}}/v'_{\text{ref}}$ and the reference timescale t'_{ref}, whereas the latter gives the ratio between the reference velocity v'_{ref} and the gravity wave speed $\sqrt{g'\,h'_{\text{ref}}}$ (celerity). In the following, we are interested in a reference time scale equal to the advection time scale of the fluid, so that $t'_{\text{ref}} = \ell'_{\text{ref}}/v'_{\text{ref}}$ and the Strouhal number becomes one ($\mathsf{Sr} = 1$).

Since the discretization of the numerical method and the stability proof discussed in this work are restricted to axiparallel rectangles $\Omega \subset \mathbb{R}^2$, we restrict ourselves to such domains in the following. However, it is current research in the authors' group to extend such a projection method to more general domains using cut-cell techniques (cf. [26]). A strategy to extend such a scheme to three dimensions was given in [28] for a similar projection method applied to the variable density zero Mach number Euler equations.

The zero Froude number limit of (1) can be analyzed by an asymptotic analysis with a small parameter Fr [41]. This is similar to the zero Mach number limit of the Euler equations (cf. [20,21]), except that in the case of the Euler equations the divergence constraint arises from the energy equation, and not from the mass equation. The resulting limit equations are given by

$$\begin{aligned} h_t + \nabla \cdot (h\boldsymbol{v}) &= 0 \\ (h\boldsymbol{v})_t + \nabla \cdot (h\boldsymbol{v} \circ \boldsymbol{v}) + h\nabla h^{(2)} &= \mathbf{0} \, . \end{aligned} \tag{2}$$

An additional unknown, the second order height $h^{(2)}$, is introduced and the leading order height becomes only dependent on time with zero gradient $h = h_0(t)$. This system of equations is no longer hyperbolic, but of mixed elliptic-hyperbolic type.

Integrating the first equation of (2) over the domain Ω and applying the divergence theorem leads to

$$\frac{1}{h_0}\frac{dh_0}{dt} = -\frac{1}{|\Omega|}\int_{\partial\Omega} \boldsymbol{v} \cdot \boldsymbol{n}\, d\sigma \, . \tag{3}$$

Thus, either the change of the leading order height is given through the normal components of the velocity field on the boundary of Ω, or the prescription of h_0 implies an integral constraint on these normal velocity components in turn. Furthermore, the integration over an arbitrary volume $V \subset \Omega$ yields

$$\int_{\partial V} (h\boldsymbol{v}) \cdot \boldsymbol{n}\, d\sigma = -|V|\frac{dh_0}{dt} \, , \tag{4}$$

which implies an integral constraint on the velocity divergence in V. Thus, in terms of optimization problems $h^{(2)}$ can be viewed as a Lagrange multiplier, which ensures that the velocity field is in compliance with the divergence constraint (4).

In the case $h_0 \equiv 1$ system (2) is equivalent to the Euler equations for incompressible flow with constant density. This underlines the applicability of numerical methods developed in this article to incompressible flows.

2 The Numerical Method

The described method is a further development of the projection method presented in [28] for the incompressible Euler Equations, which we revisit here for the case of the zero Froude number shallow water equations. The main difference between the present scheme and that of [28] lies in the discretization of the projection step (see Subsection 2.2).

2.1 Construction of the scheme

Throughout this work we assume a Cartesian space discretization of the computational domain Ω. In this discretization, the volume of a cell C is denoted by $|C|$, and two neighboring cells are separated by an interface I with area $|I|$ (cf. Figure 1). $\mathcal{C}$ and $\mathcal{I}$ are defined as the collections of all cells and interfaces, respectively. We denote the set of all interfaces, which are part of the boundary of a cell C, by $\mathcal{I}_{\partial C} \subset \mathcal{I}$.

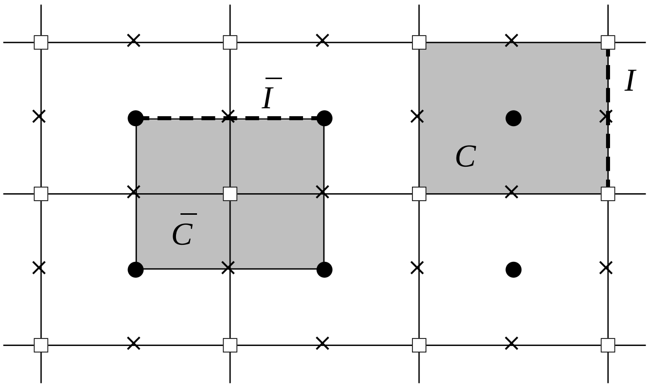

Fig. 1 Control volume C and interface I of the primary discretization and those ($\bar{C}$ and $\bar{I}$) of the dual discretization. Cell centers are denoted by *circles*, nodes by *squares* and midpoints of the interfaces by *crosses*.

For the construction of the method, a finite volume scheme in conservation form is considered, i.e.

$$\mathbf{U}_C^{n+1} = \mathbf{U}_C^n - \frac{\delta t}{|C|} \sum_{I \in \mathcal{I}_{\partial C}} |I| \, \mathbf{F}_I \quad . \tag{5}$$

In (5) $\mathbf{U}_C^n$ is a numerical approximation to the average of the exact solution $\mathbf{u}(\boldsymbol{x},t)$ of problem (2) over cell C at time t^n:

$$\mathbf{U}_C^n \approx \frac{1}{|C|} \int_C \mathbf{u}(\boldsymbol{x},t^n)\, d\boldsymbol{x} \quad , \quad \mathbf{u}(\boldsymbol{x},t) := \begin{pmatrix} h \\ h\boldsymbol{v} \end{pmatrix} .$$

The *numerical flux* $\mathbf{F}_I$ approximates the average of the flux function

$$\mathbf{f}(\mathbf{u}(\boldsymbol{x},t), h^{(2)}(\boldsymbol{x},t), \boldsymbol{n}(\boldsymbol{x})) := \begin{pmatrix} h(\boldsymbol{v}\cdot\boldsymbol{n}) \\ h\boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{n}) + h\,h^{(2)}\,\boldsymbol{n} \end{pmatrix}$$

of the zero Froude number shallow water equations. For these fluxes, the average is taken over one time step $[t^n, t^{n+1}]$, with $t^{n+1} := t^n + \delta t$, and over the interface I between two cells. The flux averages will be computed in three steps:

$$\mathbf{F}_I := \mathbf{F}_I^* + \mathbf{F}_I^{\mathsf{MAC}} + \mathbf{F}_I^{\mathsf{P2}} .$$

First, predictions of the advective fluxes $\mathbf{F}_I^*$ are computed by the numerical solution of the hyperbolic *auxiliary system*

$$\begin{aligned} h_t + \nabla\cdot(h\boldsymbol{v}) &= 0 \\ (h\boldsymbol{v})_t + \nabla\cdot\left((h\boldsymbol{v}\circ\boldsymbol{v}) + \frac{h^2}{2}\boldsymbol{I}\right) &= \mathbf{0} \, , \end{aligned} \tag{6}$$

which is the shallow water system with a rescaled Froude number. The computation of the numerical fluxes for these equations is done using an explicit high resolution upwind method for hyperbolic conservation laws (see [22] for an overview).

It should be stressed that the presented projection method is robust with respect to the particular choice of such an integration scheme. The authors have successfully implemented a version using centered-in-time approximations [40,37] with operator splitting techniques for the spatial directions [33], and one, which is based on a semi-discretization in space and Runge-Kutta time stepping [27,32]. The numerical results presented in Section 4 are based on the latter approach, and the net flux divergence at each interface $\mathbf{F}_I^*$ is given by the weighted sum of individual flux divergences computed in each Runge-Kutta stage. The stability of the numerical solution of the auxiliary system depends on a CFL time step restriction [12]. Since the eigenvalues (characteristic speeds) of this system do not depend on the Froude number, they are of order $\mathcal{O}(1)$ as $\mathsf{Fr} \to 0$, leading to $\delta t = \mathcal{O}(\delta x)$ on a regular discretization with grid spacing δx.

Then, a *MAC-type projection* [19] is applied, which corrects the advection velocity divergence by $\mathbf{F}_I^{\mathsf{MAC}}$ to be in compliance with the divergence constraint (4) applied to each grid cell. In a final *second projection* the non-convective components of the numerical fluxes, i.e., the pressure (height) contributions to the momentum fluxes, are corrected by $\mathbf{F}_I^{\mathsf{P2}}$, such that the new time level divergence of the cell-centered velocities satisfies (4) for another set of control volumes defined below. Furthermore, in the presented scheme this projection yields updates for the linear reconstructions of momentum in each grid cell.

To achieve second order accuracy in time for the flux components $\mathbf{F}_I^{\mathsf{MAC}}$ and $\mathbf{F}_I^{\mathsf{P2}}$, they are evaluated at time $t^{n+1/2} := t^n + \delta t/2$. The construction of these quantities is motivated in the following by a semi-discretization of the governing equations (2) in time (cf. [41]): Let us suppose for a moment a sufficiently smooth solution of these equations. By Taylor series expansion, height and momentum can be expressed at the new time level by

$$h(\boldsymbol{x},t^{n+1}) = h(\boldsymbol{x},t^n) - \delta t\left[\nabla\cdot(h\boldsymbol{v})(\boldsymbol{x},t^{n+1/2})\right] + \mathcal{O}\big(\delta t^3\big) \tag{7}$$

and

$$\begin{aligned}(h\boldsymbol{v})(\boldsymbol{x},t^{n+1}) = (h\boldsymbol{v})(\boldsymbol{x},t^n) - \delta t\big[&\nabla\cdot(h\boldsymbol{v}\circ\boldsymbol{v})(\boldsymbol{x},t^{n+1/2}) \\ &+ (h_0\nabla h^{(2)})(\boldsymbol{x},t^{n+1/2})\big] + \mathcal{O}\big(\delta t^3\big)\end{aligned} \tag{8}$$

for $\delta t \to 0$. Given the fluxes of the auxiliary system (6), momentum and velocity can be approximated at the half time level by

$$\begin{aligned}(h\boldsymbol{v})(\boldsymbol{x},t^{n+1/2}) &= (h\boldsymbol{v})^*(\boldsymbol{x},t^{n+1/2}) - \frac{\delta t}{2}\,(h_0\nabla h^{(2)})(\boldsymbol{x},t^{n+1/4}) + \mathcal{O}\big(\delta t^2\big) \\ \boldsymbol{v}(\boldsymbol{x},t^{n+1/2}) &= \boldsymbol{v}^*(\boldsymbol{x},t^{n+1/2}) - \frac{\delta t}{2}\,\nabla h^{(2)}(\boldsymbol{x},t^{n+1/4}) + \mathcal{O}\big(\delta t^2\big)\ .\end{aligned} \tag{9}$$

Here and below, the variables with asterisks denote the quantities of the auxiliary system, and $\boldsymbol{v}^* \equiv (h\boldsymbol{v})^*/h^*$. Note that – in order to achieve second order accuracy in time – the question at which time level the unknown $h^{(2)}$ "lives", can be relaxed to any point in the interval $[t^n, t^{n+1/2}]$. To ensure that the velocities on the left hand side of (9) satisfy the divergence constraint, we take the divergence of the first equation in (9) and obtain together with the first equation in (2) a first Poisson equation for $h^{(2)}$:

$$\frac{\delta t}{2}\,\nabla\cdot(h_0\nabla h^{(2)})(\boldsymbol{x},t^{n+1/4}) = \nabla\cdot(h\boldsymbol{v})^*(\boldsymbol{x},t^{n+1/2}) + \frac{dh_0}{dt}(t^{n+1/2}) + \mathcal{O}\big(\delta t^2\big)\ . \tag{10}$$

With the solution of this problem the right hand side of (7) and the first term in the brackets of (8) can be calculated through (9). The second term in brackets in (8) is computed by another application of the divergence constraint. Let

$$(h\boldsymbol{v})^{**}(\boldsymbol{x}) := (h\boldsymbol{v})(\boldsymbol{x},t^n) - \delta t \left[\nabla \cdot (h\boldsymbol{v} \circ \boldsymbol{v})(\boldsymbol{x},t^{n+1/2})\right] \tag{11}$$

denote a preliminary prediction of the new time level momentum that still lacks the influence of the pressure flux. Then, the momentum at the new time level is given by

$$(h\boldsymbol{v})(\boldsymbol{x},t^{n+1}) = (h\boldsymbol{v})^{**}(\boldsymbol{x}) - \delta t\,(h_0 \nabla h^{(2)})(\boldsymbol{x},t^{n+1/2}) + \mathcal{O}\big(\delta t^2\big)\ . \tag{12}$$

Imposing the divergence constraint from (2) once again at a half time step, but this time using a linear interpolation of the momentum at the full time levels, leads to

$$\frac{1}{2}\left[\nabla \cdot (h\boldsymbol{v})(\boldsymbol{x},t^{n+1}) + \nabla \cdot (h\boldsymbol{v})(\boldsymbol{x},t^n)\right] = -\frac{dh_0}{dt}(t^{n+1/2}) + \mathcal{O}\big(\delta t^2\big)\ . \tag{13}$$

Inserting (12) in (13), a second Poisson Problem for $h^{(2)}$ is obtained:

$$\begin{aligned}\delta t\,\nabla \cdot (h_0 \nabla h^{(2)})(\boldsymbol{x},t^{n+1/2}) = \nabla \cdot (h\boldsymbol{v})^{**}(\boldsymbol{x}) + \nabla \cdot (h\boldsymbol{v})(\boldsymbol{x},t^n)\\ + 2\frac{dh_0}{dt}(t^{n+1/2}) + \mathcal{O}\big(\delta t^2\big)\ .\end{aligned} \tag{14}$$

Thus, by the solution of an auxiliary hyperbolic system and two Poisson problems for the second order height $h^{(2)}$ numerical approximations to the fluxes of the zero Froude number shallow water equations can be computed up to second order accuracy in time.

At least for constant h_0, we could have also just projected at time t^{n+1}, rather than using the average of the momenta from t^n and t^{n+1} in (13). Both versions are equivalent for the present case of zero Froude number. Yet, the authors are currently working on an extension of the scheme to small but non-zero Froude number, and in this setting, evaluation at the half-time level turns out to be required in order to maintain second-order accuracy.

2.2 Discretization of the Projections

As stated above, equations (6)–(14) are a summary of the zero-Mach-number-scheme by [28] applied to the shallow water case. In this section we begin to introduce deviations from earlier work. This concerns, in particular, a new discretization of the Poisson equation, which – as we will show – leads to an exact and stable projection.

The Poisson equations (10) and (14) are discretized using a method originally proposed by Süli [34], who proves stability and convergence of the scheme in a mesh-dependent H^1 norm. In contrast to Süli, who considers a numerical method for a scalar elliptic Dirichlet problem, here we focus on the projection step of a flow solver that results in a Poisson-type problem with Neumann boundary conditions (cf. [16] and [28] for a discussion on that issue). The method can be either interpreted as a finite element or as a finite volume method. In the following, the scheme is introduced as a Petrov-Galerkin finite element method, which lays the groundwork for the stability proof of the projection given in the next section. Since the two Poisson equations are solved using slightly different discretizations, the method is first discussed for the second projection. Thereafter, modifications to be applied for the first Poisson problem are given.

