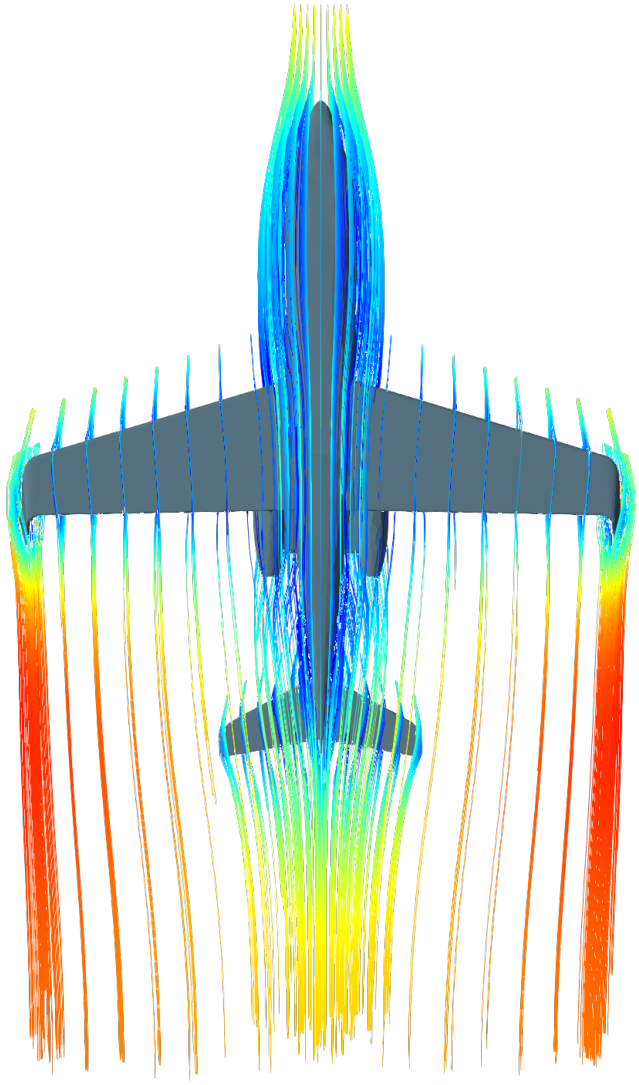
Finite Element Methods for Fluid Dynamics

Philip L. Lederer



Vienna, last update: 17. Mai 2021

Abstract

This is a preliminary version of the lecture notes for the course *Finite Element Methods in Computational Fluid Dynamics* and will be updated regularly. The notes are primarily based on:

- Lecture notes on *Numerical Methods for PDEs* (J. Schöberl, TU Wien)
- Lecture notes on *Special Topics in the Finite Element Method* (R. Stenberg, Aalto University)
- Boock Chapter: *Finite Element Methods for the Incompressible Navier-Stokes Equations* (R. Rannacher, Springer)
- Book: *Finite Element Methods for Flow Problems* (J. Donea and A. Huerta, Wiley)
- Book: *Finite Element Methods for Incompressible Flow Problems* (V. John, Springer)

Contents

1 The equations of fluid motion

This chapter is devoted to the basic principles of fluid mechanics and the derivation of the governing equations. We follow the same ideas as provided in standard literature on fluid dynamics, see [38, 5, 42, 2].

In the following we consider an Euclidean space with the independent three-dimensional variable $x = (x_1, x_2, x_3)$ and assume that the time t proceeds independently. Using the unit vectors e_1, e_2 and e_3 along the x_1, x_2 and x_3 axes, respectively, we define the vector velocity field by

$$u := u_1 e_1 + u_2 e_2 + u_3 e_3,$$

with the scalar-valued components $u_1 = u_1(x_1, x_2, x_3, t)$, $u_2 = u_2(x_1, x_2, x_3, t)$ and $u_3 = u_3(x_1, x_2, x_3, t)$. Similarly, the scalar density field and the scalar pressure is given by $\rho := \rho(x_1, x_2, x_3, t)$ and $p := p(x_1, x_2, x_3, t)$. We speak of a two-dimensional flow field, when the fluid motion is restricted to parallel planes. In this case the the velocity component, which is perpendicular to the plane is equal to zero at each point. Further, the flow is independent of deformations that are parallel to the flow. In this work a two dimensional flow is always considered in the x_1-x_2 plane, thus the velocity field is given by $u := u_1 e_1 + u_2 e_2$. Note that in order to speak of the above defined physical quantities we assumed that the *continuum assumption* holds true. This means that the physical quantities of interest of the liquid contained in a given small volume are imagined to be uniformly distributed over that volume. We can then also talk about fluid particles at a specific point, when we keep in mind that this particle is actually sufficiently large to contain enough molecules of the liquid such that an averaging, for example of the velocity, makes sense.

For the derivation of the governing equations of fluid mechanics we are using the concept of (finite) control volumes and their associated control surfaces. The main purpose of using a control volume is to focus the attention on physical events and quantities only in a small region and its boundary in order to be able to keep track of all effects. We can distinguish between two different types. A fixed control volume is specified by a given (fixed) location in space, thus the fluid passes into and out off the volume through the surface. The second type is called a material control volume. The idea is that the control volume is

moving with the liquid such that the fluid particles stay inside and do not pass the surface. This leads to two different aspects. A *Lagrangian viewpoint* focuses on the flow of fluid particles. Each particle is identified by its initial position at a specific given (start) time. When time passes all particles move and change their position. This position (trajectory) now is a function that depends on the original location and the time. Similarly, all other physical quantities only depend on the initial position and time, thus refer to one specific fluid particle. In contrast to this, the *Eulerian viewpoint* deals with fixed points in space. At a given time we can evaluate physical quantities at each point to retrieve local information on the fluid. In this work we always use the Eulerian viewpoint. The close relation of the two different viewpoints is given by the substantial derivative

$$\frac{\mathrm{D}}{\mathrm{D}t} := \frac{\partial}{\partial t} + (u \cdot \nabla), \tag{1.1}$$

which can be interpreted as the time rate of change following a fluid particle. It consists of the local time derivative at a fixed point $\partial/\partial t$ and the convective derivative $(u \cdot \nabla)$, which describes the time rate of change induced by the movement of the particle. Using the substantial derivative, also often called material derivative, we can also present the well-known *Reynolds transport theorem* which gives the relation of the time derivative of an integral over an materical control volume $\omega(t)$

$$\frac{d}{dt}\int_{\omega(t)} f(x,t)\,\mathrm{d}x = \int_{\omega\equiv\omega(t)} \frac{\partial f(x,t)}{\partial t}\,\mathrm{d}x + \int_{\partial\omega\equiv\partial\omega(t)} f(x,t)v\cdot n\,\mathrm{d}s, \tag{1.2}$$

where the $f(x,t)$ is a smooth function (we explicitly included the dependency on x and t here to make things more readable). Note, that the integrals on the right side are considered on the fixed domain ω which consides with the moving control volume $\omega(t)$ at the considered instant, t, in time.

1.1 Fundamental laws

1.1.1 The continuity equation

The fundamental physical principle that we consider in the following is the conservation of mass. To this end, let ω be an arbitrary fixed control volume, hence we assume that it is not moving with the flow. The principle of mass conservation then reads as

$$\text{Mass flow through the surface } \partial\omega = \text{time rate of decrease of mass inside } \omega \tag{1.3}$$

In the following we translate (1.3) into an explicit equation including functions and variables. We first deal with the left hand side of this equation. The mass that is transported through an infinitesimal small surface area is given by the density times the size of this area times the velocity that is perpendicular to the surface. Thus, we have, using the Gaussian theorem,

$$\text{Netto mass flow through the surface } \partial\omega := \int_{\partial\omega} \rho u \cdot n \, \mathrm{d}s = \int_{\omega} \operatorname{div}(\rho u) \, \mathrm{d}x \, .$$

The right hand side of (1.3) is given by the negative derivation with respect to time of the mass inside of ω, thus

$$\text{time rate of decrease of mass inside } \omega := -\frac{\partial}{\partial t} \int_{\omega} \rho \, \mathrm{d}x \, .$$

Note that the control volume is fixed in time, allowing us to change the order of integration and differentiation. Combining the last two results then leads to

$$\int_{\omega} \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho u) \, \mathrm{d}x = 0.$$

Taking into account that the control volume ω was arbitrary, the equation inside the integral has to be fulfilled at each point and so we finally derive the continuity equation given by

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho u) = 0. \tag{1.4}$$

This means that the time rate of change at a specific point equals the negative netto flow of the mass out of an infinitesimal small volume area (a fluid particle).

Note, that the continuity equation in integral form can also be derived by simply using the Reynolds transport theorem where $f = \rho$. The principle of conservations of mass on a time dependent domain $\omega(t)$ then simply states that

$$0 = \frac{d}{dt} \int_{\omega(t)} \rho \, \mathrm{d}x,$$

hence with (1.2), we also get

$$0 = \frac{d}{dt} \int_{\omega(t)} \rho \, \mathrm{d}x = \int_{\omega \equiv \omega(t)} \frac{\partial \rho}{\partial t} \, \mathrm{d}x + \int_{\partial\omega \equiv \partial\omega(t)} \rho \cdot n \, \mathrm{d}s = \int_{\omega \equiv \omega(t)} \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho u) \, \mathrm{d}x \, .$$

1.1.2 The momentum equation

The momentum equation, which is based on Newton's second law, relates the time rate of change of the momentum of a particle to the force acting on it. For the derivation we choose a material control volume $\omega(t)$, which is moving with the flow. Then we have

$$\text{time rate of change of momentum of } \omega(t) = \text{netto forces acting on } \omega(t). \tag{1.5}$$

For the computation of the momentum we first focus on the physical effects in the x_1-direction. The product ρu_1 is equivalent to the momentum in the direction of e_1 per unit volume,

$$\text{time rate of change of momentum in } x_1\text{-direction of } \omega(t) = \frac{d}{dt}\int_{\omega(t)} \rho u_1 \,\mathrm{d}x\,.$$

Using Reynolds transport theorem and the Gaussian theorem on the appearing surface integral, we can further write

$$\frac{d}{dt}\int_{\omega(t)} \rho u_1 \,\mathrm{d}x = \int_{\omega(t)} \frac{\partial}{\partial t}(\rho u_1)\,\mathrm{d}x + \int_{\partial\omega(t)} (\rho u_1) u \cdot n \,\mathrm{d}s = \int_{\omega(t)} \frac{\partial}{\partial t}(\rho u_1) + \operatorname{div}(\rho u_1 u)\,\mathrm{d}x,$$

hence with the matrix $[\rho u \otimes u]_{ij} = \rho u_i u_j$ and applying the same steps for the other spatial directions we get in total

$$\frac{d}{dt}\int_{\omega(t)} \rho u \,\mathrm{d}x = \int_{\omega(t)} \frac{\partial}{\partial t}(\rho u) + \operatorname{div}(\rho u \otimes u)\,\mathrm{d}x\,.$$

For the right hand side of (1.5) we first consider a volume force f and a surface force s. Thus, again restricting on the x_1-direction, we have

$$\text{Forces in } x_1\text{-direction acting on } \omega(t) = \int_{\omega(t)} \rho f_1 \,\mathrm{d}x + \int_{\partial\omega(t)} s_1 \,\mathrm{d}s\,.$$

Note that there is no density included for the boundary forces as the infinitesimal small areas contain no mass. Thus, with $f = (f_1, f_2, f_3)$ and $s = (s_1, s_2, s_3)$, in total we have

$$\int_{\omega(t)} \frac{\partial}{\partial t}(\rho u) + \operatorname{div}(\rho u \otimes u)\,\mathrm{d}x = \int_{\omega(t)} \rho f \,\mathrm{d}x + \int_{\partial\omega(t)} s \,\mathrm{d}s\,. \tag{1.6}$$

Following for example [38, chapter 5.4], one relates the appearing forces on the boundary with the Cauchy stress tensor σ such that $s = \sigma n$. Applying the Gaussian theorem for the

right integral on the left side, equation (1.6) can be written as

$$\int_{\omega(t)} \frac{\partial}{\partial t}(\rho u) + \operatorname{div}(\rho u \otimes u)\,\mathrm{d}x = \int_{\omega(t)} \rho f\,\mathrm{d}x + \int_{\omega(t)} \operatorname{div}(\sigma)\,\mathrm{d}x\,.$$

and since the control volume was arbitrary we get the differential form the momentum equation

$$\frac{\partial}{\partial t}(\rho u) + \operatorname{div}(\rho u \otimes u - \sigma) = \rho f. \tag{1.7}$$

1.1.3 The energy equation

Again we choose a material control volume $\omega(t)$ and consider the energy balance of the fluid. Let E denote the total energy per unit mass and let e be the inner energy per unit mass, i.e. we have the relation $E = e + 1/2u^2$. The first law of thermodynamics now states, that the temporal change of the total energy is balanced by the work produced by the fluid and external forces and the flow of heat across the boundary. First, as before, the Reynolds transport theorem allows to reformulate the temporal variation of the total energy in $\omega(t)$ as

$$\frac{d}{dt}\int_{\omega(t)} \rho E\,\mathrm{d}x = \int_{\omega(t)} \frac{\partial(\rho E)}{\partial t} + \operatorname{div}(\rho E u)\,\mathrm{d}x\,.$$

For a given volume function f (see section above), the work produced in the interior and on the surface is given by the integrals

$$\int_{\omega(t)} \rho f \cdot u\,\mathrm{d}x, \quad \text{and} \quad \int_{\partial\omega(t)} (\sigma u)\cdot n\,\mathrm{d}s = \int_{\omega(t)} \operatorname{div}(\sigma u)\,\mathrm{d}x\,.$$

Next, let Φ be a given function that describes the changes of the internal energy. Then, the heat flow across the boundary, can be written as

$$\int_{\partial\omega(t)} \Phi \cdot n\,\mathrm{d}x = \int_{\omega(t)} \operatorname{div}(\Phi)\,\mathrm{d}x\,.$$

The corresponding constitutive law for Φ will be given in the next section. Hence, in total we get

$$\int_{\omega(t)} \frac{\partial(\rho E)}{\partial t} + \operatorname{div}(\rho E u)\,\mathrm{d}x = \int_{\omega(t)} \rho f\cdot u\,\mathrm{d}x + \int_{\omega(t)} \operatorname{div}(\sigma u)\,\mathrm{d}x + \int_{\partial\omega(t)} \operatorname{div}(\Phi)\,\mathrm{d}x, \tag{1.8}$$

of in differential form

$$\frac{\partial(\rho E)}{\partial t} + \operatorname{div}(\rho E u) = \rho f \cdot u + \operatorname{div}(\sigma u) + \operatorname{div}(\Phi).$$

1.1.4 Constitutive laws and equation of state

The above derived equations for the conservation of mass, momentum and energy must be closed by several constitutive laws. The first equation is call Newton's viscosity law (hence we assume a Newtonian fluid) and is given by the following conditions:

1. The stress tensor σ depends only on the gradient of the velocity ∇u. Further, this dependence is linear.
2. The stress tensor σ is symmetric (conservation of angular momentum).
3. In the absence of internal friction (inviscid flows), the stress tensor σ is diagonal and proportional to the pressure (this shows that the boundary forces in the momentum equation are only applied in normal direction).

Above assumptions give the relation

$$\sigma = 2\mu\varepsilon(u) + \lambda \operatorname{div}(u)\mathrm{I} - p\mathrm{I}, \quad \text{with} \quad \varepsilon(u) = \frac{1}{2}(\nabla u + \nabla u^T).$$

Here μ is called the dynamic viscosity and λ the volume viscosity. These two coefficients are related by the definition of the bulk viscosity $\mu_B = \lambda + 2/3\mu$, which in general is negligible (Stokes hypothesis) except in the study of the structure of (for example) shock waves. In this work we will always consider the case $\mu_B = 0$. At several points the stress tensor might also be written in the more compact form $\sigma = \tau - p\mathrm{I}$ with the viscous stress tensor

$$\tau = \mu(2\varepsilon(u) - \frac{2}{3}\operatorname{div}(u)\mathrm{I}).$$

Next, we apply Fourier's law that states, that Φ is proportional to the variations of the internal energy, i.e. we have

$$\Phi = k\nabla T,$$

where k is the coefficient of thermal conductivity.

Finally, to close the system of equations, it is necessary to present an equation of state, thus give a relation between the thermodynamic variables ρ, p, T and the energy e. In the

case of a perfect gas we have the well known equation $p = \rho RT$, where R is the gas constant per unit mass. In this work we will assume that the gas (or fluid) is given as a calorically perfect gas, i.e. we assume that the specific heat at constant volume c_v, and the specific heat at constant pressure c_p, are constant. With the ratio of the specific heats given by γ we then have the relations

$$e = c_v T, \quad \gamma = \frac{c_p}{c_v}, \quad c_v = \frac{R}{\gamma - 1}, \quad c_p = \frac{\gamma R}{\gamma - 1},$$

thus, the equation of state can also be written as

$$p = (\gamma - 1)\rho e \quad \text{and} \quad T = \frac{(\gamma - 1)e}{R},$$

or with the total energy also

$$E = \frac{p}{\rho(\gamma - 1)} + \frac{1}{2}u^2.$$

1.1.5 The compressible Navier-Stokes equations

When we gather all the above equations and close them with the constitutive laws and the equation of state, we obtain the *compressible Navier-Stokes equations* given by

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho u) = 0, \tag{1.9a}$$

$$\frac{\partial \rho u}{\partial t} + \operatorname{div}(\rho u \otimes u) - \operatorname{div}(\mu(2\varepsilon(u) - \frac{2}{3}\operatorname{div}(u)\mathrm{I})) + \nabla p = \rho f, \tag{1.9b}$$

$$\frac{\partial (\rho E)}{\partial t} + \operatorname{div}(\rho E u) - \operatorname{div}(\mu(2\varepsilon(u) - \frac{2}{3}\operatorname{div}(u)\mathrm{I})u) + \operatorname{div}(pu) - \operatorname{div}(k\nabla T) = \rho f \cdot u, \tag{1.9c}$$

with

$$p = p(\rho, T), \quad \text{and} \quad E = \frac{p}{\rho(\gamma - 1)} + \frac{1}{2}u^2.$$

1.1.6 The Euler equations

In the inviscid case, thus in the limit of vanishing viscosity $\lambda = \mu = 0$, the compressible Navier-Stokes equation reduce to the so called *Euler equations* given by

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho u) = 0, \tag{1.10a}$$

$$\frac{\partial \rho u}{\partial t} + \operatorname{div}(\rho u \otimes u) + \nabla p = \rho f, \tag{1.10b}$$

$$\frac{\partial (\rho E)}{\partial t} + \operatorname{div}((\rho E + p)u) - \operatorname{div}(k \nabla T) = \rho f \cdot u, \tag{1.10c}$$

with

$$p = p(\rho, T), \quad \text{and} \quad E = \frac{p}{\rho(\gamma - 1)} + \frac{1}{2}u^2.$$

1.1.7 The incompressible Navier-Stokes and Stokes equations

In the following we derive several sets of equations that consider the incompressible case, thus we assume a constant density in space and time. To this end we define the kinematic viscosity by $\nu = \mu/\rho$ and replace the pressure p by the scaled pressure p/ρ. For simplicity we will still use the notation p for the pressure. Note, that the conservation of mass now simplifies to $\operatorname{div}(u) = 0$, and thus we also have the simplified relation of the viscous stress tensor

$$\operatorname{div}(u) = 0 \quad \Rightarrow \quad \tau = 2\mu\varepsilon(u).$$

Finally, since the equation of the conservation of energy now decouples from the other equations we get the *instationary incompressible Navier-Stokes* equations given by

$$\operatorname{div}(u) = 0, \tag{1.11a}$$

$$\frac{\partial u}{\partial t} + \operatorname{div}(u \otimes u) - 2\nu \operatorname{div}(\varepsilon(u)) + \nabla p = f. \tag{1.11b}$$

In several textbooks, this set of equations is often further simplified by using the identity

$$2\nu \operatorname{div}(\varepsilon(\mathrm{u})) = \nu \left(\Delta u + \nabla \operatorname{div}(u)\right) = \nu \Delta u,$$

which then gives

$$\operatorname{div}(u) = 0, \tag{1.12a}$$

$$\frac{\partial u}{\partial t} + \operatorname{div}(u \otimes u) - \nu \Delta(u) + \nabla p = f. \tag{1.12b}$$

Note however, that the above identity assumes a smooth enough (regular) velocity solution such that the order of differentiation can be changed. Thus, in the context of variational formulations and their discretization, one has to be very careful and maybe needs to deal with the more challenging setting where we consider the symmetric gradient $\varepsilon(u)$.

As a next step of simplification we now consider the case of a stationary flow, i.e. we consider a flow that does not change in time. Then we get the *stationary incompressible Navier Stokes equations* given by

$$\operatorname{div}(u) = 0, \tag{1.13a}$$

$$\operatorname{div}(u \otimes u) - 2\nu \operatorname{div}(\varepsilon(u)) + \nabla p = f. \tag{1.13b}$$

In order to derive the last simplification we first introduce an important characteristic quantity of fluid dynamics called the Reynolds number given by

$$\mathrm{Re} := \frac{UL}{\nu}, \tag{1.14}$$

where U and L are characteristic length and velocity scales. The Reynolds number is important as it can be interpreted as the ratio between inertia and viscous forces. If we fix the reference variables U and L to be for example $\mathcal{O}(1)$, then a high Reynolds number corresponds to a very small viscosity, i.e. the friction between fluid particles is small and the acceleration initiated by inertia forces dominates. However, in a flow characterized by a small Reynolds number, the viscous effects are crucial. Such flows are often called creeping flows and are of practical importance. This has a great impact on the governing equations of fluid motion. Using a dimension analysis for the case when $\mathrm{Re} \to 0$ shows that the nonlinear term in (1.13) vanishes, thus $\operatorname{div}(u \otimes u) \to 0$. The resulting set of partial differential equations is called the *Stokes equations* given by

$$\begin{aligned} -2\nu \operatorname{div}(\varepsilon(u)) + \nabla p &= f, \\ \operatorname{div}(u) &= 0. \end{aligned} \tag{1.15}$$

These equations are of great interest as they fit into the mathematical concept of a saddle point problem. Although the full nonlinear setting of the incompressible Navier-Stokes

equations is generally applied, a proper (numerical) treatment of (1.15) is essential since for example a lot of solving routines for the nonlinear system are based on iterations relying on the solution of (1.15).

1.1.8 Boundary and initial conditions

In order to solve the systems of partial differential equations introduced above, we might need suitable boundary and initial conditions. In particular, equations (1.11), (1.12), (1.9) and (1.10) demand an initial condition for the velocity u and the temperature T. Further, the last two demand also an initial condition for the density ρ. Since the energy equation is decoupled in the incompressible case, the initial condition for the temperature might be neglected (if one is not interested in the evolution of the temperature).

Beside the Euler equations, all other sets of equations (1.11), (1.12) and the stationary cases (1.13) and (2.6), further include a second order differential operator acting on the velocity u which allows (and demands) to prescribe boundary conditions. In a first step we consider the case where the fluid comes in contact with a wall. Since no velocity is going to pass through the wall in normal direction, we impose the condition

$$u \cdot n = u_{\mathsf{w}} \cdot n, \tag{1.16}$$

where u_{w} is the prescribed velocity of the wall. Note that in the unsteady case the boundary velocity might also depend on the time. This condition only acts on the normal component of the velocity, but has no impact on the tangential velocity. This is mainly due to the different physical effects that appear close to the wall. In history, there are several different approaches on how to deal with the tangential components of the velocity. In this work we mainly discuss the case of the so called *no-slip* condition that is commonly accepted. The idea is that the viscous effects close to the wall create a force that adhere the fluid particles and the wall together which, similar to the normal component, reads as

$$u - (u \cdot n)n = u - (u_{\mathsf{w}} \cdot n)n. \tag{1.17}$$

For a detailed discussion we refer to [38, chapter 6.4]. These two conditions together are called Dirichlet conditions.

The second type is called a Neumann boundary condition and induces a certain value for the stress tensor σ on (a part of) the boundary. Similar as in the derivation of the conservation of momentum, we can only prescribe the forces in normal direction, i.e.

$$\sigma n = (\tau - \mathrm{id}p)\, n = g, \tag{1.18}$$

with an given (vector valued) force g. An example of a Neumann condition is given by a flow through a pipe where you (want to) impose no forces ($g = 0$) on the outlet. This is also often call a *do nothing* boundary condition. For more details we refer to [21].

Finally we also want to mention the more general (Robin type) boundary conditions given by

$$\gamma_n u \cdot n + (1 - \gamma_n) n^{\mathrm{T}} \sigma n = g_n, \tag{1.19}$$

$$\gamma_t (u - (u \cdot n)n) + (1 - \gamma_t)(\sigma n - (n^{\mathrm{T}} \sigma n)n) = g_t, \tag{1.20}$$

with some given functions g_n and g_t and some fixed values $\gamma_n, \gamma_t \in [0, 1]$. The case $\gamma_n = \gamma_t = 1$ corresponds to the above discussed *no-slip* case, whereas the case $\gamma_n = 1, \gamma_t = 0$ corresponds to so called *slip conditions*.

Beside the boundary conditions for the velocity, the compressible Navier-Stokes equations (1.11) also allow to prescribe a boundary conditions for the temperature. Similarly as before, one can define Dirichlet, Neumann or Robin type boundary conditions.

2 The Stokes equations - Theory of mixed finite elements

2.1 Basic notation and fundamentals

In the following, we introduce the notation and establish properties of certain Sobolev spaces that we use throughout this work. For a more detailed discussion on this topic we refer to [1, 35, 8] and [16]. First, we introduce the notation $A \sim B$ to indicate that there exists constants $c, C > 0$ independent of the mesh size h (as defined later) and other problem parameters like the viscosity ν such that $cA \leq B \leq CA$. We also use $A \lesssim B$ when there exists a $C > 0$ independent of h and ν such that $A \leq CB$. In a similar manner we also define the symbol $\gtrsim$.

