Analysis of the Cell-Vertex Finite Volume Method for Pseudo-Incompressible Divergence Constraints on Quadrilateral and Cuboid Meshes

Dissertation zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften

am Fachbereich Mathematik und Informatik der
Freien Universität Berlin
vorgelegt von

Gottfried Hastermann

Berlin, 2022

Betreuer und 1.Gutachter

Prof. Dr.-Ing. Rupert Klein
Freie Universität Berlin
Fachbereich Mathematik und Informatik
Arnimallee 6
14195 Berlin

2.Gutachter

Prof. Colin Cotter
Department of Mathematics
Imperial College London
755 Huxley Building
South Kensington Campus, London

Tag der Disputation

4th of August, 2022

Erklärung

Ich erkläre gegenüber der Freien Universität Berlin, dass ich die vorliegende Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe. Die vorliegende Arbeit ist frei von Plagiaten. Alle Ausführungen, die wörtlich oder inhaltlich aus anderen Schriften entnommen sind, habe ich als solche kentlich gemacht. Diese Dissertation wurde in gleicher oder ähnlicher Form noch in keinem früheren Promotionsverfahren eingereicht. Mit einer Prüfung meiner Arbeit durch ein Plagiatsprüfungsprogramm erkläre ich mich einverstanden.

Mühlenbeck, den 15.3.2022
Gottfried Hastermann

Acknowledgements

First, I would like to thank Rupert Klein for supervising this thesis. Besides the stable funding and work place I would like to especially thank him for the huge amount of scientific input I gained from our discussions and meetings. Even more, I would like to thank him for his personal support which meant a lot to me. The idea of myself becoming the academic descendant of a brilliant scientist of utmost professional and personal integrity as him, fills me with joy and pride.

Although a major part of this work essentially was done in home office during the COVID-19 pandemic, I probably would not have learnt half, and enjoyed even a smaller portion of my stay at Freie Universität if not for my colleagues. I thank my dear office colleagues Tom Döerffel, Sandra Döpking and Patrick Gelß for the countless research or non-research related discussions, their open ear for whatever issue bothered me, and last but not least, coping with my sometimes over boarding enthusiasm. It was *quality* time I (will) dearly miss.

Furthermore, I wish to thank Ray Chew, Ulrike Eickers, Stefan Gerber, Thomas von Larcher, Jannes Quer, Sebastian Reich, Maria Reinhardt, Nikki Vercauteren and for being great friends, colleagues, and cooperation partners during their and my time at Freie Universiät.

Especially in the last period of this work, I was very grateful for my close friends Jonathan Gantner, Oliver Hoffmann and Florian Pribahsnik, who never gave up on me despite the long distance.

My dear friends Jana and Jan de Wiljes, I thank wholeheartedly, for all the award-winning cuisine, joyful evenings, the constant support, and so much more.

Moving to Berlin, put my/our families in a difficult situation. Thus, I am especially grateful for the effort my parents, Gabriele and Franz, as well as my and parents in law, Elena and Peter undertake regularly, to compensate for the family spread across half of Europe. Without their support, love, and care this work might not have existed after all.

Most importantly, I would like to thank my amazing children Charlotte and Constantin as well as my exciting love and wife Maria for having patience through all the ups and downs, culminating to this work and for providing me not only *how*, but also *why*.

Abstract

In this work we investigate the stability and approximation properties of the cell-vertex finite volume method applied to an elliptic partial differential equation discretized on quadrilateral or cuboid meshes in two or three dimensions respectively. The Helmholtz type equation of interest originates from the projection step in the semi-discretisation of a second order semi-implicit finite volume scheme, which is capable of resolving the pseudo-incompressible and compressible regime of the Euler equations in a unified numerical framework.

Consequently, we investigate the mixed saddle point problem determined by the pseudo-incompressible divergence constraint and include the source terms responsible for compressible effects. We provide stability and an a-priori error estimate for the projection step in the pseudo-incompressible case, as well as stability for the compressible situation. To this end we leverage an interpretation of the discrete flux variables in terms of discontinuous Galerkin method and introduce the Raviart–Thomas interpolation operator on the dual control volumes surrounding each vertex of the primary grid. This choice is motivated by the natural divergence defined via the integral normal flux passing through the boundary of a dual control volume.

Contents

List of Symbols

Symbol	Explanation
$*$	wildcard
$\mathbb{I}$	identity operator
$\mathbb{J}$	$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$
$\operatorname{adj} M$	adjugate of matrix $M \in \mathbb{R}^{n\times n}$
$\operatorname{span} A$	linear span of $A \subset \mathbb{R}^n$
$x \cdot y$	Euclidean scalar product for $x, y \in \mathbb{R}^n$
$x \times y$	Outer product for $x, y \in \mathbb{R}^3$
$X \times Y$	Cartesian product of two sets/spaces X and Y
$\lvert x \rvert$	absolut value of $x \in \mathbb{R}$
$\lvert\{i_1 \dots i_n\}\rvert$	cardinality of a countable set
$\lvert K \rvert$	Lebesgue measure of $K \subset \mathbb{R}^n$ for $n \in \mathbb{N}$
$\operatorname{dom} f$	domain of a function f
$\operatorname{ran} f$	range of a function f
$\dim X$	dimension of topological vector space X
$\overline{X}, \overline{X}^Y$	closure of some set X, with respect to norm of Y
$\operatorname{supp} f$	$\overline{\{x \in \operatorname{dom}(f) \colon f(x) \neq 0\}}$
$\partial\Omega$	the boundary of some set $\Omega \subset \mathbb{R}^n$ for $n \in \mathbb{N}$
$\operatorname{ess\,sup}/\operatorname{ess\,inf}$	essential supremum/ infimum
X'	topological dual space of X
T'	dual operator of T
$\langle \cdot, \cdot \rangle_{X'\times X}$	duality pairing of $X' \times X$
$\mathcal{B}(X,Y)/\mathcal{B}(X)$	bounded linear operators $X \to Y / X \to X$
$\lVert \cdot \rVert$	Euclidean or operator norm
$(\cdot, \cdot)_X$	inner product on some inner product space X
$L^p(\Omega)$	space of p-integrable functions on Ω
$\partial\Omega$	boundary of some set $\Omega \subset \mathbb{R}^n$ for $n \in \mathbb{N}$
τ_{n_S}	normal trace operator, c.f. Lemma A.4.3
$\mathbb{1}_K$	indicator function over K
$\operatorname{avg}_K f$	integral average of f over K

Symbol	Explanation
$\frac{\partial}{\partial x}$, ∂_x	partial derivative in x
grad, div	gradient and divergence differential operator
DIV, DIV_h	natural divergence and its discrete counterpart
D^α	distributional derivative for multi-index α, c.f. Definition A.2.2
Df	Jacobian matrix of f
$\mathcal{D}'(\Omega)$, $\mathcal{D}(\Omega)$	distributions, test functions, c.f. Section A.2
$H^k(\Omega)$	Sobolev spaces with k weak derivatives in $L^2(\Omega)$
$W^{k,p}(\Omega)$	Sobolev spaces with k weak derivatives in $L^p(\Omega)$
$H(\mathrm{div}, \Omega)$	Sobolev space with weak div in $L^2(\Omega)$
$H(\mathrm{curl}, \Omega)$	Sobolev spaces with weak curl in $L^2(\Omega)$
$\|\cdot\|_{\rho,k,p,\Omega}$	Sobolev space norm of $W^{k,p}(\Omega)^n$ for $n \in \mathbb{N}$, c.f. Definition A.3.1 based on the Lebesgue space with weight ρ, c.f. Lemma A.1.2
$\|\cdot\|_{k,p,\Omega}$	Sobolev space norm of $W^{k,p}(\Omega)^n$ for $n \in \mathbb{N}$, c.f. Definition A.3.1
$\|\cdot\|_{k,\Omega}$	Sobolev space norm $\|\cdot\|_{k,2,\Omega}$
$\|\cdot\|_{1,\Omega}$	$H^1(\Omega)$ semi-norm, c.f. Remark A.3.3
$\mathcal{T}_h$	a grid as introduced in Definition 2.1.1
$\mathcal{N}_K$	vertices attached/in $K \in \mathcal{T}_h$ or $K \subseteq \mathcal{T}_h$
$\mathcal{E}_K$	edges attached/in $K \in \mathcal{T}_h$ or $K \subseteq \mathcal{T}_h$
$\mathcal{F}_K$	faces attached/in $K \in \mathcal{T}_h$ or $K \subseteq \mathcal{T}_h$
$\mathcal{U}$	space for the analytical velocity variable
$\mathcal{H}$	space for the analytical pressure variable
$\mathcal{D}_h$	approximation space for the velocity variable
$\mathcal{R}_h$	approximation space for the gradient of the pressure
$\mathcal{W}_h^1 / \mathcal{W}_{h,0,\mathrm{bc}}^1$	approximation space for the pressure variable
$\mathcal{W}_h'^0 / \mathcal{W}_{h,0,\mathrm{bc}}'^0$	test space for the divergence constraint
κ_1, κ_2	$\ker \mathrm{grad}'$ and $\ker \mathrm{DIV}_h$ in the sense of Eqs. (3.11) and (3.12)
$\mathcal{I}_h / \mathcal{I}_K$	global/local interpolation operator for first order Lagrangian finite elements
$\mathcal{I}_h^d / \mathcal{I}_K^d$	global/local interpolation operator, c.f. Definition 2.4.9
$\mathbb{P}_k$,$\mathbb{Q}_k$	multivariate polynomials, c.f. Definition 2.1.12
$L^{r,d}$	isomorphism identifying $\mathcal{R}_h$ and $\mathcal{D}_h$
L	*steepening* operator $\kappa_1 \to \kappa_2$
Λ,Λ_0	Lumping operator, see Lemma 3.2.5

1 Introduction

In the following, we give a brief introduction to the mathematical models used to predict the time evolution of fluid flow in Earth's atmosphere. We focus on continuum models described by partial differential equations, which are derived by conservation laws and the Newtonian axioms. Furthermore, we point out some technical and practical difficulties arising in these models. Subsequently, we outline and position the contribution of this work in the context of numerical methods for the Euler equations.

1.1 Background

1.1.1 Fluid flow in earth's atmosphere

The mathematical description of one of the most common phenomena tangible to humankind, the airflow in the atmosphere, gives rise to models that pose severe analytical and numerical challenges. Although there was some progress on the matter of Hilbert's 6th problem [63] we mainly have to rely upon experimentally validated continuum models from meteorology and physics derived by the conservation of mass, momentum and (some) energy.

The title of this chapter might suggest there is a unified model to describe the whole atmosphere, but it comes as no surprise that this is not the case. If we restrict ourselves to the homosphere i.e., the atmosphere up to approximately $80\,\mathrm{km}$ height, then we can at least assume the chemical components of air to be well mixed and, therefore, treat it as one Newtonian fluid described by its thermodynamic quantities' density ρ, temperature T, and pressure p as well as its velocity field v. More specifically, we assume dry air to be a perfect gas i.e., to satisfy the ideal gas equation

$$p = \rho R T, \tag{1.1}$$

where $R = 287\,\mathrm{m^2 s^{-2} K^{-1}}$ denotes the specific gas constant for dry air. Furthermore, a perfect gas can be described by constant heat capacities c_V and c_P, where $R = c_P - c_V$.

Therefore, the internal energy ε satisfies

$$\varepsilon = c_V T. \tag{1.2}$$

The isentropic exponent is given by $\gamma = \frac{c_p}{c_V}$ and, as the air consists mainly of diatomic nitrogen, $\gamma = \frac{7}{5}$ is a reasonable assumption for temperature ranges occurring in the homosphere.

Let $\Omega \subseteq \mathbb{R}^3$ be a domain. Then the conservation of mass, momentum and energy on Ω delivers the Navier–Stokes equations (see e.g., Chorin and Marsden [35])

$$\partial_t \rho + \operatorname{div}(\rho v) = 0, \tag{1.3a}$$

$$\partial_t (\rho v) + \operatorname{div}\left(\rho v \otimes v + p\,\mathbb{I} - \tau\right) = f, \tag{1.3b}$$

$$\partial_t e + \operatorname{div}\left(ev + pv - k \operatorname{grad} T - \tau \cdot v\right) = q + f \cdot v. \tag{1.3c}$$

Here f and q describe external volume forces and heating, respectively. Furthermore, $\tau(\operatorname{grad} u)$ denotes the (Newtonian) viscous stress tensor and $k > 0$ the thermal conductivity. The total energy per unit volume e collects the internal and kinetic energy i.e.,

$$e = \rho\varepsilon + \rho\frac{|v|^2}{2} = \rho c_V T + \rho\frac{|v|^2}{2}. \tag{1.4}$$

For sufficiently smooth solutions, we can equivalently state Eq. (1.3c) in non-conservative form and in terms of the pressure p:

$$\partial_t p + v \cdot \operatorname{grad} p + \gamma p \operatorname{div} v = \gamma \left(\operatorname{grad}(v) \colon \tau + \operatorname{div}(k \operatorname{grad} T) + \rho q\right). \tag{1.5}$$

Remark 1.1.1. Henceforth, we only consider adiabatic motion and, therefore, assume $q = 0$.

Remark 1.1.2. For convenience, we do not discuss physically plausible boundary conditions to supplement Eq. (1.3). As the choice of boundary conditions is a delicate matter and rely on the specific modifications of Eq. (1.3) as well as the concrete physical scenario, we, henceforth, ignore physical boundaries and assume periodic boundary conditions except when mentioned differently.

Gravity and fictitious forces

To model the atmosphere and determine the relevant volume forces we assume the earth to be a rotating sphere with homogeneous mass distribution. More realistic models are available [69], but not meteorologically relevant. As the atmosphere is shallow we assume

gravity as constant force characterized by $g \approx 9.81\,\mathrm{ms}^{-2}$ and pointing towards the centre of the earth. We, furthermore, consider a rotating Cartesian frame of reference with its origin located at earth's surface at a certain longitude and latitude and the x_3-axis being aligned with the gravitational force i.e.,

$$f_g = -\rho g e_3. \tag{1.6}$$

In the rotating frame of reference we have to consider the fictitious centripetal and Coriolis forces. The first one is of minor importance as opposed to the second [99]. Given the angular velocity $\omega \in \mathbb{R}^3$, the Coriolis force reads

$$f_C(\rho, v) = -2\rho\omega \times v. \tag{1.7}$$

Due to the rotated frame of reference, the angular speed ω depends on the latitude, but we will ignore the Coriolis component orthogonal to the surface from now on i.e., project onto the tangent plane. The magnitude of the components in the tangent plane (the horizontal components) then depends on the latitude ϕ and e.g., in [99] the force is given by

$$f_C(\rho, v) \approx -2\rho \sin(\phi)|\Omega| e_3 \times v. \tag{1.8}$$

We, furthermore, assume the change in latitude $\sin(\phi)$ to be negligible in Ω i.e., we fix a certain Coriolis parameter $f_0 = 2|\omega| \sin(\phi)$. This then gives the so called f-plane approximation.

The total volume forces in the rotating coordinate frame now read

$$f = f_g + f_C. \tag{1.9}$$

Inviscid flow

The Navier-Stokes equations in the presented form Eq. (1.3) still are far from a complete model for atmosphere. Several important quantities and processes e.g., temperature, radiation, and moisture transport are not included and the treatment of boundary conditions is not determined at all so far. Nevertheless, instead of incorporating these effects, we on the contrary, ignore the viscosity present in the Navier–Stokes equations as justified by the following reasoning. First we rewrite the system as presented in [93] in terms of the non-dimensional variables from Table 1.1 and choose $\mathrm{Sr} = 1$. This gives

$$\partial_t \rho + \mathrm{div}(\rho v) = 0, \tag{1.10a}$$

Parameter	Sr	Ma	Re	Ro	Fr	Pr
Definition	$\frac{l_{\mathrm{ref}}}{t_{\mathrm{ref}} v_{\mathrm{ref}}}$	$\sqrt{\frac{\rho_{\mathrm{ref}} v_{\mathrm{ref}}^2}{p_{\mathrm{ref}}}}$	$\frac{\rho_{\mathrm{ref}} v_{\mathrm{ref}} l_{\mathrm{ref}}}{\mu}$	$\frac{v_{\mathrm{ref}}}{f_0 l_{\mathrm{ref}}}$	$\sqrt{\frac{v_{\mathrm{ref}}^2}{l_{\mathrm{ref}} g}}$	$\frac{c_p \mu}{k}$

Table 1.1: Summarizing the definitions of the non-dimensional parameters depending on the reference and physical quantities.

$$\partial_t(\rho v) + \operatorname{div}(\rho v \otimes v) + \frac{1}{\mathrm{Ma}^2} \operatorname{grad} p = \frac{1}{\mathrm{Re}} \tau - \rho \frac{1}{\mathrm{Fr}^2} e_3 - \frac{1}{\mathrm{Ro}} \rho e_3 \times v, \tag{1.10b}$$

$$\partial_t p + v \operatorname{grad} p + \gamma p \operatorname{div} v = \frac{\gamma}{\mathrm{Pr}\,\mathrm{Re}} \operatorname{div}(\operatorname{grad} T). \tag{1.10c}$$

Here the Reynolds number Re determines the amount of viscosity in the fluid flow compared to the non-linear advection term. To compute concrete exemplary values of the non-dimensional parameters which occur on earth, we have to fix the reference variables. Meteorologically plausible reference values are derived by Klein [80]. Choosing the smallest reference length $l_{\mathrm{ref}} = 11\,\mathrm{km}$ and reference velocity $v_{\mathrm{ref}} = 12\,\mathrm{ms}^{-1}$ as well as the reference density $\rho_{\mathrm{ref}} = 1.25\,\mathrm{kgm}^{-3}$ we are left to determine the dynamical viscosity μ of dry air. For this, as well as the heat conductivity k, we obtain values from [72]. Using these reference values, we roughly assume $\mu \approx 10^{-5}\,\mathrm{kgms}^{-1}$ and $k \approx 10^{-2}\,\mathrm{Wm}^{-1}\mathrm{K}^{-1}$ for a temperature range of $\pm 10^2\,\mathrm{K}$ around $273.15\,\mathrm{K}$ i.e., the freezing point of water. Therefore, we obtain the lowest plausible Reynolds number with $\mathrm{Re} \approx 10^{10}$ and assume a specific heat capacity $c_p \approx 1 \times 10^3\,\mathrm{Jkg}^{-1}\mathrm{K}^{-1}$. This leads to $\mathrm{Pr} \approx 1$.

Since the viscous and conductivity terms seem to nearly vanish, one might suggest to approximate solutions of Eq. (1.10) by solutions of the compressible Euler equations

$$\partial_t \rho + \operatorname{div}(\rho v) = 0, \tag{1.11a}$$

$$\mathrm{Sr}\, \partial_t(\rho v) + \operatorname{div}(\rho v \otimes v) + \frac{1}{\mathrm{Ma}^2} \operatorname{grad} p = -\rho \frac{1}{\mathrm{Fr}^2} e_3 - \frac{1}{\mathrm{Ro}} \rho e_3 \times v, \tag{1.11b}$$

$$\partial_t p + v \operatorname{grad} p + \gamma p \operatorname{div} v = 0. \tag{1.11c}$$

Indeed, Swann [114] and Kato [75] rigorously prove that solutions of the Navier–Stokes equations approach solutions of the Euler equations as the viscosity vanishes for the incompressible $\operatorname{div}(v) = 0$ case and in the absence of boundaries. In presence of boundaries even flows with very little viscosity may substantially differ from an inviscid flow. This is due the to emergence of boundary layers [100] which lead for example, to the non-symmetric flow configuration known as Kármán-Vortex street [74], although the data is

initially symmetric.

Large Reynolds number flows also exhibit turbulence, a complex motion dissipating energy across different eddy length scales [82]. The occurrence of turbulence was the original motivation to study flows depending on the Reynolds number by Reynolds [104] and still counts as one of the great open problems in physics [95].

Both phenomena play important roles in the time evolution of the atmosphere, but inherently are driven or triggered by processes on the molecular scale. These still cannot be resolved by current methods and computational capabilities in simulations of the atmosphere on planetary scales. This results in the unsatisfactory situation that the artificial viscosity introduced by the numerical method might be magnitudes larger than the actual physical one. As $\mathrm{Re} \to \infty$, we therefore consider inviscid fluids governed by the compressible Euler equations Eq. (1.11) only. Finally, we point out that in practical applications the boundary effects and turbulence from the sub grid scales require additional modelling assumptions to be tractable numerically.

To adopt the notation of the numerical method in the following chapter, we introduce the potential temperature (see e.g., [45])

$$\theta \coloneqq T \left(\frac{p_0}{p} \right)^{R/c_p} \tag{1.12}$$

given a reference pressure p_0, as well as the Exner pressure

$$\pi \coloneqq \left(\frac{p}{p_0} \right)^{R/c_p} = \left(\frac{p}{p_0} \right)^{\frac{\gamma-1}{\gamma}}. \tag{1.13}$$

Furthermore, we introduce the additional pressure variable

$$P \coloneqq \frac{p_0}{R} \pi^{\frac{1}{\gamma-1}} \propto p^{1/\gamma} \tag{1.14}$$

The preceding definitions imply $\pi\theta = T$ and $P = \frac{p}{\pi R}$. Therefore, the equation of state, Eq. (1.1), now becomes

$$P = \rho\theta. \tag{1.15}$$

We follow Klein et al. [81] and recast the Euler equations in these new variables to obtain

$$\partial_t \rho + \operatorname{div}(\rho v) = 0, \tag{1.16a}$$

$$\partial_t(\rho v) + \operatorname{div}(\rho v \otimes v) + c_p P \operatorname{grad} \pi = -f_0 e_3 \times \rho v - \rho g e_3, \tag{1.16b}$$

$$\partial_t P + \operatorname{div}(P v) = 0. \tag{1.16c}$$

1.1.2 Analytical results

Before discussing the specific numerical method proposed for the solution of Eq. (1.16), we recall that the question of well-posedness of the Euler equations cannot be answered satisfactorily until now. For short times the local existence and uniqueness of smooth solutions is positively answered in e.g., [88]. As pointed out by Feireisl, Lukáčová-Medvidová, and Mizerová [58], the textbook of Benzoni-Gavage and Serre [18] provides a state-of-the-art overview about the theory of systems of hyperbolic conservation laws.

Nevertheless, classical solutions might lose regularity (blow up) and, therefore, one has to consider a more general functional framework i.e., weak solutions. Unfortunately the issue of uniqueness is negatively answered in such class of solutions. Recent work by Chiodaroli, De Lellis, and Kreml [33] proves that there are infinitely many weak solutions to the Cauchy problem for the Euler equations. As usual for conservation laws, one might nevertheless find physical selection principles to identify a unique physically plausible solution [14].

Following this line of a selection principle, the functional analytic framework of dissipative measure valued solutions provided in [26] is suitable to investigate convergence of numerical methods to smooth solutions for the compressible Euler equations [58] as well as the Navier–Stokes equations [59]. The general class of measure valued solutions for hyperbolic conservation laws originates from the seminal work by DiPerna [50] in the context of scalar hyperbolic conservation laws. For details on the historic development, we refer to the excellent introduction of the matter in Feireisl, Lukáčová-Medvidová, and Mizerová [58].

As already hinted, one key ingredient of this framework is the concept of weak-strong uniqueness [64] to assure pointwise convergence to classical solutions if they exist.

1.1.3 The pseudo-incompressible model

The Euler equations including source terms as given in Eq. (1.11) or equivalently in Eq. (1.16) allow for a multitude of phenomena. Many of them occur on different time and spatial scales. Ignoring the Coriolis term in the equations for a moment, the equations give rise to three different time scales as argued by Klein [80]. Given parameters as valid for Earth's atmosphere (see again [80]) the slowest phenomenon modelled by Eq. (1.16) is advection $v_{\mathrm{ref}} = 12\,\mathrm{ms}^{-1}$, followed by the internal gravity waves $c_{\mathrm{int}} \approx 110\,\mathrm{ms}^{-1}$. The fastest waves described by the Euler equations are acoustic waves $c_{\mathrm{ac}} \approx 330\,\mathrm{ms}^{-1}$ which are usually negligible in context of the weather system [80]. Although the use of an overly complicated model, including sound, might seem acceptable at this point, the infamous condition

named after Courant, Friedrichs, and Lewy [44], short CFL stability condition, does imply severe practical limitations in terms of spatial resolution for numerical approximations of solutions for the discussed system of non-linear partial differential equations. Depending on the approximation algorithm and the specific equation the CFL condition bounds the ratio of temporal resolution Δt and spatial resolution Δx by a constant over the (fastest) characteristic speed c_{ac}.

$$\frac{\Delta t}{\Delta x} \leq \frac{\text{const.}}{c_{\text{ac}}} \tag{1.17}$$

As mentioned the fastest characteristic speed in the Euler system is the speed of sound, which now leads to the unsatisfactory situation of the need to pay huge computational effort to the resolution of a phenomenon which does not need to be resolved for the required numerical prediction.

One remedy widely adopted in meteorological simulations is the use of reduced models [99, 117, 89] as opposed to the integration of the full Euler system Eq. (1.16). A certain class of those are the so called *soundproof* models which aim to include only advection and internal gravity waves.

One of these models is the pseudo incompressible model introduced by Durran [53] and Klein et al. [81]. Following Klein et al. [81], we assume $\partial_t P = 0$ and the equations are then given by

$$\partial_t \rho + \operatorname{div}(\rho v) = 0, \tag{1.18a}$$

$$\partial_t(\rho v) + \operatorname{div}(\rho v \otimes v) + c_p P \operatorname{grad} \pi = -f_0 e_3 \times \rho v - \rho g e_3, \tag{1.18b}$$

$$\operatorname{div}(P v) = 0. \tag{1.18c}$$

In line with the seminal works of Klainerman and Majda [76, 77], which address the low Mach number limit for the compressible Euler equations, we might ask in which sense solutions to Eq. (1.18) exist and under which conditions they can be obtained by an asymptotic limit of Eq. (1.16). In contrast to a somewhat similar anelastic approximation [90, 31] and to the knowledge of the author this question is not yet rigorously answered as long as we assume physically plausible distinguished asymptotic limits. In comparison to previously mentioned works [76, 77] the stratification plays a crucial role, when passing to the limit. This issue as well as the validity of Eq. (1.18) as reduced model for (1.16) is discussed in [81] in terms of an Eigenmode analysis. A different formal reasoning presented in [79] identifies the zero Mach limit of the Euler equations including variable density with the small scale limit of the pseudo incompressible model. This is beneficial, in comparison to the anelastic model, as the corresponding limit thereof is the Boussinesq

approximation, which does not allow for large density variations. Furthermore, the use of the pseudo incompressible model for the investigation of gravity wave breaking with arbitrary background stratification is justified in [1].

In comparison to the anelastic model, which would give the Boussinesq equations as corresponding limit, this is beneficial, as the pseudo incompressible model is valid over a wider range of For the latter, one obtains the (see again [79]), which does only allow for small density variations.

The small spatial scale limit of the anelastic model on the other hand allows for large Despite some lack of rigorous understanding, the pseudo incompressible model has favourable properties as the . in the asymptotic limit of small spatial scales allows for large density in comparison with the anelastic model is a surprisingly sufficient soundproof model for numerical weather forecasting [80, 81].

1.1.4 Uniform numerical treatment

In the following, we introduce a state of the art numerical method provided by Benacchio and Klein [16]. It allows the numerical integration of both previously introduced models, the Euler equations as given in Eq. (1.16) and Durran's pseudo incompressible model Eq. (1.18). It is a second order semi-implicit finite volume method which leverages two structurally identical projections to ensure the discrete solution complies with the divergence constraint on Pv and furthermore allow us to treat the gravity source term implicitly. This not only enables the method to overcome step size restrictions caused by acoustic, but also gravity waves (see e.g. [117]). In this chapter we merely introduce the temporal semi-discretisation. This falls short on the careful bespoke design of the spatial discretization provided in [16]. On the other hand we will elaborate on details of the spatial discretization, whenever it will serve the purpose of this work in subsequent chapters.

Considering the Euler equations as given in Eq. (1.16) we first introduce fast and slow version of the inverse potential temperature χ and the Exner pressure π via

$$\chi = \chi' + \overline{\chi}, \qquad \pi = \pi' + \overline{\pi}. \tag{1.19}$$

Here we assume the slow background states to depend only on the vertical axis i.e., $\overline{\chi}(t, x) = \overline{\chi}(t, x_3)$ and $\overline{\pi}(t, x) = \overline{\pi}(t, x_3)$ and to satisfy

$$\partial_{x_3}\overline{\pi} = -\frac{g}{c_p}\overline{\chi}, \qquad \overline{\pi}(0) = 1. \tag{1.20}$$

The time evolution of fast variables χ', π' follow the transport equations

$$\partial_t(P\chi') + \operatorname{div}(Pv\chi') = -Pv_3\partial_{x_3}\overline{\chi} \tag{1.21}$$

$$\partial_t\pi + \frac{\partial\pi}{\partial P}\operatorname{div}(Pv) = 0, \tag{1.22}$$

respectively.

Remark 1.1.3. The necessity to consider the time evolution of the fast *auxiliary* variables originates in their implicit treatment in the projection step in Eq. (1.28b) at time t^{n+1}.

Subsequently, we consider an abstract formulation as suggested in [110]. For this purpose we collect the transported variables

$$\Psi \coloneqq (\chi, \chi v, \chi') \tag{1.23}$$

and realize that we can express Eq. (1.16) in the following abstract form

$$\partial_t(P\Psi) + \mathcal{A}_{Pv}(\Psi) = S(P, \Psi) \tag{1.24a}$$

$$\partial_t P + \operatorname{div}(Pv) = 0, \tag{1.24b}$$

where we denote the non-linear advection by $\mathcal{A}_{Pv}(\Psi)$ and by $S(P, \Psi)$ the source terms on the right-hand side of Eq. (1.16). In the context of the advection operator $\mathcal{A}_u(g)$ we refer to the subscript index u, as advecting vector field and to g as the advected quantities.

In this notation the update for one step of time integration is given by the following semi-discrete update formulas

$$(P\Psi)^{n+1} = \mathcal{A}^{\Delta t}_{(Pv)^{n+1/2}}\left(\Psi + \Delta\frac{t}{2}\,S\left(P^n, \Psi^n\right)\right) + \Delta\frac{t}{2}\,S\left(P^{n+1}, \Psi^{n+1}\right) \tag{1.25a}$$

$$P^{n+1} = P^n - \Delta t\,\operatorname{div}(Pv)^{n+1/2}. \tag{1.25b}$$

$$\pi^{n+1} = \pi^n - \Delta t\left(\frac{\partial\pi}{\partial P}\right)^{\circ}\operatorname{div}\frac{Pv^{n+1} + Pv^n}{2}. \tag{1.25c}$$

Next, we either solve implicitly c.f. Remark 1.1.4 and Remark 1.1.5 via

$$\left(\frac{\partial\pi}{\partial P}\right)^{\circ} \approx \frac{\pi^{n+1} - \pi^n}{P^{n+1} - P^n} \tag{1.26}$$

or estimate the $(\partial_P\pi)^{\circ}$ from previously computed quantities.

Remark 1.1.4. The redundant integration of Eq. (1.24b) by the means of Eq. (1.25b) and

Eq. (1.25c) might seem superfluous at first sight. This, however, conveniently allows us to evaluate the fluxes by the means of P and at the same time solve implicitly for the stiff pressure perturbation π'. To this end the pressure variables P and π, differ in their spatial discretization. In [16] the degrees of freedom of P are cell centred as opposed to the node centred values of π. Considering Eq. (1.26) the authors of [16] propose the use of a spatial average via linear interpolation K and to compute

$$\left(\frac{\partial \pi}{\partial P}\right)^{\circ} := \frac{\pi^{n+1} - \pi^n}{KP^{n+1} - KP^n}. \tag{1.27}$$

For the first set of equations i.e., Eq. (1.25a) we can interpret this discretization as trapezoidal rule along the advecting field Pv. Indeed, Smolarkiewicz and Margolin [111] provide a proof for second order consistency for this strategy and propose the use of a multidimensional positive definite advection transport algorithm, *MPDATA* [109] as advection operator $\mathcal{A}$. Benacchio and Klein [16] adopt this idea to the extent as their proposed algorithm solves the linear advection equation and chooses the advecting velocity field to be $(Pv)^{n+1/2}$. Their advection method, however, differs and leverages a MUSCL-type finite volume method for spatial discretization.

Apart from the specific discretization of $\mathcal{A}$ two questions arise. Firstly, how do we obtain the advecting velocity field $(Pv)^{n+1/2}$ and secondly, how do we solve Eq. (1.25)?

The first question can be answered by restating a slightly modified version of Eq. (1.25). More specifically we have to shorten the time step to $\Delta t/2$ and choose the advecting field at the old time level t^n. This leaves us to discuss the solution of Eq. (1.25).

Projection

To address the question on how to solve Eq. (1.25), Benacchio and Klein [16] suggest using an auxiliary implicit Euler discretization of Eq. (1.16c). After division of the momentum equations in Eq. (1.25a) by $\chi^{n+1} = \chi^*$ and in combination with the equation for the fast inverse potential temperature and Eq. (1.25c) we obtain

$$(P\psi)^* = \mathcal{A}^{\Delta t}_{(Pv)^{n+1/2}} \left(\Psi + \Delta t/2\, S\left(P^n, \Psi^n\right)\right), \tag{1.28a}$$

$$\begin{aligned}(Pv)^{n+1} = (Pv)^* &- \frac{\Delta t}{2}\left(\frac{\rho}{\chi}\right)^{n+1} \frac{c_p}{\chi^{n+1}} \operatorname{grad} \pi'^{n+1} \\ &- \frac{\Delta t}{2}\left(\frac{\rho}{\chi}\right)^{n+1} f e_3 \times v^{n+1} - \frac{\Delta t}{2}\left(\frac{\rho}{\chi}\right)^{n+1} g \frac{\chi'^{n+1}}{\chi^{n+1}} e_3,\end{aligned} \tag{1.28b}$$

$$(P\chi')^{n+1} = (P\chi')^* - \frac{\Delta t}{2} \partial_{x_3} \overline{\chi}^{} (Pv_3)^{n+1}, \tag{1.28c}$$

$$\left(\frac{\partial P}{\partial \pi}\right)^{\circ} \pi^{n+1} = \left(\frac{\partial P}{\partial \pi}\right)^{\circ} \pi^{*} - \frac{\Delta t}{2} \operatorname{div}(Pv)^{n+1}. \tag{1.28d}$$

These equations already provide a framework to enforce the hydrostatic balance and the pseudo compressible divergence constraint. More concretely we can realize those limit regimes by eliminating terms via scalar prefactors $\alpha_P \in \{0, 1\}$ and $\alpha_W \in \{0, 1\}$. For notational convenience we additionally introduce

$$\Sigma_W = \operatorname{diag}(1, 1, \alpha_W). \tag{1.29}$$

We now put those prefactors in front of the terms in Eq. (1.28) which originate from the time derivative. The modified version of Eq. (1.28) reads

$$\begin{aligned} \Sigma_W (Pv)^{n+1} &= \Sigma_W (Pv)^{*} - \left(\frac{\rho}{\chi}\right)^{n+1} \frac{\Delta t}{2} \frac{c_p}{\chi^{n+1}} \operatorname{grad} \pi'^{n+1}, && \text{(1.30a)} \\ &\quad - \left(\frac{\rho}{\chi}\right)^{n+1} \frac{\Delta t}{2} f e_3 \times v^{n+1} - \left(\frac{\rho}{\chi}\right)^{n+1} \frac{\Delta t}{2} g \frac{\chi'^{n+1}}{\chi^{n+1}} e_3, \\ (P\chi')^{n+1} &= (P\chi')^{*} - \frac{\Delta t}{2} (Pv_3)^{n+1} \partial_{x_3} \overline{\chi}, && \text{(1.30b)} \\ \alpha_P \pi^{n+1} \left(\frac{\partial P}{\partial \pi}\right)^{\circ} &= \alpha_P \pi^{n} \left(\frac{\partial P}{\partial \pi}\right)^{\circ} - \frac{\Delta t}{2} \operatorname{div}(Pv)^{n+1}. && \text{(1.30c)} \end{aligned}$$

In this form we already can observe that $\alpha_P = 0$ enforces the pseudo incompressible balance at the new time level i.e.,

$$\operatorname{div}(Pv)^{n+1} = 0. \tag{1.31}$$

In the case of $\alpha_W = 0$ the update equations Eq. (1.30) enforce the hydrostatic balance

$$\frac{c_p}{\chi^{n+1}} \partial_z \pi^{n+1} - g = 0. \tag{1.32}$$

Remark 1.1.5. The system stated in Eq. (1.30) indeed provides a second order update for the momentum. For the Exner pressure the authors of [16] also provide a second order update (c.f. Remark 1.1.4) to non-linearly correct the potential mismatch between the two pressure variables. Furthermore, they point out that in many situations it is not necessary to do so and the mismatch of Eq. (1.14) between the nodal Exner pressure π and cell centred variable P is negligible.

In the following we mostly consider the case $\alpha_P = 0$ and henceforth use the linear

formulation i.e., assume $\partial_\pi P$ to be given prior to the resolution of Eq. (1.30).

Remark 1.1.6. Let $\alpha_w > 0$ or $\partial_{x_3}\overline{\chi} \neq 0$. Then the system of equations Eq. (1.30) constitutes a quasilinear Helmholtz equation as provided in [16]. To this end Benacchio and Klein inverted the rotation operator responsible for the Coriolis term in Eq. (1.30a).

1.1.5 Numerical methods

Semi-implicit (projection) methods

The idea of solving a Poisson equation in the pressure variable to obtain a velocity field obeying the divergence constraint originates from the seminal works of Harlow and Welch [66] and Chorin [34]. Bell, Colella, and Glaz [15] provides a second order version of this discrete projection onto the divergence free manifold.

All preceding references consider viscous fluid flow, and we remark that the choice of boundary conditions in the pressure variable that ensure optimal (second order) convergence is non-trivial and the issue is discussed e.g., in [30]. Nevertheless, for the case of inviscid flow the issue vanishes as does viscosity and there is a natural choice of homogeneous Neumann boundary conditions for the pressure.

Numerical solutions for the close low-Mach regime i.e., $0 < \mathrm{Ma} \ll 1$ on the other hand are non-trivial to obtain [119]. This difficulty is caused by the pressure term which becomes stiff in the low-Mach regime. In his seminal work [78], Klein resolves the issue in one spatial dimension by splitting the pressure variable using reasoning from asymptotic analysis. Subsequently, the fast pressure variable is treated implicitly in contrast to the slow one which is treated explicitly. This idea is also fundamental for the previously discussed numerical method developed in [16] and is extended to the multidimensional situation in [106].

The question if one can numerically pass to the limit (without loss of stability and consistency) coined the term *asymptotic preserving*. Recent works in context of incompressible/compressible Euler equations are e.g., [41, 24].

In contrast to the classical numerical treatment of the Poisson equation, where one is not necessarily interested in the gradient of the solution, the momentum correction and therefore some variant of the gradient of the pressure is the main concern. This leads to the question, how the degrees of freedom of the velocity field are located in relation to the ones of the discrete pressure. In the context of finite difference methods for the shallow water equations the location of the variable relative to each other was classified by Arakawa and Lamb [4]. In the same year Raviart and Thomas [101] and Raviart and Thomas [102] introduced their mixed finite elements on triangular grids for second order

elliptic problems. More recently, Cotter and Shipton [42] discuss the connection between C-grid finite differences and mixed finite elements. (Un)surprisingly it turns out that if we define the elements via nodal values, the position of those coincides with the corresponding finite difference grid staggering. In context of discontinuous or hybrid methods, where the jump across the individual element is a degree of freedom on its own (see e.g., [27]), this suggests that we might provide at least some parts of the analysis locally on each element/cell.

1.1.6 Discontinuous Petrov Galerkin and finite volume methods

Historically, discontinuous Galerkin methods originate from the seminal work by Reed and Hill [103] in the context of neutron transport. According to the introduction of [70], however, the idea to use such in the context of second order elliptic problems originates to Nitsche [98], Wheeler [121], and Arnold [5].

In contrast to the continuous finite element methods, the discontinuous approximation and test functions allow for the same favourable conservation properties as finite volumes methods do. Both are closely related and for the low order methods we can indeed restate the same discretization in each of both frameworks. For high order approximation the local nature of the discontinuous functions also gives the computational advantage of block diagonal mass matrices. Nevertheless, in this work we consider only piecewise (bi)linear functions and do not investigate the possibility of a high order version. In general, we need to introduce some kind of stabilization for the application of a discontinuous Galerkin method to second order partial differential equations. For a review and unified analytical framework of common discontinuous Galerkin methods applicable to the mixed Poisson problem we refer to [9]. A quite recent textbook on Discontinuous Galerkin methods applied to different scenarios is [49].

The famous pair of mixed finite elements by [102] is a convenient choice to discretize the flux conservatively, however suboptimal convergence rates are observed on quadrilateral grids in [7]. For low order elements, where these suboptimal convergence rates would be catastrophic, this can be avoided by using the correct natural divergence defined by integration along the boundary, instead of the classical distributional divergence [25]. This might seem undesirable when considering the goal of a $H(\mathrm{div}, \Omega)$ conforming situation, nevertheless from the perspective of finite volume methods (c.f. [87, 57]), this choice appears to be quite natural.

Although closely related, a pure finite volume approach demands the (re)construction of a consistent gradient to be discretized by integration of the relevant variable, multiplied by

the outwards normal of the control volume (c.f. [57]). The choice on how to do so depends on the grid geometry and is non-trivial in general. One rather recent suggestion by authors of [56] is the use of a variational formulation to recover an approximate gradient similar to the one, one would expect considering the mixed problem in the sense of [102].