For the derivation of the method, consider a Poisson problem with Neumann boundary conditions:

$$\begin{cases} -\nabla \cdot \nabla p = f & \text{in } \Omega, \\ \dfrac{\partial p}{\partial \boldsymbol{n}} = 0 & \text{on } \partial\Omega. \end{cases} \tag{15}$$

Given the r.h.s. $f \in L^2(\Omega)$ with $\int_\Omega f \, d\boldsymbol{x} = 0$, this problem has a unique solution $p \in H^1(\Omega)/\mathbb{R}$. Note that the space $H^1(\Omega)/\mathbb{R}$ is equivalent to the space $\{p \in H^1(\Omega) \,|\, \int_\Omega v \, d\boldsymbol{x} = 0\}$, the latter formulation being commonly used in practical computations. Since the right hand side f is of the form $-\nabla \cdot \boldsymbol{v}$ with a given velocity field $\boldsymbol{v}$ in the equation to be solved in the projection method, f is substituted with this term in the following discussion. The weak formulation of this problem is derived by multiplication of (15) with a test function ψ and integration over the whole domain Ω. Thus, we have to find p, such that

$$\int_\Omega \psi \, \nabla \cdot \nabla p \, d\boldsymbol{x} = \int_\Omega \psi \, \nabla \cdot \boldsymbol{v} \, d\boldsymbol{x} \quad \forall \psi \quad . \tag{16}$$

In (16) it is left open which trial and test spaces are considered. In contrast to the classical finite element theory, where the test function ψ is chosen to be (weakly) differentiable and Green's formula is applied to shift one derivative to the test function, here, a test space containing piecewise constant test functions is considered. In this case – assuming for a moment that p and $\boldsymbol{v}$ are sufficiently smooth – the divergence theorem can be applied.

In particular, for the construction of the test space, a *dual discretization* of the computational domain Ω is introduced, in which $\bar{\mathcal{C}}$ is the set of control volumes $\bar{C}$ centered about nodes of the original grid (see Figure 1). In the given Cartesian space discretization each grid cell is cut into four equally sized rectangles, and a dual control volume is the union of all rectangles, which have one grid node in common. Thus, control volumes which are associated with nodes on the boundary of the domain, have either half or fourth size compared to the inner control volumes. Notice that usage of the dual cells in formulating the projection is in line with [7,28]. The difference will lie in how we account for piecewise linear in-cell distributions of momentum and how they are affected by the divergence correction. The interfaces between these control volumes and the set of all such interfaces is denoted – in analogy to the primal discretization – by $\bar{I}$ and $\bar{\mathcal{I}}$, respectively. Then, the test space is given by all functions in $L^2(\Omega)$, which are constant on the dual control volumes. This space is defined by

$$\mathcal{Q}^h := \left\{ q \in L^2(\Omega) \,|\, \forall \bar{C} \in \bar{\mathcal{C}} : q|_{\bar{C}} \in \mathcal{P}_0(\bar{C}) \right\} \ , \tag{17}$$

in which

$$\mathcal{P}_k(U) := \left\{ p \in C^\infty(U) \,\middle|\, p(x,y) = \sum_{\substack{i+j \le k \\ i,j \ge 0}} c_{ij} x^i y^j, c_{ij} \in \mathbb{R} \right\} \tag{18}$$

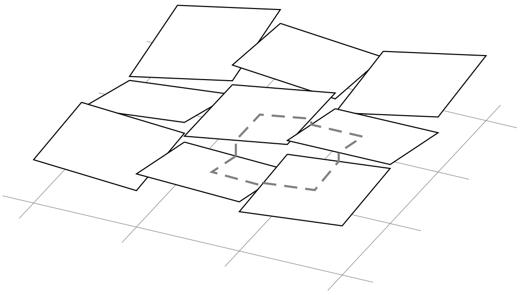

Fig. 2 Piecewise linear function for the velocity. The *dashed line* visualizes the projection of the integration path of the boundary integral onto the graph of the integrated function, which is evaluated in the discrete divergence. Thus, only piecewise linear functions have to be integrated.

is the space of polynomial functions on $U \subset \mathbb{R}^2$ of degree less than or equal to k. A basis of $\mathcal{Q}^h$ is given by $\bigcup_{\bar{C}\in\bar{\mathcal{C}}}\{\chi_{\bar{C}}\}$, where χ_U is the characteristic function on the set U. Therefore, a test function can be decomposed into $\psi(x,y) = \sum_{\bar{C}} \psi_{\bar{C}}\, \chi_{\bar{C}}(x,y)$, and equation (16) becomes

$$\sum_{\bar{C}\in\bar{\mathcal{C}}} \psi_{\bar{C}} \left(\int_{\bar{C}} \nabla\cdot\nabla p \, d\boldsymbol{x} - \int_{\bar{C}} \nabla\cdot\boldsymbol{v} \, d\boldsymbol{x} \right) = 0 \quad \forall \psi \in \mathcal{Q}^h \; .$$

Now, the divergence theorem can be applied, and we have to find p, such that

$$\sum_{\bar{C}\in\bar{\mathcal{C}}} \psi_{\bar{C}} \left(\int_{\partial\bar{C}} \nabla p\cdot\boldsymbol{n} \, d\sigma - \int_{\partial\bar{C}} \boldsymbol{v}\cdot\boldsymbol{n} \, d\sigma \right) = 0 \quad \forall \psi \in \mathcal{Q}^h \; , \tag{19}$$

Since all of the $\bar{C}$ are pairwise disjoint, this problem is a linear combination of the local problems to find p, such that

$$\int_{\partial\bar{C}} \nabla p\cdot\boldsymbol{n} \, d\sigma - \int_{\partial\bar{C}} \boldsymbol{v}\cdot\boldsymbol{n} \, d\sigma = 0 \quad \forall \bar{C} \in \bar{\mathcal{C}} \; , \tag{20}$$

and the solution p satisfies (19), if and only if it satisfies (20).

Using the latter formulation, the trial spaces for the unknown p and the vector valued function $\boldsymbol{v}$ are now defined as follows: Let us denote by $\mathcal{Q}_k(U)$ the space of all polynomials on $U \subset \mathbb{R}^2$ that are of degree $\leq k$ with respect to each, x and y. Choosing for p a trial space of continuous functions, which are piecewise bilinear on the primal control volumes $C \in \mathcal{C}$, i.e.

$$\mathcal{H}^h := \left\{ p \in H^1(\Omega)/\mathbb{R} \mid \forall C \in \mathcal{C} : p|_C \in \mathcal{Q}_1(C) \right\} \; , \tag{21}$$

the gradient of such functions is piecewise linear in each component on a control volume of the primal discretization, but discontinuous at the interfaces. Thus, for the velocity $\boldsymbol{v}$ a finite element space is chosen, which contains such gradients. It is defined by

$$\mathcal{U}^h := \left\{ \boldsymbol{v} = (u,v) \in [L^2(\Omega)]^2 \mid \forall C \in \mathcal{C} : \boldsymbol{v}|_C \in [\mathcal{P}_1(C)]^2 \right\} \; . \tag{22}$$

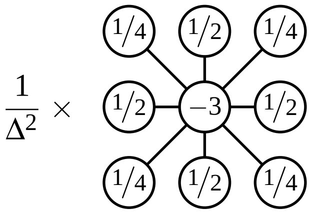

Fig. 3 Stencil of the discrete Laplacian on a uniform Cartesian grid with the same grid spacing Δ in both coordinate directions.

Although this space allows for discontinuities along cell interfaces, all the integrals in (20) are well defined. This is true, because the normal component of $\boldsymbol{v}$ and ∇p are piecewise linear along the boundaries of the dual control volumes (cf. Figure 2), and the expressions can be exactly evaluated. Note, that piecewise linear velocity or momentum components are the natural ansatz to obtain a second order Godunov-type scheme used in the explicit predictor step. To account for the boundary conditions, we separately treat the integrals over $\partial\Omega \cap C$ in (19) and replace $\nabla p \cdot \boldsymbol{n}$ by 0 as set in (15). This is in line with the procedure used for classical finite elements.

Using a suitable normalization, the integrals on the left hand side of (20) define a discrete Laplacian and divergence. Specifically, let us define the discrete Laplacian by

$$\mathsf{L} : \mathcal{H}^h \to \mathcal{Q}^h \text{ with } \mathsf{L}(p) := \sum_{\bar{C}\in\bar{\mathcal{C}}} \chi_{\bar{C}} \frac{1}{|\bar{C}|} \int_{\partial\bar{C}} \nabla p \cdot \boldsymbol{n}\, d\sigma \tag{23}$$

and the discrete divergence by

$$\mathsf{D} : \mathcal{U}^h \to \mathcal{Q}^h \text{ with } \mathsf{D}(\boldsymbol{v}) := \sum_{\bar{C}\in\bar{\mathcal{C}}} \chi_{\bar{C}} \frac{1}{|\bar{C}|} \int_{\partial\bar{C}} \boldsymbol{v} \cdot \boldsymbol{n}\, d\sigma\ . \tag{24}$$

Since each basis function of the test space is only nonzero on one dual control volume, the resulting stencil of the Laplacian is compact, i.e. it only uses next neighbors to the grid point for which the differential operator is discretized. As a consequence, the associated linear system can be easily computed with standard iterative methods. On a uniform Cartesian grid with the same grid spacing in both coordinate directions the stencil is given in Figure 3.

The property that the analytical gradient of $p \in \mathcal{H}^h$ is in the space $\mathcal{U}^h$ almost everywhere suggests that the discrete gradient operator is defined by

$$\mathsf{G} : \mathcal{H}^h \to \mathcal{U}^h \text{ with } \mathsf{G}(p) := \nabla p \quad \text{a.e.} \tag{25}$$

These discrete operators inherit from their analytic counterparts the property that they satisfy the equality $\mathsf{L} = \mathsf{D}(\mathsf{G})$.

The discretization of the first projection is done in a similar way. However, this time the advection velocity has to be corrected at the boundary of the primary control volumes. Thus, the test functions are chosen to be piecewise constant on each grid cell, which means that the divergence is applied to each such control

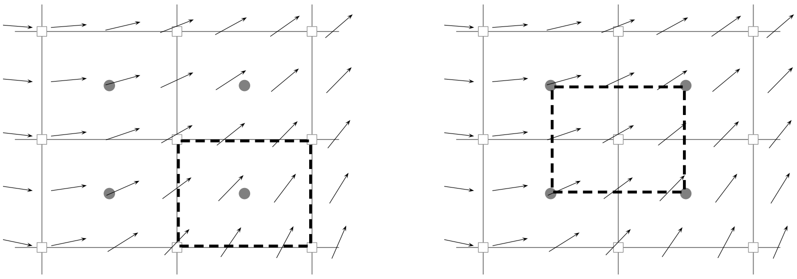

Fig. 4 Application of the divergence constraint in the MAC (*left*) and the second projection (*right*).

volume (see Figure 4). On a Cartesian grid, the discretization is essentially shifted by half a grid cell in each coordinate direction.

The resulting flux, arising from the MAC projection, is given by

$$\mathbf{F}_I^{\mathsf{MAC}} = -\frac{\delta t}{2}\begin{pmatrix} h_0 \nabla h^{(2)} \cdot \boldsymbol{n} \\ (h\boldsymbol{v})^* \nabla h^{(2)} \cdot \boldsymbol{n} + h_0 \nabla h^{(2)} \boldsymbol{v}^* \cdot \boldsymbol{n} \end{pmatrix}_I .$$

In the second projection, the local updates of the momentum are given by

$$(h\boldsymbol{v})^{n+1}(\boldsymbol{x})|_C = (h\boldsymbol{v})^{**}(\boldsymbol{x})|_C - \delta t\, h_0 \nabla h^{(2)}(\boldsymbol{x})|_C \quad C \in \mathcal{C}$$

(cf. (12)). This results in the flux contribution $\mathbf{F}_I^{\mathsf{P2}} = (0, h_0 h^{(2)} \boldsymbol{n})_I^T$, and conservation of momentum is guaranteed.

We emphasize that the update of the second projection not only involves the cell mean values, but also the gradient within a cell. This can be seen by a decomposition of the quantities into their mean value, linear and bilinear fractions, i.e.:

$$h^{(2)}(x,y)|_C = h_C^{(2)} + (x - x_C) h_{x,C}^{(2)} + (y - y_C) h_{y,C}^{(2)} + (x - x_C)(y - y_C) h_{xy,C}^{(2)} ,$$

where (x_C, y_C) is the center of cell C. Then, the gradient in each grid cell is given by

$$\nabla h^{(2)}(x,y)|_C = \begin{pmatrix} h_{x,C}^{(2)} \\ h_{y,C}^{(2)} \end{pmatrix} + \begin{pmatrix} y - y_C \\ x - x_C \end{pmatrix} h_{xy,C}^{(2)} ,$$

and the update of the mean values is

$$(h\boldsymbol{v})_C^{n+1} = (h\boldsymbol{v})_C^{**} - \delta t\, h_0 \begin{pmatrix} h_{x,C}^{(2)} \\ h_{y,C}^{(2)} \end{pmatrix} ,$$

whereas the correction of the gradients is computed by

$$(h\boldsymbol{v})_{x,C}^{n+1} = (h\boldsymbol{v})_{x,C}^{**} - \delta t\, h_0 \begin{pmatrix} 0 \\ h_{xy,C}^{(2)} \end{pmatrix}$$

and

$$(h\boldsymbol{v})^{n+1}_{y,C} = (h\boldsymbol{v})^{**}_{y,C} - \delta t\, h_0 \begin{pmatrix} h^{(2)}_{xy,C} \\ 0 \end{pmatrix} .$$

Additionally, a reconstruction step is introduced after the first projection, which reconstructs piecewise linear functions from cell averages of the intermediate momentum components $(hu)^{**}_C$ and $(hv)^{**}_C$. The second projection is then applied to this vector field to obtain a final momentum distribution. Note that the total variation diminishing (TVD) property could be destroyed in the projection step, even if it was installed in the reconstruction step before.

2.3 Exact Projection Method

Using the discretization described above for the second Poisson equation, the numerical method is formulated as an *exact projection method*. This means that the incompressibility condition on the velocity

$$(\nabla \cdot \boldsymbol{v}^n)_{\bar{C}} \coloneqq \frac{1}{\bar{C}} \int\limits_{\partial \bar{C}} \boldsymbol{v} \cdot \boldsymbol{n}\, d\sigma = -\frac{1}{h_0} \frac{dh_0}{dt}$$

is theoretically satisfied to machine precision at each full time level (i.e. in practice to the precision of the iterative solver for the discrete Poisson equation). As noted above, this definition of the divergence not only incorporates the cell mean values, but also the gradients of the velocity within each cell intersecting $\bar{C}$.

To derive an exact projection method the piecewise linear functions for the momentum have to be used throughout the whole scheme. For the solution of the semidiscrete equations arising from the auxiliary system Heun's method

$$\begin{aligned} \mathbf{U}^* &= \mathbf{U}^n + \frac{\delta t}{2} \left(f(\mathbf{U}^n) + f(\mathbf{U}^{*,\mathrm{int}}) \right) \quad \text{with} \\ \mathbf{U}^{*,\mathrm{int}} &= \mathbf{U}^n + \delta t\, f(\mathbf{U}^n) \end{aligned}$$

is applied for the integration in time. This approach leads to second-order accuracy in time. To obtain second-order accuracy in space as well, the cell average values in $\mathbf{U}^n$ and $\mathbf{U}^{*,\mathrm{int}}$ are reconstructed as piecewise linear functions on each grid cell. The numerical fluxes are then evaluated with the reconstructed values on the two sides of any particular interface.