For the rest of the work let $\Omega \subset \mathbb{R}^d$, $d = 2$ or 3, be an open bounded subset such that the boundary $\Gamma := \partial\Omega$ is smooth, i.e. $\Gamma \in C^\infty$,

Let $\mathcal{C}^k(\Omega, \mathbb{R})$ be the function space consisting of real-valued k-times continuously differentiable functions on Ω. Then we define $\mathcal{D}(\Omega, \mathbb{R}) := \mathcal{C}_0^\infty(\Omega, \mathbb{R})$ as the set of infinitely differentiable, compactly supported, real-valued functions on Ω and denote by $\mathcal{D}'(\Omega)$ the space of distributions. To inidicate vector and matrix-valued functions we include the range in the notation, thus $\mathcal{D}(\Omega, \mathbb{R}^d) := \{\phi : \Omega \to \mathbb{R}^d \text{ with } \phi_i \in \mathcal{D}(\Omega, \mathbb{R})\}$ and $\mathcal{D}(\Omega, \mathbb{R}^{d\times d}) := \{\phi : \Omega \to \mathbb{R}^{d\times d} \text{ with } \phi_{ij} \in \mathcal{D}(\Omega, \mathbb{R})\}$ indicate vector and matrix-valued infinitely differentiable, compactly supported, real-valued functions, respectively. This notation is extended to other functions spaces as needed. Whereas

$$L^2(\Omega, \mathbb{R}) := \{f : \int_\Omega |f|^2 \,\mathrm{d}x < \infty\} \tag{2.1}$$

denotes the space of square integrable functions with the inner product and the norm

$$(f, g)_{L^2(\Omega)} := \int_\Omega fg \,\mathrm{d}x, \qquad \|f\|^2_{L^2(\Omega)} := (f, f)_{L^2(\Omega)}, \qquad \forall f, g \in L^2(\Omega), \tag{2.2}$$

the spaces $L^2(\Omega, \mathbb{R}^d)$ and $L^2(\Omega, \mathbb{R}^{d\times d})$ denote its vector and matrix-valued versions. At several points in the later chapters we make use of the local L^2-norm defined on subsets

$\omega \subset \Omega$. For a better readability we introduce the following notation

$$\|\cdot\|_\omega := \|\cdot\|_{L^2(\omega)}.$$

Certain differential operators have different definitions depending on the context. We define the "curl" operator by

$$\begin{aligned}
&\operatorname{curl}(\phi) = (-\partial_2\phi, \partial_1\phi)^{\mathrm{T}}, && \text{for } \phi \in \mathcal{D}'(\Omega,\mathbb{R}) \text{ and } d = 2,\\
&\operatorname{curl}(\phi) = -\partial_2\phi_1 + \partial_1\phi_2, && \text{for } \phi \in \mathcal{D}'(\Omega,\mathbb{R}^2) \text{ and } d = 2,\\
&\operatorname{curl}(\phi) = (\partial_2\phi_3 - \partial_3\phi_2, \partial_3\phi_1 - \partial_1\phi_3, \partial_1\phi_2 - \partial_2\phi_1)^{\mathrm{T}} && \text{for } \phi \in \mathcal{D}'(\Omega,\mathbb{R}^3) \text{ and } d = 3,
\end{aligned}$$

where $(\cdot)^{\mathrm{T}}$ denotes the transpose and ∂_i abbreviates ∂/∂_i. Similarly, $\nabla\phi$ has different meanings depending on the context and results either in a vector $[\nabla\phi]_i = \partial_i\phi$ for $\phi \in \mathcal{D}'(\Omega,\mathbb{R})$ or in a matrix $[\nabla\phi]_{ij} = \partial_i\phi_j$ for $\phi \in \mathcal{D}'(\Omega,\mathbb{R}^d)$. Finally, we denote by $\operatorname{div}(\phi) = \sum_{i=1}^3 \partial_i\phi_i$ the standard divergence operator for $\phi \in \mathcal{D}'(\Omega,\mathbb{R}^d)$ and by $[\operatorname{div}(\phi)]_j = \sum_{i=1}^3 \partial_i\phi_{ji}$ the vector-valued divergence operator applied to $\phi \in \mathcal{D}'(\Omega,\mathbb{R}^{d\times d})$.

Let $\tilde{d} := d(d-1)/2$ (such that $\tilde{d} = 1$ and $\tilde{d} = 3$ for $d = 2$ and $d = 3$, respectively). The standard Sobolev spaces are denoted by

$$\begin{aligned}
H^1(\Omega,\mathbb{R}) &:= \{u \in L^2(\Omega,\mathbb{R}) : \|\nabla u\|_{L^2(\Omega)} < \infty\},\\
H^1(\Omega,\mathbb{R}^d) &:= \{u \in L^2(\Omega,\mathbb{R}^d) : \|\nabla u\|_{L^2(\Omega)} < \infty\},\\
H(\operatorname{div},\Omega) &:= \{u \in L^2(\Omega,\mathbb{R}^d) : \|\operatorname{div}(u)\|_{L^2(\Omega)} < \infty\},\\
H(\operatorname{curl},\Omega) &:= \{u \in L^2(\Omega,\mathbb{R}^d) : \|\operatorname{curl}(u)\|_{L^2(\Omega)} < \infty\},
\end{aligned}$$

with the associated norms given by $\|\cdot\|_{H^1(\Omega)}$, $\|\cdot\|_{H(\operatorname{div},\Omega)}$ and $\|\cdot\|_{H(\operatorname{curl},\Omega)}$, respectively. Note that we will not distinguish between the dimension of the ordinary Sobolev space in the definition of the norm, thus we use $\|\cdot\|_{H^1(\Omega)}$ as the symbol for the norm on $H^1(\Omega,\mathbb{R})$ and $H^1(\Omega,\mathbb{R}^d)$. In the same fashion we also denote the seminorms by $|\cdot|_{H^1(\Omega)}$, $|\cdot|_{H(\operatorname{div},\Omega)}$ and $|\cdot|_{H(\operatorname{curl},\Omega)}$. Sobolev spaces with higher regularity are similarly given by

$$\begin{aligned}
H^m(\Omega,\mathbb{R}) &:= \{u \in L^2(\Omega,\mathbb{R}) : \|\nabla^m u\|_{L^2(\Omega)} < \infty\},\\
H^m(\Omega,\mathbb{R}^d) &:= \{u \in L^2(\Omega,\mathbb{R}^d) : \|\nabla^m u\|_{L^2(\Omega)} < \infty\},\\
H^m(\operatorname{div},\Omega) &:= \{u \in H^m(\Omega,\mathbb{R}^d) : \|\operatorname{div}(u)\|_{L^2(\Omega)} < \infty\},\\
H^m(\operatorname{curl},\Omega) &:= \{u \in H^m(\Omega,\mathbb{R}^d) : \|\operatorname{curl}(u)\|_{L^2(\Omega)} < \infty\},
\end{aligned}$$

and we use the notation $\|\cdot\|_{H^m(\Omega)}$, $\|\cdot\|_{H^m(\operatorname{div},\Omega)}$ and $\|\cdot\|_{H^m(\operatorname{curl},\Omega)}$ for the corresponding norms. Note that the Sobolev spaces above can also be defined as the closure of $\mathcal{C}^\infty(\overline{\Omega},\cdot)$

(for sufficiently smooth boundaries) with the according norms, see for example in [23] for spaces with more regularity and for the standard spaces [18, 20, 16]. The equivalence of those definitions is not trivial and goes back to the famous theorem of N. Meyers and J. Serrin, see [34]. A detailed proof can also be found in the book [14, 1].

We continue with the definition of appropriate Sobolev spaces on the boundary. Using the notations from above the space of square integrable functions on the boundary Γ is denoted by $L^2(\Gamma, \mathbb{R})$. Now let n denote the outward unit normal on Ω, then we introduce the following trace operators for smooth functions

$$\begin{aligned}
\gamma\phi &:= \phi|_\Gamma && \forall \phi \in C^1(\overline{\Omega}, \mathbb{R}), && \gamma_n\phi := \phi|_\Gamma \cdot n && \forall \phi \in C^1(\overline{\Omega}, \mathbb{R}^d),\\
\gamma_t\phi &:= \phi|_\Gamma \times n && \forall \phi \in C^1(\overline{\Omega}, \mathbb{R}^d), && \pi_t\phi := (\phi|_\Gamma - (\phi|_\Gamma \cdot n)n) && \forall \phi \in C^1(\overline{\Omega}, \mathbb{R}^d),\\
\gamma_{nn}\phi &:= \gamma_n(\phi|_\Gamma n)|_\Gamma && \forall \phi \in C^1(\overline{\Omega}, \mathbb{R}^{d\times d}), && \pi_{nt}\phi := \pi_t(\phi|_\Gamma n) && \forall \phi \in C^1(\overline{\Omega}, \mathbb{R}^d).
\end{aligned}$$

Note that in three dimensions there holds $\pi_t\phi = n \times (\phi \times n)|_\Gamma$ and that in two dimensions γ_t does not exist. For the ease of notation we omit the symbols of the corresponding trace operator if it is clear from the context, e.g. where ϕ_n, ϕ_t represent the normal part and the tangential projection (with respect to π_t) of a vector-valued function. Similarly, ϕ_{nn} and ϕ_{nt} are the normal-normal and the normal-tangential projection of a matrix-valued function.

Next, recall that γ can be extended to the Sobolev space $H^1(\Omega, \mathbb{R})$ such that

$$\gamma : H^1(\Omega, \mathbb{R}) \to H^{1/2}(\Gamma, \mathbb{R}),$$

is a linear, continuous and surjective operator. Here, $H^{1/2}(\Gamma, \mathbb{R})$, denotes the standard trace space of H^1. Next, let $\Gamma_i \subset \Gamma$ be an arbitrary subset, then we define the closed subspaces with vanishing trace

$$\begin{aligned}
H^1_0(\Omega, \mathbb{R}) &:= \{u \in H^1(\Omega, \mathbb{R}) : u = 0 \text{ on } \partial\Omega\},\\
H^1_{0,\Gamma_i}(\Omega, \mathbb{R}) &:= \{u \in H^1(\Omega, \mathbb{R}) : u = 0 \text{ on } \partial\Gamma_i\},
\end{aligned}$$

and similarly the vector-valued versions $H^1_0(\Omega, \mathbb{R}^d)$ and $H^1_{0,\Gamma_i}(\Omega, \mathbb{R}^d)$. For the definition of further trace operators we first need some dual spaces. We use the superscript * in the case of a Hilbert space, whereas the dual spaces of the above defined Sobolev spaces are simply defined using the well known notation with negative indices. Thus we have for example

$$H^{-1}(\Omega, \mathbb{R}) := [H^1_0(\Omega, \mathbb{R})]^* \quad \text{and} \quad H^{-1}_{\Gamma_i}(\Omega, \mathbb{R}) := [H^1_{0,\Gamma_i}(\Omega, \mathbb{R})]^*,$$

and similarly on the boundary

$$H^{-1/2}(\Gamma,\mathbb{R}) := [H^{1/2}(\Gamma,\mathbb{R})]^*.$$

Further we introduce the following notation: the action of a continuous linear functional f on an element g belonging to a topological space X is denoted by $\langle f, g\rangle_X$. We omit the subscript in $\langle\cdot,\cdot\rangle$ when it is obvious from the context. For the Soblev space $H(\mathrm{div},\Omega)$ the appropriate trace operator is given by γ_n such that

$$\gamma_n : H(\mathrm{div},\Omega) \to H^{-1/2}(\Gamma,\mathbb{R}),$$

is a linear, continuous and surjective operator. We define the closed subspaces with vanishing normal trace

$$\begin{aligned} H_0(\mathrm{div},\Omega) &:= \{u \in H(\mathrm{div},\Omega) : \langle u\cdot n, \phi\rangle = 0 \ \forall \phi \in H^1(\Omega,\mathbb{R})\},\\ H_{0,\Gamma_i}(\mathrm{div},\Omega) &:= \{u \in H(\mathrm{div},\Omega) : \langle u\cdot n, \phi\rangle = 0 \ \forall \phi \in H^1_{0,\Gamma\setminus\overline{\Gamma}_i}(\Omega,\mathbb{R})\}. \end{aligned}$$

Finally, the operators γ_t and π_t can be extended to $H(\mathrm{curl},\Omega)$ such that they are linear, continuous and surjective with respect to appropriate trace spaces. Since their construction demands a lot of notation they are neglected for now and will be introduced if necessary.

Finally, similarly to the differential operators above, we define the operator $\mathrm{skw}\cdot$ depending on the context. To this end let $\phi \in \mathcal{D}'(\Omega,\mathbb{R})$ and $\psi \in \mathcal{D}'(\Omega,\mathbb{R}^3)$ then we have

$$\mathrm{skw}\phi = \begin{pmatrix} 0 & -\phi \\ \phi & 0 \end{pmatrix}, \quad \text{and} \quad \mathrm{skw}\psi = \begin{pmatrix} 0 & \psi_3 & -\psi_2 \\ -\psi_3 & 0 & \psi_1 \\ \psi_2 & -\psi_1 & 0 \end{pmatrix}.$$

For matrix valued functions $\phi \in \mathcal{D}'(\Omega,\mathbb{R}^{d\times d})$ we simply set $\mathrm{skw}\phi := \frac{1}{2}\phi - \phi^{\mathrm{T}}$.

We conclude this section by introducing some important inequalities.

Theorem 1 (Inverse inequality for polynomials)**.** *Let $\omega \subset \mathbb{R}^d$ and let $p_h \in \mathbb{P}^k(\omega,\mathbb{R})$. There holds the inverse inequality*

$$\|p_h\|_{\partial\omega} \lesssim \frac{k}{\sqrt{\mathrm{diam}(\omega)}}\|p_h\|_{\omega}.$$

Theorem 2 (Cauchy Schwarz inequality)**.** *Let V be an inner product space, and let $f, g \in$*

V. There holds

$$|(f,g)_V| \leq \|f\|_V \|g\|_V$$

Theorem 3 (Youngs inequality)**.** *There holds the arithmetic-geometric-mean ineqaulity*

$$|ab| \leq \frac{\varepsilon}{2}a^2 + \frac{1}{2\varepsilon}b^2 \quad a,b \in \mathbb{R}, \varepsilon > 0,$$

or as we will often use

$$-|ab| \geq -\frac{\varepsilon}{2}a^2 - \frac{1}{2\varepsilon}b^2 \quad a,b \in \mathbb{R}, \varepsilon > 0.$$

Theorem 4 (Poincaré inequality)**.** *Let $\Omega \subset \mathbb{R}^d$, $d = 2$ or 3, be an arbitrary bounded and connected Lipschitz domain with $\operatorname{diam}(\Omega) = 1$. For a function $u \in H^1(\Omega)$ there holds*

$$\|u\|^2_{H^1(\Omega)} \leq c_P \left(|u|^2_{H^1(\Omega)} + \left(\int_\Omega u \, \mathrm{d}x \right)^2 \right),$$

where c_p only depends on the shape of Ω.

Theorem 5 (Friedrichs inequality)**.** *Let $\Omega \subset \mathbb{R}^d$, $d = 2$ or 3, be an arbitrary bounded and connected Lipschitz domain with $\operatorname{diam}(\Omega) = 1$. Let $\Gamma_D \subset \partial\Omega$ be of positive measure $|\Gamma_D| > 0$. There holds*

$$\|u\|_{H^1(\Omega)} \leq c_F |u|_{H^1(\Omega)} \quad \forall u \in H^1_{0,\Gamma_D}(\Omega),$$

where c_F only depends on the shape of Ω.

Theorem 6 (Korn inequality)**.** *Let $\Omega \subset \mathbb{R}^d$, $d = 2$ or 3, be an arbitrary bounded and connected Lipschitz domain. For $u \in H^1(\Omega, \mathbb{R}^d)$ there holds*

$$\|\varepsilon(u)\|^2_{L^2(\Omega)} + \|u\|^2_{L^2(\Omega)} \geq c_k \|u\|^2_{H^1(\Omega)},$$

where the constant c_k depends on the domain Ω. Now let $\Gamma_D \subset \partial\Omega$ be of positive measure $|\Gamma_D| > 0$, and let $u \in H^1_{0,\Gamma_D}(\Omega, \mathbb{R}^d)$, then

$$\|\varepsilon(u)\|^2_{L^2(\Omega)} \geq c_k \|\nabla u\|^2_{L^2(\Omega)}.$$

Proof. For a detailed proof for a smooth boundary we refer to chapter 3.3 in [11], and for non-smooth boundaries see [9]. □

2.2 Preliminaries and notation for finite element methods

We start with the introduction of several preliminaries that we shall use within this work. Given a domain $\Omega \subset \mathbb{R}^d$ with $d = 2$ or 3 with a Lipschitz boundary, let $\mathcal{T}_h$ be a partition of Ω into triangles and tetrahedrons in two and three dimensions, respectively. Throughout this work we assume that the triangulation $\mathcal{T}_h$ is

- shape regular: There exists a constant $c_s > 0$ such that

$$\max_{T \in \mathcal{T}_h} \frac{\operatorname{diam}(T)^d}{|T|} \leq c_s \quad \text{for all } T \in \mathcal{T}_h,$$

  and

- quasi-uniform: There exists a constant $c_q > 0$ such that

$$\operatorname{diam}(T) \geq c_q h \quad \text{for all } T \in \mathcal{T}_h,$$

  where $h := \max_{T \in \mathcal{T}_h} \operatorname{diam}(T)$.

For a given element $T \in \mathcal{T}_h$ we denote by $\mathcal{V}_h(T)$ the set of vertices of the element T, and by $\mathcal{F}_h(T)$ the set of faces, so the $d-1$ subsimplices, of the element T. In a similar manner we then denote by $\mathcal{F}_h$ the set of all element interfaces and boundaries of the given triangulation $\mathcal{T}_h$. This set can further be split into two parts. The first part is denoted by $\mathcal{F}_h^{\text{ext}}$ and is given by all facets that lie on the boundary of the domain, thus $\mathcal{F}_h^{\text{ext}} := \{F \in \mathcal{F}_h : F \cap \Gamma \neq \emptyset\}$. The second part, denoted by $\mathcal{F}_h^{\text{int}}$, contains all facets that are in the interior of the domain, thus $\mathcal{F}_h^{\text{int}} = \mathcal{F}_h \setminus \mathcal{F}_h^{\text{ext}}$. Finally, we denote by $\mathcal{V}_h$ the set of the nodes of the triangulation $\mathcal{T}_h$ which we split as before into nodes on the boundary $\mathcal{V}_h^{\text{ext}}$ and nodes in the interior $\mathcal{V}_h^{\text{int}}$.

With a slight abuse of notation, we use the same symbol n for the outward unit normal vector on each element boundary ∂T and for the normal vector defined on the boundary Γ. Then, the corresponding normal and tangential traces of smooth vector-valued functions, and the normal-normal and normal-tangential traces of smooth matrix-valued functions on element boundaries and facets are equivalently defined as in section 2.1.

At several points in the definition of the finite elements and also in the numerical analysis we make use of a mapping from a physical element $T \in \mathcal{T}_h$ to a so called reference

element denoted by $\widehat{T}$. To this end we define

$$\begin{aligned}
\widehat{T} &:= \{(x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_1, x_2 \text{ and } x_1 + x_2 \leq 1\} && \text{for} \quad d = 2, \\
\widehat{T} &:= \{(x_1, x_2, x_3) \in \mathbb{R}^3 : 0 \leq x_1, x_2, x_3 \text{ and } x_1 + x_2 + x_3 \leq 1\} && \text{for} \quad d = 3.
\end{aligned}$$

Although one could define a different reference element, it is important that the diameter is approximately one, thus $\operatorname{diam}(\widehat{T}) = \mathcal{O}(1)$. On these reference elements we denote the vertices by

$$V_0 := (0,0), \quad V_1 := (1,0), \quad V_2 := (0,1),$$

and

$$V_0 := (0,0,0), \quad V_1 := (1,0,0), \quad V_2 := (0,1,0), \quad V_3 := (0,0,1),$$

for two and three dimensions, respectively. Next, we further define the follwoing reference faces and the associated normal and tangential vectors. In two dimensions we have

$$\begin{aligned}
\hat{F}_0 &:= \{(x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_1, x_2 \leq 1, x_1 + x_2 = 1\}, \\
\hat{F}_1 &:= \{(0, x_2) \in \mathbb{R}^2 : 0 \leq x_2 \leq 1\}, \\
\hat{F}_2 &:= \{(x_1, 0) \in \mathbb{R}^2 : 0 \leq x_1 \leq 1\},
\end{aligned}$$

with

$$\begin{aligned}
\hat{n}_0 &:= \frac{1}{\sqrt{2}}(1,1)^{\mathrm{T}}, \quad \hat{n}_1 := (-1,0)^{\mathrm{T}}, \quad \hat{n}_2 := (0,-1)^{\mathrm{T}}, \\
\hat{t}_0 &:= \frac{1}{\sqrt{2}}(-1,1)^{\mathrm{T}}, \quad \hat{t}_1 := (0,-1)^{\mathrm{T}}, \quad \hat{t}_2 := (1,0)^{\mathrm{T}}.
\end{aligned}$$

For the three dimensional case we have

$$\begin{aligned}
\hat{F}_0 &:= \{(x_1, x_2, x_3) \in \mathbb{R}^3 : 0 \leq x_1, x_2, x_3 \leq 1, x_1 + x_2 + x_3 = 1\}, \\
\hat{F}_1 &:= \{(0, x_2, x_3) \in \mathbb{R}^3 : 0 \leq x_2, x_3 \leq 1, 0 \leq x_2 + x_3 \leq 1\}, \\
\hat{F}_2 &:= \{(x_1, 0, x_3) \in \mathbb{R}^2 : 0 \leq x_1, x_3 \leq 1, 0 \leq x_1 + x_3 \leq 1\}, \\
\hat{F}_3 &:= \{(x_1, x_2, 0) \in \mathbb{R}^2 : 0 \leq x_1, x_2 \leq 1, 0 \leq x_1 + x_2 \leq 1\},
\end{aligned}$$

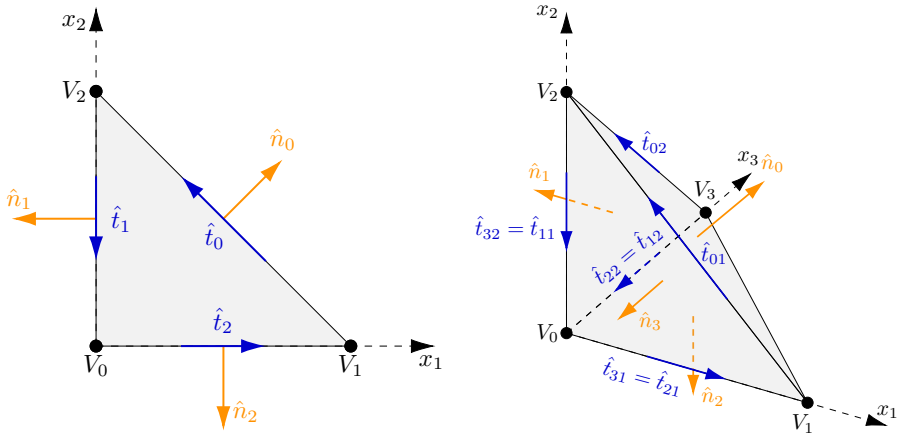


Figure 2.1: The reference element $\hat{T}$ and the corresponding normal and tangential vectors in two dimensions (left) and in three dimensions (right).

with

$$\begin{aligned}
\hat{n}_0 &:= \frac{1}{\sqrt{3}}(1,1,1)^{\mathrm{T}}, \quad \hat{t}_{01} := \frac{1}{\sqrt{2}}(-1,1,0)^{\mathrm{T}}, \quad \hat{t}_{02} := \frac{1}{\sqrt{2}}(0,1,-1)^{\mathrm{T}},\\
\hat{n}_1 &:= (-1,0,0)^{\mathrm{T}}, \quad \hat{t}_{11} := (0,-1,0)^{\mathrm{T}}, \quad\quad\ \hat{t}_{12} := (0,0,-1)^{\mathrm{T}},\\
\hat{n}_2 &:= (0,-1,0)^{\mathrm{T}}, \quad \hat{t}_{21} := (1,0,0)^{\mathrm{T}}, \quad\quad\ \hat{t}_{22} := (0,0,-1)^{\mathrm{T}},\\
\hat{n}_3 &:= (0,0,-1)^{\mathrm{T}}, \quad \hat{t}_{31} := (1,0,0)^{\mathrm{T}}, \quad\quad\ \hat{t}_{32} := (0,-1,0)^{\mathrm{T}}.
\end{aligned}$$

In figure 2.1 we illustrated the reference elements in both dimensions.

By the definition of the reference element we are now able to define the associated element mappings. For an arbitrary element $T \in \mathcal{T}_h$ let $\phi_T : \widehat{T} \to T$ be an affine homeomorphism, with the Jacobi matrix denoted by $F_T := \phi_T'$. As we assumed that the triangulation $\mathcal{T}_h$ is shape regular and quasi-uniform we have

$$||F_T||_\infty \approx h \quad \text{and} \quad ||F_T^{-1}||_\infty \approx h^{-1} \quad \text{and} \quad |\det(F_T)| \approx h^d. \tag{2.3}$$

Similarly, we can restrict the mapping ϕ_T to a reference face $\hat{F} \in \mathcal{F}_h(\widehat{T})$ and reference edge $\hat{E} \subset \partial\hat{F}$ (in three dimensions) whose gradients are then denoted by $F_T^F := (\phi_T|_{\hat{F}})'$ and $F_T^E := (\phi_T|_{\hat{E}})'$. Using these quantities the unit normals and tangents of the reference element and its mapped configurations on the physical element T are related by

$$n = \frac{\det(F_T)}{\det(F_T^F)} F_T^{-\mathrm{T}} \hat{n} \quad \text{and} \quad t = \frac{1}{\det(F_T^E)} F_T \hat{t}, \tag{2.4}$$

where in two dimensions we have to replace F_T^E by F_T^F.

We continue with the definition of polynomial spaces. For a given element $T \in \mathcal{T}_h$ we denote by $\mathbb{P}^k(T)$ the space of polynomials defined on T whose total order is less or equal k. Again, we use the same notation as for function spaces for non scalar-valued polynomial spaces, e.g. where $\mathbb{P}^k(T, \mathbb{R}^d)$ denotes the space of vector-valued polynomials, we use $\mathbb{P}^k(T, \mathbb{R}^{d\times d})$ for the space of matrix-valued polynomials. Using these notations we further define polynomials on the triangulation by

$$\mathbb{P}^k(\mathcal{T}_h, \mathbb{R}) := \prod_{T\in\mathcal{T}_h} \mathbb{P}^k(T, \mathbb{R}),$$

and similarly $\mathbb{P}^k(\mathcal{T}_h, \mathbb{R}^d)$ and $\mathbb{P}^k(\mathcal{T}_h, \mathbb{R}^{d\times d})$. Beside this we make use of homogeneous polynomials denoted by $\mathbb{P}^k_{\mathrm{hom}}(\mathcal{T}_h, \mathbb{R})$ and the space of matrix-valued skew symmetric polynomials defined by

$$\mathbb{P}^k_{\mathrm{skw}}(\mathcal{T}_h, \mathbb{R}^{d\times d}) := \{\eta \in \mathbb{P}^k(\mathcal{T}_h, \mathbb{R}^{d\times d}) : (\eta + \eta^{\mathrm{T}})|_T = 0 \text{ on all } T \in \mathcal{T}_h\}.$$

Finally, we introduce the space of rigid displacements by

$$\mathrm{RM}(\mathcal{T}_h) := \{a + Bx : a \in \mathbb{P}^0(T, \mathbb{R}^d), B \in \mathbb{P}^0_{\mathrm{skw}}(T, \mathbb{R}^{d\times d})\}. \tag{2.5}$$

At several points in the analysis we make use of polynomials defined in the tangent plane of a face of a given element T. To this end let $F \in \mathcal{F}_h(T)$, then with a slight abuse of notation we do not distinguish between the tangent plane parallel to the facet F and the isomorphic $\mathbb{R}^{d-1}$ and write instead $\mathbb{P}^k(F, \mathbb{R}^{d-1})$. Note that for example the tangential projection of a polynomial $\mu \in \mathbb{P}^k(T, \mathbb{R}^d)$ is in this space, thus $\mu_t \in \mathbb{P}^k(F, \mathbb{R}^{d-1})$.