A way to avoid the reconstruction, is provided by the finite volume element method. The analysis of this method dates back to [65, 32]. The latter suggests the use of a (continuous) finite element approximation space, but poses the equation in form of boundary integrals. Along this line Süli [113] proves optimal convergence of the cell centre (a node centred) scheme for the Poisson problem on Cartesian grids. The work of [115] discuss discretization and error estimates for a mixed Poisson problem in the $H(\mathrm{div}, \Omega)$ conforming setting on triangular and rectangular grids. Finally, the work of Angermann [3] and Vater and Klein [118] extended this work to a non-conforming approximation of the fluxes on triangular and Cartesian grids respectively. Another line of work for cell centred finite volume methods is presented in [13, 51, 52].

1.2 Outline and Scope of this Work

1.2.1 Motivation

Given the preceding introduction the presented work aims to refine and develop numerical analysis for semi-implicit numerical methods based on finite volume approximations specifically developed for the compressible Euler equations and variants thereof.

The overarching theme is the question if a semi-implicit finite volume method [87] in the spirit of the work of Benacchio and Klein [16] is converging in some sense to solutions of the Euler equations. Gallouët et al. [61], Feireisl, Lukáčová-Medvidová, and Mizerová [58], and Feireisl et al. [59] show that convergence proofs for numerical methods for Navier–Stokes and Euler equations are possible in principle, but they leverage either a fully implicit and therefore dissipative time discretization or consider the semi-discretisation in space obtained by entropy stable finite volume methods.

In the linear case the seminal work by Lax and Richtmyer [84] answered the question of convergence for finite differences. In the case of non-linear systems of partial differential equations the situation is less straight forward, nevertheless in analogy to the linear case we expect two necessary conditions. The first one is consistency to ensure the numerical method indeed approximates its analytical counterpart. The second one is stability which in the non-linear case essentially is covered by discrete a priori estimates. In this work we concentrate on the latter.

For our concrete situation one step towards the goal of stability is to establish a bound on the discrete in time evolution of Pv in an appropriate norm. We ask for a constant $C > 0$ independent of the computational effort i.e., the number of time steps $n = T/\Delta t$ such that the $(Pv)^n$ as approximation of $(Pv)(T)$ is bound by

$$\|Pv^n\|_X \leq C\|Pv^0\|_X + \mathcal{O}(T) \tag{1.33}$$

for some appropriate norm $\|\cdot\|_X$. The constant C, however, depends on time and the other physical parameters in the model.

Given half-time fluxes $Pv^{n+1/2}$ we can decompose the algorithm proposed in [16] conceptionally into two steps. First we advect $P\psi$. Subsequently, we solve for the divergence corrected field Pv. For the latter we utilize the implicit update for the Exner pressure π.

One way of establishing Eq. (1.33) is to provide an individual stability estimate for every step of $\Delta t = T/n$ by e.g.,

$$\|(Pv)^n\|_X \leq C_n\|(Pv)^{n-1}\|_X + \Delta t D_n \tag{1.34}$$

where $D_n > 0$. Implicitly we assumed the norm does not change for every time step, this however, might not be a useful assumption. Considering this simplification, nevertheless, we resolve the recursion i.e.,

$$\|Pv^n\|_X \leq \prod_{j=1}^{n} C_j\|(Pv)^0\|_X + \Delta t \sum_{i=1}^{n} D_i \prod_{j=i}^{n-1} C_j \tag{1.35}$$

and ask for a uniform bound

$$D_i \prod_{j=i}^{n-1} C_j \leq K \tag{1.36}$$

for every $i \in \{1 \dots n\}$. Subsequently, we consider the non-negative sequence $C_n \in \mathbb{R}$ and observe that convergence of its product

$$C \coloneqq \lim_{n\to\infty} \prod_{i=1}^{n} C_i \tag{1.37}$$

is sufficient to ensure some $\varepsilon > 0$ and $n_0 \in \mathbb{N}$ with

$$\|Pv^n\| \leq (C+\varepsilon)\|(Pv)^0\|_X + \Delta t \frac{T}{\Delta T} K = (C+\varepsilon)\|(Pv)^0\|_X + KT \tag{1.38}$$

for every $n > n_0$.

The concrete problem of interest, is the question, if in the pseudo incompressible case $\alpha_P = 0$ the implicit projection step in Eq. (1.30), is almost a contraction mapping i.e., it satisfies

$$\|Pv^{n+1}\|_X \leq (1 + O(\Delta t))\|Pv^*\|_X + O(\Delta t) \tag{1.39}$$

for some appropriate norm $\|\cdot\|_X$ and step size parameter $\Delta t > 0$.

So far we have not discussed what discrete solutions to Eq. (1.30) look like. Although Benacchio and Klein [16] use a five point finite-difference stencil for the two-dimensional case, they mention the projection proposed in [118] as potential alternative.

The advantage of the latter is its proven inf-sup stability. This, however, comes at the expense of a mixed discontinuous Petrov–Galerkin framework which utilizes a piecewise linear approximation space for the momenta and piecewise linear $H^1(\Omega)$ conforming finite elements for the pressure. Different from a classical finite difference stencil, this framework allows the use of more sophisticated, but computationally more expensive stencils derived from a well-posed finite volume approach by Süli [113]. Vater and Klein [118] build upon the mentioned result as well as on [115, 3] and provide well-posedness for the mentioned discrete mixed formulation.

As the original goal of the additional degrees of freedom in form of the gradient information is stabilization only, the work in [118] does not emphasize the details of the connections between the underlying finite volume discretization and a formulation in terms of finite elements. As a consequence they neither relate their result to the underlying analytical saddle point problem nor do they provide a compatible interpolation operator to obtain the discrete functions.

1.2.2 Goals

In this work we aim to provide a more complete picture on the functional analytic framework for the implicit part of finite volume method provided in [16]. With future prospects in mind this might be considered helpful, when establishing a priori bounds on the complete step of the aforementioned scheme.

However, as we do consider the implicit part only at this point, we remark that it is not necessary to adopt exactly the same functional framework in the advection and the projection ($\alpha_P = 0$) steps as the work of Lehrenfeld and Schöberl [85] illustrates in the case of a hybrid discontinuous method for the incompressible Navier–Stokes equations. Having that in mind we concentrate on the above-mentioned questions and additionally prove the resulting discretization to satisfy an estimate in the spirit of Eq. (1.39).

We aim to provide finite elements which result in the same discretization as presented in [118]. One of the key aspects of the aforementioned work is the absence of any additional stabilization mechanism as used in e.g., [91] in the case of Darcy flow. In this sense, the proposed method also differs conceptionally from discontinuous Galerkin methods reviewed in [9].

For the definition of such elements we leverage ideas borrowed by from the $H(\mathrm{div}, \Omega)$ and $H(\mathrm{curl}, \Omega)$ conforming finite elements introduced by the seminal works of Raviart and Thomas in [102] and Nédélec in [94] respectively. The former work provides a mixed formulation for the Poisson equation leveraging discontinuous pressure elements. These elements are a formidable choice for the discretization of conservation laws as they are $H(\mathrm{div}, \Omega)$ conforming and therefore allow for Gauss' theorem in the discrete setting. The authors of [22], however, provide a counter example for the discrete $\mathrm{inf} - \mathrm{sup}$ stability of the mixed formulation of the Poisson problem. In contrast to the classical situation the finite elements provided in this work are non-conforming and depend on piecewise linear but completely discontinuous functions and contour integrals on a dual grid.

1.2.3 Contribution

This work extends the method provided in [118] and applicable to two-dimensional Cartesian grids in the following ways.

- We provide a compatible interpolation operator for the discontinuous momentum variable. Using this interpolation, the interpolant of a divergence free field is divergence free in the discrete sense.
- We rephrase and extend the existing methodology in terms of finite element theory in the sense of Ciarlet [38]. Together with the careful modification of the classical theory we present a line of reasoning, which not only follows the guiding principle of the discrete de Rham complex for $H(\mathrm{div}, \Omega)$, but also generalizes the method to three spatial dimensions as well as to unstructured quadrilateral and cuboid grids.
- Applied to the projection in the pseudo-incompressible case, we prove the resulting numerical method to be stable and consistent in a mesh dependent norm. Furthermore, we prove stability of the projection in the compressible case.
- Another consequence of the presented refined analysis is the identification of the null spaces associated with the dual gradient operator and the discrete divergence operator. This in turn allows us to establish an analogue to integration by parts

and the required a priori bound on the solution in the pseudo-incompressible case, c.f. Eq. (1.39).

2 Dual grid finite elements

In the following we introduce some basic notation and assumptions as, e.g., the polynomial spaces and the grid. We follow the textbook strategy [54] to introduce the proposed pair of elements. Therefore, we employ the definitions and properties first on a reference element and subsequently use bilinear or trilinear images of such a reference element to cover the whole grid. In this global picture we aim to use these elements to recover the inf-sup stable discontinuous Galerkin discretization provided in [118].

2.1 Preliminaries

2.1.1 Grid

In this work we consider grids, consisting of quadrilaterals and cubiods, which are diffeomorphic and positively oriented images of a reference element. An exemplary situation for $n = 2$ is depicted in Fig. 2.1. We denote the vertices, edges, and faces of any entity K in the grid by $\mathcal{N}_K$, $\mathcal{E}_K$ and $\mathcal{F}_K$ respectively. Furthermore, we define the diameter of any compact set $K \subset \mathbb{R}^n$ by

$$h_K := \operatorname{diam}(K) = \max_{x_1, x_2 \in K} \|x_1 - x_2\|. \tag{2.1}$$

Now we define a grid as follows.

Definition 2.1.1 (Grid)**.** Let $n = 3$ ($n = 2$) and let $\Omega \subset \mathbb{R}^n$ be a bounded Lipschitz domain with piecewise bilinear (linear) boundary $\partial\Omega$. A collection of trilinear (bilinear) images $K \subset \Omega$ of $\hat{K} = [-1, 1]^n$ with diameter $h_K < h$ is a grid and denoted by $\mathcal{T}_h$, if all $K_i, K_j \in \mathcal{T}_h$ are connected by a path not intersecting any edge (node) and they only overlap at their boundary i.e., $\int_\Omega \mathbb{1}_{K_1 \cap K_2} \, dx = 0$. Furthermore, a grid has to cover the domain i.e., $\bigcup_{K \in \mathcal{T}_h} K = \overline{\Omega}$. A part of such a grid is illustrated in Fig. 2.1.

Remark 2.1.2. Let $\hat{K} = [-1, 1]^2$ and let K be a non-degenerate quadrangle, then there is a unique bilinear map $T_K \colon \hat{K} \to K$ with $T_K(\hat{K}) = K$ and $\det T_K(x) > 0$ for every $x \in \hat{K}$. If $\hat{K} = [-1, 1]^3$ and K is a non-degenerate cuboid, then T_K is trilinear.

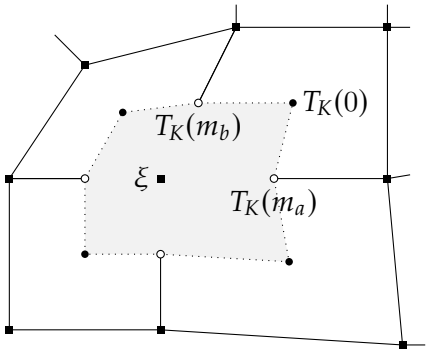


Figure 2.1: The figure depict parts of a grid of quadrilaterals and one of its dual shapes (c.f. also [112])

Definition 2.1.3 (Quasi-uniform)**.** A family of grids $(\mathcal{T}_{\mathfrak{h}_i})_{i\in\mathbb{N}}$ is called quasi-uniform, if there is a constant $c > 0$ satisfying $c\mathfrak{h}_i < h_K$ for every $K \in \mathcal{T}_{\mathfrak{h}_i}$ for every $i \in \mathbb{N}$.

Definition 2.1.4 (Shape-regular)**.** Let $T_K \colon \hat{K} \to K$ denote the map from the reference element $\hat{K}$ to some element K and let $\lambda_j(DT_K)$ denote the $j - th$ eigenvalue of the Jacobian matrix for $j \in \{1 \dots n\}$.

A family of grids $(\mathcal{T}_{\mathfrak{h}_i})_{i\in\mathbb{N}}$ is called shape-regular if there is a constant $c > 0$ satisfying

$$\frac{|\lambda_j(DT_K)|}{h_K} \leq c \tag{2.2}$$

for every $j \in \{1 \dots n\}$, $K \in \mathcal{T}_{\mathfrak{h}_i}$ and $i \in \mathbb{N}$.

Remark 2.1.5. Henceforth, we consider only shape-regular and quasi-uniform families of grids. In abuse of notation we denote such family by $\mathcal{T}_h$ and omit the index for h.

Remark 2.1.6. The dual grid is denoted by $\mathcal{T}_h'$ and constructed around the nodes $\xi \in \mathcal{N}_{\mathcal{T}_h}$. To be more precise, consider the case of $n = 2$. For every $\xi \in \mathcal{N}_{\mathcal{T}_h}$ there is at least one $K = T_K(\hat{K})$ containing ξ. Two of the edges of $\hat{K}$ have images containing ξ, labelled by, say, $a, b \in \mathcal{F}_K$. Denot the center of these edges by m_a, m_b. To construct the boundary of the dual shape K' we connect $T_K(m_a)$, $T_K(0)$ and $T_K(m_b)$. The geometries of dual grids obtained by this strategy are non-trivial and depend on the connectivity of the shapes K, as one can readily see in Fig. 2.1.

Remark 2.1.7. The boundary of every dual cell $\partial K'_\xi$ is composed of the interior faces of the neighbouring cells $K \in \mathcal{T}_h$ with $K \cap \xi \neq \emptyset$ (c.f. Definition 2.2.3). We denote the interior faces around ξ by $\mathcal{F}_{K,\xi}$.

Remark 2.1.8. As we map from certain reference configurations to the dual grid, we only allow shapes that lead to a closed control volume. This requirement is always satisfied by a bilinear (trilinear) transformation for $n = 2$ ($n = 3$).

Remark 2.1.9. Every element K of a Cartesian grid is given by a linear map by

$$T_K(x) = \sum_{i=1}^{n} \frac{\Delta x_i}{2} x_i + b, \tag{2.3}$$

where $\Delta x_i > 0$ denotes the cell size in coordinate direction e_i for $i \in \{1 \dots n\}$ and $b \in \mathbb{R}^n$ is some offset.

Remark 2.1.10. For notational convenience, we introduce the following set

$$\mathcal{K}_{\xi,\eta} \coloneqq \{K \in \mathcal{T}_h \colon \xi, \eta \in \mathcal{N}_K\} \tag{2.4}$$

of shapes K containing two nodes $\xi, \eta \in \mathcal{N}_{\mathcal{T}_h}$. This set satisfies

$$\mathcal{K}_{\xi,\eta} = \mathcal{K}_{\xi,\xi} \cap \mathcal{K}_{\eta,\eta}. \tag{2.5}$$

Furthermore, we introduce the matrices transforming a element local index to a global one. For cell based degrees of freedom we use $G_K^d \in \{0,1\}^{|I||\mathcal{T}_h| \times |I|}$ with

$$G_{K_1}^{d\ T} G_{K_2}^d = \delta_{K_1,K_2} \mathbb{I}_{\mathbb{R}^{|I|}}, \quad \left(G_{K_1}^d G_{K_2}^{d\ T}\right)_{i,j} = \delta_{i,j}\delta_{K_1,K_2} \tag{2.6}$$

for every $K_1, K_2 \in \mathcal{T}_h$ and $i, j \in \{1 \dots |I||\mathcal{T}_h|\}$. The indices of node based degrees of freedom are transformed by $G_K \in \{0,1\}^{|\mathcal{N}_{\mathcal{T}_h}| \times 2^n}$. As the nodal degrees of freedom interact with all neighbouring elements, $G_{K_1} G_{K_2}^{\ T}$ as well as $G_{K_1}^{\ T} G_{K_2}$ do not vanish for $K_1 \neq K_2 \in \mathcal{T}_h$. Nevertheless, the matrices satisfy

$$G_K^{\ T} G_K = \mathbb{I}_{\mathbb{R}^{2^n}}, \quad \left(G_K G_K^{\ T}\right)_{i,j} = \delta_{i,j} \tag{2.7}$$

for every $K \in \mathcal{T}_h$ and $i, j \in \{1 \dots |\mathcal{N}_{\mathcal{T}_h}|\}$.

2.1.2 Spaces of polynomials

As usual for finite element methods we introduce approximation function spaces for the discrete versions of the unknowns i.e., the velocity and the pressure variable. The classical approach to do so, is to utilize polynomial spaces [54]. In multiple dimensions there are different ways to restrict ourselves to certain subspaces of all multivariate polynomials, as

presented in the following.

Remark 2.1.11. Wherever multiple indices are required, and it serves readability, we employ multi-index notation i.e., for every n-tuple $\alpha \in \mathbb{N}^n$ we have

$$f_\alpha = f_{\alpha_1,\alpha_2,\dots,\alpha_n}. \tag{2.8}$$

In abuse of notation we additionally allow scalar operands to be applied to multi-indices, these then are applied component wise. Furthermore, we define

$$|\alpha| \coloneqq \sum_{k=1}^{n} \alpha_k. \tag{2.9}$$

Definition 2.1.12 (Multivariate polynomials)**.** Henceforth, we denote the multivariate polynomials in $\mathbb{R}^n$ up to degree k by

$$\mathbb{P}_k \coloneqq \left\{ f \in C^\infty : \exists c \in \mathbb{R}^{k^n} : f(x) = \sum_{|\alpha|\leq k} c_\alpha \prod_{i=1}^{n} x_i^{\alpha_i} \right\}. \tag{2.10}$$

Additionally, we define the polynomials which have degree up to k in each coordinate

$$\mathbb{Q}_k \coloneqq \left\{ f \in C^\infty : \exists c \in \mathbb{R}^{k^n} : f(x) = \sum_{\max \alpha \leq k} c_\alpha \prod_{i=1}^{n} x_i^{\alpha_i} \right\}. \tag{2.11}$$

The vectorial versions are constructed by the Cartesian product, but allow for intermediate spaces which are e.g., linear in one of the local coordinates x_i and constant in another.

Remark 2.1.13. Let $\alpha \in \mathbb{N}^n$ be a multi index. The polynomials with degree in i coordinate of at most α_i are denoted by

$$\mathbb{P}_\alpha \coloneqq \times_{i=1}^{n} (\mathbb{P}_{\alpha_i} \circ \pi_i), \tag{2.12}$$

where $\pi_i : \mathbb{R}^n \to \mathbb{R}^{n-1}$ denotes the projection onto the plane defined by $x_i = 0$.

Example 2.1.14. A monomial basis for $\mathbb{P}_{0,1} \times \mathbb{P}_{1,0}$ is given by

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} x_2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ x_1 \end{pmatrix} \right\}. \tag{2.13}$$

2.1.3 Some geometric properties

In the subsequent chapters, we repeatedly leverage integral transformations to transfer results established on a reference shape to a mesh of more general shapes. To this end, we introduce the appropriate maps between the function spaces defined on different shapes in the mesh. Consider a regular C^s-diffeomorphism $f : \hat{\Omega} \to \Omega$. Let $p \in [1, \infty]$, then one transformation appropriate for a scalar field (c.f. Theorem 37.1 in [36]) is given by

$$\psi_f : \begin{cases} L^p(\Omega) \to L^p(\hat{\Omega}) \\ q \mapsto q \circ f \end{cases} . \tag{2.14}$$

This transformation is motivated by the existence of $c > 0$ depending on the derivatives of f and $0 \leq k \leq s$ such that

$$\|q\|_{k,p,\Omega} \leq c \|q\|_{k,p,\hat{\Omega}}. \tag{2.15}$$

For the case of affine mappings one obtains stronger results in terms of semi-norms (c.f. [36] again). As the construction of Lagrangian finite elements is classical, we do not further elaborate on ψ at this point.

The vector valued case does need more care if we consider the function spaces $H(\mathrm{div}, \Omega)$ and $H(\mathrm{curl}, \Omega)$. For this sake we introduce the covariant and contravariant Piola transformation and recall some classical results as presented e.g., in [38, 54] with some minor modifications. Let f be a regular C^1-diffeomorphism as before, then we introduce contravariant and covariant Piola transformation as

$$\psi_f^d : \begin{cases} L^p(K) \to L^p(\hat{K}) \\ v \mapsto \left(\frac{Df^{-1} v}{\det(Df^{-1})} \right) \circ f \end{cases} \quad \text{and} \tag{2.16}$$

$$\psi_f^r : \begin{cases} L^p(K) \to L^p(\hat{K}) \\ v \mapsto \left(Df^T v \right) \circ f \end{cases} \tag{2.17}$$

respectively. The superscripts ^{d} and ^{r} hint at the invariance of div and rot under the respective transformation.

All the previously mentioned transformations satisfy the following basic properties.

Remark 2.1.15. Let $n \in \mathbb{N}$ and let $\Omega, \hat{\Omega}, \widetilde{\Omega} \subset \mathbb{R}^n$ be bounded domains. Furthermore, let $g : \Omega \to \hat{\Omega}$ and $f : \hat{\Omega} \to \widetilde{\Omega}$ be C^1-diffeomorphisms with $|\det Dg| > 0$ and $|\det Df| > 0$.

Then $v \to \psi_f^*(v)$ is a bijective linear map and via the multivariate chain rule we have

$$\psi_g^* \circ \psi_f^* = \psi_{f\circ g}^* \tag{2.18}$$

$$\psi_f^{*\,-1} = \psi_{f^{-1}}^*. \tag{2.19}$$

Here, we denote the transformations ψ^d, ψ^r and ψ by ψ^*.

Remark 2.1.16. On uniform Cartesian grids with grid spacing $(\Delta x)_i = h > 0$ for every $i \in \{1, \dots, n\}$, the transformations are linearly dependent. For arbitrary Cartesian grid we still observe

$$\psi_{T_C}^d(v) = \left(\prod_{i=1}^n (\Delta x)_i\right) \operatorname{diag}(1/(\Delta x)_1, \dots, 1/(\Delta x)_n)\psi_{T_C}(v) \tag{2.20}$$

$$\psi_{T_C}^r(v) = \operatorname{diag}((\Delta x)_1, \dots, (\Delta x)_n)\psi_{T_C}(v) \tag{2.21}$$

We now establish bounds on the transformations introduced in Eqs. (2.16) and (2.17). To this end, we recall some basic estimates first [55].

Lemma 2.1.17. *Let $n \in \mathbb{N}$ and let $\Omega, \widetilde{\Omega} \subset \mathbb{R}^n$ be bounded domains. Furthermore let $g\colon \Omega \to \widetilde{\Omega}$ be a C^s-diffeomorphism with $|\det Dg| > 0$ and $f \in W^{s,p}(\widetilde{\Omega})^n$. If α is a multi-index with $s := |\alpha|$, then there is a constant $c > 0$ independent of f and g such that*

$$\|D^\alpha(f \circ g)\|_{0,p,\Omega} \le c\||\det(Dg)|^{-\frac{1}{p}}\|_{0,\infty,\Omega} \max_{1\le r,|\beta|\le s} \|D^\beta g\|_{0,\infty,\Omega}^r \sum_{1\le\delta\le s} \|(D^\delta f)\|_{0,p,\widetilde{\Omega}}. \tag{2.22}$$

Let $h \in C^s(\widetilde{\Omega}, \mathbb{R}^{n\times n})$ then the derivative of the product is bound by

$$\|D^\alpha(hf)\|_{0,p,\widetilde{\Omega}} \le \sum_{\beta+\gamma=\alpha} \|D^\beta f\|_{0,p,\widetilde{\Omega}} \|D^\gamma h\|_{0,\infty,\widetilde{\Omega}}. \tag{2.23}$$

Proof. We find an explicit form of high order partial differentials in [40]. More specifically we find

$$D^\alpha h = \sum_{1\le\beta\le|\alpha|} G_{\alpha,\beta}(g)\,(D^\beta f)\circ g, \tag{2.24}$$

where $G_{\alpha,\beta}$ is a polynomial of partial derivatives. The order of this polynomial is at most s as well as the maximal order of partial derivatives. Changing variables we obtain

$$\|D^\alpha(f \circ g)\|_{0,p,\Omega} \le \sum_{1\le\beta\le|\alpha|} \|G_{\alpha,\beta}(g)\|_{0,\infty,\Omega} \|(D^\beta f)\circ g\|_{0,p,\Omega} \tag{2.25}$$

$$\leq \sum_{1\leq\beta\leq|\alpha|} \|G_{\alpha,\beta}(g)\|_{0,\infty,\Omega} \| |\det(Dg)|^{-\frac{1}{p}} \|_{0,\infty,\Omega} \|(D^\beta f)\|_{0,p,\widetilde{\Omega}} \tag{2.26}$$

$$\leq c \| |\det(Dg)|^{-\frac{1}{p}} \|_{0,\infty,\Omega} \max_{1\leq r,|\beta|\leq s} \|D^\beta g\|^r_{0,\infty,\Omega} \sum_{1\leq\delta\leq|\alpha|} \|(D^\delta f)\|_{0,p,\widetilde{\Omega}}. \tag{2.27}$$

For the second observation we consider the classical product rule which also applies to a smooth function multiplied by a distribution. We therefore obtain

$$\partial_{x_l}(hf)_i = \partial_{x_l} \sum_k h_{ik} f_k = \sum_{k=1}^{p} (\partial_{x_l} h_{ik}) g_k + h_{ik}(\partial_{x_l} f_k) \tag{2.28}$$

$$= \left(h(\partial_{x_l} f) + (\partial_{x_l} h) f\right)_i \quad \forall i \in \{1, \dots, d\}. \tag{2.29}$$

We repeat this line of reasoning iteratively to conclude

$$\|D^\alpha(hf)\|_{0,p,\widetilde{\Omega}} = \| \sum_{\beta+\gamma=\alpha} D^\beta h D^\gamma f \|_{0,p,\widetilde{\Omega}} \leq \sum_{\beta+\gamma=\alpha} \|D^\gamma f\|_{0,p,\widetilde{\Omega}} \|D^\beta h\|_{0,\infty,\widetilde{\Omega}}. \tag{2.30}$$

□

Remark 2.1.18. The same statement is true for $s = 1$ and a locally Lipschitz homeomorphism g, which is diffeomorphic on Ω except on a set of Lebesgue measure zero.

Lemma 2.1.19. *Let $\Omega, \widetilde{\Omega} \subset \mathbb{R}^n$ be two bounded domains. Let $g : \Omega \to \widetilde{\Omega}$ be a C^s-diffeomorphism with $|\det Dg(x)| > 0$ for every $x \in \Omega$. Then ψ_g^d and ψ_g^r are homeomorphisms $W^{s,p}(\widetilde{\Omega})^n \to W^{s,p}(\Omega)^n$. More specifically there are constants $c_{g,q} > 0$ with*

$$\|\psi_g^*(v)\|_{0,p,\Omega} \leq c^*_{g,0} \|v\|_{0,p,\widetilde{\Omega}}, \tag{2.31}$$

$$|\psi_g^*(v)|_{1,p,\Omega} \leq c^*_{g,1} |v|_{1,p,\widetilde{\Omega}}, \tag{2.32}$$

$$\|\psi_g^*(v)\|_{q,p,\Omega} \leq c^*_{g,q} \|v\|_{q,p,\widetilde{\Omega}} \tag{2.33}$$

for every $v \in W^{q,p}(\widetilde{\Omega})^n$ and $q \in \{0, \dots, s\}$. Furthermore, there is $c > 0$ is independent of f and g allowing the explicit characterization of the lower order constants as

$$c^d_{g,0} = \|\operatorname{adj} Dg\|_{0,\infty,\Omega} \| |\det(Dg)|^{-\frac{1}{p}} \|_{0,\infty,\Omega}, \tag{2.34}$$

$$c^d_{g,1} = c \|\operatorname{adj} Dg\|_{0,\infty,\Omega} \|Dg\|_{0,\infty,\Omega} \| |\det(Dg)|^{-\frac{1}{p}} \|_{0,\infty,\Omega}, \tag{2.35}$$

$$c^r_{g,0} = \|Dg\|_{0,\infty,\Omega} \| |\det(Dg)|^{-\frac{1}{p}} \|_{0,\infty,\Omega}, \tag{2.36}$$

$$c^r_{g,1} = c \|Dg\|^2_{0,\infty,\Omega} \| |\det(Dg)|^{-\frac{1}{p}} \|_{0,\infty,\Omega}. \tag{2.37}$$

Proof. As discussed in Remark 2.1.15, both transformations are linear bijections from $W^{s,p}(\widetilde{\Omega})^n$ to $W^{s,p}(\Omega)^n$. Next, we prove continuity and consider the contravariant Piola transformation first. After applying the implicit function theorem and recalling Cramer's rule we see

$$\psi_g^d(w) = \frac{(Dg)^{-1}}{\det(Dg)^{-1}}(w \circ g) = \operatorname{adj}(Dg)(w \circ g), \tag{2.38}$$

where $\operatorname{adj}(Dg)$ denotes the adjugate matrix (i.e., the transposed cofactor matrix).

Let α be a multi index with $1 \leq |\alpha| \leq s$ and let $v \in W^{s,p}(\widetilde{\Omega})$. Changing variables we obtain

$$\|\psi_g^d(v)\|_{0,p,\Omega} \leq \|\operatorname{adj}(Dg)\|_{0,\infty,\Omega} \| |\det(Dg)|^{-\frac{1}{p}} \|_{0,\infty,\Omega} \|v\|_{0,p,\widetilde{\Omega}}. \tag{2.39}$$

Subsequently we use Lemma 2.1.17 to establish

$$\begin{aligned} \|D^\alpha(\psi_g^d(w))\|_{0,p,\Omega} &\leq \sum_{\alpha=\beta+\gamma} \|D^\beta \operatorname{adj}(Dg)\|_{0,\infty,\Omega} \|D^\gamma(w \circ g)\|_{0,p,\Omega} && (2.40) \\ &\leq c \| |\det(Dg)|^{-\frac{1}{p}} \|_{0,\infty,\Omega} \sum_{\alpha=\beta+\gamma} \Bigg(\|D^\beta \operatorname{adj}(Dg)\|_{0,\infty,\Omega} \max_{1\leq r, |\delta|\leq|\gamma|} \\ &\qquad \|D^\delta g\|_{0,\infty,\Omega}^r \sum_{1\leq\eta\leq|\gamma|} \|(D^\eta w)\|_{0,p,\widetilde{\Omega}} \Bigg). && (2.41) \end{aligned}$$

Similarly we obtain

$$\begin{aligned} \|D^\alpha(\psi_g^r(w))\|_{0,p,\Omega} \leq c \| |\det(Dg)|^{-\frac{1}{p}} \|_{0,\infty,\Omega} \sum_{\alpha=\beta+\gamma} \Bigg(\|D^\beta Dg^T\|_{0,\infty,\Omega} \max_{1\leq r, |\delta|\leq|\gamma|} \\ \|D^\delta g\|_{0,\infty,\Omega}^r \sum_{1\leq\eta\leq|\gamma|} \|(D^\eta w)\|_{0,p,\widetilde{\Omega}} \Bigg). \end{aligned} \tag{2.42}$$

Collecting these estimates, we find that there is $\tilde{c} > 0$ independent of w with

$$\|\psi_g^*(w)\|_{s,p,\Omega} \leq c\|w\|_{s,p,\widetilde{\Omega}} \quad \forall w \in W^{s,p}(\widetilde{\Omega}). \tag{2.43}$$

Exchanging the role of Ω and $\widetilde{\Omega}$ and replacing g by its inverse delivers the analogous estimate for ${\psi_g^r}^{-1}$ and ${\psi_g^d}^{-1}$. As the Piola transformation and its inverse are linear they both are continuous. From Eqs. (2.39) and (2.42) we see the constants only depending on derivatives of g or g^{-1}. Therefore, they are invariant under translations of g, and we

conclude the statement. □

Remark 2.1.20. Furthermore, there are constants $c_1, c_2 > 0$ such that

$$\|\psi_g^d(v)\|_{H(\mathrm{div},\Omega)} \leq c_1 \|v\|_{H(\mathrm{div},\widetilde{\Omega})}, \tag{2.44}$$

$$\|\psi_g^r(w)\|_{H(\mathrm{curl},\Omega)} \leq c_2 \|w\|_{H(\mathrm{curl},\widetilde{\Omega})} \tag{2.45}$$

for every $v \in H(\mathrm{div}, \widetilde{\Omega}), w \in H(\mathrm{curl}, \widetilde{\Omega})$. The first, and in our case more important result, is a direct consequence of

$$\mathrm{div}\, \psi_g^d(v) = \det(Dg)\, \widetilde{\mathrm{div}} g, \tag{2.46}$$

which in turn can be proven [38, Thm. 1.7.1.] via the Piola identity

$$\mathrm{div}(Dg^{-T} \det Dg) = \mathrm{div}\, \mathrm{adj}\, Dg^T. \tag{2.47}$$

Remark 2.1.21. Subsequently, one can conclude

$$\int_{\partial\Omega} \psi_g^d(v)\, d\mu = \int_{\partial\tilde{\Omega}} v\, d\mu \tag{2.48}$$

by change of variables and Gauß' theorem. Unfortunately this does not provide a similar result for only some part of the boundary in $H(\mathrm{div}, \Omega)$. One could use this only for a subspace of $H(\mathrm{div}, \Omega)$ where the trace vanishes on parts of the complement of the part of interest. On the other hand, one can restrict the trace operator, but obtaining a trace theorem including a continuous lifting Lemma A.4.3, is not completely straightforward, as one needs to consider the extension of the fractional Sobolev trace space from parts to the whole boundary. This is, however, possible [48].

Following Remark 2.1.21, we avoid using the trace on the whole boundary as well as the use of Gauß' theorem and characterize the surface integral over different domains, by the means of transformed normal vectors only. For convenience, we consider only continuously differentiable surfaces as we do not need less regular structures.

Lemma 2.1.22. *Let $\hat{\Omega}, \Omega \subset \mathbb{R}^3$ such that $\hat{S} = [0,1]^{n-1} \times \{0\} \subset \hat{\Omega}$. Let $f : \hat{\Omega} \to \Omega$ be a C^1-diffeomorphism with $\det Df(x) > 0$ for every $x \in \hat{S}$ and denote the image of $\hat{S}$ by $S = f(\hat{S})$. Then the surface integral over S is characterized by*

$$\int_S \tau_{n_S}(v)\, d\mu = \int_{\hat{S}} (\tau_{n_{\hat{S}}} \circ \psi_Q^d)(v)\, d\mu \tag{2.49}$$

for every $v \in H^1(\Omega)^n$.

Proof. We first remark that S is a compact simply connected subset of a smooth manifold and the trace operator $\tau_{n_S} \colon H^1(\Omega)^n \to S$ is well-defined. The latter we can conclude constructing a Lipschitz domain $\tilde{\Omega} \subset \Omega$ such that $S \subset \partial\tilde{\Omega}$ and subsequently applying Theorem A.3.12. Therefore, the surface integral is well-defined.

As next step we observe

$$\mathbb{J}Df\mathbb{J}^T e_2 = \operatorname{adj}(Df)e_2 \tag{2.50}$$

$$(e_1^T Df) \times (e_2^T Df) = \operatorname{adj}(Df)e_3 = \operatorname{adj}(Df)(e_1 \times e_2), \tag{2.51}$$

for $n \in \{2, 3\}$ respectively. This allows to express the unit normal as

$$n_S = \frac{\operatorname{adj}(Df)e_n}{\|\operatorname{adj}(Df)e_n\|} \tag{2.52}$$

We then argue per definition of the surface measure [73] and direct verification

$$dS = \sqrt{\det Df^T Df}\, dx = \|\operatorname{adj}(Df)e_n\|\, dx. \tag{2.53}$$

Using this and Eq. (2.52) the surface integral becomes

$$\int_S v \cdot n_S \, dS = \int_{[0,1]^{n-1}} v \circ f \cdot n_S \|\operatorname{adj}(Df)e_n\| \, dx \tag{2.54}$$

$$= \int_{[0,1]^{n-1}} v \circ f \cdot \operatorname{adj}(Df)e_n \, dx \tag{2.55}$$

$$= \int_{[0,1]^{n-1}} \operatorname{adj}(Df)^T v \circ f \cdot e_n \, dx \tag{2.56}$$

$$= \int_{\hat{S}} \operatorname{adj}(Df)^T v \circ f \cdot n_{\hat{S}} \, d\mu \tag{2.57}$$

$$= \int_{\hat{S}} \psi_f^d(v) \cdot n_{\hat{S}} \, d\mu \tag{2.58}$$

for every $v \in C^\infty(\Omega)^n$.

Considering Lemma 2.1.19 and Theorem A.3.12 we find both integrals in Eq. (2.49) to be

in $\left(H^1(\Omega)^n\right)'$ by

$$\left|\int_S \tau_{n_S}(v)\,d\mu\right| \leq \int_S |\tau_{n_S}(v)|\,d\mu \tag{2.59}$$

$$\leq \sqrt{|S|}\|\tau_{n_S}(v)\|_{0,S} \tag{2.60}$$

$$\leq c\sqrt{|S|}\|v\|_{1,\tilde{\Omega}} \tag{2.61}$$

$$\leq c\sqrt{|S|}\|v\|_{1,\Omega} \tag{2.62}$$

and

$$\left|\int_{\hat{S}} (\tau_{n_{\hat{S}}} \circ \psi_Q^d)(v)\,d\mu\right| \leq \int_{\hat{S}} |(\tau_{n_{\hat{S}}} \circ \psi_Q^d)(v)|\,d\mu \tag{2.63}$$

$$\leq \sqrt{|\hat{S}|}\|\tau_{n_{\hat{S}}}(\psi_Q^d(v))\|_{0,\hat{S}} \tag{2.64}$$

$$\leq c\sqrt{|\hat{S}|}\|\psi_Q^d(v)\|_{1,\hat{\Omega}\cap\mathbb{R}^3_+} \tag{2.65}$$

$$\leq \tilde{c}\sqrt{|\hat{S}|}\|v\|_{1,\hat{\Omega}} \tag{2.66}$$

for every $v \in H^1(\Omega)^n$. As the integrals are linear, continuous and coincide on the dense subset $C^\infty(\Omega)^n \subset H^1(\Omega)^n$ they already coincide on $H^1(\Omega)^n$. □

Corollary 2.1.23. *Let $g\colon \Omega \to \mathbb{R}^n$ C^1-diffeomorphism with $\det Dg(x) > 0$ for every $x \in S$. Let $T = g(S)$ denote the image of $S \subseteq \Omega$. Then the surface integral satisfies*

$$\int_T \tau_{n_T}(v)\,d\mu = \int_S (\tau_{n_S} \circ \psi_g^d)(v)\,d\mu \tag{2.67}$$

for every $v \in H^1(g(\Omega))^n$.

Proof. $g \circ f$ is a C^1-diffeomorphism with $\det D(g \circ f)(x) > 0$ for every $x \in \hat{S}$. Using Remark 2.1.15 we see

$$\int_T \tau_{n_T}(v)\,d\mu = \int_{\hat{S}} (\tau_{n_{\hat{S}}} \circ \psi_{g\circ f}^d)(v)\,d\mu \tag{2.68}$$

$$= \int_S (\tau_{n_S} \circ \psi_{f^{-1}}^d \circ \psi_{g\circ f}^d)(v)\,d\mu \tag{2.69}$$

$$= \int_S (\tau_{n_S} \circ \psi_g^d)(v)\, d\mu. \tag{2.70}$$

□

The results in Corollary 2.1.23 can be extended to $H(\mathrm{div}, \Omega)$ in terms of the trace space.

Corollary 2.1.24. *The map $\tau_{n_S} \circ \psi_g^d$ is continuous i.e., $\mathcal{B}(H(\mathrm{div}, g(\Omega)), H^{-1/2}(S))$ and we have*

$$\langle \tau_{n_T}(u), v \rangle = \langle \tau_{n_S}(\psi_g^d(u)), v \circ g \rangle \tag{2.71}$$

for every $u \in H(\mathrm{div}, g(\Omega))$ and $v \in H^{1/2}(T)$.

Proof. We apply the classical result on the divergence of the contravariant Piola transformation presented e.g., in [38, Thm. 1.7.1.]. This allows us to conclude $\psi_g^d(v) \in H(\mathrm{div}, \Omega)$ and

$$\|\psi_g^d(v)\|_{H(\mathrm{div},\Omega)} \leq \tilde{c} \|v\|_{H(\mathrm{div}, g(\Omega))}. \tag{2.72}$$

This implies continuity of $\psi_g^d \colon H_{g(\Omega)}(\mathrm{div}, \Omega) \to H(\mathrm{div}, \Omega)$ due to its linearity. As composition of continuous functions, $\tau_{n_S} \circ \psi_g^d$ is then continuous too.

If me multiply v by some arbitrary $\varphi \in C^\infty(g(\Omega))$, Corollary 2.1.23 now reads

$$\langle \tau_n(v), \varphi \rangle = \langle \tau_{n_S}(\psi_g^d(v)), \varphi \circ g \rangle. \tag{2.73}$$

Since g is a diffeomorphism and the space C^∞ and therefore C^1 is dense in H^1 we conclude the statement. □

Remark 2.1.25. Modifying the proof of Lemma 2.1.22 only slightly we obtain a similar statement for the tangential component. Together with Stokes theorem we realize

$$\int_T \tau_{t_T}(v)\, d\mu = \int_S (\tau_{t_S} \circ \psi_g^r)(v)\, d\mu \tag{2.74}$$

for every $v \in H^1(g(\Omega))^n$.

Finally, we present another version of Corollary 2.1.23 applied to the $L^2(T)$-product.

Lemma 2.1.26. *Let $g \colon \Omega \to \mathbb{R}^n$ C^1-diffeomorphism with $\det Dg(x) > 0$ for every $x \in S$. Let $T = g(S)$ denote the image of $S \subseteq \Omega$. Then*

$$\int_T u \cdot \mathrm{grad}\, p\, dx = \int_S \psi_g^d(u) \cdot \mathrm{grad}(p \circ g)\, dx \tag{2.75}$$

for all $p \in H^1(T)$ *and* $u \in H^1(T)^n$.