Since the momentum components are already piecewise linear at time level t^n, they do not have to be reconstructed from the cell mean values and the gradients of the momentum components are used for the calculation of the numerical fluxes of the auxiliary system. These gradients are not only used for $\mathbf{U}^n$, but for $\mathbf{U}^{*,\mathrm{int}}$ as well. This does not reduce the scheme's order, because a Taylor series expansion for the gradient of $\mathbf{U}^{*,\mathrm{int}}$ yields

$$\mathbf{U}^{*,\mathrm{int}}_{\boldsymbol{x},C} = \mathbf{U}^n_{\boldsymbol{x},C} + \mathcal{O}(\delta t) .$$

In this scheme $\mathbf{U}_{\boldsymbol{x},C}$ is always multiplied by δx to yield the numerical fluxes of the auxiliary system. Therefore, the second order accuracy in space and time is retained.

With these modifications, we have a velocity field at each time level, which satisfies the discrete divergence constraint up to the accuracy of the elliptic solver, i.e. we have constructed an exact projection method.

3 Stability of the second projection

In proving stability of our semi-implicit method, the stability of the second projection step is an important prerequisite. Furthermore, as stated in the introduction, the final projection often led to a velocity-pressure decoupling in former projection methods. By using the theory of mixed finite element methods, we demonstrate that such instabilities cannot occur in the presented method.

In the second projection, the second order height $h^{(2)}$ is computed to correct the intermediate momentum update $(h\boldsymbol{v})^{**}$ in a post-processing step (cf. (12)). Thus, we are not only interested in a stable approximation of $h^{(2)}$, but rather in one of the momentum at the new time step. The associated Poisson-type problem is derived by imposing the additional requirement that the momentum at the new time step shall satisfy a discrete version of the divergence constraint

$$\int_{\partial V} (h\boldsymbol{v}) \cdot \boldsymbol{n}\, d\sigma = -|V| \frac{dh_0}{dt} \quad . \tag{26}$$

In the context of finite element methods, this leads to the theory of *saddle point problems* (mixed finite elements), which arise from minimization problems with additional side constraints. Starting with the fundamental work of Babuška [5] and Brezzi [10], this theory provides conditions for existence and uniqueness of solutions and for stable discretizations of such problems.

After having introduced the functional analytic framework, the discrete Poisson-type problem

$$\delta t\, \mathsf{D}\left(h_0\, \mathsf{G}(h^{(2)})\right) = \mathsf{D}\left((h\boldsymbol{v})^{**}\right) + \mathsf{D}\left((h\boldsymbol{v})^{n}\right) + 2\frac{dh_0}{dt} \tag{27}$$

is reformulated for the new projection method as a generalized saddle point problem, which is the starting point for the subsequent stability analysis.

3.1 Generalized Saddle Point Problems – Theory

The theory of finite element methods heavily benefits from the utilization of *Sobolev spaces*. These are based on the Hilbert space $L^2(\Omega)$, which includes all square integrable functions on Ω. The latter is defined by

$$L^2(\Omega) \coloneqq \left\{ q \;\middle|\; \int_\Omega |q(\boldsymbol{x})|^2\, d\boldsymbol{x} < +\infty \right\} ,$$

and the inner product and norm on this space are given by

$$(p,q)_{0,\Omega} \coloneqq \int_\Omega p(\boldsymbol{x}) q(\boldsymbol{x})\, d\boldsymbol{x} \, , \quad \|q\|_{0,\Omega} \coloneqq \sqrt{(q,q)_{0,\Omega}} \, .$$

Then, the first order Sobolev space is

$$H^1(\Omega) \coloneqq \left\{ q \in L^2(\Omega) \mid \nabla q \in [L^2(\Omega)]^2 \right\} .$$

We put $|q|_{1,\Omega} \coloneqq \|\nabla q\|_{0,\Omega}$ and $\|q\|_{1,\Omega} \coloneqq (\|q\|^2_{0,\Omega} + |q|^2_{1,\Omega})^{1/2}$, which define a seminorm and a norm on $H^1(\Omega)$, respectively. Note that $|\cdot|_{1,\Omega}$ defines a norm on the aforementioned quotient space $H^1(\Omega)/\mathbb{R}$, the space of equivalence classes of functions that differ only by a constant. We also refer to spaces of vector valued functions. For this reason, let us introduce

$$H(\mathrm{div};\Omega) \coloneqq \{ \boldsymbol{v} \in [L^2(\Omega)]^2 \mid \nabla \cdot \boldsymbol{v} \in L^2(\Omega) \} .$$

For a vector function $\boldsymbol{v} \in H(\mathrm{div};\Omega)$ it is possible to define its normal component on the boundary $\partial\Omega$ [15], and the subspace with vanishing normal component on $\partial\Omega$ is denoted by

$$H_0(\mathrm{div};\Omega) \coloneqq \{ \boldsymbol{v} \in H(\mathrm{div};\Omega) \mid \boldsymbol{v} \cdot \boldsymbol{n} = 0 \text{ on } \partial\Omega \} .$$

These spaces are equipped with the Hilbertian graph norm

$$\|\boldsymbol{v}\|_{\mathrm{div},\Omega} \coloneqq \left(\|\boldsymbol{v}\|^2_{0,\Omega} + \|\nabla \cdot \boldsymbol{v}\|^2_{0,\Omega} \right)^{1/2} .$$

For the analysis of the second projection we are interested in generalized mixed formulations with three distinct bilinear forms a, b_1, b_2. That is, to find $(u,p) \in \mathcal{U} \times \mathcal{H}$, such that

$$\begin{cases} a(u,v) + b_1(p,v) = \langle v', v \rangle & \forall v \in \mathcal{V} \\ \qquad\qquad b_2(u,q) = \langle q', q \rangle & \forall q \in \mathcal{Q} . \end{cases} \tag{28}$$

In this formulation, $\mathcal{H}$, $\mathcal{Q}$, $\mathcal{U}$ and $\mathcal{V}$ are four Hilbert spaces (or, more generally, reflexive Banach spaces) with norms $\|\cdot\|_{\mathcal{H}}$, $\|\cdot\|_{\mathcal{Q}}$, $\|\cdot\|_{\mathcal{U}}$ and $\|\cdot\|_{\mathcal{V}}$. The bilinear form a is defined on $\mathcal{U} \times \mathcal{V}$, b_1 on $\mathcal{H} \times \mathcal{V}$ and b_2 on $\mathcal{U} \times \mathcal{Q}$. Furthermore, v' and q' are elements of $\mathcal{V}'$ and $\mathcal{Q}'$, the dual spaces of $\mathcal{V}$ and $\mathcal{Q}$. The abstract theory of such problems is given in Nicolaïdes [25] and developed further in [8].

To obtain conditions for existence, uniqueness and stability of problem (28), let us introduce for any $r' \in \mathcal{H}'$ and $q' \in \mathcal{Q}'$ the closed affine spaces

$$\mathcal{K}_1(r') \coloneqq \{ v \in \mathcal{V} \mid \forall r \in \mathcal{H} : b_1(r,v) = \langle r', r \rangle \}$$

and

$$\mathcal{K}_2(q') \coloneqq \{ w \in \mathcal{U} \mid \forall q \in \mathcal{Q} : b_2(w,q) = \langle q', q \rangle \} .$$

We denote by $\mathcal{K}_i \coloneqq \mathcal{K}_i(0)$ $(i = 1,2)$ the kernel of the operator induced by b_i. With these definitions the following Theorem can be stated:

Theorem 1 **[25]** *Let a and b_i $(i=1,2)$ be bounded bilinear forms. Assume that there exists a constant $\alpha > 0$, such that*

$$\inf_{w\in\mathcal{K}_2} \sup_{v\in\mathcal{K}_1} \frac{a(w,v)}{\|w\|_{\mathcal{U}}\,\|v\|_{\mathcal{V}}} \geq \alpha \tag{29}$$

and

$$\sup_{w\in\mathcal{K}_2} a(w,v) > 0 \quad \forall v \in \mathcal{K}_1 \setminus \{0\} \quad . \tag{30}$$

Furthermore, assume that the b_i $(i=1,2)$ satisfy the inf-sup conditions

$$\inf_{r\in\mathcal{H}} \sup_{v\in\mathcal{V}} \frac{b_1(r,v)}{\|r\|_{\mathcal{H}}\,\|v\|_{\mathcal{V}}} \geq \beta_1 > 0 \tag{31}$$

and

$$\inf_{q\in\mathcal{Q}} \sup_{w\in\mathcal{U}} \frac{b_2(w,q)}{\|w\|_{\mathcal{U}}\,\|q\|_{\mathcal{Q}}} \geq \beta_2 > 0 \quad . \tag{32}$$

Then, problem (28) *has a unique solution (u,p) for all $v' \in \mathcal{V}'$ and $q' \in \mathcal{Q}'$ and the following estimate holds:*

$$\|u\|_{\mathcal{U}} + \|p\|_{\mathcal{H}} \leq c\left(\|v'\|_{\mathcal{V}'} + \|q'\|_{\mathcal{Q}'}\right) \quad . \tag{33}$$

For the discretization of problem (28), it is assumed that there are finite-dimensional subspaces $\mathcal{H}^h \subset \mathcal{H}$, $\mathcal{Q}^h \subset \mathcal{Q}$, $\mathcal{U}^h \subset \mathcal{U}$ and $\mathcal{V}^h \subset \mathcal{V}$ and bilinear forms $a_h : \mathcal{U}^h \times \mathcal{V}^h \to \mathbb{R}$, $b_{1h} : \mathcal{H}^h \times \mathcal{V}^h \to \mathbb{R}$ and $b_{2h} : \mathcal{U}^h \times \mathcal{Q}^h \to \mathbb{R}$. Given the linear functionals $v_h' \in (\mathcal{V}^h)'$ and $q_h' \in (\mathcal{Q}^h)'$, we are looking for the solution $(u_h,p_h) \in \mathcal{U}^h \times \mathcal{H}^h$ of the discrete problem

$$\begin{cases} a_h(u_h,v_h) + b_{1h}(p_h,v_h) = \left\langle v_h', v_h \right\rangle & \forall v_h \in \mathcal{V}^h \\ \\ b_{2h}(u_h,q_h) = \left\langle q_h', q_h \right\rangle & \forall q_h \in \mathcal{Q}^h \ , \end{cases} \tag{34}$$

approximating the solution of the continuous problem. With the definition of the discrete affine spaces $\mathcal{K}_1^h$ and $\mathcal{K}_2^h$, in analogy to the continuous case, Theorem 1 can be applied to problem (34), and existence, uniqueness and stability are obtained given the constants α, β_1 and β_2 in (29), (31) and (32) are independent of the grid parameter h. Examples of mixed finite element discretizations of such type are given in [25] and [8]. A nonconforming discretization, where $\mathcal{U}^h \not\subseteq \mathcal{U}$, is constructed in [4]. Moreover, error estimates are provided in these references for both, the conforming and the nonconforming situation.

In the following, such a formulation is derived for the new projection in order to analyze its stability concerning the corrected momentum field.

3.2 Reformulation of the problem

The continuous counterpart of the discrete Poisson-type problem (27) is obtained by a combination of the momentum update and the divergence constraint, i.e.

$$\begin{aligned} (h\boldsymbol{v})^{n+1} &= (h\boldsymbol{v})^{**} - \delta t\,(h_0 \nabla h^{(2)}) \\ \frac{1}{2}\left[\nabla\cdot(h\boldsymbol{v})^{n+1} + \nabla\cdot(h\boldsymbol{v})^{n}\right] &= -\frac{dh_0}{dt} \quad . \end{aligned} \tag{35}$$

A variational formulation of these two equations is derived by the usual procedure: $(35)_1$ and $(35)_2$ are multiplied with test functions $\boldsymbol{\varphi}$ and ψ, respectively, and the resulting equations are integrated over the whole domain Ω. This leads to

$$\begin{aligned} \left((h\boldsymbol{v})^{n+1}, \boldsymbol{\varphi}\right)_{0,\Omega} + \left(\delta t\, h_0 \nabla h^{(2)}, \boldsymbol{\varphi}\right)_{0,\Omega} &= \left((h\boldsymbol{v})^{**}, \boldsymbol{\varphi}\right)_{0,\Omega} \\ \left(\nabla\cdot(h\boldsymbol{v})^{n+1}, \psi\right)_{0,\Omega} &= -\left(\nabla\cdot(h\boldsymbol{v})^{n} + 2\frac{dh_0}{dt}, \psi\right)_{0,\Omega} \quad . \end{aligned} \tag{36}$$

This formulation can be already interpreted as a generalized problem as formulated in (28). The discrete method, equivalent to the Poisson-type problem (27), is derived by introducing appropriate finite dimensional trial and test spaces. For the choice of the trial spaces, we are confined to our selection for the momentum $(h\boldsymbol{v})$ and the height $h^{(2)}$. In the projection method, the momentum distribution is approximated by discontinuous piecewise linear functions belonging to the space $\mathcal{U}^h$ defined in (22). The second order height $h^{(2)} \in \mathcal{H}^h$ is given by continuous piecewise bilinear functions (cf. (21)).

To obtain the same divergence as in (27), also the test functions ψ for the second equation of (36) are fixed to be piecewise constant on dual control volumes, forming the space $\mathcal{Q}^h$ defined in (17). The selection of the test space $\mathcal{V}^h$ for the first equation is yet undetermined. Let us choose $\mathcal{V}^h = \mathcal{U}^h$, the space which is also used for the momentum variable. A basis of $\mathcal{V}^h$ is given by

$$\begin{aligned} \bigcup_{C\in\mathcal{C}} \Bigg\{ &\begin{pmatrix}\chi_C\\0\end{pmatrix}, \begin{pmatrix}0\\\chi_C\end{pmatrix}, \begin{pmatrix}(x-x_C)\chi_C\\0\end{pmatrix}, \begin{pmatrix}(y-y_C)\chi_C\\0\end{pmatrix}, \\ &\qquad \begin{pmatrix}0\\(x-x_C)\chi_C\end{pmatrix}, \begin{pmatrix}0\\(y-y_C)\chi_C\end{pmatrix} \Bigg\}, \end{aligned} \tag{37}$$

where (x_C, y_C) is the center of the cell C. In our Cartesian space discretization grid cells are defined by $C_{i,j}$, $i = 1,\ldots,m$, $j = 1,\ldots,n$, with cell centers (x_i, y_j). Because of the linearity of the equations (36) in $\boldsymbol{\varphi}$ and ψ, it is sufficient to "test" them with only a basis of $\mathcal{U}^h$ and $\mathcal{Q}^h$, respectively. Let us consider the first equation in conjunction with the test function $\boldsymbol{\varphi} = (\chi_{C_{i,j}}, 0)^T$. Because the second component of $\boldsymbol{\varphi}$ is zero and its support is $C_{i,j}$, this yields

$$\int\limits_{C_{i,j}} (hu)^{n+1}\, d\boldsymbol{x} + \delta t\, h_0 \int\limits_{C_{i,j}} \frac{\partial h^{(2)}}{\partial x}\, d\boldsymbol{x} = \int\limits_{C_{i,j}} (hu)^{**}\, d\boldsymbol{x} \quad . \tag{38}$$

Furthermore, by expanding the height $h^{(2)}$ in a volumewise representation, i.e.

$$h^{(2)}(x,y)|_{C_{i,j}} = h^{(2)}_{i,j} + (x-x_i)h^{(2)}_{x,i,j} + (y-y_j)h^{(2)}_{y,i,j} + (x-x_i)(y-y_j)h^{(2)}_{xy,i,j} \quad , \tag{39}$$

the calculation of the second integral in (38) leads to

$$\int\limits_{C_{i,j}} \frac{\partial h^{(2)}}{\partial x}\, d\boldsymbol{x} = \int\limits_{C_{i,j}} \left(h^{(2)}_{x,i,j} + (y-y_j)h^{(2)}_{xy,i,j} \right) d\boldsymbol{x} = \delta x\, \delta y\, h^{(2)}_{x,i,j} \quad .$$

The integral of the second term vanishes, because it is an odd function in y with respect to y_j. With similar results for the other terms in (38), we finally obtain

$$(hu)^{n+1}_{i,j} + \delta t\, h_0\, h^{(2)}_{x,i,j} = (hu)^{**}_{i,j} \quad . \tag{40}$$

By using the other five test functions in (37), this procedure yields the equations

$$\begin{aligned}
(hv)^{n+1}_{i,j} &+ \delta t\, h_0\, h^{(2)}_{y,i,j} &&= (hv)^{**}_{i,j} \\
(hu)^{n+1}_{x,i,j} & &&= (hu)^{**}_{x,i,j} \\
(hu)^{n+1}_{y,i,j} &+ \delta t\, h_0\, h^{(2)}_{xy,i,j} &&= (hu)^{**}_{y,i,j} \\
(hv)^{n+1}_{x,i,j} &+ \delta t\, h_0\, h^{(2)}_{xy,i,j} &&= (hv)^{**}_{x,i,j} \\
(hv)^{n+1}_{y,i,j} & &&= (hv)^{**}_{y,i,j} \quad .
\end{aligned} \tag{41}$$

Therefore, six equations are obtained for each cell $C_{i,j}$. They represent the discretization of $(36)_1$.