With respect to a triangulation we introduce for each element $T \in \mathcal{T}_h$ the local element-wise L^2-projection on polynomials of order k by Π_T^k. Note that we do not distinguish between scalar-, vector- or matrix-valued functions, but always use the same symbol. Following the notations from above the corresponding global L^2-projection onto the space $\mathbb{P}^k(\mathcal{T}_h)$ is given by $\Pi^k_{\mathcal{T}_h}$. Similarly, on each facet $F \in \mathcal{F}_h$, let Π_F^k denote the L^2-projection onto the space of polynomials of order k on F. Again, we use the same symbols for projections with different ranges. For example, the projection into the tangent plane of F is also given by Π_F^k, i.e., with the notation from above we have for any vector-valued function $v \in L^2(F, \mathbb{R}^{d-1})$ that the projection $\Pi_F^k v \in \mathbb{P}^k(F, \mathbb{R}^{d-1})$ satisfies $(\Pi_F^k v, q)_F = (v, q)_F$ for all $q \in \mathbb{P}^k(F, \mathbb{R}^{d-1})$.

Similarly, we also define function spaces with respect to the triangulation $\mathcal{T}_h$, e.g.

$$H^m(\mathcal{T}_h, \mathbb{R}) := \{u \in L^2(\Omega, \mathbb{R}) : u|_T \in H^m(T, \mathbb{R}) \text{ for all } T \in \mathcal{T}_h\},$$

denotes the broken Sobolev space of order m. Note that we use the same symbols for a broken differential operator applied on each element for functions in a broken Sobolev space and the continuous operator applied on functions in the corresponding standard Sobolev space, e.g. we write $(\nabla u)|_T = \nabla(u|_T)$ for functions $u \in H^1(\mathcal{T}_h, \mathbb{R})$.

Now let $I_{\mathcal{V}_h(T)}$ be the index set of the vertices $\mathcal{V}_h(T)$, then we use the standard notation for the barycentric coordinate functions given by λ_i, thus we have

$$\lambda_i \in \mathbb{P}^1(T, \mathbb{R}) \quad \text{such that} \quad \lambda_i(V_j) = \delta_{ij} \quad \forall i, j \in I_{\mathcal{V}_h(T)},$$

where δ_{ij} is the Kronecker delta.

2.3 The variational formulation of the Stokes equations

Before we can deal with the time dependent non-linear versions of the fluid equations we have to develop some basics knowledge of the Stokes equations and its descretization techniques. For the ease we only consider the case of homogenoues Dirichlet boundary condition, i.e. we have the problem: Find u, p such that

$$\begin{aligned} -\nu \operatorname{div}(\varepsilon(u)) + \nabla p &= f, \\ \operatorname{div}(u) &= 0. \end{aligned} \tag{2.6}$$

Note, that for simplicity we neglect the scaling of the viscosity with the constant 2 in this section. In a first step we are going to prove that the Stokes equations have an unique solution (if it exists). To this end we need to take a closer look onto the kernel of the symmetric gradient.

Theorem 7. *The strain $\tau = \varepsilon(u)$ vanishes if and only if the velocity is a rigid body motion, i.e. for $d = 3$ we have*

$$\varepsilon(u) = 0 \Leftrightarrow u(x) = a + b \times x$$

where $a, b \in \mathbb{R}^3$ and for $d = 2$ we have

$$\varepsilon(u) = 0 \Leftrightarrow u(x) = a + b \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix},$$

with $a \in \mathbb{R}^2$ *and* $b \in \mathbb{R}$*.*

Note, that the term “rigid body motion” is motivated from the theory of elasticity. For the Stokes equations, a rigid body motion equals a flow where the velocity of every fluid particle is a linear combination of a constant velocity and a constant rotation. Obviously, this induced no diffusive forces and the strain vanishes. Understanding the kernel, we can now proof the following uniqueness result

Theorem 8. *Assuming enough regularity, the Stokes problem* (2.6) *has an, up to an additive constant pressure, unique solution.*

Proof. Since the Stokes equations are a linear problem, we have to show that if the right hand side vanishes $f = 0$, the solution is given by $u = 0$ and $p = c$, with $c \in \mathbb{R}$. In a first step we multiply the first equation with the exact solution u, integrate over the domain Ω, and apply integration by parts. This gives

$$0 = \int_\Omega \nu\varepsilon(u) : \nabla(u)\,\mathrm{d}x - \int_\Omega p\,\mathrm{div}(u)\,\mathrm{d}x\,.$$

Due to the incompressibility constraint the second integral vanishes and we obtain (using the symmetry of $\varepsilon(u)$)

$$\begin{aligned} 0 = \int_\Omega \nu\varepsilon(u) : \nabla(u)\,\mathrm{d}x &= \int_\Omega \nu\varepsilon(u) : \frac{1}{2}\nabla(u)\,\mathrm{d}x + \int_\Omega \nu\varepsilon(u) : \frac{1}{2}\nabla(u)\,\mathrm{d}x \\ &= \int_\Omega \nu\varepsilon(u) : \frac{1}{2}\nabla(u)\,\mathrm{d}x + \int_\Omega \nu\varepsilon(u) : \frac{1}{2}\nabla(u)^{\mathrm{T}}\,\mathrm{d}x \\ &= \int_\Omega \nu|\varepsilon(u)|^2\,\mathrm{d}x\,. \end{aligned}$$

By Theorem 7, the vanishing L^2-norm of the strain implies that u equals a rigid body motion. However, since we consider homogeneous Dirichlet boundary condition on $\partial\Omega$ this finally gives that $u = 0$ and further

$$-\nu\,\mathrm{div}(\varepsilon(u)) + \nabla p = \nabla p = 0,$$

from what we conclude the proof, since the first equations now gives $\nabla p = 0$, i.e. p is constant. □

We continue with the derivation of the weak formulation of the Stokes equations which, as usual, follows very similar steps as in the proof above. The second order differential operator in the Stokes equations motivates to choose the (vector valued!) space $V := H_0^1(\Omega, \mathbb{R}^d)$ for the velocity. In order to guarantee uniqueness of the variational formulation

we consider the closed subspace space of square integratable functions with vanishing mean value

$$Q := L_0^2(\Omega, \mathbb{R}) := \{f \in L^2(\Omega, \mathbb{R}) : \int_\Omega f \, \mathrm{d}x = 0\},$$

as space for the pressure. Multiplying (2.6) with test functions $v, q \in V \times Q$, integrating over the domain Ω and integrate by parts gives the weak formulation: Find $u, p \in V \times Q$ such that

$$\begin{aligned} \int_\Omega \nu\varepsilon(u) : \varepsilon(v) \, \mathrm{d}x - \int_\Omega \mathrm{div}(v)p &= \int_\Omega f \cdot v \, \mathrm{d}x \quad \forall v \in V \\ - \int_\Omega \mathrm{div}(u)q &= 0 \qquad\qquad \forall q \in Q. \end{aligned} \tag{2.7}$$

In order to prove uniqueness (in the spaces V and Q) we follow similar steps as before. Choosing $f = 0$ and the test functions $v = u$ and $q = p$ gives (including the incompressibility constraint)

$$\int_\Omega \nu\varepsilon(u) : \varepsilon(v) \, \mathrm{d}x = 0,$$

which again implies $u = 0$. What remains is the condition

$$- \int_\Omega \mathrm{div}(v)p = 0 \quad \forall v \in V,$$

or, assuming a smooth solution, integration by parts also gives

$$\int_\Omega v \cdot \nabla p = 0 \quad \forall v \in V,$$

Now let $b(x)$ be a positive function $b(x) > 0$ for all $x \in \Omega$ that vanishes on the boundary, i.e. $b(x)|_{\partial\Omega} = 0$ (the b stands for *bubble function*). Choosing $v = \nabla p b$ (the b was needed for the boundary conditions) gives

$$\int_\Omega b|\nabla p|^2 \, \mathrm{d}x = 0,$$

and thus (since this reads as an (equivalent) weighted L^2-norm) $\nabla p = 0$ implying a constant pressure. Due to the choice of the space Q, this shows that $p = 0$. In the case of a non-smooth solution (p only in L^2), "integration by parts" gives the duality pair $\langle \nabla p, v \rangle_V = 0$ for all $v \in V$. Proving that this again implies that $p = 0$ is not that simple and requires some applied functional analysis which we will discuss later.

Since the symmetric bilinear form in the upper left part of (2.7) is elliptic (as we will discuss in the next section), we can interpret the weak formulation of the Stokes equations as the Euler-Lagrange equations of a constrained optimization problem. In particular, the velocity solution $u \in V$ is given as the solution of

$$\min_{v \in V} \int_\Omega \frac{\nu}{2} |\varepsilon(v)|^2 - f \cdot v \, \mathrm{d}x,$$

subject to the constraint

$$\mathrm{div}(v) = 0.$$

To solve this problem we can define the Lagrange function $\mathcal{L} : V \times Q \to \mathbb{R}$ by

$$\mathcal{L}(v, q) := \frac{\nu}{2} \|\varepsilon(v)\|_\Omega^2 - (\mathrm{div}(v), q)_\Omega.$$

The variation with respect to the velocity test function gives the first equation of (2.7), and the variation of the scalar pressure test function gives the incompressibility constraint, i.e. the second line of (2.7). This shows, that the physical meaning (and also in a mathematical sense) of the pressure is the Lagrange multiplier enforcing the divergence constraint of the velocity. Further, the solution of the minimization problem is a saddle point, i.e. the velocity u is a minimizer and p is a maximizer

$$\mathcal{L}(u, q) \leq \mathcal{L}(u, p) \leq \mathcal{L}(v, p) \quad \forall (v, q) \in V \times Q.$$

In the next section we discuss the stability of mixed methods in a general setting. Further, we will focus on the case where, motivated by above findings, the solution corresponds to a saddle point problem.

2.4 Stability theory of mixed methods

2.4.1 The inf-sup theorem

In this section we discuss the stability of variational problems in a more general framework. To give a proper name to the results and theorems, a detailed study of the history is needed and even then there might be some disagreement. One can find these results for example in the survey of lectures by Babuška and Aziz, see [4] where they also refer to the works of Nirenberg, see [37]. At several points in the literature one also finds references of the work by Nečas, see [36]. In [13], the theorem is called the BNB theorem, since

beside Babuška and Nečas, it can also be seen as rephrasing two fundamental results of Banach (the closed range and open mapping theorem). For the ease, we will call it the *inf-sup theorem*.

Now let H be a Hilbert space with the inner product $(\cdot,\cdot)_H$ and the corresponding norm $\|\cdot\|_H$. Assume a given bilinear form $K : H \times H \to \mathbb{R}$ and a given right hand side $F \in H'$. We consider the problem: Find $u \in H$ such that

$$K(u,v) = F(v) \quad \forall v \in H. \tag{2.8}$$

Theorem 9 (inf-sup)**.** *Consider the above setting, and suppose that the bilinear form K fulfills the following conditions:*

- *Continuity: there exists a positive constant α such that*

$$|K(u,v)| \leq \alpha \|u\|_H \|v\|_H \quad \forall u,v \in H.$$

- *The "inf-sup" condition: there exists a positive constant β such that*

$$\sup_{v\in H, v\neq 0} \frac{K(u,v)}{\|v\|_H} \geq \beta \|u\| \quad \forall u \in H.$$

- *There holds*

$$\sup_{u\in H} K(u,v) \neq 0 \quad \forall v \in V.$$

Then, the variational problem (2.8) *has an unique solution depending continuously on the data, i.e.*

$$\|u\|_H \leq \frac{1}{\beta} \|F\|_{H'}.$$

Proof. Step 1: Let $w \in H$ be arbitrary and define the functional

$$\phi_w(v) := K(w,v) \quad \forall v \in H.$$

By the Riesz representation theorem, there exists a function $z \in H$ such that

$$(z,v)_H = \phi_w(v).$$

Hence, we can define a linear mapping $\mathcal{K} : H \to H, \mathcal{K}(w) = z$ such that

$$(\mathcal{K}(w), v) = K(w, v) \quad \forall v \in H.$$

Using the continuity properties of the bilinear form we have for all $w \in H$

$$\|\mathcal{K}(w)\|_H^2 = (\mathcal{K}(w), \mathcal{K}(w)) = K(w, \mathcal{K}(w)) \leq \alpha \|w\|_H \|\mathcal{K}(w)\|_H, \tag{2.9}$$

thus $\mathcal{K}$ is bounded $\|\mathcal{K}\| \leq \alpha$ (where $\|\cdot\|$ is the operator norm).
Step 2: We continue by proving that $\mathcal{K}$ is also bounded from below and that the range $R(\mathcal{K})$ is closed. The first statement follows immediately by the inf-sup condition

$$\|\mathcal{K}(w)\|_H = \sup_{v \in H, v \neq 0} \frac{(\mathcal{K}(w), v)_H}{\|v\|_H} = \sup_{v \in H, v \neq 0} \frac{K(w, v)_H}{\|v\|_H} \geq \beta \|w\|.$$

No let $\mathcal{K}(w_n)$ be a Cauchy sequence in $R(\mathcal{K})$. Using above estimate we have

$$\|\mathcal{K}(w_n) - \mathcal{K}(w_m)\|_H \geq \beta \|w_n - w_m\|_H,$$

and thus w_n is also a Cauchy sequence (in the Hilbert space H). For the ease, let w denote the limit of w_n. Since $\mathcal{K}$ is bounded we also have that $\mathcal{K}(w_n) \to \mathcal{K}(w)$, i.e. the range $R(\mathcal{K})$ is closed.
Step 3: We proof $R(\mathcal{K}) = H$ by contradiction: Assume there exists an element $v_0 \in H$ with $v_0 \neq 0$ such that

$$(\mathcal{K}(w), v_0)_H = 0 \quad \forall w \in H, \quad (\Leftrightarrow v_0 \perp^H R(\mathcal{K})).$$

By definition, this is equivalent to $K(w, v_0) = 0$ for all $w \in H$, thus we have a contradiction to the third assumption of the theorem.
Step 4: We apply the Riesz theorem to the right hand side of problem (2.8), i.e. we find a function u_F such that

$$F(v) = (u_F, v)_H \quad \forall v \in H,$$

thus the variational problem is equivalent to the operator problem $\mathcal{K}(u) = u_f$, with the solution

$$u = \mathcal{K}^{-1}(u_F),$$

where $\mathcal{K}$ is invertible since $\mathcal{K}$ is a bijective bounded linear operator, thus the existence

of the inverse follows from the bounded inverse theorem (which is equivalent to the open mapping and closed graph theorem, see comment above).
Step 5: From the inf-sup condition we finally have

$$\beta\|u\|_H \leq \sup_{v\in H, v\neq 0} \frac{K(u,v)}{\|v\|_H} = \sup_{v\in H, v\neq 0} \frac{F(v)}{\|v\|_H} = \|F\|_{H'}.$$

□

Remark 1. The inf-sup theorem as stated above only considers the case where the bilinear form K is defined on $H \times H$, where H is a Hilbert space, thus it can be seen as a generalized version of the Lax-Milgram theorem. The universal case considers the a bilinear form $K : U \times V \to \mathbb{R}$, with two Banach spaces U, V (i.e. the name BNB-theorem, see above).

Remark 2. The first part of step 2 in the above proof showed that the operator $\mathcal{K}$ is injective, i.e., the inf-sup condition gives the uniqueness of the problem. The second part of step 2 and step 3 showed the surjectivity.

Remark 3. There are alternative versions of stability (the inf-sup) condition

- There exists a positive constant β such that

$$\inf_{u\in H, u\neq 0} \sup_{v\in H, v\neq 0} \frac{K(u,v)}{\|u\|_H \|v\|_H} \geq \beta.$$

- There exists a positive constant β such that for every $u \in H$ there exists a $v \in H$ such that

$$K(u,v) = \|u\|_H^2 \quad \text{and} \quad \|v\|_H \leq \beta\|u\|_H.$$

Considering the variational formulation of the Stokes equations (2.7), we could now set $H := H_0^1(\Omega, \mathbb{R}^3) \times L_0^2(\Omega)$ and define the bilinear form

$$K((u,p),(v,q)) := \int_\Omega \nu\varepsilon(u) : \varepsilon(v)\,\mathrm{d}x - \int_\Omega \operatorname{div}(u) q\,\mathrm{d}x - \int_\Omega \operatorname{div}(v) p\,\mathrm{d}x,$$

and the linear form $F((v,q)) := f(v)$. In order to guarantee that the problem is well posed, we have to check the stability conditions of Theorem 9. Nevertheless, as already discussed in the previous section, the solution of the variational formulation of the Stokes problem is a saddle point, i.e. we have a saddle point problem. In this case simplified (inf-sup like) conditions can be considered which are discussed in the next section.

2.4.2 The Brezzi theorem for saddle point problems

We consider the abstract setting of a saddle point problem. To this end let V, Q be two Hilbert spaces with the inner product $(\cdot,\cdot)_V$ and $(\cdot,\cdot)_Q$, and the corresponding norms $\|\cdot\|_V$ and $\|\cdot\|_Q$. We consider the problem: Find $(u,p) \in V \times Q$ such that:

$$\begin{aligned} a(u,v) + b(v,p) &= f(v) \quad \forall v \in V \\ b(u,q) &= g(q) \quad \forall q \in Q \end{aligned} \tag{2.10}$$

for given right hand sides $f \in V'$ and $g \in Q'$, and the bilinear forms $a : V \times V \to \mathbb{R}$ and $b : V \times Q \to \mathbb{R}$.

Theorem 10. *Consider the above settings and assume that the bilinear forms fulfill the conditions:*

- *The bilinear forms are continuous*

$$\begin{aligned} a(u,v) &\leq \alpha_1 \|u\|_V \|v\|_V \quad \forall v, u \in V \\ b(u,q) &\leq \alpha_2 \|u\|_V \|q\|_Q \quad \forall v \in V, \forall q \in Q. \end{aligned}$$

- *The bilinear form a is elliptic on the kernel of the bilinear form b, i.e. we have*

$$a(u,u) \geq \beta_1 \|u\|_V^2 \quad \forall u \in V_0 := \{v \in V : b(v,q) = 0 \; \forall q \in Q\}.$$

- *The bilinear form b fulfills the LBB (Ladyshenskaya-Babuška-Brezzi) condition, i.e.*

$$\sup_{u \in V, u \neq 0} \frac{b(u,q)}{\|u\|_V} \geq \beta_2 \|q\|_Q \quad \forall q \in Q.$$

Then, the variational problem (2.10) has a unique solution depending continuously on the data, i.e.

$$\|u\|_V + \|p\|_Q \lesssim \beta_2^{-2}(\|f\|_{V'} + \|g\|_{Q'}) \quad \textit{and} \quad \|u\|_V \lesssim \beta_2^{-1}(\|f\|_{V'} + \|g\|_{Q'}).$$

where the constants depends on $\alpha_1, \alpha_2, \beta_1$.

Proof. We want to show that the conditions of Theorem 10 imply that the conditions of Theorem 9 are valid. To this end we define the “big” bilinear form on the product space $H = V \times Q$ with the norm $\|(u,p)\|_H^2 = \|u\|_V^2 + \|p\|_Q^2$ by

$$K((u,p),(v,q)) = a(u,v) + b(u,q) + b(v,p).$$

In the proof we want to explicitly keep track of the LBB constant as it plays an important role (for example in the theory of preconditioners).
Step 1: The continuity of $\mathcal{K}$ on H is a direct consequence of the continuity of the bilinear forms a and b., i.e. we have

$$|\mathcal{K}((u,p),(v,q))| \leq |a(u,v)| + |b(u,q)| + |b(v,p)| \lesssim \|(u,p)\|_H \|(v,q)\|_H.$$

Step 2: Let $v \in V$ and $q \in Q$ be arbitrary (but fixed). First, the LBB condition (surjectivity) shows that there exists an u_1 (not unique!) such that

$$b(u_1,q) = (q,q)_Q \quad \text{and} \quad \|u_1\|_V \lesssim \beta_2^{-1}\|q\|_Q.$$

Next, we solve the following problem on the kernel: Find $u_0 \in V_0$ such that

$$a(u_0,v_0) = (v,v_0) - a(u_1,v_0) \quad \forall v_0 \in V_0.$$

By Lax-Milgram this problem has a unique solution with the stability estimate

$$\|u_0\|_V \lesssim \|v\|_V + \|u_1\|_V.$$

Step 3: We set $u = u_0 + u_1$, and define the functional $(v,\cdot)_V - a(u,\cdot) \in V'$. Then, using the Riesz isomorphism we find a function $z \in V$ such that

$$(z,w)_V = (v,w)_V - a(u,w) \quad \forall w \in V.$$

By construction, we have that for all $v_0 \in V_0$

$$(z,v_0)_V = (v,v_0)_V - a(u,v_0) = (v,v_0)_V - a(u_0,v_0) - a(u_1,v_0) = 0,$$

thus, $z \in V_0^\perp$. As in the proof of Theorem 9, we now define the operator $B^* : Q \to V$ such that $(u,B^*p)_V = b(u,p)$. The LBB conditions now shows that we can bound B^* from below,

$$\beta_2\|p\|_Q \leq \sup_{w\in V} \frac{b(w,p)}{\|w\|_V} = \sup_{w\in V} \frac{(w,B^*p)_V}{\|w\|_V} \leq \|B^*p\|_V,$$

and thus, similarly to before, this shows that the range $R(B^*) = B^*Q$ is closed. Since

$$(v_0,B^*p)_V = b(v_0,p) = 0 \quad \forall v_0 \in V_0$$

shows that B^*Q is V-orthogonal onto V_0, and since V_0 is closed (kernel of cont. operator) we have the orthogonal decomposition $V = V_0 \oplus B^*Q$. In total this gives $z \in B^*Q$ and so we can find a $p \in Q$ such that $z = B^*p$. Further, we have the stability estimate

$$\|p\|_Q \leq \beta_2^{-1}\|z\|_V \lesssim \beta_2^{-1}(\|v\|_V + \|u_1\|_V) \lesssim \beta_2^{-2}(\|v\|_V + \|q\|_Q),$$

or all together

$$\|u\|_V + \|p\|_Q \lesssim \beta_2^{-2}(\|v\|_V + \|q\|_Q).$$

Since we also have

$$\begin{aligned} K((u,p),(v,q)) &= a(u,v) + b(u,q) + b(v,p) \\ &= a(u,v) + b(u,q) + (z,v)_V \\ &= a(u,v) + b(u,q) + (v,v)_V - a(u,v) \\ &= b(u_0,q) + b(u_1,q) + \|v\|_V^2 \\ &= \|q\|_Q^2 + \|v\|_V^2. \end{aligned}$$

This concludes the well posedness proof (see second point of Remark 3) and gives the continuity estimate

$$\|u\|_V + \|p\|_Q \lesssim C\beta_2^{-2}(\|f\|_{V'} + \|g\|_{Q'}).$$

For the second estimate, follow the same steps as above but scale the solution p with the LBB constant. This gives the stability estimate

$$\|u\|_V \leq \|u\|_V + \|\beta_2 p\|_Q \lesssim \beta_2^{-1}(\|v\|_V + \|q\|_Q).$$

The rest follows as before. □

Remark 4. In contrast to the proof of Theorem 9, where the stability conditions where used to show that the corresponding operator is bijective, the conditions of the Brezzi theorem can be interpreted in the following sense: the kernel ellpiticity simply provides a condition for the solvability in the case where the constraint given by the b bilinear form vanishes. Here we can simply use the standard theory of elliptic problems given by the Lax-Milgram theorem. The LBB condition has to be valid because it guarantees that there are “enough” functions in V such that the second line of the saddle point problem can be fulfilled (i.e. we have surjectivity of the operator corresponding to the constraint).

Theorem 10, also often just called Brezzi's theorem, shows that there are two crucial conditions that we have to check: the kernel ellipticity and the LBB condition. If we apply the above setting to the Stokes equations we set $V := H_0^1(\Omega, \mathbb{R}^3)$ and $Q = L_0^2(\Omega)$ and define for all $u, v \in V$ and $q \in Q$ the bilinear forms

$$a(u, v) := \int_\Omega \nu\varepsilon(u) : \varepsilon(v)\,\mathrm{d}x \tag{2.11a}$$

$$b(u, q) := -\int_\Omega \mathrm{div}(u)q\,\mathrm{d}x\,. \tag{2.11b}$$

The continuity reads by a proper scaling with the viscosity and reads as

$$\begin{aligned} a(u, v) &\leq \nu\|u\|_1\|v\|_1 \quad \forall v, u \in V \\ b(u, q) &\leq \|u\|_1\|q\|_Q \quad \forall v \in V, \forall q \in Q. \end{aligned}$$

and the kernel is given by all divergence free functions

$$\begin{aligned} V_0 :=& \{u \in H_0^1(\Omega, \mathbb{R}^3) : b(u, q) = 0\ \forall q \in Q\} \\ =& \{u \in H_0^1(\Omega, \mathbb{R}^3) : \mathrm{div}(u) = 0 \text{ in } L^2\}. \end{aligned}$$

The coercivity for the Stokes equations follows immediately using Korn's inequality, see Theorem 6, i.e. we have

$$a(u, u) \gtrsim \nu\|u\|_1^2 \quad \forall u \in V.$$

Note, that we even have ellipticity on the whole space V (and not only on the kernel), but keep in mind that this is not the usual case. The LBB condition now reads as

$$\sup_{u \in V} \frac{\int_\Omega \mathrm{div}(u)q\,\mathrm{d}x}{\|u\|_1} \gtrsim \|q\|_0 \quad \forall q \in Q. \tag{2.12}$$

Unfortunately, there is no simple proof of the above theorem for arbitrary domains. We refer for example to [13] where the LBB condition is discussed in more details including different boundary conditions. Nevertheless, a proof can be constructed in the case where we assume the surjectivity of the H^2 trace operator, i.e. we consider a domain Ω such that the operators

$$(\gamma(\cdot), \gamma_n(\nabla(\cdot)) : H^2(\Omega, \mathbb{R}) \to (H^{3/2}(\partial\Omega, \mathbb{R}), H^{1/2}(\partial\Omega, \mathbb{R}))$$

(evaluation of a function and its normal derivative) is continuous and surjective. This im-

plies that the boundary of the domain has to be smooth enough, see [32, 12, 19, 6] for a detailed discussion. The LBB condition reads as the surjectivity of the divergence operator (see remarks above), i.e. consider an arbitrary $q \in Q$, then we have to find a $u \in V$ such that $\operatorname{div}(u) = q$ and $\|u\|_1 \lesssim \|q\|_0$. In a first step we solve the auxiliary Poisson problem $-\Delta\varphi = -q$ in Ω with Neumann boundary conditions $\nabla\varphi \cdot n = 0$ on $\partial\Omega$. Due to the zero mean value of q this problem has a unique solution in $H^1(\Omega, \mathbb{R})/\mathbb{R}$. Now set $u := \nabla\varphi$ to get $\operatorname{div}(u) = \Delta\varphi = q$ and using a regularity result for the Poisson problem we get $\|u\|_1 = \|\varphi\|_2 \lesssim \|q\|_0$. Further note, that we already have $u \cdot n = \nabla\varphi \cdot n = 0$ on $\partial\Omega$. In the next step we are going to correct the tangential component such that the resulting velocity satisfies the zero boundary conditions of V. Thus, we seek for a function $\psi \in H^2(\Omega, \mathbb{R}^3)$ that fulfills

$$\psi = 0 \quad \text{on } \partial\Omega \quad \text{and} \quad \frac{\partial\psi}{\partial n} = -u \cdot t \quad \text{on } \partial\Omega \quad \text{and} \quad \|\psi\|_2 \lesssim \|u\|_1.$$

Since we assumed that the H^2 trace operator is surjective, the existence of such a function is guaranteed, see Theorem 1.12 in [7]. Now set $\tilde{u} := u + \operatorname{curl}\psi$ to get $\operatorname{div}\tilde{v} = \operatorname{div}(u) + \operatorname{div}(\operatorname{curl}(\psi)) = q$ in Ω. On the boundary $\partial\Omega$ we observe

$$\tilde{u} \cdot n = u \cdot n + \operatorname{curl}(\psi) \cdot n = \nabla\psi \cdot t = 0 \quad \text{and} \quad \tilde{u} \cdot t = u \cdot t + \operatorname{curl}(\psi) \cdot t = u \cdot t + \nabla\psi \cdot nn = 0.$$

Finally, due to the H^2-continuity of ψ, we get $\|\tilde{u}\|_1 = \|u\|_1 + \|\operatorname{curl}(\psi)\|_1 \lesssim \|u\|_1 \lesssim \|q\|_0$.