Proof.

$$\int_T u \cdot \operatorname{grad} p \, dx = \int_S (\operatorname{grad} p) \circ g \cdot (u \circ g) \det Dg \, dx \tag{2.76}$$

$$= \int_S \operatorname{grad}(p \circ g)^T Dg^{-1} (u \circ g) \frac{1}{\det Dg^{-1}} \, dx \tag{2.77}$$

$$= \int_S \operatorname{grad}(p \circ g) \cdot \psi_g^d(u) \, dx \tag{2.78}$$

□

2.1.4 Discrete deRham complex

For the $H^1(\Omega)$, $H(\operatorname{curl}, \Omega)$ and $H(\operatorname{div}, \Omega)$-conforming finite elements on affine simplicial grids we recall the famous discrete de Rham complex illustrated by the following commuting diagram Fig. 2.2. W_h^1 denotes the global approximation space of piecewise linear

$$\begin{array}{ccccccc} H^1(\Omega) & \xrightarrow{\text{grad}} & H(\operatorname{curl}, \Omega) & \xrightarrow{\text{curl}} & H(\operatorname{div}, \Omega) & \xrightarrow{\text{div}} & L^2(\Omega) \\ \downarrow \mathcal{I}^1 & & \downarrow \mathcal{I}^N & & \downarrow \mathcal{I}^{\mathbb{RT}} & & \downarrow \mathcal{I}^0 \\ W_h^1 & \xrightarrow{\text{grad}} & R_h^N & \xrightarrow{\text{curl}} & D_h^{\mathbb{RT}} & \xrightarrow{\text{div}} & W_h^0 \end{array}$$

Figure 2.2: Illustration of the discrete de Rham complex using global Approximations spaces.

and continuous functions, R_h^N the global approximation space spaced spanned by the transformed basis of the Nédélec element [94] and $D_h^{\mathbb{RT}}$ the same but spanned by the transformed basis of the Raviart–Thomas elements and finally W_h^0 the space of piecewise constant functions. The transformation for a single Nédélec element on $K = T_K(\hat{K})$ is given by the covariant Piola transformation $\psi_{T_K}^r$ whereas the transformation for a Raviart–Thomas element on K is given by the contravariant version $\psi_{T_K}^d$ c.f. Eqs. (2.16) and (2.17). Furthermore, $\mathcal{I}^{\mathbb{RT}}$, $\mathcal{I}^N$ and $\mathcal{I}^{\mathbb{RT}}$ denote the corresponding interpolation operators.

For the mixed Poisson problem and, therefore, the compatible discretization of Theorem A.5.1 mostly the left and right parts of the commuting diagram are relevant, as we do not compute the vector potential for curl, but aim to compute the velocity field directly.

Remark 2.1.27. On one hand, the above stated properties help to satisfy conservation properties and resemble qualitative features of the exact solution of the above-mentioned

partial differential equations even in the case of low resolution. This might be especially important in large scale weather simulations, where the grid size is bound due to limitations of computational resources [97, 42, 43]. On the other hand, these properties are valuable when proving stability estimates for the discrete problem. The idea of transferring the geometric properties from the analytical setting into the discrete one, leads to the notion of finite element exterior calculus as presented e.g., in [8].

2.2 A compatible pair of reference elements

We present a pair of reference elements for the velocity and pressure variables.

For the pressure element, we use classical bilinear or trilinear Lagrangian finite elements for $n = 2$ or $n = 3$, respectively. The velocity element is strongly inspired by the Raviart–Thomas elements [101], but *living* on the dual grid. More specifically, we use the functionals developed in [101] and polynomial basis functions akin to the ones proposed in [94]. Our choice leads to bounds on both interpolation errors of equal (first) order in corresponding mesh dependent norms.

2.2.1 Pressure element

Let $n \in \{2, 3\}$. For the pressure reference element we consider Lagrangian finite elements of first order. See e.g., [36, 54] for an overview and [6] for question of approximation order on quadrilateral grids. We define the reference element on a reference shape $\hat{K} := [-1, 1]^n$ and choose the local polynomial space to be $\mathbb{Q}_1$. The functionals are given by the point evaluation

$$\sigma^1_\xi \colon q \to q(\xi) \tag{2.79}$$

for every $\xi \in \mathcal{N}_{\hat{K}}$. We denote the family of functionals on one element by $\Sigma^1 := \{\sigma^1_\xi \colon \xi \in \mathcal{N}_{\hat{K}}\}$. The finite element in the sense of Ciarlet [36] then is given by the triple $\{\hat{K}, P_1, \Sigma^1\}$. We recall the shape functions $\theta_\xi \in \mathbb{Q}_1$, to be uniquely determined by

$$\sigma^1_{\xi_j}(\theta^1_{\xi_i}) = \delta_{i,j} \quad \forall \xi_i, \xi_j \in \mathcal{N}_{\hat{K}} \tag{2.80}$$

and collected in the family $\Theta^1_{\hat{K}}$. Subsequently, we introduce the local interpolation operator

$$\mathcal{I}^1_{\hat{K}} \colon \begin{cases} W^{s,p}(\hat{K}) \to \operatorname{span} \Theta^1_{\hat{K}} \\ q \mapsto \sum\limits_{\xi \in \mathcal{N}_{\hat{K}}} \sigma^1_\xi(q) \theta^1_\xi \end{cases} \tag{2.81}$$

where $s > n/p$. This interpolation operator is well-defined due to the Sobolev embedding theorem [2, Chapter 4]. For $s > n/p$ this theorem guarantees a continuous representative for every $f \in W^{s,p}(\hat{K})$ and therefore pointwise evaluation of f is well defined.

2.2.2 Velocity element

Having established the pressure element, we introduce a version of the (lowest order) Raviart–Thomas [101] elements defined by curve or surface integrals on the dual grid. The construction of the presented elements therefore shares strong resemblance with these classical elements of the $H(\mathrm{div}, \Omega)$ conforming situation. In contrast to the classical situation the polynomial space is changed. Additionally, we do not aim to impose normal continuity conditions in our approximation spaces as the underlying finite volume method does not expose such properties.

The shift to the dual grid is responsible for an exchange of the role of the primal element-wise differential operators div and curl in the sense that they act on exchanged polynomial spaces. Therefore, one needs to introduce discrete analogues to the divergence operator on the dual grid. This approach shares conceptually some ideas of [116], where the authors consider continuous finite elements on the dual grid, but differ in the sense that we do not introduce the finite element spaces on the dual grid itself and act in a discontinuous context.

Raviart–Thomas elements on quadrilaterals do have some caveats their simplicial counterparts do not possess. The approximation of the (distributional) divergence on bilinear transformations might converge suboptimally [7]. In our case of low order elements this essentially leads to loss of convergence (c.f. [7, Table 2]). The remedy presented in [25] is proposed by the means of the natural divergence, defined via boundary integrals. In the resulting norm, one can achieve optimal convergence rates and additionally this definition is insensitive to the discontinuous nature of our method.

Remark 2.2.1. The following notation allows us to state subsequent results for $n = 3$ and still recover the appropriate results for $n = 2$ with relative ease.

Let $\pi_n : \mathbb{R}^n \to \mathbb{R}^{n-1} : x \mapsto \sum_{i=1}^{n-1} x_i e_i$ be the orthogonal projection onto the plane defined by $x_n = 0$. Next, let $\iota_n : \mathbb{R}^{n-1} \to \mathbb{R}^n : x \mapsto x \oplus 0e_n$ be the canonical embedding into the same plane defined by $x_n = 0$.

Finally, let $f : \mathbb{R}^n \to \mathbb{R}^n$, $g : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$, then $\pi_n f \circ \iota_n$ is a vector field $\mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ and $\iota_n g \circ \pi_n$ is a vector field $\mathbb{R}^n \to \mathbb{R}^n$.

Remark 2.2.2. The matrices

$$\mathbb{J}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \quad \mathbb{J}_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix} \quad \mathbb{J}_3 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.82}$$

are $\pi/2$-rotations around the coordinate axis e_i and satisfy $\mathbb{J}_i^T = \mathbb{J}_i^{-1}$ and $\det(\mathbb{J}_i) = 1$ for $i \in \{0, 1, 2, 3\}$.

The following definition of interior faces allows us to use the element wise construction of a classical finite element method without necessarily involving the dual grid. However, the sole use of these geometric objects is their composition to assemble parts of the boundary of a dual grid cell.

Definition 2.2.3 (Interior faces)**.** Let $n = 3$ we first consider

$$R_i : \begin{cases} \mathbb{J}_1^{i-1}\mathbb{J}_2\mathbb{J}_3 & 1 \le i \le 4 \\ \mathbb{J}_2^{i-1}\mathbb{J}_1\mathbb{J}_3 & 5 \le i \le 8 \\ \mathbb{J}_3^{i-1} & 9 \le i \le 12 \end{cases} \tag{2.83}$$

and subsequently define the interior faces as rotations of the unit square embedded in $\mathbb{R}^n$ i.e., $F_i \coloneqq R_i([0,1]^2 \times \{0\})$.

Lemma 2.2.4. *Let $I_1 = \{1 \dots 4\}$, $I_2 = \{5 \dots 8\}$ and $I_3 = \{9 \dots 12\}$. Then $R_i^T R_j$ satisfies the following identity:*

$$\mathbb{J}_3{}^{i-1} R_i^T R_j \mathbb{J}_3^{T\,j-1} = \begin{array}{c|ccc} i \diagdown j & I_1 & I_2 & I_3 \\ \hline I_1 & \mathbb{I} & \mathbb{J}_2^T\mathbb{J}_1^T\mathbb{J}_2\mathbb{J}_1 & \mathbb{J}_2^T\mathbb{J}_1^T \\ I_2 & \mathbb{J}_1^T\mathbb{J}_2^T\mathbb{J}_1\mathbb{J}_2 & \mathbb{I} & \mathbb{J}_1^T\mathbb{J}_2^T \\ I_3 & \mathbb{J}_1\mathbb{J}_2 & \mathbb{J}_2\mathbb{J}_1 & \mathbb{I} \end{array}. \tag{2.84}$$

Furthermore, $R_i e_3$ is orthogonal to $R_j e_3$ if and only if $(i, j) \in \{1 \dots 12\}^2 \setminus (I_1^2 \cup I_2^2 \cup I_3^2)$.

Proof. We first realize $\mathbb{J}_3^T\mathbb{J}_2^T\mathbb{J}_1\mathbb{J}_2\mathbb{J}_3 = \mathbb{J}_3^T\mathbb{J}_1^T\mathbb{J}_2\mathbb{J}_1\mathbb{J}_3 = \mathbb{J}_3$, and for $(i, j) \in I_1^2$ we have

$$R_i^T R_j = \mathbb{J}_3^T\mathbb{J}_2^T\mathbb{J}_1{}^{j-i}\mathbb{J}_2\mathbb{J}_3 = (\mathbb{J}_3^T\mathbb{J}_2^T\mathbb{J}_1\mathbb{J}_2\mathbb{J}_3)^{j-i} = \mathbb{J}_3^{j-i}. \tag{2.85}$$

Via the same reasoning this holds true for I_2^2 and I_3^2 too. Now let $i \in I_1$ and $j \in I_2$, then

factoring the powers gives

$$\begin{aligned} R_i^T R_j &= \mathbb{J}_3^T \mathbb{J}_2^T \mathbb{J}_1^T{}^{i-1} \mathbb{J}_2{}^{j-1} \mathbb{J}_1 \mathbb{J}_3 \\ &= (\mathbb{J}_3^T \mathbb{J}_2^T \mathbb{J}_1^T \mathbb{J}_2 \mathbb{J}_3)^{i-2} (\mathbb{J}_3^T \mathbb{J}_2^T \mathbb{J}_1^T \mathbb{J}_2 \mathbb{J}_1 \mathbb{J}_3)(\mathbb{J}_3^T \mathbb{J}_1^T \mathbb{J}_2 \mathbb{J}_1 \mathbb{J}_3)^{j-2} \\ &= \mathbb{J}_3^T{}^{i-1} (\mathbb{J}_2^T \mathbb{J}_1^T \mathbb{J}_2 \mathbb{J}_1) \mathbb{J}_3{}^{j-1}. \end{aligned} \tag{2.86}$$

Similarly, for we obtain

$$R_i^T R_j = \mathbb{J}_3^T \mathbb{J}_2^T \mathbb{J}_1^T{}^{i-1} \mathbb{J}_3^{j-1} = (\mathbb{J}_3^T \mathbb{J}_2^T \mathbb{J}_1^T \mathbb{J}_2 \mathbb{J}_3)^{i-2} (\mathbb{J}_3^T \mathbb{J}_2^T \mathbb{J}_1^T) \mathbb{J}_3{}^{j-1} = \mathbb{J}_3^T{}^{i-1} (\mathbb{J}_2^T \mathbb{J}_1^T) \mathbb{J}_3{}^{j-1} \tag{2.87}$$

for $(i, j) \in I_1 \times I_3$ and

$$R_i^T R_j = \mathbb{J}_3^T \mathbb{J}_1^T \mathbb{J}_2^T{}^{i-1} \mathbb{J}_3^{j-1} = (\mathbb{J}_3^T \mathbb{J}_1^T \mathbb{J}_2^T \mathbb{J}_1 \mathbb{J}_3)^{i-2} (\mathbb{J}_3^T \mathbb{J}_1^T \mathbb{J}_2^T) \mathbb{J}_3{}^{j-1} = \mathbb{J}_3^T{}^{i-1} (\mathbb{J}_1^T \mathbb{J}_2^T) \mathbb{J}_3{}^{j-1} \tag{2.88}$$

for $(i, j) \in I_2 \times I_3$. The other cases follow from transposition as $(R_i^T R_j)^T = R_j^T R_i$. To conclude the second statement, we first observe $\mathbb{J}_3 e_3 = \mathbb{J}_3^T e_3 = e_3$. Subsequently, we conclude

$$(R_i e_3) \cdot (R_j e_3) = e_3^T R_i^T R_j e_3 = e_3^T \mathbb{J}_3{}^{i-1} R_i^T R_j \mathbb{J}_3^T{}^{j-1} e_3. \tag{2.89}$$

As $e_3^T \mathbb{J}_1 \mathbb{J}_2 e_3 = 0$,$e_3^T \mathbb{J}_2 \mathbb{J}_1 e_3 = 0$ and $e_3^T \mathbb{J}_1^T \mathbb{J}_2^T \mathbb{J}_1 \mathbb{J}_2 e_3 = 0$ this already concludes the proof. □

Remark 2.2.5. Henceforth, we denote the polynomial space of interest, by

$$\mathcal{P}^d := \mathbb{Q}_{0,1,1} \times \mathbb{Q}_{1,0,1} \times \mathbb{Q}_{1,0,1}. \tag{2.90}$$

Let $\tilde{\theta} \colon x \mapsto 1/4 + x_1/2 + x_2/2 + x_1 x_2$ and let $\theta_9^d \colon x \mapsto \tilde{\theta}(x) e_3$, then

$$\theta_i^d \colon x \mapsto R_i \theta_9^d((R_i^T) x) \tag{2.91}$$

for $i \in \{1, \ldots, 12\}$, form a basis of $\mathcal{P}^d$.

Example 2.2.6. Let $n = 3$, then the basis constructed in Remark 2.2.5 of $\mathcal{P}^d$ is determined

by the rows of

$$\theta^d = \begin{pmatrix} \frac{1}{4}-\frac{x_2}{2}-\frac{x_3}{2}+x_2x_3 & 0 & 0 \\ \frac{1}{4}-\frac{x_2}{2}+\frac{x_3}{2}-x_2x_3 & 0 & 0 \\ \frac{1}{4}+\frac{x_2}{2}+\frac{x_3}{2}+x_2x_3 & 0 & 0 \\ \frac{1}{4}+\frac{x_2}{2}-\frac{x_3}{2}-x_2x_3 & 0 & 0 \\ 0 & \frac{1}{4}+\frac{x_1}{2}+\frac{x_3}{2}+x_1x_3 & 0 \\ 0 & \frac{1}{4}+\frac{x_1}{2}-\frac{x_3}{2}-x_1x_3 & 0 \\ 0 & \frac{1}{4}-\frac{x_1}{2}-\frac{x_3}{2}+x_1x_3 & 0 \\ 0 & \frac{1}{4}-\frac{x_1}{2}+\frac{x_3}{2}-x_1x_3 & 0 \\ 0 & 0 & \frac{1}{4}+\frac{x_1}{2}+\frac{x_2}{2}+x_1x_2 \\ 0 & 0 & \frac{1}{4}+\frac{x_1}{2}-\frac{x_2}{2}-x_1x_2 \\ 0 & 0 & \frac{1}{4}-\frac{x_1}{2}-\frac{x_2}{2}+x_1x_2 \\ 0 & 0 & \frac{1}{4}-\frac{x_1}{2}+\frac{x_2}{2}-x_1x_2 \end{pmatrix}. \tag{2.92}$$

Remark 2.2.7. The projection of the polynomial space $\mathcal{P}^d$ onto the e_3-plane gives $\pi_3\mathcal{P}^d \circ \iota_3 = \mathbb{P}_{0,1} \times \mathbb{P}_{1,0}$. Therefore, $\dim(\pi_3\mathcal{P}^d \circ \iota_3) = 4$.

To obtain a basis for $\mathbb{P}_{0,1} \times \mathbb{P}_{1,0}$ one can choose $2\pi_3\theta_i^d \circ \iota_3$ for $i \in \{1, 3, 5, 7\}$. The elements of this basis are given as rows of the following matrix.

$$\theta^{d,2} := \begin{pmatrix} -x_2 + \frac{1}{2} & 0 \\ 0 & x_1 + \frac{1}{2} \\ -x_2 - \frac{1}{2} & 0 \\ 0 & x_1 - \frac{1}{2} \end{pmatrix} \tag{2.93}$$

The geometric objects introduced in Definition 2.2.3 constitute the boundary of an intersecting dual grid element. We now use this to define degrees of freedoms akin to those introduced by Raviart and Thomas in their seminal work [101].

Lemma 2.2.8. *Let $n = 3$ and choose the domain as $V^{\mathrm{div}}(\hat{K}) = C^\infty(K)^n \subset H(\mathrm{div}, K)$. Consider the functionals*

$$\sigma_i^d : \begin{cases} V^{\mathrm{div}}(\hat{K}) \to \mathbb{R} \\ v \mapsto \int\limits_{F_i} v \cdot n_{F_i}\, d\mu \end{cases}, \quad \text{for every } i \in \{1, \dots, 12\}. \tag{2.94}$$

Then the family of σ_i^d forms a basis of the algebraic dual space $\left(\mathcal{P}^d\right)^*$. More specifically, we have $\sigma_i^d(\theta_j^d) = \delta_{i,j}$.

Proof. Let

$$J_i : \begin{cases} x \mapsto \left(\mathbb{J}^T\right)^{i-1} x & n = 2 \\ x \mapsto R_i x & n = 3 \end{cases}. \tag{2.95}$$

Each of the functionals is linear as they are integrals. By change of variables and Defini-

tion 2.2.3 we identify

$$\sigma_i^d(v) = \int_{F_i} v \cdot n_{F_i}\, ds = \int_{F_1} R_i^T(v \circ R_i) \cdot n_{F_1}\, ds = \sigma_{d,1}(R_i^T(v \circ R_i)). \tag{2.96}$$

Since $R_i^T = R_i^{-1}$ and by virtue of Remark 2.2.5 we conclude

$$\sigma_i^d(\theta_j^d) = \sigma_1^d\left(R_i^T R_j \theta_1 \circ (R_j^T R_i)\right). \tag{2.97}$$

Due to Lemma 2.2.4 and Remark 2.2.5 we know

$$\sigma_i^d(\theta_j^d) = \int_{[0,1]^2} e_3^T R_i^T R_j e_3 (\tilde{\theta} \circ R_j^T R_i)(x_1, x_2, 0)\, dx = 0 \tag{2.98}$$

for all $(i,j) \notin I_1^2 \cup I_2^2 \cup I_3^2$. Furthermore, we have

$$\sigma_i^d(\theta_j^d) = \int_{[0,1]^2} (\tilde{\theta} \circ \mathbb{J}_3{}^{i-j})\, dx = \int_{[0,1]^2} (\tilde{\theta} \circ \mathbb{J}^{i-j})\, dx \tag{2.99}$$

for every $(i,j) \in I_1^2 \cup I_2^2 \cup I_3^2$. As the $\pi/2$-rotations around the e_3 generate a finite group of only four elements, we are left to prove

$$\sigma_i^d(\theta_j^d) = \delta_{ij} \tag{2.100}$$

for every $i - j \in I_1$. This we do by the following verification. $\tilde{\theta} \circ \mathbb{J}_3^k$ is an odd function with respect to either one of the axis or both for $k \in \{1 \dots 3\}$. $\tilde{\theta} \circ \mathbb{I}$ is positive, therefore the integral does not vanish and one can compute the value to be 1. □

Lemma 2.2.9. *Let $n = 2$, then the family of functionals defined by $v \mapsto \sigma_i^d(\iota_3 v \circ \pi_3)$ for $i \in \{1,3,5,7\}$ forms a basis of the algebraic dual space $\left(\pi_3 \mathcal{P}^d \circ \iota_3\right)^*$. More specifically, we have $2(\sigma_i^d)(\iota_3\pi_3\theta_j^d \circ (\iota_3\pi_3)) = \delta_{i,j}$ for every $i,j \in \{1,3,5,7\}$.*

Proof. For convenience, we denote $\mathbb{I}_{1,2} = \iota_3\pi_3$.

$$(\sigma_i^d)(\iota_3\pi_3\theta_j^d \circ (\iota_3\pi_3)) = \sigma_1^d\left(R_i^T\, \mathbb{I}_{1,2}\, R_j \theta_1 \circ (R_j^T\, \mathbb{I}_{1,2}\, R_i)\right). \tag{2.101}$$

Consulting Lemma 2.2.4 for the above-mentioned indices $i,j \in \{1 \dots 4\}$ shows $R_i\theta_1^d(x) = R_i\tilde{\theta}(x)e_3 \notin \operatorname{span} e_3$. Additionally, θ_1^d does not depend on x_3 and σ_1^d only integrates with

respect to x_1, x_2 we therefore have

$$(\sigma_i^d)(\mathbb{I}_{1,2}\,\theta_j^d \circ (\mathbb{I}_{1,2})) = \sigma_1^d\left(R_i^T\,\mathbb{I}_{1,2}\,R_j\theta_1^d \circ (R_j^T\,\mathbb{I}_{1,2}\,R_i)\right) \tag{2.102}$$

$$= \sigma_1^d\left(R_i^T R_j e_3\tilde{\theta} \circ (R_j^T\,\mathbb{I}_{1,2}\,R_i)\right) \tag{2.103}$$

$$= \sigma_1^d\left(R_i^T R_j e_3\tilde{\theta} \circ (\mathbb{I}_{1,2}\,R_j^T\,\mathbb{I}_{1,2}^2\,R_i\,\mathbb{I}_{1,2})\right) \tag{2.104}$$

$$= \int_{[0,1]^2} e_3 R_i^T R_j e_3\tilde{\theta} \circ (\mathbb{I}_{1,2}\,R_j^T\,\mathbb{I}_{1,2}^2\,R_i\,\mathbb{I}_{1,2})\,dx \tag{2.105}$$

We check

$$\tilde{\theta} \circ (\mathbb{I}_{1,2}\,R_j^T\,\mathbb{I}_{1,2}^2\,R_i\,\mathbb{I}_{1,2}) = \begin{cases} \frac{1}{4} + \frac{1}{2}x_1 & i = j \in \{1,3\}, \\ \frac{1}{4} + \frac{1}{2}x_2 & i = j \in \{5,7\}, \\ \frac{1}{4} - \frac{1}{2}x_1 & (i,j) \in \{(1,3),(3,1)\}, \\ \frac{1}{4} - \frac{1}{2}x_2 & (i,j) \in \{(5,7),(7,5)\}, \\ \frac{1}{4} & \text{else} \end{cases} \tag{2.106}$$

as well as

$$e_3 R_i^T R_j e_3 = \begin{cases} 1 & (i,j) \in \{1,3\}^2 \cup \{5,7\}^2, \\ 0 & \text{else} \end{cases} \tag{2.107}$$

to subsequently evaluate the integral and obtain a positive, odd (with respect to $1/2$) or constant integrand multiplied by the corresponding constant scalar.

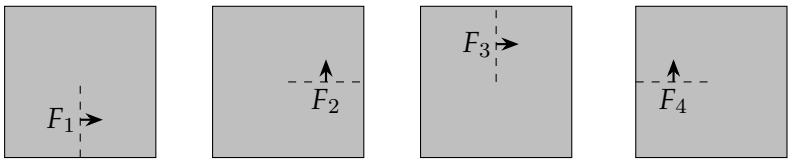
□

Proposition 2.2.10. *Let $\hat{K} = [-1,1]^n$ be a reference shape. For $n = 3$ the triple $\left\{\hat{K}, \mathcal{P}^d, \sigma^d\right\}$ is a finite element as defined by Ciarlet in [36]. Furthermore, the corresponding shape functions are given by θ^d.*

In the case of $n = 2$ the triple $\left\{\hat{K}, \mathbb{P}_{0,1} \times \mathbb{P}_{1,0}, \sigma^{d,2}\right\}$ is a finite element as defined by Ciarlet in [36]. Where $\sigma^{d,2}$ denotes the family of $\pi_3\sigma_i^d \circ \iota_3$ for $i \in \{1,3,5,7\}$. The corresponding shape functions are given by $\theta^{d,2}$.

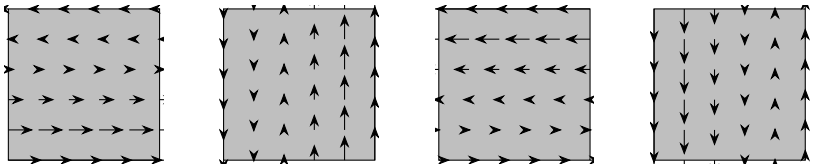
Remark 2.2.11. Whenever the statements are independent of $n \in \{2,3\}$ and in abuse of notation we refer to $\sigma^{d,2}$,$\theta^{d,2}$ and $\pi_3\mathcal{P}^d \circ \iota_3$ by σ^d,θ^d and $\mathcal{P}^d$ respectively. For convenience, we also define $I = \{1 \dots 4\}$ and $I = \{1 \dots 12\}$ for $n = 2$ and $n = 3$ respectively. The family of functionals $\sigma^{d,2}$ is depicted in Fig. 2.3. The shape functions $\theta^{d,2}$ we can see in Fig. 2.4.

Remark 2.2.12. One arguably could have found a more straightforward definition, different from the projection onto the e_3-plane for the case of $n = 2$. Nevertheless, this way the unit



(a) σ_1^d (b) σ_2^d (c) σ_3^d (d) σ_4^d

Figure 2.3: The degrees of freedom are depicted reference over the reference element $\hat{K} = [-1, 1]^2$. Hereby the dashed lines show the domain of the corresponding integral and the arrows indicate the unit normal vector.



(a) θ_1^d (b) θ_2^d (c) θ_3^d (d) θ_4^d

Figure 2.4: Above the shape functions θ^d for $n = 2$, are depicted as vector field over the reference element $\hat{K} = [-1, 1]^2$.

normals are consistent with the three-dimensional case and the definition of the surface integral.

Lemma 2.2.13. *Let $n \in \{2, 3\}$ Then the functionals σ_d can be uniquely extended to $H(\mathrm{div}, \hat{K})'$. Furthermore, the local interpolation operator*

$$\mathcal{I}_{\hat{K}}^d : \begin{cases} H(\mathrm{div}, \hat{K}) \to \mathcal{P}^d \\ \theta \mapsto \sum\limits_{i \in I} \sigma_i^d(\theta)\theta_i^d \end{cases} \tag{2.108}$$

is well-defined, a projection onto $\mathcal{P}^d$ and bounded i.e., $\mathcal{I}_{\hat{K}}^d \in \mathcal{B}(H(\mathrm{div}, \hat{K}), L^2(\hat{K})^n)$.

Proof. The extensions of the functionals are well-defined and continuous as discussed in Lemma A.4.3 i.e., $\sigma_i^d \in H(\mathrm{div}, \hat{K})'$. The same line of reasoning allows us to establish

$$\|\mathcal{I}_{\hat{K}}^d\theta\|_{L^2(\hat{K})}^2 \leq c \sum_{i \in I} \left(\sigma_i^d(\theta)\right)^2 \|\theta_i^d\|_0^2 \leq \tilde{c} \sum_{i \in I} \left(\sigma_i^d(\theta)\right)^2 \leq c' \sum_{i \in I} \|\theta\|_{H(\mathrm{div}, \hat{K})}^2 \tag{2.109}$$

which already provides the statement. □

Lemma 2.2.14. *Let $v \in H^1(\hat{K})^n \subset H(\mathrm{div}, \hat{K})$, then we have*

$$\|v - \mathcal{I}_{\hat{K}}^d(v)\|_{0,\hat{K}} \leq c|v|_{1,\hat{K}}. \tag{2.110}$$

Proof. The proof follows a standard strategy, c.f. [54]. Lemma 2.2.13 implies the continuity of the error i.e.,

$$F\colon v \mapsto v - \mathcal{I}_K^d(v) \in \mathcal{B}(H^1(\hat{K})^n, L^2(\hat{K})^n). \tag{2.111}$$

As next step we prove, $\mathbb{P}_0^n \subseteq \ker F$. To this end let $k = |\dim \mathcal{P}^d|/n$, we then observe

$$e_j = \sum_{i=(j-1)k+1}^{jk} \theta_i^d. \tag{2.112}$$

Let $v = \sum_{i=1}^n c_i e_i \in \mathbb{P}_0^n$, then by linearity of $\mathcal{I}_{\hat{K}}^d$ we see $\mathcal{I}_{\hat{K}}^d v = v$. Convinced of $\mathbb{P}_0^n \subseteq \ker F$, we apply the Deny-Lions lemma Lemma B.0.3 for each component to establish the following bound

$$\begin{aligned}
\|v - \mathcal{I}_{\hat{K}}^d v\|_{0,\hat{K}} &= \|F(v)\|_{0,\hat{K}} && (2.113)\\
&= \inf_{p\in\mathbb{P}_0^n} \|F(v+p)\|_{0,\hat{K}} && (2.114)\\
&\leq \|F\|_{\mathcal{B}(H^1(\hat{K})^n, L^2(\hat{K})^n)} \inf_{p\in\mathbb{P}_0^n} \|v+p\|_{1,\hat{K}} && (2.115)\\
&\leq \tilde{c}|v|_{1,\hat{K}}. && (2.116)
\end{aligned}$$

□

Remark 2.2.15. Although constructed via the normal of curve integrals too, the polynomial approximation spaces change their role in comparison to the classical Raviart–Thomas elements. To convince ourselves that the present elements indeed are different from the original Raviart–Thomas elements we consider $s(x_1, x_2) = ax_2e_1 + bx_1e_2$ where $a, b \in \mathbb{R}$. We denote edges of $\hat{K} = [-1, 1]^2$ by $\mathcal{E}_{\hat{K}}$. The Raviart–Thomas degrees of freedom $\sigma_e^{\mathbb{RT}}$ then are given by the curve integrals

$$\sigma_e^{\mathbb{RT}_0}\colon \theta \mapsto \int_e \tau_{n_e}(\theta)\, d\mu \quad \forall e \in \mathcal{E}_{\hat{K}}. \tag{2.117}$$

By the means of Lemma 2.2.13 we observe $\mathcal{I}_K^d(s) = s$ first. Subsequently, we realize $\mathcal{I}_K^{\mathbb{RT}}(s) = 0$ due to $\sigma_e^{\mathbb{RT}}(s) = 0$ for every $e \in \mathcal{E}_{\hat{K}}$. Therefore, the two elements differ.

Lemma 2.2.16. *Let $p \in \mathbb{Q}_n$ then $\operatorname{grad} p \in \mathcal{P}^d$.*

Proof. For $n \in \{2, 3\}$ this is straightforward to verify when p is expressed by the means of a monomial basis. □

Remark 2.2.17. Let $p \in H^1(\hat{K})$, then classical vector analysis gives $\operatorname{curl} \operatorname{grad} p = 0$.

2.3 Transformed elements

To transfer the results from the reference elements on $\hat{K}$ to more general geometries $K \subset \mathbb{R}^n$, we introduce a diffeomorphism on some bounded domain $\Omega \subset \mathbb{R}^n$ containing the compact reference element $\hat{K}$. More specifically we assume $T_K : \hat{K} \to K$ with $K = T_K(\hat{K})$, $\det DT_K(x) > 0$ for every $x \in \hat{K}$. In abuse of notation, we do not mention the surrounding Ω and only consider the restriction of such a diffeomorphism to $\hat{K}$. Subsequently, we use suitable transformations to map the degrees of freedom as well as the shape functions. In the case of the Lagrangian pressure elements, this transformation is given by Eq. (2.14). The degrees of freedom and shape functions on K are then obtained by (see e.g., [36, 54])

$$\sigma_{K,i} \coloneqq \sigma_i \circ \psi_{T_K} \tag{2.118}$$

$$\theta_{K,i} \coloneqq \psi_{T_K}^{-1} \circ \theta_i. \tag{2.119}$$

Similarly, we use the contravariant Piola transformation from Eq. (2.16) to define degrees of freedom and shape functions for the velocity component on K via

$$\sigma_{K,i}^d \coloneqq \sigma_i^d \circ \psi_{T_K}^d \tag{2.120}$$

$$\theta_{K,i}^d \coloneqq \left(\psi_{T_K}^d\right)^{-1} \circ \theta_i^d. \tag{2.121}$$

They obviously guarantee $\sigma_{K,i}^d(\theta_{K,j}^d) = \sigma_{K,i}^d(\theta_j^d) = \delta_{i,j}$ c.f. Lemmas 2.2.8 and 2.2.9.

In principle, we can choose $\psi_{T_K}^d$ differently. However, some restrictions occur. On one hand, we have to ensure to obtain a finite element in the sense of Ciarlet [36]. On the other hand it is desirable for the analysis as well as the implementation, to not structurally change the degrees of freedom i.e., preserve their form in terms of boundary integrals. Given these requirements, the transformation is essentially determined to be Eq. (2.16). This is well established (see e.g., the textbook of Ern and Guermond [54]) and, subsequently, we follow the classical strategy with some modifications, to employ the corresponding interpolation results on general (smooth) surfaces and on potentially curved geometries.

Proposition 2.3.1 (Parametric finite element)**.** *Let $\{\hat{K}, \mathcal{P}^d, \sigma_d\}$ be the reference finite element introduced in Proposition 2.2.10. Let $T_K : \hat{K} \to K$ be a C^1-diffeomorphism, with $K = T_K(\hat{K})$ and $\det DT_K(x) > 0$ for $x \in \hat{K}$. Furthermore, denote the family of functionals and shape functions transformed according to Eqs.* (2.16) *and* (2.120) *by σ_K^d and θ_K^d respectively. Then $\left\{K, span\, \theta_K^d, \sigma_K^d\right\}$ is a finite element in the sense of Ciarlet [36].*

Proof. K is compact with Lipschitz boundary. $\operatorname{span} \theta_K^d$ is a vector space and σ_K^d a basis of its algebraic dual space by construction of Eqs. (2.120) and (2.121). □

We aim to characterize the transformed functionals σ_d by surface integrals.

Lemma 2.3.2. *Let $T_K : \hat{K} \to K$ be a C^1-diffeomorphism with $K = T_K(\hat{K})$ and $\det DT_K(x) > 0$ for every $x \in \hat{K}$. Then for $i \in I$ the structure of σ_i^d is preserved under the transformation via $\psi_{T_K}^d$ i.e., $\sigma_{K,i}^d$ satisfies*

$$\langle \sigma_{K,i}^d(v), \phi \rangle = \langle \sigma_i^d(\psi_{T_K}^d(v)), \phi \circ T_K \rangle \tag{2.122}$$

for every $v \in H(\operatorname{div}, K)$ and $\phi \in H^{1/2}(K)$. More specifically

$$\sigma_{K,i}^d(v) = \sigma_i^d \circ \psi_{T_K}^d(v) \tag{2.123}$$

for every $v \in H^1(K)$.

Proof. This is an immediate consequence of Corollaries 2.1.23 and 2.1.24. □

Our next step is to introduce the local interpolation operators analogously to the ones on the reference element (c.f. again [54]).

Definition 2.3.3 (Local interpolation)**.** Let $T_K : \hat{K} \to K$ be a C^1-diffeomorphism with $K = T_K(\hat{K})$ and $\det DT_K(x) > 0$ for every $x \in \hat{K}$. Let $w \in H^1(K)^n$, then the local interpolation operator is given by

$$\mathcal{I}_K^d(w) \coloneqq \sum_{i \in I} \sigma_{K,i}^d(w) \theta_{K,i}^d \tag{2.124}$$

Lemma 2.3.4. *Let $T_K : \hat{K} \to K$ be a C^1-diffeomorphism with $\det DT_K(x) > 0$ for every $x \in \hat{K}$ and let $g : K \to \widetilde{K}$ be a C^1-diffeomorphism with $\widetilde{K} = g(K)$ and $\det Dg > 0$ for every $x \in K$. Then interpolation and transformation commute i.e.,*

$$\psi_g^d \, \mathcal{I}_{\widetilde{K}}^d(v) = \mathcal{I}_K^d \, \psi_g^d(v) \tag{2.125}$$

for every $v \in H(\operatorname{div}, K)$.

Proof. Let $v \in H(\operatorname{div}, \tilde{K})$ be arbitrary. We use Remark 2.1.15 and Definition 2.3.3 to obtain

$$\psi_g^d \mathcal{I}_{\widetilde{K}}^d(v) = \psi_g^d \sum_{i \in I} \sigma_{\widetilde{K}}^d(v) \theta_{\widetilde{K},i}^d \tag{2.126}$$

$$= \psi_g^d \sum_{i\in I} (\sigma_i^d \circ \psi_{T_{\tilde{K}}}^d)(v) \big(\psi_{T_{\tilde{K}}}^d\big)^{-1} \circ \theta_i^d \tag{2.127}$$

$$= \sum_{i\in I} (\sigma_i \circ \psi_{T_K}^d \circ \psi_g^d)(v) \big(\psi_g^d\big)^{-1} \circ \big(\psi_{T_K}^d\big)^{-1} \circ \theta_i^d \tag{2.128}$$

$$= \sum_{i\in I} \sigma_{K,i}^d(\psi_g^d(v)) \psi_{T_K^{-1}}^d(\theta_i) \tag{2.129}$$

$$= \mathcal{I}_K^d \psi_g^d(v). \tag{2.130}$$

□

In the following, we adopt the strategy presented in [7]. This allows us to use rather arbitrary geometries without considering their shape explicitly. To this end, we first establish a bound for elements of diameter $h = 1$ and scale accordingly afterwards.

Remark 2.3.5. Given any C^1-diffeomorphism $T_K : \hat{K} \to K$ with $K = T_K(\hat{K})$ and $\det DT_K(x) > 0$ for every $x \in \hat{K}$. Let $S_K \colon x \mapsto h_K x$, then $\operatorname{diam}\big((S_K^{-1} \circ T_K)(\hat{K})\big) = 1$. Furthermore, we have the following estimates

$$\|\operatorname{adj}(DS_K^{-1})\|_{0,\infty,K} = h_K^{n-1}, \quad \|DS_K^{-1}\|_{0,\infty,K} = \frac{1}{h_K}, \quad \||\det DS_K^{-1}|^{-1/2}\|_{0,\infty,K} = \sqrt{h_K^n} \tag{2.131}$$

and

$$\|\operatorname{adj}(DS_K)\|_{0,\infty,K} = \frac{1}{h_K^{n-1}}, \quad \|DS_K\|_{0,\infty,\tilde{K}} = h_K, \quad \||\det DS_K|^{-1/2}\|_{0,\infty,\tilde{K}} = \frac{1}{\sqrt{h_K^n}}. \tag{2.132}$$

Lemma 2.3.6. *Let $T_K \colon \hat{K} \to K$ be a C^1-diffeomorphism with $K = T_K(\hat{K})$, $\det DT_K(x) > 0$ for every $x \in \hat{K}$ and $h_K = 1$. Then there is c such that*

$$\|v - \mathcal{I}_K^d v\|_{0,K} \le c|v|_{1,K} \tag{2.133}$$

holds true for every $v \in H^1(K)$.

Proof. Let $w := \psi_K^d(v)$ then we apply Lemmas 2.1.19 and 2.2.14 as well as the construction of $\mathcal{I}_K^d$ to conclude

$$\|v - \mathcal{I}_K^d v\|_{0,K} \le c_{T_K} \|\psi_{T_K}^d (v - \mathcal{I}_K^d v)\|_{0,\hat{K}} \tag{2.134}$$

$$= c_{T_K} \|\psi_{T_K}^d(v) - \psi_{T_K}^d(\mathcal{I}_K^d v)\|_{0,\hat{K}} \tag{2.135}$$

$$= c_{T_K} \|w - \mathcal{I}_K^d w\|_{0,\hat{K}} \tag{2.136}$$

$$\le c_{T_K} c |w|_{1,\hat{K}} \tag{2.137}$$

$$\leq c_{T_K^{-1}} c_{T_K} c |v|_{1,K}. \tag{2.138}$$

□

Proposition 2.3.7. *Let $T_K \colon \hat{K} \to K$ be a C^1-diffeomorphism with $K = T_K(\hat{K})$ and $\det DT_K(x) > 0$ for every $x \in \hat{K}$. Then there is $c > 0$ depending only on the shape of K such that*

$$\|v - \mathcal{I}_K^d v\|_{0,K} \leq c h_K |v|_{1,K} \tag{2.139}$$

holds true for every $w \in H^1(K)$.