The discretization of the second equation in (36) is done as follows. The application of the test function $\psi = \chi_{\bar{C}}$ and the divergence theorem yields for the terms involving the momentum the key ingredient of the discrete divergence $\mathsf{D}(\cdot)$. Thus, multiplying this equation by $\chi_{\bar{C}}/|\bar{C}|$ and summation over $\bar{C} \in \bar{\mathcal{C}}$ leads to

$$\mathsf{D}\big((h\boldsymbol{v})^{n+1}\big) = -\mathsf{D}((h\boldsymbol{v})^n) - 2\,\frac{dh_0}{dt} \quad . \tag{42}$$

Let us recall that $h^{(2)}$ is uniquely defined by its nodal values and that each velocity component has three degrees of freedom per grid cell. Then there are $7 \cdot m \cdot n$ unknowns in case of periodic boundary conditions, where m and n are the number of cells in x and y direction, respectively. The analysis above yielded the same number of linear equations. Finally, by inserting the equations from (40) and (41) into (42), the second discrete Poisson-type problem from our new projection method is obtained. We have derived a *Petrov-Galerkin* mixed formulation, which utilizes different trial and test spaces for the scalar variables.

3.3 Stability analysis of the mixed formulation

In order to apply the theory from Section 3.1 to the mixed formulation (36), the corresponding continuous problem is defined which has been shown to have a unique solution in [41]. Here, the main investigation will be on the stability of the discrete mixed formulation.

For the analysis of the continuous problem appropriate function spaces for the trial and test functions have to be chosen. In the Poisson-type problem

$$\delta t \, \nabla \cdot (h_0 \nabla h^{(2)}) = \nabla \cdot (h\boldsymbol{v})^{**} + \nabla \cdot (h\boldsymbol{v})^n + 2\,\frac{dh_0}{dt} \ ,$$

the continuous counterpart of (27), the second order height $h^{(2)}$ is only determined up to an additive constant. This constant can be fixed by the additional condition of a zero mean value, i.e., $\int_\Omega h^{(2)} d\boldsymbol{x} = 0$. Thus, a suitable trial space for $h^{(2)}$ is given by $\mathcal{H} \coloneqq H^1(\Omega)/\mathbb{R}$. An appropriate space for the momentum should also bound the divergence of the unknown variable. Furthermore, the boundary conditions are given by the integral constraint (26). For simplicity, let us assume that there is no flux across the boundary, i.e. there are impermeable rigid walls and $dh_0/dt \equiv 0$. Then, the momentum is sought in the space $\mathcal{U} = H_0(\mathrm{div};\Omega)$. The test functions of the discrete problem are discontinuous at the interfaces either of the primal or of the dual discretization. Therefore, no particular regularity is assumed for the test spaces in the continuous problem as well, and they are defined by $\mathcal{V} = [L^2(\Omega)]^2$ and $\mathcal{Q} = L^2(\Omega)$, respectively.

With the definition of the bilinear forms

$$\begin{aligned} a : \mathcal{U} \times \mathcal{V} \to \mathbb{R} \quad &\text{with} \quad a(\boldsymbol{w},\boldsymbol{v}) \coloneqq (\boldsymbol{w},\boldsymbol{v})_{0,\Omega} \\ b_1 : \mathcal{H} \times \mathcal{V} \to \mathbb{R} \quad &\text{with} \quad b_1(r,\boldsymbol{v}) \coloneqq (\nabla r,\boldsymbol{v})_{0,\Omega} \\ b_2 : \mathcal{U} \times \mathcal{Q} \to \mathbb{R} \quad &\text{with} \quad b_2(\boldsymbol{w},q) \coloneqq (\nabla \cdot \boldsymbol{w},q)_{0,\Omega} \ , \end{aligned} \tag{43}$$

problem (36) can be reformulated to obtain the following continuous saddle point problem. Find $((h\boldsymbol{v})^{n+1}, \delta t\, h_0\, h^{(2)}) \in \mathcal{U} \times \mathcal{H}$, such that

$$\begin{aligned} a\big((h\boldsymbol{v})^{n+1},\boldsymbol{\varphi}\big) + b_1\Big(\delta t\, h_0\, h^{(2)},\boldsymbol{\varphi}\Big) &= ((h\boldsymbol{v})^{**},\boldsymbol{\varphi})_{0,\Omega} \quad \forall\, \boldsymbol{\varphi} \in \mathcal{V} \\ b_2\big((h\boldsymbol{v})^{n+1},\psi\big) &= -b_2((h\boldsymbol{v})^n,\psi) \quad \forall\, \psi \in \mathcal{Q} \quad . \end{aligned} \tag{44}$$

This obviously defines a problem of the form (28). The formulation is also referred to as a *primal-dual* formulation [36,4]. In [41] it is shown that the given bilinear forms are bounded and that the inf-sup conditions (29)–(32) are satisfied. Thus, the following theorem can be stated:

Theorem 2 [41] *The generalized saddle point problem defined by* (44) *has a unique solution* $((h\boldsymbol{v})^{n+1}, \delta t\, h_0\, h^{(2)})$ *in* $\mathcal{U} \times \mathcal{H}$.

Since $\mathcal{H}^h \subset \mathcal{H}$, $\mathcal{U}^h \subset [L^2(\Omega)]^2$ and $\mathcal{V}^h \subset \mathcal{V}$, and the discrete gradient G is equal to its continuous counterpart on each grid cell, the bilinear forms a and b_1 are well defined on $\mathcal{U}^h \times \mathcal{V}^h$ and $\mathcal{H}^h \times \mathcal{V}^h$, respectively. This is different for b_2, since $\mathcal{U}^h \not\subseteq \mathcal{U}$. The bilinear form which represents the discrete divergence from (24) is defined by

$$b_{2h} : \mathcal{U}^h \times \mathcal{Q}^h \to \mathbb{R} \quad \text{with} \quad b_{2h}(\boldsymbol{v}_h, q_h) \coloneqq \sum_{\bar{C} \in \bar{\mathcal{C}}} q_{h,\bar{C}} \int_{\partial \bar{C}} \boldsymbol{v}_h \cdot \boldsymbol{n} \, d\sigma \;, \tag{45}$$

where $q_{h,\bar{C}}$ is the (constant) value of q_h on $\bar{C}$. This definition is consistent with the definition of its continuous counterpart b_2, since for functions $\boldsymbol{v} \in H(\operatorname{div};\Omega)$ they both give the same result. Furthermore, the $H(\operatorname{div};\Omega)$ norm is no longer appropriate for the space $\mathcal{U}^h$, and a suitable mesh dependent norm $\|\cdot\|_{\mathcal{U}^h}$ has to be introduced (cf. [9]).

Proposition 3 *A norm on the finite element space $\mathcal{U}^h$ is given by*

$$\|\boldsymbol{w}_h\|_{\mathcal{U}^h} \coloneqq \|\boldsymbol{w}_h\|_{0,\Omega} + \sup_{z_h \in \mathcal{Q}^h} \frac{b_{2h}(\boldsymbol{w}_h, z_h)}{\|z_h\|_{\mathcal{Q}}} \quad \textit{for } \boldsymbol{w}_h \in \mathcal{U}^h \;.$$

Proof We have to show definiteness, homogeneity, and the triangle inequality of $\|\cdot\|_{\mathcal{U}^h}$:

- First, it follows by the definition of the norm that for $\boldsymbol{w}_h \in \mathcal{U}^h$ with $\|\boldsymbol{w}_h\|_{\mathcal{U}^h} = 0$ one obtains $\|\boldsymbol{w}_h\|_{0,\Omega} = 0$. Since $\boldsymbol{w}_h$ is piecewise linear, i.e., piecewise continuous, $\boldsymbol{w}_h$ has to be zero almost everywhere.
- For $\lambda \in \mathbb{R}$ and $\boldsymbol{w}_h \in \mathcal{U}^h$ we have

$$\|\lambda \boldsymbol{w}_h\|_{\mathcal{U}^h} = \|\lambda \boldsymbol{w}_h\|_{0,\Omega} + \sup_{z_h \in \mathcal{Q}^h} \frac{b_{2h}(\lambda \boldsymbol{w}_h, z_h)}{\|z_h\|_{\mathcal{Q}}} = |\lambda| \, \|\boldsymbol{w}_h\|_{\mathcal{U}^h} \;.$$

- The triangle inequality holds for $\boldsymbol{w}_h, \tilde{\boldsymbol{w}}_h \in \mathcal{U}^h$, since

$$\begin{aligned}
\|\boldsymbol{w}_h + \tilde{\boldsymbol{w}}_h\|_{\mathcal{U}^h} &= \|\boldsymbol{w}_h + \tilde{\boldsymbol{w}}_h\|_{0,\Omega} + \sup_{z_h \in \mathcal{Q}^h} \frac{b_{2h}(\boldsymbol{w}_h + \tilde{\boldsymbol{w}}_h, z_h)}{\|z_h\|_{\mathcal{Q}}} \\
&\le \|\boldsymbol{w}_h\|_{0,\Omega} + \|\tilde{\boldsymbol{w}}_h\|_{0,\Omega} + \sup_{z_h \in \mathcal{Q}^h} \frac{b_{2h}(\boldsymbol{w}_h, z_h) + b_{2h}(\tilde{\boldsymbol{w}}_h, z_h)}{\|z_h\|_{\mathcal{Q}}} \\
&\le \|\boldsymbol{w}_h\|_{\mathcal{U}^h} + \|\tilde{\boldsymbol{w}}_h\|_{\mathcal{U}^h}
\end{aligned}$$

□

In this norm, the bilinear form b_{2h} is continuous, since for arbitrary $q_h \in \mathcal{Q}^h$ and $\boldsymbol{w}_h \in \mathcal{U}^h$ it follows that

$$\begin{aligned}
b_{2h}(\boldsymbol{w}_h, q_h) &= \frac{\|q_h\|_{\mathcal{Q}} \, b_{2h}(\boldsymbol{w}_h, q_h)}{\|q_h\|_{\mathcal{Q}}} \\
&\le \|q_h\|_{\mathcal{Q}} \sup_{z_h \in \mathcal{Q}^h} \frac{b_{2h}(\boldsymbol{w}_h, z_h)}{\|z_h\|_{\mathcal{Q}}} \\
&\le \|q_h\|_{\mathcal{Q}} \, \|\boldsymbol{w}_h\|_{\mathcal{U}^h}
\end{aligned}$$

Proposition 4 *For $\boldsymbol{w}_h \in \mathcal{U}^h$ one has*

$$\sup_{z_h \in \mathcal{Q}^h} \frac{b_{2h}(\boldsymbol{w}_h, z_h)}{\|z_h\|_{\mathcal{Q}}} = \left(\sum_{\bar{C} \in \bar{\mathcal{C}}} \frac{1}{|\bar{C}|} \left(\int_{\partial \bar{C}} \boldsymbol{w}_h \cdot \boldsymbol{n} \, d\sigma \right)^2 \right)^{1/2} .$$

Proof Taking $\boldsymbol{w}_h \in \mathcal{U}^h$ and $z_h \in \mathcal{Q}^h$ it follows from the Cauchy-Schwarz inequality that

$$\begin{aligned}
\frac{b_{2h}(\boldsymbol{w}_h, z_h)}{\|z_h\|_{\mathcal{Q}}} &= \frac{\sum_{\bar{C}} z_h|_{\bar{C}} \int_{\partial \bar{C}} \boldsymbol{w}_h \cdot \boldsymbol{n} \, d\sigma}{\left(\sum_{\bar{C}} |\bar{C}| (z_h|_{\bar{C}})^2\right)^{1/2}} \\
&= \frac{\sum_{\bar{C}} \left(|\bar{C}|^{1/2} z_h|_{\bar{C}}\right) \left(|\bar{C}|^{-1/2} \int_{\partial \bar{C}} \boldsymbol{w}_h \cdot \boldsymbol{n} \, d\sigma\right)}{\left(\sum_{\bar{C}} |\bar{C}| (z_h|_{\bar{C}})^2\right)^{1/2}} \\
&\le \frac{\left(\sum_{\bar{C}} |\bar{C}| (z_h|_{\bar{C}})^2\right)^{1/2} \left(\sum_{\bar{C}} |\bar{C}|^{-1} \left(\int_{\partial \bar{C}} \boldsymbol{w}_h \cdot \boldsymbol{n} \, d\sigma\right)^2\right)^{1/2}}{\left(\sum_{\bar{C}} |\bar{C}| (z_h|_{\bar{C}})^2\right)^{1/2}} \\
&= \left(\sum_{\bar{C}} \frac{1}{|\bar{C}|} \left(\int_{\partial \bar{C}} \boldsymbol{w}_h \cdot \boldsymbol{n} \, d\sigma \right)^2 \right)^{1/2}
\end{aligned}$$

Since z_h is arbitrary, this provides the proof in one direction. On the other hand, setting $z_h|_{\bar{C}} := |\bar{C}|^{-1} \int_{\partial \bar{C}} \boldsymbol{w}_h \cdot \boldsymbol{n} \, d\sigma$ gives

$$\frac{b_{2h}(\boldsymbol{w}_h, z_h)}{\|z_h\|_{\mathcal{Q}}} = \left(\sum_{\bar{C}} \frac{1}{|\bar{C}|} \left(\int_{\partial \bar{C}} \boldsymbol{w}_h \cdot \boldsymbol{n} \, d\sigma \right)^2 \right)^{1/2}$$

Taking the supremum over all $z_h \in \mathcal{Q}^h$ leads to the desired result. □

With the definition of the bilinear form in (45), the discrete mixed formulation derived in Section 3.2 is to find $((h\boldsymbol{v})^{n+1}, \delta t \, h_0 \, h^{(2)}) \in \mathcal{U}^h \times \mathcal{H}^h$, such that

$$\begin{aligned}
a\big((h\boldsymbol{v})^{n+1}, \boldsymbol{\varphi}_h\big) + b_1\Big(\delta t \, h_0 \, h^{(2)}, \boldsymbol{\varphi}_h\Big) &= ((h\boldsymbol{v})^{**}, \boldsymbol{\varphi}_h)_{0,\Omega} \quad && \forall \boldsymbol{\varphi}_h \in \mathcal{V}^h \\
b_{2h}\big((h\boldsymbol{v})^{n+1}, \psi_h\big) &= -b_{2h}((h\boldsymbol{v})^n, \psi_h) \quad && \forall \psi_h \in \mathcal{Q}^h \, .
\end{aligned} \tag{46}$$

Note that the trial space $\mathcal{U}^h$ is not contained in its continuous counterpart $\mathcal{U}$. Therefore, the discrete problem (46) is an approximation using *nonconforming finite elements*.