2.5 Conforming Finite element methods for the Stokes equations

In this section we want to derive a (conforming) finite element method in order to discretize the variational formulation (2.7). To this end let $V_h \subset V$ and $Q_h \subset Q$ be two finite dimensional spaces, then we have the problem: Find $(u_h, p_h) \in V_h \times Q_h$ such that

$$a(u_h, v_h) + b(v_h, p_h) = (f, v_h) \qquad \forall v_h \in V_h \tag{2.13}$$

$$b(u_h, q_h) = 0 \qquad \forall q_h \in Q_h, \tag{2.14}$$

where the bilinear forms are given by (2.11). The stability conditions of Theorem 10, show that there is a strong connection between the continuous velocity space V and the pressure space Q. Unfortunately, in contrast to standard elliptic problems where the solveability of a (conforming) discrete method is inherited from the continuous setting, this is not the case for saddle point problems. Thus, the discrete spaces V_h and Q_h can not be

chosen independently. Since the kernel ellipticity of the Stokes problem on the continuous level holds on the whole space V, this condition is indeed inherited due to the conformity $V_h \subset V$. Thus, in order to prove well posedness (i.e. unique solveability) of (2.13) we can focus on the LBB condition.

However, since the LBB condition might be difficult to prove, we will first discuss a uniqueness (but not existence) property that might be easier to check.

Theorem 11. *The solution of the discrete problem* (2.13) *is unique if and only if the discrete finite element spaces* V_h *and* Q_h *fulfill the condition*

$$b(v_h, p_h) = -\int_\Omega \operatorname{div}(v_h) p_h \, \mathrm{d}x = 0 \quad \forall v_h \in V_h \Rightarrow p_h = 0. \tag{2.15}$$

Proof. The proof follows with the same steps as in the continuous setting. □

Remark 5. In general, above theorem should be stated such that the implication gives that the discrete pressure is constant, i.e. $p_h = c \in \mathbb{R}$ (and not zero as above). Here, we explicitly have stated that p_h should vanish as we enforced uniqueness by the zero mean value constraint $Q_h \subset Q = L^2_0(\Omega)$. Note, that this is just a mathematical "grounding" technique. One could have also enforced (for example) a different (fixed) non-zero mean value to guarantee uniqueness.

Above condition is not always true and can help to check if a pair of finite element spaces is a suitable couple for the discretization of the Stokes equations.

Example 1. Probably the first most trivial choice of a finite element discretization might be to choose a standard linear Lagrangian finite element approximation for the velocity and the pressure, i.e. we choose the space

$$\begin{aligned} V_h &:= \{v_h \in H^1_0(\Omega, \mathbb{R}^d) : v_h|_T \in \mathbb{P}^1(T, \mathbb{R}^d) \; \forall T \in \mathcal{T}_h\}, \\ Q_h &:= \{q_h \in L^2_0(\Omega, \mathbb{R}) \cap C^0(\Omega, \mathbb{R}) : q_h|_T \in \mathbb{P}^1(T, \mathbb{R}) \; \forall T \in \mathcal{T}_h\}. \end{aligned}$$

We give a simple counter example which proves that this discretization does not provide a unique solution. Let Ω be a square and let the triangulation be given as in Figure 2.2. We set p_h such that its evaluation equals either -1 or 1 on nodes that are on common vertical lines. Any function $v_h \in V_h$ is uniquely defined by fixing the values on the nodes in the interior, which are further associated to the corresponding nodal hat functions. For simplicity let $v_h = \varphi$, where φ is the hat function of the blue vertex with support on the vertex patch $\omega = \omega_r \cup \omega_o$ which is split into an orange part ω_o and a red part ω_r. Since ∇p_h

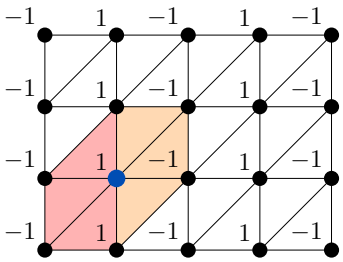


Figure 2.2: Considered triangulation and nodal values of p_h

is 2 and -2 on ω_r and ω_o, respectively, integration by parts shows

$$-\int_\Omega \operatorname{div}(v_h)p_h \,\mathrm{d}x = -\int_\omega \operatorname{div}(v_h)p_h \,\mathrm{d}x = \int_\omega v_h \cdot \nabla p_h \,\mathrm{d}x = 2\int_{\omega_r} \varphi - 2\int_{\omega_o} \varphi = 0,$$

where we used, that $|\omega_o| = |\omega_r|$ and that φ is point symmetric on ω with respect to the blue vertex (i.e. the integrals have the same value). A similar argument can be used for any other hat function (and linear combination) which shows that

$$-\int_\Omega \operatorname{div}(v_h)p_h \,\mathrm{d}x = 0 \quad \forall v_h \in V_h.$$

Example 2 (The MINI element)*.* The MINI element uses the same pressure approximation as before, but the linear Lagrangian velocity space is augmented by local element wise bubble functions such that it admits a unique solution. To this end we define for each element $T \in \mathcal{T}_h$ the bubble space given by $B(T,\mathbb{R}) = \mathbb{P}^{d+1}(T,\mathbb{R}) \cap H_0^1(T,\mathbb{R})$, thus cubic or quartic polynomials for $d = 2$ and $d = 3$ respectively, that vanish on the boundary of the element. A local basis function of (the one dimensional space) $B(T,\mathbb{R})$ is simply given by the bubble $b_T = \Pi_{i=1}^{d+1}\lambda_i$ where λ_i are the barycentric coordinate functions on T (i.e. linear polynomials). With a slight abuse of notation let $B(T,\mathbb{R}^d)$ be the vector valued bubble space where each component is given by $B(T,\mathbb{R})$. We choose the spaces

$$\begin{aligned} V_h &:= \{v_h \in H_0^1(\Omega,\mathbb{R}^d) : v_h|_T \in [\mathbb{P}^1(T,\mathbb{R}^d) + B(T,\mathbb{R}^d)]\ \forall T \in \mathcal{T}_h\}, \\ Q_h &:= \{q_h \in L_0^2(\Omega,\mathbb{R}) \cap C^0(\Omega,\mathbb{R}) : q_h|_T \in \mathbb{P}^1(T,\mathbb{R})\ \forall T \in \mathcal{T}_h\}. \end{aligned}$$

Now assume that for a given p_h the condition (2.15) is satisfied . Let $T \in \mathcal{T}_h$ be arbitrary, then we choose the discrete velocity such that $v_h = 0$ on $\Omega \setminus T$ and $v_h|_T = b_T \nabla p_h$, where

b_T is the local element bubble defined as above. Integration by parts then gives

$$0 = b(v_h, p_h) = -\int_\Omega \operatorname{div}(v_h) p_h \,\mathrm{d}x = \int_T b_T |\nabla p_h|^2 \,\mathrm{d}x,$$

from which we obtain that $p_h|_T = c$ on T (since b_T is a positive weighting). As p_h is continuous it has to be the same constant on the whole domain Ω, and as further $p_h \in L^2_0(\Omega, \mathbb{R})$ we have $c = 0$.

Example 3. Now we consider a method with a discontinuous pressure approximation. We choose the spaces

$$\begin{aligned} V_h &:= \{v_h \in H^1_0(\Omega, \mathbb{R}^d) : v_h|_T \in \mathbb{P}^d(T, \mathbb{R}^d)\ \forall T \in \mathcal{T}_h\}, \\ Q_h &:= \{q_h \in L^2_0(\Omega, \mathbb{R}) : q_h|_T \in \mathbb{P}^0(T, \mathbb{R})\ \forall T \in \mathcal{T}_h\}. \end{aligned}$$

Let $T_1, T_2 \in \mathcal{T}_h$ be two adjacent elements with common face $F \in \mathcal{F}_h$. Set v_h such that it vanished on $\Omega \setminus (T_1 \cup T_2)$ (i.e. in two dimensions where we have the second order Lagrangian finite element space for the velocity, just the edge bubble has a non zero coefficient). Now let $p_h^1 := p_h|_{T_1}$ and $p_h^2 := p_h|_{T_2}$ be the constant values on T_1 and T_2, respectively. Condition (2.15) gives

$$\begin{aligned} 0 = b(v_h, p_h) &= -\int_{T_1 \cup T_2} \operatorname{div}(v_h) p_h \,\mathrm{d}x \\ &= -p_h^1 \int_{T_1} \operatorname{div}(v_h) \,\mathrm{d}x - p_h^2 \int_{T_2} \operatorname{div}(v_h) \,\mathrm{d}x \\ &= -p_h^1 \int_F v_h \cdot n_1 \,\mathrm{d}s - p_h^2 \int_F v_h \cdot n_2 \,\mathrm{d}s = (p_h^2 - p_h^1) \int_F v_h \cdot n_1 \,\mathrm{d}s\,. \end{aligned}$$

Since v_h is equivalent to the edge bubble, the integral on the edge is not zero and we conclude that $p_h^1 = p_h^2$. This shows that p_h equals a global constant, and as $p_h \in L^2_0(\Omega)$ is has to vanish.

Example 4. The last example is also based on a discontinuous pressure approximation. Note, that this choice only works in two space dimensions (but a similar version also exists for $d = 3$). We choose the spaces

$$\begin{aligned} V_h &:= \{v_h \in H^1_0(\Omega, \mathbb{R}^2) : v_h|_T \in [\mathbb{P}^2(T, \mathbb{R}^2) + B(T, \mathbb{R}^2)]\ \forall T \in \mathcal{T}_h\}, \\ Q_h &:= \{q_h \in L^2_0(\Omega, \mathbb{R}) : q_h|_T \in \mathbb{P}^1(T, \mathbb{R})\ \forall T \in \mathcal{T}_h\}. \end{aligned}$$

The uniqueness follows with the same techniques as before.

After providing a simple check if a couple $V_h \times Q_h$ of finite element spaces is suitable,

i.e. provides a unique solution, the next two sections are dedicated to present a detailed stability analysis. Note, that since we only consider the case with homogeneous Dirichlet boundary conditions given on the whole boundary $\partial\Omega$ we will use at several point in the analysis the equivalence (see Theorem 3.2)

$$\|v\|_1 \sim \|\nabla v\|_0 \quad \forall v \in V.$$

2.5.1 Discrete stability by mesh dependent norms

Since the kernel ellipticity of the bilinear form a is inherited from the continuous setting, we aim to provide a proof for the discrete LBB condition given by

$$\sup_{v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_1} \gtrsim \|q_h\|_0 \quad \forall q_h \in Q_h. \tag{2.16}$$

It turns out that a simple technique for proving that (2.16) holds true is based on defining a new mesh dependent norm for the pressure space. To this end we define the norm

$$\|p_h\|_{0,h}^2 := \sum_{T \in \mathcal{T}_h} h^2 \|\nabla p_h\|_T^2 + \sum_{F \in \mathcal{F}_h} h \|[\![p_h]\!]\|_F^2 \quad \forall p_h \in Q_h,$$

where $[\![\cdot]\!]$ denotes the jump operator, as defined in (2.33a). The modified LBB condition now reads as

$$\sup_{v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_1} \gtrsim \|q_h\|_{0,h} \quad \forall q_h \in Q_h. \tag{2.17}$$

Before we provide a proof that the modified stability condition is sufficient, we introduce the so called Clément quasi interpolation operator. To this end let $V_i \in \mathcal{V}_h$ be the nodes of the triangulation. Then we define the vertex patch by

$$\omega_i := \bigcup_{T : V_i \in T} T.$$

For a function $v \in L^2(\omega_i)$ let $\Pi^0_{\omega_i} v$ be the L^2 projection onto constant functions on ω_i, i.e. we have

$$\Pi^0_{\omega_i} v := \frac{1}{|\omega_i|} \int_{\omega_i} v \, \mathrm{d}x \, .$$

Now let $I_{\mathcal{V}_h^{\text{int}}}$ denote the index set of nodes in the interior of Ω, and let $\varphi_i \in \mathbb{P}^1(\mathcal{T}_h)$ be the corresponding nodal hat functions. We define the Clément quasi interpolation operator $I_{\mathcal{C}}$

by

$$I_{\mathcal{C}} v := \sum_{i \in I_{\mathcal{V}_h^{\text{int}}}} (\Pi^0_{\omega_i} v) \varphi_i. \tag{2.18}$$

Note, that the result is a piece-wise linear polynomial, i.e. we have $I_{\mathcal{C}} v \in \mathbb{P}^1(\mathcal{T}_h)$.

Theorem 12. *Let $v \in H^1_0(\Omega)$. The Clément quasi interpolation operator is continuous, i.e. $\|I_{\mathcal{C}} v\|_1 \lesssim \|v\|_1$ and there holds the approximation result*

$$\left(\sum_{T \in \mathcal{T}_h} h^{-2} \|v - I_{\mathcal{C}} v\|^2_T + h^{-1} \|v - I_{\mathcal{C}} v\|^2_{\partial T} \right)^{1/2} \lesssim \|\nabla v\|_0.$$

Proof. The proof is based on the Bramble-Hilbert Lemma, standard scaling arguments and a partition of unity argument. A proof can be found for example in [8]. □

Remark 6. The operator $I_{\mathcal{C}}$ is called a quasi interpolation operator because $I_{\mathcal{C}} v_h = v_h$ does not hold true for all $v_h \in \mathbb{P}^1(\mathcal{T}_h)$.

Remark 7. In the case where we only have partial Dirichlet boundary condition, the definition of the Clément quasi interpolation operator considers all nodes in the interior and all nodes that are on non Dirichlet boundary parts.

Theorem 13. *Suppose that the couple $V_h \times Q_h$ fulfills the modified stability condition (2.17). Then (2.16) is valid*

Proof. Let $q_h \in Q_h$ be arbitrary. Since q_h is in Q we can use the continuous Stokes-LBB (2.12) to find a function $v \in H^1_0(\Omega, \mathbb{R}^2)$ such that $b(v, q_h) \geq C_1 \|v\|_1 \|q_h\|_0$. Now let $v_h := I_{\mathcal{C}} v \in V_h$ be the Clément interpolant of the continuous velocity v, then we have

$$b(v_h, q_h) = b(v, q_h) - b(v - v_h, q_h).$$

Using an element by element integration by parts argument and Cauchy-Schwarz yields

$$\begin{aligned} b(v - v_h, q_h) &= \sum_{T \in \mathcal{T}_h} \int_T \operatorname{div}(v_h - v) q_h \, \mathrm{d}x \\ &= - \sum_{T \in \mathcal{T}_h} \int_T (v_h - v) \cdot \nabla q_h \, \mathrm{d}x + \sum_{F \in \mathcal{F}_h} \int_F (v_h - v) \cdot n [\![q_h]\!] \, \mathrm{d}s \\ &\lesssim \left(\sum_{T \in \mathcal{T}_h} h^{-2} \|v_h - v\|^2_T + \sum_{F \in \mathcal{F}_h} h^{-1} \|(v_h - v) \cdot n\|^2_F \right)^{1/2} \|q_h\|_{0,h} \end{aligned}$$

Using the interpolation properties of the Clément operator, see Theorem 12, we finally get $b(v - v_h, q_h) \leq C_2\|v\|_1\|q_h\|_{0,h}$, thus in total

$$b(v_h, q_h) \geq (C_1\|q_h\|_0 - C_2\|q_h\|_{0,h})\|v\|_1.$$

By the continuity of the Clément operator $\|v_h\|_1 \leq C_3^{-1}\|v\|_1$ we obtain

$$\frac{b(v_h, q_h)}{\|v_h\|_1} \geq C_3(C_1\|q_h\|_0 - C_2\|q_h\|_{0,h}),$$

and thus with $\tilde{C}_i := C_iC_3$ we have

$$\sup_{v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_1} \geq \tilde{C}_1\|q_h\|_0 - \tilde{C}_2\|q_h\|_{0,h}.$$

Using the modified LBB condition, there exists a constant $\tilde{C}_3$ such that

$$\sup_{v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_1} \geq \tilde{C}_3\|q_h\|_{0,h},$$

and thus a convex combination with $0 \leq t \leq 1$ this finally gives

$$\begin{aligned}\sup_{v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_1} &= t \sup_{v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_1} + (1-t) \sup_{v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_1} \\ &\geq (t(\tilde{C}_3 + \tilde{C}_2) - \tilde{C}_2)\|q_h\|_{0,h} + (1-t)\tilde{C}_1\|q_h\|_0.\end{aligned}$$

By the choice $1 > t > \tilde{C}_2/(\tilde{C}_2 + \tilde{C}_3)$ we can conclude the proof. □

2.5.2 Examples of stable Stokes discretizations

We can now prove the stability proof for the methods discussed before. To this end we show that the modified LBB condition (2.17) holds, since Theorem 13 then provides stability.

Example 5 (The MINI element). We have the spaces

$$\begin{aligned}V_h &:= \{v_h \in H_0^1(\Omega, \mathbb{R}^d) : v_h|_T \in [\mathbb{P}^1(T, \mathbb{R}^d) + B(T, \mathbb{R}^d)]\ \forall T \in \mathcal{T}_h\}, \\ Q_h &:= \{q_h \in L_0^2(\Omega, \mathbb{R}) \cap C^0(\Omega, \mathbb{R}) : q_h|_T \in \mathbb{P}^1(T, \mathbb{R})\ \forall T \in \mathcal{T}_h\}.\end{aligned}$$

Now let $q_h \in Q_h$ be given. We choose $v_h \in V_h$ such that $v_h|_T := -h^2 b_T \nabla q_h$ for all elements $T \in \mathcal{T}_h$. This choice is possible because we augmented the velocity space with the local element bubbles. Since we consider a continuous pressure approximation, integration by

parts gives

$$b(v_h, q_h) = -\int_\Omega \operatorname{div}(v_h) q_h \,\mathrm{d}x = \int_\Omega v_h \cdot \nabla q_h \,\mathrm{d}x$$
$$= \sum_{T\in\mathcal{T}_h} h^2 \|b_T^{1/2} \nabla q_h\|_T^2 \gtrsim \sum_{T\in\mathcal{T}_h} h^2 \|\nabla q_h\|_T^2 = \|q_h\|_{0,h}^2.$$

Using that $v_h \in H_0^1(T, \mathbb{R}^d)$, we have on each element the estimate (use scaling arguments)

$$\|\nabla v_h\|_T \lesssim h^{-1} \|v_h\|_T \lesssim h \|b_T \nabla q_h\| \lesssim h \|\nabla q_h\|_T,$$

and so $\|v_h\|_1 \lesssim \|\nabla v\|_0 \lesssim \|q_h\|_{0,h}$, which proves that (2.17) holds true.

Example 6 (The $\mathbb{P}2\mathbb{P}0$ element). Consider the case $d = 2$. We have the spaces

$$V_h := \{v_h \in H_0^1(\Omega, \mathbb{R}^2) : v_h|_T \in \mathbb{P}^2(T, \mathbb{R}^2)\ \forall T \in \mathcal{T}_h\},$$
$$Q_h := \{q_h \in L_0^2(\Omega, \mathbb{R}) : q_h|_T \in \mathbb{P}^0(T, \mathbb{R})\ \forall T \in \mathcal{T}_h\}.$$

Let $q_h \in Q_h$ be arbitrary and set $v_h|_F := h b_F [\![q_h]\!] n$, where b_F is the edge bubble. Note, that this choice was only possible because we included the element bubble in the velocity space. This shows why the $\mathbb{P}^1 \times \mathbb{P}^0$ combination does not work. The above choice then gives

$$b(v_h, q_h) = -\int_\Omega \operatorname{div}(v_h) q_h \,\mathrm{d}x = -\sum_{T\in\mathcal{T}_h} q_h|_T \int_T \operatorname{div}(v_h) \,\mathrm{d}x$$
$$= \sum_{F\in\mathcal{F}_h} \int_F v_h \cdot n [\![q_h]\!] \,\mathrm{d}s = \sum_{F\in\mathcal{F}_h} h \int_F b_F |[\![q_h]\!]|^2 \,\mathrm{d}s \sim \|q_h\|_{0,h}^2.$$

The inverse inequality, see Theorem 1, and scaling then also gives $\|v_h\|_1 \leq \|q_h\|_{0,h}$.

Example 7 (The Bernardi Raugel (BR) element). Consider the case $d = 2$. Above example shows, that we only need to control the normal velocity at the edge, i.e. adding the edge bubble for both components of the velocity seems to be sub optimal (with respect to computational costs and the expected approximation properties). The idea now is to only add the normal edge bubble. To this ed we define

$$B_n(\mathcal{T}_h) := \{v_h \in H_0^1(\Omega, \mathbb{R}^2) \cap \mathbb{P}^2(\mathcal{T}_h, \mathbb{R}^2) : v_h|_F = c b_F n, c \in \mathbb{R}, \forall F \in \mathcal{F}_h\}.$$

Then we set

$$V_h := \{v_h \in H_0^1(\Omega, \mathbb{R}^2) : v_h|_T \in \mathbb{P}^1(T, \mathbb{R}^2) \forall T \in \mathcal{T}_h\} \cup B_n(\mathcal{T}_h),$$
$$Q_h := \{q_h \in L_0^2(\Omega, \mathbb{R}) : q_h|_T \in \mathbb{P}^0(T, \mathbb{R}) \; \forall T \in \mathcal{T}_h\}.$$

The proof for the stability follows as before.

Example 8 (The $\mathbb{P}3\mathbb{P}0$ element). Consider the case $d = 3$. We have the spaces

$$V_h := \{v_h \in H_0^1(\Omega, \mathbb{R}^2) : v_h|_T \in \mathbb{P}^3(T, \mathbb{R}^3) \; \forall T \in \mathcal{T}_h\},$$
$$Q_h := \{q_h \in L_0^2(\Omega, \mathbb{R}) : q_h|_T \in \mathbb{P}^0(T, \mathbb{R}) \; \forall T \in \mathcal{T}_h\}.$$

The proof follows with the same steps as before and is given as an exercise for the reader.

Example 9 (The $\mathbb{P}2$-bubble element). Consider the case $d = 2$. We choose the spaces

$$V_h := \{v_h \in H_0^1(\Omega, \mathbb{R}^2) : v_h|_T \in [\mathbb{P}^2(T, \mathbb{R}^2) + B(T, \mathbb{R}^2)] \; \forall T \in \mathcal{T}_h\},$$
$$Q_h := \{q_h \in L_0^2(\Omega, \mathbb{R}) : q_h|_T \in \mathbb{P}^1(T, \mathbb{R}) \; \forall T \in \mathcal{T}_h\}.$$

We combine the results from before. Let $q_h \in Q_h$ be arbitrary. On each element T we can decompose $q_h = q_h^0 + q_h^1$ such that $q_h^0 \in \mathbb{P}^0(T, \mathbb{R})$ and $q_h^1 \in \mathbb{P}^1(T, \mathbb{R}) \cap L_0^2(T, \mathbb{R})$, i.e. we set

$$q_h^0|_T := \frac{\int_T q_h \, \mathrm{d}x}{|T|}.$$

The idea now is to use the additional bubble to control the element wise linear polynomials with vanishing mean value, and the edge dofs to control the constants. From the stability (i.e. surjectivity of the divergence) of the $\mathbb{P}2\mathbb{P}0$ element there exists a function $v_h^0 \in V_h$ such that

$$b(v_h^0, q_h^0) = \|q_h^0\|_0^2 \quad \text{and} \quad \|v_h^0\|_1 \leq C_0 \|q_h^0\|_0.$$

Next, using the stability result of the MINI element (on each element separately) we find another function $v_h^1 \in V_h$ such that (by scaling we can use the same constant C_0 here)

$$b(v_h^1, q_h^1) = \|q_h^1\|_T^2 \quad \text{and} \quad \|v_h^1\|_1 \leq C_0 \|q_h^1\|_0.$$

Note, that it was crucial that $q_h^1 \in L_0^2(T)$. Further, since v_h^1 vanishes on the boundary of

each element we have

$$b(v_h^1, q_h^0) = 0.$$

Now set $v_h := v_h^1 + \alpha v_h^0$ where $\alpha > 0$ is a constant yet to be set. By Cauchy Schwarz and Young's inequality we have

$$\begin{aligned} b(v_h, q_h) &= b(v_h^1, q_h^1) + \alpha b(v_h^0, q_h^0) + \alpha b(v_h^0, q_h^1) \\ &= \|q_h^1\|_0^2 + \alpha \|q_h^0\|_0^2 + \alpha b(v_h^0, q_h^1) \\ &\gtrsim \|q_h^1\|_0^2 + \alpha \|q_h^0\|_0^2 - \alpha \|v_h^0\|_1 \|q_h^1\|_0 \\ &\gtrsim \|q_h^1\|_0^2 + \alpha \|q_h^0\|_0^2 - \frac{\alpha\varepsilon}{2} \|v_h^0\|_1^2 - \frac{\alpha}{\varepsilon 2} \|q_h^1\|_0^2 \\ &\gtrsim (1 - \frac{\alpha}{\varepsilon 2}) \|q_h^1\|_0^2 + \alpha (1 - \frac{\varepsilon}{2C_0^2}) \|q_h^0\|_0^2 \end{aligned}$$

Hence, in a first step we choose ε such that $1 - \varepsilon/(2C_0^2) > 0$, and then α such that $1 - \alpha/(2\varepsilon) > 0$, which gives

$$b(v_h, q_h) \gtrsim (\|q_h^0\|_0^2 + \|q_h^1\|_0^2) \gtrsim \|q_h\|_0^2.$$

Since we also have $\|v_h\|_1 \lesssim \|q_h\|_0$, we have proven the stability (here without using directly the modified LBB).