Proof. Let $\tilde{K} = (S_K^{-1} \circ T_K)(\hat{K})$. Furthermore, for every $v \in H^1(K)$ there is $w = \psi_{S_Q}^d(v) \in H^1(\tilde{K})$. Leveraging Remarks 2.1.15 and 2.3.5 and Lemmas 2.1.19 and 2.3.4 we have

$$\begin{aligned}
\|v - \mathcal{I}_K^d(v)\|_{0,K} &= \|\psi_{S_Q^{-1}}^d(w) - \psi_{S_Q^{-1}}^d(\mathcal{I}_K^d(w))\|_{0,K} && (2.140)\\
&\leq h_K^{n-1} \sqrt{h_K^n} \|w - \mathcal{I}_K^d(w)\|_{0,\tilde{K}} && (2.141)\\
&\leq h_K^{n-1} \sqrt{h_K^n} \tilde{c} |w|_{1,\tilde{K}} && (2.142)\\
&= h_K^{n-1} \sqrt{h_K^n} \tilde{c} |\psi_{S_K}^d(v)|_{1,K} && (2.143)\\
&\leq c h_K |v|_{1,K}, && (2.144)
\end{aligned}$$

where Eq. (2.141) follows from Eq. (2.131) as well as Eq. (2.144) follows from Eq. (2.132). □

Having established the approximation properties of the interpolation operator, we aim to investigate the polynomial spaces spanned by θ_K and θ_K^d.

Lemma 2.3.8. *Let $T_K \colon \hat{K} \to K$ be a C^1-diffeomorphism with $K = T_K(\hat{K})$ and $\det DT_K(x) > 0$ for every $x \in \hat{K}$. Let $p \in \operatorname{span} \psi_{T_K}^{-1}(\theta)$ then*

$$\operatorname{grad} p \in \operatorname{span} (\psi_{T_K}^r)^{-1}(\theta^d). \tag{2.145}$$

Furthermore, there is a $L^2(K)$-isomorphism $d_K \colon \operatorname{span}\big((\psi_{T_K}^r)^{-1}(\theta^d)\big) \to \operatorname{span}\big((\psi_{T_K}^d)^{-1}(\theta^d)\big)$ with constants $c_1, c_2 > 0$ independent of $\operatorname{diam} K$ such that

$$c_1 \|v\|_{0,K} \leq \|d_K v\|_{0,K} \leq c_2 \|v\|_{0,K} \tag{2.146}$$

Proof. For every $p \in \operatorname{span} \psi_{T_K}^{-1}(\theta)$ there is $\hat{p} = p \circ T_K$. As $\hat{p} \in \mathbb{Q}_n$ we apply Lemma 2.2.16, to

conclude $\operatorname{grad}\hat{p} \in \mathcal{P}^d$. The chain rule gives

$$\operatorname{grad}\hat{p} = \operatorname{grad}(p \circ T_K) = DT_K^T(\operatorname{grad} p) \circ T_K = \psi^r_{T_K}(\operatorname{grad} p), \tag{2.147}$$

which in turn implies $\psi^r_{T_K}(\operatorname{grad} p) \in \mathbb{Q}_n$ and therefore Eq. (2.145).

Obviously ${\psi^d_{T_K}}^{-1} \circ \psi^r_{T_K}$ is an isomorphism with $\operatorname{span}\psi^r_{T_K}(\theta^d) \to \operatorname{span}\psi^d_{T_K}(\theta^d)$, as both transformations are bijective and the spaces are finite dimensional. For the constants, we use the same strategy as discussed for the proof of Proposition 2.3.7 (c.f. also Remark 2.3.5 and Lemma 2.3.6). We split $T_K = S_K \circ T_{\tilde{K}}$ where $\operatorname{diam} T_{\tilde{K}}(\hat{K}) = 1$ and $S_K \colon x \mapsto h_K x$. As ${\psi^d_{T_{\tilde{K}}}}^{-1} \circ \psi^r_{T_{\tilde{K}}}$ does not depend on $\operatorname{diam} K$, we then have

$$\begin{aligned}
\|\big({\psi^d_{T_K}}^{-1} \circ \psi^r_{T_K}\big)(v)\|_{0,K} &= \|\big({\psi^d_{S_K \circ T_{\tilde{K}}}}^{-1} \circ \psi^r_{S_K \circ T_{\tilde{K}}}\big)(v)\|_{0,K} && (2.148)\\
&= \|\big({\psi^d_{S_K}}^{-1} \circ {\psi^d_{T_{\tilde{K}}}}^{-1} \circ \psi^r_{T_{\tilde{K}}} \circ \psi^r_{S_K}\big)(v)\|_{0,K} && (2.149)\\
&\leq h_K^{n-1}\sqrt{h_K^n}\|\big({\psi^d_{T_{\tilde{K}}}}^{-1} \circ \psi^r_{T_{\tilde{K}}} \circ \psi^r_{S_K}\big)(v)\|_{0,\tilde{K}} && (2.150)\\
&\leq h_K^{n-1}\sqrt{h_K^n}c_2\|\psi^r_{S_K}(v)\|_{0,\tilde{K}} && (2.151)\\
&= c_2\|v\|_{0,K} && (2.152)
\end{aligned}$$

by Lemma 2.1.19. The second estimates follow by the same strategy and the choice

$$v = \left(\left({\psi^d_{T_K}}^{-1}\psi^r_{T_K}\right)^{-1}\left({\psi^d_{T_K}}^{-1}\psi^r_{T_K}\right)\right)(v). \tag{2.153}$$

□

Remark 2.3.9. In comparison to Fig. 2.2, the polynomial space induced by the covariant and contravariant transformation originates from the same basis θ^d. In this regard the Eq. (2.145) might appear confusing, but actually hints that θ^d is closely related to the basis functions of the Nédélec element [94] on a cube, rather than the Raviart–Thomas analogues (c.f. Remark 2.2.15 and [3]). Furthermore, it is straightforward to verify that the Nédélec basis functions span a subspace of $\operatorname{span}\theta_d$.

For the case of Cartesian grids, we additionally identify the polynomial space induced by the corresponding transformation and their counterparts on the reference element.

Remark 2.3.10. Let $\mathcal{T}_h$ be a Cartesian grid. Let $C \in \mathcal{T}_h$ and $T_C \colon \hat{K} \to C$ denote the transformation discussed in Remark 2.1.9 then θ^d_C and ${\psi^r_{T_C}}^{-1} \circ \sigma^d$ is a basis of $\mathcal{P}^d$.

Proof. The Jacobian matrix DT_C is constant and of diagonal form. Next we observe that in

this case $\theta_i^d \circ T_C^{-1} \in \mathcal{P}^d$ as we simply scale and shift the coordinates. Multiplication by a constant diagonal matrix obviously does not change the polynomial degree too. Therefore, we find

$$\theta_{C,i}^d(x) = ({\psi_{T_C}^d}^{-1}\theta_i^d)(x) = \frac{1}{\det DT_C} DT_C(\theta_i^d \circ T_C^{-1})(x) \in \mathcal{P}^d \tag{2.154}$$

The second part of the proof follows the same reasoning. □

Remark 2.3.11. There is an affine transformation T_Q and $i \in I$ such that $\theta_{Q,i}^d$ or $\theta_{Q,i}^r$ are not element of $\mathcal{P}^d$.

Remark 2.3.12. Without additional restriction, we obtain similar results as presented in Proposition 2.3.7 for the pressure element. Unfortunately this is a suboptimal result in the sense that we only bound the $L^2(K)$-norm. The classical result indeed gives

$$\|q - \mathcal{I}_K q\|_{1,K} \leq c h_K \|q\|_{2,K} \tag{2.155}$$

even in the isoparametric case [37, 62], which therefore is also applicable to the pressure variable in our situation.

The fact that we bound the interpolation error of the pressure to a higher degree, then the velocity variable seems undesirable at this point. As it will turn out in Proposition 2.4.17, we indeed can improve the error estimate of the velocity variable in a mesh dependent analogue to the $H(\mathrm{div}, K)$ norm.

2.4 A pair of finite element spaces

Having established the local construction on the reference element and as well as elements induced by smooth transforms thereof, we conveniently are able to construct the global interpolation operators via their local counterparts. For the sake of notation, we introduce the following approximation spaces.

Definition 2.4.1 (Approximation spaces)**.** Let $\mathcal{T}_h$ be a grid as introduced in Definition 2.1.1. Then the parametric finite element spaces for the velocity and pressure variable are defined as

$$\mathcal{D}_h := \left\{ v \subset \Omega \times \mathbb{R}^n : {\psi_K^d}^{-1}\left(v|_K\right) \in \operatorname{span} \theta^d, \forall K \in \mathcal{T}_h \right\}, \tag{2.156}$$

$$\mathcal{W}_h^1 := \left\{ q \in H^1(\Omega) : \psi_K^{-1}\left(q|_K\right) \in \operatorname{span} \theta_K, \forall K \in \mathcal{T}_h \right\}, \tag{2.157}$$

$$\mathcal{W}_{h,0}^1 \coloneqq \left\{ q \in \mathcal{W}_h^1 \colon \int_\Omega q \, dx = 0 \right\}. \tag{2.158}$$

Here, elements of $\mathcal{D}_h$ are possibly multivalued. Additionally, we define

$$\mathcal{R}_h \coloneqq \left\{ v \subset \Omega \times \mathbb{R}^n \colon {\psi_{T_K}^r}^{-1}\big(v|_K\big) \in \operatorname{span} \theta^d, \forall K \in \mathcal{T}_h \right\}, \tag{2.159}$$

$$\mathcal{W}_h'^0 \coloneqq \left\{ q \subset \Omega \times \mathbb{R} \colon \psi_{T_K}^{-1}\big(q|_K\big) \in \operatorname{span}\{1\}, \forall K \in \mathcal{T}_h' \right\}. \tag{2.160}$$

Considering Remark 2.3.10 we observe the following

Remark 2.4.2. If $\mathcal{T}_h$ is a Cartesian grid as introduced in Remark 2.1.9, then $\mathcal{D}_h$ and $\mathcal{R}_h$ can be identified by

$$\mathcal{D}_h = \mathcal{R}_h = \left\{ f \in L^1(\Omega)^n \colon \ f|_C \in \mathcal{P}^d, \forall C \in \mathcal{T}_h \right\}. \tag{2.161}$$

Analogue simplifications occur in the Cartesian case for the other approximation spaces as well.

Remark 2.4.3. $\mathcal{R}_h$ and $\mathcal{D}_h$ are not necessarily subspaces of $H^1(\Omega)^n$ or $H(\operatorname{div}, \Omega)$. This is an immediate consequence of the discontinuous (normal and tangential) traces across the element boundary.

2.4.1 Differential operators

For our discretization, we aim to use the coordinate free representation of the divergence div. On one hand, this choice fits the conservation property to the finite volume type discretization discussed in this work. On the other hand, we avoid the issue of deteriorated convergence rates. In the context of low order methods based on Raviart–Thomas interpolation the mimetic divergence [108, 116] does achieve convergence [7, 25]. This will also hold true for our method.

In line with [10] we recall the definition as

$$\operatorname{DIV}(w)(x) \coloneqq \lim_{r \to 0} \frac{1}{|B_r|} \int_{\partial B_r(x)} \tau_{n_{\partial B_r(x)}}(w) \, ds. \tag{2.162}$$

Remark 2.4.4. Let $\Omega \subset \mathbb{R}^n$ be a bounded Lipschitz domain. Then the differential operator in Eq. (2.162) is well-defined on $W^{1,1}(U) \cap W^{1,\infty}(\Omega)$. This is more restrictive than the previously used space $H^1(\Omega)$. Nevertheless, any piecewise polynomial function on Ω satisfies this additional requirement.

Remark 2.4.5. Let $\Omega \subset \mathbb{R}^n$ be a bounded Lipschitz domain. Let $w \in \mathcal{C}^1(\Omega)^n$. Then $\mathrm{DIV}(w)$ and $\mathrm{div}(w)$ coincide.

Proof. This is an immediate consequence of Gauß' and Stokes' theorem as well as the mean value theorem i.e., for every open ball $B_r(x) \subset \Omega$ there is $\xi \in B_r(x) \subset \Omega$ with

$$\begin{aligned}
\mathrm{DIV}(w)(x) &= \lim_{r\to 0} \frac{\int\limits_{\partial B_r(x)} \tau_{n_{\partial B_r(x)}}(w)\, ds}{|B_r|} && (2.163)\\
&= \lim_{r\to 0} \frac{\int\limits_{B_r} \mathrm{div}(w)\, dx}{|B_r|} && (2.164)\\
&= \lim_{r\to 0} \mathrm{div}(w)(\xi) \frac{\int\limits_{B_r} dx}{|B_r|} && (2.165)\\
&= div(w)(x) && (2.166)
\end{aligned}$$

□

Definition 2.4.6. Let $\mathcal{T}_h$ be a grid as introduced in Definition 2.1.1 and let $\mathcal{T}_h'$ denote its dual. Let $w \in H^1(\mathcal{T}_h)$, the broken Sobolev space introduced in Definition A.6.1. Then we define the discrete analogue (c.f. [25]) to DIV as

$$\mathrm{DIV}_h : \begin{cases} H^1(\mathcal{T}_h) \to {\mathcal{W}_h'}^0 \\ w \mapsto \sum\limits_{K' \in \mathcal{T}_h'} \frac{\mathbb{1}_{K'}}{|K'|} \int\limits_{\partial K'} \tau_{n_{\partial K'}}(w)\, ds. \end{cases} \tag{2.167}$$

The unique extension of this operator (c.f. Lemma A.4.3) to $\mathcal{B}(H(\mathrm{div}, \mathcal{T}_h), {\mathcal{W}_h'}^0)$ is denoted by DIV_h too.

As next step, we introduce a mesh dependent semi-norm based on DIV_h. The following result is a generalization of the work in [3, 118].

Proposition 2.4.7. *Let $\mathcal{T}_h$ be a grid as defined in Definition 2.1.1, then*

$$|v|_{\mathrm{DIV}_h} \coloneqq \|\mathrm{DIV}_h\, v\|_{L^2(\Omega)} \tag{2.168}$$

is a semi-norm on $H(\mathrm{div}, \mathcal{T}_h)$.

Proof. $\mathrm{DIV}(w)$ is well-defined and determined by integrals i.e. linear operators. We, therefore, obtain homogeneity, and by applying the triangle inequality, we obtain subadditivity

too, since $\|\cdot\|_{L^2(T(\Omega))}$ is already a norm. □

Remark 2.4.8. In line with [3, 118] and again due to the fact that $\|\cdot\|_{L^2(\Omega)}$ is a norm on $L^2(\Omega)$ we can introduce a norm on $H(\mathrm{div}, \mathcal{T}_h)$ by

$$\|w\|^2_{\mathrm{DIV}_h} \coloneqq \|w\|^2_0 + |w|^2_{\mathrm{DIV}_h}. \tag{2.169}$$

2.4.2 Global interpolation

In line with standard finite element theory [54], we construct the global interpolation operators by their local counterparts.

Definition 2.4.9 (Global interpolation)**.** Let $\mathcal{T}_h$ be a grid as given in Definition 2.1.1 and let $s > n/2$. Then we define the global interpolation operators via

$$\mathcal{I}^d_h \colon \begin{cases} H(\mathrm{div}, \mathcal{T}_h) \to \mathcal{D}_h \\ v \mapsto \sum\limits_{K \in \mathcal{T}_h} \mathbb{1}_K \mathcal{I}^d_K v \end{cases} \quad , \quad \mathcal{I}_h \colon \begin{cases} H^s(\Omega) \to \mathcal{W}^1_h \\ v \mapsto \sum\limits_{K \in \mathcal{T}_h} \mathbb{1}_K \mathcal{I}_K v. \end{cases} \tag{2.170}$$

Additionally, we introduce the lowest order Clement quasi interpolation operator [39, 20, 19] onto the piecewise constant functions $\mathcal{W}'^0_h$

$$\mathcal{I}^0_{\mathcal{T}'_h} \colon \begin{cases} L^1(\Omega) \to \mathcal{W}'^0_h \\ q \mapsto \sum\limits_{K' \in \mathcal{T}'_h} \mathbb{1}_{K'} \operatorname{avg}_{K'} q. \end{cases} \tag{2.171}$$

Remark 2.4.10. The global approximation space for the pressure variable is conforming to $H^1(\Omega)$ i.e., $\mathcal{W}^1_h \subset H^1(\Omega)$.

Remark 2.4.11. The family of tent functions is a basis of $\mathcal{W}^1_h$. In abuse of notation we denote the basis functions by θ_ξ for every $\xi \in \mathcal{N}_{\mathcal{T}_h}$, since each basis function is composed by the shape functions $\theta_{\xi,K}$ for $K \in \mathcal{K}_{\xi,\xi}$. Their support of θ_ξ is included in the union of all elements $K \in \mathcal{K}_{\xi,\xi}$ and $\theta_\xi(\xi) = 1$.

Remark 2.4.12. One readily observes the following element wise commuting property

$$\mathcal{I}^0_{K'} \operatorname{avg}_{K'} q = \operatorname{avg}_{K'} \mathcal{I}^0_{K'} q \tag{2.172}$$

for every $q \in L^1(\Omega)$ and $K' \in \mathcal{T}'_h$.

The following result delivers convergence in the mesh dependent norm $\|\cdot\|_{\mathrm{DIV}_h}$ and follows the idea of [25] in the conforming setting.

Lemma 2.4.13. *Let $\mathcal{T}_h$ be as given in Definition 2.1.1 and let $\mathcal{T}_h'$ denote its dual. Then we have*

$$\mathrm{DIV}_h(\mathcal{I}_h^d(v)) = \mathcal{I}^0_{\mathcal{T}_h'} \mathrm{DIV}_h(v), \tag{2.173}$$

for every $v \in H(\mathrm{div}, \mathcal{T}_h)$. If we assume $v \in H(\mathrm{div}, \Omega)$, we additionally have

$$\mathrm{DIV}_h(\mathcal{I}_h^d(v)) = \mathcal{I}^0_{\mathcal{T}_h'} \mathrm{div}(v). \tag{2.174}$$

Proof. Using Lemma 2.3.2 we establish the statement for elements of $H^1(K)$ first. To this end we see

$$\begin{aligned}
\mathrm{DIV}_h\, \mathcal{I}_h^d(w) &= \mathrm{DIV} \sum_{K\in\mathcal{T}_h} \sum_{i\in I} \sigma_{K,i}^d(w)\theta_{K,i}^d && (2.175)\\
&= \sum_{K\in\mathcal{T}_h} \sum_{K\cap K'\neq\emptyset} \frac{\mathbb{1}_{K'}}{|K'|} \sum_{i\in I} \sigma_{K,i}^d(w) \int_{\partial K'} \tau_{n_{\partial K'}}(\theta_{K,i}^d)\, d\mu && (2.176)\\
&= \sum_{K\in\mathcal{T}_h} \sum_{K\cap K'\neq\emptyset} \frac{\mathbb{1}_{K'}}{|K'|} \sum_{T_K(F_i)\subseteq\partial K'} \sigma_{K,i}^d(w) && (2.177)\\
&= \sum_{K\in\mathcal{T}_h} \sum_{K\cap K'\neq\emptyset} \frac{\mathbb{1}_{K'}}{|K'|} \int_{\partial K'\cap K} \tau_{n_{\partial K'}}(w)\, d\mu && (2.178)\\
&= \sum_{K'\in\mathcal{T}_h'} \sum_{K\cap K'\neq\emptyset} \frac{\mathbb{1}_{K'}}{|K'|} \int_{\partial K'\cap K} \tau_{n_{\partial K'}}(w)\, ds && (2.179)\\
&= \mathcal{I}^0_{\mathcal{T}_h'}(\mathrm{DIV}_h\, w). && (2.180)
\end{aligned}$$

If additionally $w \in H^1(\Omega)$, we are able to use Gauß' theorem on K' such that

$$\begin{aligned}
\mathcal{I}^0_{\mathcal{T}_h'}(\mathrm{DIV}_h\, w) &= \sum_{K'\in\mathcal{T}_h'} \frac{\mathbb{1}_{K'}}{|K'|} \sum_{K\cap K'\neq\emptyset} \int_{\partial K'\cap K} \tau_{n_{\partial K'}}(w)\, d\mu && (2.181)\\
&= \sum_{K'\in\mathcal{T}_h'} \frac{\mathbb{1}_{K'}}{|K'|} \int_{\partial K'} \tau_{n_{\partial K'}}(w)\, d\mu && (2.182)\\
&= \sum_{K'\in\mathcal{T}_h'} \frac{\mathbb{1}_{K'}}{|K'|} \int_{K'} \mathrm{div}(w)\, d\mu && (2.183)\\
&= \mathcal{I}^0_{\mathcal{T}_h'} \mathrm{div}(w). && (2.184)
\end{aligned}$$

By density of $H^1(K)$ in $H(\mathrm{div}, K)$, $H^1(\Omega)$ in $H(\mathrm{div}, \Omega)$ and Corollary 2.1.24 we obtain the statements. □

Remark 2.4.14. Although not stated in the form of DIV_h, and not done locally on a reference element, the interpolation operator provided in [3] is constructed by requiring exactly the aforementioned commuting property.

Corollary 2.4.15. *Let $\mathcal{T}_h$ be a shape-regular and quasi-uniform family of grids, each one of the form as defined in Definition 2.1.1 and with $\operatorname{diam}(K) \leq h$ for every $K \in \mathcal{T}_h$. Then the interpolation error satisfies*

$$\|w - \mathcal{I}^d_{h_n} w\|_{\mathrm{DIV}_{h_n}} \leq c h_n |w|_{H^1(\mathcal{T}_h)} \tag{2.185}$$

for every $w \in H^1(\Omega)$.

Proof. For every $h > 0$ we have

$$\begin{aligned}
\|w - \mathcal{I}^d_h w\|^2_{\mathrm{DIV}_h} &= \|w - \mathcal{I}^d_h w\|^2_{0,\Omega} + \|\mathrm{DIV}_h w - \mathrm{DIV}_h \mathcal{I}^d_h w\|^2_{0,\Omega} && (2.186)\\
&= \|w - \mathcal{I}^d_h w\|^2_{0,\Omega} + \|\mathcal{I}^0_{\mathcal{T}_h'} \mathrm{DIV}_h w - \mathcal{I}^0_{\mathcal{T}_h'} \mathrm{DIV}_h \mathcal{I}^d_h w\|^2_{0,\Omega} && (2.187)\\
&= \|w - \mathcal{I}^d_h w\|^2_{0,\Omega} + \|\mathcal{I}^0_{\mathcal{T}_h'} \mathrm{DIV}_h w - \mathcal{I}^0_{\mathcal{T}_h'} \mathrm{DIV}_h w\|^2_{0,\Omega} && (2.188)\\
&= \|w - \mathcal{I}^d_h w\|^2_{0,\Omega} && (2.189)\\
&\leq c h |w|^2_{H^1(\mathcal{T}_h)}, && (2.190)
\end{aligned}$$

where the last bound follows by Lemma 2.2.14, shape-regularity and quasi-uniformity. □

Remark 2.4.16. One helpful conclusion we can gain from Eq. (2.174) is that interpolants of $u \in H(\mathrm{div}, \Omega)$ with $\mathrm{div}\, u = 0$ are divergence free in the discrete sense of DIV_h.

Using Corollary 2.4.15 we now are able to state a bound on the global interpolation error. Having Remark 2.3.12 in mind we avoid the question of explicit characterizations of sufficiently shape regular grids and use the implicit definition using the constants of the local interpolation operators. Besides the obvious candidates of sufficiently shape regular families of affine grids also grids consisting of quadrilaterals [37] or cuboids can satisfy the assumptions of the following.

Proposition 2.4.17. *Let $\mathcal{T}_h$ be a shape-regular, quasi-uniform family of grids, each one of the form as defined in Definition 2.1.1 and with $\operatorname{diam}(K) \leq h$ for every $K \in \mathcal{T}_h$. Let $v \in H^1(\Omega)$ and $q \in H^2(\Omega)$, then the global interpolation error is bound by*

$$\|v - \mathcal{I}^d_h v\|_{\mathrm{DIV}_h} + \|q - \mathcal{I}_h q\|_{1,\Omega} \leq c h_K \big(|v|_{1,\Omega} + |q|_{2,\Omega}\big). \tag{2.191}$$

Proof. We simply sum the squares of the element wise estimates Proposition 2.3.7 and Remark 2.3.12 and apply Corollary 2.4.15 obtained by the commuting property of DIV_h and $\mathcal{I}_h^d$. □

Lemma 2.4.18. *Let $\mathcal{T}_h$ be a grid as given in Definition 2.1.1, then there is a linear isomorphism $L^{r,d}$ independent of h such that*

$$L^{r,d}\colon \begin{cases} \mathcal{R}_h \to \mathcal{D}_h \\ v \mapsto \sum\limits_{K\in\mathcal{T}_h} {\psi^d_{T_K}}^{-1} \circ \psi^r_{T_K}(v|_K). \end{cases} \tag{2.192}$$

Furthermore, $L^{r,d}$ is bounded by a constant $c > 0$ independent of h such that

$$\|L^{r,d}v\|_0 \le ch^{2-n}\|v\|_0 \tag{2.193}$$

for every $v \in \mathcal{R}_h$.

Proof. This map is an isomorphism due to Remark 2.1.15. Every transformation T_K can be decomposed in a scaling S_K and a deformation D_K which does not change the diameter and a constant offset. Without loss of generality, we ignore the offset and assume $T_K = S_K \circ D_K$. For the scaling $S_K\colon x \mapsto h_K(x - \mathrm{avg}_K(x))$ we obtain

$$\begin{aligned} {\psi^d_{T_K}}^{-1} \circ \psi^r_{T_K} &= \psi^d_{{S_K}^{-1}} \psi^d_{{D_K}^{-1}} \circ \psi^r_{D_K} \circ \psi^r_{S_K} && (2.194)\\ &= \frac{h_K\,\mathbb{I}}{h_K^2} \circ \psi^d_{{D_K}^{-1}} \circ \psi^r_{D_K} \circ (h_K\,\mathbb{I}) && (2.195)\\ &= \psi^d_{{D_K}^{-1}} \circ \psi^r_{D_K}. && (2.196) \end{aligned}$$

and as D_K is independent of h_K therefore ${\psi^d_{T_K}}^{-1} \circ \psi^r_{T_K}$ and $L^{r,d}$.

The continuity constant can be determined by

$$\begin{aligned} \|L^{r,d}v\|_0^2 &= \sum_{K\in\mathcal{T}_h} \int_K \left({\psi^d_{T_K}}^{-1} \circ \psi_{r,K}(v)\right)^T {\psi^d_{T_K}}^{-1} \circ \psi^r_K(v)\,dx && (2.197)\\ &= \sum_{K\in\mathcal{T}_h} \int_K v^T \frac{\left(DT_K DT_K^T DT_K DT_K^T\right) \circ T_K^{-1}}{\det(DT_K \circ T_K^{-1})^2} v\,dx && (2.198) \end{aligned}$$

Splitting the transformation into a pure scaling by h and a transformation to $T_{\tilde{K}}$ with

$\operatorname{diam} \tilde{K} = 1$ the last expression can be bound from above by

$$\sum_{K \in \mathcal{T}_h} h_K^{4-2n} c_K \int_K v^T v \, dx \tag{2.199}$$

where $c_K > 0$ does not depend on h. Subsequently, quasi-uniformity and shape-regularity deliver the statement. □

Remark 2.4.19. $\operatorname{grad}: \mathcal{W}_h^1 \to R_h$ and the classical identity

$$\operatorname{curl} \operatorname{grad} q = 0 \tag{2.200}$$

is satisfied for every $q \in \mathcal{W}_h^1$.

3 Analysis of the projection step

Henceforth, we investigate the mixed saddle point formulation of the projection step by the means of the approximation spaces and interpolation operators introduced in the previous chapter.

3.1 Variational formulation

As first step we introduce the exact formulation and appropriate function spaces for the analytical problem.

3.1.1 Analytical problem

Consider the Hilbert spaces $\mathcal{U} = H(\mathrm{div}, \Omega)$, $\mathcal{H} = \{q \in H^1(\Omega) \colon \int_\Omega q = 0\}$. We aim to state Eq. (1.30) in variational form. To this end we first divide by $c_p(P\theta)^{n+1}$, which we assume to be positive almost everywhere. Subsequently, we search for ${\pi'}^{n+1} \in \mathcal{H}$ and $(Pv)^{n+1} \in \mathcal{U}$ such that

$$a((Pv)^{n+1}, w) + b_1({\pi'}^{n+1}, w) = f(w), \tag{3.1a}$$

$$b_2((Pv)^{n+1}, q) + c({\pi'}^{n+1}, q) \quad = g(q) \tag{3.1b}$$

for every $w \in L^2(\Omega)^n$ and $q \in L^2(\Omega)$. Here we substitute Eq. (1.28c) into Eq. (1.28b) to obtain

$$\begin{aligned} a((Pv)^{n+1}, w) = &\int_\Omega \frac{w^T \Sigma_W (Pv)^{n+1}}{c_p (P\theta)^{n+1}}\, dx + \frac{\Delta t}{2} \int_\Omega \frac{f_0}{c_p \theta^{n+1}} (w \times e_3)^T (Pv)^{n+1}\, dx \\ &- \frac{\Delta t^2}{4} \int_\Omega w^T \frac{g}{c_p} \partial_{x_3} \overline{\chi}\, e_3 \otimes e_3 (Pv)^{n+1}\, dx \end{aligned} \tag{3.2a}$$

$$b_1({\pi'}^{n+1}, w) = \int_\Omega w^T \frac{\Delta t}{2} \operatorname{grad} {\pi'}^{n+1}\, dx \tag{3.2b}$$

$$b_2((Pv)^{n+1}, q) = \int_\Omega \frac{\Delta t}{2} \operatorname{div}(Pv)^{n+1}\, q dx \tag{3.2c}$$

$$c(p, q) = \int_\Omega \alpha_P \left(\frac{\partial P}{\partial \pi}\right)^\circ q {\pi'}^{n+1}\, dx \tag{3.2d}$$

$$f(w) = \int_\Omega \frac{w^T \Sigma_W (Pv)^*}{c_p (P\theta)^{n+1}} - \frac{\Delta t}{2} \frac{g}{c_p} (P\chi')^* w^T e_3\, dx \tag{3.2e}$$

$$g(q) = \int_\Omega \alpha_P \left(\frac{\partial P}{\partial \pi}\right)^\circ q {\pi'}^{n}\, dx. \tag{3.2f}$$

Remark 3.1.1. In Eq. (3.2a) there is some $(Pv)^{n+1} \in H(\operatorname{div}, \Omega)$ such that $e_3 \times (Pv)^{n+1} \notin H(\operatorname{div}, \Omega)$. Although not necessary at this point we rotate the test function w instead, using $(w \times e_3)^T (Pv)^{n+1} = w^T (e_3 \times (Pv)^{n+1})$, to avoid any confusion and make continuity of a more obvious.

To simplify notation, we henceforth will only investigate the following system. Let $\delta, \omega \in L^\infty(\Omega)^{n \times n}$ such that δ is a symmetric positive definite matrix almost everywhere. Let $\tau = \Delta t/2 > 0$ represent the time step parameter, $\alpha_P \geq 0$ and let $\zeta \in L^\infty(\Omega)$ with $\operatorname{ess\,inf} \zeta > 0$. Ignoring the physical meaning of u, v, p, we introduce the following bilinear forms

$$a(u, v) = \int_\Omega v^T (\delta + \tau\omega) u\, dx \tag{3.3a}$$

$$b_1(p, v) = \int_\Omega v^T \operatorname{grad} p\, dx \tag{3.3b}$$

$$b_2(u, q) = \int_\Omega q \operatorname{div}(u)\, dx \tag{3.3c}$$

$$c(p, q) = \int_\Omega \zeta q p\, dx. \tag{3.3d}$$

Given two continuous linear functionals $f \in \mathcal{U}', g \in \mathcal{H}'$, we now search for $u \in \mathcal{U}$ and $p \in \mathcal{H}$ such that Eq. (3.1) is true for every $v \in \mathcal{U}$ and $q \in \mathcal{H}$.

Discrete problem

For the discretization we aim to generalize a discrete approximation of Eq. (3.1) originally developed by the means of the cell centre finite volume method [113]. For this purpose

we use the discontinuous Petrov Galerkin finite element proposed by Vater and Klein in [118]. They suggest replacing the divergence operator acting on u by boundary integrals around dual cells. More specifically they propose to replace $b_2(v, q)$ by the mesh dependent bilinear form

$$b_{2,h}(v, q) = \int_\Omega q \, \mathrm{DIV}_h(v) \, dx = \sum_{K' \in \mathcal{T}_h'} \mathrm{avg}_{K'}(q) \int_{\partial K'} \tau_{n_{K'}}(v) \, d\mu \tag{3.4}$$

where $\xi_{K'}$ is the node inside the dual cell K'.

Remark 3.1.2. In contrast to the discontinuous Petrov Galerkin methods introduced by [46, 47] the scheme introduced by [118] is not based on an ultra weak formulation, where all differential operators act only on test functions.

Remark 3.1.3. Henceforth, assume the functions δ, ω and ζ to be piecewise constant on the reference element. In case of the general setting one can extend the following error analysis by introducing additional linear- and bilinear forms a_h,c_h, g_h and f_h. Under suitable assumptions the error between those and their analytical counterparts can be bound by terms of $O(h)$. This however is out of scope for the present work.

3.1.2 Boundary conditions

As we have not yet discussed any boundary conditions, we will do so for two important cases. The first one is determined by an impermeable wall i.e., $u \cdot n_{\partial\Omega} = 0$ on $\partial_n\Omega$. This is similar to the Neumann boundary condition in the pressure, which are so-called natural boundary conditions. Therefore, we do not impose this boundary condition on the functions $u \in \mathcal{D}_h$ directly, but instead apply it to the bilinear form $b_{2,h}$. This is in contrast to the essential Dirichlet boundary conditions on the pressure variable $p \in \mathcal{W}_h^1$. Therefore, we restrict $\mathcal{W}_h^1$ to a subspace of functions satisfying the boundary condition at the corresponding part of the boundary. Although we do not consider a Dirichlet boundary explicitly, concerning the second case of periodic boundary conditions, it might help to illustrate that they are a combination of natural and essential boundary conditions. To this end, we consider $\partial_{\mathrm{per}}\Omega \subset \partial\mathcal{T}_h$ containing an even number of nodes and identify each node $\xi \in \mathcal{N}_{\partial_{\mathrm{per}}\Omega}$ with a single other node $\xi' \in \mathcal{N}_{\partial_{\mathrm{per}}\Omega}$. This identification is done in such a way that for every continuous function $q \in C^1(\Omega)$ the periodic extension defined by this identification is continuous. Consequently, the approximation space for the pressure with

periodic boundary conditions reads

$$\mathcal{W}^1_{h,0,\mathrm{bc}} = \left\{ q \in \mathcal{W}^1_h \colon \int_\Omega q \, dx = 0 \colon q(\xi) = q(\xi') \quad \forall \xi \in \mathcal{N}_{\partial_{\mathrm{per}}\Omega} \right\}. \tag{3.5}$$

Remark 3.1.4. The continuity of the periodic extension is important, as we discretize the pressure by a continuous variable and do not want to have any additional side constraints. Another way of expressing this requirement is that

$$\dim \mathcal{W}^1_{h,\mathrm{bc}} = \dim \mathcal{W}^1_h - \frac{|\mathcal{N}_{\partial_{\mathrm{per}}\Omega}|}{2}. \tag{3.6}$$

The velocity space, however, does not obtain any further restrictions as it is discontinuous across element boundaries.

Similarly to the pressure space we introduce the periodic version of $\mathcal{W}'^0_h$ via

$$\mathcal{W}'^0_{h,\mathrm{bc}} = \left\{ q \in \mathcal{W}'^0_h \colon q(\xi) = q(\xi') \quad \forall \xi \in \mathcal{N}_{\partial_{\mathrm{per}}\Omega} \right\}, \tag{3.7}$$

$$\mathcal{W}'^0_{h,0,\mathrm{bc}} = \left\{ q \in \mathcal{W}'^0_h \colon \int_\Omega q \, dx = 0 \colon q(\xi) = q(\xi') \quad \forall \xi \in \mathcal{N}_{\partial_{\mathrm{per}}\Omega} \right\}. \tag{3.8}$$

Finally, we observe the mesh dependent bilinear form respecting the boundary conditions to be

$$b_{2,h} \colon \mathcal{D}_h \times \mathcal{W}'^0_{h,0,\mathrm{bc}} \colon (u; q) \mapsto \sum_{\xi \in \mathcal{N}_{\mathcal{T}_h}} q(\xi) \tau \sum_{K \in \mathcal{K}_{\xi,\xi}} \sum_{\substack{F \in \mathcal{F}_{K,\xi} \\ F \not\subset \partial_n \Omega}} \int_F \tau_{n_F}(u) \, d\mu. \tag{3.9}$$

The proposed discrete problem then is to find $u \in \mathcal{D}_h$ and $p \in \mathcal{W}^1_{h,0,\mathrm{bc}}$ such that

$$a(u, v) + \tau b_1(p, v) = f(v) \tag{3.10a}$$

$$\tau b_{2,h}(u, q) + \alpha_P c(p, q) = g(q) \tag{3.10b}$$

for every $v \in \mathcal{D}_h$ and $q \in \mathcal{W}'^0_h$ as well as arbitrary continuous linear $f \colon (\mathcal{D}_h, \|\cdot\|_0) \to \mathbb{R}$ and $g \colon (\mathcal{W}'^0_h, \|\cdot\|_0) \to \mathbb{R}$.

In the following chapters we prove the existence of a unique solution of the approximate solutions provided by Eq. (3.10). In the case of $\alpha_P = 0$, we also prove convergence towards solutions of Eq. (3.1).

3.2 Stability

We aim to prove existence of a unique solution to Eq. (3.10). Since for the case $\alpha_P = 0$, Eq. (3.10) becomes a generalized saddle point problem, we follow the approach presented in [96, 21, 118] which is a generalization of the theory developed by Babuška and Brezzi in [11, 12, 28] and briefly summarized in Section A.7.

As it turns out, we have to investigate the properties and relations of the approximation spaces in some detail to prove well-posedness. As far as possible, we use functional analytic arguments instead of relying on concrete discrete reasoning. This, however, is not possible for every bilinear form. Especially in the case of c we need to use the coordinate form. The structure of the following chapter, aligns with the approach to prove each individual LBB-condition Definition A.7.4.

Remark 3.2.1. We denote the associated operators (c.f. Section A.7) for the bilinear forms a, b_1 and $b_{2,h}$ by A, B_1 and B_2, each mapping the approximation space into the topological dual of the test function space. Due to our definition of b_1, b_2 and the order of their arguments, we will give hints on the domain and range of the associated operators whenever there is cause of confusion.

Consulting Section A.7 again, we first identify the following null spaces

$$\kappa_1 = \left\{ v \in \mathcal{D}_h : b_1(q, v) = 0 \quad \forall q \in \mathcal{W}^1_{h,\mathrm{bc}} \right\} \tag{3.11}$$

$$\kappa_2 = \left\{ v \in \mathcal{D}_h : b_{2,h}(v, q) = 0 \quad \forall q \in \mathcal{W}'^0_h \right\}. \tag{3.12}$$

3.2.1 Properties of the null spaces

Remark 3.2.2. Both null spaces are already determined by a smaller set of test functions i.e.,

$$\kappa_1 = \left\{ v \in \mathcal{D}_h : b_1(q, v) = 0 \quad \forall q \in \mathcal{W}^1_{h,0,\mathrm{bc}} \right\}, \tag{3.13a}$$

$$\kappa_2 = \left\{ v \in \mathcal{D}_h : b_2(v, q) = 0 \quad \forall q \in \mathcal{W}'^0_{h,0,\mathrm{bc}} \right\}. \tag{3.13b}$$

Proof. $\kappa_1 \subseteq \left\{ v \in \mathcal{D}_h : b_1(q, v) = 0 \quad \forall q \in \mathcal{W}^1_{h,0,\mathrm{bc}} \right\}$ is clear as $\mathcal{W}^1_{h,0,\mathrm{bc}} \subset \mathcal{W}^1_{h,\mathrm{bc}}$. On the other hand we know for each $p \in \mathcal{W}^1_{h,\mathrm{bc}}$ there is $q \in \mathcal{W}^1_{h,0,\mathrm{bc}}$ such that $p - \operatorname{avg}_\Omega p = q$ and therefore $\operatorname{grad} p = \operatorname{grad}(p - \operatorname{avg}_\Omega p) = \operatorname{grad} q$. This provides the other inclusion.

Again we obtain $\kappa_2 \subseteq \left\{ v \in \mathcal{D}_h : b_2(v, q) = 0 \quad \forall q \in \mathcal{W}'^0_{h,0,\mathrm{bc}} \right\}$ due to $\mathcal{W}'^0_{h,0,\mathrm{bc}} \subset \mathcal{W}'^0_h$. Let

$q_c \equiv c \in \mathbb{R}$ a constant function, then by continuity of v (on every element K) we obtain

$$b_2(v, q_c) = \sum_{K' \in \mathcal{T}_h{}'} c \int_{\partial K' \setminus \partial\Omega} \tau_{n_{K'}}(v)\, d\mu = 0 \tag{3.14}$$

for every $v \in \mathcal{D}_h$. As $q \in \mathcal{W}_h'^0$ if and only if $q - \operatorname{avg}_\Omega q \in \mathcal{W}_{h,0,\mathrm{bc}}'^0$ we conclude $b_2(v, q - \operatorname{avg}_\Omega q) = b_2(v, q)$ for every $v \in \mathcal{D}_h$ and every $q \in \mathcal{W}_h'^0$. Therefore, the remaining inclusion $\kappa_2 \supseteq \left\{ v \in \mathcal{D}_h : b_2(v, q) = 0 \quad \forall q \in \mathcal{W}_{h,0,\mathrm{bc}}'^0 \right\}$ holds true.