Now, the verification of the inf-sup-conditions can be carried out. The proof for the b_1 form is nearly identical to the continuous case (cf. [41]).

Proposition 5 *There exists a constant $\beta_1^* > 0$ independent of the mesh size, h, such that*

$$\inf_{r_h \in \mathcal{H}^h} \sup_{\boldsymbol{v}_h \in \mathcal{V}^h} \frac{b_1(r_h, \boldsymbol{v}_h)}{\|r_h\|_{\mathcal{H}} \, \|\boldsymbol{v}_h\|_{\mathcal{V}}} \ge \beta_1^*$$

Proof It has been already pointed out that $r_h \in \mathcal{H}^h$ implies $\nabla r_h \in \mathcal{V}^h$. Thus, we have for arbitrary $r_h \in \mathcal{H}^h$, $\|r_h\|_{\mathcal{H}} \neq 0$

$$\sup_{\boldsymbol{v}_h \in \mathcal{V}^h} \frac{b_1(r_h, \boldsymbol{v}_h)}{\|\boldsymbol{v}_h\|_{\mathcal{V}}} \geq \frac{b_1(r_h, \nabla r_h)}{\|\nabla r_h\|_{0,\Omega}} = \frac{|r_h|^2_{1,\Omega}}{|r_h|_{1,\Omega}} = \|r_h\|_{\mathcal{H}} \quad .$$

□

Next, it is proved what is normally known as coercivity for the bilinear form a. Since we deal with a Petrov-Galerkin method, the characterization has to be generalized to the two conditions (29) and (30). Let us define the subspaces

$$\begin{aligned}
\mathcal{K}_1^h &:= \{\boldsymbol{v}_h \in \mathcal{V}^h \mid \forall r_h \in \mathcal{H}^h : b_1(r_h, \boldsymbol{v}_h) = 0\} \\
\mathcal{K}_2^h &:= \{\boldsymbol{w}_h \in \mathcal{U}^h \mid \forall q_h \in \mathcal{Q}^h : b_{2h}(\boldsymbol{w}_h, q_h) = 0\} \ .
\end{aligned}$$

Our strategy is to show that there is a one-to-one mapping between these spaces, and an estimate can be given between corresponding elements. To characterize the spaces, it suffices to test the bilinear forms that are used in defining them against a complete set of basis functions of the test spaces. Thus, let $r_h \in \mathcal{H}^h$ with $r_h(x_{k+1/2}, y_{l+1/2}) = \delta_{ik}\delta_{jl}$ for a given node $(x_{i+1/2}, y_{j+1/2})$. Assuming a cell wise representation of r_h (cf. (39)), a careful investigation of such a basis function reveals that $r_{x,l,k} = \pm\frac{1}{2\delta x}$, $r_{y,l,k} = \pm\frac{1}{2\delta y}$ and $r_{xy,l,k} = \pm\frac{1}{\delta x\delta y}$ for $l \in \{i, i+1\}$, $k \in \{j, j+1\}$. Thus, $\boldsymbol{v}_h = (u, v) \in \mathcal{V}^h$ is in $\mathcal{K}_1^h$, if and only if for all possible (i, j)

$$\begin{aligned}
0 = b_1(r_h, \boldsymbol{v}_h) &= \sum_{l,k} \int_{C_{lk}} \nabla r_h \cdot \boldsymbol{v}_h \, d\boldsymbol{x} \\
&= \sum_{l=i}^{i+1} \sum_{k=j}^{j+1} \delta x \delta y \left(u_{l,k} r_{x,l,k} + v_{l,k} r_{y,l,k} + \frac{1}{12}(\delta y^2 u_{y,l,k} + \delta x^2 v_{x,l,k}) r_{xy,l,k} \right) \\
&= -\frac{\delta y}{2} u_{i+1,j+1} - \frac{\delta x}{2} v_{i+1,j+1} + \frac{\delta y^2}{12} u_{y,i+1,j+1} + \frac{\delta x^2}{12} v_{x,i+1,j+1} \\
&\quad + \frac{\delta y}{2} u_{i,j+1} - \frac{\delta x}{2} v_{i,j+1} - \frac{\delta y^2}{12} u_{y,i,j+1} - \frac{\delta x^2}{12} v_{x,i,j+1} \\
&\quad + \frac{\delta y}{2} u_{i,j} + \frac{\delta x}{2} v_{i,j} + \frac{\delta y^2}{12} u_{y,i,j} + \frac{\delta x^2}{12} v_{x,i,j} \\
&\quad - \frac{\delta y}{2} u_{i+1,j} + \frac{\delta x}{2} v_{i+1,j} - \frac{\delta y^2}{12} u_{y,i+1,j} - \frac{\delta x^2}{12} v_{x,i+1,j}
\end{aligned} \tag{47}$$

Similarly, let $q_h \in \mathcal{Q}^h$ with $q_h = \chi_{\bar{C}_{i+1/2,j+1/2}}$ be arbitrary. Then, $\boldsymbol{w}_h = (u,v) \in \mathcal{U}^h$ is in $\mathcal{K}_2^h$, if and only if for all possible (i,j)

$$
\begin{aligned}
0 = -b_{2h}(\boldsymbol{w}_h, q_h) &= - \int\limits_{\partial \bar{C}_{i+1/2,j+1/2}} \boldsymbol{w}_h \cdot \boldsymbol{n}\, d\sigma \\
&= -\frac{\delta y}{2} u_{i+1,j+1} - \frac{\delta x}{2} v_{i+1,j+1} + \frac{\delta y^2}{8} u_{y,i+1,j+1} + \frac{\delta x^2}{8} v_{x,i+1,j+1} \\
&\quad + \frac{\delta y}{2} u_{i,j+1} \quad - \frac{\delta x}{2} v_{i,j+1} \quad - \frac{\delta y^2}{8} u_{y,i,j+1} \quad - \frac{\delta x^2}{8} v_{x,i,j+1} \\
&\quad + \frac{\delta y}{2} u_{i,j} \quad + \frac{\delta x}{2} v_{i,j} \quad + \frac{\delta y^2}{8} u_{y,i,j} \quad + \frac{\delta x^2}{8} v_{x,i,j} \\
&\quad - \frac{\delta y}{2} u_{i+1,j} \quad + \frac{\delta x}{2} v_{i+1,j} \quad - \frac{\delta y^2}{8} u_{y,i+1,j} \quad - \frac{\delta x^2}{8} v_{x,i+1,j}
\end{aligned}
\tag{48}
$$

Comparing (47) and (48), we observe that these conditions only differ by a constant factor in the terms, which include partial derivatives of the velocity components. This means that a one-to-one mapping between $\mathcal{K}_1^h$ and $\mathcal{K}_2^h$ can be defined by multiplying the partial derivatives of an element with $8/12 = 2/3$, and the spaces have the same dimension. Furthermore, the following estimates can be given for corresponding elements $\boldsymbol{v}_h \in \mathcal{K}_1^h$ and $\boldsymbol{w}_h \in \mathcal{K}_2^h$ (i.e. with the same mean values $\bar{\boldsymbol{w}}_h = \bar{\boldsymbol{v}}_h$, and linear variations $\tilde{\boldsymbol{w}}_h = 2/3\tilde{\boldsymbol{v}}_h$):

$$
\begin{aligned}
a(\boldsymbol{w}_h, \boldsymbol{v}_h) &= a(\bar{\boldsymbol{w}}_h, \bar{\boldsymbol{v}}_h) + a(\tilde{\boldsymbol{w}}_h, \tilde{\boldsymbol{v}}_h) \\
&= a(\bar{\boldsymbol{v}}_h, \bar{\boldsymbol{v}}_h) + \frac{2}{3}\, a(\tilde{\boldsymbol{v}}_h, \tilde{\boldsymbol{v}}_h) \geq \frac{2}{3}\, a(\boldsymbol{v}_h, \boldsymbol{v}_h)
\end{aligned}
$$

and

$$
a(\boldsymbol{w}_h, \boldsymbol{v}_h) = a(\bar{\boldsymbol{w}}_h, \bar{\boldsymbol{w}}_h) + \frac{3}{2}\, a(\tilde{\boldsymbol{w}}_h, \tilde{\boldsymbol{w}}_h) \geq a(\boldsymbol{w}_h, \boldsymbol{w}_h)
$$

and

$$
\begin{aligned}
a(\boldsymbol{v}_h, \boldsymbol{v}_h) &\leq \frac{3}{2} a(\boldsymbol{w}_h, \boldsymbol{v}_h) = \frac{3}{2} \left(a(\bar{\boldsymbol{w}}_h, \bar{\boldsymbol{w}}_h) + \frac{3}{2}\, a(\tilde{\boldsymbol{w}}_h, \tilde{\boldsymbol{w}}_h) \right) \\
&\leq \frac{9}{4} a(\boldsymbol{w}_h, \boldsymbol{w}_h)
\end{aligned}
$$

With these estimates, we can prove the desired properties for the a form in the discrete case:

Proposition 6 *There exists a constant $\alpha^* > 0$ independent of the mesh size, h, such that*

$$
\inf_{\boldsymbol{w}_h \in \mathcal{K}_2^h} \sup_{\boldsymbol{v}_h \in \mathcal{K}_1^h} \frac{a(\boldsymbol{w}_h, \boldsymbol{v}_h)}{\|\boldsymbol{w}_h\|_{\mathcal{U}^h} \, \|\boldsymbol{v}_h\|_{\mathcal{V}}} \geq \alpha^* \quad . \tag{49}
$$

Furthermore,

$$
\sup_{\boldsymbol{w}_h \in \mathcal{K}_2^h} a(\boldsymbol{w}_h, \boldsymbol{v}_h) > 0 \quad \forall \boldsymbol{v}_h \in \mathcal{K}_1^h \setminus \{0\} \quad . \tag{50}
$$

Proof For $\boldsymbol{w}_h \in \mathcal{K}_2^h$, $\|\boldsymbol{w}_h\|_{\mathcal{U}^h} \neq 0$ it holds $\|\boldsymbol{w}_h\|_{\mathcal{U}^h} = \|\boldsymbol{w}_h\|_{0,\Omega}$. Thus, using the estimates derived from the one-to-one mapping above, for each such $\boldsymbol{w}_h$ we have

$$\sup_{\boldsymbol{v}_h \in \mathcal{K}_1^h} \frac{a(\boldsymbol{w}_h,\boldsymbol{v}_h)}{\|\boldsymbol{v}_h\|_{\mathcal{V}^h}} \geq \frac{a(\boldsymbol{w}_h,\boldsymbol{w}_h)}{\frac{3}{2}\|\boldsymbol{w}_h\|_{0,\Omega}} = \frac{2}{3}\frac{\|\boldsymbol{w}_h\|_{0,\Omega}^2}{\|\boldsymbol{w}_h\|_{0,\Omega}} = \frac{2}{3}\|\boldsymbol{w}_h\|_{\mathcal{U}^h} \quad ,$$

and for $\boldsymbol{v}_h \in \mathcal{K}_1^h \setminus \{0\}$

$$\sup_{\boldsymbol{w}_h \in \mathcal{K}_2^h} a(\boldsymbol{w}_h,\boldsymbol{v}_h) \geq \frac{2}{3} a(\boldsymbol{v}_h,\boldsymbol{v}_h) > 0 \quad .$$

Therefore, the conditions (49) and (50) are satisfied. □

Before the inf-sup condition for the bilinear form b_{2h} is also proved, a *lumping operator* $\Lambda : \mathcal{H}^h \to \mathcal{Q}^h$ is introduced, which is given by

$$\Lambda r_h := \sum_{\bar{C} \in \bar{\mathcal{C}}} \chi_{\bar{C}}\, r_h(x_{\bar{C}}, y_{\bar{C}}) \quad \forall r_h \in \mathcal{H}^h \ ,$$

where $(x_{\bar{C}}, y_{\bar{C}})$ again is the midpoint of $\bar{C}$, i.e. the coordinate of the grid node around which $\bar{C}$ is centered. Thus, in each dual control volume, the value of Λr_h is the value of r_h at the corresponding node in the middle of the control volume. This operator has the following properties, which are proven by Propositions 9 and 10 in the Appendix 6.2:

For $r_h \in \mathcal{H}^h$ with $\nabla r_h \cdot \boldsymbol{n} \equiv 0$ on $\partial\Omega$ we have

$$\|\nabla r_h\|_{0,\Omega}^2 \leq -b_{2h}(\nabla r_h, \Lambda r_h) \ .$$

and

For $r_h \in \mathcal{H}^h$ the estimate

$$\|\Lambda r_h\|_{0,\Omega} \leq C\,\|r_h\|_{0,\Omega}$$

where C is a constant, holds.

Now, we are in the position to prove the inf-sup condition for b_{2h}. The general idea is adapted from a proof of a similar problem in [4].