Example 10 (Taylor-Hood element). In all above examples it was possible to prove the stability by a local construction of the discrete velocity. Unfortunately, this is not possible for the famous element called Taylor-Hood element. Here we choose the spaces

$$\begin{aligned} V_h &:= \{v_h \in H_0^1(\Omega, \mathbb{R}^d) : v_h|_T \in \mathbb{P}^2(T, \mathbb{R}^d) \; \forall T \in \mathcal{T}_h\}, \\ Q_h &:= \{q_h \in L_0^2(\Omega, \mathbb{R}) \cap C^0(\Omega, \mathbb{R}) : q_h|_T \in \mathbb{P}^1(T, \mathbb{R}) \; \forall T \in \mathcal{T}_h\}, \end{aligned}$$

thus, similar to the MINI element, we consider a continuous pressure approximation. The stability analysis is based on the construction of a Fortin interpolation operator (see next section) and is based on a macro element technique. See for example in [33].

2.5.3 Discrete stability by Fortin-Interpolation operators

Another very common technique to prove discrete stability of a finite element method is based on the introduction of a Fortin operator denoted by I_F.

Theorem 14 (Fortin operator)**.** *Assume there exists an operator* $I_F : V \to V_h$ *such that*

$$b(I_F v, q_h) = b(v, q_h) \quad \forall q_h \in Q_h, \quad \text{and} \quad \|I_F v\|_1 \lesssim \|v\|_1.$$

Then the discrete LBB condition (2.16) *follows from the continuous LBB condition* (2.12)*.*

Proof. Using the above properties we get

$$\sup_{v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_1} \gtrsim \sup_{v \in V} \frac{b(I_F v, q_h)}{\|I_F v\|_1} = \sup_{v \in V} \frac{b(v, q_h)}{\|I_F v\|_1} \gtrsim \sup_{v \in V} \frac{b(v, q_h)}{\|v\|_1} \gtrsim \|q_h\|_0 \quad \forall q_h \in Q_h,$$

where we used (2.12) in the last step (since $Q_h \subset Q$). □

Note that the construction of a Fortin operator has to be done for each discretization separately. As we will see, this will be done with the same techniques that we already used in the previous section.

Example 11 (The $\mathbb{P}2\mathbb{P}0$ element)*.* Consider the case $d = 2$. We have the spaces

$$V_h := \{v_h \in H_0^1(\Omega, \mathbb{R}^2) : v_h|_T \in \mathbb{P}^2(T, \mathbb{R}^2)\ \forall T \in \mathcal{T}_h\},$$
$$Q_h := \{q_h \in L_0^2(\Omega, \mathbb{R}) : q_h|_T \in \mathbb{P}^0(T, \mathbb{R})\ \forall T \in \mathcal{T}_h\}.$$

The construction of a Fortin operator is split into two steps. First, let $I_F^1 := I_C$ be given as the Clément operator. Note, that I_F^1 only gives a linear approximation, i.e. we have only defined the nodal values. Next, we choose I_F^2 to be the operator defined by the equations

$$I_F^2 v(x_V) = 0 \quad \forall x_V \in \mathcal{V}_h,$$
$$\int_F I_F^2 v \cdot n \, \mathrm{d}s = \int_F v \cdot n \, \mathrm{d}s \quad \forall F \in \mathcal{F}_h$$

Note, that this can be done by setting

$$I_F^2 v := \sum_{F \in \mathcal{F}_h} \frac{\int_F v \cdot n \, \mathrm{d}s}{\int_F b_F \cdot n \, \mathrm{d}s} b_F,$$

where b_F is (now a vector valued) edge bubble. Next note, that $\int_F b_F \cdot n \, \mathrm{d}s \sim h$ and by a standard scaling arguments $\|\nabla b_F\|_T \sim 1$ and

$$\|u \cdot n\|_F^2 \lesssim h^{-1}\|u\|_T^2 + h\|\nabla u\|_T^2.$$

Thus, in total we get (using Cauchy Schwarz)

$$\|\nabla I_F^2 v\|_T^2 \lesssim \sum_{F\in\mathcal{F}_h} \frac{1}{h^2}\Big(\int_F u\cdot n\,\mathrm{d}s\Big)^2$$
$$\sum_{F\in\mathcal{F}_h} \frac{h}{h^2}\int_F (u\cdot n)^2\,\mathrm{d}s \lesssim h^{-2}\|u\|_T^2 + \|\nabla u\|_T^2.$$

Combining these two operators we define the Fortin operator as

$$I_F v := I_F^1 v + I_F^2(v - I_F^1 v).$$

Now let $q_h \in Q_h$ be arbitrary, then we have on each element using the Gaussian theorem (since q_h is a piece wise constant)

$$\begin{aligned}\int_T \operatorname{div}(I_F v) q_h\,\mathrm{d}x &= q_h\int_{\partial T} I_F v\cdot n\,\mathrm{d}s = \\ &= q_h\sum_{F\subset\partial T}\int_F I_F v\cdot n\,\mathrm{d}s \\ &= q_h\sum_{F\subset\partial T}\int_F I_F^1 v\cdot n\,\mathrm{d}s + \int_F I_F^2(v - I_F^1 v)\cdot n\,\mathrm{d}s \\ &= q_h\sum_{F\subset\partial T}\int_F I_F^1 v\cdot n\,\mathrm{d}s + \int_F (v - I_F^1 v)\cdot n\,\mathrm{d}s \\ &= q_h\int_{\partial T} v\cdot n\,\mathrm{d}s = \int_T \operatorname{div}(v) q_h\,\mathrm{d}x\,.\end{aligned}$$

Further, by Theorem 12 we get on each element with above estimates

$$\begin{aligned}\|\nabla I_F v\|_T &\le \|\nabla I_F^1 v\|_T + \|\nabla I_F^2(v - I_F^1 v)\|_T \\ &\le \|\nabla v\|_T + \frac{1}{h}\|(v - I_F^1 v)\|_T + \|\nabla(v - I_F^1 v)\|_T \le \|\nabla v\|_T.\end{aligned}$$

The construction of a Fortin operator for the other elements follows with very similar ideas and will be left as examples for the reader.

2.5.4 Stabilized methods

In the previous section we saw that the choice of the discretization spaces V_h and Q_h is not straight forward and, with respect to the discrete LBB condition (2.16), stability can either be forced by decreasing the dimension of pressure space or increasing the dimension of the velocity space. In this section we will introduce the idea of stabilization techniques.

The idea is to use a pair $V_h \times Q_h$ that is not inf-sup stable but can be made well posed by weakening the incompressibility constraint such that $\operatorname{div} u_h = g_h$ for some appropriate g_h. The stabilization can also be motivated by looking at the saddle point structure of the discrete problem given by

$$\begin{pmatrix} A & B^{\mathrm{T}} \\ B & 0 \end{pmatrix} \begin{pmatrix} u_h \\ p_h \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}.$$

Here A and B represent the finite element matrices of the bilinear forms a and b, respectively. A simple calculation shows that the pressure Schur complement is given by

$$BA^{-1}B^{\mathrm{T}}p = BA^{-1}f,$$

where we assumed that A is invertible which is fulfilled due to the ellipiticity of the bilinear form a on the whole space $V_h \subset V$. Here we can see that a discrete method is well posed if and only if the symmetric positive semi-definite matrix $BA^{-1}B^{\mathrm{T}}$ only has the constants in the null space which is the same constraint as Theorem 11. The idea of a stabilization is now to replace the above matrix by

$$\begin{pmatrix} A & B^{\mathrm{T}} \\ B & 0 \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} A & B^{\mathrm{T}} \\ B & -\beta C \end{pmatrix},$$

which gives the modified Schur complement

$$BA^{-1}B^{\mathrm{T}}p + \beta Cp = BA^{-1}f.$$

The motivation now is that the stabilization βC allows to remove non constant pressure which lie in the kernel of the original Schur complement. In order to motivate the structure of C we will revisit the MINI finite element. To this end we define the space

$$\begin{aligned} V_h^l &:= H_0^1(\Omega, \mathbb{R}^d) \cap \mathbb{P}^1(\mathcal{T}_h, \mathbb{R}^d) \quad \text{and} \quad V_h^b := \bigcup_{T \in \mathcal{T}_h} B(T, \mathbb{R}^d) \\ V_h &:= V_h^l \oplus V_h^b \\ Q_h &:= L_0^2(\Omega, \mathbb{R}) \cap \mathbb{P}^1(\mathcal{T}_h, \mathbb{R}). \end{aligned}$$

Here, V_h^l represent the low order space of linear approximations and V_h^b is the space of local bubbles. Now let u_h be the solution of the discrete Stokes problem, then we can split the solution into $u_h := u_h^l + u_h^b$ where $u_h^l \in V_h^l$ and $u_h^b \in V_h^b$. It turns out, that the two parts fulfill an orthogonality property in the momentum balance. To this end let $v_h^l \in V_h^l$

be a linear test function, then since the bubbles vanish at element interfaces we get using integration by parts

$$a(u_h^b, v_h^l) = \sum_{T\in\mathcal{T}_h} \nu \int_T \varepsilon(u_h^b) : \varepsilon(v_h^l)\,\mathrm{d}x = \sum_{T\in\mathcal{T}_h} -\nu \int_T u_h^b \cdot \operatorname{div}(\varepsilon(v_h^l))\,\mathrm{d}x = 0, \tag{2.19}$$

and thus we have

$$a(u_h^l, v_h^l) + b(v_h^l, p_h) = (f, v_h^l) \quad \forall v_h^l \in V_h^l.$$

Now let $c_T \in \mathbb{R}^d$ be the coefficient of the solution of u_h^b such that

$$u_h^b = \sum_{T\in\mathcal{T}_h} c_T b_T \in V_h^b,$$

where b_T are the (scalar) bubble functions on each element T. Using that the discrete pressure is continuous, integration by parts and choosing a bubble $b_{T'}$, where $T' \in \mathcal{T}_h$ is arbitrary, as test function we get with (2.19) in the momentum equation

$$\begin{aligned} a(u_h, b_{T'}) + b(b_{T'}, p_h) &= a(u_h^b, b_{T'}) + (b_{T'}, \nabla p_h) \\ &= \int_{T'} \nu c_{T'} |\varepsilon(b_{T'})|^2\,\mathrm{d}x + \int_{T'} b_{T'} \cdot \nabla p_h\,\mathrm{d}x = \int_{T'} f \cdot b_{T'}\,\mathrm{d}x\,. \end{aligned}$$

Since this can be done for all elements separately, we get an explicit formula for the coefficients given by

$$c_T := \frac{\int_T (f - \nabla p_h)\cdot b_T\,\mathrm{d}x}{\int_T \nu |\varepsilon(b_{T'})|^2\,\mathrm{d}x} \quad \forall T \in \mathcal{T}_h.$$

For the ease let us define $\gamma_T := (\int_T |\varepsilon(b_T)|^2\,\mathrm{d}x)^{-1}$, then the incompressibility constraint gives for all $q_h \in Q_h$

$$\begin{aligned} 0 = b(u_h, q_h) &= b(u_h^l, q_h) + b(u_h^b, q_h) \\ &= b(u_h^l, q_h) + \sum_{T\in\mathcal{T}_h} \int_T c_T b_T \cdot \nabla q_h\,\mathrm{d}x \\ &= b(u_h^l, q_h) + \sum_{T\in\mathcal{T}_h} \gamma_T \left(\int_T b_T \cdot \nabla q_h\,\mathrm{d}x\right)\left(\int_T (f - \nabla p_h)\cdot b_T\,\mathrm{d}x\right) \end{aligned}$$

In total this shows, that the linear part $(u_h^l, p_h) \in V_h^l \times Q_h$ of the solution of MINI finite element method solves the problem

$$\begin{aligned} a(u_h^l, v_h^l) + b(v_h^l, p_h) &= (f, v_h^l) \quad \forall v_h^l \in V_h^l \\ b(u_h^l, q_h) - \sum_{T\in\mathcal{T}_h} \tilde{\gamma}_T \int_T b_T^2 \nabla p_h \cdot \nabla q_h \,\mathrm{d}x &= \sum_{T\in\mathcal{T}_h} \gamma_T \int_T b_T \cdot \nabla q_h \,\mathrm{d}x \int_{T'} f \cdot b_T \,\mathrm{d}x \quad \forall q_h \in Q_h, \end{aligned}$$

where we used that ∇p_h and ∇q_h are constant and $\tilde{\gamma}_T := \gamma_T |T|^{-1} (\int_T b_T \,\mathrm{d}x)^2$. This can be interpreted as a $P^1 \times P^1$ approximation of the partial differential equation

$$\begin{aligned} -\nu \operatorname{div}(\varepsilon(u)) + \nabla p &= f \\ \operatorname{div}(u) - \rho \Delta p &= -\rho \operatorname{div}(f), \end{aligned}$$

with some constant ρ. Note, that since u_h^l is linear, we may also add the additional term $\operatorname{div}(\varepsilon(v_h^l))$ to the left hand side of the second equations. Now let $V_h \times Q_h$ be arbitrary. Since $\gamma_T \sim h^2$ and $b_T = \mathcal{O}(1)$ above derivations motivates to define for all $(u_h, p_h), (v_h, q_h) \in V_h \times Q_h$ the bilinear form

$$\begin{aligned} c((u_h, p_h), (v_h, q_h)) =& \alpha \sum_{T\in\mathcal{T}_h} h^2 \int_T (-\nu \operatorname{div}(\varepsilon(u_h)) + \nabla p_h) \cdot (-\nu \operatorname{div}(\varepsilon(v_h)) + \nabla q_h) \,\mathrm{d}x \\ &+ \beta \sum_{F\in\mathcal{F}_h} h \int_F [\![q_h]\!][\![p_h]\!] \,\mathrm{d}s \,. \end{aligned}$$

Note, that the jump term is only essential for a lowest order discontinuous pressure approximation and when the velocity space does not contain polynomials of order d, i.e. we have

$$Q_h \subset C(\Omega) \quad \text{or} \quad \mathbb{P}^d(\Omega, \mathbb{R}^d) \cap H^1(\Omega, \mathbb{R}^d) \subset V_h \quad \Rightarrow \quad \beta = 0.$$

Then we have the stabilized problem: Find $(u_h, p_h) \in V_h \times Q_h$ such that

$$\begin{aligned} &a(u_h, v_h) + b(u_h, q_h) + b(v_h, p_h) - c((u_h, p_h), (v_h, q_h)) \\ &\qquad = (f, v_h) - \sum_{T\in\mathcal{T}_h} \alpha h^2 \int_T f \cdot (-\nu \operatorname{div}(\varepsilon(v_h)) + \nabla q_h) \,\mathrm{d}x \quad \forall (v_h, q_h) \in V_h \times Q_h. \end{aligned}$$

Note, that stability of above method then depends on a proper choice of the stabilization parameters α and β. The proof follows similar steps as the proof of Theorem 13. A detailed analysis is presented in Chapter 4 of [22]. Further note, that the new stabilized terms are

consistent, i.e. for the exact solution $u_h = u$ and $p_h = p$ we have

$$-c((u_h, p_h), (v_h, q_h)) = -\sum_{T \in \mathcal{T}_h} \alpha h^2 \int_T f \cdot (-\nu \operatorname{div}(\varepsilon(v_h)) + \nabla q_h) \, \mathrm{d}x \quad \forall (v_h, q_h) \in V_h \times Q_h,$$

thus the exact solution still solves above stabilized problem.

2.5.5 Error analysis

In this section we derive a priori error estimates for the solution of the discrete problem (2.13). Similarly as for standard elliptic problems, the derivation is based on a best approximation result and by means of appropriate interpolation operators. Similarly as for the continuous setting we define the space of discrete divergence-free velocity functions

$$V_{0,h} := \{v_h \in V_h : b(v_h, q_h) = 0 \; \forall q_h \in Q_h\}.$$

Further, assuming stability, let $\beta_{2,h}$ be the discrete LBB condition in (2.16).

Lemma 1. *Let $(u, p) \in V \times Q$ be the exact solution of weak formulation of the Stokes equation (2.7), and let $(u_h, p_h) \in V_h \times Q_h$ be the discrete solution of (2.13). There holds the best approximation result*

$$\|u - u_h\|_1 \lesssim \inf_{v_h \in V_{h,0}} \|u - v_h\|_1 + \frac{1}{\nu} \inf_{q_h \in Q_h} \|p - q_h\|_0.$$

If there holds the kernel inclusion property $V_{0,h} \subset V_0$ we further have

$$\|u - u_h\|_1 \leq \inf_{v_h \in V_{h,0}} \|u - v_h\|_1.$$

Proof. Let $v_h \in V_{0,h}$ be arbitrary. In a first step we use the triangle inequality to get

$$\|u - u_h\|_1 \leq \|u - v_h\|_1 + \|v_h - u_h\|_1.$$

Since we also have $u_h \in V_{h,0}$ we get for the difference $v_h - u_h$ by the coercivity of the bilinear form a

$$\nu \|v_h - u_h\|_1^2 \lesssim a(v_h - u_h, v_h - u_h) = a(v_h - u, v_h - u_h) + a(u - u_h, v_h - u_h).$$

For the second term on the right hand side we get by linearity

$$\begin{aligned} a(u-u_h, v_h-u_h) &= a(u, v_h-u_h) - a(u_h, v_h-u_h) \\ &= (f, v_h-u_h) - b(v_h-u_h, p) - (f, v_h-u_h) - b(v_h-u_h, p_h) \\ &= -b(v_h-u_h, p-p_h). \end{aligned}$$

Next, let $q_h \in Q_h$ be arbitrary. Since $v_h - u_h \in V_{h,0}$ we can write

$$b(v_h-u_h, p-p_h) = b(v_h-u_h, p-q_h),$$

and thus in total

$$\begin{aligned} \nu\|v_h-u_h\|_1^2 &\leq a(v_h-u, v_h-u_h) - b(v_h-u_h, p-q_h) \\ &\lesssim \nu\|v_h-u\|_1\|v_h-u_h\|_1 + \nu\|v_h-u_h\|_1\frac{1}{\nu}\|p-q_h\|_0. \end{aligned}$$

Dividing by $\nu\|v_h-u_h\|_1$ gives the first result. The second estimate follows with the same steps and using $b(v_h-u_h, p-p_h) = 0$ which follows from $V_{0,h} \subset V_0$. □

Lemma 2. *Let $(u,p) \in V \times Q$ be the exact solution of weak formulation of the Stokes equation (2.7), and let $(u_h, p_h) \in V_h \times Q_h$ be the discrete solution of (2.13). There holds the best approximation result*

$$\inf_{v_h \in V_{h,0}} \|u-v_h\|_1 \lesssim \left(1 + \frac{1}{\beta_{2,h}}\right) \inf_{v_h \in V_h} \|u-v_h\|_1.$$

Proof. We aim to follow similar steps as in the proof of the Brezzi theorem. To this end let $w_h \in V_h$ be arbitrary. We solve the variational problem: Find $r_h \in V_h$ such that

$$b(r_h, q_h) = b(u-w_h, q_h) \quad \forall q_h \in Q_h.$$

Note that this problem admits a (non unique!) solution due to the discrete LBB condition (2.16) with the stability estimate

$$\|r_h\|_1 \leq \beta_{2,h}^{-1}\|b(u-w_h, \cdot)\|_{Q_h^*} = \beta_{2,h}^{-1} \sup_{q_h \in Q_h} \frac{b(u-w_h, q_h)}{\|q_h\|_0} \lesssim \beta_{2,h}^{-1}\|u-w_h\|_1.$$

Now let $v_h = r_h + w_h$, and observe

$$b(v_h, q_h) = b(r_h, q_h) + b(w_h, q_h) = b(u, q_h) - b(w_h, q_h) + b(w_h, q_h) = 0,$$

which shows that $v_h \in V_{h,0}$. Together with the estimate

$$\|u - v_h\|_1 \leq \|u - w_h\|_1 + \|r_h\|_1 \leq (1 + \beta_{2,h}^{-1})\|u - w_h\|_1,$$

we conclude the proof. □

Lemma 3. *Let $(u, p) \in V \times Q$ be the exact solution of weak formulation of the Stokes equation (2.7), and let $(u_h, p_h) \in V_h \times Q_h$ be the discrete solution of (2.13). There holds the best approximation result*

$$\|p - p_h\|_0 \lesssim (1 + \beta_{2,h}^{-1}) \inf_{q_h \in Q_h} \|p - q_h\|_0 + \nu\beta_{2,h}^{-1}\|u - u_h\|_1.$$

Proof. Since $V_h \subset V$ and $Q_h \subset Q$, Galerkin orthogonality gives for all $v_h \in V_h$ the equation $b(v_h, p - p_h) = -a(u - u_h, v_h)$ and thus

$$b(v_h, q_h - p_h) = -a(u - u_h, v_h) - b(v_h, p - q_h).$$

Using the discrete LBB condition (2.16) then provides the estimate

$$\begin{aligned}\beta_{2,h}\|q_h - p_h\|_0 &\leq \sup_{v_h \in V_h} \frac{b(v_h, q_h - p_h)}{\|v_h\|_1} \\ &= \sup_{v_h \in V_h} \frac{-a(u - u_h, v_h) - b(v_h, p - q_h)}{\|v_h\|_1} \lesssim \nu\|u - u_h\|_1 + \|p - q_h\|_0.\end{aligned}$$

By the triangle inequality we finally get

$$\begin{aligned}\|p - p_h\|_0 &\leq \|p - q_h\|_0 + \|q_h - p_h\|_0 \\ &\lesssim (1 + \beta_{2,h}^{-1})\|p - q_h\|_0 + \nu\beta_{2,h}^{-1}\|u - u_h\|_1.\end{aligned}$$

□

Theorem 15 (Best approximation)**.** *Let $(u, p) \in V \times Q$ be the exact solution of weak formulation of the Stokes equation (2.7), and let $(u_h, p_h) \in V_h \times Q_h$ be the discrete solution of (2.13). There holds the best approximation result*

$$\|u - u_h\|_1 + \nu^{-1}\|p - p_h\|_0 \lesssim \inf_{v_h \in V_h} \|u - v_h\|_1 + \nu^{-1} \inf_{q_h \in Q_h} \|p - q_h\|_0.$$

In order to study the convergence orders, we introduce appropriate interpolation operators. In the case of a conforming discretization, these are given by the standard nodal

Lagrange interpolation operator for the velocity space, and the L^2-projection for the discrete pressure space.

Theorem 16 (Interpolation operator)**.** *Assume that the discrete velocity space includes polynomials of order k_V, and the discrete pressure space polynomials of order k_Q, i.e. we have*

$$\mathbb{P}^{k_V}(\mathcal{T}_h, \mathbb{R}^d) \cap V \subset V_h \quad and \quad \mathbb{P}^{k_Q}(\mathcal{T}_h, \mathbb{R}) \cap Q \subset Q_h$$

Assume that $(u, p) \in H^l(\Omega, \mathbb{R}^d) \times H^r(\Omega, \mathbb{R})$. There exists interpolation operators I_V and I_Q such that

$$\|u - I_V u\|_1 \lesssim h^s \|u\|_{s+1}, \quad and \quad \|p - I_Q p\|_0 \lesssim h^t \|p\|_t,$$

where $s = \min(k_V, l-1)$ and $t = \min(k_Q + 1, r)$.

Proof. Let I_V be the standard Lagrange interpolation operator and let I_Q be defined as the L^2-projection. The result follows by scaling arguments and the Bramble-Hilbert lemma, see for example in [13]. □

In view of the best approximation results given by Theorem 15 and the interpolation results, we see that there is a relation between the approximation order of the velocity space k_V and the order of the pressure space k_Q. To see this, let $r = l - 1$, thus assume the regularity $(u, p) \in H^l(\Omega, \mathbb{R}^d) \times H^{l-1}(\Omega, \mathbb{R})$, then we have the convergence results

$$\|u - u_h\|_1 + \nu^{-1} \|p - p_h\|_0 \lesssim h^s (\|u\|_{s+1} + \frac{1}{\nu} \|p\|_s),$$

where $s = \min(k_V, k_Q + 1, l - 1)$. This shows that, in an optimal setting, the pressure order is one order smaller compared to the velocity error. In Table 2.1 we can see the expected order of convergence for several Stokes discretizations. Note, that k_V and k_Q correspond to the polynomial orders that are **completely** (locally on each element) included in the corresponding approximation spaces.

2.5.6 Pressure robustness

This section deals with a property of Stokes discretizations called “pressure robustness” which was first discussed in the work [31]. Before revealing the mechanisms in detail we aim, to motivate pressure robustness in the following.

With respect to the error estimates and the best approximation results of Theorem 15 in the previous section we see, that the velocity error depends on the pressure error with a

element	k_V	k_Q	conv. order
$\mathbb{P}2\mathbb{P}0$	2	0	$\mathcal{O}(h)$
MINI	1	1	$\mathcal{O}(h)$
BR	1	1	$\mathcal{O}(h)$
stab. $\mathbb{P}1\mathbb{P}1$	1	1	$\mathcal{O}(h)$
stab. $\mathbb{P}2\mathbb{P}2$	2	2	$\mathcal{O}(h^2)$
$\mathbb{P}2$-bubble	2	1	$\mathcal{O}(h^2)$
Taylor-Hood	2	1	$\mathcal{O}(h^2)$

Table 2.1: Expected convergence order for various Stokes elements

scaling factor ν^{-1}. This shows, that there might occur a blow up in the case of a vanishing viscosity $\nu \to 0$. For a closer investigation we consider a simple example. Let $\Omega = (0,1)^2$ and $f = -\operatorname{div}(\nu\varepsilon(u)) + \nabla p$ with the exact solutions

$$u = \operatorname{curl}(\psi), \quad \text{and} \quad p := x^5 + y^5 - \frac{1}{3},$$

where the potential is given by $\psi := x^2(x-1)^2y^2(y-1)^2$. In Figure 2.3 we compare the H^1-semi norm error $\|\nabla u - \nabla u_h\|_0$ for the standard non pressure robust Taylor-Hood (TH) element of order $k = 2$ (for the velocity) and a pressure robust method abbreviated by MCS (mass conserving mixed stress methods, see [17]). Note, that although the Taylor-Hood element provides optimal orders of convergence, we see that the error shows the unwanted scaling with respect to ν and can get arbitrary big.