□

A slightly richer approximation space for velocity variable is given by

$$\mathcal{U}_h \coloneqq \left\{ v \subset \Omega \times \mathbb{R}^n : {\psi_{T_K}^d}^{-1}\left(v|_K\right) \in \mathbb{Q}^n_{1,1,1}, \forall K \in \mathcal{T}_h \right\}. \tag{3.15}$$

Lemma 3.2.3. *Let $u \in \mathcal{U}_h / \mathcal{D}_h$ then*

$$b_1(p, u) = 0, \quad b_{2,h}(u, q) = 0 \tag{3.16}$$

for every $p \in \mathcal{W}_h^1$ and $q \in \mathcal{W}_h'^0$.

Proof. After transformation, using $\psi_{T_K}^d$, we consider the reference element $\hat{K}$ only. Let $u \in \mathbb{Q}^n_{\{1\}^n} / \mathcal{P}^d$, then there is $i, j, k \in \{1 \dots n\}$ with $k \neq i \neq j$ and $u = e_i x_i f(x_j, x_k)$. For every interior face F_l we then have

$$\int_{F_l} n_{F_l} \cdot e_i x_i f(x_j, x_k)\, d\mu = 0 \tag{3.17}$$

as either the normal n_{F_l} is orthogonal to e_i or $x \in F_l$ implies $x_i = 0$.

It is straightforward to verify that for every $u \in \mathbb{Q}^n_{\{1\}^n} / \mathcal{P}^d$ and every gradient of a shape function shape $\operatorname{grad} \theta_{\hat{\xi}}$, each term of $u \cdot \operatorname{grad} \theta_{\hat{\xi}}$ is odd in at least one component. Therefore, its integral over $\hat{K}$ vanishes i.e.,

$$\int_{\hat{K}} u \operatorname{grad} \theta_{\hat{\xi}} = 0 \tag{3.18}$$

for every $\hat{\xi} \in \mathcal{N}_{\hat{K}}$. □

Remark 3.2.4. As the additional degrees of freedom are included in κ_1, κ_2 the stability of Eq. (3.10) posed on $\mathcal{U}_h$ can be discussed by revisiting the proof of the $inf - \sup$ conditions of a. However, this out of the scope of this work.

3.2.2 Integration by parts

Before being able to state a stability result for the compressible regimes (i.e., $\alpha_P \neq 0$), we have to investigate a connection between DIV_h' and the classical gradient grad defined by integration by parts. We start by identification of the $\mathcal{W}_h'^0$ and $\mathcal{W}_h^1$ via a (mass) lumping operator treated e.g., in [71, Chapter 2].

Lemma 3.2.5. *Let $\mathcal{T}_h$ a shape-regular and quasi-uniform family grids as given in Definition 2.1.1. Then*

$$\Lambda\colon \begin{cases} \mathcal{W}_h^1 \to \mathcal{W}_h'^0 \\ q = \sum\limits_{\xi\in\mathcal{N}_{\mathcal{T}_h}} q_\xi \sigma_\xi \mapsto \sum\limits_{\xi\in\mathcal{N}_{\mathcal{T}_h}} q(\xi)\, \mathbb{1}_{K'_\xi} \end{cases} \tag{3.19}$$

is a linear isomorphism with constants $c_\Lambda, c'_\Lambda > 0$ and

$$\sup_{p\in\mathcal{W}_h^1\setminus\{0\}} \frac{(\Lambda^{-1}(p), q)}{\|p\|_0} \geq c_\Lambda \|q\|_0 \quad \forall q \in \mathcal{W}_h'^0 \tag{3.20}$$

$$\sup_{q\in\mathcal{W}_h'^0\setminus\{0\}} \frac{(\Lambda^{-1}(p), q)}{\|q\|_0} \geq c'_\Lambda \|p\|_0 \quad \forall p \in \mathcal{W}_h^1. \tag{3.21}$$

Furthermore, there are constants $c_i > 0$ independent of h for $i = \{1 \dots 4\}$ satisfying

$$\|\Lambda p\|_0 \leq c_1 \|p\|_0 \tag{3.22}$$

for every $p \in \mathcal{W}_h^1$ and

$$\|\Lambda^{-1} q\|_0 \leq c_2 \|q\|_0 \tag{3.23}$$

for every $q \in \mathcal{W}_h'^0$ and more specifically

$$\|\Lambda^{-1} q\|_0 \leq c_3 \sqrt{h_k^n} \sqrt{\sum_{\xi\in\mathcal{N}_{\mathcal{T}_h}} q_\xi^2} \tag{3.24}$$

$$\|\Lambda p\|_0 \leq c_4 \sqrt{h_k^n} \sqrt{\sum_{\xi\in\mathcal{N}_{\mathcal{T}_h}} p_\xi^2}. \tag{3.25}$$

Proof. First we observe $\dim \mathcal{W}_h'^0 = \dim \mathcal{W}_h^1$ as both spaces contain only nodal degrees of freedom and therefore are isomorphic to $\mathbb{R}^m$, where $m = |\mathcal{N}_{\mathcal{T}_h}| - |\mathcal{N}_{\partial_{\mathrm{per}}}|$. The map is linear by construction and as it is finite dimensional $\ker \Lambda = \{0\}$ already provides the existence of a unique inverse. Since Λ changes only the basis vectors the former and therefore the latter is true. Λ being a linear isomorphism provides existence of two constants c_Λ, c'_Λ in

Eqs. (3.20) and (3.21).

We denote the dual control volume around the node ξ by K'_ξ and observe

$$\|\Lambda p\|_0^2 = \int_\Omega \Big(\sum_{\xi\in\mathcal{N}_{\mathcal{T}_h}} p_\xi \, \mathbb{1}_{K'_\xi}\Big)^2 \, dx \tag{3.26}$$

$$= \sum_{\xi\in\mathcal{N}_{\mathcal{T}_h}} \int_{K'_\xi} p_\xi^2 \, dx \tag{3.27}$$

$$= \sum_{\xi\in\mathcal{N}_{\mathcal{T}_h}} |K'_\xi| p_\xi^2 \tag{3.28}$$

$$= \sum_{\xi\in\mathcal{N}_{\mathcal{T}_h}} \sum_{K\in\mathcal{K}_{\xi,\xi}} |K\cap K'_\xi| p_\xi^2 \tag{3.29}$$

and therefore there are constants $c_1, c_2 > 0$ such that

$$c_1 h^n \sum_{\xi\in\mathcal{N}_{\mathcal{T}_h}} p_\xi^2 \le \|\Lambda p\|_0^2 \le c_2 h^n \sum_{\xi\in\mathcal{N}_{\mathcal{T}_h}} p_\xi^2. \tag{3.30}$$

Next we express the norm in terms of the element-wise mass matrix $\hat{m}_K \in \mathbb{R}^{2^n\times 2^n}$ satisfying

$$\hat{m}_{K,i,j} = \int_{\hat{K}} \theta_i\theta_j \det DT_K \, dx \tag{3.31}$$

for every $i, j \in \{1 \dots 2^n\}$.

$$\|p\|_0^2 = \int_\Omega \Big(\sum_{\xi\in\mathcal{N}_{\mathcal{T}_h}} p_\xi\theta_\xi\Big)^2 \, dx \tag{3.32}$$

$$= \int_\Omega \Big(\sum_{K\in\mathcal{T}_h} \mathbb{1}_K \sum_{\xi\in\mathcal{N}_K} p_\xi\theta_\xi\Big)^2 \, dx \tag{3.33}$$

$$= \sum_{K\in\mathcal{T}_h} \int_K \Big(\sum_{\xi\in\mathcal{N}_K} p_\xi\theta_\xi\Big)^2 \, dx \tag{3.34}$$

$$= \sum_{K\in\mathcal{T}_h} \int_{\hat{K}} \Big(\sum_{\xi_i\in\mathcal{N}_K} p_{\xi_i}\theta_i\Big)^2 \det DT_K \, dx \tag{3.35}$$

$$= \sum_{K\in\mathcal{T}_h} \sum_{\xi_i,\xi_j\in\mathcal{N}_K} p_{\xi_i} m_{K,i,j} p_{\xi_j} \tag{3.36}$$

as usual we split the transformation T_K into a scaling S_K and $T_{\tilde{K}}$ such that $T_K = S_K T_{\tilde{K}}$ and

$\operatorname{diam} \tilde{K} = 1$. Then we observe $m_K = h_K^n m_{\tilde{K}}$. This enables us to provide bounds in terms of smallest and largest eigenvalues i.e.,

$$h_K^n \min_{l \in \{1 \dots 2^n\}} \lambda_l(m_{\tilde{K}}) \sum_{\xi_i \in \mathcal{N}_K} p_{\xi_i}^2 \leq \sum_{\xi_i, \xi_j \in \mathcal{N}_K} p_{\xi_i} m_{K,i,j} p_{\xi_j} \leq h_K^n \max_{l \in \{1 \dots 2^n\}} \lambda_l(m_{\tilde{K}}) \sum_{\xi_i \in \mathcal{N}_K} p_{\xi_i}^2. \tag{3.37}$$

It is straightforward to verify that $m_{\tilde{K}}$ is symmetric and has only positive eigenvalues and therefore this applies equally to m_K. Considering shape-regularity and quasi-uniformity, we collect the contributions of neighbouring cells to each node and obtain two constants $c_1, c_2 > 0$ independent of h such that

$$c_1 h^n \sum_{\xi \in \mathcal{N}_{\mathcal{T}_h}} p_\xi^2 \leq \|p\|_0^2 \leq h^n c_2 \sum_{\xi \in \mathcal{N}_{\mathcal{T}_h}} p_\xi^2. \tag{3.38}$$

Together with Eq. (3.30) we obtain

$$\tilde{c}_1 \|\Lambda p\|_0 \leq \|p\|_0 \leq \tilde{c}_2 \|\Lambda p\|_0 \tag{3.39}$$

and therefore the statement. □

Remark 3.2.6. The linear map

$$\Lambda_0 \colon \begin{cases} \mathcal{W}^1_{h,0,\mathrm{bc}} \to \mathcal{W}'^0_{h,0,\mathrm{bc}} \\ p \mapsto \Lambda p - \operatorname{avg}_\Omega(\Lambda p) \end{cases} \tag{3.40}$$

is an isomorphism and its inverse is given by $\Lambda_0^{-1} = \Lambda^{-1} - \operatorname{avg}_\Omega \Lambda^{-1}$.

Proof. The map is linear since it is a linear combination composed of linear operators. Λ_0 is well-defined as $\Lambda_0 p \in \mathcal{W}_h^1$ and

$$\operatorname{avg}_\Omega \Lambda_0 p = \operatorname{avg}_\Omega \Lambda p - \operatorname{avg}_\Omega \operatorname{avg}_\Omega \Lambda p = 0. \tag{3.41}$$

Next we remark that for every constant function k, $\Lambda k = k = \Lambda^{-1} k$. Subsequently, we conclude by

$$\begin{aligned} \Lambda_0^{-1} \Lambda_0 p &= \Lambda^{-1} \Lambda p - \Lambda^{-1} \operatorname{avg}_\Omega \Lambda p - \operatorname{avg}_\Omega \Lambda^{-1} \Lambda p + \operatorname{avg}_\Omega \Lambda^{-1} \operatorname{avg}_\Omega \Lambda p && (3.42) \\ &= p - \operatorname{avg}_\Omega \Lambda p - \operatorname{avg}_\Omega p + \operatorname{avg}_\Omega \Lambda p && (3.43) \\ &= p && (3.44) \end{aligned}$$

for every $p \in \mathcal{W}^1_{h,0,\mathrm{bc}}$ and

$$\Lambda_0 \Lambda_0^{-1} q = \Lambda \Lambda^{-1} q - \Lambda \operatorname{avg}_\Omega \Lambda^{-1} q - \operatorname{avg}_\Omega \Lambda \Lambda^{-1} q + \operatorname{avg}_\Omega \Lambda \operatorname{avg}_\Omega \Lambda^{-1} q \tag{3.45}$$

$$= q - \operatorname{avg}_\Omega \Lambda^{-1} p - \operatorname{avg}_\Omega q - \operatorname{avg}_\Omega \Lambda^{-1} q \tag{3.46}$$

$$= q \tag{3.47}$$

for every $q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}$. □

Remark 3.2.7. We can express the B_1 by

$$\langle B_1 \theta_{\xi_j}, \theta^d_{K,i} \rangle = \sum_{K \in \mathcal{K}_{\xi_j, \xi_j}} (G_K \tilde{B}_1 (G^d_K)^T)_{j,i} \tag{3.48}$$

for $i \in \{1 \dots |\mathcal{T}_h||\mathcal{P}^d|\}$ and $j \in \{1 \dots |\mathcal{N}_{\mathcal{T}_h}|\}$. For B_2 this holds true by an analogous sum.

Lemma 3.2.8. *Let $\mathcal{T}_h$ be a grid as defined by in Definition 2.1.1. Consider the map*

$$L \colon \begin{cases} \mathcal{D}_h \to \mathcal{D}_h \\ \sum\limits_{K \in \mathcal{T}_h} v_K \mathbb{1}_K \mapsto \sum\limits_{K \in \mathcal{T}_h} \mathbb{1}_K L_K(v_K) \end{cases} \tag{3.49}$$

where $L_K = \psi^{d\,-1}_{T_K} \circ \hat{L} \circ \psi^d_{T_K}$ and $L_{\hat{K}}(v) = v \circ (x \mapsto \frac{2}{3}x)$. Then we have an analogue for integration by parts by

$$\langle B_1 \Lambda^{-1} q, v \rangle + \langle B_2 L v, q \rangle = 0 \tag{3.50}$$

for every $v \in \mathcal{D}_h$ and $q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}$. Furthermore, the operator $L \in \mathcal{B}(\mathcal{D}_h)$ is invertible and the inverse of L is determined element wise by $L_K^{-1}(v) = v \circ (x \mapsto \frac{3}{2}x)$. Finally, the continuity constants of L,L' and L^{-1} are independent of h.

Proof. We first observe the linearity of L due to its construction via linear maps. Subsequently, we observe the transformed monomial basis being invariant under L i.e., for every $v = \mathbb{1}_K \psi^d_{T_K}(\hat{v}) \in \mathcal{D}_h$, where $\hat{v}$ is a monomial vector, we have $Lv \in \operatorname{span} v$. As this implies, that the image of a basis under L is a basis again, we conclude $L \colon \mathcal{D}_h \to \mathcal{D}_h$ is bijective.

Besides the fact that every finite dimensional linear mapping is continuous, we have to establish the independence of h for the continuity bounds. Let $v = \sum\limits_{K \in \mathcal{T}_h} v_K \mathbb{1}_K \in \mathcal{D}_h$. Furthermore, consider $S_K \colon \hat{K} \to \tilde{K} \colon x \mapsto (h_K / h_{\hat{K}})x$. Then $\operatorname{diam} \tilde{K} = \operatorname{diam} K$ which together with the constant shape and position of $\tilde{K}$ implies that $\tilde{T}_K \colon \tilde{K} \to K \colon x \mapsto (T_K \circ S_K^{-1})(x)$ is

independent of h_K. Subsequently, we realize

$$(\psi_{S_K}^{d\ -1} L_{\hat{K}} \psi_{S_K}^d)(v) = \frac{h}{h^n}(L_{\hat{K}}(\frac{h}{h^n}(v \circ S_K))) \circ S_K^{-1} \tag{3.51}$$

$$= (L_{\hat{K}}(v \circ S_K)) \circ S_K^{-1} \tag{3.52}$$

$$= v(\frac{h_n}{h_n}\frac{3}{2}x) \tag{3.53}$$

$$= L_{\hat{K}}(v). \tag{3.54}$$

The element-wise operator now also is independent of h_K due to

$$L_K = \psi_{T_K}^{d\ -1} \hat{L} \psi_{T_K}^d = \psi_{\tilde{T}_K}^{d\ -1} \psi_{S_K}^{d\ -1} \hat{L} \psi_{S_K}^d \psi_{\tilde{T}_K}^d = \psi_{\tilde{T}_K}^{d\ -1} \hat{L} \psi_{\tilde{T}_K}^d. \tag{3.55}$$

Finally, we therefore have constants $l_K > 0$ independent of h satisfying

$$\|Lv\|_0^2 = \sum_{K \in \mathcal{T}_h} \|L_K v_K\|_{0,K}^2 \leq \sum_{K \in \mathcal{T}_h} l_K^2 \|v_K\|_{0,K}^2 \leq \max_{K \in \mathcal{T}_h}(l_K^2) \|v_K\|_0^2 \tag{3.56}$$

The statement $L^{-1}(v) = \sum_{K \in \mathcal{T}_h} \mathbb{1}_K L_K^{-1}(v_K)$ follows directly by

$$L^{-1}(L(v)) = \sum_{K \in \mathcal{T}_h} \mathbb{1}_K L_K^{-1}(L_K(v_K)) = \sum_{K \in \mathcal{T}_h} \mathbb{1}_K v_K) = v \tag{3.57}$$

and $L(L^{-1}(v)) = v$. By completely analogous reasoning as in the case of L we argue the continuity constant of L^{-1} to be independent of h. The dual operator of L with regard to $(\,\cdot\,,\,\cdot\,)_0$ is given in terms of the element-wise inverse too i.e.,

$$\langle Lu, v \rangle = \sum_{K \in \mathcal{T}_h} \int_K (Lu)^T v \, dx = \frac{3}{2} \sum_{K \in \mathcal{T}_h} \int_{\frac{3}{2}K} u^T L_K^{-1} v \, dx =: \langle u, L'v \rangle \tag{3.58}$$

where the latter expression is independent of h again. Therefore, the continuity constant of L' is too.

Finally, we prove the statement with regard to integration by parts. As in Remark 3.2.7 we use the element wise representation of B_1, B_2 and L, the latter given by

$$\tilde{L}_{i,j} = \sum_{j \in I} \left(V_{i,j} \, (2/3)^{\deg(\theta_j^m)} \, V_{j,k}^{-1} \right) \tag{3.59}$$

where V is the transformation from θ^d to the monomial basis θ^m and $i, j \in I$. Now we

have for every global index $i, j \in \{1 \dots |I||\mathcal{T}_h|\}$

$$L_{i,j} = \sum_{K \in \mathcal{T}_h} (G_K^d \tilde{L} G_K^{d^T})_{i,j}. \tag{3.60}$$

Due to the properties of G_K^d the composition of B_2 and L then reads

$$\langle B_2 L u_i, \mathbb{1}_{K'_{\xi_j}} \rangle = \sum_{K \in \mathcal{K}_{\xi_j, \xi_j}} (G_K \tilde{B}_2 \tilde{L} G_K^{d^T})_{j,i} \tag{3.61}$$

for every node $\xi_j \in \mathcal{N}_{\mathcal{T}_h}$ and global basis $u_i \in \mathcal{D}_h$. Via direct verification, we observe $\tilde{B}_2 \tilde{L}_K + \tilde{B}_1 = 0$. Subsequently, we establish

$$\begin{aligned}
\langle B_2 L u_i, \mathbb{1}_{K'_{\xi_j}} \rangle &= \sum_{K \in \mathcal{K}_{\xi_j, \xi_j}} (G_K \tilde{B}_2 \tilde{L} G_K^{d^T})_{j,i} && (3.62)\\
&= - \sum_{K \in \mathcal{K}_{\xi_j, \xi_j}} (G_K \tilde{B}_1 G_K^{d^T})_{j,i} && (3.63)\\
&= -\langle B_1 \theta_{\xi_j}, u_i \rangle && (3.64)
\end{aligned}$$

and therefore the missing statement. □

Corollary 3.2.9. *There are constants c_1, c_2 satisfying*

$$c_1 \|B_2' q\|_0 \leq \sup_{v \in \mathcal{D}_h} \frac{|\langle B_1 \Lambda^{-1} q, v \rangle|}{\|v\|_0} \leq c_2 \|B_2' q\|_0 \tag{3.65}$$

for every $q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}$.

Proof. We first establish

$$\begin{aligned}
\sup_{v \in \mathcal{D}_h} \frac{|\langle B_1 \Lambda^{-1} q, v \rangle|}{\|v\|_0} &= \sup_{v \in \mathcal{D}_h} \frac{|\langle q, B_2 L v \rangle|}{\|v\|_0} && (3.66)\\
&= \sup_{v \in \mathcal{D}_h} \frac{|\langle B_2' q, v \rangle|}{\|L^{-1} v\|_0} && (3.67)\\
&\geq \sup_{v \in \mathcal{D}_h} \frac{|\langle B_2' q, v \rangle|}{\|L^{-1}\| \|v\|_0} && (3.68)\\
&= \frac{\|B_2' q\|_0}{\|L^{-1}\|} && (3.69)
\end{aligned}$$

and recall the continuity constant of L^{-1} being independent of h. The other direction

follows by

$$\sup_{v\in\mathcal{D}_h}\frac{|\langle B_2'q,v\rangle|}{\|L^{-1}v\|_0}=\sup_{v\in\mathcal{D}_h}\frac{|\langle L'B_2'q,v\rangle|}{\|v\|_0} \tag{3.70}$$

$$=\|L'B_2'q\|_0 \tag{3.71}$$

$$\leq\|L'\|\|B_2'q\|_0 \tag{3.72}$$

and the same observation. □

Lemma 3.2.10. *The operator L is self-adjoint on $\mathcal{D}_h$, positive definite and the Eigenvalues are independent of h.*

Proof. - For each shape $K\subset\mathcal{T}_h$, $DT_K^TDT_K\in C^\infty(\hat{K},\mathbb{R}^{dxd})$ has only symmetric positive definite images on whole $\hat{K}$. Let $\phi_{K,i}=\mathbb{1}_K((\psi_{T_K}^d)^{-1}\circ\theta_i^m)$ be a transformed monomial basis function of $\mathcal{P}^d$ and degree n. Then $\lambda_{K,i}=(\frac{2}{3})^n$ is an eigenvalue of L and $\phi_{K,i}$ the corresponding eigenfunction.

In case of $d=3$ the algebraic multiplicity of the eigenvalues is 3,6 and 3 for the eigenfunctions originating from the constant, linear and bilinear basis functions respectively. Similarly, for $d=2$ we conclude the algebraic multiplicity of both eigenvalues to be 2. As all Eigenfunctions $\phi_{K,i}$ are linearly independent we conclude that the geometric multiplicity is the same. Furthermore, the $\phi_{K,i}$ are a (non-orthogonal) Eigenbasis. Therefore, L has only real positive Eigenvalues and is a self-adjoint operator. □

Lemma 3.2.11. *The element wise operators L_K and $\psi_{T_K}^d$ commute in the following sense i.e.,*

$$\psi_{T_K}^dL_Kv=L_{\hat{K}}\psi_{T_K}^dv \tag{3.73}$$

for every $v\in\mathcal{D}_h$.

Proof. L $v\in\mathcal{D}_h$ then we have

$$(\psi_{T_K}^dL_K)v=(\psi_{T_K}^d\psi_{T_K}^{d\,-1}L_{\hat{K}}\psi_{T_K}^d)v=(L_{\hat{K}}\psi_{T_K}^d)v \tag{3.74}$$

by definition of L_K. □

This property now allows us to establish the following interpolation property of L.

Lemma 3.2.12. *Let $\mathcal{T}_h$ be a shape-regular and quasi-uniform family of grids as given in Definition 2.1.1. Then there is a constant $c>0$ independent of h satisfying*

$$\|Lv-v\|_0\leq ch_n\|v\|_{H^1(\mathcal{T}_h)} \tag{3.75}$$

for every $v \in H^1(\mathcal{T}_h)^n$.

Proof. We first observe $L_{\hat{K}} v = v$ for every constant v. Next we apply Lemma B.0.3 on the reference element $\hat{K}$ i.e.,

$$\begin{aligned}
\|L_{\hat{K}} v - v\|_{0,\hat{K}} &= \|(\mathbb{I} - L_{\hat{K}})(v)\|_{0,\hat{K}} && (3.76)\\
&= \inf_{u \in \mathbb{P}_0{}^n} \|(\mathbb{I} - L_{\hat{K}})(v-u)\|_{0,\hat{K}} && (3.77)\\
&\leq \|\mathbb{I} - L_{\hat{K}}\|_{\mathcal{B}(H^1(\hat{K})^n, L^2(\hat{K})^n)} \inf_{u \in \mathbb{P}_0{}^n} \|v-u\|_{1,\hat{K}} && (3.78)\\
&\leq \hat{c}|v|_{1,\hat{K}} && (3.79)
\end{aligned}$$

Using this result and Lemma 3.2.11 we obtain a local estimate

$$\|Lv - v\|_{0,K} \leq ch|v|_{1,K} \tag{3.80}$$

if we follow the proofs of Lemma 2.3.6 and Proposition 2.3.7 line by line, but replace $\mathcal{I}_K^d$ by L_K. In resemblance of Proposition 2.4.17 we subsequently conclude the result

$$\|Lv - v\|_0^2 = \sum_{K \in \mathcal{T}_h} \|Lv - v\|_{0,K}^2 \leq \sum_{K \in \mathcal{T}_h} c_K^2 h_K^2 |v|_{1,K}^2 = c^2 h^2 |v|_1^2 \tag{3.81}$$

for every $v \in \mathcal{D}_h$ due to quasi-uniformity and shape-regularity. □

3.2.3 Coercivity on the null space

Remark 3.2.13. Let $\delta = \sum_{K \in \mathcal{T}_h} \psi_{T_K}^{d\,-1} \delta_K \psi_{T_K}^d \mathbb{1}_K$ then M_δ, i.e. multiplication by δ, and L commute. This also holds true if δ is matrix valued.

Proof. We first observe

$$M_{\psi_{T_K}^{d\,-1} \delta_K \psi_{T_K}^d} = \psi_{T_K}^{d\,-1} M_{\delta_K} \psi_{T_K}^d$$

and $L_{\hat{K}} M_{\delta_K} = M_{\delta_K} L_{\hat{K}}$ as δ_K is constant. Therefore,

$$\begin{aligned}
M_{\psi_{T_K}^{d\,-1} \delta_K \psi_{T_K}^d} L_K &= \psi_{T_K}^{d\,-1} M_{\delta_K} \psi_{T_K}^d \psi_{T_K}^{d\,-1} L_{\hat{K}} \psi_{T_K}^d \\
&= \psi_{T_K}^{d\,-1} M_{\delta_K} L_{\hat{K}} \psi_{T_K}^d \\
&= \psi_{T_K}^{d\,-1} \psi_{T_K}^d \\
&= \psi_{T_K}^{d\,-1} L_{\hat{K}} \psi_{T_K}^d \psi_{T_K}^{d\,-1} M_{\delta_K} \psi_{T_K}^d \\
&= L_K M_{\psi_{T_K}^{d\,-1} \delta_K \psi_{T_K}^d}
\end{aligned}$$

for every $K \in \mathcal{T}_h$. Summing over all elements gives the desired result. □

Lemma 3.2.14. *Let $(\mathcal{T}_h)_{n\in\mathbb{N}}$ quasi-uniform and shape-regular family of grids as defined in Definition 2.1.1. Let $\delta, \omega \in L^\infty(\Omega)^{n\times n}$ be piecewise constant and assume δ is symmetric and positive definite almost everywhere. Then $a : (\mathcal{D}_h, \|\cdot\|_0^n) \times (\mathcal{D}_h, \|\cdot\|_0^n) \to \mathbb{R}$ is a continuous bilinear form. Additionally, there exists $\tau_0 > 0$ such that the bilinear form is coercive on whole $\mathcal{D}_h$ i.e., there is a constant $c_a > 0$ independent of h_n satisfying*

$$\sup_{u\in\kappa_2\setminus\{0\}} \frac{a(u,v)}{\|u\|_{\mathrm{DIV}_h}} \geq c_a\|v\|_0 \quad \forall v \in \kappa_1 \setminus \{0\}, \tag{3.82a}$$

$$\sup_{v\in\kappa_1\setminus\{0\}} \frac{a(u,v)}{\|v\|_0} \geq c_a\|u\|_{\mathrm{DIV}_h} \quad \forall u \in \kappa_2 \setminus \{0\} \tag{3.82b}$$

and for all $\tau \in [0, \tau_0]$.

Proof. a is bilinear as the real inner product is bilinear. Continuity follows directly from the Cauchy-Schwarz inequality and

$$|a(u,v)| \leq \|\delta + \tau\omega\|_\infty \|u\|_0\|v\|_0 = \|\delta + \tau\omega\|_\infty \|u\|_0\|v\|_{\mathrm{DIV}_h} \tag{3.83}$$

for every $u, v \in \mathcal{D}_h$. As L is symmetric, positive definite and commutes with δ, the positive definite and symmetric square root $\sqrt{L}$ exists and commutes with δ too. Therefore, the first condition follows by

$$\begin{aligned}
\sup_{u\in\kappa_2\setminus\{0\}} \frac{a(u,v)}{\|u\|_{\mathrm{DIV}_h}} &\geq \frac{a(Lv,v)}{\|Lv\|_0} && (3.84)\\
&= \frac{(\delta Lv, v)_0 + \tau(\omega Lv, v)_0}{\|Lv\|_0} && (3.85)\\
&= \frac{(\delta\sqrt{L}v, \sqrt{L}v)_0 + \tau(\omega Lv, v)_0}{\|Lv\|_0} && (3.86)\\
&\geq \frac{(c_\delta \min\lambda_i - \tau\|\|\omega L\|\|_\infty)\,\|v\|_0^2}{\|Lv\|_0} && (3.87)\\
&\geq \frac{(c_\delta \min\lambda_i - \tau\|\|\omega L\|\|_\infty)\,\|v\|_0^2}{\|L\|\|v\|_0} && (3.88)\\
&\geq c_a\|v\|_0 && (3.89)
\end{aligned}$$

for every $v \in \kappa_1 \setminus \{0\}$. The other condition follows by the same reasoning using L^{-1} and the fact that $\|u\|_0 = \|u\|_{\mathrm{DIV}}$ for every $u \in \kappa_2$. □

3.2.4 Stability of the gradient

Lemma 3.2.15. *Let $\mathcal{T}_h$ denote a shape-regular and quasi-uniform family of grids. Then $b_1 \colon (\mathcal{W}^1_{h,0,\mathrm{bc}}, \|\cdot\|_1) \times (\mathcal{D}_h, \|\cdot\|_0) \to \mathbb{R}$ is a continuous bilinear form and there is $c_{b_1} > 0$ independent of h such that*

$$\sup_{v \in \kappa_1^\perp} \frac{b_1(p,v)}{\|v\|_0} \geq c_{b_1} \|p\|_1 \tag{3.90}$$

for every $p \in \mathcal{W}^1_{h,0,\mathrm{bc}}$.

Proof. b_1 is a bilinear form by construction via linear operators and a real inner product. We conclude continuity by Cauchy-Schwarz and

$$|b_1(p,v)| \leq \|\operatorname{grad} p\|_0 \|v\|_0 \leq \|p\|_1 \|v\|_0 \tag{3.91}$$

Applying the isomorphism $L^{rd} \colon \mathcal{R}_h \to \mathcal{D}_h$ introduced in Lemma 2.4.18 we can choose $v = L^{r,d} \operatorname{grad} p$ and obtain

$$\begin{aligned}
\sup_{v \in \mathcal{D}_h} \frac{b_1(p,v)}{\|v\|_0} &\geq \frac{b_1(p, L^{r,d}(\operatorname{grad} p))}{\|L^{r,d}(\operatorname{grad} p)\|_0} && (3.92)\\
&= \frac{\sum_K \int_K \left(\psi^{d\,-1}_{T_K} \psi^r_{T_K}(\operatorname{grad} p)\right) \cdot \operatorname{grad} p \, dx}{\|L^{r,d}(\operatorname{grad} p)\|_0} && (3.93)\\
&= \frac{\sum_K \int_K \frac{1}{\det(DT_K \circ T_K^{-1})} \operatorname{grad} p^T \left(DT_K DT_K^T\right) \circ T_K^{-1} \operatorname{grad} p \, dx}{\|L^{rd}(\operatorname{grad} p)\|_0}. && (3.94)
\end{aligned}$$

We observe $\left(DT_K DT_K^T\right)$ being symmetric and positive definite almost everywhere, as all scalar fields are positive almost everywhere and DT_K is regular due the assumption $\det DT_K > 0$. Furthermore, by splitting T_K into a pure scaling and $T_{\tilde{K}}$ with $\operatorname{diam} \tilde{K} = 1$ we have a constant $c_{b_1} > 0$ independent of h (c.f. Lemma 2.4.18) such that

$$\begin{aligned}
&\sum_K \int_K \frac{\operatorname{grad} p^T \left(DT_K DT_K^T\right) \circ T_K^{-1} \operatorname{grad} p}{\det(DT_K \circ T_K^{-1})} \, dx \\
&\geq \min_{K \in \mathcal{T}_h} c_{b_{1,K}} h^{2-n} \|\operatorname{grad} p\|_0^2
\end{aligned} \tag{3.95}$$

Subsequently, we apply the Poincaré inequality to obtain a constant $c_{b_1} > 0$ independent

of h such that

$$\min_{K\in\mathcal{T}_h} c_{b_{1,K}} \operatorname{ess\,inf} \frac{1}{\det(DT_K \circ T_K^{-1})} \|\operatorname{grad} p\|_0^2 \geq c_{b_1} h^{2-n} \|\operatorname{grad} p\|_0 \|p\|_1. \tag{3.96}$$

In combination with Eq. (3.94) and Eq. (2.193) we obtain

$$\sum_K \int_K \frac{\operatorname{grad} p^T \left(DT_K DT_K^T\right) \circ T_K^{-1} \operatorname{grad} p \, dx}{\det(DT_K \circ T_K^{-1}) \|L^{r,d} \operatorname{grad} p\|_0} \geq \frac{\tilde{c}_{b_1} h^{2-n} \|\operatorname{grad} p\|_0 \|p\|_1}{h^{2-n} \|\operatorname{grad} p\|_0} \tag{3.97}$$

$$\geq \tilde{c}_{b_1} \|p\|_1. \tag{3.98}$$

□

Remark 3.2.16. As subspace of $L^2(\Omega)$, we can equip $\mathcal{W}_h^1$ with $\|\cdot\|_0$ instead of $\|\cdot\|_1$. The inner product space $(\mathcal{W}_h^1, (\cdot,\cdot)_0)$ is a Hilbert space, as it is finite dimensional and therefore closed. This allows us to still apply the framework of abstract saddle point problems, despite the *wrong* inner product. This choice, however comes at the expense of a stability constant depending on some negative power of h.

Lemma 3.2.17. *Let $\mathcal{T}_h$ a quasi-uniform shape-regular family of grids as given in Definition 2.1.1. Then $b_1 \colon (\mathcal{W}_{h,0,\mathrm{bc}}^1, \|\cdot\|_0) \times (\mathcal{D}_h, \|\cdot\|_0) \to \mathbb{R}$ is a continuous bilinear form and there is $c_{b_1} > 0$ such that*

$$\sup_{v\in\kappa_1^\perp} \frac{b_1(p,v)}{\|v\|_0} \geq c_{b_1} \|p\|_0 \tag{3.99}$$

for every $p \in \mathcal{W}_{h,0,\mathrm{bc}}^1$.

Proof. See proof of Lemma 3.2.15, but bound b_1 by an inverse inequality Lemma B.0.4 and

$$|b_1(p,v)| \leq \|\operatorname{grad} p\|_0 \|v\|_0 \leq \|p\|_1 \|v\|_0 \leq \frac{c}{h} \|p\|_0 \|v\|_0. \tag{3.100}$$

Furthermore, use the Poincaré inequality to apply

$$\min_{K\in\mathcal{T}_h} c_{b_{1,K}} \operatorname{ess\,inf} \frac{1}{\det(DT_K \circ T_K^{-1})} \|\operatorname{grad} p\|_0^2 \geq c_{b_1} h^{2-n} \|\operatorname{grad} p\|_0 \|p\|_1, \tag{3.101}$$

instead of Eq. (3.96) and finally conclude

$$\sum_K \int_K \frac{\beta_1 \operatorname{grad} p^T \left(DT_K DT_K^T\right) \circ T_K^{-1} \operatorname{grad} p \, dx}{\det(DT_K \circ T_K^{-1}) \|L^{r,d} \operatorname{grad} p\|_0} \geq \frac{\tilde{c}_{b_1} h^{2-n} \|\operatorname{grad} p\|_0 \|p\|_0}{h^{2-n} \|\operatorname{grad} p\|_0} \tag{3.102}$$

$$\geq \tilde{c}_{b_1} \|p\|_0. \tag{3.103}$$

□

3.2.5 Stability of the divergence

Lemma 3.2.18 (c.f. Bochev and Ridzal [25, Lemma 3.2])**.** *The operator* $\mathrm{DIV}_h : \mathcal{D}_h \to \mathcal{W}_h'^0$ *is surjective and with continuous lifting from* $\mathcal{W}_h'^0$ *into* $\mathcal{D}_h$ *i.e., for every* $q \in \mathcal{W}_h'^0$ *there is* $u_q \in \mathcal{D}_h$ *satisfying* $q = \mathrm{DIV}_h(u_q)$ *and* $c > 0$ *independent of* u_q,q *such that*

$$\|u_q\|_{\mathrm{DIV}_h} \leq c\|q\|_0. \tag{3.104}$$

Proof. We follow the proof of [25, Lemma 3.2] line by line. We simply exchange the respective function spaces and interpolation operators. We first observe the surjectivity of $\mathrm{div} : \mathcal{U} \to L^2(\Omega)$ by solving a Poisson problem (c.f. [27, pp. 135–137]). Now $q^h \in \mathcal{W}_h'^0 \subset L^2(\Omega)$ and therefore there is $u_q \in \mathcal{U}$ such that

$$\mathrm{div}(u_q) = q^h. \tag{3.105}$$

The map $q^h \to u_q$ is continuous (c.f. [27, §IV, Remark 1.1]), i.e. u_q additionally satisfies

$$\|u_q\|_{\mathrm{div}} \leq \hat{c}\|q^h\|_0. \tag{3.106}$$

Furthermore, there is $u_q^h = \mathcal{I}_h^d u_q \in \mathcal{D}_h$ and using the commuting property in Lemma 2.4.13, it satisfies

$$\mathrm{DIV}_h(u_q^h) = \mathrm{DIV}_h(\mathcal{I}_h^d u_q) = \mathcal{I}^0_{\mathcal{T}_h'}\,\mathrm{div}(u_q) = \mathcal{I}^0_{\mathcal{T}_h'}\,q^h = q^h. \tag{3.107}$$

To establish continuity of the lifting, we consider the analytical property Eq. (3.105), continuity of the interpolation established in Lemma 2.2.13 and Eq. (3.106). We then see

$$\begin{aligned}
\|u_q^h\|^2_{\mathrm{DIV}_h} &= \|u_q^h\|_0^2 + \|\mathrm{DIV}_h\, u_q^h\|_0^2 && (3.108)\\
&= \|\mathcal{I}_h^d\, u_q\|_0^2 + \|q^h\|_0^2 && (3.109)\\
&\leq \tilde{c}\|u_q\|^2_{\mathrm{div}} + \|q^h\|_0^2 && (3.110)\\
&\leq c\|q^h\|_0^2 + \|q^h\|_0^2 && (3.111)\\
&= (c+1)\|q^h\|_0^2. && (3.112)
\end{aligned}$$

□

Lemma 3.2.19. *Let* $\mathcal{T}_h$ *a quasi-uniform and shape-regular family of grids as given in Definition 2.1.1. The bilinear form* $b_{2,h} : (\mathcal{D}_h; \|\cdot\|_{\mathrm{DIV}_h}) \times (\mathcal{W}_h'^0; \|\cdot\|_0) \to \mathbb{R}$ *is continuous and there is a constant*

$c_{b_2} > 0$ *independent of* h *satisfying*

$$\sup_{u \in \kappa_2^\perp} \frac{b_{2,h}(u,q)}{\|u\|_{\mathrm{DIV}_h}} \geq c_{b_2} \|q\|_0 \quad \forall q \in \mathcal{W}'^0_{h,0,\mathrm{bc}} \tag{3.113}$$

$$\sup_{q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}} b_{2,h}(u,q) > 0 \qquad \forall u \in \kappa_2^\perp \setminus \{0\}. \tag{3.114}$$

Proof. Continuity follows by Cauchy-Schwarz and

$$|b_{2,h}(u,q)| \leq |(\mathrm{DIV}_h\, u, q)| \leq \|\mathrm{DIV}_h\, u\|_0 \|q\|_0 \leq \|u\|_{\mathrm{DIV}_h} \|q\|_0. \tag{3.115}$$

The statement is trivially satisfied for $q = 0$. Let $q \neq 0$, then using Lemma 3.2.18 we have $u_q \in \kappa_2^\perp \subset \mathcal{D}_h$ with $\mathrm{DIV}(u_q) = q$. This enables us to establish

$$\sup_{u \in \kappa_2^\perp} \frac{b_{2,h}(u,q)}{\|u\|_{\mathrm{DIV}_h}} \geq \frac{b_{2,h}(u_q,q)}{\|u_q\|_{\mathrm{DIV}_h}} \tag{3.116}$$

$$\geq \frac{(\mathrm{DIV}_h\, u_q, q)_0}{\|u_q\|_{\mathrm{DIV}_h}} \tag{3.117}$$

$$\geq \frac{\|q\|_0^2}{\|u_q\|_{\mathrm{DIV}_h}} \tag{3.118}$$

$$\geq \frac{\|q\|_0^2}{c\|q\|_0} \tag{3.119}$$

$$\geq \frac{1}{c}\|q\|_0. \tag{3.120}$$

The second bound follows by

$$\sup_{q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}} b_{2,h}(u,q) \geq b_{2,h}(u, \mathrm{DIV}_h\, u) = \|\mathrm{DIV}_h\, u\|_0^2 > 0 \tag{3.121}$$

for every $u \in \kappa_2^\perp \setminus \{0\}$. □

Corollary 3.2.20. *There is a constant* $c = 1/c_{b_2} > 0$ *independent of* h *such that*

$$\|u\|_0 \leq c|u|_{\mathrm{DIV}_h} \tag{3.122}$$

for every $u \in \kappa_2^\perp$.