Proposition 7 *There exists a constant $\beta_2^* > 0$ independent of the mesh size, h, such that*

$$\inf_{q_h \in \mathcal{Q}^h} \sup_{\boldsymbol{w}_h \in \mathcal{U}^h} \frac{b_{2h}(\boldsymbol{w}_h, q_h)}{\|\boldsymbol{w}_h\|_{\mathcal{U}^h}\,\|q_h\|_{\mathcal{Q}}} \geq \beta_2^*$$

Proof To show the inf-sup condition for the b_{2h} form an auxiliary mapping $G_h : \mathcal{Q}^h \to \mathcal{U}^h$ is introduced. It is defined by the solution of the Poisson problem

$$r_h \in \mathcal{H}^h : \quad -\mathsf{L}(r_h) = q_h$$

for $q_h \in \mathcal{Q}^h$, where $G_h q_h := \nabla r_h \in \mathcal{U}^h$. This is to find $r_h \in \mathcal{H}^h$, such that

$$b_{2h}(\nabla r_h, z_h) = (q_h, z_h)_{0,\Omega} \quad \forall z_h \in \mathcal{Q}^h \ . \tag{51}$$

Using the properties of the lumping operator Λ and a not so common version of the Poincaré inequality for H^1 functions (see e.g. [14]), the following estimate can be given for the solution r_h of the Poisson problem (51) (which can be shown to have a unique solution for fixed mesh size h):

$$\begin{aligned}
\|\nabla r_h\|_{0,\Omega}^2 &\le -b_{2h}(\nabla r_h, \Lambda r_h) && \text{(Proposition 9)}\\
&= (q_h, \Lambda r_h)_{0,\Omega} && \text{(Poisson problem (51))}\\
&\le \|q_h\|_{0,\Omega}\,\|\Lambda r_h\|_{0,\Omega} && \text{(Cauchy-Schwarz inequality)}\\
&\le C_1\,\|q_h\|_{0,\Omega}\,\|r_h\|_{0,\Omega} && \text{(Proposition 10)}\\
&\le C_2\,\|q_h\|_{0,\Omega}\,\|\nabla r_h\|_{0,\Omega}\ . && \text{(Poincaré inequality)}
\end{aligned}$$

Thus, we have

$$\|\nabla r_h\|_{0,\Omega} \le C_2\,\|q_h\|_{0,\Omega}\ .$$

Furthermore, this solution satisfies

$$\sup_{z_h\in\mathcal{Q}^h} \frac{b_{2h}(\nabla r_h, z_h)}{\|z_h\|_{\mathcal{Q}}} = \sup_{z_h\in\mathcal{Q}^h} \frac{(q_h, z_h)_{0,\Omega}}{\|z_h\|_{\mathcal{Q}}} = \|q_h\|_{\mathcal{Q}}\ .$$

By the definition of the norm on $\mathcal{U}^h$, it then follows that

$$\|G_h q_h\|_{\mathcal{U}^h} = \|\nabla r_h\|_{0,\Omega} + \sup_{z_h\in\mathcal{Q}^h} \frac{b_{2h}(\nabla r_h, z_h)}{\|z_h\|_{\mathcal{Q}}} \le C\,\|q_h\|_{\mathcal{Q}}$$

where $C = 1 + C_2$, and

$$\|G_h q_h\|_{\mathcal{U}^h}\,\|q_h\|_{\mathcal{Q}} \le C\,\|q_h\|_{\mathcal{Q}}^2 = C\,b_{2h}(\nabla r_h, q_h) = C\,b_{2h}(G_h q_h, q_h)$$

which leads to

$$\frac{1}{C} \le \frac{b_{2h}(G_h q_h, q_h)}{\|G_h q_h\|_{\mathcal{U}^h}\,\|q_h\|_{\mathcal{Q}}} \le \sup_{\boldsymbol{w}_h\in\mathcal{U}^h} \frac{b_{2h}(\boldsymbol{w}_h, q_h)}{\|\boldsymbol{w}_h\|_{\mathcal{U}^h}\,\|q_h\|_{\mathcal{Q}}}$$

Since q_h was chosen arbitrarily, this proves the inf-sup condition for b_{2h}. □

As a summary of this section, we can conclude with the main result of this work:

Theorem 8 *The generalized mixed formulation* (46) *has a unique and stable solution* $((h\boldsymbol{v})^{n+1}, \delta t\, h_0\, h^{(2)})$ *in* $\mathcal{U}^h \times \mathcal{H}^h$.

We have successfully established a mixed formulation equivalent to the second projection of the new scheme. Using this formulation for the stability analysis of the projection step, stability has been shown for the discrete problem. This gives approximations, in which the solution of the Poisson problem $h^{(2)}$ and the momentum update $(h\boldsymbol{v})^{n+1}$ cannot decouple.

In comparison with other projection methods like the ones from [7] and [28], the stability is established by an augmented velocity space in the projection. With this modification, the discrete gradient operator has only a one dimensional kernel. Thus, essentially all modes of the pressure have an influence on the velocity correction. Furthermore, as outlined above, Süli [34] showed that the nine point stencil of the discrete Laplacian results in a stable and robust solution scheme for the Poisson problem. In contrast, the discretizations in [7] and [28] result in gradient operators with higher dimensional kernels. These admit a local decoupling, which is basically the reason for their failure concerning the stability.

4 Numerical Results

To illustrate the performance of the described projection method, the results of two test cases are presented. The main goal is to assess its accuracy and to compare it with a previous version of the method, which rests on standard discretizations for the differential operators used in the projection step and was introduced in [28]. Furthermore, the differences between an exact and an approximate projection formulation are assessed. In the first test case, the second-order convergence of the method is demonstrated for smooth solutions. The second test deals with the translation of a vortex.

For both test cases the exact solution of the particular problem is known, and the error of the numerical approximation can be computed. The computations are performed on a uniform Cartesian grid with equal grid spacing $\delta x = \delta y$. The boundary conditions are those discussed in [41]. So far, we have only investigated the case of constant background height $h_0 \equiv 1$. Thus, in all calculations, the term dh_0/dt is set to zero. To start with initial data, which have zero divergence, i.e.

$$\mathsf{D}(\boldsymbol{v}^0) = -\frac{1}{h_0}\frac{dh_0}{dt}\bigg|_{t=0} = 0\ ,$$

the given values for the momentum are corrected by the solution of the Poisson problem

$$\mathsf{L}(\varphi) = \mathsf{D}\big((h\boldsymbol{v})^{0,r}\big)$$

for $\varphi \in \mathcal{H}^h$. Here, $(h\boldsymbol{v})^{0,r}$ is a linear reconstruction of the exact solution $(h\boldsymbol{v})$ at time $t = 0$. The initial momentum distribution is then given by

$$(h\boldsymbol{v})^0 = (h\boldsymbol{v})^{0,r} - \mathsf{G}(\varphi) \quad .$$

As mentioned earlier, the auxiliary system is solved using an explicit standard second-order Godunov-type method for hyperbolic conservation laws. Since the stability of this method strongly relies on a CFL time step restriction, in all the computations presented in this chapter a time step has been chosen, which is at most $C = 0.8$ of the maximum allowed by the CFL condition.

The discrete divergence and gradient operators, which are used in the two elliptic correction steps, are those given in Appendix 6.1. The linear systems for computing the height $h^{(2)}$ on the primary and on the dual discretizations are solved using the Bi-CGSTAB algorithm [38]. In each iteration, the Euclidean norm

$$\|r_{\mathcal{C}}\|_2 \coloneqq \sqrt{\sum_{C\in\mathcal{C}} r_C^2}$$

(similarly for the second Poisson problem with $\|r_{\bar{C}}\|_2$) of the residual vector

$$\begin{aligned} r_{\mathrm{P1}}\left(h^{(2)}\right) &:= \mathsf{D}((h\boldsymbol{v})^*) - \frac{\delta t}{2}\,\mathsf{D}\left(h_0^{n+1/4}\,\mathsf{G}(h^{(2)})\right) \\ r_{\mathrm{P2}}\left(h^{(2)}\right) &:= \mathsf{D}((h\boldsymbol{v})^{**}) + \mathsf{D}((h\boldsymbol{v})^n) - \delta t\,\mathsf{D}\left(h_0^{n+1/2}\,\mathsf{G}(h^{(2)})\right) \end{aligned}$$

is calculated. The algorithm is terminated when either this absolute value or the ratio between the norm of the current residual and that of the initial residual is less than 10^{-11}.

4.1 Convergence study

The first test case demonstrates the second-order convergence of numerical solutions to the exact solution for smooth data. This test, which involves a Taylor vortex being translated at a constant speed, was originally proposed in [23] and [1] for the incompressible flow equations. Here it has been adapted for the zero Froude number shallow water equations.

For constant height h_0 and an initial velocity distribution

$$\begin{aligned} u_0(x,y) &= 1 - 2\cos(2\pi x)\sin(2\pi y) \\ v_0(x,y) &= 1 + 2\sin(2\pi x)\cos(2\pi y) \quad , \end{aligned}$$

the exact solution of the zero Froude number shallow water equations is given by

$$\begin{aligned} u(x,y,t) &= 1 - 2\cos(2\pi(x-t))\sin(2\pi(y-t)) \\ v(x,y,t) &= 1 + 2\sin(2\pi(x-t))\cos(2\pi(y-t)) \\ h^{(2)}(x,y,t) &= -\cos(4\pi(x-t)) - \cos(4\pi(y-t)) \quad . \end{aligned}$$

The problem is solved on the unit square with $(x,y) \in [0,1]^2$ and periodic boundary conditions. It describes the advection of four vortices in the $(1,1)$ direction. The piecewise linear reconstruction of the momentum field components is done using central differences with no slope limiter.

The numerical solution is computed on three different grids with 32×32, 64×64 and 128×128 cells. We start the calculation at $t = 0$, and the error vector in the velocity $\boldsymbol{e}^N$ with elements

$$e_{i,j}^N := \left|\overline{u(x,y,t^N)}^{C_{i,j}} - u_{i,j}^N\right| + \left|\overline{v(x,y,t^N)}^{C_{i,j}} - v_{i,j}^N\right|$$

is evaluated at time $t^N = 3$. This corresponds to 750, 1500 and 3000 time steps, respectively. Note that we could have also incorporated the linear variation of the velocity on each grid cell in the error analysis of the new projection. We do not choose this alternative in favor of a better comparison with the original method. The global error is measured using a discrete L^2 norm and the L^∞ norm. These are defined by

$$\|\boldsymbol{e}^N\|_0 := \left(\sum_{i,j} |C_{i,j}|\,|e_{i,j}^N|^2\right)^{1/2} \qquad \text{and} \qquad \|\boldsymbol{e}^N\|_\infty := \max_{i,j}\{e_{i,j}^N\} \quad .$$

Table 1 Errors and convergence rates for the different projection methods.

Method	Norm	32x32	Rate γ	64x64	Rate γ	128x128
Projection from [28]	L^2	0.292096	2.16	0.065415	2.17	0.014566
	L^∞	0.419370	2.16	0.094106	2.18	0.020747
Approximate projection	L^2	0.291967	2.16	0.065412	2.17	0.014566
	L^∞	0.419130	2.16	0.094098	2.18	0.020747
Exact projection	L^2	0.082379	2.65	0.013129	2.23	0.002796
	L^∞	0.126207	2.46	0.022999	2.33	0.004573

We have summarized these error measures for the aforementioned previous version of the projection method [28] as well as for the approximate and the exact projection methods in Table 1. Here, the “approximate projection method” utilizes the same stencil as the exact projection method, but it leaves slope computations for the in-cell distributions of momentum entirely to classical slope limiting procedures instead of letting several components of these derivatives be determined by the projection step.

Additionally, the corresponding convergence rate γ is given, which is calculated by

$$\gamma := \frac{\log(\|\boldsymbol{e}_c^N\|/\|\boldsymbol{e}_f^N\|)}{\log(\delta x_c/\delta x_f)} \quad . \tag{52}$$

In this definition, $\boldsymbol{e}_c^N$ and $\boldsymbol{e}_f^N$ are the computed error vectors of the solution on the coarse and the fine grid and δx_c and δx_f are the corresponding grid spacings. Clearly, second order accuracy is obtained in the L^2 as well as in the L^∞ norm. Also note that the absolute error obtained with the exact projection is about four times smaller than the one obtained with the approximate projection method and with the scheme from [28].

4.2 Advection of a vortex

Let us consider the advection of a vortex by a constant background flow. For the implementation of this test case, originally proposed in [17], a rectangular domain with size $[0,4]\times[0,1]$ is examined. The domain has periodic boundary conditions at the short sides and walls at the long sides. The initial conditions are defined to be

$$u_0(x,y) = 1 - v_\theta(r)\sin\theta \quad \text{and} \quad v_0(x,y) = v_\theta(r)\cos\theta \quad ,$$

in which

$$v_\theta(r) = \begin{cases} 5r\,v_{\max} & \text{for } 0 \le r < \frac{1}{5} \\ (2-5r)\,v_{\max} & \text{for } \frac{1}{5} \le r < \frac{2}{5} \\ 0 & \text{for } \frac{2}{5} \le r \end{cases} \tag{53}$$

and

$$r = \sqrt{\left(x-\tfrac{1}{2}\right)^2 + \left(y-\tfrac{1}{2}\right)^2} \quad .$$

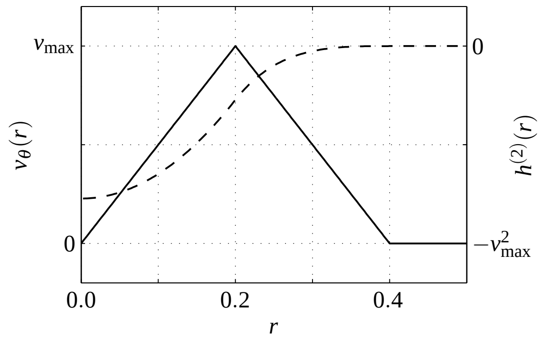

Fig. 5 Advection of a vortex: tangential velocity (*solid line*) and height profile (*dashed line*) with respect to the distance r from the center of the vortex.

In equation (53) $v_{\max}$ is the maximum tangential velocity of the vortex. The height $h^{(2)}$ must then satisfy the constraint $\partial_r h^{(2)} = v_\theta^2/r$. This relationship is visualized in Figure 5.

The test is set up with $v_{\max} = 1$ and background height $h_0 \equiv 1$. The computational domain consists of 80×20 grid cells. Three different strategies for the linear reconstruction of the components in the momentum variable are investigated. In particular, we consider central differences (no limiter), the *monotonized central difference (MC)* limiter and *Sweby's* limiter [31] with $k = 1.8$, the latter being a convex combination of the *minmod* ($k = 1$) and the *superbee* limiter ($k = 2$).

For comparison, the results for the scheme from [28] are given in Figure 6, in which the stream function of the velocity distribution is displayed at four different times of the simulation. Similar to the results in [28] for the incompressible Euler equations, the core is advected almost along the center line of the channel. Also, the vortex experiences a considerable deformation due to the coarse discretization we have chosen for this test.

As in the convergence studies, the new exact projection method shows some improvement in the numerical results for this test (cf. Figure 7). All reconstruction strategies show less deviation from the center line of the channel than in the original method. Furthermore, the loss in vorticity is slightly reduced. Again, the results of the approximate projection method (not shown) are comparable to the ones obtained by the method from [28].

The presented numerical test cases are primarily given to show the applicability of the presented method. They are not designed to highlight the improved stability properties of the method. To demonstrate that the described projection method provides significant gains over existing schemes is a matter of more extensive studies and goes beyond the scope of this paper.

5 Conclusions

In this paper, we demonstrate that it is possible to formulate a finite volume projection method for incompressible flows with an exact *and* stable projection step. No further stabilization techniques are required to prevent a velocity-pressure decoupling, which is often observed in former exact projection methods. This is

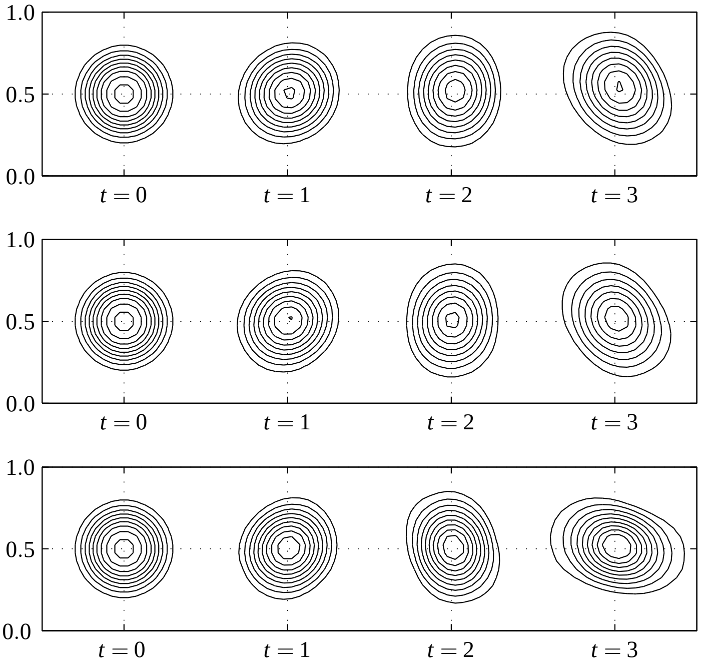

Fig. 6 Advection of a vortex at times $t = 0, 1, 2$ and 3 for the method by [28]. Contour lines of the stream function are shown at [-0.02, -0.04, ..., -0.18] starting from outside of the vortex. *Top:* unlimited slopes, *middle:* monotonized central difference (MC) limiter, *bottom:* Sweby's limiter ($k = 1.8$).

achieved by using a Petrov-Galerkin finite element discretization of the associated Poisson problem, originally proposed in [34]. Furthermore, the method locally conserves mass and momentum.