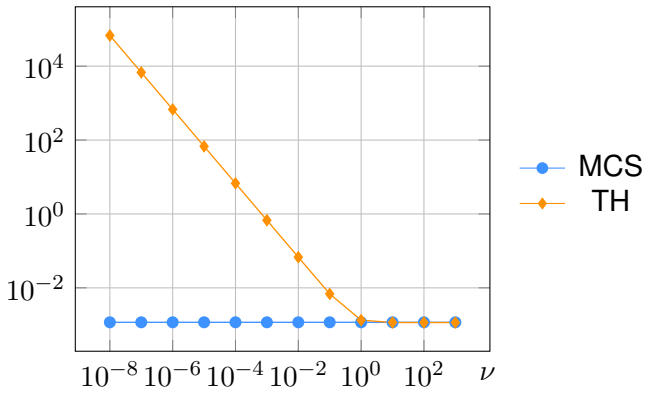


Figure 2.3: The H^1-seminorm error for the MCS method and a Taylor-Hood approximation for varying viscosities ν.

To identify the problem, we consider now a more general setting with an arbitrary domain Ω. We want to solve the Stokes problem (2.7) with homogeneous Dirichlet boundary

conditions where the right hand side is given by a gradient field, i.e. we have $f := \nabla\Psi$. Using integration by parts we see that the exact solution is given by $(0, \Psi)$ as

$$a(0, v) + b(v, \Psi) = -\int_\Omega \operatorname{div}(v)\Psi \,\mathrm{d}x = \int_\Omega v \cdot \nabla\Psi \,\mathrm{d}x = (f, v) \quad \forall v \in V,$$

and $b(0, q) = 0$ for all $q \in Q$. This shows that arbitrary gradient fields are totally balanced by the pressure. The question that arises is, if this physical property is also given in the discrete setting, thus if the discrete solution is given by $(0, \Pi_{Q_h}\Psi)$, where Π_{Q_h} is the L^2-projection onto the discrete pressure space. The problem can be easily seen if the discrete system is tested with a discretely divergence free test function $v_h \in V_{0,h}$. Similarly as before, integration by parts (now on the right side) gives

$$-\int_\Omega \operatorname{div}(v_h)\Pi_{Q_h}\Psi \,\mathrm{d}x = -\int_\Omega \Psi \operatorname{div}(v_h) \,\mathrm{d}x\,.$$

Since v_h is discretely divergence-free and $\Pi_{Q_h}\Psi \in Q_h$, the left hand side vanishes. Nevertheless, the right hand side only vanishes if either $\Psi \in Q_h$ or if v_h is also exactly divergence-free, thus if the Stokes discretization fulfills the kernel inclusion property $V_{0,h} \subset V_0$. Indeed, Lemma 1 and Lemma 2 show that one can then deduce a velocity error estimate

$$\|u - u_h\|_1 \leq \inf_{v_h \in V_h} \|u - v_h\|_1.$$

that is independent of the best approximation of the pressure and independent of the viscosity. In general, the author of [31] calls a finite element method for the Stokes problem **pressure robust** if one can deduce a pressure independent velocity error estimate. Note, that this then also corresponds to the structure preserving property mentioned above that gradient fields (forces) are only balanced by the discrete pressure.

As shown above, pressure robustness is immediately given in the case when $V_{0,h} \subset V_0$. A finite element method that yields the kernel inclusion is given by the Scott-Vogelius finite element methods given by the choice

$$V_h := \{v_h \in H_0^1(\Omega, \mathbb{R}^2) : v_h|_T \in \mathbb{P}^2(T, \mathbb{R}^2)\ \forall T \in \mathcal{T}_h\},$$
$$Q_h := \{q_h \in L_0^2(\Omega, \mathbb{R}) : q_h|_T \in \mathbb{P}^1(T, \mathbb{R})\ \forall T \in \mathcal{T}_h\}.$$

Since $\operatorname{div}(V_h) \subset Q_h$ we have that

$$\int_\Omega \operatorname{div}(v_h) q_h \,\mathrm{d}x = 0 \quad \forall q_h \in Q_h \qquad \overset{q_h := \operatorname{div}(v_h)}{\Rightarrow} \qquad \operatorname{div}(v_h) = 0,$$

and thus discretely divergence-free functions are also exactly divergence-free. Unfortunately is the Scott-Vogelius method not stable on arbitrary triangulations but only on barycentric refined ones. This is derived by splitting each triangle (in two dimensions for example) $T \in \mathcal{T}_h$ into three sub triangles by connecting the barycenter with the vertices. Note however, that this procedure might produce elements with a very bad aspect ratio if boundary layers need to be approximated.

Unluckily, all other methods discussed so far, which are used in many (industrial) codes for computational fluid dynamics are not pressure robust in general. To this end many authors as in [24, 10, 30, 27, 26, 15, 41, 43] have studied a technique to "repair" pressure robustness for standard methods by means of the introduction of a reconstruction operator. For simplicity we now assume that Q is discretized by a discontinuous approximation space. Note, that the continuous setting is also possible, see [29], but is much more difficult. Now let $k_{\mathcal{R}} := k_Q + 1$, where k_Q is the polynomial order of the discrete pressure space. We assume that there exists an operator $\mathcal{R} : V_h \to \tilde{V}_h$, with some $H(\mathrm{div})$-conforming space $\tilde{V}_h$, that fulfills the properties

$$\|v_h - \mathcal{R}v_h\|_T \lesssim h\|\nabla v_h\|_T \quad \forall v_h \in V_h, \forall T \in \mathcal{T}_h \tag{2.20}$$

$$(\mathcal{R}v_h - v_h, l_h) = 0 \quad \forall l_h \in \mathbb{P}^{k_{\mathcal{R}}-2}(\mathcal{T}_h, \mathbb{R}^d), \tag{2.21}$$

$$\mathrm{div}(\mathcal{R}v_h) = 0 \quad \forall v_h \in V_{h,0}, \tag{2.22}$$

By means of this operator we now define modified Stokes problem: Find $(u_h, p_h) \in V_h \times Q_h$ such that

$$a(u_h, v_h) + b(v_h, p_h) = (f, \mathcal{R}v_h) \quad \forall v_h \in V_h \tag{2.23}$$

$$b(u_h, q_h) = 0 \quad \forall q_h \in Q_h. \tag{2.24}$$

Here, we only introduced a consistency error by changing the right hand side. Note, that by standard scaling argument and with (2.20) we have

$$\|\mathcal{R}v_h\|_0 \leq \|\mathcal{R}v_h - v_h\|_0 + \|v_h\|_0 \lesssim (\sum_{T\in\mathcal{T}_h} h^2\|\nabla v_h\|_T^2)^{1/2} + \|v_h\|_0 \lesssim \|v_h\|_0,$$

and thus since $\|v_h\|_0 \lesssim \|v_h\|_1$ we have that $(f, \mathcal{R}v_h)$ is still a continuous functional (needed for solveability). This allows us to derive the following pressure robust error estimate.

Theorem 17. *Let $(u, p) \in V \times Q$ be the exact solution of weak formulation of the Stokes equation (2.7), and let $(u_h, p_h) \in V_h \times Q_h$ be the discrete solution of (2.23). Further assume the regularity estimate $\Delta u \in L^2(\Omega)$. There holds the pressure robust best approximation*

result

$$\|u - u_h\|_1 \lesssim \inf_{v_h \in V_h} \|u - v_h\|_1 + h\|(\mathrm{id} - \Pi_{\mathcal{T}_h}^{k_{\mathcal{R}}-2}) \operatorname{div}(\varepsilon(u))\|_0,$$

where $\Pi_{\mathcal{T}_h}^{k_{\mathcal{R}}-2} = 0$ *if* $k_{\mathcal{R}} \leq 1$.

Proof. The proof follows with very similar steps as in Lemma 1. To this end let $v_h \in V_{0,h}$, then the triangle inequality gives

$$\|u - u_h\|_1 \leq \|u - v_h\|_1 + \|v_h - u_h\|_1.$$

Now let $w_h := v_h - u_h$ then the coercivity of the bilinear form a induces

$$\nu\|v_h - u_h\|_1^2 = \nu\|w_h\|_1^2 \lesssim a(v_h - u_h, v_h - u_h) = a(v_h - u, w_h) + a(u - u_h, w_h).$$

For the second term on the right hand side we get by linearity

$$\begin{aligned} a(u - u_h, w_h) &= a(u, w_h) - a(u_h, w_h) \\ &= a(u, w_h) - (f, \mathcal{R}w_h) - b(v_h - u_h, p_h). \end{aligned}$$

Now, since $w_h \in V_{0,h}$ we have $b(v_h - u_h, p_h) = 0$, and by property (2.22) integration by parts shows that

$$(f, \mathcal{R}w_h) = (-\nu\Delta u, \mathcal{R}w_h) + (\nabla p, \mathcal{R}w_h) = (-\operatorname{div}(\nu\varepsilon(u)), \mathcal{R}w_h).$$

In total we have the estimate, again by integration by parts we get

$$a(u - u_h, w_h) = (-\operatorname{div}(\nu\varepsilon(u)), w_h - \mathcal{R}w_h) = ((\mathrm{id} - \Pi_{\mathcal{T}_h}^{k_{\mathcal{R}}-2})(-\operatorname{div}(\nu\varepsilon(u))), w_h - \mathcal{R}w_h),$$

where we used (2.21) in the last step. Note, that in the case where the reconstruction operator fulfills no orthogonality properties ($k_{\mathcal{R}} \leq 1$) we simply set $\Pi_{\mathcal{T}_h}^{k_{\mathcal{R}}-2} = 0$. Using the approximation results (2.20) and the Cauchy-Schwarz inequality then further gives

$$a(u - u_h, w_h) \lesssim \|(\mathrm{id} - \Pi_{\mathcal{T}_h}^{k_{\mathcal{R}}-2}) \operatorname{div}(\nu\varepsilon(u))\|_0 h\|w_h\|_1.$$

By the continuity of a and a division by ν we conclude

$$\|w_h\|_1^2 \lesssim \|v_h - u\|_1\|w_h\|_1 + h\|(\mathrm{id} - \Pi_{\mathcal{T}_h}^{k_{\mathcal{R}}-2}) \operatorname{div}(\varepsilon(u))\|_0,$$

which proves the statement. □

In the following we aim to define a reconstruction operator $\mathcal{R}$ that fulfills above properties. For the ease we only consider the case $d = 2$ but note that these findings can also be extended to three dimensions. In the case of a conforming velocity approximation and a discontinuous pressure approximation, the reconstruction operator is given by an interpolation operator into the $H(\mathrm{div})$-conforming Brezzi-Douglas-Marini space of appropriate order, where (as a short reminder)

$$H(\mathrm{div}, \Omega) := \{v \in L^2(\Omega, \mathbb{R}^d) : \mathrm{div}(v) \in L^2(\Omega)\},$$

hence L^2 functions whose **weak divergence** is also in $L^2(\Omega)$. Note (see also Section 2.1) that the normal trace operator γ_n can be continuously extended onto $H(\mathrm{div}, \Omega)$. This motivates to approximate the $H(\mathrm{div}, \Omega)$ space by a normal continuous polynomial space. To this end we define the space

$$\begin{aligned}\mathrm{BDM}^k :=& \{v_h \in H(\mathrm{div}, \Omega) : v_h|_T \in \mathbb{P}^k(T, \mathbb{R}^d)\} \\ =& \{v_h \in \mathbb{P}^k(\mathcal{T}_h, \mathbb{R}^d) : [\![v_h \cdot n]\!] = 0 \text{ on all } F \in \mathcal{F}_h\},\end{aligned}$$

where the jump is defined as in (2.33a). Whereas the "one to one" mapping is the proper mapping for standard H^1-conforming finite element spaces (because it preserves continuity) the correct mapping for the BDM-space is given by the Piola mapping. To this end let $\phi_T : \widehat{T} \to T$ be the (affine) mapping from the reference to the physical element, and let $F_T := \phi_T'$ denote its Jacobian. For a functions $\hat{\sigma} \in L^2(\hat{T})$ we define the Piola mapping by

$$\mathcal{P}(\hat{\sigma})(x) := \frac{1}{\det(F_T)} F_T \hat{\sigma}(\hat{x}) \quad \text{with} \quad x = \phi_T(\hat{x}).$$

Lemma 4. *Let $\hat{\sigma} \in H(\mathrm{div}, \hat{T})$ and set $\sigma = \mathcal{P}(\hat{\sigma})$. Then we have*

$$\mathrm{div}(\sigma)(x) = \frac{1}{\det(F_T)} \mathrm{div}(\hat{\sigma})(\hat{x}) \quad \textit{with} \quad x = \phi_T(\hat{x}).$$

Proof. Follows immediately using the definition of the weak divergence and is left for the reader as exercise. □

The corresponding finite element for BDM^k and every $T \in \mathcal{T}_h$ is based (for example) on

the following set of functionals

$$\Phi^F(v) := \left\{ \int_F v \cdot n r_h \,\mathrm{d}s : r_h \in \mathbb{P}^k(F), \forall F \subset \partial T \right\}, \tag{2.25}$$

$$\Phi^T_{\mathrm{div}}(v) := \left\{ \int_T \mathrm{div}(v) s_h \,\mathrm{d}x : s_h \in \mathbb{P}^{k-1}(T)/\mathbb{R} \right\}, \tag{2.26}$$

$$\Phi^T_{\mathrm{curl}}(v) := \left\{ \int_T v \cdot \begin{pmatrix} x_2 \\ -x_1 \end{pmatrix} l_h \,\mathrm{d}x : l_h \in \mathbb{P}^{k-2}(T) \right\}. \tag{2.27}$$

Remark 8. The last group is named curl since the function $(x_2, -x_1)^{\mathrm{T}} l_h$ all have a zero divergence but a nonzero curl.

In the following we prove that these functionals are linearly independent. To this end we first show that we can map them to the reference element $\hat{T}$.

Lemma 5. *Let $\hat{v}$ be such that $v = \mathcal{P}(\hat{v})$, then the functionals* (2.25),(2.26) *and* (2.27) *are equivalent to*

$$\Phi^{\hat{F}}(v) := \left\{ \int_{\hat{F}} \hat{v} \cdot \hat{n} \hat{r}_h \,\mathrm{d}s : \hat{r}_h \in \mathbb{P}^k(\hat{F}), \forall \hat{F} \subset \partial \hat{T} \right\}, \tag{2.28}$$

$$\Phi^{\hat{T}}_{\mathrm{div}}(v) := \left\{ \int_{\hat{T}} \mathrm{div}(\hat{v}) \hat{s}_h \,\mathrm{d}\hat{x} : \hat{s}_h \in \mathbb{P}^{k-1}(\hat{T})/\mathbb{R} \right\}, \tag{2.29}$$

$$\Phi^{\hat{T}}_{\mathrm{curl}}(v) := \left\{ \int_{\hat{T}} \hat{v} \cdot \begin{pmatrix} \hat{x}_2 \\ -\hat{x}_1 \end{pmatrix} \hat{l}_h \,\mathrm{d}\hat{x} : \hat{l}_h \in \mathbb{P}^{k-2}(\hat{T}) \right\}. \tag{2.30}$$

Proof. In the following we use a one to one mapping for the testing polynomials, i.e. we have $r_h(x) = \hat{r}_h(\hat{x}), s_h(x) = \hat{s}_h(\hat{x})$ and $l_h(x) = \hat{l}_h(\hat{x})$. Now let $\hat{F}$ be a facet of the reference element $\hat{T}$ such that $F = \phi_T(\hat{F})$. Following (2.4), the normal vector has the relation

$$n = \frac{\det(F_T)}{\det(F_T^F)} F_T^{-\mathrm{T}} \hat{n},$$

where F_T^F is the Jacobian of $\phi_T|_{\hat{F}}$. This shows that for all r_h we have

$$\int_F v \cdot n r_h \,\mathrm{d}s = \int_{\hat{F}} \frac{1}{\det(F_T)} F_T \hat{v} \cdot \frac{\det(F_T)}{\det(F_T^F)} F_T^{-\mathrm{T}} \hat{n} \hat{r}_h \det(F_T^F) \,\mathrm{d}\hat{s} = \int_{\hat{F}} \hat{v} \cdot \hat{n} \hat{r}_h \,\mathrm{d}\hat{s}\,.$$

Similarly we have by Lemma 4 for all s_h

$$\begin{aligned}\int_T \operatorname{div}(v) s_h \,\mathrm{d}x &= \int_T \frac{1}{\det(F_T)} \operatorname{div}(\hat{v}) s_h \,\mathrm{d}x \\ &= \int_{\hat{T}} \frac{1}{\det(F_T)} \operatorname{div}(\hat{v}) \hat{s}_h \det(F_T) \,\mathrm{d}\hat{x} = \int_{\hat{T}} \operatorname{div}(\hat{v}) \hat{s}_h \,\mathrm{d}\hat{x}\,.\end{aligned}$$

For the last group we first have to observe what mapping has to be chosen. To this end we set $\hat{m}(\hat{x}) := (\hat{x}_2, -\hat{x}_1)^{\mathrm{T}} \hat{l}_h(\hat{x})$ and define $m(x) := F_T^{-\mathrm{T}} \hat{m}(\hat{x})$ (this is called a covariant transformation and is used for $H(\mathrm{curl})$-conforming functions). We will now show that this mapping preserves the space. For the ease we now only consider the case where $\phi_T(\hat{x}) = x = F_T \hat{x}$ (hence no translation is included). Together with the rotation matrix

$$R := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

we see that

$$m(x) = F_T^{-\mathrm{T}} \hat{m}(\hat{x}) = F_T^{-\mathrm{T}} \begin{pmatrix} \hat{x}_2 \\ -\hat{x}_1 \end{pmatrix} \hat{l}_h(\hat{x}) = F_T^{-\mathrm{T}} R \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} \hat{l}_h(\hat{x}) = F_T^{-\mathrm{T}} R F_T^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \hat{l}_h(\hat{x}).$$

Since R is skew-symmetric, and $F_T^{-\mathrm{T}} R F_T^{-1}$ is also skew-symmetric, there exists a constant $c \in \mathbb{R}$ such that $F_T^{-\mathrm{T}} R F_T^{-1} = cR$, and thus since $\hat{l}_h = l_h$ is arbitrary we get with the substitution $\hat{l}_h \to c\hat{l}_h$

$$m(x) = F_T^{-\mathrm{T}} R F_T^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \hat{l}_h(\hat{x}) = R \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} l_h(x) = \begin{pmatrix} x_2 \\ -x_1 \end{pmatrix} l_h(x).$$

In total this gives

$$\int_T v \cdot \begin{pmatrix} x_2 \\ -x_1 \end{pmatrix} l_h \,\mathrm{d}x = \int_{\hat{T}} \frac{1}{\det(F_T)} F_T \hat{v} \cdot F_T^{-\mathrm{T}} \begin{pmatrix} \hat{x}_2 \\ -\hat{x}_1 \end{pmatrix} \hat{l}_h \det(F_T) \,\mathrm{d}\hat{x} = \int_{\hat{T}} \hat{v} \cdot \begin{pmatrix} \hat{x}_2 \\ -\hat{x}_1 \end{pmatrix} \hat{l}_h \,\mathrm{d}\hat{x}\,.$$

□

Next we continue with the proof of the linearly independence of the first two groups.

Lemma 6. *The functionals* (2.28) *and* (2.29) *are linearly independent.*

Proof. Let $\hat{r}_h \in \mathbb{P}^k(\partial\hat{T}, \mathbb{R})$ and $\hat{s}_h \in \mathbb{P}^{k-1}(\hat{T}, \mathbb{R})/\mathbb{R}$ such that

$$\int_{\partial\hat{T}} \hat{r}_h \hat{v}_h \cdot \hat{n} \,\mathrm{d}\hat{s} + \int_{\hat{T}} \operatorname{div}(\hat{v}_h) \hat{s}_h \,\mathrm{d}\hat{x} = 0 \quad \forall v_h \in \mathbb{P}^k(\hat{T}, \mathbb{R}^2).$$

We show that this induces $\hat{r}_h = \hat{s}_h = 0$. In a first step we take the choice

$$\hat{v}_h = (\hat{x}_2 \partial_{\hat{x}_1} \hat{s}_h, \hat{x}_2 \partial_{\hat{x}_2} \hat{s}_h)^{\mathrm{T}} (1 - \hat{x}_1 - \hat{x}_2)$$

. This gives that $\hat{v}_h \cdot \hat{n} = 0$ on the boundary, and so using integration by parts above conditions gives

$$\int_{\hat{T}} [\hat{x}_1 (\partial_{\hat{x}_1} \hat{s}_h)^2 + \hat{x}_2 (\partial_{\hat{x}_2} \hat{s}_h)^2](1 - \hat{x}_1 - \hat{x}_2) \, \mathrm{d}\hat{x} = 0,$$

thus $\nabla \hat{s}_h = 0$ (since all terms are positive) which gives $s_h = 0$. On the face $\hat{F}_0$, see Figure 2.1, we now set $\hat{v}_h = (\hat{x}_1 q_h, 0)^{\mathrm{T}}$ or $\hat{v}_h = (0, \hat{x}_2 q_h)^{\mathrm{T}}$, where $q_h \in \mathbb{P}^{k-1}(\hat{F}_0)$. This shows since

$$\int_{\hat{F}_0} \hat{r}_h \hat{x}_1 q_h \, \mathrm{d}\hat{s} = \int_{\hat{F}_0} \hat{r}_h \hat{x}_2 q_h \, \mathrm{d}\hat{s} = \int_{\hat{F}_0} \hat{r}_h q_h \, \mathrm{d}\hat{s} = 0,$$

where we used that $\hat{x}_1 + \hat{x}_2 = 1$ on $\hat{F}_0$. In total this shows that $\hat{r}_h$ vanishes on $\hat{F}_0$. In a similar way we continue on $\hat{F}_1$ and $\hat{F}_2$, to conclude that $\hat{r}_h = 0$ on $\partial \hat{T}$. □

We are now in the position of proving the linear independence, to this end we first further introduce the space

$$\hat{\mathbb{H}}^k := \{\hat{u}_h \in P^k(\hat{T}, \mathbb{R}^2) : \hat{u}_h \cdot \hat{n} = 0 \text{ on } \partial \hat{T}, \operatorname{div}(\hat{u}_h) = 0\}.$$

Lemma 7. *The functionals* (2.28), (2.29) *and* (2.30) *are linearly independent on* $P^k(\hat{T}, \mathbb{R}^2)$.

Proof. For a given $\hat{v}_h \in \mathbb{P}^k(\hat{T}, \mathbb{R}^2)$ assume that all functionals (2.28),(2.29) and (2.30) vanish. In the following we show that this induces that $\hat{v}_h = 0$. A counting argument will conclude the proof. Since $\hat{v}_h \in \mathbb{P}^k(\hat{T}, \mathbb{R}^2)$ we first choose $\hat{r}_h := \hat{v}_h \cdot \hat{n}$. Then, the first group shows that the normal trace of $\hat{v}_h$ vanishes. Next, set $\hat{s}_h := \operatorname{div}(\hat{v}_h) - c$ where $c \in \mathbb{R}$ is such that $\hat{s}_h$ has a zero mean value. Then the second group and the Gaussian theorem show

$$0 = \int_{\hat{T}} \operatorname{div}(\hat{v}_h) \hat{s}_h \, \mathrm{d}\hat{x} = \int_{\hat{T}} \operatorname{div}(\hat{v}_h) \operatorname{div}(\hat{v}_h) \, \mathrm{d}\hat{x} - c \int_{\partial \hat{T}} \hat{v}_h \cdot \hat{n} \, \mathrm{d}\hat{s} = \int_{\hat{T}} \operatorname{div}(\hat{v}_h)^2 \, \mathrm{d}\hat{x} \,.$$

and so v_h has a zero divergence. A counting argument shows that the first and second group given by (2.28) and (2.29) result in $3(k+1) + \frac{k(k+1)}{2} - 1$ constraints. This shows,

that the dimension of $\hat{\mathbb{H}}^k$ is given by

$$\dim(\hat{\mathbb{H}}^k) = \dim(P^k(\hat{T},\mathbb{R}^2)) - 3(k+1) - \frac{k(k+1)}{2} + 1 = \frac{k(k-1)}{2} = \dim(P^{k-2}(\hat{T},\mathbb{R})).$$

This shows that

$$\hat{\mathbb{H}}^k = \{\hat{u}_h \in P^k(\hat{T},\mathbb{R}^2) : \hat{u}_h = \operatorname{curl}(b_{\hat{T}}\hat{\xi}_h), \hat{\xi}_h \in \mathbb{P}^{k-2}(\hat{T},\mathbb{R})\}$$

because every $\operatorname{curl}(b_{\hat{T}}\hat{\xi}_h)$ is divergence free and has a zero normal trace. In total this shows that we find a (fixed) function $\hat{\xi}_h \in \mathbb{P}^{k-2}(T)$ such that $\hat{v}_h = \operatorname{curl}(b_{\hat{T}}\hat{\xi}_h)$. By choosing $\hat{l}_h = \hat{\xi}_h$ we have

$$\begin{aligned}
0 = \int_{\hat{T}} \operatorname{curl}(b_{\hat{T}}\hat{\xi}_h)\cdot\begin{pmatrix}\hat{x}_2\\-\hat{x}_1\end{pmatrix}\hat{\xi}_h\,\mathrm{d}\hat{x} &= \int_{\hat{T}} \nabla(b_{\hat{T}}\hat{\xi}_h)\cdot\begin{pmatrix}\hat{x}_1\\\hat{x}_2\end{pmatrix}\hat{\xi}_h\,\mathrm{d}\hat{x}\\
&= -\int_{\hat{T}} b_{\hat{T}}\hat{\xi}_h \operatorname{div}(\begin{pmatrix}\hat{x}_1\\\hat{x}_2\end{pmatrix}\hat{\xi}_h)\,\mathrm{d}\hat{x}\\
&= -\int_{\hat{T}} b_{\hat{T}}\hat{\xi}_h(\operatorname{div}(\begin{pmatrix}\hat{x}_1\\\hat{x}_2\end{pmatrix})\hat{\xi}_h + \begin{pmatrix}\hat{x}_1\\\hat{x}_2\end{pmatrix}\cdot\nabla(\hat{\xi}_h))\,\mathrm{d}\hat{x}\\
&= -\int_{\hat{T}} 2b_{\hat{T}}\hat{\xi}_h^2\,\mathrm{d}\hat{x} - \frac{1}{2}\int_{\hat{T}} b_{\hat{T}}\nabla(\hat{\xi}_h^2)\cdot\begin{pmatrix}\hat{x}_1\\\hat{x}_2\end{pmatrix}\mathrm{d}\hat{x}\\
&= -\int_{\hat{T}} 2b_{\hat{T}}\hat{\xi}_h^2\,\mathrm{d}\hat{x} + \frac{1}{2}\int_{\hat{T}} \hat{\xi}_h^2 \operatorname{div}(\begin{pmatrix}\hat{x}_1\\\hat{x}_2\end{pmatrix} b_{\hat{T}})\,\mathrm{d}\hat{x}\\
&= -\int_{\hat{T}} b_{\hat{T}}\hat{\xi}_h^2\,\mathrm{d}\hat{x} + \frac{1}{2}\int_{\hat{T}} \begin{pmatrix}\hat{x}_1\\\hat{x}_2\end{pmatrix}\cdot\nabla(b_{\hat{T}})\hat{\xi}_h^2\,\mathrm{d}\hat{x}
\end{aligned}$$

On the reference element we have $\lambda_1 = \hat{x}_1$, $\lambda_2 = \hat{x}_2$ and $\lambda_0 = (1-\hat{x}_1-\hat{x}_2)$. Further, the bubble is given by $b_{\hat{T}} = \lambda_0\lambda_1\lambda_2$. This gives

$$\begin{aligned}
-b_{\hat{T}} + \frac{1}{2}\begin{pmatrix}\hat{x}_1\\\hat{x}_2\end{pmatrix}\cdot\nabla(b_{\hat{T}}) &= -\lambda_0\lambda_1\lambda_2 + \frac{1}{2}\lambda_1\partial_1(\lambda_0\lambda_1\lambda_2) + \frac{1}{2}\lambda_2\partial_2(\lambda_0\lambda_1\lambda_2)\\
&= -\lambda_0\lambda_1\lambda_2 + \frac{1}{2}\lambda_1[\lambda_0\lambda_2 + \partial_1(\lambda_0\lambda_2)] + \frac{1}{2}\lambda_2[\lambda_0\lambda_1 + \partial_2(\lambda_0\lambda_1)]\\
&= \frac{1}{2}\lambda_1\partial_1(\lambda_0\lambda_2) + \frac{1}{2}\lambda_2\partial_2(\lambda_0\lambda_1) = -\hat{x}_1\hat{x}_2.
\end{aligned}$$

In total we have (by a scaling with -1)

$$\int_{\hat{T}} \hat{x}_1 \hat{x}_2 \hat{\xi}_h^2 \, \mathrm{d}\hat{x} = 0,$$

and since $\hat{x}_1, \hat{x}_2 \geq 0$ on $\hat{T}$ this shows that $\hat{\xi}_h = 0$. We conclude the proof by a simple counting argument. □

Remark 9. Above proof shows that the third group can be changed to the following set

$$\Phi^T_{\mathrm{curl}}(v) := \left\{ \int_T v \cdot \mathrm{curl}(b_T l_h) \, \mathrm{d}x : l_h \in \mathbb{P}^{k-2}(T) \right\}.$$

Based on the functionals (2.25),(2.26) and (2.27) we define the reconstruction operator $\mathcal{R} : V_h \to \mathrm{BDM}^{k_\mathcal{R}}$ such that for an arbitrary $v_h \in V_h$ we have

$$\begin{aligned}
\int_F (v_h - \mathcal{R}v_h) \cdot n r_h \, \mathrm{d}s = 0 \quad & \forall r_h \in \mathbb{P}^{k_\mathcal{R}}(F) \forall F \subset \partial T, \\
\int_T \mathrm{div}(v_h - \mathcal{R}v_h) s_h \, \mathrm{d}x = 0 \quad & \forall s_h \in \mathbb{P}^{k_\mathcal{R}-1}(T)/\mathbb{R}, \\
\int_T (v - \mathcal{R}v_h) \cdot \begin{pmatrix} x_2 \\ -x_1 \end{pmatrix} l_h \, \mathrm{d}x = 0 \quad & \forall l_h \in \mathbb{P}^{k_\mathcal{R}-2}(T),
\end{aligned}$$

then we have the following properties.