Proof. Let ι denote the isometric Riesz isomorphism of $L^2(\Omega)$ as discussed in Remark A.7.2. The dual operator B_2' i.e. the negative gradient induced by DIV_h is a linear functional on

$\mathcal{D}_h \subset L^2(\Omega)$. Therefore, $\iota B_2' \in \mathcal{D}_h$ and using the estimate from Lemma 3.2.19 we obtain

$$\|\iota B_2' q\|_0 = \sup_{v \in \kappa_2^\perp} \frac{\langle B_2' q, v\rangle}{\|v\|_0} \geq \sup_{v \in \kappa_2^\perp} \frac{\langle B_2' q, v\rangle}{\|v\|_{\mathrm{DIV}_h}} \geq c_{b_2} \|q\|_0. \tag{3.123}$$

Combining Lemma 3.2.19 and Theorem A.7.5 we realize there is exactly one $q_v \in \mathcal{W}_{h,\mathrm{bc}}'^0$ for every $v \in \kappa_2^\perp$ with $v = \iota B_2' q_v$. Using the aforementioned estimate we therefore have

$$\begin{aligned}
\|v\|_0 = \|\iota B_2' q_v\|_0 &\leq \frac{1}{c_{b_2}} \frac{\|\iota B_2' q_v\|_0^2}{\|q_v\|_0} && (3.124)\\
&= \frac{1}{c_{b_2}} \frac{\langle B_2' q_v, \iota B_2' q_v\rangle}{\|q_v\|_0} && (3.125)\\
&= \frac{1}{c_{b_2}} \frac{\langle B_2 \iota B_2' q_v, q_v\rangle}{\|q_v\|_0} && (3.126)\\
&= \frac{1}{c_{b_2}} \frac{\langle B_2 v, q_v\rangle}{\|q_v\|_0} && (3.127)\\
&\leq \sup_{q \in \mathcal{W}_{h,0,\mathrm{bc}}'^0} \frac{1}{c_{b_2}} \frac{\langle B_2 v, q\rangle}{\|q\|_0} && (3.128)\\
&= \frac{1}{c_{b_2}} \|\iota B_2 v\|_0 && (3.129)\\
&= \frac{1}{c_{b_2}} \|\mathrm{DIV}_h\, v\|_0. && (3.130)
\end{aligned}$$

□

Corollary 3.2.21. *Under the assumptions of Lemma 3.2.19 there is a constant $c > 0$ independent of h satisfying*

$$\sup_{q \in \mathcal{W}_{h,0,\mathrm{bc}}'^0} \frac{b_{2,h}(v, q)}{\|q\|_0} \geq c \|v\|_{\mathrm{DIV}_h} \tag{3.131}$$

for every $v \in \kappa_2^\perp$.

Proof.

$$\sup_{q \in \mathcal{W}_{h,0,\mathrm{bc}}'^0} \frac{b_{2,h}(v, q)}{\|q\|_0} \geq \frac{b_{2,h}(v, \mathrm{DIV}_h\, u)}{\|\mathrm{DIV}_h\, v\|_0} = \|\mathrm{DIV}_h\, v\|_0 \geq c \|v\|_{\mathrm{DIV}_h} \tag{3.132}$$

□

3.2.6 The pseudo-incompressible regime

Proposition 3.2.22 (c.f. Süli [113])**.** *The Poisson problem* $\mathrm{DIV}\,\mathrm{grad}\,p = -g$ *with Neumann and periodic boundary conditions in the sense of* b_1, b_2 *has unique solution* $p \in \mathcal{W}^1_{h,0,\mathrm{bc}}$ *for every right-hand side* $g \in {\mathcal{W}'^0_{h,\mathrm{bc}}}'$*. More specifically there is exactly one* $(u;p) \in \mathcal{D}_h \times \mathcal{W}^1_{h,0,\mathrm{bc}}$ *for every* $g \in {\mathcal{W}'^0_{h,0,\mathrm{bc}}}'$ *such that*

$$\langle u, v\rangle + \langle B_1 p, v\rangle = 0, \tag{3.133}$$

$$\langle B_2 u, q\rangle \qquad\qquad = \langle g, q\rangle \tag{3.134}$$

for every $q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}$ *and* $v \in \mathcal{D}_h$*.*

Proof. Due to Lemmas 3.2.14, 3.2.15 and 3.2.19 and the fact that $\mathbb{I}$ commutes with L we can apply the stability result in [21, Thm. 2.1]. □

Proposition 3.2.23 (c.f. Vater and Klein [118])**.** *Let* $\alpha_P = 0$*, then the mixed saddle point problem Eq.* (3.10) *with Neumann and periodic boundary conditions in the sense of* a, b_1, b_2 *has unique solution* $p \in \mathcal{W}^1_{h,0,\mathrm{bc}}$ *for every right-hand side. More specifically there is exactly one* $(u;p) \in \mathcal{D}_h \times \mathcal{W}^1_{h,0,\mathrm{bc}}$ *for every* $(f;g) \in (\mathcal{D}_h; \|\cdot\|_0)' \times (\mathcal{W}'^0_{h,0,\mathrm{bc}}; \|\cdot\|_0)'$ *such that*

$$\langle Au, v\rangle + \tau\langle B_1 p, v\rangle = \langle f, v\rangle \tag{3.135a}$$

$$\tau\langle B_2 u, q\rangle \qquad\qquad = \langle g, q\rangle \tag{3.135b}$$

for every $q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}$ *and* $v \in \mathcal{D}_h$*. Furthermore, this solution satisfies*

$$\|u\|_{\mathrm{DIV}_h} \leq \frac{1}{c_a}\|f\|_{\mathcal{D}_h'} + \frac{1}{\tau c_{b_2}}\left(1 + \frac{\|a\|}{c_a}\right)\|g\|_{{\mathcal{W}'^0_{h,0,\mathrm{bc}}}'} \tag{3.136}$$

$$\|p\|_1 \leq \frac{1}{\tau c_{b_1}}\left(1 + \frac{\|a\|}{c_a}\right)\|f\|_{\mathcal{D}_h'} + \frac{\|a\|}{\tau^2 c_{b_2} c_{b_1}}\left(1 + \frac{\|a\|}{c_a}\right)\|g\|_{{\mathcal{W}'^0_{h,0,\mathrm{bc}}}'}. \tag{3.137}$$

Proof. Due to Lemmas 3.2.14, 3.2.15 and 3.2.19, we can apply the stability result in [21, Thm. 2.1]. □

Corollary 3.2.24. *Let* $g = 0$*, then the solution obtain in Proposition 3.2.23 and denoted by* $(u;p)$ *satisfies* $u \in \kappa_2$ *and therefore* $(\mathrm{grad}\,p, L^{-1}u)_0 = 0$*.*

3.2.7 Error estimates

After establishing existence and uniqueness of the discrete solution, we aim to prove consistency.

Remark 3.2.25. As consequence of Theorem A.3.12 and the Cauchy-Schwarz inequality, $b_{2,h}\colon \mathcal{U} \cap H^1(\Omega)^n \times L^2(\Omega) \to \mathbb{R}\colon (u;q) \mapsto b_2(u,q)$ is well-defined and continuous.

Although the following results can be generalized to $u \in \mathcal{U}$ we avoid doing so, as in this case we do not have an upper bound in terms of a positive power of h on the best approximation $\inf_{v_h \in \mathcal{D}_h} \|u - v_h\|_0$ available.

Proposition 3.2.26. *The operator $B_2'\Lambda\colon \mathcal{W}_h^1 \to \mathcal{D}_h'$ is a consistent approximation to $B_1\colon \mathcal{W}_h^1 \to \mathcal{D}_h'$ i.e., there is a constant $c > 0$ independent of h satisfying*

$$\sup_{v \in \mathcal{D}_h} \frac{|\langle B_1 p + B_2'\Lambda p, v\rangle|}{\|v\|_0} \leq ch\|v\|_{H^1(\mathcal{T}_h)} \tag{3.138}$$

for every $p \in \mathcal{W}_h^1$.

Proof.

$$\sup_{v \in \mathcal{D}_h} \frac{|\langle B_1\Lambda^{-1}q + B_2'q, v\rangle|}{\|v\|_0} \tag{3.139}$$

$$= \sup_{v \in \mathcal{D}_h} \frac{|\langle B_1\Lambda^{-1}q + (\mathbb{I} - L' + L')B_2'q, v\rangle|}{\|v\|_0} \tag{3.140}$$

$$\leq \sup_{v \in \mathcal{D}_h} \frac{|\langle B_1\Lambda^{-1}q + L'B_2'q, v\rangle|}{\|v\|_0} + \sup_{v \in \mathcal{D}_h} \frac{|\langle (\mathbb{I} - L')B_2'q, v\rangle|}{\|v\|_0} \tag{3.141}$$

$$= \sup_{v \in \mathcal{D}_h} \frac{|\langle B_1\Lambda^{-1}q, v\rangle + \langle q, B_2 Lv\rangle|}{\|v\|_0} + \sup_{v \in \mathcal{D}_h} \frac{|\langle B_2'q, (\mathbb{I} - L)v\rangle|}{\|v\|_0} \tag{3.142}$$

$$\leq 0 + \|B_2'q\| \sup_{v \in \mathcal{D}_h} \frac{\|(\mathbb{I} - L)v\|_0}{\|v\|_0} \tag{3.143}$$

Therefore, there is a constant $c > 0$ independent of h satisfying

$$\sup_{v \in \mathcal{D}_h} \frac{|\langle B_1\Lambda^{-1}q + B_2'q, v\rangle|}{\|v\|_0} \leq ch\|v\|_{H^1(\mathcal{T}_h)}. \tag{3.144}$$

We conclude by substitution $q = \Lambda p$. □

We observe the following straightforward, but crucial consistency results in

Lemma 3.2.27. *Let $u \in \mathcal{U} \cap H^1(\Omega)^n$ with $b_2(u,q) = 0$ for every $q \in L^2(\Omega)$ then*

$$b_{2,h}(u, q_h) = 0 \quad \forall q_h \in \mathcal{W}_{h,0,\mathrm{bc}}'^0. \tag{3.145}$$

Proof. By assumption, we have

$$b_2(u, q_h) = 0 \quad \forall q_h \in \mathcal{W}'^0_{h,0,\mathrm{bc}}. \tag{3.146}$$

As every element $q_h \in \mathcal{W}'^0_{h,0,\mathrm{bc}}$ is a linear combination of indicator functions $\mathbb{1}_{K'}$ with support $K' \in \mathcal{T}_h{}'$ and b_2 is linear in the second argument we only need to prove

$$b_{2,h}(u, \mathbb{1}_K) = 0 \quad \forall K' \in \mathcal{T}_h{}'. \tag{3.147}$$

For this sake we first realize $u|_{K'} \in H^1(K')^n$. Subsequently, we apply Gauß' theorem (c.f. Theorem A.3.15 and [62, (2.17)] for the extension to $H(\mathrm{div}, \Omega)$) to conclude

$$\begin{aligned} b_{2,h}(u, \mathbb{1}_{K'}) &= \int_{\partial K'} \tau_{n_{\partial K'}}(u)\, dx && (3.148)\\ &= \int_{K'} \mathrm{div}\, u\, dx && (3.149)\\ &= b_2(u, \mathbb{1}_{K'}) = 0. && (3.150) \end{aligned}$$

□

Corollary 3.2.28. *Let $u \in \mathcal{U} \cap H^1(\Omega)^n$ such that $b_2(u, q) = 0$ for every $q \in L^2(\Omega)$, then*

$$\inf_{u_h \in \kappa_2} \|u - u_h\|_{\mathrm{DIV}_h} = \inf_{v_h \in \mathcal{D}_h} \|u - v_h\|_0. \tag{3.151}$$

Proof. We first realize $u \in \kappa_2 \cap \mathcal{U} \cap H^1(\Omega)^n$ by virtue of Lemma 3.2.27. Next we recall that $v_h \in \mathcal{D}_h$, as element of $L^2(\Omega)$ can be uniquely decomposed into two parts. One element of κ_2 and the other element of the orthogonal complement $\kappa_2^\perp$. Subsequently, we see

$$\begin{aligned} \inf_{u_h \in \kappa_2} \|u - u_h\|_{\mathrm{DIV}_h} &= \inf_{u_h \in \kappa_2} \|u - u_h\|_0 && (3.152)\\ &= \inf_{v_h \in \mathcal{D}_h} \|u - (v_h - v_{h,\kappa_2^\perp})\|_0 && (3.153)\\ &\leq \inf_{v_h \in \mathcal{D}_h} (\|u - v_h\|_0 + \|v_{h,\kappa_2^\perp}\|_0) && (3.154)\\ &= \inf_{v_h \in \mathcal{D}_h} \|u - v_h\|_0 && (3.155) \end{aligned}$$

The lower bound follows by

$$\inf_{v_h \in \mathcal{D}_h} \|u - v_h\|_0 = \inf_{\substack{v_{h,\kappa_2} \in \kappa_2 \\ v_{h,\kappa_2}^\perp \in \kappa_2^\perp}} \|u - (v_{h,\kappa_2^\perp} + v_{h,\kappa_2})\|_0 \tag{3.156}$$

$$\leq \inf_{\substack{v_{h,\kappa_2} \in \kappa_2 \\ v_{h,\kappa_2}^\perp \in \kappa_2^\perp}} \|v_{h,\kappa_2^\perp}\|_0 + \|u - v_{h,\kappa_2}\|_0 \tag{3.157}$$

$$= \inf_{u_h \in \kappa_2} \|u - u_h\|_0. \tag{3.158}$$

□

Proposition 3.2.29. *Let $\mathcal{T}_h$ be a quasi-uniform and shape-regular family of grids as given in Definition 2.1.1. Let $(u;p) \in (\boldsymbol{\mathcal{U}} \cap H^1(\Omega)^n) \times H^2(\Omega)$ be a solution to the analytical problem Eq.* (3.1)*. Then the approximate solution of Proposition 3.2.23, denoted by $(u_h;p_h)$ satisfies*

$$\|u - u_h\|_{\mathrm{DIV}_h} \leq \left(1 + \frac{\|a\|}{c_a}\right) ch\|u\|_1 \tag{3.159}$$

$$\|p - p_h\|_1 \leq \frac{\|a\|}{c_{b_1}} \left(1 + \frac{\|a\|}{c_a}\right) c\frac{h}{\tau}\|u\|_1 + \left(1 + \frac{\|b_1\|}{c_{b_1}}\right) ch\|p\|_1 \tag{3.160}$$

for a constant $c > 0$ independent of a,b_1,b_2,h and τ.

Proof. We follow the standard procedure presented in [27, p. II.2.2]. The following holds true for every $v_h \in \kappa_2$ uniformly due to Eq. (3.82) as well as for every $q_h \in \mathcal{W}_{h,0,\mathrm{bc}}^1$.

$$\|u - u_h\|_{\mathrm{DIV}_h} \leq \|u - v_h\|_{\mathrm{DIV}_h} + \|v_h - u_h\|_{\mathrm{DIV}_h} \tag{3.161}$$

$$\leq \|u - v_h\|_{\mathrm{DIV}_h} + \frac{1}{c_a} \sup_{w_h \in \kappa_1} \frac{a(u_h - v_h, w_h)}{\|w_h\|_0} \tag{3.162}$$

$$= \|u - v_h\|_{\mathrm{DIV}_h} + \frac{1}{c_a} \sup_{w_h \in \kappa_1} \frac{a(-v_h, w_h) + f(w_h)}{\|w_h\|_0} \tag{3.163}$$

$$= \|u - v_h\|_{\mathrm{DIV}_h} + \frac{1}{c_a} \sup_{w_h \in \kappa_1} \frac{a(u - v_h, w_h) + \tau b_1(p - p_h, w_h)}{\|w_h\|_0} \tag{3.164}$$

$$= \|u - v_h\|_{\mathrm{DIV}_h} + \frac{1}{c_a} \sup_{w_h \in \kappa_1} \frac{a(u - v_h, w_h) + \tau b_1(p - q_h, w_h)}{\|w_h\|_0} \tag{3.165}$$

$$\leq \|u - v_h\|_{\mathrm{DIV}_h} + \frac{\|a\|}{c_a}\|u - v_h\|_{\mathrm{DIV}_h} + \tau\frac{\|b_1\|}{c_a}\|p - q_h\|_1. \tag{3.166}$$

Consulting Corollary 3.2.28 we therefore obtain

$$\|u-u_h\|_{\mathrm{DIV}_h} \leq \left(1+\frac{\|a\|}{c_a}\right) \inf_{v_h\in\mathcal{D}_h} \|u-v_h\|_{\mathrm{DIV}_h} + \tau\frac{\|b_1\|}{c_a} \inf_{q_h\in\mathcal{W}^1_{h,0,\mathrm{bc}}} \|p-q_h\|_1 \tag{3.167}$$

By assumption we have

$$a(u-u_h,v_h)+\tau b_1(p-p_h,v_h)=0 \tag{3.168}$$

for every $v_h\in\mathcal{D}_h$. Let $q_h\in\mathcal{W}^1_{h,0,\mathrm{bc}}$, then by Lemma 3.2.15

$$\begin{aligned}
\|p-p_h\|_1 &\leq \|p-q_h\|_1+\|q_h-p_h\|_1 && (3.169)\\
&\leq \|p-q_h\|_1+\frac{1}{c_{b_1}}\sup_{v_h\in\kappa_1^\perp}\frac{b_1(q_h-p_h,v_h)}{\|v_h\|_0} && (3.170)\\
&= \|p-q_h\|_1+\frac{1}{\tau c_{b_1}}\sup_{v_h\in\kappa_1^\perp}\frac{a(u-u_h,v_h)+\tau b_1(p-q_h,v_h)}{\|v_h\|_0} && (3.171)\\
&\leq \|p-q_h\|_1+\frac{\|a\|}{\tau c_{b_1}}\|u-u_h\|_{\mathrm{DIV}_h}+\frac{\|b_1\|}{c_{b_1}}\|p-q_h\|_1. && (3.172)
\end{aligned}$$

Thus, we obtain

$$\begin{aligned}
&\|p-p_h\|_1+\|u-u_h\|_{\mathrm{DIV}_h}\\
&\leq \left(1+\frac{\|a\|}{c_a}+\frac{\|a\|}{\tau c_{b_1}}\right)\inf_{v_h\in\mathcal{D}_h}\|u-v_h\|_{\mathrm{DIV}_h}+\left(1+\frac{\|b_1\|}{c_{b_1}}+\tau\frac{\|b_1\|}{c_a}\right)\inf_{q_h\in\mathcal{W}^1_{h,0,\mathrm{bc}}}\|p-q_h\|_1 \qquad (3.173)
\end{aligned}$$

and by Proposition 2.4.17 we conclude the statement bounding the best approximation error from above by the specific choice of the interpolated solution in the sense of Proposition 2.4.17. □

3.2.8 The compressible regimes

Remark 3.2.30. As the finite dimension of $\mathcal{W}_h'^0$ and $\mathcal{W}_h^1$ coincide, they are obviously isomorphic. This fact, however, does not give any hint how the two subspaces of $L^2(\Omega)$ are oriented in the terms of the inner product $(\,\cdot\,,\,\cdot\,)_0$. The following lemma implies that the $L^2(\Omega)$ projection

$$\begin{cases}\mathcal{W}_h^1\to\mathcal{W}_h'^0\\ p\mapsto\sum_{\xi_i\in\mathcal{N}_{\mathcal{T}_h}}(p,\mathbb{1}_{K'_{\xi_i}})_0\,\mathbb{1}_{K'_{\xi_i}}\end{cases} \tag{3.174}$$

already provides such an isomorphism and therefore the spaces are not orthogonal to each other.

Lemma 3.2.31. *Let $\zeta \in L^\infty(\Omega)$ positive almost everywhere. The matrix $c_{i,j} = (\zeta\theta_{\xi_i}, \mathbb{1}_{K'_{\xi_j}})_0$ is symmetric and positive definite. More specifically there is $c_c > 0$ independent of h such that*

$$\sum_{\xi_i,\xi_j \in \mathcal{N}_{\mathcal{T}_h}} r_i r_j c_{i,j} \geq c_c \|r\|^2 \tag{3.175}$$

for every $r \in \mathbb{R}^{|\mathcal{N}_{\mathcal{T}_h}|}$.

Proof. Let $r, s \in r \in \mathbb{R}^{|\mathcal{N}_{\mathcal{T}_h}|}$ be arbitrary. We transform to the reference element by

$$\sum_{\xi_i,\xi_j \in \mathcal{N}_{\mathcal{T}_h}} r_i s_j (\zeta\theta_{\xi_i}, \mathbb{1}_{K'_{\xi_j}})_0 = \sum_{K\in\mathcal{T}_h} \sum_{\xi_i,\xi_j\in\mathcal{N}_K} r_i s_j \int_K \zeta\theta_{\xi_i} \mathbb{1}_{K'_{\xi_j}}\, dx \tag{3.176}$$

$$= \sum_{K\in\mathcal{T}_h} \sum_{\xi_i,\xi_j\in\mathcal{N}_K} r_i s_j \int_{\hat{K}} (\zeta\theta_{\xi_i} \mathbb{1}_{K'_{\xi_j}}) \circ T_K \det DT_K\, dx. \tag{3.177}$$

As both basis functions are positive almost everywhere, we can use the mean value theorem to find constants c_K for each element K with

$$0 < \operatorname{ess\,inf} \zeta < c_K < \operatorname{ess\,sup} \zeta \tag{3.178}$$

such that

$$\sum_{K\in\mathcal{T}_h} \sum_{\xi_i,\xi_j\in\mathcal{N}_K} r_i s_j \int_{\hat{K}} (\zeta\theta_{\xi_i} \mathbb{1}_{K'_{\xi_j}}) \circ T_K \det DT_K\, dx \tag{3.179}$$

$$= \sum_{K\in\mathcal{T}_h} c_K h^n \sum_{\xi_i,\xi_j\in\mathcal{N}_K} r_i s_j \int_{\hat{K}} \theta_{\xi_i} \circ T_K \mathbb{1}_{K'_{\xi_j}} \circ T_K\, dx \tag{3.180}$$

$$= \sum_{K\in\mathcal{T}_h} c_K h^n \sum_{i,j=1}^{|\mathcal{N}_K|} r_i s_j (G_K \tilde{C} G_K^T)_{i,j} \tag{3.181}$$

where $G_K \in \mathbb{R}^{|\mathcal{N}_{\mathcal{T}_h}|\times 2^n}$ transforms from a local index to a global one and

$$\tilde{C} = \frac{1}{64}\begin{pmatrix} 27 & 9 & 9 & 3 & 9 & 3 & 3 & 1 \\ 9 & 27 & 3 & 9 & 3 & 9 & 1 & 3 \\ 9 & 3 & 27 & 9 & 3 & 1 & 9 & 3 \\ 3 & 9 & 9 & 27 & 1 & 3 & 3 & 9 \\ 9 & 3 & 3 & 1 & 27 & 9 & 9 & 3 \\ 3 & 9 & 1 & 3 & 9 & 27 & 3 & 9 \\ 3 & 1 & 9 & 3 & 9 & 3 & 27 & 9 \\ 1 & 3 & 3 & 9 & 3 & 9 & 9 & 27 \end{pmatrix}. \tag{3.182}$$

This element-wise matrix is symmetric and therefore the assembled one is so too. Furthermore, the element-wise matrix has only positive eigenvalues, the smallest of them being $1/8$ and when choosing $r = s$ we have

$$\sum_{K\in\mathcal{T}_h} c_K h^n \sum_{i,j=1}^{|\mathcal{N}_K|} r_i r_j (G_K \tilde{C} G_K^T)_{i,j} \geq \sum_{K\in\mathcal{T}_h} \frac{c_K}{8} h^n \sum_{i=1}^{|\mathcal{N}_K|} r_i^2 (G_K G_K^T)_{i,i} \tag{3.183}$$

$$\geq \sum_{K\in\mathcal{T}_h} \frac{c_K}{8} h^n \sum_{\xi_i\in\mathcal{N}_K} r_i^2 \tag{3.184}$$

As the square of every coefficient appears in the sum at least once, there is a constant $c > 0$ independent of h such that

$$(\zeta p, q)_0 \geq c \operatorname{ess\,inf}(\zeta)\, h^n \sum_{i=1}^{|\mathcal{N}_{\mathcal{T}_h}|} q_i^2. \tag{3.185}$$

□

Corollary 3.2.32. *The identity*

$$c(p, q) = c(\Lambda^{-1} q, \Lambda p) \tag{3.186}$$

as well as

$$c(p, q)^2 \leq c(p, \Lambda p) c(\Lambda^{-1} q, q) \tag{3.187}$$

holds for every $q \in \mathcal{W}_h'^0$ and $p \in \mathcal{W}_h^1$.

Proof. The first statement follows from the symmetry established in Lemma 3.2.31. Following an idea of [27, Thm. 1.2], we establish the estimate by considering Lemma 3.2.31 and the non-negative

$$0 \leq c(p + s\Lambda^{-1} q, \Lambda p + sq) \tag{3.188}$$

for every $s \in \mathbb{R}$. This implies

$$0 \leq c(p + s\Lambda^{-1} q, \Lambda p + sq) \tag{3.189}$$

$$= c(p, \Lambda p) + sc(\Lambda^{-1} q, \Lambda p) + sc(p, q) + s^2 c(\Lambda^{-1} q, q) \tag{3.190}$$

$$= c(p, \Lambda p) + 2sc(p, q) + s^2 c(\Lambda^{-1} q, q) \tag{3.191}$$

for every $s \in \mathbb{R}$. Considering Lemma 3.2.31 again we additionally realize $c(\Lambda^{-1} q, q) > 0$ and $c(p, \Lambda p)$ for every $q \in \mathcal{W}_h'^0$ and $p \in \mathcal{W}_h^1$. Therefore, $c(p, q) = 0$ trivially satisfies the

required statement. Now let $c(p,q) \neq 0$ then the choice

$$s = -\frac{c(p,\Lambda p)}{c(p,q)} \tag{3.192}$$

gives

$$c(p,q) \leq c(p,\Lambda p)c(\Lambda^{-1}q,q) \tag{3.193}$$

after simple algebraic manipulation. □

Proposition 3.2.33. *Let $\zeta \in \mathbb{R}$, $\alpha_P > 0$ and let $h > 0$ be sufficiently small, then problem Eq.* (3.10) *with Neumann and (potentially) periodic boundary conditions in the sense of a, b_1, b_2, c has unique solution $p \in \mathcal{W}^1_{h,0,\mathrm{bc}}$ for every right-hand side. More specifically there is exactly one $(u;p) \in \mathcal{D}_h \times \mathcal{W}^1_{h,0,\mathrm{bc}}$ for every $(f;g) \in (\mathcal{D}_h; \|\cdot\|_0)' \times (\mathcal{W}'^0_{h,0,\mathrm{bc}}; \|\cdot\|_0)'$ such that*

$$\langle Au, v\rangle + \tau\langle B_1 p, v\rangle = \langle f, v\rangle \tag{3.194}$$

$$\tau\langle B_2 u, q\rangle + \quad \langle Cp, q\rangle \ = \langle g, q\rangle \tag{3.195}$$

for every $q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}$ and $v \in \mathcal{D}_h$.

Proof. For convenience, we denote

$$\mathcal{B}((u;p),(v;q)) = a(u,v) + \tau b_1(p,v) + \tau b_2(u,q) + c(p,q). \tag{3.196}$$

We aim to apply Theorem A.7.5 to the combined problem

$$\mathcal{B}((u;p),(v;q)) = f(v) + g(q). \tag{3.197}$$

The linear forms f, g are continuous functionals by definition and $\mathcal{B}$ due to Lemmas 3.2.14, 3.2.15 and 3.2.19 i.e., there is a constant $\|B\| \geq 0$ obeying

$$|\mathcal{B}((u;p),(v;q))| \leq \|\mathcal{B}\|\sqrt{\|p\|_1^2 + \|u\|^2_{\mathrm{DIV}_h}}\sqrt{\|q\|_0^2 + \|v\|_0^2} \tag{3.198}$$

for every $u, v \in \mathcal{D}_h$, $p \in \mathcal{W}^1_{h,0,\mathrm{bc}}$ and $q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}$.

Next we observe $\Lambda^{-1}q - \operatorname{avg}_\Omega q \in \mathcal{W}^1_{h,0,\mathrm{bc}}$ and realize $b_1(\Lambda^{-1}q - \operatorname{avg}_\Omega q, v) = b_1(\Lambda^{-1}q, v)$ as the gradient of every constant function vanishes. As $\zeta \in \mathbb{R}$ we also have

$$c(\operatorname{avg}_\Omega \Lambda^{-1}q, q) = \zeta \operatorname{avg}_\Omega(\Lambda^{-1}q)(1,q)_0 = 0. \tag{3.199}$$

Now we can choose a specific pair $(u;p) = (Lv;\Lambda_0^{-1}q)$ and consider Lemma 3.2.8 to find

$$\sup_{(u,p)\in\mathcal{D}_h\times\mathcal{W}^1_{h,0,\mathrm{bc}}} \mathcal{B}((u;p),(v;q)) \tag{3.200}$$

$$\geq a(Lv,v)+\tau b_1(\Lambda^{-1}q,v)+\tau b_2(Lv,q)+c(\Lambda^{-1}q-\operatorname{avg}_\Omega\Lambda^{-1}q,q) \tag{3.201}$$

$$= a(Lv,v)-a(Lv,v)+a(v,v)+0+c(\Lambda^{-1}q,q)-c(\operatorname{avg}_\Omega\Lambda^{-1}q,q)\geq \tag{3.202}$$

$$= a(v,v)+c(\Lambda^{-1}q,q)-|a(Lv-v,v)|-0 \tag{3.203}$$

for every $v\in\mathcal{D}_h$ and $q\in\mathcal{W}'^0_{h,0,\mathrm{bc}}$. Similarly to the proof of Lemma 3.2.14 there is a bound $\|a\|$ independent of h such that $|a(w,v)|\leq\|a\|\|w\|_0\|v\|_0$. This in turn allows another lower bound of Eq. (3.202) by the means of Lemma 3.2.12 and

$$a(v,v)+c(\Lambda^{-1}q,q)-|a(Lv-v,v)| \tag{3.204}$$

$$\geq c_a\|v\|_0^2+\zeta c_c\|q\|_0^2-\|a\|\|Lv-v\|_0\|v\|_0 \tag{3.205}$$

$$\geq c_a\|v\|_0^2+c_c\|q\|_0^2-\|a\|\|v\|_0 c_1 h \tag{3.206}$$

$$\geq c_a\|v\|_0(\|Lv\|_0-|\|Lv\|_0-\|v\|_0|)+c_c{\|q\|_0}^2-\|a\|\|v\|_0 c_1 h \tag{3.207}$$

$$\geq c_a\|v\|_0(\|Lv\|_0-c_3 h)+c_c{\|q\|_0}^2-\|a\|\|v\|_0 c_1 h \tag{3.208}$$

Therefore there is $h_0>0$ such that

$$\sup_{(u,p)\in\mathcal{D}_h\times\mathcal{W}^1_{h,0,\mathrm{bc}}} \mathcal{B}((u;p),(v;q))\geq\tilde{c}_a\|v\|_0\|u\|_0+\tilde{c}_c\|p\|_0\|q\|_0 \tag{3.209}$$

for every $h\in(0,h_0]$, $v\in\mathcal{D}_h$ and $q\in\mathcal{W}'^0_{h,0,\mathrm{bc}}$. As in finite dimensional spaces all norms are equivalent we obtain a constant, depending on h (c.f. Lemma B.0.4) and satisfying

$$\sup_{(u,p)\in\mathcal{D}_h\times\mathcal{W}^1_{h,0,\mathrm{bc}}} \frac{\mathcal{B}((u;p),(v;q))}{\sqrt{\|u\|^2_{\mathrm{DIV}_h}+\|p\|_1^2}}\geq c\sqrt{\|v\|_0^2+\|q\|_0^2}, \tag{3.210}$$

for every $h\leq h_0$, $v\in\mathcal{D}_h$ and $q\in\mathcal{W}'^0_{h,0,\mathrm{bc}}$.

Our last step is to prove the second condition of Definition A.7.4. To this end we recall $L\colon\mathcal{D}_h\to\mathcal{D}_h$ is bijective and therefore.

$$\sup_{(v,q)\in\mathcal{D}_h\times\mathcal{W}^1_{h,0,\mathrm{bc}}} \frac{\mathcal{B}((u;p),(v;q))}{\sqrt{\|v\|_0^2+\|q\|_0^2}}=\sup_{(\tilde{v},q)\in L\mathcal{D}_h\times\mathcal{W}^1_{h,0,\mathrm{bc}}} \frac{\mathcal{B}((u;p),(\tilde{v};q))}{\sqrt{\|\tilde{v}\|_0^2+\|q\|_0^2}} \tag{3.211}$$

$$= \sup_{(v,q)\in\mathcal{D}_h\times\mathcal{W}^1_{h,0,\mathrm{bc}}} \frac{\mathcal{B}((u;p),(Lv;q))}{\sqrt{\|Lv\|_0^2+\|q\|_0^2}} \tag{3.212}$$

Choosing $q = \Lambda p - \operatorname{avg}_\Omega \Lambda p$, $v = u$ we deduce

$$\sup_{(v,q)\in\mathcal{D}_h\times\mathcal{W}^1_{h,0,\mathrm{bc}}} \frac{\mathcal{B}((u;p),(Lv;q))}{\sqrt{\|Lv\|_0^2+\|q\|_0^2}} \tag{3.213}$$

$$\geq a(v,Lu) + b_1(p,Lu) + b_2(u,\Lambda p) + c(p,\Lambda p - \operatorname{avg}_\Omega \Lambda p) \tag{3.214}$$

$$\geq c_a\|u\|_0^2 + c_c\|p\|_0^2 - \|a\|\|Lu-u\|_0\|u\|_0 \tag{3.215}$$

$$\geq \tilde{c}_a\|u\|_0^2 + \tilde{c}_c\|p\|_0^2 \tag{3.216}$$

as before for every $u \in \mathcal{D}_h$ and $p \in \mathcal{W}^1_{h,0,\mathrm{bc}}$. Also, along the same lines we obtain a constant $c' > 0$, depending on h and such that

$$\sup_{(u,p)\in\mathcal{D}_h\times\mathcal{W}^1_{h,0,\mathrm{bc}}} \frac{\mathcal{B}((u;p),(v;q))}{\sqrt{\|v\|_0^2+\|q\|_0^2}} \geq c'\sqrt{\|u\|^2_{\mathrm{DIV}_h}+\|p\|_1^2}. \tag{3.217}$$

□

Proposition 3.2.34. *Let $(u;p)$ the solution provided by Proposition 3.2.33 and let $h_0 > 0$ be sufficiently small, then there is $c > 0$ independent of $h \in (0, h_0]$ with*

$$\|p\|_1 + \|u\|_0 \leq c\|f\|_0 \tag{3.218}$$

Proof. Let $(u;p)$ be the solution to Eq. (3.10). We follow the ideas of [27, Thm. 1.2] and decompose the solution $u = u_{\mathcal{K}_2} + u_{\mathcal{K}_2^\perp}$. Lemma 3.2.15 gives

$$\|p\|_1 \leq \frac{1}{c_{b_1}} \sup_{v\in\mathcal{D}_h} \frac{b_1(p,v)}{\|v\|_0} \tag{3.219}$$

$$= \frac{1}{c_{b_1}} \sup_{v\in\mathcal{D}_h} \frac{\langle f,v\rangle - a(u,v)}{\|v\|_0} \tag{3.220}$$

$$\leq \frac{1}{c_{b_1}}\|a\|\|u\|_{\mathrm{DIV}_h} + \sup_{v\in\mathcal{D}_h} \frac{\langle f,v\rangle}{\|v\|_0}. \tag{3.221}$$

Due to Lemma 3.2.8 we have

$$a(u,L^{-1}u) + c(p,\Lambda p - \operatorname{avg}\Lambda p) = f(L^{-1}u) + g(\Lambda p - \operatorname{avg}\Lambda p) \tag{3.222}$$

by summing both equations Eq. (3.10) and choosing the test functions appropriately. This additionally implies

$$a(u, L^{-1}u) + c(p, \Lambda p - \operatorname{avg} \Lambda p) \tag{3.223}$$

$$\leq \sup_{v \in \mathcal{D}_h} \frac{f(v)}{\|v\|_0} \|L^{-1}u\|_0 + \sup_{q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}} \frac{g(q)}{\|q\|_0} \|\Lambda p - \operatorname{avg} \Lambda p\|_0 \tag{3.224}$$

$$\leq \sup_{v \in \mathcal{D}_h} \frac{f(v)}{\|v\|_0} \|L^{-1}\| \|u\|_0 + \sup_{q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}} \frac{g(q)}{\|q\|_0} (\|\Lambda\| + \|\operatorname{avg} \Lambda\|) \|p\|_0. \tag{3.225}$$

Now consider the orthogonal decomposition of u. We apply Corollary 3.2.21 and see

$$\|u_{\kappa_2^\perp}\|_{\mathrm{DIV}_h} \leq \frac{c_{b'}}{\tau} \sup_{q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}} \frac{b_{2,h}(u_{\kappa_2^\perp}, q)}{\|q\|_0} \tag{3.226}$$

$$= \frac{c_{b'}}{\tau} \sup_{q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}} \frac{\langle g, q \rangle - c(p, q)}{\|q\|_0} \tag{3.227}$$

$$\leq \frac{c_{b'}}{\tau} \sup_{q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}} \frac{|\langle g, q \rangle|}{\|q\|_0} + \frac{c_{b'}}{\tau} \sup_{q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}} \frac{|c(p, q)|}{\|q\|_0} \tag{3.228}$$

$$\leq \frac{c_{b'}}{\tau} \sup_{q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}} \frac{|\langle g, q \rangle|}{\|q\|_0} + \frac{c_{b'}}{\tau} \sup_{q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}} \frac{\sqrt{c(p, \Lambda p) c(\Lambda^{-1} q, q)}}{\|q\|_0} \tag{3.229}$$

$$\leq \frac{c_{b'}}{\tau} \sup_{q \in \mathcal{W}'^0_{h,0,\mathrm{bc}}} \frac{|\langle g, q \rangle|}{\|q\|_0} + c_2 \sqrt{c(p, \Lambda p)} \sqrt{\|c\|} \tag{3.230}$$

for some constants $c_2 > 0$ determined by Eq. (3.23). On the orthogonal complement of κ_2 we leverage Eq. (3.82) such that

$$\|u_{\kappa_2}\|_{\mathrm{DIV}_h} \leq \frac{1}{c_a} \sup_{v \in \kappa_1} \frac{a(u_{\kappa_2}, v)}{\|v\|_0} \tag{3.231}$$

$$= \frac{1}{c_a} \sup_{v \in \kappa_1} \frac{a(u - u_{\kappa_2^\perp}, v)}{\|v\|_0} \tag{3.232}$$

$$= \frac{1}{c_a} \sup_{v \in \kappa_1} \frac{f(v) - a(u_{\kappa_2^\perp}, v)}{\|v\|_0} \tag{3.233}$$

$$\leq \frac{\|a\|}{c_a} \|u_{\kappa_2^\perp}\|_{\mathrm{DIV}_h} + \sup_{v \in \kappa} \frac{f(v)}{\|v\|_0}. \tag{3.234}$$

Combining Eqs. (3.219) and (3.223) and the recently established bound on $u = u_{\kappa_2} + u_{\kappa_2^\perp}$ we find some constant $\mathfrak{C} > 0$ which is bound only by the constants in the aforementioned

estimates and satisfying

$$a(u, L^{-1}u) + c(p, \Lambda p - \operatorname{avg} \Lambda p) \leq \mathfrak{C}(\sqrt{c(p, \Lambda p)} + 1). \tag{3.235}$$

For sufficiently small $h \in (0, h_0]$ we know $a(u, L^{-1}u) > 0$ as $a(u, u) > 0$ and due to Lemma 3.2.12 therefore we also have

$$c(p, \Lambda p) = c(p, \Lambda p - \operatorname{avg} \Lambda p) \leq \mathfrak{C}(\sqrt{c(p, \Lambda p)} + 1) \tag{3.236}$$

such that $c(p, \Lambda p)$ is bounded uniformly (with respect to h) by a constant. Subsequently, we find a bound on u by combining Eqs. (3.226) and (3.231) which ultimately gives the bound on p. □

Remark 3.2.35. Given the a-priori estimate, an error estimate [27, Prop. 2.11] is in reach, however out of scope of this work.

Remark 3.2.36. To avoid the rather restrictive assumption on ζ and still be able to follow the presented strategy of proof, we would have to find a lower bound $c_c(h)$ for $c(p, \Lambda_0 p) > c_c$, which is either non-negative or $\lim_{h \to 0} c_c(h) = 0$. The latter can be achieved by the approximation properties of $\mathcal{I}^0_{\mathcal{T}_h}$ in $\|\cdot\|_0$ and continuity of $\operatorname{avg}_\Omega \colon L^2(\Omega) \to \mathbb{R}$. Furthermore, we would have to refine estimate Corollary 3.2.32 to be applicable to Λ_0 instead of Λ.

4 Conclusion and future plans

In summary, we provided a refined analysis of the ideas developed by Vater and Klein [118], an appropriate interpolation operator and extend the method to quadrilateral and cuboid meshes in two and three spatial dimensions respectively. We prove stability and a bound on the approximation error in suitable mesh dependent norms for the pseudo-incompressible case, but only stability for the compressible scenario. This analysis does apply to the hydrostatic situation as well as to the general case. Furthermore, we established orthogonality of the projection in the incompressible case.

In the following the author would like to point out future plans and context of the presented results.

- Currently, a reference implementation for the presented general grid geometries is still under development and should supplement the presented work by numerical experiments.

- The missing error estimate in the compressible situation is somewhat unsatisfying. However, as our method is non-conforming the a priori estimate Proposition 3.2.34 alone is not sufficient to provide an estimate in line with [27] and therefore further work is required to resolve this issue.