In order to prove stability of the second projection step, which corrects the cell-centered momentum to be in compliance with the divergence constraint, we have used the theory of mixed finite element methods, the latter providing strong results about the stability of discretizations. This technique is well known from finite element methods for *viscous* incompressible flows, where the Laplacian of the velocity field interacts with the pressure gradient. Here, the theory is applied in the case of a finite volume method for *inviscid* incompressible flows, which means that the velocity directly interacts with the pressure gradient.

The numerical results, obtained from the application of the new method, show practical accuracy improvements on fixed grids compared to the method presented in [28], with both methods being second order accurate. The discretization for the new projection can be also used for the first projection of the method, yielding a unified discretization for both Poisson-type problems. Furthermore, the linear systems associated with the Poisson equations can be solved with the same algorithms that are used for standard second order discretizations of the differential operators.

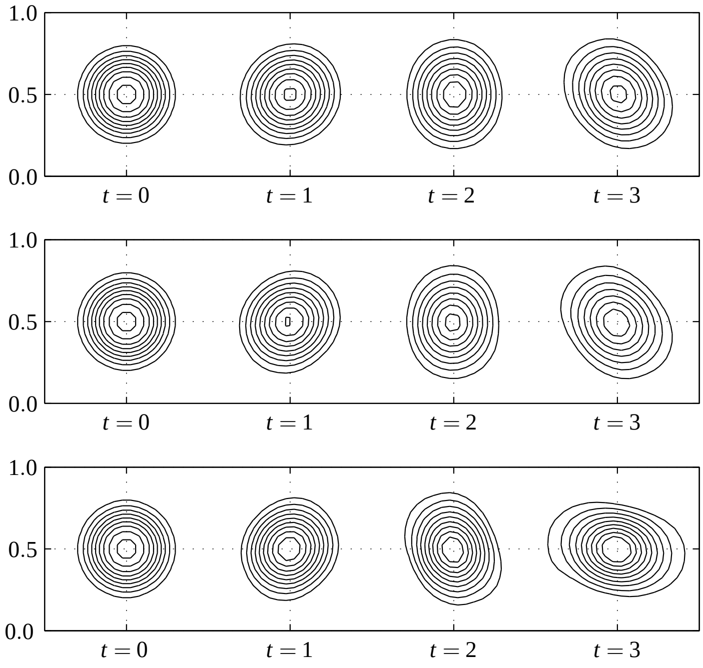

Fig. 7 Same as Figure 6 for the new exact projection method.

However, there are still some open questions, and the analysis of them is ongoing research. It was mentioned that the second projection adjusts the piecewise linear portions of the momentum field, which, in turn, results in a possible loss of the TVD property of the whole method. This is a delicate issue, because it concerns the stability of the predictor step. So far, we have only numerical evidence that this still results in stable approximations.

In the present paper, it was only proved that the projection step yields a stable approximation and does not admit any pressure-velocity decoupling. One of the next steps is to investigate the convergence of this mixed formulation to the continuous solution (mentioned in Theorem 2). Furthermore, it would be desirable to extend our results to zero Mach number variable density flows, where the Projection results in a Poisson-type problem with a weighted Laplace operator.

The overall motivation for this work stems from meteorological and combustion applications. The solution of such problems requires large scale computation techniques such as locally refined meshes. Coupling the presented method to such technologies is ongoing research. Furthermore, one has small, but non-zero Mach numbers (resp. Froude numbers) in these problems. The extension of the current method to allow for smooth transitions from fully compressible to zero Froude number flows would hopefully yield favorable results for these application areas. Such attempts were already reported in Klein [21] for one space dimension and in Geratz [13] and Munz et al. [24] for higher dimensional problems. We are planning to advance the ideas outlined in these references.

Acknowledgements The authors are grateful to Eberhard Bänsch and Nicola Botta for helpful discussions and remarks that improved the content of the paper. Furthermore, Nicola Botta provided the software environment (the compressible flow solver), in which the algorithm was implemented and the presented calculations were accomplished. This work benefitted greatly from free software products. Without these tools – such as LaTeX, the GNU C/C++ compiler and the Linux operating system – a lot of tasks would not have been so easy to realize. It is our pleasure to thank all developers for their excellent products.

The authors thank Deutsche Forschungsgemeinschaft for their partial support of this work through grants KL 611/6 and KL 611/14.

6 Appendix

6.1 Discretization of the new projection

Here, the discrete gradient, divergence and Laplacian of the second projection are given for a two-dimensional Cartesian grid with constant grid spacings δx and δy. The operators for the first projection are derived by shifting the indices by one half. The double index (i,j) is used to refer to a cell value, while the index $(i+1/2, j+1/2)$ is used for node values.

Let us define

$$\begin{aligned}
p_{x,i,j} &:= \frac{1}{2\,\delta x}\left(p_{i+1/2,j+1/2} - p_{i-1/2,j+1/2} + p_{i+1/2,j-1/2} - p_{i-1/2,j-1/2}\right)\\
p_{y,i,j} &:= \frac{1}{2\,\delta y}\left(p_{i+1/2,j+1/2} - p_{i+1/2,j-1/2} + p_{i-1/2,j+1/2} - p_{i-1/2,j-1/2}\right)\\
p_{xy,i,j} &:= \frac{1}{\delta x\,\delta y}\left(p_{i+1/2,j+1/2} - p_{i-1/2,j+1/2} - p_{i+1/2,j-1/2} + p_{i-1/2,j-1/2}\right)\ .
\end{aligned}$$

The discrete gradient G is then given by

$$\mathsf{G}\,(p)\,|_{C_{i,j}} = \begin{pmatrix} p_{x,i,j} \\ p_{y,i,j} \end{pmatrix} + \begin{pmatrix} y - y_j \\ x - x_i \end{pmatrix} p_{xy,i,j} \quad .$$

The divergence D is defined by

$$\begin{aligned}
\mathsf{D}(\boldsymbol{v})|_{\bar{C}_{i+1/2,j+1/2}} = &\frac{1}{2\,\delta x}\left(u_{i+1,j+1} - u_{i,j+1} + u_{i+1,j} - u_{i,j}\right)\\
&+ \frac{\delta y}{8\,\delta x}\left(-u_{y,i+1,j+1} + u_{y,i,j+1} + u_{y,i+1,j} - u_{y,i,j}\right)\\
&+ \frac{1}{2\,\delta y}\left(v_{i+1,j+1} - v_{i+1,j} + v_{i,j+1} - v_{i,j}\right)\\
&+ \frac{\delta x}{8\,\delta y}\left(-v_{x,i+1,j+1} + v_{x,i+1,j} + v_{x,i,j+1} - v_{x,i,j}\right)\ .
\end{aligned}$$

With the above definitions $\mathsf{D}(\mathsf{G}(\cdot))$ is the 9-points Laplacian proposed by [34] (cf. Figure 3):

$$\mathsf{L}(p)|_{\bar{C}_{i+1/2,j+1/2}} = \mathsf{D}(\mathsf{G}(p))|_{\bar{C}_{i+1/2,j+1/2}}$$

$$
\begin{aligned}
&= \frac{1}{8}\left(\triangle_{xx,i+1/2,j+3/2}(p) + 6\triangle_{xx,i+1/2,j+1/2}(p) + \triangle_{xx,i+1/2,j-1/2}(p)\right) \\
&\quad + \frac{1}{8}\left(\triangle_{yy,i+3/2,j+1/2}(p) + 6\triangle_{yy,i+1/2,j+1/2}(p) + \triangle_{yy,i-1/2,j+1/2}(p)\right)
\end{aligned}
$$

with

$$
\begin{aligned}
\triangle_{xx,i+1/2,j+1/2}(p) &\coloneqq \frac{1}{\delta x^2}\left(p_{i+3/2,j+1/2} - 2p_{i+1/2,j+1/2} + p_{i-1/2,j+1/2}\right) \\
\triangle_{yy,i+1/2,j+1/2}(p) &\coloneqq \frac{1}{\delta y^2}\left(p_{i+1/2,j+3/2} - 2p_{i+1/2,j+1/2} + p_{i+1/2,j-1/2}\right) \ .
\end{aligned}
$$

6.2 Properties of the Lumping-Operator Λ

Proposition 9 *For $r_h \in \mathcal{H}^h$ with $\nabla r_h \cdot \boldsymbol{n} \equiv 0$ on $\partial\Omega$ we have*

$$
\|\nabla r_h\|_{0,\Omega}^2 \leq -b_{2h}(\nabla r_h, \Lambda r_h)
$$

Proof Let us consider a cell-wise representation of r_h, i.e. on a control volume $C_{i,j}$ of the primary discretization r_h can be also represented by

$$
r_h(x,y)|_{C_{i,j}} = r_{i,j} + (x-x_i)r_{x,i,j} + (y-y_j)r_{y,i,j} + (x-x_i)(y-y_j)r_{xy,i,j} \quad ,
$$

in which $r_{i,j}$ is the mean value of r_h on $C_{i,j}$, and $r_{x,i,j}$, $r_{y,i,j}$ and $r_{xy,i,j}$ are the partial and mixed derivatives of r_h in (x_i, y_j), respectively. With this definition, we have

$$
\begin{aligned}
[\nabla r_h(x,y)]^2|_{C_{i,j}} &= r_{x,i,j}^2 + 2(y-y_j)r_{x,i,j}\, r_{xy,i,j} + (y-y_j)^2 r_{xy,i,j}^2 \\
&\quad + r_{y,i,j}^2 + 2(x-x_i)r_{y,i,j}\, r_{xy,i,j} + (x-x_i)^2 r_{xy,i,j}^2
\end{aligned}
$$

Furthermore, we obtain

$$
\begin{aligned}
\|\nabla r_h\|_{0,\Omega}^2 &= \sum_{i,j} \int_{C_{i,j}} [\nabla r_h]^2 dx \\
&= \delta x\, \delta y \sum_{i,j} \left(r_{x,i,j}^2 + r_{y,i,j}^2 + \frac{\delta x^2 + \delta y^2}{12} r_{xy,i,j}^2 \right)
\end{aligned}
$$

To compare this result with the expression in the b_{2h} form, the bilinear form has to be written as sum over the primary cells. Using partial summation, this leads to

$$
\begin{aligned}
b_{2h}(\nabla r_h, \Lambda r_h) = r_{1/2,1/2} &\left[\frac{\delta y}{2} r_{x,1,1} + \frac{\delta x}{2} r_{y,1,1} - \frac{\delta x^2 + \delta y^2}{8} r_{xy,1,1} \right] \\
&+ \sum_{j=1}^{n-1} r_{1/2,j+1/2} \left[\frac{\delta y}{2}(r_{x,1,j+1} + r_{x,1,j}) + \frac{\delta x}{2}(r_{y,1,j+1} - r_{y,1,j}) + \frac{\delta x^2 + \delta y^2}{8}(-r_{xy,1,j+1} + r_{xy,1,j}) \right] \\
&+ r_{1/2,n+1/2} \left[\frac{\delta y}{2} r_{x,1,n} - \frac{\delta x}{2} r_{y,1,n} + \frac{\delta x^2 + \delta y^2}{8} r_{xy,1,n} \right]
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{i=1}^{m-1}\Bigg(r_{i+1/2,1/2}\Bigg[\frac{\delta y}{2}(r_{x,i+1,1}-r_{x,i,1})+\frac{\delta x}{2}(r_{y,i+1,1}+r_{y,i,1})+\frac{\delta x^2+\delta y^2}{8}(-r_{xy,i+1,1}+r_{xy,i,1})\Bigg]\\
&\qquad+\sum_{j=1}^{n-1}r_{i+1/2,j+1/2}\Bigg[\frac{\delta y}{2}(r_{x,i+1,j+1}-r_{x,i,j+1}+r_{x,i+1,j}-r_{x,i,j})\\
&\qquad\qquad\qquad+\frac{\delta x}{2}(r_{y,i+1,j+1}+r_{y,i,j+1}-r_{y,i+1,j}-r_{y,i,j})\\
&\qquad\qquad\qquad+\frac{\delta x^2+\delta y^2}{8}(-r_{xy,i+1,j+1}+r_{xy,i,j+1}+r_{xy,i+1,j}-r_{xy,i,j})\Bigg]\\
&\qquad+r_{i+1/2,n+1/2}\Bigg[\frac{\delta y}{2}(r_{x,i+1,n}-r_{x,i,n})+\frac{\delta x}{2}(-r_{y,i+1,n}-r_{y,i,n})+\frac{\delta x^2+\delta y^2}{8}(r_{xy,i+1,n}-r_{xy,i,n})\Bigg]\Bigg)\\
&+r_{m+1/2,1/2}\Bigg[-\frac{\delta y}{2}r_{x,m,1}+\frac{\delta x}{2}r_{y,m,1}+\frac{\delta x^2+\delta y^2}{8}r_{xy,m,1}\Bigg]\\
&+\sum_{j=1}^{n-1}r_{m+1/2,j+1/2}\Bigg[\frac{\delta y}{2}(-r_{x,m,j+1}-r_{x,m,j})+\frac{\delta x}{2}(r_{y,m,j+1}-r_{y,m,j})+\frac{\delta x^2+\delta y^2}{8}(r_{xy,m,j+1}-r_{xy,m,j})\Bigg]\\
&+r_{m+1/2,n+1/2}\Bigg[-\frac{\delta y}{2}r_{x,m,n}-\frac{\delta x}{2}r_{y,m,n}-\frac{\delta x^2+\delta y^2}{8}r_{xy,m,n}\Bigg]\\
=&\sum_{j=1}^{n}\Bigg[\frac{\delta y}{2}r_{x,1,j}(r_{1/2,j-1/2}+r_{1/2,j+1/2})+\frac{\delta x}{2}r_{y,1,j}(r_{1/2,j-1/2}-r_{1/2,j+1/2})\\
&\qquad+\frac{\delta x^2+\delta y^2}{8}r_{xy,1,j}(-r_{1/2,j-1/2}+r_{1/2,j+1/2})\Bigg]\\
&+\sum_{i=1}^{m-1}\Bigg(\sum_{j=1}^{n}\Bigg[\frac{\delta y}{2}r_{x,i,j}(-r_{i+1/2,j-1/2}-r_{i+1/2,j+1/2})+\frac{\delta y}{2}r_{x,i+1,j}(r_{i+1/2,j-1/2}+r_{i+1/2,j+1/2})\\
&\qquad\qquad+\frac{\delta x}{2}r_{y,i,j}(r_{i+1/2,j-1/2}-r_{i+1/2,j+1/2})+\frac{\delta x}{2}r_{y,i+1,j}(r_{i+1/2,j-1/2}-r_{i+1/2,j+1/2})\\
&\qquad\qquad+\frac{\delta x^2+\delta y^2}{8}r_{xy,i,j}(r_{i+1/2,j-1/2}-r_{i+1/2,j+1/2})\\
&\qquad\qquad+\frac{\delta x^2+\delta y^2}{8}r_{xy,i+1,j}(-r_{i+1/2,j-1/2}+r_{i+1/2,j+1/2})\Bigg]\Bigg)\\
&+\sum_{j=1}^{n}\Bigg[\frac{\delta y}{2}r_{x,m,j}(-r_{m+1/2,j-1/2}-r_{m+1/2,j+1/2})+\frac{\delta x}{2}r_{y,m,j}(r_{m+1/2,j-1/2}-r_{m+1/2,j+1/2})\\
&\qquad+\frac{\delta x^2+\delta y^2}{8}r_{xy,m,j}(r_{m+1/2,j-1/2}-r_{m+1/2,j+1/2})\Bigg]\\
=&\sum_{i=1}^{m}\sum_{j=1}^{n}\Bigg[\frac{\delta y}{2}r_{x,i,j}(-r_{i+1/2,j-1/2}-r_{i+1/2,j+1/2}+r_{i-1/2,j-1/2}+r_{i-1/2,j+1/2})\\
&\qquad+\frac{\delta x}{2}r_{y,i,j}(r_{i+1/2,j-1/2}-r_{i+1/2,j+1/2}+r_{i-1/2,j-1/2}-r_{i-1/2,j+1/2})\\
&\qquad+\frac{\delta x^2+\delta y^2}{8}r_{xy,i,j}(r_{i+1/2,j-1/2}-r_{i+1/2,j+1/2}-r_{i-1/2,j-1/2}+r_{i-1/2,j+1/2})\Bigg]\\
=&\sum_{i=1}^{m}\sum_{j=1}^{n}\Bigg[\frac{\delta y}{2}r_{x,i,j}(-2\delta x\, r_{x,i,j})+\frac{\delta x}{2}r_{y,i,j}(-2\delta y\, r_{y,i,j})+\frac{\delta x^2+\delta y^2}{8}r_{xy,i,j}(-\delta x\delta y\, r_{xy,i,j})\Bigg]\\
=&-\delta x\,\delta y\sum_{i,j}\left(r_{x,i,j}^2+r_{y,i,j}^2+\frac{\delta x^2+\delta y^2}{8}r_{xy,i,j}^2\right)
\end{aligned}
$$