Lemma 8. *The operator $\mathcal{R} : V_h \to \mathrm{BDM}^{k_\mathcal{R}}$ fulfills the properties* (2.20)*,* (2.21) *and* (2.22)*.*

Proof. Since the first group (2.25) shows that $\mathcal{R}$ preserves constants, the approximation property (2.20) follows by standard scaling arguments and the Bramble Hilbert Lemma. Now let $v_h \in V_{0h}$ and $q_h \in Q_h$ be arbitrary. On every element $T \in \mathcal{T}_h$ we can split $q_h = q_h^0 + q_h^1$ with $q_h^1 \in \mathbb{P}^{k_\mathcal{R}-1}(T) \setminus \mathbb{R}$ (since $k_\mathcal{R} = k_Q + 1$) and $q_h^0 \in \mathbb{R}$. Then, the first two groups show that

$$\begin{aligned}
0 = \int_\Omega \mathrm{div}(\mathcal{R}v_h) q_h \, \mathrm{d}x &= \sum_{T \in \mathcal{T}_h} \int_T \mathrm{div}(\mathcal{R}v_h) q_h^0 \, \mathrm{d}x + \int_T \mathrm{div}(\mathcal{R}v_h) q_h^1 \, \mathrm{d}x \\
&= \sum_{T \in \mathcal{T}_h} \int_{\partial T} \mathcal{R}v_h \cdot n q_h^0 \, \mathrm{d}s + \int_T \mathrm{div}(v_h) q_h^1 \, \mathrm{d}x \\
&= \sum_{T \in \mathcal{T}_h} \int_{\partial T} v_h \cdot n q_h^0 \, \mathrm{d}s + \int_T \mathrm{div}(v_h) q_h^1 \, \mathrm{d}x = \int_\Omega \mathrm{div}(v_h) q_h \, \mathrm{d}x = 0.
\end{aligned}$$

Next note that the reconstruction operator preserves the homogeneous Dirichlet boundary conditions in normal direction, i.e. $\mathcal{R}(v_h) \cdot n = 0$ on $\partial\Omega$ which shows that $\mathrm{div}(\mathcal{R}v_h) \in Q_h$.

Choosing $q_h = \operatorname{div}(\mathcal{R}v_h)$ in above observations then gives $\operatorname{div}(\mathcal{R}v_h) = 0$, thus (2.22) holds. For (2.21) first note that on each element we can split the polynomial space $\mathbb{P}^{k_\mathcal{R}-2}(T, \mathbb{R}^d)$ into (see for example in [29])

$$\mathbb{P}^{k_\mathcal{R}-2}(T, \mathbb{R}^d) = \nabla \mathbb{P}^{k_\mathcal{R}-1}(T, \mathbb{R}) \oplus \begin{pmatrix} x_2 \\ -x_1 \end{pmatrix} \mathbb{P}^{k_\mathcal{R}-3}(T, \mathbb{R}),$$

and thus, the orthogonality follows by the definition of the functionals. □

Example 12. We now consider the $\mathbb{P}2\mathbb{P}0$ example. To this end we set $k_\mathcal{R} = 1$, thus the reconstruction operator maps into the space of linear $H(\operatorname{div})$-conforming polynomials. Theorem 17 gives the best approximation result

$$\|u - u_h\|_1 \lesssim \inf_{v_h \in V_h} \|u - v_h\|_1 + h\|\operatorname{div}(\varepsilon u)\|_0.$$

Although Theorem 16 shows that the infimum can be bounded by $\mathcal{O}(h^2)$, the second term limits the order and we get in total

$$\|u - u_h\|_1 \lesssim h\|u\|_2.$$

Nevertheless, since the $\mathbb{P}2\mathbb{P}0$ element in general only shows a linear convergence, this is the result we expected.

Example 13. Now we consider the $\mathbb{P}2$-bubble element. Here we have $k_\mathcal{R} = 2$ and so Theorem 17 gives the best approximation result

$$\|u - u_h\|_1 \lesssim \inf_{v_h \in V_h} \|u - v_h\|_1 + h\|(\mathrm{id} - \Pi^0_{\mathcal{T}_h}) \operatorname{div}(\varepsilon u)\|_0.$$

Using the approximation properties of the L^2-projection we can bound the second term by

$$h\|(\mathrm{id} - \Pi^0_{\mathcal{T}_h}) \operatorname{div}(\varepsilon u)\|_0 \lesssim h \left(\sum_{T \in \mathcal{T}_h} h^2 |u|_3^2 \right)^{1/2},$$

and thus by Theorem 17 we have again in total (assuming enough regularity)

$$\|u - u_h\|_1 \lesssim h^2 \|u\|_3.$$

2.6 (Hybrid) Discontinuous Galerkin methods for the Stokes equation

2.6.1 (Hybrid-) Discontinuous Galerkin methods for the Poisson equation

In this section we aim to derive a new non-conforming finite element method for the approximation of second order problems. For the ease, we only consider the scalar Poisson equation for now and extend the results to the Stokes equations later. We aim to solve the model problem: Find u such that

$$-\Delta u = f \quad \text{in } \Omega \tag{2.31}$$

$$u = u_D \quad \text{on } \partial\Omega. \tag{2.32}$$

Since there are several different definitions of the jump and the mean value in the literature, we give a precise definition as we use it within these notes in the following. To this end T_1 and T_2 be two elements with a common edge F, and let n_1 and n_2 be the two outward pointing normal vectors. Further, for functions $v \in H^1(T_1, \mathbb{R}) \cup H^1(T_2, \mathbb{R})$ and $\tau \in H^1(T_1, \mathbb{R}^d) \cup H^1(T_2, \mathbb{R}^d)$ we set $v_i := v|_{T_i}, \tau_i := \tau|_{T_i}$ with $i = 1, 2$. Then we define

$$\{\!\{v\}\!\} := \frac{1}{2}(v_1 + v_2), \tag{2.33a}$$

$$[\![v]\!]^* := v_1 - v_2, \tag{2.33b}$$

$$\{\!\{\tau\}\!\}^* := \frac{1}{2}(\tau_1 n_1 - \tau_2 n_2), \tag{2.33c}$$

$$[\![\tau \cdot n]\!] := \tau_1 n_1 + \tau_2 n_2. \tag{2.33d}$$

In the case where F is on the boundary $\partial\Omega$ we further set

$$\begin{aligned} \{\!\{v\}\!\} &:= v_1, \\ [\![v]\!]^* &:= v_1, \\ \{\!\{\tau\}\!\}^* &:= \tau_1 n_1, \\ [\![\tau \cdot n]\!] &:= \tau_1 n_1. \end{aligned}$$

Remark 10. Here the symbol $\cdot^*$ should highlight that there is a direction included in the definition. However, as we will see later, changing the direction in both terms $[\![\cdot]\!]^*$ and $\{\!\{\cdot\}\!\}^*$ will give us the same formulations later.

The Nitsche penalty method

Before we start with the derivation of the final method we first discuss two discretisation techniques which are often called the Nitsche penalty method. The first method shows how we can incorporate above the Dirichlet boundary conditions in a weak sense. To this end we multiply the first equation of (2.31) with a test function that does not vanish at the boundary. Integration by parts then gives

$$\int_\Omega \nabla u \cdot \nabla v \,\mathrm{d}x - \int_{\partial\Omega} \nabla u \cdot nv \,\mathrm{d}s = \int_{\partial\Omega} fv \,\mathrm{d}x\,.$$

Using that $u - u_D = 0$ on the boundary, we can add a consistent term to get

$$\int_\Omega \nabla u \cdot \nabla v \,\mathrm{d}x - \int_{\partial\Omega} \nabla u \cdot nv \,\mathrm{d}s - \int_{\partial\Omega} \nabla v \cdot nu \,\mathrm{d}s = \int_{\partial\Omega} fv \,\mathrm{d}x - \int_{\partial\Omega} \nabla v \cdot nu_D \,\mathrm{d}s\,.$$

In order to obtain stability of the method (as proven below) we further add a stabilization integral to define the bilinear form and linear form

$$\begin{aligned} a^{N1}(u,v) &= \int_\Omega \nabla u \cdot \nabla v \,\mathrm{d}x - \int_{\partial\Omega} \nabla u \cdot nv \,\mathrm{d}s - \int_{\partial\Omega} \nabla v \cdot nu \,\mathrm{d}s + \frac{\alpha k^2}{h} \int_{\partial\Omega} uv \,\mathrm{d}s \\ f^{N1}(v) &= \int_{\partial\Omega} fv \,\mathrm{d}x - \int_{\partial\Omega} \nabla v \cdot nu_D \,\mathrm{d}s + \frac{\alpha k^2}{h} \int_{\partial\Omega} u_D v \,\mathrm{d}s, \end{aligned}$$

where α has to be chosen sufficiently large. Note, that above bilinear and linear forms are not well defined for functions in H^1 since, beside evaluating the traces at the boundary we further need the values of the normal derivative which is only well defined if $\nabla u \in H(\mathrm{div})$. Now let $V_h^{N1} := \mathcal{P}^k(\mathcal{T}_h, \mathbb{R}) \cap H^1(\Omega)$, then we define the problem: Find u_h such that

$$a^{N1}(u_h, v_h) = f^{N1}(v_h) \quad \forall v_h \in V_h^{N1}.$$

For the analysis we now define the discrete H^1-like Nitsche norm

$$\|u_h\|_{N1}^2 := \|\nabla u_h\|_\Omega^2 + \frac{k^2}{h}\|u\|_{\partial\Omega}^2.$$

Lemma 9. *Assume that $\alpha > 0$ is sufficiently large, then above bilinear form $a^{N1}(\cdot,\cdot)$ is coercive and continuous on V_h^{N1} with respect to the norm $\|\cdot\|_{N1}$.*

Proof. The crucial ingredient for the stability analysis is the inverse inequality for polynomials as given in Theorem 1. Using this estimate on each boundary element separately

we get in total for the normal flux

$$\|\nabla u_h \cdot n\|_{\partial\Omega}^2 \leq c_1 \|\nabla u_h\|_{\partial\Omega}^2 \lesssim \frac{k^2}{h}\|\nabla u_h\|_\Omega^2 \quad \forall u_h \in V_h^{N1}.$$

By the Cauchy-Schwarz inequality we then immediately derive continuity. Next, applying Cauchy Schwarz and Young's inequality for the integral including the normal derivative we then further get with above inverse inequality

$$\begin{aligned} a^{N1}(u_h, u_h) &= \|\nabla u_h\|_\Omega^2 - 2\int_{\partial\Omega} \nabla u_h \cdot n u_h \,\mathrm{d}s + \frac{\alpha k^2}{h}\|u_h\|_{\partial\Omega}^2 \\ &\geq \|\nabla u_h\|_\Omega^2 - \frac{h}{\varepsilon k^2}\|\nabla u_h \cdot n\|_{\partial\Omega}^2 - \frac{\varepsilon k^2}{h}\|u_h\|_{\partial\Omega}^2 + \frac{\alpha k^2}{h}\|u_h\|_{\partial\Omega}^2 \\ &\geq (1 - \frac{c_1}{\varepsilon})\|\nabla u_h\|_\Omega^2 + \frac{(\alpha - \varepsilon)k^2}{h}\|u_h\|_{\partial\Omega}^2. \end{aligned}$$

Choosing $\alpha > c_1$ and $\varepsilon < \alpha$, shows coercivity. □

Note that we can not apply the standard theory to derive an apriori error estimate since a is not continuous on H^1 and so we can not derive Céa like best approximation results. Nevertheless we could directly estimate the interpolation error $\|u_h - I_h u\|_{N1}$.

Above technique provided a method that incorporates the boundary conditions in a weak sense. In a similar way we can also derive a method that enforces weak continuity between two domains. To this end assume that we split the domain into two parts, i.e. we have $\Omega = \Omega_1 \cup \Omega_2$ with $\gamma := \overline{\Omega}_1 \cap \overline{\Omega}_2$. Further we consider for simplicity the case where the (for the ease homogeneous) Dirichlet boundary conditions are incorporated in a strong sense. With $u_1 := u|_{\Omega_1}$ and $u_2 := u|_{\Omega_2}$ we have the problem

$$\begin{aligned} -\Delta u &= f && \text{in } \Omega, \\ u_1 &= u_2 && \text{on } \gamma, \\ \nabla u_1 \cdot n_1 &= -\nabla u_2 \cdot n_2 && \text{on } \gamma, \\ u &= 0 && \text{on } \partial\Omega, \end{aligned}$$

where n_1 and n_2 are the outward pointing normal vectors on Ω_1 and Ω_2, respectively. Testing the first line of above problem with a domain wise smooth test function that vanishes on $\partial\Omega$ and applying integration by parts on each subdomain gives

$$\int_{\Omega_1} \nabla u_1 \cdot \nabla v_1 \,\mathrm{d}x - \int_\gamma \nabla u_1 \cdot n_1 v_1 \,\mathrm{d}s + \int_{\Omega_2} \nabla u_2 \cdot \nabla v_2 \,\mathrm{d}x - \int_\gamma \nabla u_2 \cdot n_2 v_2 \,\mathrm{d}s = \int_\Omega f v,$$

where as before $v_i := v|_{\Omega_i}$ with $i = 1, 2$. Extending the definition of the jump and the mean

value in (2.33a) to the above case (i.e. set $T_1 = \Omega_1$ and $T_2 = \Omega_2$) we see that

$$-\int_\gamma \nabla u_1 \cdot n_1 v_1 \,\mathrm{d}s - \int_\gamma \nabla u_2 \cdot n_2 v_2 \,\mathrm{d}s = -\int_\gamma \{\!\{\nabla u\}\!\}^* [\![v]\!]^* + \{\!\{v\}\!\} [\![\nabla u \cdot n]\!] \,\mathrm{d}s\,.$$

Thus, using the continuity of the normal flux $[\![\nabla u \cdot n]\!] = 0$ we get in total

$$\int_\Omega \nabla u \cdot \nabla v \,\mathrm{d}x - \int_\gamma \{\!\{\nabla u\}\!\}^* [\![v]\!]^* \,\mathrm{d}s = \int_\Omega fv.$$

Note, that if we change the numbering of the two subdomains, the definition of the mean value $\{\!\{\cdot\}\!\}^*$ and the jump $[\![\cdot]\!]^*$ changes in the same manner, thus in total we get the same formulation, see also Remark 10. As before we add a consistent symmetric term (using that the exact solution u is continuous) and a stability term to get the variational formulation: Find $u_h \in V_h^{N2}$ such that

$$a^{N2}(u_h, v_h) = f^{N2}(v_h) \quad \forall v_h \in V_h^{N2},$$

with the discrete space

$$V_h^{N2} := \{v_h \in H^1(\Omega_1, \mathbb{R}) \cup H^1(\Omega_2, \mathbb{R}) : v_h|_T \in \mathbb{P}^k(T, \mathbb{R}), v_h = 0 \text{ on } \partial\Omega\}.$$

and the bilinear and linear form

$$\begin{aligned} a^{N2}(u_h, v_h) &:= \int_\Omega \nabla u_h \cdot \nabla v_h \,\mathrm{d}x - \int_\gamma \{\!\{\nabla u_h\}\!\}^* [\![v_h]\!]^* \,\mathrm{d}s \\ &\quad - \int_\gamma \{\!\{\nabla v_h\}\!\}^* [\![u_h]\!]^* \,\mathrm{d}s + \frac{\alpha k^2}{h} \int_\gamma [\![u_h]\!]^* [\![v_h]\!]^* \,\mathrm{d}s, \\ f^{N2}(v_h) &:= \int_\Omega f v_h \,\mathrm{d}x\,. \end{aligned}$$

Above method is now stable in the norm

$$\|u_h\|_{N2}^2 := \|\nabla u_h\|_{\Omega_1}^2 + \|\nabla u_h\|_{\Omega_2}^2 + \frac{k^2}{h} \|[\![u_h]\!]^*\|_\gamma^2.$$

The stability proof is left as an exercise.

The discontinuous Galerkin method

The idea of the discontinuous Galerkin (DG) method is to use a Nitsche penalty technique to enforce weak continuity on each facet of the triangulation separately and to further

enforce Dirichlet boundary conditions in a weak sense. The final result is the symmetric interior penalty discontinuous Galerkin (SIP-DG) bilinear form given by

$$a^{DG}(u,v) := \sum_{T\in\mathcal{T}_h}\int_T \nabla u\cdot\nabla v\,\mathrm{d}x - \sum_{F\in\mathcal{F}_h}\int_F \{\!\{\nabla u\}\!\}^*[\![v]\!]^*\,\mathrm{d}s \\ - \sum_{F\in\mathcal{F}_h}\int_F \{\!\{\nabla v\}\!\}^*[\![u]\!]^*\,\mathrm{d}s + \sum_{F\in\mathcal{F}_h}\frac{\alpha k^2}{h}\int_F [\![u]\!]^*[\![v]\!]^*\,\mathrm{d}s\,.$$

and the right hand side by

$$f^{DG}(v) := \int_\Omega fv\,\mathrm{d}x - \sum_{F\in\mathcal{F}_h^{\text{ext}}}\int_F u_D\nabla v\cdot n + \frac{\alpha k^2}{h}u_D v\,\mathrm{d}s\,.$$

Using the space of piece wise polynomials $\mathbb{P}^k(\mathcal{T}_h,\mathbb{R})$ as approximation space we then have the problem: Find $u_h\in\mathbb{P}^k(\mathcal{T}_h,\mathbb{R})$ such that

$$a^{DG}(u_h,v_h) = f^{DG}(v_h)\quad \forall v_h\in\mathbb{P}^k(\mathcal{T}_h,\mathbb{R}).$$

For the analysis we extend the ideas of the previous section and define the norm

$$\|u_h\|_{DG}^2 := \sum_{T\in\mathcal{T}_h}\|\nabla u_h\|_T^2 + \sum_{F\in\mathcal{F}_h}\frac{k^2}{h}\|[\![u_h]\!]^*\|_F^2.$$

Above norm can be interpreted as a discrete H^1-like semi norm. Note, that in the lowest order case, i.e. $k=0$, the first sum vanishes. Then the norm of the jump divided by the h can be interpreted as a difference quotient at each facet, hence we still measure a derivative like quantity. Following similar steps as in the proof of Lemma 9 one can show that the bilinear form a^{DG} is again coercive and continuous (on $\mathbb{P}^k(\mathcal{T}_h,\mathbb{R})$). The error analysis needs a detailed investigation but will not be presented since it follows similar ideas as the analysis presented in the next section.

Note that beside the SIP-DG method a lot of other DG schemes can be found in the literature. An overview and a unified analysis can be found for example in [3].

There are several different motivations for using a DG method instead of a standard continuous Galerkin (CG) approximation as discussed so far. Particularly, as we will see later, DG methods are well suited for convection equations since they allow to incorporate a very smart stabilization mechanism. Nevertheless, although DG methods earned a lot of attention in computational fluid dynamics, they have a crucial disadvantage when we consider standard second order elliptic problems. First of all, compared to a CG method

the number of degrees of freedom is much higher (on the same mesh) and secondly, even worse, the number of non-zero entries per row in the system matrix is much higher. In Figure 2.4 we have plotted the sparsity pattern of two discretizations of problem (2.31) where $\Omega = (0,1)^2$. We have fixed the polynomial order $k = 5$ and compare the non-zero entries of a standard H^1-conforming approximation (left) and a SIP-DG method (right). As mentioned above we observe that the inter element coupling, i.e. the number of non-zero entries per row, is much worse for DG. In the next section we present a technique how this increased coupling can be eliminated.

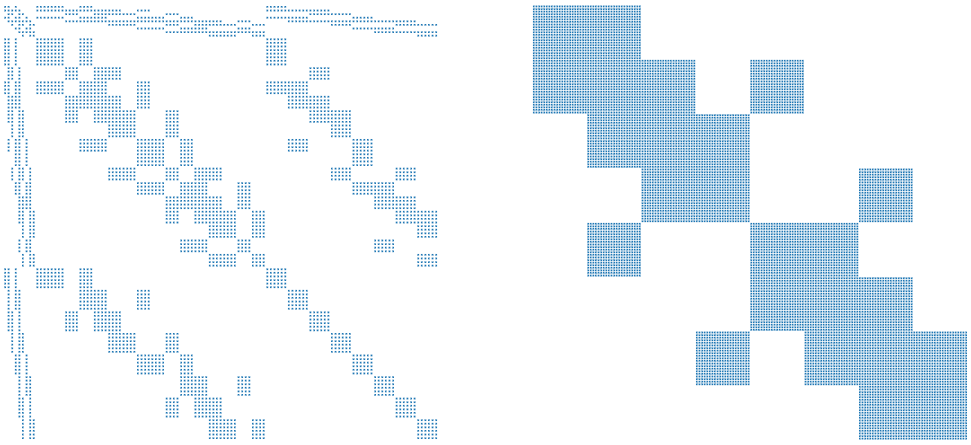


Figure 2.4: Sparsity patterns of a continuous Galerkin and a discontinuous Galerkin approximation of the Poisson problem with $k = 5$ on a regular triangulation with 8 elements on the domain $\Omega = (0,1)^2$. Left we see the pattern of the system matrix (CG) of size 121×121 of the CG approach and right the pattern of the system matrix of size 168×168 of the DG approach.

The hybrid discontinuous Galerkin method

The main idea of a hybridized discontinuous Galerkin approximation is to reduce the inter element coupling of two adjacent elements by introducing additional unknowns at the facets. Although this further increases the number of unknowns, we can apply a static condensation technique to eliminate all local element unknowns. For this then only small local element matrices need to be inverted (what can be done in parallel manner). The final system that is solved then only includes the facet unknowns.

Let $v_h \in V_h$ with $V_h := \mathbb{P}^k(\mathcal{T}_h, \mathbb{R})$ be an element wise smooth test function. Assuming enough regularity of the exact solution, testing (2.31) with v_h and using integration by parts

on all $T \in \mathcal{T}_h$ gives

$$\sum_{T\in\mathcal{T}_h} \int_T \nabla u \cdot \nabla v_h dx - \int_{\partial T} \nabla u \cdot n v_h \,\mathrm{d}s = (f, v_h).$$

Since the normal flux of the exact solution is continuous we have

$$\sum_{T\in\mathcal{T}_h} \int_{\partial T} \nabla u \cdot n \hat{v}_h \,\mathrm{d}s = 0 \quad \forall \hat{v}_h \in \hat{V}_h := \{\hat{v}_h \in \mathbb{P}^k(\mathcal{F}_h, \mathbb{R}) : \hat{v}_h = 0 \text{ on } \partial\Omega\}.$$

Note that similarly as in the derivation of a CG methods, the facet test functions vanish on the (Dirichlet-) boundary. Adding these two equations gives

$$\sum_{T\in\mathcal{T}_h} \int_T \nabla u \cdot \nabla v_h dx - \int_{\partial T} \nabla u \cdot n(v_h - \hat{v}_h) \,\mathrm{d}s = (f, v_h).$$

Here the terms $(v_h - \hat{v}_h)$ read as a hybrid version of the jumps used in the derivation of the DG method. Since the exact solution is continuous across element interfaces we may again add a consistent symmetric and stabilizing term to define the bilinear form

$$\begin{aligned} a^{HDG}((u_h, \hat{u}_h), (v_h, \hat{v}_h)) := &\sum_{T\in\mathcal{T}_h} \int_T \nabla u_h \cdot \nabla v_h dx - \int_{\partial T} \nabla u_h \cdot n(v_h - \hat{v}_h) \,\mathrm{d}s \\ &- \int_{\partial T} \nabla v_h \cdot n(u_h - \hat{u}_h) \,\mathrm{d}s + \frac{\alpha k^2}{h} \int_{\partial T} (u_h - \hat{u}_h)(v_h - \hat{v}_h) \,\mathrm{d}s, \end{aligned}$$

and the problem: Find $(u_h, \hat{u}_h) \in V_h \times \hat{V}_h$ such that

$$a^{HDG}((u_h, \hat{u}_h), (v_h, \hat{v}_h)) = (f, v_h) \quad \forall (v_h, \hat{v}_h) \in V_h \times \hat{V}_h. \tag{2.34}$$

For the stability analysis we introduce the broken Sobolev spaces

$$H^s(\mathcal{T}_h, \mathbb{R}) := \{u \in L^2(\Omega, \mathbb{R}) : u|_T \in H^s(T, \mathbb{R}) \; \forall T \in \mathcal{T}_h\},$$

with the broken norm $\|u\|^2_{H^s(\mathcal{T}_h)} := \sum_{T\in\mathcal{T}_h} \|u\|^2_{H^s(T)}$. Since the boundary integrals of a^{HDG} demand for a higher regularity we define the following continuous spaces

$$\begin{aligned} V^{\mathrm{reg}} &:= H^1(\Omega, \mathbb{R}) \cap H^2(\mathcal{T}_h, \mathbb{R}), \\ \hat{V}^{\mathrm{reg}} &:= \{\hat{u} \in L^2(\mathcal{F}_h, \mathbb{R}) \text{ with } \hat{u} = 0 \text{ on } \partial\Omega\}. \end{aligned}$$

In a first step we show that the HDG method is consistent.