- So far we provide only lowest order elements. Despite the fact that for Cartesian grids we expect second order convergence in the pressure variable (c.f. [113]), the estimate for the momentum variable does require further investigation. Furthermore, the question arises if the current numerical method can be extended to a higher order setting. To the knowledge of the author a positive answer to this question requires at least the restriction to affine grids, as one cannot establish the interpolation error in the required higher order semi-norms otherwise. The issue can be pinned to Lemma 2.1.19, which in our case is only stated for the full Sobolev norms or the concrete case of $|\cdot|_{1,p,\Omega}$. When considering quadrilaterals and cuboids, affine grids are, however, rather restrictive with respect to domain geometry. Nevertheless, this might be a worthwhile pursuit in context of the next open question.

- In context of the original goals and motivation, the arguably most intriguing question concerns the interplay between the bespoke advection method of [16] and the projection established in the work at hand. Here, the goal is to obtain a bound on the advected momentum variable in terms of the original state and the discrete divergence of the advecting field. Initial research hints at the averaging operator, which the authors of the aforementioned work apply to the advecting field.

Bibliography

[1] U. Achatz, R. Klein, and F. Senf. "Gravity Waves, Scale Asymptotics and the Pseudo-Incompressible Equations". In: *Journal of Fluid Mechanics* 663 (2010), pp. 120–147. DOI: 10.1017/S0022112010003411.

[2] R. A. Adams and J. J. F. Fournier. *Sobolev Spaces*. 2. ed., Reprinted. Pure and Applied Mathematics 140. Amsterdam: Acad. Press, 2008. 305 pp.

[3] L. Angermann. "Node-Centered Finite Volume Schemes and Primal-Dual Mixed Formulations". In: *Commun. Appl. Anal.* 7.4 (2003), pp. 529–565.

[4] A. Arakawa and V. R. Lamb. "Computational Design of the Basic Dynamical Processes of the UCLA General Circulation Model". In: *Methods in Computational Physics: Advances in Research and Applications*. Vol. 17. Elsevier, 1977, pp. 173–265. DOI: 10.1016/B978-0-12-460817-7.50009-4.

[5] D. N. Arnold. "An Interior Penalty Finite Element Method with Discontinuous Elements". In: *SIAM J. Numer. Anal.* 19.4 (1982), pp. 742–760. DOI: 10.1137/0719052.

[6] D. N. Arnold, D. Boffi, and R. S. Falk. "Approximation by Quadrilateral Finite Elements". In: *Math. Comp.* 71.239 (2002), pp. 909–922. DOI: 10.1090/S0025-5718-02-01439-4.

[7] D. N. Arnold, D. Boffi, and R. S. Falk. "Quadrilateral $H(\mathrm{div})$ Finite Elements". In: *SIAM J. Numer. Anal.* 42.6 (2005), pp. 2429–2451. DOI: 10.1137/S0036142903431924.

[8] D. N. Arnold, R. S. Falk, and R. Winther. "Finite Element Exterior Calculus, Homological Techniques, and Applications". In: *Acta Numerica* 15 (2006), pp. 1–155. DOI: 10.1017/S0962492906210018.

[9] D. N. Arnold et al. "Unified Analysis of Discontinuous Galerkin Methods for Elliptic Problems". In: *SIAM J. Numer. Anal.* 39.5 (2002), pp. 1749–1779. DOI: 10.1137/S0036142901384162.

[10] V. I. Arnold. *Mathematical Methods of Classical Mechanics*. 2nd ed. Graduate Texts in Mathematics 60. New York: Springer, 1997. 516 pp.

[11] I. Babuška. "Error-Bounds for Finite Element Method". In: *Numer. Math.* 16.4 (1971), pp. 322–333. DOI: 10.1007/BF02165003.

[12] I. Babuška. "The Finite Element Method with Lagrangian Multipliers". In: *Numer. Math.* 20.3 (1973), pp. 179–192. DOI: 10.1007/BF01436561.

[13] J. Baranger, J.-F. Maitre, and F. Oudin. "Connection between Finite Volume and Mixed Finite Element Methods". In: *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique* 30.4 (1996), pp. 445–465.

[14] C. Bardos, E. S. Titi, and E. Wiedemann. "The Vanishing Viscosity as a Selection Principle for the Euler Equations: The Case of 3D Shear Flow". In: *Comptes Rendus Mathematique* 350.15-16 (2012), pp. 757–760. DOI: 10.1016/j.crma.2012.09.005.

[15] J. B. Bell, P. Colella, and H. M. Glaz. "A Second-Order Projection Method for the Incompressible Navier-Stokes Equations". In: *Journal of Computational Physics* 85.2 (1989), pp. 257–283. DOI: 10.1016/0021-9991(89)90151-4.

[16] T. Benacchio and R. Klein. "A Semi-Implicit Compressible Model for Atmospheric Flows with Seamless Access to Soundproof and Hydrostatic Dynamics". In: *Mon. Wea. Rev.* 147.11 (2019), pp. 4221–4240. DOI: 10.1175/MWR-D-19-0073.1.

[17] M. Benzi, G. H. Golub, and J. Liesen. "Numerical Solution of Saddle Point Problems". In: *Acta Numerica* 14 (2005), pp. 1–137. DOI: 10.1017/S0962492904000212.

[18] S. Benzoni-Gavage and D. Serre. *Multidimensional Hyperbolic Partial Differential Equations: First-Order Systems and Applications*. Oxford Mathematical Monographs. Oxford ; New York: Clarendon Press, 2007. 508 pp.

[19] C. Bernardi and V. Girault. "A Local Regularization Operator for Triangular and Quadrilateral Finite Elements". In: *SIAM J. Numer. Anal.* 35.5 (1998), pp. 1893–1916. DOI: 10.1137/S0036142995293766.

[20] C. Bernardi. "Optimal Finite-Element Interpolation on Curved Domains". In: *SIAM J. Numer. Anal.* 26.5 (1989), pp. 1212–1240. DOI: 10.1137/0726068.

[21] C. Bernardi, C. Canuto, and Y. Maday. "Generalized Inf-Sup Conditions for Chebyshev Spectral Approximation of the Stokes Problem". In: *SIAM J. Numer. Anal.* 25.6 (1988), pp. 1237–1271. DOI: 10.1137/0725070.

[22] F. Bertrand and D. Boffi. "A Counterexample for the Inf-sup Stability of the $RT^0 - P^1 \subset L^2(\Omega) \times H_0^1(\Omega)$ Finite Element Combination for the Mixed Poisson Equation". In: *Proc. Appl. Math. Mech.* 19.1 (2019). DOI: 10.1002/pamm.201900426.

[23] H. Bhatia et al. "The Helmholtz-Hodge Decomposition—A Survey". In: *IEEE Trans. Visual. Comput. Graphics* 19.8 (2013), pp. 1386–1404. DOI: 10.1109/TVCG.2012.316.

[24] G. Bispen, M. Lukáčová-Medvidová, and L. Yelash. "Asymptotic Preserving IMEX Finite Volume Schemes for Low Mach Number Euler Equations with Gravitation". In: *Journal of Computational Physics* 335 (2017), pp. 222–248. DOI: 10.1016/j.jcp.2017.01.020.

[25] P. B. Bochev and D. Ridzal. "Rehabilitation of the Lowest-Order Raviart–Thomas Element on Quadrilateral Grids". In: *SIAM J. Numer. Anal.* 47.1 (2009), pp. 487–507. DOI: 10.1137/070704265.

[26] J. Březina and E. Feireisl. "Measure-Valued Solutions to the Complete Euler System Revisited". In: *Z. Angew. Math. Phys.* 69.3 (2018), p. 57. DOI: 10.1007/s00033-018-0951-8.

[27] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer Series in Computational Mathematics 15. New York: Springer-Verlag, 1991. 350 pp.

[28] F. Brezzi. "On the Existence, Uniqueness and Approximation of Saddle-Point Problems Arising from Lagrangian Multipliers." In: *Rev. Franc. Automat. Inform. Rech. Operat., R* 8.2 (1974), pp. 129–151.

[29] F. Brezzi. "The Inf-Sup Condition, the Bubble, and the Subgrid" (Gainesville). 2004.

[30] D. L. Brown, R. Cortez, and M. L. Minion. "Accurate Projection Methods for the Incompressible Navier–Stokes Equations". In: *Journal of Computational Physics* 168.2 (2001), pp. 464–499. DOI: 10.1006/jcph.2001.6715.

[31] G. Bruell and E. Feireisl. "On a Singular Limit for Stratified Compressible Fluids". In: *Nonlinear Analysis: Real World Applications* 44 (2018), pp. 334–346. DOI: 10.1016/j.nonrwa.2018.05.004.

[32] Z. Cai. "On the Finite Volume Element Method". In: *Numer. Math.* 58.1 (1990), pp. 713–735. DOI: 10.1007/BF01385651.

[33] E. Chiodaroli, C. De Lellis, and O. Kreml. "Global Ill-Posedness of the Isentropic System of Gas Dynamics". In: *Comm. Pure Appl. Math* 68.7 (2015), pp. 1157–1190. DOI: 10.1002/cpa.21537.

[34] A. J. Chorin. "The Numerical Solution of Navier-Stokes Equations for an Incompressible Fluid". In: *Bulletin of the American Mathematical Society* 73.6 (1967), pp. 928–931.

[35] A. J. Chorin and J. E. Marsden. *A Mathematical Introduction to Fluid Mechanics*. Red. by J. E. Marsden, L. Sirovich, and M. Golubitsky. Vol. 4. Texts in Applied Mathematics. New York, NY: Springer New York, 1993. DOI: 10.1007/978-1-4612-0883-9.

[36] P. Ciarlet. "Basic Error Estimates for Elliptic Problems". In: *Handbook of Numerical Analysis*. Vol. 2. Elsevier, 1991, pp. 17–351. DOI: 10.1016/S1570-8659(05)80039-0.

[37] P. Ciarlet and P.-A. Raviart. "Interpolation Theory over Curved Elements, with Applications to Finite Element Methods". In: *Computer Methods in Applied Mechanics and Engineering* 1.2 (1972), pp. 217–249. DOI: 10.1016/0045-7825(72)90006-0.

[38] P. G. Ciarlet. *Three-Dimensional Elasticity.* Burlington: Elsevier, 1988.

[39] Ph. Clément. "Approximation by Finite Element Functions Using Local Regularization". In: *R.A.I.R.O. Analyse Numérique* 9.R2 (1975), pp. 77–84. DOI: 10.1051/m2an/197509R200771.

[40] G. M. Constantine and T. H. Savits. "A Multivariate Faa Di Bruno Formula with Applications". In: *Trans. Amer. Math. Soc.* 348.2 (1996), pp. 503–520. DOI: 10.1090/S0002-9947-96-01501-2.

[41] F. Cordier, P. Degond, and A. Kumbaro. "An Asymptotic-Preserving All-Speed Scheme for the Euler and Navier–Stokes Equations". In: *Journal of Computational Physics* 231.17 (2012), pp. 5685–5704. DOI: 10.1016/j.jcp.2012.04.025.

[42] C. Cotter and J. Shipton. "Mixed Finite Elements for Numerical Weather Prediction". In: *Journal of Computational Physics* 231.21 (2012), pp. 7076–7091. DOI: 10.1016/j.jcp.2012.05.020.

[43] C. Cotter and J. Thuburn. "A Finite Element Exterior Calculus Framework for the Rotating Shallow-Water Equations". In: *Journal of Computational Physics* 257 (2014), pp. 1506–1526. DOI: 10.1016/j.jcp.2013.10.008.

[44] R. Courant, K. Friedrichs, and H. Lewy. "Über die partiellen Differenzengleichungen der mathematischen Physik". In: *Math. Ann.* 100.1 (1928), pp. 32–74. DOI: 10.1007/BF01448839.

[45] J. A. Curry and P. J. Webster. *Thermodynamics of Atmospheres and Oceans*. International Geophysics Series v. 65. San Diego: Academic Press, 1999. 471 pp.

[46] L. Demkowicz and J. Gopalakrishnan. "A Class of Discontinuous Petrov–Galerkin Methods. Part I: The Transport Equation". In: *Computer Methods in Applied Mechanics and Engineering* 199.23-24 (2010), pp. 1558–1572. DOI: 10.1016/j.cma.2010.01.003.

[47] L. Demkowicz and J. Gopalakrishnan. "Analysis of the DPG Method for the Poisson Equation". In: *SIAM J. Numer. Anal.* 49.5 (2011), pp. 1788–1809. DOI: 10.1137/100809799.

[48] E. Di Nezza, G. Palatucci, and E. Valdinoci. "Hitchhiker's Guide to the Fractional Sobolev Spaces". In: *Bulletin des Sciences Mathématiques* 136.5 (2012), pp. 521–573. DOI: 10.1016/j.bulsci.2011.12.004.

[49] D. A. Di Pietro and A. Ern. *Mathematical Aspects of Discontinuous Galerkin Methods*. Mathématiques et Applications 69. Berlin ; New York: Springer, 2012. 384 pp.

[50] R. J. DiPerna. "Measure-Valued Solutions to Conservation Laws". In: *Arch. Rational Mech. Anal.* 88.3 (1985), pp. 223–270. DOI: 10.1007/BF00752112.

[51] F. Dubois. "Finite Volumes and Mixed Petrov-Galerkin Finite Elements: The Unidimensional Problem". In: *Numerical Methods for Partial Differential Equations* 16.3 (2000), pp. 335–360. DOI: 10.1002/(SICI)1098-2426(200005)16:3<335::AID-NUM5>3.0.CO;2-X.

[52] F. Dubois, I. Greff, and C. Pierre. "Raviart–Thomas Finite Elements of Petrov–Galerkin Type". In: *ESAIM: M2AN* 53.5 (2019), pp. 1553–1576. DOI: 10.1051/m2an/2019020.

[53] D. R. Durran. "Improving the Anelastic Approximation". In: *Journal of the Atmospheric Sciences* 46.11 (1989), pp. 1453–1461. DOI: 10.1175/1520-0469(1989)046<1453:ITAA>2.0.CO;2.

[54] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements*. 2004.

[55] L. C. Evans. *Partial Differential Equations*. 2nd ed. Graduate Studies in Mathematics v. 19. Providence, R.I: American Mathematical Society, 2010. 749 pp.

[56] R. Eymard, T. Gallouët, and R. Herbin. "Finite Volume Approximation of Elliptic Problems and Convergence of an Approximate Gradient". In: *Applied Numerical Mathematics* 37.1-2 (2001), pp. 31–53. DOI: 10.1016/S0168-9274(00)00024-6.

[57] R. Eymard, T. Gallouët, and R. Herbin. "Finite Volume Methods". In: *Handbook of Numerical Analysis*. Vol. 7. Elsevier, 2000, pp. 713–1018. DOI: 10.1016/S1570-8659(00)07005-8.

[58] E. Feireisl, M. Lukáčová-Medvidová, and H. Mizerová. "Convergence of Finite Volume Schemes for the Euler Equations via Dissipative Measure-Valued Solutions". In: *Found Comput Math* 20.4 (2020), pp. 923–966. DOI: 10.1007/s10208-019-09433-z.

[59] E. Feireisl et al. “On the Convergence of a Finite Volume Method for the Navier–Stokes–Fourier System”. In: *IMA Journal of Numerical Analysis* (2020). DOI: 10.1093/imanum/draa060.

[60] E. Gagliardo. “Caratterizzazioni delle tracce sulla frontiera relative ad alcune classi di funzioni in n variabili”. In: *Rendiconti del Seminario Matematico della Università di Padova* 27 (1957), pp. 284–305.

[61] T. Gallouët et al. “Convergence of the Marker and Cell Scheme for the Incompressible Navier–Stokes Equations on Non-uniform Grids”. In: *Found Comput Math* 18.1 (2018), pp. 249–289. DOI: 10.1007/s10208-016-9338-4.

[62] V. Girault and P.-A. Raviart. *Finite Element Methods for Navier-Stokes Equations: Theory and Algorithms*. Springer Series in Computational Mathematics 5. Berlin: Springer, 1986. 374 pp.

[63] F. Golse and L. Saint-Raymond. “The Navier–Stokes Limit of the Boltzmann Equation for Bounded Collision Kernels”. In: *Invent. math.* 155.1 (2004), pp. 81–161. DOI: 10.1007/s00222-003-0316-5.

[64] P. Gwiazda, A. Świerczewska-Gwiazda, and E. Wiedemann. “Weak-Strong Uniqueness for Measure - Valued Solutions of Some Compressible Fluid Models”. In: *Nonlinearity* 28.11 (2015), pp. 3873–3890. DOI: 10.1088/0951-7715/28/11/3873.

[65] W. Hackbusch. “On First and Second Order Box Schemes”. In: *Computing* 41.4 (1989), pp. 277–296. DOI: 10.1007/BF02241218.

[66] F. H. Harlow and J. E. Welch. “Numerical Calculation of Time-Dependent Viscous Incompressible Flow of Fluid with Free Surface”. In: *Phys. Fluids* 8.12 (1965), p. 2182. DOI: 10.1063/1.1761178.

[67] H. Helmholtz. “Über Integrale Der Hydrodynamischen Gleichungen, Welche Den Wirbelbewegungen Entsprechen.” In: *Journal für die reine und angewandte Mathematik* 1858.55 (1858), pp. 25–55. DOI: 10.1515/crll.1858.55.25.

[68] W. V. D. Hodge. “A Dirichlet Problem for Harmonic Functionals, with Applications to Analytic Varities”. In: *Proceedings of the London Mathematical Society* s2-36.1 (1934), pp. 257–303. DOI: 10.1112/plms/s2-36.1.257.

[69] S. Holmes et al. “Earth Gravitational Model 2008”. In: 2008.

[70] P. Houston, C. Schwab, and E. Süli. “Discontinuous *Hp* -Finite Element Methods for Advection-Diffusion-Reaction Problems”. In: *SIAM J. Numer. Anal.* 39.6 (2002), pp. 2133–2163. DOI: 10.1137/S0036142900374111.

[71] T. Ikeda. *Maximum Principle in Finite Element Models for Convection-Diffusion Phenomena*. North-Holland Mathematics Studies 76. Amsterdam New York Oxford Tokyo: North-Holland Kinokuniya, 1983.

[72] K. Kadoya, N. Matsunaga, and A. Nagashima. "Viscosity and Thermal Conductivity of Dry Air in the Gaseous Phase". In: *Journal of Physical and Chemical Reference Data* 14.4 (1985), pp. 947–970. DOI: 10.1063/1.555744.

[73] M. Kaltenbäck. *Aufbau Analysis*. Berliner Studienreihe zur Mathematik 27. Lemgo: Heldermann Verlag, 2021. 392 pp.

[74] Th. von Kármán. "Ueber Den Mechanismus Des Widerstandes, Den Ein Bewegter Körper in Einer Flüssigkeit Erfährt". In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1911 (1911), pp. 509–517.

[75] T. Kato. "Nonstationary Flows of Viscous and Ideal Fluids in R3". In: *Journal of Functional Analysis* 9.3 (1972), pp. 296–305. DOI: 10.1016/0022-1236(72)90003-1.

[76] S. Klainerman and A. Majda. "Singular Limits of Quasilinear Hyperbolic Systems with Large Parameters and the Incompressible Limit of Compressible Fluids". In: *Comm. Pure Appl. Math.* 34.4 (1981), pp. 481–524. DOI: 10.1002/cpa.3160340405.

[77] S. Klainerman and A. Majda. "Compressible and Incompressible Fluids". In: *Comm. Pure Appl. Math.* 35.5 (1982), pp. 629–651. DOI: 10.1002/cpa.3160350503.

[78] R. Klein. "Semi-Implicit Extension of a Godunov-Type Scheme Based on Low Mach Number Asymptotics I: One-dimensional Flow". In: *Journal of Computational Physics* 121.2 (1995), pp. 213–237. DOI: 10.1016/S0021-9991(95)90034-9.

[79] R. Klein. "Asymptotics, Structure, and Integration of Sound-Proof Atmospheric Flow Equations". In: *Theor. Comput. Fluid Dyn.* 23.3 (2009), pp. 161–195. DOI: 10.1007/s00162-009-0104-y.

[80] R. Klein. "Scale-Dependent Models for Atmospheric Flows". In: *Annu. Rev. Fluid Mech.* 42.1 (2010), pp. 249–274. DOI: 10.1146/annurev-fluid-121108-145537.

[81] R. Klein et al. "Regime of Validity of Soundproof Atmospheric Flow Models". In: *Journal of the Atmospheric Sciences* 67.10 (2010), pp. 3226–3237. DOI: 10.1175/2010JAS3490.1.

[82] A. Kolmogorov. "The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds' Numbers". In: *Akademiia Nauk SSSR Doklady* 30 (1941), pp. 301–305.

[83] O. A. Ladyženskaja. *The Mathematical Theory of Viscous Incompressible Flow / by O. A. Ladyzhenskaya.* Rev. engl. ed., 2. print. New York [u.a.]: Gordon & Breach, 1964. XIV, 184 S.

[84] P. D. Lax and R. D. Richtmyer. "Survey of the Stability of Linear Finite Difference Equations". In: *Comm. Pure Appl. Math.* 9.2 (1956), pp. 267–293. DOI: 10.1002/cpa.3160090206.

[85] C. Lehrenfeld and J. Schöberl. "High Order Exactly Divergence-Free Hybrid Discontinuous Galerkin Methods for Unsteady Incompressible Flows". In: *Computer Methods in Applied Mechanics and Engineering* 307 (2016), pp. 339–361. DOI: 10.1016/j.cma.2016.04.025.

[86] J. Leray. "Sur Le Mouvement d'un Liquide Visqueux Emplissant l'espace". In: *Acta Math.* 63.0 (1934), pp. 193–248. DOI: 10.1007/BF02547354.

[87] R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. 1st ed. Cambridge University Press, 2002. DOI: 10.1017/CBO9780511791253.

[88] A. Majda. *Compressible Fluid Flow and Systems of Conservation Laws in Several Space Variables*. 1984.

[89] A. Majda. *Introduction to PDEs and Waves for the Atmosphere and Ocean*. Courant Lecture Notes in Mathematics 9. New York : Providence, R.I: Courant Institute of Mathematical Sciences ; American Mathematical Society, 2003. 234 pp.

[90] N. Masmoudi. "Rigorous Derivation of the Anelastic Approximation". In: *Journal de Mathématiques Pures et Appliquées* 88.3 (2007), pp. 230–240. DOI: 10.1016/j.matpur.2007.06.001.

[91] A. Masud and T. J. Hughes. "A Stabilized Mixed Finite Element Method for Darcy Flow". In: *Computer Methods in Applied Mechanics and Engineering* 191.39-40 (2002), pp. 4341–4370. DOI: 10.1016/S0045-7825(02)00371-7.

[92] N. G. Meyers and J. Serrin. "H = W". In: *Proceedings of the National Academy of Sciences* 51.6 (1964), pp. 1055–1056. DOI: 10.1073/pnas.51.6.1055.

[93] C.-D. Munz et al. "The Extension of Incompressible Flow Solvers to the Weakly Compressible Regime". In: *Computers & Fluids* 32.2 (2003), pp. 173–196. DOI: 10.1016/S0045-7930(02)00010-5.

[94] J. C. Nedelec. "Mixed Finite Elements in $\mathbb{R}^3$". In: *Numer. Math.* 35.3 (1980), pp. 315–341. DOI: 10.1007/BF01396415.

[95] M. Nelkin. "In What Sense Is Turbulence an Unsolved Problem?" In: *Science* 255.5044 (1992), pp. 566–570. DOI: 10.1126/science.255.5044.566.

[96] R. A. Nicolaides. "Existence, Uniqueness and Approximation for Generalized Saddle Point Problems". In: *SIAM J. Numer. Anal.* 19.2 (1982), pp. 349–357. DOI: 10.1137/0719021.

[97] R. A. Nicolaides. "Direct Discretization of Planar Div-Curl Problems". In: *SIAM J. Numer. Anal.* 29.1 (1992), pp. 32–56. DOI: 10.1137/0729003.

[98] J. Nitsche. "Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind". In: *Abh.Math.Semin.Univ.Hambg.* 36.1 (1971), pp. 9–15. DOI: 10.1007/BF02995904.

[99] J. Pedlosky. *Geophysical Fluid Dynamics*. New York, NY: Springer New York, 1987. DOI: 10.1007/978-1-4612-4650-3.

[100] L. Prandtl. "Über Flüssigkeitsbewegung Bei Sehr Kleiner Reibung. Verhandlungen Des III. Internationalen Mathematiker Kongresses". In: *Verhandlungen Des III. Internationalen Mathematiker Kongresses*. Heidelberg: B. G. Teubner, Leipzig, 1904.

[101] P. A. Raviart and J. M. Thomas. "A Mixed Finite Element Method for 2nd Order Elliptic Problems". In: *Mathematical Aspects of Finite Element Methods*. Ed. by I. Galligani and E. Magenes. Vol. 606. Berlin, Heidelberg: Springer Berlin Heidelberg, 1977, pp. 292–315. DOI: 10.1007/BFb0064470.

[102] P.-A. Raviart and J. M. Thomas. "Primal Hybrid Finite Element Methods for 2nd Order Elliptic Equations". In: *Math. Comp.* 31.138 (1977), pp. 391–391. DOI: 10.1090/S0025-5718-1977-0431752-8.

[103] W. H. Reed and T. R. Hill. "Triangular Mesh Methods for the Neutron Transport Equation". In: United States, 1973.

[104] O. Reynolds. "XXIX. An Experimental Investigation of the Circumstances Which Determine Whether the Motion of Water Shall Be Direct or Sinuous, and of the Law of Resistance in Parallel Channels". In: *Phil. Trans. R. Soc.* 174 (1883), pp. 935–982. DOI: 10.1098/rstl.1883.0029.

[105] W. Rudin. *Functional Analysis*. 2nd ed. International Series in Pure and Applied Mathematics. New York: McGraw-Hill, 1991. 424 pp.

[106] T. Schneider et al. "Extension of Finite Volume Compressible Flow Solvers to Multi-dimensional, Variable Density Zero Mach Number Flows". In: *Journal of Computational Physics* 155.2 (1999), pp. 248–286. DOI: 10.1006/jcph.1999.6327.

[107] G. Schwarz. *Hodge Decomposition: A Method for Solving Boundary Value Problems*. Lecture Notes in Mathematics 1607. Berlin ; New York: Springer-Verlag, 1995. 155 pp.

[108] M. Shashkov and S. Steinberg. *Conservative Finite-Difference Methods on General Grids*. Symbolic and Numeric Computation Series. Boca Raton: CRC Press, 1996. 359 pp.

[109] P. K. Smolarkiewicz. "A Fully Multidimensional Positive Definite Advection Transport Algorithm with Small Implicit Diffusion". In: *Journal of Computational Physics* 54.2 (1984), pp. 325–362. DOI: 10.1016/0021-9991(84)90121-9.

[110] P. K. Smolarkiewicz, C. Kühnlein, and N. P. Wedi. "A Consistent Framework for Discrete Integrations of Soundproof and Compressible PDEs of Atmospheric Dynamics". In: *Journal of Computational Physics* 263 (2014), pp. 185–205. DOI: 10.1016/j.jcp.2014.01.031.

[111] P. K. Smolarkiewicz and L. O. Margolin. "On Forward-in-Time Differencing for Fluids: Extension to a Curvilinear Framework". In: *Monthly Weather Review* 121.6 (1993), pp. 1847–1859. DOI: 10.1175/1520-0493(1993)121<1847:OFITDF>2.0.CO;2.

[112] P. K. Smolarkiewicz, J. Szmelter, and A. A. Wyszogrodzki. "An Unstructured-Mesh Atmospheric Model for Nonhydrostatic Dynamics". In: *Journal of Computational Physics* 254 (2013), pp. 184–199. DOI: 10.1016/j.jcp.2013.07.027.

[113] E. Süli. "Convergence of Finite Volume Schemes for Poisson's Equation on Nonuniform Meshes". In: *SIAM J. Numer. Anal.* 28.5 (1991), pp. 1419–1430. DOI: 10.1137/0728073.

[114] H. S. G. Swann. "The Convergence with Vanishing Viscosity of Nonstationary Navier-Stokes Flow to Ideal Flow in R 3". In: *Transactions of the American Mathematical Society* 157 (1971), p. 373. DOI: 10.2307/1995853.

[115] J.-M. Thomas and D. Trujillo. "Mixed Finite Volume Methods". In: *International Journal for Numerical Methods in Engineering* 46.9 (1999), pp. 1351–1366. DOI: 10.1002/(SICI)1097-0207(19991130)46:9<1351::AID-NME702>3.0.CO;2-0.

[116] J. Thuburn and C. J. Cotter. "A Primal–Dual Mimetic Finite Element Scheme for the Rotating Shallow Water Equations on Polygonal Spherical Meshes". In: *Journal of Computational Physics* 290 (2015), pp. 274–297. DOI: 10.1016/j.jcp.2015.02.045.

[117] G. K. Vallis. *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation*. 2nd ed. Cambridge: Cambridge University Press, 2017. DOI: 10.1017/9781107588417.

[118] S. Vater and R. Klein. "Stability of a Cartesian Grid Projection Method for Zero Froude Number Shallow Water Flows". In: *Numer. Math.* 113.1 (2009), pp. 123–161. DOI: 10.1007/s00211-009-0224-8.

[119] G. Volpe. "Performance of Compressible Flow Codes at Low Mach Numbers". In: *AIAA Journal* 31.1 (1993), pp. 49–56. DOI: 10.2514/3.11317.

[120] H. Weyl. "The Method of Orthogonal Projection in Potential Theory". In: *Duke Math. J.* 7.1 (1940), pp. 411–444. DOI: 10.1215/S0012-7094-40-00725-6.

[121] M. F. Wheeler. "An Elliptic Collocation-Finite Element Method with Interior Penalties". In: *SIAM J. Numer. Anal.* 15.1 (1978), pp. 152–161. DOI: 10.1137/0715010.

Appendices

A Function spaces

Throughout this chapter we denote a bounded domain $\Omega \subseteq \mathbb{R}^n$ and its boundary Γ. The outer normal vector onto the boundary is denoted by $\nu|_\Gamma$ and for $n = 2$ the tangential vector of the boundary is denoted by $t|_\Gamma$.

Borrowing the function spaces from the treatment of the Stokes equation

$$\begin{aligned} \nu \Delta u + \operatorname{grad} p &= f \\ \operatorname{div} u &= 0 \end{aligned}', \tag{A.1}$$

we follow [62] and introduce the (classical) Lebesgue and Sobolev spaces.

A.1 Lebesgue spaces

Let $\Omega \subseteq \mathbb{R}^n$ be open and bounded and let $p \in [1, \infty)$. Then the space of p-integrable functions is given by equivalence classes of Lebesgue measurable functions $\Omega \to \mathbb{R}$ that satisfy

$$\|f\|_{L^p(\Omega)} \coloneqq \left(\int_\Omega |f|^p \, dx \right)^{1/p} < \infty. \tag{A.2}$$

Here, the equivalence classes are given by the identification of functions that coincide almost everywhere. The map (A.2) is a norm on these functions and the resulting normed space, denoted by $L^p(\Omega)$ is a Banach space. In the special case of $p = 2$ on can equip the space with the classical $L^2(\Omega)$ inner product

$$(f, g) \coloneqq \int_\Omega f g \, dx \tag{A.3}$$

to obtain a Hilbert space.

Remark A.1.1. For the vector valued case $f, g \in L^2(\Omega)^n$ it is common to use the same symbols to denote the inner product

$$(f, g) \coloneqq \int_\Omega f \cdot g \, dx, \tag{A.4}$$

which then naturally induces the norm

$$\|f\|_{L^2(\Omega)^n} \coloneqq \sqrt{(f,f)} = \sqrt{\sum_{i=1}^{n} \|f_i\|^2_{L^2(\Omega)}}. \tag{A.5}$$

Lemma A.1.2. *Let $\omega \in L^\infty(\Omega)$ with $\operatorname{ess\,inf} \omega > 0$, then $L^2(\Omega)$ equipped with the inner product*

$$(f,g)_\omega \coloneqq (\omega f, g) \tag{A.6}$$

is a Hilbert space. Furthermore, the norms induced by $(\,\cdot\,,\,\cdot\,)$ and $(\,\cdot\,,\,\cdot\,)_\omega$ are equivalent i.e.,

$$\operatorname{ess\,inf}(\sqrt{\omega})\|f\|_0 \leq \|f\|_{\omega,0} \leq \operatorname{ess\,sup}(\sqrt{\omega})\|f\|_0 \quad \forall f \in L^2(\Omega) \tag{A.7}$$

Definition A.1.3. Let $f \in L^1(\Omega)$ and let $C \subset \Omega$ then the average of f over C given by

$$f_C = \frac{1}{|C|} \int_C f \, dx, \tag{A.8}$$

where $|C|$ denotes the Lebesgue measure of C.

Lemma A.1.4. *Let $p \in [1,\infty)$ and let $f \in L^p(\Omega)$, then*

$$\|\operatorname{avg} f\|_0 \leq c\|f\|_0. \tag{A.9}$$

Remark A.1.5. As we only consider bounded domains we have

$$L^q(\Omega) \subset L^p(\Omega), \tag{A.10}$$

for every combination $1 \leq p < q \leq \infty$. Therefore, the average is well-defined on $L^p(\Omega)$, for every $p \in [0,\infty]$.

A.1.1 Surface measure

Let $\mathcal{M} \subset \mathbb{R}^n$ be a $(n-1)$ dimensional C^1 manifold. In the following we introduce the notion of the surface measure and the corresponding space $L^p(\mathcal{M})$.

As we only require to discuss relatively simple manifolds we present the result only for the case where M can be covered by one diffeomorphic chart $\varphi\colon B_r(0) \to \mathcal{M}$ for some radius $r > 0$. In this case the surface measure of a set A, element of the mapped

$n-1$-dimensional Borel sigma algebra on $\mathcal{M}$, is given [73] by

$$\mu(A) = \int\limits_{\varphi^{-1}(A)} \sqrt{\det D\varphi(s)^T D\varphi(s)}\, d\lambda(s). \tag{A.11}$$

Therefore, the surface integral of $f : \mathcal{M} \to \mathbb{R}$ over compatible set $S \subset \mathcal{M}$ (see before) is given by

$$\int\limits_S f\, d\mu = \int\limits_{\varphi^{-1}(S)} f \circ \phi(s) \sqrt{\det D\varphi(s)^T D\varphi(s)}\, d\lambda(s). \tag{A.12}$$

This surface integral gives rise to the definition of $L^p(S)$ which consists of all the Borel-measurable functions with $\int\limits_S |f|^p\, d\mu \leq \infty$.

Remark A.1.6. As Eq. (A.12) relates the surface integral with usual Lebesgue integral these spaces behave exactly as their previously defined counterparts.

A.2 Distributions

Let $\Omega \subset \mathbb{R}^n$ be open and bounded, then the space of compactly supported infinitely often differentiable functions from Ω into the real numbers is denoted by

$$\mathcal{D}(\Omega) \coloneqq \left\{ \phi \in C^\infty(\Omega) : \operatorname{supp} \phi \subset \Omega \wedge \operatorname{supp} \phi = \overline{\operatorname{supp} \phi} \right\}. \tag{A.13}$$

As it turns out, we can approximate every member of $L^p(\Omega)$ by a sequence of such functions.

Lemma A.2.1. *Let $p \in [1, \infty)$, then $\mathcal{D}(\Omega)$ is dense in $L^p(\Omega)$.*

The dual space of $\mathcal{D}(\Omega)$ i.e., the space of all linear continuous functionals $\mathcal{D}(\Omega) \to \mathbb{R}$ are the distributions[1] $\mathcal{D}'(\Omega)$. Some distributions $\Lambda_f \in \mathcal{D}'(\Omega)$ allow for a locally integrable representation $f \in L^1(\Omega)_{\text{loc}}$ via

$$\Lambda_f(\phi) = \int_\Omega \phi f\, dx. \tag{A.14}$$

If applicable, we henceforth identify the distribution Λ_f and f.

[1] A concise, but rigorous introduction on the matter of distributions is presented in the excellent textbook by Rudin [105].

Definition A.2.2 (Distributional derivative)**.** One defines the distributional derivative $g \in \mathcal{D}'(\Omega)$ of $f \in \mathcal{D}'(\Omega)$ via the duality pairing $\langle f, \phi \rangle := f(\phi)$ such that

$$\langle D^\alpha f, \phi \rangle = (-1)^{|\alpha|} \langle f, D^\alpha \phi \rangle \quad \forall \phi \in \mathcal{D}(\Omega) \tag{A.15}$$

If additionally $f \in C^1$, then this coincides with the well known integration by parts

$$(D^\alpha f, \phi) = (-1)^{|\alpha|} (f, D^\alpha \phi) \quad \forall \phi \in \mathcal{D}(\Omega) \tag{A.16}$$

As e.g., shown in [105], one can do calculus with the distributional derivative. Therefore, we can define the vector calculus operators by just replacing the classical differential by the distributional derivative.

Definition A.2.3 (Distributional vector calculus)**.** Let $f \in \mathcal{D}'(\Omega)$ and $g \in \mathcal{D}'(\Omega)^n$, then the distributional gradient, and divergence are given by

$$\operatorname{grad} f := \sum_{i=1}^{n} e_i D^{e_i} f \tag{A.17}$$

$$\operatorname{div} g := \sum_{i=1}^{n} D^{e_i} g_i. \tag{A.18}$$

For $n = 2$, $n = 3$ we additionally define

$$\operatorname{curl} g := D^{e_1} g_2 - D^{e_2} g_1, \quad \operatorname{curl} g := \begin{pmatrix} D^{e_2} g_3 - D^{e_3} g_2 \\ D^{e_3} g_1 - D^{e_1} g_3 \\ D^{e_1} g_2 - D^{e_2} g_1 \end{pmatrix} \text{respectively.} \tag{A.19}$$

The following Lemma is a direct consequence of the definition of the distributional derivative.

Lemma A.2.4. *Let $f \in \mathcal{D}'(\Omega)^n$, then distributional divergence is characterized by $g \in \mathcal{D}'(\Omega)$ such that*

$$\langle g, \phi \rangle = -\sum_{i=0}^{n} \langle f_i, (\operatorname{grad} \phi)_i \rangle \quad \forall \phi \in \mathcal{D}(\Omega)\,. \tag{A.20}$$

A.3 Sobolev spaces

In this section we introduce Sobolev spaces. Each of them is the subspace of $L^p(\Omega)$, whose elements distributional derivatives are again in $L^p(\Omega)$.

Definition A.3.1 (Sobolev spaces)**.** Let $\Omega \subseteq \mathbb{R}^n$ be a domain and let $k \in \mathbb{N}$ and $p \in [1, \infty)$, then we define the norm

$$\|f\|_{k,p} := \left(\sum_{|\alpha| \leq k} \|D^\alpha f\|^p_{L^p(\Omega)} \right)^{1/p}, \tag{A.21}$$

and introduce the following spaces

$$W^{k,p}(\Omega) := \{f \in L^p(\Omega) : D^\alpha f \in L^p(\Omega),\ \forall \alpha : |\alpha| \leq k\}, \tag{A.22}$$

$$H^{k,p}(\Omega) := \overline{C^k}^{\|\cdot\|_{k,p}}, \tag{A.23}$$

$$H_0^{k,p}(\Omega) := \overline{\mathcal{D}(\Omega)}^{\|\cdot\|_{k,p}}. \tag{A.24}$$

Lemma A.3.2 (Poincare [55])**.** *Let $v \in H_0^{1,p}(\Omega)$ then there is a constant $C_1 > 0$ depending only on the domain Ω and $p \in [1, \infty)$ such that*

$$\|v - v_\Omega\|_{L^p(\Omega)} \leq c_1 \|\operatorname{grad} v\|_{L^p(\Omega)}. \tag{A.25}$$

Let $v \in H_0^{1,p}(\Omega)$ then there is another constant $C_2 > 0$ depending only on the domain Ω and $p \in [1, \infty)$ such that

$$\|v\|_{L^p(\Omega)} \leq c_2 \|\operatorname{grad} v\|_{L^p(\Omega)}. \tag{A.26}$$

Remark A.3.3. Using the preceding Poincare inequality one concludes that the semi-norm

$$|f|_{1,p} := \left(\sum_{|\alpha| = k} \|D^\alpha f\|^p_{L^p(\Omega)} \right)^{1/p}, \tag{A.27}$$

is a norm on $H_0^{1,p}(\Omega)$.

Theorem A.3.4 ([2, Theorem 3.3])**.** *Let $k \in \mathbb{N}$ and $p \in [1, \infty)$, then $W^{k,p}(\Omega)$ equipped with the norm $\|\cdot\|_{k,p}$ is a Banach space.*

Theorem A.3.5 ($H = W$ [92])**.** *Let $\Omega \subseteq \mathbb{R}^n$ be a domain, $k \in \mathbb{N}$ and $p \in [1, \infty)$, then the definitions introduced above coincide i.e.,*

$$H^{k,p}(\Omega) = W^{k,p}(\Omega). \tag{A.28}$$

Remark A.3.6. In the special case $p = 2$ we denote the Sobolev spaces by $H^k(\Omega)$ and

introduce a shortened notation by

$$\|f\|_k \coloneqq \|f\|_{H^k(\Omega)}. \tag{A.29}$$

Additionally, we obtain a similar result to Theorem A.3.5 for smooth functions on the closure of Lipschitz domains.