These results lead to the desired estimate:

$$
\|\nabla r_h\|_{0,\Omega}^2=\delta x\,\delta y\sum_{i,j}\left(r_{x,i,j}^2+r_{y,i,j}^2+\frac{\delta x^2+\delta y^2}{12}r_{xy,i,j}^2\right)
$$

$$\leq \delta x\, \delta y \sum_{i,j} \left(r_{x,i,j}^2 + r_{y,i,j}^2 + \frac{\delta x^2 + \delta y^2}{8} r_{xy,i,j}^2 \right)$$
$$= -b_{2h}(\nabla r_h, \Lambda r_h)$$

□

Proposition 10 *For $r_h \in \mathcal{H}^h$ the estimate*

$$\|\Lambda r_h\|_{0,\Omega} \leq C\, \|r_h\|_{0,\Omega}$$

where C is a constant, is true.

Proof Since r_h is piecewise bilinear, its L^2-norm can be rewritten as

$$\begin{aligned}
\|r_h\|_{0,\Omega}^2 &= \int_{\Omega} r_h^2 \, dx \\
&= \sum_{i,j} \int_{C_{i,j}} [r_{i,j} + (x-x_i) r_{x,i,j} + (y-y_j) r_{y,i,j} + (x-x_i)(y-y_j) r_{xy,i,j}]^2 \, dx \\
&= \sum_{i,j} \int_{C_{i,j}} \left[r_{i,j}^2 + (x-x_i)^2 r_{x,i,j}^2 + (y-y_j)^2 r_{y,i,j}^2 + (x-x_i)^2 (y-y_j)^2 r_{xy,i,j}^2 \right] dx \\
&= \delta x \delta y \sum_{i,j} \left[r_{i,j}^2 + \frac{\delta x^2}{3 \cdot 4} r_{x,i,j}^2 + \frac{\delta y^2}{3 \cdot 4} r_{y,i,j}^2 + \frac{\delta x^2\, \delta y^2}{9 \cdot 16} r_{xy,i,j}^2 \right] \\
&= \delta x \delta y \sum_{i,j} \left[\frac{1}{16} (r_{i+1/2,j+1/2} + r_{i+1/2,j-1/2} + r_{i-1/2,j+1/2} + r_{i-1/2,j-1/2})^2 \right. \\
&\qquad + \frac{1}{48} (r_{i+1/2,j+1/2} + r_{i+1/2,j-1/2} - r_{i-1/2,j+1/2} - r_{i-1/2,j-1/2})^2 \\
&\qquad + \frac{1}{48} (r_{i+1/2,j+1/2} - r_{i+1/2,j-1/2} + r_{i-1/2,j+1/2} - r_{i-1/2,j-1/2})^2 \\
&\qquad \left. + \frac{1}{144} (r_{i+1/2,j+1/2} - r_{i+1/2,j-1/2} - r_{i-1/2,j+1/2} + r_{i-1/2,j-1/2})^2 \right] \\
&= \delta x \delta y \sum_{i,j} \left[\frac{1}{9} (r_{i+1/2,j+1/2}^2 + r_{i+1/2,j-1/2}^2 + r_{i-1/2,j+1/2}^2 + r_{i-1/2,j-1/2}^2) \right. \\
&\qquad + \frac{1}{9} (r_{i+1/2,j+1/2} r_{i+1/2,j-1/2} + r_{i+1/2,j+1/2} r_{i-1/2,j+1/2} \\
&\qquad\quad + r_{i+1/2,j-1/2} r_{i-1/2,j-1/2} + r_{i-1/2,j+1/2} r_{i-1/2,j-1/2}) \\
&\qquad \left. + \frac{1}{18} (r_{i+1/2,j+1/2} r_{i-1/2,j-1/2} + r_{i+1/2,j-1/2} r_{i-1/2,j+1/2}) \right]
\end{aligned}$$

$$= \frac{\delta x \delta y}{18} \sum_{i,j} \Big[(r^2_{i+1/2,j+1/2} + r^2_{i+1/2,j-1/2} + r^2_{i-1/2,j+1/2} + r^2_{i-1/2,j-1/2})$$

$$+ (r_{i+1/2,j+1/2} + r_{i+1/2,j-1/2} + r_{i-1/2,j+1/2} + r_{i-1/2,j-1/2})^2$$

$$- (r_{i+1/2,j+1/2} r_{i-1/2,j-1/2} + r_{i+1/2,j-1/2} r_{i-1/2,j+1/2}) \Big]$$

Since

$$r_{i+1/2,j+1/2}\, r_{i-1/2,j-1/2} + r_{i+1/2,j-1/2}\, r_{i-1/2,j+1/2} \le \frac{1}{2} \left(r^2_{i+1/2,j+1/2} + r^2_{i+1/2,j-1/2} + r^2_{i-1/2,j+1/2} + r^2_{i-1/2,j-1/2} \right)$$

it follows that

$$\|r_h\|^2_{0,\Omega} \ge \frac{\delta x \delta y}{18} \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{1}{2} (r^2_{i+1/2,j+1/2} + r^2_{i+1/2,j-1/2} + r^2_{i-1/2,j+1/2} + r^2_{i-1/2,j-1/2})$$

$$= \frac{\delta x \delta y}{36} \Big[r^2_{1/2,1/2} + \sum_{j=1}^{n-1} 2 r^2_{1/2,j+1/2} + r^2_{1/2,n+1/2}$$

$$+ 2 \sum_{i=1}^{m-1} \left(r^2_{i+1/2,1/2} + \sum_{j=1}^{n-1} 2 r^2_{i+1/2,j+1/2} + r^2_{i+1/2,n+1/2} \right)$$

$$+ r^2_{m+1/2,1/2} + \sum_{j=1}^{n-1} 2 r^2_{m+1/2,j+1/2} + r^2_{m+1/2,n+1/2} \Big]$$

$$= \frac{1}{9} \|\Lambda r_h\|^2_{0,\Omega}$$

□

References

1. Almgren, A.S., Bell, J.B., Colella, P., Howell, L.H., Welcome, M.L.: A conservative adaptive projection method for the variable density incompressible Navier-Stokes equations. Journal of Computational Physics **142**(1), 1–46 (1998) 26
2. Almgren, A.S., Bell, J.B., Crutchfield, W.Y.: Approximate projection methods: Part I. inviscid analysis. SIAM Journal on Scientific Computing **22**(4), 1139–1159 (2000) 2
3. Almgren, A.S., Bell, J.B., Szymczak, W.G.: A numerical method for the incompressible Navier-Stokes equations based on an approximate projection. SIAM Journal on Scientific Computing **17**(2), 358–369 (1996) 2
4. Angermann, L.: Node-centered finite volume schemes and primal-dual mixed formulations. Communications in Applied Analysis **7**(4), 529–566 (2003) 15, 18, 23
5. Babuška, I.: Error-bounds for finite element method. Numerische Mathematik **16**, 322–333 (1971) 2, 13
6. Bell, J.B., Colella, P., Glaz, H.M.: A second-order projection method for the incompressible Navier-Stokes equations. Journal of Computational Physics **85**(2), 257–283 (1989) 2
7. Bell, J.B., Marcus, D.L.: A second-order projection method for variable-density flows. Journal of Computational Physics **101**, 334–348 (1992) 2, 8, 24, 25

8. Bernardi, C., Canuto, C., Maday, Y.: Generalized inf-sup conditions for the Chebyshev spectral approximation of the Stokes problem. SIAM Journal on Numerical Analysis **25**(6), 1237–1271 (1988) 14, 15
9. Braess, D.: Finite Elemente: Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie, 3 edn. Springer, Berlin (2003) 19
10. Brezzi, F.: On the existence, uniqueness and approximation of saddle point problems arising from Lagrangian multipliers. RAIRO Analyse numérique **8**, 129–151 (1974) 2, 13
11. Chorin, A.J.: Numerical solution of the Navier-Stokes equations. Mathematics of Computation **22**(104), 745–762 (1968) 2
12. Courant, R., Friedrichs, K.O., Lewy, H.: Über die partiellen Differenzengleichungen der mathematischen Physik. Mathematische Annalen **100**, 32–74 (1928) 6
13. Geratz, K.J.: Erweiterung eines Godunov-Typ-Verfahrens für zwei-dimensionale kompressible Strömungen auf die Fälle kleiner und verschwindender Machzahl. PhD dissertation, Rheinisch-Westfälische Technische Hochschule Aachen (1997) 30
14. Gilbarg, D., Trudinger, N.S.: Elliptic Partial Differential Equations of Second Order, 2 edn. Springer (1983) 24
15. Girault, V., Raviart, P.A.: Finite Element Methods for Navier-Stokes Equations, *Springer Series in Computational Mathematics*, vol. 5. Springer, Berlin (1986) 14
16. Gresho, P.M.: On the theory of semi-implicit projection methods for viscous incompressible flow and its implementation via a finite element method that also introduces a nearly consistent mass matrix. Part 1: Theory. International Journal for Numerical Methods in Fluids **11**(5), 587–620 (1990) 7
17. Gresho, P.M., Chan, S.T.: On the theory of semi-implicit projection methods for viscous incompressible flow and its implementation via a finite element method that also introduces a nearly consistent mass matrix. Part 2: Implementation. International Journal for Numerical Methods in Fluids **11**(5), 621–659 (1990) 27
18. Guermond, J.L., Minev, P., Shen, J.: An overview of projection methods for incompressible flows. Computer Methods in Applied Mechanics and Engineering **195**(44–47), 6011–6045 (2006) 2
19. Harlow, F.H., Welch, J.E.: Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface. The Physics of Fluids **8**(12), 2182–2189 (1965) 6
20. Klainerman, S., Majda, A.: Singular limits of quasilinear hyperbolic systems with large parameters and the incompressible limit of compressible fluids. Communications in Pure Applied Mathematics **34**, 481–524 (1981) 2, 3
21. Klein, R.: Semi-implicit extension of a Godunov-Type scheme based on low Mach number asymptotics I: One-dimensional flow. Journal of Computational Physics **121**, 213–237 (1995) 2, 3, 30
22. LeVeque, R.J.: Finite Volume Methods for Hyperbolic Problems, *Cambridge Texts in Applied Mathematics*, vol. 31. Cambridge University Press, Cambridge (2002) 5
23. Minion, M.L.: A projection method for locally refined grids. Journal of Computational Physics **127**(1), 158–178 (1996) 26
24. Munz, C.D., Roller, S., Klein, R., Geratz, K.J.: The extension of incompressible flow solvers to the weakly compressible regime. Computers & Fluids **32**(2), 173–196 (2003) 30
25. Nicolaïdes, R.A.: Existence, uniqueness and approximation for generalized saddle point problems. SIAM Journal on Numerical Analysis **19**(2), 349–357 (1982) 14, 15
26. Oevermann, M., Klein, R.: A cartesian grid finite volume method for elliptic equations with variable coefficients and embedded interfaces. Journal of Computational Physics **219**(2), 749–769 (2006) 3
27. Osher, S.: Convergence of generalized MUSCL schemes. SIAM Journal on Numerical Analysis **22**(5), 947–961 (1985) 6
28. Schneider, T., Botta, N., Geratz, K.J., Klein, R.: Extension of finite volume compressible flow solvers to multi-dimensional, variable density zero Mach number flows. Journal of Computational Physics **155**, 248–286 (1999) 2, 3, 4, 7, 8, 24, 25, 27, 28, 29
29. Schochet, S.: Fast singular limits of hyperbolic PDEs. Journal of Differential Equations **114**(2), 476–512 (1994) 2
30. Schochet, S.: The mathematical theory of low Mach number flows. RAIRO Modélisation Mathématique et Analyse Numérique **39**(3), 441–458 (2005) 2
31. Schulz-Rinne, C.W.: The Riemann problem for two-dimensional gas dynamics and new limiters for high-order schemes. PhD dissertation, Eidgenössische Technische Fachhochschule (ETH) Zürich (1993). Diss. ETH No. 10297 28

32. Shu, C.W., Osher, S.: Efficient implementation of essentially non-oscillatory shock-capturing schemes. Journal of Computational Physics **77**(2), 439–471 (1988) 6
33. Strang, G.: On the construction and comparison of difference schemes. SIAM Journal on Numerical Analysis **5**(3), 506–517 (1968) 6
34. Süli, E.: Convergence of finite volume schemes for Poisson's equation on nonuniform meshes. SIAM Journal on Numerical Analysis **28**(5), 1419–1430 (1991) 2, 7, 25, 29, 31
35. Temam, R.: Une méthode d'approximation de la solution des équations de Navier-Stokes. Bulletin de la Société Mathématique de France **96**, 115–152 (1968) 2
36. Thomas, J.M., Trujillo, D.: Mixed finite volume methods. International Journal for Numerical Methods in Engineering **46**(9), 1351–1366 (1999) 18
37. van Albada, G.D., van Leer, B., Roberts Jr., W.W.: A comparative study of computational methods in cosmic gas dynamics. Astronomy and Astrophysics **108**(1), 76–84 (1982) 6
38. van der Vorst, H.A.: Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems. SIAM Journal on Scientific and Statistical Computing **13**(2), 631–644 (1992) 25
39. van Kan, J.: A second-order accurate pressure-correction scheme for viscous incompressible flow. SIAM Journal on Scientific and Statistical Computing **7**(3), 870–891 (1986) 2
40. van Leer, B.: Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method. Journal of Computational Physics **32**(1), 101–136 (1979) 6
41. Vater, S.: A new projection method for the zero Froude number shallow water equations. PIK Report 97, Potsdam Institute for Climate Impact Research (2005) 3, 6, 18, 20, 25