Theorem A.3.7 ([2, Thereom 3.22])**.** *Let $\Omega \subseteq \mathbb{R}^n$ be a Lipschitz domain, $k \in \mathbb{N}$ and $p \in [1, \infty)$ then $\mathcal{D}\left(\overline{\Omega}\right)$ is dense in $W^{k,p}(\Omega)$.*

Definition A.3.8 (Fractional Sobolev Spaces)**.** Let $s \leq 0$ and $\sigma \in (0, 1)$ with $k = \lfloor s \rfloor$ and $s = k + \sigma$ then we generalize the notion of a Sobolev space using the norm

$$\|f\|_{s,p} = \left(\|f\|_{k,p}^p + \sum_{|\alpha|=k} \int_\Omega \int_\Omega \frac{|D^\alpha f(x) - D^\alpha f(y)|^p}{\|x - y\|^{n+\sigma p}} \right)^{1/p} \tag{A.30}$$

and define

$$W^{s,p}(\Omega) \coloneqq \left\{ f \in L^p(\Omega) : \|f\|_{s,p} < \infty \right\}. \tag{A.31}$$

Remark A.3.9. Let $s > 0$, then one again can define appropriate spaces $H^s(\Omega)$ via the Fourier transformation on the whole space and subsequent restriction to Ω. On Lipschitz domains we can conclude the equivalence $W^{s,2}(\Omega) = H^s(\Omega)$.

Remark A.3.10. Let $s > 0$). We denote the dual space of $H^s(\Omega)$ by $H^{-s}(\Omega)$. The operator norm on $H^{-s}(\Omega)$ is given as

$$\|\cdot\|_{H^{-s}(\Omega)} \coloneqq \sup_{\|f\|_{H^s(\Omega)}=1} \langle \cdot, f \rangle \tag{A.32}$$

Remark A.3.11. Generally elements of $H^{-s}(\Omega)$ are only in $\mathcal{D}'(\Omega)$. If $f' \in H^{-s}(\Omega)$ additionally satisfies $f \in L^2(\Omega)$ then we have

$$\langle f, \cdot \rangle = (f, \cdot). \tag{A.33}$$

The preceding Definition A.3.8 and Theorem A.3.7 enables us to define a trace operator which allows the evaluation at the boundary. A priori, the elements of $W^{k,p}(\Omega) \subset L^p(\Omega)$ do not allow a well-defined point evaluation. Nevertheless, we can evaluate in the sense of some dual space.

Theorem A.3.12 (Trace theorem [60])**.** *Let $p \in (1, \infty)$. Let $\Omega \subseteq \mathbb{R}^n$ be a Lipschitz domain then*

the trace operator

$$\tau_\Gamma : \begin{cases} \mathcal{D}\left(\overline{\Omega}\right) \to W^{1-1/p,p}(\Gamma) \\ v \mapsto v|_\gamma \end{cases} \tag{A.34}$$

can be uniquely extended to $\mathcal{B}(W^{1,p}(\Omega), W^{1-1/p,p}(\Gamma))$. *Furthermore, for every* $f W^{1-1/p,p}(\Gamma)$ *there is at least one* $v \in W^{1,p}(\Omega)$ *with* $\tau(v) = f$ *and*

$$\|v\|_{W^{1,p}(\Omega)} \leq c\|f\|_{W^{1-1/p,p}(\Omega)}, \tag{A.35}$$

where $c > 0$ *is independent of* f *and* v.

In the following we adapt this result to the situation in the presented work.

Corollary A.3.13. *Assume the situation of Theorem A.3.12. Consider a Borel measurable set (c.f. Eq.* (A.12)*)* $S \subset \Gamma$ *then the restricted trace*

$$\tau_S \begin{cases} \mathcal{D}\left(\overline{\Omega}\right) \to H^{1/2}(S) \\ v \mapsto v|_S \end{cases} \tag{A.36}$$

can be uniquely extended to $\mathcal{B}(H^1(\Omega), H^{1/2}(S))$. *Furthermore, there is at least one* $v \in H^1(\Omega)$ *with* $\tau_S(v) = f$ *and*

$$\|v\|_{H^1(\Omega)} \leq c\|f\|_{H^{1/2}(S)}, \tag{A.37}$$

where $c > 0$ *is independent of* f *and* v.

Proof. We first observe

$$\tau_\Gamma(v)|_S = \tau_S(v) \tag{A.38}$$

by density i.e., by

$$(\tau_\Gamma(\varphi), v)_S = (\tau_S(\varphi), v)_S \tag{A.39}$$

for every $v \in H^{1/2}(S)$, and $\varphi \in \mathcal{D}\left(\overline{S}\right)$. We also have $\|v\|_{H^{1/2}(S)} \leq \|v\|_{H^{1/2}(\Gamma)}$ for every $v \in H^{1/2}(\Gamma)$. Therefore, we conclude

$$\|\tau_S(v)\|_{H^{1/2}(S)} \leq \|\tau_\Gamma(v)\|_{H^{1/2}(\Gamma)} \leq c\|v\|_{H^1(\Omega)}. \tag{A.40}$$

Minor modifications of [48, Lem.5.1, Lem.5.2 and Thm 5.4] show the continuous embedding of $W^{s,p}(S)$ in $W^{s,p}(\Gamma)$ for $s \in (0,1)$, $p \in [1,\infty)$. Therefore, we have for every $v \in W^{s,p}(S)$ some $w_v \in W^{s,p}(\Gamma)$ such that $w_v|_S = v$ and there is $c > 0$ independent of v

with

$$\|v\|_{W^{s,p}(\Gamma)} \leq c\|v\|_{W^{s,p}(S)}. \tag{A.41}$$

In combination with Theorem A.3.12 we obtain some $u_v \in W^{1,2}(\Omega)$ with

$$\tau_S(u_v) = \tau_\Gamma(u_v)|_S = w_v|_S = v \tag{A.42}$$

and

$$\|u_v\|_{W^{1,2}(\Omega)} \leq \tilde{c}\|v\|_{W^{1/2,2}(S)} \tag{A.43}$$

for every $v \in W^{1/2,2}(S)$. The constant $\tilde{c} > 0$ is independent of v. □

Remark A.3.14. Every bounded smooth manifold is part of the boundary of a Lipschitz domain.

Theorem A.3.15 (Gauß-Green-Ostrogradski)**.** *Let $n \in \mathbb{N}$, $\Omega \subset \mathbb{R}^n$ be a Lipschitz domain and let Γ denote its boundary. Let $p \in [1, \infty)$ and $u \in H^{1,p}(\Omega)^n$ then*

$$\int_\Omega \operatorname{div} u \, dx = \int_\Gamma \tau_\Gamma(u) \cdot n_\Gamma \, dx. \tag{A.44}$$

Proof. As Ω is bounded, the continuous differentiable functions $C^1(\Omega)$ are dense in $H^{1,p}(\Omega)$. For every sequence $(u_k)_{k\in\mathbb{N}}$ with $u_k \in C^1(\Omega)$ and

$$\lim_{k\to\infty} \|u_k - u\|_1 = 0 \tag{A.45}$$

the classical result

$$\int_\Omega \operatorname{div} u_k \, dx = \int_\Gamma u_k \cdot n_\Gamma \, dx \tag{A.46}$$

is true. Applying Hölder's inequality we obtain

$$\left|\int_\Omega \operatorname{div} u - \operatorname{div} u_k \, dx\right| \leq c\|u - u_k\|_{1,p} \tag{A.47}$$

as well as we obtain

$$\left|\int_\Gamma (u_k \tau_\Gamma(u)) \cdot n_\Gamma \, dx\right| \leq c\|u - u_k\|_{W^{1,p}(\Omega)} \tag{A.48}$$

Passing to the limit delivers the desired result. □

Lemma A.3.16.

$$\ker \tau_\Gamma = H_0^1(\Omega) \tag{A.49}$$

For the treatment of the following Neumann boundary value problem of the weighted Poisson equation i.e.,

$$-\operatorname{div}(\rho \operatorname{grad} u(x)) = f(x) \quad x \in \Omega \tag{A.50a}$$

$$\frac{\partial u}{\partial n}(x) = g \quad x \in \Gamma \tag{A.50b}$$

one considers the weak formulation

$$(\operatorname{grad} u, \operatorname{grad} v)_\rho = (f, v) \quad \forall v \in H^1(\Omega). \tag{A.51}$$

As it turns out, the standard space $H^1(\Omega)$ is not the right choice to obtain well-posedness of Eq. (A.51). The crucial ingredient to apply the Lax-Milgram theorem is the coercivity of the bilinear form on the left-hand side. This property in our case requires an estimate of the type

$$\|u\|_0 \leq c\|\operatorname{grad} u\|_0, \tag{A.52}$$

which is not true on $H^1(\Omega)$, but on the subspace $H_0^1(\Omega)$. As we do want to allow for non-zero boundary values, this is not the correct choice either. The remedy is the use of the subspace

$$H_{*,0}^1(\Omega) = \left\{ v \in H^1(\Omega) \colon \int_\Omega v \, dx = 0 \right\}. \tag{A.53}$$

Remark A.3.17. As null space of a bounded linear map, $H_{*,0}^1(\Omega)$ is closed and as closed subspace of the Hilbert space $H^1(\Omega)$ again a Hilbert space equipped with the same inner product and norm. Furthermore, for every element $v \in H^1(\Omega)$ we find a real number $r \coloneqq \operatorname{avg}_\Omega v \in \mathbb{R}$ and $v_0 \in H_{*,0}^1(\Omega)$ such that $\|v - (v_0 + r)\|_0 = 0$. In other words $H_{*,0}^1(\Omega)$ is the factor space containing the equivalence classes of functions in $H^1(\Omega)$ which differ only by a real number i.e., which have the same gradient almost everywhere or equivalently

$$\|\operatorname{grad}(v) - \operatorname{grad}(v_0)\|_0 = 0. \tag{A.54}$$

Theorem A.3.18. *Let $f \in L^2(\Omega)$ such that $\int_\Omega f \, dx = 0$ then Eq.* (A.51) *has a unique solution u with*

$$\|u\|_1 \leq c\|f\|_0 \tag{A.55}$$

Proof. This is a classical result therefore we omit some details. We first restate the problem

by

$$(\operatorname{grad} u, \operatorname{grad} v)_\rho = (f, v) \qquad \forall v \in H^1(\Omega) \tag{A.56}$$

$$\Leftrightarrow (\operatorname{grad} u, \operatorname{grad}(v - v_\Omega))_\rho = (f, (v - v_\Omega)) \quad \forall v \in H^1(\Omega) \tag{A.57}$$

$$\Leftrightarrow (\operatorname{grad} u, \operatorname{grad} w)_\rho = (f, w) \qquad \forall w \in H^1_{*,0}(\Omega) \tag{A.58}$$

As both sides are bounded, Eq. (A.7) and by the virtue of the Poincaré inequality Eq. (A.25), we can now apply the Lax-Milgram theorem to obtain a unique $u \in H^1_{*,0}(\Omega)$ satisfying Eq. (A.58) and

$$\|u\|_1 \leq c\|f\|_0. \tag{A.59}$$

Since $H^1_{*,0}(\Omega) \subset H^1(\Omega)$ the equivalence of Eq. (A.51) and Eq. (A.58) concludes the proof. □

Remark A.3.19. The condition $\int_\Omega f \, dx$ is necessary as one can easily see by $v \equiv 1$, which is in $H^1(\Omega)$ on bounded domains.

Remark A.3.20. For the vector valued case we again use the straightforward construction from Remark A.1.1.

A.4 Sobolev spaces for divergence and rotation

It turns out that for use of the vector calculus operators div and curl one does not need the full regularity of $H^1(\Omega)$ to stay in $L^2(\Omega)$. In resemblance of Definition A.3.1 we introduce the following spaces.

Definition A.4.1. Let $n \in \mathbb{N}$ and $\Omega \subseteq \mathbb{R}^n$ be a Lipschitz domain. We introduce the Hilbert spaces

$$H(\operatorname{div}, \Omega) \coloneqq \left\{ f \in L^2(\Omega)^n : \operatorname{div} f \in L^2(\Omega) \right\} \tag{A.60}$$

$$H_0(\operatorname{div}, \Omega) \coloneqq \overline{\mathcal{D}(\Omega)^n}^{H(\operatorname{div},\Omega)} \tag{A.61}$$

and their inner product

$$(u, v)_{H(\operatorname{div},\Omega)} \coloneqq (u, v) + (\operatorname{div} u, \operatorname{div} v) \tag{A.62}$$

As in the case of classical Sobolev spaces we can characterize the elements of $H(\operatorname{div}, \Omega)$ by functions smooth up to the boundary.

Lemma A.4.2. *Let $\Omega \subseteq \mathbb{R}^n$ be a Lipschitz domain, then*

$$H(\mathrm{div}, \Omega) = \overline{\mathcal{D}\left(\overline{\Omega}\right)^n}^{H(\mathrm{div},\Omega)}. \tag{A.63}$$

Using a minor modification to the classical result (see e.g., [62]) we can evaluate an element from $H(\mathrm{div}, \Omega)$ at some sufficiently smooth surface inside the domain.

Lemma A.4.3. *Let $\Omega \subseteq \mathbb{R}^n$ be a bounded Lipschitz domain and let S smooth $n-1$-manifold inside a compact set $K \subset \Omega$. Then we can extend the normal trace on S*

$$\tau_{n_S} : \begin{cases} \mathcal{D}\left(\overline{\Omega}\right)^n \to \mathcal{D}(S) \\ v \mapsto v|_S \cdot n_S \end{cases} \tag{A.64}$$

to $\mathcal{B}(H(\mathrm{div}, \Omega), H^{-1/2}(S))$.

Proof. Let

$$\Omega_a^b = \{y \in \mathbb{R}^n : \exists x \in S, \xi \in [-a, b] : y = x + \xi n_S(x)\}. \tag{A.65}$$

As S is a compact set inside the open set Ω, there is $\varepsilon > 0$ such that

$$\Omega_0^\varepsilon \subset \Omega_\varepsilon^\varepsilon \subset \Omega. \tag{A.66}$$

Obviously S is contained in the boundary of this new set i.e., $S \in \partial\Omega_0^\varepsilon$. For every $v \in H(\mathrm{div}, \Omega)$ and Lipschitz domain $\tilde{\Omega} \subset \Omega$ we have $v \in H_{\tilde{\Omega}}(\mathrm{div}, \Omega)$ and $\|v\|_{H_{\tilde{\Omega}}(\mathrm{div},\Omega)} \leq \|v\|_{H(\mathrm{div},\Omega)}$. After integration by parts and the use of the Cauchy-Schwarz inequality we therefore obtain

$$\begin{aligned}
\left|\int_S \phi v \cdot n_S \, dS\right| &= \left|\int_{\partial\Omega_\varepsilon^\varepsilon} \phi v \cdot n_S \, dS - \int_{\partial\Omega_0^\varepsilon} \phi v \cdot n_S \, dS\right| && \text{(A.67)} \\
&\leq \left|\int_{\Omega_\varepsilon^\varepsilon} \phi \, \mathrm{div}(v) \, dx\right| + \left|\int_{\Omega_\varepsilon^\varepsilon} v \cdot \mathrm{grad}\, \phi \, dx\right| \\
&\quad + \left|\int_{\Omega_0^\varepsilon} \phi \, \mathrm{div}(v) \, dx\right| + \left|\int_{\Omega_0^\varepsilon} v \cdot \mathrm{grad}\, \phi \, dx\right| && \text{(A.68)} \\
&\leq \|\phi\|_{L^2(\Omega_\varepsilon^\varepsilon)} \|\mathrm{div}(v)\|_{L^2(\Omega_\varepsilon^\varepsilon)} + \|\mathrm{grad}\, \phi\|_{L^2(\Omega_\varepsilon^\varepsilon)} \|v\|_{L^2(\Omega_\varepsilon^\varepsilon)} && \text{(A.69)} \\
&\quad + \|\phi\|_{L^2(\Omega_0^\varepsilon)} \|\mathrm{div}(v)\|_{L^2(\Omega_0^\varepsilon)} + \|\mathrm{grad}\, \phi\|_{L^2(\Omega_0^\varepsilon)} \|v\|_{L^2(\Omega_0^\varepsilon)} && \text{(A.70)} \\
&\leq 2\|v\|_{H(\mathrm{div},\Omega)} \|\phi\|_{H^1(\Omega)} && \text{(A.71)}
\end{aligned}$$

for every $v \in \mathcal{D}\left(\overline{\Omega}\right)^n$ and $\phi \in \mathcal{D}\left(\overline{\Omega}\right)$. Since the $\mathcal{D}\left(\overline{\Omega}\right)$ is dense in $H^1(\Omega)$, we obtain the same result for every $\phi \in H^1(\Omega)$.

Next we use Corollary A.3.13 (c.f. also [62]) to conclude there is $\varphi \in H^{1/2}(S)$ with $\tau_S(\varphi) = \phi$ and

$$\left| \int_S \varphi v \cdot n_S \, dS \right| \leq c\|v\|_{H(\mathrm{div},\Omega)}\|\phi\|_{H^1(\Omega)} \leq \tilde{c}\|v\|_{H(\mathrm{div},\Omega)}\|\varphi\|_{H^{1/2}(S)}. \tag{A.72}$$

Therefore, we have

$$\|\tau_S(v)\|_{H^{-1/2}(S)} = \sup_{\phi \in H^{1/2}(S)} \frac{\langle \tau_S(v), \phi \rangle}{\|\phi\|_{H^{1/2}(S)}} \tag{A.73}$$

$$= \sup_{\phi \in H^{1/2}(S)} \frac{(\tau_S(v), \phi)_{L^2(S)}}{\|\phi\|_{H^{1/2}(S)}} \tag{A.74}$$

$$\leq \tilde{c}\|v\|_{H(\mathrm{div},\Omega)} \tag{A.75}$$

i.e., the τ_S is continuous. The trace is furthermore linear and therefore there is a unique extension of τ_S to $\mathcal{B}(H(\mathrm{div},\Omega), H^{-1/2}(S))$. □

Lemma A.4.4. *Let $\Omega \subseteq \mathbb{R}^n$ be Lipschitz domain, then we can identify the elements with vanishing trace by*

$$H_0(\mathrm{div},\Omega) = \{f \in H(\mathrm{div},\Omega) \colon f \cdot n|_\Gamma = 0\}. \tag{A.76}$$

Remark A.4.5. We denote the closed subspace of divergence free elements of $H_0(\mathrm{div},\Omega)$ by

$$H_{0,0}(\mathrm{div},\Omega) \coloneqq \left\{f \in H_0(\mathrm{div},\Omega) \colon f \in \ker \mathrm{div}\right\}. \tag{A.77}$$

Lemma A.4.6 (Gauss)**.** *For every $v \in H(\mathrm{div},\Omega)$ and every $q \in H^1(\Omega)$ we have*

$$(p, \mathrm{div}\, v) + (\mathrm{grad}\, p, v) = (\tau_\Gamma p, \tau_\nu v)_{L^2(\Gamma)}. \tag{A.78}$$

Corollary A.4.7. *Let $\omega \in \Omega$ and let $u \in H(\mathrm{div},\Omega)$ then*

$$\int_\omega \mathrm{div}\, u \, dx = \int_{\partial\omega} \tau_{\nu_\omega}(u) \, dx. \tag{A.79}$$

Lemma A.4.8. *The orthogonal complement of $H_{0,0}(\mathrm{div},\Omega)^\perp$ are exactly the gradient fields of $L^2(\Omega)^n$ i.e.,*

$$H_{0,0}(\mathrm{div},\Omega)^\perp = \{q \in L^2(\Omega) \colon \exists p \in H^1(\Omega) \colon q = \mathrm{grad}\, p\} \tag{A.80}$$

$H_{0,0}(\mathrm{div},\Omega)$ is closed and for every $v \in H_{0,0}(\mathrm{div},\Omega)$ there is a sequence of $\phi_k \in \mathcal{D}(\Omega) \cap H_{0,0}(\mathrm{div},\Omega)$ such that $\lim_{k\to\infty}\|v-\phi_k\|_0$. Furthermore, for every $v \in H_{0,0}(\mathrm{div},\Omega)^\perp$ there is a sequence of $\phi_k \in \mathcal{D}(\Omega) \cap H_{0,0}(\mathrm{div},\Omega)^\perp$ such that $\lim_{k\to\infty}\|v-\phi_k\|_0$.

Definition A.4.9. Let $n \in \{2,3\}$, and $\Omega \subseteq \mathbb{R}^n$ be a Lipschitz domain. We introduce the spaces

$$H(\mathrm{curl},\Omega) \coloneqq \left\{ f \in L^2(\Omega)^n : \mathrm{curl}\, f \in L^2(\Omega)^n \right\} \tag{A.81}$$

$$H_0(\mathrm{curl},\Omega) \coloneqq \overline{\mathcal{D}(\Omega)^n}^{H(\mathrm{curl},\Omega)} \tag{A.82}$$

and their norm by

$$\|f\|_{H(\mathrm{curl},\Omega)} \coloneqq \left(\|f\|_0^2 + \|\mathrm{curl}\, f\|_0^2.\right)^{1/2} \tag{A.83}$$

Again we can characterize the elements of $H(\mathrm{curl},\Omega)$ by functions smooth up to the boundary.

Lemma A.4.10. *Let $n \in \{2,3\}$ and $\Omega \subseteq \mathbb{R}^n$ be a Lipschitz domain, then*

$$H(\mathrm{curl},\Omega) = \overline{\mathcal{D}\left(\overline{\Omega}\right)^n}^{H(\mathrm{curl},\Omega)}. \tag{A.84}$$

Lemma A.4.11. *Let $n \in \{2,3\}$ and let $\Omega \subseteq \mathbb{R}^n$ be a Lipschitz domain, then we can extend the tangential derivative*

$$\tau_t : \begin{cases} \mathcal{D}(\Omega)^2 & \to H^{-1/2}(\Gamma) \\ v & \to v \cdot t|_\Gamma \end{cases} \qquad \tau_t : \begin{cases} \mathcal{D}(\Omega)^3 & \to H^{-1/2}(\Gamma) \\ v & \to v \times \nu|_\Gamma \end{cases} \tag{A.85}$$

uniquely to $\mathcal{B}(H(\mathrm{curl},\Omega), H^{-1/2}(\Gamma))$.

Lemma A.4.12. *Under the assumptions of the preceding lemma we have*

$$\ker \tau_t = H_0(\mathrm{curl},\Omega). \tag{A.86}$$

A.5 Helmholtz decomposition

Helmholtz [67] introduces the decomposition of a smooth three-dimensional vector field into an irrotational component governed by a potential and a solenoidal part. This decomposition, also known as the Helmholtz decomposition, is crucial for the analysis of the

incompressible Navier-Stokes and Euler equations [86] as well as for numerical approximation of their solutions [62, 34]. On compact domains, the decomposition is generalized by Weyl [120] based on the work of Hodge [68]. In the introductory part in [107] one can find a more detailed overview over the historical development and [23] provides a rather recent review on the Helmholtz decomposition focused on perspective of applications.

On bounded connected Lipschitz domains, one can state the result for Sobolev spaces as presented in [62]. Henceforth, we refer to the latter version which is stated in the following.

Theorem A.5.1 (Helmholtz-Hodge-Weyl decomposition)**.** *Let Ω be a bounded, connected, but not necessarily simply connected, Lipschitz domain. Then for every $v \in L^2(\Omega)^3$ there is the orthogonal decomposition*

$$v = \operatorname{curl} \phi + \operatorname{grad} q, \tag{A.87}$$

where $q \in H^1(\Omega)/\mathbb{R}$ is the only solution of

$$(\operatorname{grad} q, \operatorname{grad} \mu) = (v, \operatorname{grad} \mu) \quad \forall \mu \in H^1(\Omega)^3 \tag{A.88}$$

and $\phi \in H^1(\Omega)^3$ with $\operatorname{curl} \phi \in \left\{ v \in H_0(\operatorname{div}, \Omega) : \operatorname{div} v = 0 \right\}$. If Ω is additionally simply connected, then ϕ is uniquely determined in $H(\operatorname{curl}, \Omega)$.

A.6 Broken Sobolev spaces

As this work considers discontinuous functions, the classical Sobolev spaces are not necessarily an appropriate analytical framework. As usual in discontinuous Galerkin methods [49], we consider these spaces element wise. For this purpose we recall the following broken Sobolev spaces.

Definition A.6.1. Let $\mathcal{T}_h$ be a grid as given in Definition 2.1.1 and let $n \in \{2, 3\}$, then

$$H^1(\mathcal{T}_h) \coloneqq \left\{ f \in L^2(\Omega) \colon \operatorname{grad} f|_K \in L^2(K)^n \quad \forall K \in \mathcal{K} \right\} \tag{A.89}$$

$$H(\operatorname{div}, \mathcal{T}_h) \coloneqq \left\{ f \in L^2(\Omega)^n \colon \operatorname{div} f|_K \in L^2(K) \quad \forall K \in \mathcal{K} \right\} \tag{A.90}$$

$$H(\operatorname{curl}, \mathcal{T}_h) \coloneqq \left\{ f \in L^2(\Omega)^n \colon \operatorname{curl} f|_K \in L^2(K) \vee \operatorname{curl} f|_K \in L^2(K)^n \quad \forall K \in \mathcal{K} \right\} \tag{A.91}$$

Remark A.6.2. Equipped with the norms

$$\|f\|_{H^1(\mathcal{T}_h)} \coloneqq \|f\|_0 + \sum_{K \in \mathcal{T}_h} \|\operatorname{grad} f|_K\|_0, \tag{A.92}$$

$$\|f\|_{H(\operatorname{div}, \mathcal{T}_h)} \coloneqq \|f\|_0 + \sum_{K \in \mathcal{T}_h} \|\operatorname{div} f|_K\|_0, \tag{A.93}$$

$$\|f\|_{H(\mathrm{curl},\mathcal{T}_h)} \coloneqq \|f\|_0 + \sum_{K\in\mathcal{T}_h} \|\mathrm{curl}\, f|_K\|_0, \tag{A.94}$$

$H^1(\mathcal{T}_h)$, $H(\mathrm{div},\mathcal{K})$ and $H(\mathrm{curl},\mathcal{K})$ are Banach spaces (see e.g., [49]).

The question when a Sobolev space contains its broken counterpart can be characterized by the question of the continuity properties across the element boundaries.

Remark A.6.3. Let $\Omega = \bigcup_{K\in\mathcal{T}_h} K$ the domain and denote its boundary by Γ. Every $f \in H(\mathrm{div},\mathcal{K}) \cap H^1(\mathcal{K})$ is in $H(\mathrm{div},\Omega)$ if and only if the normal jump across the elements vanishes (see e.g., [49]). This we can readily see by

$$\begin{aligned}
\langle \mathrm{div}\, u, \phi\rangle &= -\int_\Omega u\, \mathrm{grad}\, \phi \\
&= \sum_{K\in\mathcal{T}_h} \int_K \phi\, \mathrm{div}\, u\, dx - \int_{\partial K} \phi u\cdot \nu\, dx \\
&= \sum_{K\in\mathcal{T}_h} \int_K \phi\, \mathrm{div}\, u\, dx - \int_\Gamma \phi u\cdot \nu_\Gamma\, dx - \sum_{E\in\mathcal{E}} \int_E [\![u]\!]_E \cdot \nu_E \phi\, dx \\
&= \sum_{K\in\mathcal{T}_h} \int_K \phi\, \mathrm{div}\, u\, dx - \sum_{E\in\mathcal{E}} \int_E [\![u]\!]_E \cdot \nu_E \phi\, dx \quad \forall \phi \in \mathcal{D}(\Omega)\,.
\end{aligned} \tag{A.95}$$

For the spaces $H^1(\mathcal{T}_h)$ and $H(\mathrm{curl},\mathcal{T}_h)$ one obtains similar conditions for jumps in every and tangential direction respectively.

A.7 Abstract framework: Saddle point problem

Let $(X,(\cdot,\cdot)_X),(Y,(\cdot,\cdot)_Y)$ be Hilbert spaces. Let furthermore be $b\colon X\times Y\to\mathbb{R}$ a continuous bilinear form i.e., linear in both arguments and

$$|b(x,y)| \le \|b\|\|x\|\|y\| \quad \forall x\in X, y\in Y, \tag{A.96}$$

where

$$\|b\| \coloneqq \sup_{\substack{x\in X\colon \|x\|_X=1\\ y\in Y\colon \|y\|_Y=1}} b(x,y). \tag{A.97}$$

Remark A.7.1. Each bounded bilinear form $b\colon X\times Y\to\mathbb{R}$ uniquely determines a bounded operator $B\in\mathcal{B}(X,Y')$ by

$$B \coloneqq \begin{cases} X\to Y' \\ x\mapsto b(x,\cdot) \end{cases} \tag{A.98}$$

B is linear due to the linearity of b in the first argument. The bound on b implies continuity of B as

$$\|Bx\|_{Y'} = \sup_{\substack{y\in Y \\ \|y\|_Y=1}} b(x,y) \leq \|b\|\|x\|_X \tag{A.99}$$

Using the same construction on the other variable one obtains the transpose operator $B^t : Y \to X'$ which is uniquely determined by

$$\langle Bx, y\rangle_{Y'\times Y} = \langle x, B'y\rangle_{X\times X'} \quad \forall x \in X, y \in Y. \tag{A.100}$$

Remark A.7.2. The Riesz representation theorem allows identification the dual of $L^2(\Omega)$ by itself. In light of a Hilbert space $Z \subset L^2(\Omega)$ it is important to distinguish between the dual operator, defined by the duality product and the adjoint operator defined by the respective inner product. In general both do not coincide. However, $L^2(\Omega)$ is contained in Z' in the sense of the Riesz representation theorem. Therefore, sufficiently regular elements allow for the following identification

$$\langle x', \cdot\rangle_{Z'\times Z} = (z', \cdot)_0. \tag{A.101}$$

Remark A.7.3. Let X, Y, Z be finite dimensional Hilbert spaces, with basis elements a_i, b_i, c_i respectively. Let $T : X \to Y$,$S : Y \to Z'$ be linear operators, with matrix representations

$$T_{i,j} = \langle Ta_i, b_j\rangle_{Y'\times Y}, \quad S_{j,k} = \langle Sb_j, c_k\rangle_{Z'\times Z} \tag{A.102}$$

for every $i \in \{1 \dots \dim X\}$, $j \in \{1 \dots \dim Y\}$ and $k \in \{1 \dots \dim Z\}$. Then we can express the composition of S and T by

$$\langle S\iota Tx, z\rangle_{Z'Z} = \sum_{i=1}^{\dim X}\sum_{j=1}^{\dim Y}\sum_{k=1}^{\dim Z} \gamma_k c_i S_{j,k} T_{i,j} \alpha_i a_i \tag{A.103}$$

for every $x = \sum_{i=1}^{\dim X} \alpha_i a_i \in X$ and $y = \sum_{j=1}^{\dim Z} \gamma\gamma_j c_i \in Z$, where ι denotes the Riesz isomorphism.

If T^{-1} exists then its matrix representation is $T^{-1}_{i,j}$.

Let $f \in X'$, $g \in Y'$ and $a : X \times X \to \mathbb{R}$ be a continuous symmetric positive semi definite bilinear form i.e., as before, but additionally we assume

$$a(x,y) = a(y,x) \text{ and } a(x,x) \geq 0 \quad \forall x, y \in X. \tag{A.104}$$

The critical points of the Lagrangian functional

$$\mathcal{L}(x,y) = \frac{1}{2}a(x,x) - f(x) + b(x,y) - g(y) \tag{A.105}$$

are given by the solution $(x^*; y^*)$ of the problem

$$a(x^*, x) + b(x, y^*) = f(x) \quad \forall x \in X, \tag{A.106a}$$

$$b(x^*, y) = g(y) \quad \forall y \in Y. \tag{A.106b}$$

Each of the critical points $(x^*; y^*)$ is a saddle point i.e., satisfies

$$\mathcal{L}(x^*, y) \leq \mathcal{L}(x^*, y^*) \leq \mathcal{L}(x, y^*) \quad \forall x \in X, y \in Y. \tag{A.107}$$

The first inequality in Eq. (A.107) is trivially fulfilled as we obtain even equality by Eq. (A.106b). The second inequality follows by the first variation of Eq. (A.106a) with respect to x and the fact that A is positive semi definite. For this reason the system Eq. (A.106) is called a saddle point problem. One can also state the problem on the level of operators, which becomes especially useful if one considers a finite dimensional setting, where the operators A, B, B' become matrices and directly tell the structure of the system matrix one has to invert numerically as

$$\begin{pmatrix} A & B' \\ B & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}. \tag{A.108}$$

In both cases, the infinite dimensional and the finite dimensional setting, we would like to learn about solvability of the linear system at hand. Well-posedness of the abstract problem Eq. (A.106) is proven in the seminal work by Brezzi [28]. Similarly, Nicolaides [96] provides criteria for the well-posedness of generalized saddle point problems Eq. (A.116).

In both cases, the classical Eq. (A.106) and the generalized saddle point problems Eq. (A.116), the famous Ladyzenskaja–Babuška–Brezzi (LBB) inf-sup condition[2] turns out to be the crucial criterion to determine if the problem is well-posedness.

Definition A.7.4 (Ladyzenskaja–Babuška–Brezzi)**.** Let X, Y be Hilbert spaces and let $c : X \times Y \to \mathbb{R}$ a continuous bilinear form then c satisfies the LBB condition if there is a positive

[2]Babuška used inf-sup conditions in [11, 12] to generalize the Lax-Milgram theorem. Brezzi applied these techniques to the abstract saddle point structure in [28]. Ladyzenskaja stated and proved the inf-sup somewhat hidden condition to be true for the Stokes problem [83].

constant $\gamma_c > 0$ such that

$$\sup_{y \in Y} \frac{c(x,y)}{\|y\|_Y} \geq \gamma_c \|x\|_X \quad \forall x \in X \setminus \{0\}, \tag{A.109a}$$

$$\sup_{x \in X} c(x,y) > 0 \quad \forall y \in Y \setminus \{0\}. \tag{A.109b}$$

Theorem A.7.5 (Babuška [12])**.** *Let the assumptions be as in Definition A.7.4 and let $f \in X'$ and c fulfil the LBB condition then the problem*

$$c(x^*, y) = f(y) \quad \forall y \in Y \tag{A.110}$$

can be solved uniquely, and the solution can be bound by

$$\|x^*\|_X \leq \gamma_c^{-1} \|f\|. \tag{A.111}$$

Proof. We follow the reasoning presented in [29]. Let $C \in \mathcal{B}(X, Y')$ denote the associated operator. Condition Eq. (A.109a) implies

$$\|Cx\|_{Y'} = \sup_{y \in Y} \frac{c(x,y)}{\|y\|_Y} \geq \gamma_c \|x\|_X. \tag{A.112}$$

Therefore, any Cauchy sequence $(y_n)_{n \in \mathbb{N}}$ in $\operatorname{ran}(C)$ implies the sequence of preimages $(x_n)_{n \in \mathbb{N}}$ with $Cx_n = y_n$ to be a Cauchy sequence too. Since X is complete there exists a limit $x^* \in X$ and which gives $y = Cx \in \operatorname{ran}(C)$. Therefore, the range of C is closed i.e., $\overline{\operatorname{ran}(C)} = \operatorname{ran}(C)$, and we can apply Banach's closed range theorem.

Additionally, Eq. (A.112) implies $\ker C = \{0\}$ i.e., C is injective. Due to the closed range theorem this is equivalent to $\operatorname{ran}(C') = X$ i.e., C' is surjective.

Condition Eq. (A.109b) implies injectivity of $C' \colon Y \to X'$ and applying the closed range theorem again this gives surjectivity of C and therefore the existence of C^{-1}.

Using Eq. (A.112) again we obtain

$$\|C^{-1}y\|_X \leq \gamma_c \|CC^{-1}y\|_{Y'} = \|y\|_{Y'} \tag{A.113}$$

for every $y \in Y'$. The solution to Eq. (A.110) now is given by $x^* = C^{-1}f$. □

Remark A.7.6. The proof of Theorem A.7.5 allows us to highlight the reasons for the two parts of Eq. (A.109). The first i.e., Eq. (A.109a) provides injectivity and the bound on the inverse. The second i.e., Eq. (A.109b) surjectivity.

Instead of Eq. (A.109) one can equivalently ask for two $\inf - \sup$ conditions, but often it is not necessary to have the second constant available.

Remark A.7.7. The converse of Theorem A.7.5 is also true. Let $C\colon X \to Y'$ be a linear isomorphism with

$$\|C^{-1}y'\|_X \leq \gamma \|y'\|_{Y'} \tag{A.114}$$

for every $y' \in Y'$. Therefore, we also have

$$\|x\|_X = \|C^{-1}Cx\|_X \leq \gamma \|Cx\|_{Y'} \tag{A.115}$$

for every $x' \in X'$, which in turn implies Eq. (A.109a). Now we apply the closed range theorem and the surjectivity of C gives the injectivity of C', hence Eq. (A.109b).

Definition A.7.8 (Generalized saddle point problem)**.** Let $i \in \{1, 2\}$ and let X_i, Y_i be Hilbert spaces and $f \in X_1'$ as well as $g \in Y_2'$. Furthermore, let $a\colon X_2 \times X_1 \to \mathbb{R}$, $b_1\colon X_1 \times Y_1 \to \mathbb{R}$ and $b_2\colon X_2 \times Y_2 \to \mathbb{R}$ bounded bilinear forms. $(x^*, y^*) \in (X_2, Y_1)$ is a solution to the generalized saddle point problem if it satisfies

$$a(x^*, x) + b_1(x, p) = f(x) \quad \forall x \in X_1, \tag{A.116a}$$

$$b_2(x^*, y) = g(y) \quad \forall y \in Y_2. \tag{A.116b}$$

The associated null spaces are

$$\kappa_i \coloneqq \{x \in X_i\colon b_i(x, y) = 0,\ \forall y \in Y_i\} \tag{A.117}$$

for $i \in \{1, 2\}$.

Theorem A.7.9 (Well-posedness [28, 96, 21])**.** *Let a,b_1,b_2,f and g be given as in Definition A.7.8. If b_1 and b_2 satisfy Eq.* (A.109a) *and $a|_{\kappa_2 \times \kappa_1}$ satisfy Eq.* (A.109)*, then the generalized saddle point problem Eq.* (A.116) *has unique solution $(x^*; y^*) \in X_2 \times Y_1$. This solution is bounded by*

$$\|x\|_{X_2} \leq \frac{1}{\gamma_a}\|f\|_{X_1'} + \frac{c}{\gamma_{b_2}}\|g\|_{Y_2'}, \tag{A.118}$$

$$\|y\|_{Y_1} \leq \frac{c}{\gamma_{b_1}}\|f\|_{X_1'} + \frac{c\|a\|}{\gamma_{b_2}\gamma_{b_1}}\|g\|_{Y_2'}, \tag{A.119}$$

where $c = (\|a\|/\gamma_a + 1)$.

The framework partly introduced above, proved to be useful [17] and especially in the context of partial differential equations related to the Stokes equation it forms the foundation for the analytical and numerical treatment [62].

B Finite Element Method

Definition B.0.1 (Finite Elements in the sense of Ciarlet)**.** Besides a compact set with Lipschitz boundary $K \subset \mathbb{R}^n$, a finite element as defined by Ciarlet [36] contains additionally a function space P containing elements $p : K \to \mathbb{R}^m$ and the local degrees of freedom Σ a basis of $\mathcal{B}(P, \mathbb{R})$. For each finite element one can obtain the so-called shape functions Θ determined by $\sigma_i(\theta_j) = \delta_{ij}$. Given the shape functions Θ as well as Σ we are able to map them via T_k to an arbitrary grid cell. For every $C \in \mathcal{C}$, $\theta \in \Theta$ and $\sigma \in \Sigma$ we therefore define

$$\sigma_{C_k} : f \mapsto \sigma(f \circ T_k) \tag{B.1}$$

$$\theta_{C_k} := \theta \circ T_k^{-1}. \tag{B.2}$$

Remark B.0.2. For arbitrary meshes the transformed shape function is not always a polynomial again.[54]

The following useful and classical lemma is presented as in [54, Lemma B.67].

Lemma B.0.3 (Deny-Lions)**.** *Let Ω be a Lipschitz domain. Let $l \geq 0$ and $1 \leq p \leq \infty$, then there exists $c > 0$ such that*

$$\inf_{q \in \mathbb{P}_l} \|v + q\|_{W^{l+1,p}(\Omega)} \leq c|v|_{W^{l+1,p}(\Omega)}. \tag{B.3}$$

Lemma B.0.4. *Let $\mathcal{T}_{h_n})_{n\in\mathbb{N}}$ be a family of shape regular quasi uniform grids with $h_n > 0$ for every $n \in \mathbb{N}$. Then there is $c > 0$ independent of h_n satisfying*

$$\|\operatorname{grad} q\|_0 \leq c \frac{1}{h_n} \|q\|_0 \tag{B.4}$$

for every $q \in \mathcal{W}_h^1$.

Zusammenfassung

Diese Arbeit untersucht die Stabilitäts und Approximationseigenschaften der Anwendung des Zell-Knoten Finite Volumen Verfahrens auf eine elliptische partielle Differentialgleichung zweiter Ordnung welche auf Gittern diskretisiert wird, welche aus bilinearen Bildern von Quadraten oder trilinearen Bildern von Würfel bestehen. Diese Gleichung entsteht aus der Semi-Diskretisierung des Projektionsschrittes einer semi-impliziten Finiten Volumen Methode zweiter Ordnung, welche sowohl das pseudo-inkompressible als auch das kompressible Regime der Euler Gleichungen auflösen kann.

Infolgedessen wird das gemischte Sattelpunktsproblem untersucht, das durch die pseudo inkompressible Divergenzbedingung und die für die kompressiblen Effekte verantwortlichen Quellterme bestimmt wird untersucht. Im pseudo-inkompressiblen Fall wird Stabilität und eine a-priori Fehlerabschätzung für den Projektionsschritt bewiesen, im kompressiblen Fall zumindest Stabiltät. Zu diesem Zweck wird eine Interpretation der diskreten Flussvariablen im Sinne einer unstetige Galerkin Methode und führen einen Raviart-Thomas Interpolationsoperator auf den dualen Kontrollvolumina, welche die Knoten des primären Gitters umschließen, eingeführt. Hierbei wird die natürliche Divergenz über das Integral des Flusses durch den Rand der dualen Kontrollvolumina definiert.