Lemma 10. *Let $u \in H_0^1(\Omega, \mathbb{R}) \cap V^{\text{reg}}$ be the weak solution of (2.31) (with $u_D = 0$) and let $\hat{u} := u|_{\mathcal{F}_h}$. The HDG formulation (2.34) is consistent, i.e.*

$$a^{HDG}((u, \hat{u}), (v, \hat{v}_h)) = (f, v_h) \quad \forall (v_h, \hat{v}_h) \in V_h \times \hat{V}_h.$$

Proof. By the continuity of the exact solution ($u - \hat{u} = 0$ on all $F \in \mathcal{F}_h$) we have

$$a^{HDG}((u, \hat{u}), (v_h, \hat{v}_h)) = \sum_{T \in \mathcal{T}_h} \int_T \nabla u \cdot \nabla v_h dx - \int_{\partial T} \nabla u \cdot n(v_h - \hat{v}_h) \, \mathrm{d}s \, .$$

Next, since we assumed that $f \in L^2(\Omega)$ we also have $f = \operatorname{div}(\nabla u) \in L^2(\Omega)$ thus $\nabla u \in H(\operatorname{div}, \Omega)$. Since this implies that the gradient is normal continuous we have as $\hat{v}_h$ is single valued on the edges

$$\sum_{T \in \mathcal{T}_h} \int_{\partial T} \nabla u \cdot n \hat{v}_h \, \mathrm{d}s = 0.$$

We conclude by an integration by parts argument. □

Next we define two norms

$$\|(u_h, \hat{u}_h)\|_{1,h}^2 := \sum_{T \in \mathcal{T}_h} \|\nabla u_h\|_T^2 + \frac{k^2}{h} \|u_h - \hat{u}_h\|_{\partial T}^2$$

$$\|(u_h, \hat{u}_h)\|_{1,h,*}^2 := \sum_{T \in \mathcal{T}_h} \|\nabla u_h\|_T^2 + \frac{k^2}{h} \|u_h - \hat{u}_h\|_{\partial T}^2 + \frac{h}{k^2} \|\nabla u_h \cdot n\|_{\partial T}^2.$$

By means of these norms we can proof the following stability results.

Lemma 11. *Let $(v_h, \hat{v}_h) \in V_h \times \hat{V}_h$. There holds the norm equivalence*

$$\|(v_h, \hat{v}_h)\|_{1,h} \sim \|(v_h, \hat{v}_h)\|_{1,h,*}.$$

Proof. Follows immediately by the inverse inequality1. □

Since the inverse inequality only holds for discrete functions the second norm is needed to prove continuity on $V \times \hat{V}$.

Lemma 12. *Let the stabilization parameter $\alpha > 0$ be sufficiently large. The bilinear form a^{HDG} is continuous on $(V^{\text{reg}} \times \hat{V}^{\text{reg}}) + (V_h \times \hat{V}_h)$, i.e. there holds*

$$a^{HDG}((u, \hat{u}), (v, \hat{v})) \lesssim \|(u, \hat{u})\|_{1,h,*} \|(v, \hat{v})\|_{1,h,*} \quad \forall (u, \hat{u}), (v, \hat{v}) \in (V^{\text{reg}} \times \hat{V}^{\text{reg}}) + (V_h \times \hat{V}_h).$$

Proof. Follows by the Cauchy-Schwarz inequality and is left as exercise. □

In contrast to the continuity result, the bilinear form is only coercive on the discrete space.

Lemma 13. *There holds the coercivity estimate*

$$a^{HDG}((u_h,\hat{u}_h),(u_h,\hat{u}_h)) \gtrsim \|(u_h,\hat{u}_h)\|_{1,h}^2 \quad \forall (u_h,\hat{u}_h) \in V_h \times \hat{V}_h. \tag{2.35}$$

Proof. The proof follows with the Cauchy Schwarz, the Young and the inverse inequality for polynomials similarly as in the stability proof of the Nitsche penalty methods, see Lemma 9. This immediately shows why coercivity only holds on the discrete space. □

Theorem 18. *There exists a unique solution of the HDG variational formulation (2.34). Further, let $u \in H_0^1(\Omega,\mathbb{R}) \cap V^{\text{reg}}$ be the exact solution of (2.31) (with $u_D = 0$) and let $\hat{u} := u|_{\mathcal{F}_h}$. There holds the Cea-like best approximation result*

$$\|(u-u_h,\hat{u}-\hat{u}_h)\|_{1,h,*} \lesssim \inf_{(v_h,\hat{v}_h)\in V_h\times\hat{V}_h} \|(u-v_h,\hat{u}-\hat{v}_h)\|_{1,h,*}.$$

Proof. Existence and uniqueness (of the discrete method) follows with the Lax-Milgram theorem, Lemma 12 and Lemma 13. For the best approximation results let $(v_h,\hat{v}_h) \in V_h \times \hat{V}_h$ be arbitrary, then the triangle inequality gives

$$\|(u-u_h,\hat{u}-\hat{u}_h)\|_{1,h,*} \leq \|(u-v_h,\hat{u}-\hat{v}_h)\|_{1,h,*} + \|(v_h-u_h,\hat{v}_h-\hat{u}_h)\|_{1,h,*}.$$

Using the continuity of the exact solution Lemma 10 gives the Galerkin orthogonality

$$a^{HDG}((u-u_h,\hat{u}-\hat{u}_h),(v_h,\hat{v}_h)) = 0 \quad \forall (v_h,\hat{v}_h) \in V_h \times \hat{V}_h,$$

thus using Lemma 13 and Lemma 11 we have

$$\begin{aligned}
\|(v_h-u_h,\hat{v}_h-\hat{u}_h)\|_{1,h,*}^2 &\sim \|(v_h-u_h,\hat{v}_h-\hat{u}_h)\|_{1,h}^2 \\
&\lesssim a^{HDG}((v_h-u_h,\hat{v}_h-\hat{u}_h),(v_h-u_h,\hat{v}_h-\hat{u}_h)) \\
&= a^{HDG}((v_h-u,\hat{v}_h-\hat{u}),(v_h-u,\hat{v}_h-\hat{u})) \\
&\lesssim \|(u-v_h,\hat{u}-\hat{v}_h)\|_{1,h,*}\|(v_h-u_h,\hat{v}_h-\hat{u}_h)\|_{1,h,*}.
\end{aligned}$$

□

Lemma 14. *Let $u \in H_0^1(\Omega,\mathbb{R}) \cap H^l(\mathcal{T}_h,\mathbb{R})$ be the exact solution of (2.31) (with $u_D = 0$)*

and let $\hat{u} := u|_{\mathcal{F}_h}$*, there holds the approximation result*

$$\|(u - u_h, \hat{u} - \hat{u}_h)\|_{1,h,*} \lesssim h^s \|u\|_{H^{s+1}(\mathcal{T}_h)},$$

where $s = \min(k, l-1)$*.*

Proof. Let $I_{HDG} : V^{\mathrm{reg}} \times \hat{V}^{\mathrm{reg}} \to V_h \times \hat{V}_h$ be the element and facet wise L^2-projection, i.e. we have $I_{HDG}(u, \hat{u}) = (\Pi^k_{\mathcal{T}_h} u, \Pi^k_{\mathcal{F}_h} \hat{u})$. Scaling arguments and the Bramble-Hilbert Lemma show that there holds the approximation result (assuming enough regularity)

$$\|I_{HDG}(u, \hat{u}) - (u, \hat{u})\|_{1,h,*} \lesssim h^s |u|_{H^{s+1}(\mathcal{T}_h)}.$$

Then the result follows by Theorem 18. □

We finish this section with a discussion regarding the computational costs and the sparsity pattern. As mentioned in the previous section, a main disadvantage of a DG approach is the increased coupling between neighbouring elements, see right picture of Figure 2.4. Although an HDG further increases the number of unknowns, the sparsity pattern, of the condensed system, is much smaller. To analyse this in detail we define for all $(u_h, \hat{u}_n), (v_h, \hat{v}_h) \in V_h \times \hat{V}_h$ the bilinear forms

$$\begin{aligned}
a^{TT}((u_h, 0), (v_h, 0)) &:= \sum_{T \in \mathcal{T}_h} \int_T \nabla u_h \cdot \nabla v_h \,\mathrm{d}x + \int_{\partial T} -\nabla u_h \cdot n v_h - \nabla v_h \cdot n u_h + \frac{\alpha k^2}{h} u_h v_h \,\mathrm{d}s \\
a^{TF}((u_h, 0), (0, \hat{v}_h)) &:= \sum_{T \in \mathcal{T}_h} \int_{\partial T} \nabla u_h \cdot n \hat{v}_h \,\mathrm{d}s - \frac{\alpha k^2}{h} \int_{\partial T} u_h \hat{v}_h \,\mathrm{d}s \\
a^{FT}((0, \hat{u}_h), (v_h, 0)) &:= \sum_{T \in \mathcal{T}_h} \int_{\partial T} \nabla v_h \cdot n \hat{u}_h \,\mathrm{d}s - \frac{\alpha k^2}{h} \int_{\partial T} v_h \hat{u}_h \,\mathrm{d}s \\
a^{FF}((0, \hat{u}_h), (0, \hat{v}_h)) &:= \sum_{T \in \mathcal{T}_h} \frac{\alpha k^2}{h} \int_{\partial T} \hat{v}_h \hat{u}_h \,\mathrm{d}s \,.
\end{aligned}$$

Note that $a^{HDG} = a^{TT} + a^{TF} + a^{FT} + a^{FF}$. Using again $u_h, \hat{u}_h$ as symbols for the coefficients of the solutions, the discrete problem (2.34) can be written as

$$\begin{pmatrix} A^{TT} & A^{FT} \\ A^{TF} & A^{FF} \end{pmatrix} \begin{pmatrix} u_h \\ \hat{u}_h \end{pmatrix} = \begin{pmatrix} f_h \\ 0 \end{pmatrix}$$

where $A^{TT}, A^{TF}, A^{FT}, A^{FF}$ are the corresponding system matrices of $a^{TT}, a^{TF}, a^{FT}, a^{FF}$, respectively and f_h is the right hand side vector. Since A^{TT} is block diagonal we can invert it on each element separately (computational cheap!). This allows to condense the local

variable $u_h = (A^{TT})^{-1}(f - A^{FT}\hat{u}_h)$ and thus we get

$$(A^{FF} - A^{TF}(A^{TT})^{-1}A^{FT})\hat{u}_h = -A^{TF}(A^{TT})^{-1}f.$$

In Figure 2.5 we can see the sparsity pattern of the corresponding matrices. On the left side we can clearly see the sub matrices $A^{TT}, A^{TF}, A^{FT}, A^{FF}$ and the block structure of A^{TT}. On the right side we see the much smaller condensed system of $(A^{FF} - A^{TF}(A^{TT})^{-1}A^{FT})$ that needs to be solved.

Remark 11. Note, that the local matrices are invertible since on each element $T \in \mathcal{T}_h$ the bilinear form a^{TT} equals the Nietsche bilinear form a^{N1} for the case $\Omega = T$. Hence, inverting A^{TT} corresponds to solving a Poisson problem on T with a weak incorporation of homogeneous Dirichlet boundary conditions on ∂T.

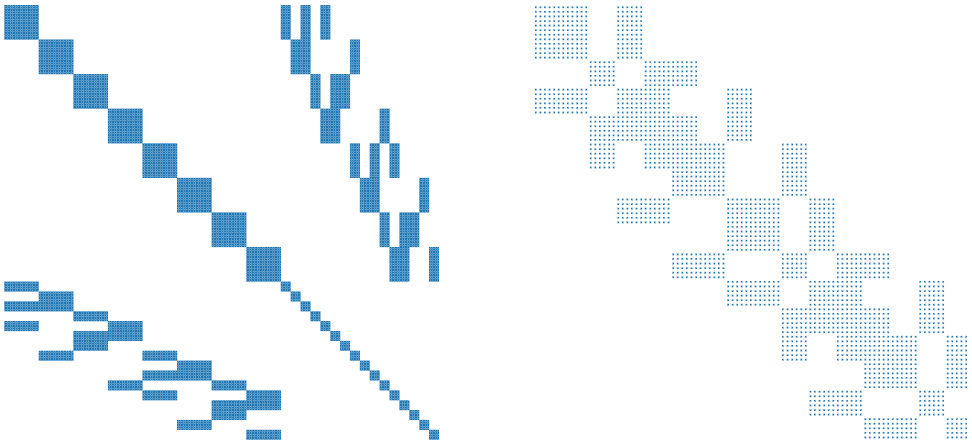


Figure 2.5: Sparsity pattern of a hybridized discontinuous Galerkin approximation of the Poisson problem with $k = 5$ on a regular triangulation with 8 elements on the domain $\Omega = (0, 1)^2$. Left we see the pattern of the system matrix of size 264×264 before static condensation and right the pattern of the system matrix of size 96×96 after static condensation.

2.6.2 Hybrid discontinuous Galerkin method for the Stokes equation

We now want to apply the HDG techniques from the previous section also to the Stokes equations.

A fully discontinuous approach

We define the discrete spaces as

$$\begin{aligned} V_h &:= \mathbb{P}^k(\mathcal{T}_h, \mathbb{R}^d), \\ \hat{V}_h &:= \{\hat{v}_h \in \mathbb{P}^k(\mathcal{F}_h, \mathbb{R}^d) : \hat{v}_h = 0 \text{ on } \partial\Omega\}, \\ Q_h &:= \mathbb{P}^{k-1}(\mathcal{T}_h, \mathbb{R}) \cap Q. \end{aligned}$$

On these spaces we define for all $(u_h, \hat{u}_h), (v_h, \hat{v}_h) \in V_h \times \hat{V}_h$ and $q_h \in Q_h$ the bilinear forms

$$\begin{aligned} a^{HDG}((u_h, \hat{u}_h), (v_h, \hat{v}_h)) := \sum_{T\in\mathcal{T}_h} \int_T \nu\varepsilon(u_h) : \varepsilon(v_h) dx - \int_{\partial T} \nu\varepsilon(u_h) \cdot n(v_h - \hat{v}_h)\,\mathrm{d}s \\ - \int_{\partial T} \nu\varepsilon(v_h) \cdot n(u_h - \hat{u}_h)\,\mathrm{d}s + \frac{\nu\alpha k^2}{h} \int_{\partial T} (u_h - \hat{u}_h)(v_h - \hat{v}_h)\,\mathrm{d}s, \\ b^{HDG}((u_h, \hat{u}_h), q_h) := \sum_{T\in\mathcal{T}_h} - \int_T \operatorname{div}(u_h) q_h\,\mathrm{d}x + \int_{\partial T} (u_h - \hat{u}_h) \cdot n q_h ds. \end{aligned}$$

The definition of the incompressibility constraint follows the same ideas as in the derivation of a^{HDG} (see also the proof of the consistency Lemma below). Now we have the problem: Find $((u_h, \hat{u}_h), p_h) \in (V_h \times \hat{V}_h) \times Q_h$ such that

$$a^{HDG}((u_h, \hat{u}_h), (v_h, \hat{v}_h)) + b^{HDG}((v_h, \hat{v}_h), p_h) = (f, v_h) \quad \forall (v_h, \hat{v}_h) \in V_h \times \hat{V}_h \tag{2.36a}$$

$$b^{HDG}((u_h, \hat{u}_h), q_h) = 0 \qquad \forall q_h \in Q_h. \tag{2.36b}$$

Again there holds the following consistency results.

Lemma 15. *Let $u \in H_0^1(\Omega, \mathbb{R}^d) \cap H^2(\mathcal{T}_h, \mathbb{R}^d)$ with $\hat{u} := u|_{\mathcal{F}_h}$ and $p \in Q \cap H^1(\mathcal{T}_h, \mathbb{R})$ be the exact solution of (2.7). There holds the consistency result*

$$\begin{aligned} a^{HDG}((u, \hat{u}), (v_h, \hat{v}_h)) + b^{HDG}((v_h, \hat{v}_h), p) = (f, v_h) \quad \forall (v_h, \hat{v}_h) \in V_h \times \hat{V}_h \\ b^{HDG}((u, \hat{u}), q_h) = 0 \qquad \forall q_h \in Q_h. \end{aligned}$$

Proof. Similarly as in the proof of Lemma 10 the regularity $f \in L^2(\Omega, \mathbb{R}^d)$ shows that $f = -\operatorname{div}(\nu\varepsilon(u) + pI) \in L^2(\Omega, \mathbb{R}^d)$ which gives that the stress $\nu\varepsilon(u) + pI$ is normal continuous and thus since $\hat{v}_h$ is single valued we have

$$\sum_{T\in\mathcal{T}_h} \int_{\partial T} (-\nu\varepsilon(u) + pI) \cdot n\hat{v}_h\,\mathrm{d}s = 0$$

Using the continuity of the velocity solution and integration by parts locally on each element then gives

$$\begin{aligned} a^{HDG}((u,\hat{u}),(v_h,\hat{v}_h)) + b^{HDG}((v_h,\hat{v}_h),p) &= \sum_{T\in\mathcal{T}_h}\int_T \nu\varepsilon(u)\cdot\varepsilon(v_h)dx - \int_{\partial T}\nu\varepsilon(u)\cdot n v_h \,\mathrm{d}s \\ &\quad + \sum_{T\in\mathcal{T}_h} -\int_T \operatorname{div}(v_h)p\,\mathrm{d}x + \int_{\partial T} v_h\cdot npds \\ &= \sum_{T\in\mathcal{T}_h}\int_T -\operatorname{div}(\nu\varepsilon(u)+pI)v_h dx \\ &= \sum_{T\in\mathcal{T}_h}\int_T f\cdot v_h\,\mathrm{d}x\,. \end{aligned}$$

□

Since all the techniques from the previous section can be adapted to prove continuity and (kernel) coercivity of a^{HDG}, we only discuss well posedness, i.e. continuity and the inf-sup condition, for the incompressibility constraint b^{HDG}. To this end we extend the definition of the norms $\|\cdot\|_{1,h}$ and $\|\cdot\|_{1,h,*}$ onto the vector valued velocity spaces $V_h\times\hat{V}_h$. On the pressure space we use the L^2-norm and further introduce the norm

$$\|q\|_{0,*}^2 = \sum_{T\in\mathcal{T}_h}\|q_h\|_T^2 + h\|q\|_{\partial T}^2.$$

Similarly as in the previous section, the norm $\|q\|_{0,*}$ is needed to prove continuity of above bilinear forms with respect to the spaces

$$\begin{aligned} V^{\mathrm{reg}} &:= H^1(\Omega,\mathbb{R}^d)\cap H^2(\mathcal{T}_h,\mathbb{R}^d), \\ \hat{V}^{\mathrm{reg}} &:= \{\hat{u}\in L^2(\mathcal{F}_h,\mathbb{R}^d) \text{ with } \hat{u}=0 \text{ on } \partial\Omega\}, \\ Q^{\mathrm{reg}} &:= L_0^2(\Omega)\cap H^1(\mathcal{T}_h,\mathbb{R}), \end{aligned}$$

where the pressure space includes a local H^1 regularity such that the evaluation on element boundaries is applicable. As for the velocity space, a scaling argument shows that on the discrete pressure space there holds the norm equivalence

$$\|q_h\|_0 \sim \|q_h\|_{0,*} \quad \forall q_h\in Q_h.$$

There holds the following stability result.

Lemma 16. *The bilinear form b^{HDG} is continuous*

$$b^{HDG}((u,\hat{u}),q) \lesssim \|u,\hat{u}\|_{1,h,*}\|q\|_{0,*} \quad \forall (u_h,\hat{u}_h) \in (V^{\mathrm{reg}} \times \hat{V}^{\mathrm{reg}}) + (V_h \times \hat{V}_h), \forall q \in Q^{\mathrm{reg}} + Q_h.$$

and there holds the discrete LBB condition

$$\sup_{(u_h,\hat{u}_h)\in V_h\times\hat{V}_h} \frac{b^{HDG}((u_h,\hat{u}_h),q_h)}{\|u_h,\hat{u}_h\|_{1,h}} \gtrsim \|q_h\|_0 \quad \forall q_h \in Q_h.$$

Proof. The continuity result follows simply by the Cauchy Schwarz inequality. The LBB proof follows with the technique of the mesh dependent norms and an adaption of the results of Theorem 13 (where we need to exchange b by b^{HDG}). For this note that the Clément interpolant $v_h := I_{\mathcal{C}}v$ for all $v \in H_0^1(\Omega,\mathbb{R}^d)$ is continuous and thus particularly also an element of V_h. Thus by setting $\hat{v}_h := v_h|_{\mathcal{F}_h}$ we further have $\|v_h,\hat{v}_h\|_{1,h} = \|v_h\|_1$ and the arguments of Theorem 13 also hold for b^{HDG}. This shows that it is sufficient to prove the modified LBB

$$\sup_{(u_h,\hat{u}_h)\in V_h\times\hat{V}_h} \frac{b^{HDG}((u_h,\hat{u}_h),q_h)}{\|u_h,\hat{u}_h\|_{1,h}} \gtrsim \|q_h\|_{0,h} \quad \forall q_h \in Q_h.$$

Now let $q_h \in Q_h$ be arbitrary, then we define on each element $u_h := h^2\nabla q_h$ and with a fixed normal vector on each facet further $\hat{u}_h := -h[\![q_h]\!]^* n$. Using integration by parts then gives

$$\begin{aligned} b^{HDG}((u_h,\hat{u}_h),q_h) &= \sum_{T\in\mathcal{T}_h}\int_T u_h\cdot\nabla q_h\,\mathrm{d}x + \int_{\partial T} -\hat{u}_h\cdot n q_h ds \\ &= \sum_{T\in\mathcal{T}_h}\int_T u_h\cdot\nabla q_h\,\mathrm{d}x + \sum_{F\in\mathcal{F}_h}\int_F -\hat{u}_h\cdot n[\![q_h]\!]^* ds \geq \|q_h\|_{0,h}^2, \end{aligned}$$

Further we have

$$\|u_h,\hat{u}_h\|_{1,h}^2 = \sum_{T\in\mathcal{T}_h}\|u_h\|_T^2 + \frac{k^2}{h}\|u_h-\hat{u}_h\|_{\partial T}^2 = \sum_{T\in\mathcal{T}_h} h^4\|\nabla^2 q_h\|_T^2 + \frac{k^2}{h}\|h^2\nabla q_h + h[\![q_h]\!]^* n\|_{\partial T}^2.$$

By a scaling argument we have on each element $h^4\|\nabla^2 q_h\|_T^2 \leq h^2\|\nabla q_h\|_T^2$. Next we use the triangle inequality (and that $|n|=1$) to split the boundary term into two parts

$$\sum_{T\in\mathcal{T}_h}\frac{k^2}{h}\|h^2\nabla q_h + h[\![q_h]\!]^* n\|_{\partial T}^2 \leq \sum_{T\in\mathcal{T}_h}\frac{k^2}{h}\|h^2\nabla q_h\|_{\partial T}^2 + \frac{k^2}{h}\|h[\![q_h]\!]^*\|_{\partial T}^2.$$

By the inverse inequality, the first sum can be bounded again by the element terms since $\|h^2\nabla q_h\|^2_{\partial T} \lesssim h^{-1}\|h^2\nabla q_h\|^2_T$. In total this gives

$$\|u_h, \hat{u}_h\|^2_{1,h} \lesssim \sum_{T\in\mathcal{T}_h} h^2\|\nabla q_h\|^2_T + \frac{k^2}{h}\|h[\![q_h]\!]\|^2_{\partial T} \lesssim \|q_h\|^2_{0,h},$$

where the constants depend on the polynomial order k. This proves that the modified LBB condition holds true and thus we conclude the proof. □

Theorem 19. *There exists an unique solution $(u_h, \hat{u}_h), p_h \in (V_h \times \hat{V}_h) \times Q_h$ of problem (2.36). Let $u \in H^1_0(\Omega, \mathbb{R}^2) \cap H^l(\mathcal{T}_h, \mathbb{R}^d)$ with $\hat{u} := u|_{\mathcal{F}_h}$ and $p \in L^2_0(\Omega) \cap H^{l-1}(\mathcal{T}_h, \mathbb{R})$ be the exact solution of (2.7), there holds the approximation result*

$$\|(u - u_h, \hat{u} - \hat{u}_h)\|_{1,h,*} + \frac{1}{\nu}\|p - p_h\|_{0,*} \lesssim h^s(\|u\|_{H^{s+1}(\mathcal{T}_h)} + \frac{1}{\nu}\|p\|_{H^s(\mathcal{T}_h)}).$$

where $s = \min(k, l-1)$.

Proof. The existence follows by Theorem 10 and above stability results. The approximation results are derived with the same techniques as in Section 2.5.5, above stability results and the consistency results of Lemma 15. □

An $H(\mathrm{div})$-conforming approach

The definition of the bilinear form b^{HDG} above shows that in order to guarantee consistency we needed to add the additional terms

$$\sum_{T\in\mathcal{T}_h} \int_{\partial T} (u_h - \hat{u}_h) \cdot n q_h \,\mathrm{d}s\,.$$

From a more mathematical point of view, the integrals on the boundary can be interpreted as additional edge distributions that results from taking the weak divergence of a discontinuous function u_h. These findings motivate to define an HDG method that lies between a fully H^1-conforming and a fully discontinuous approach as above such that the weak divergence (but not the full gradient) is well defined. To this end we define the following

discrete spaces

$$\begin{aligned}
V_h :=& \mathrm{BDM}^k(\mathcal{T}_h, \Omega) \cap H_0(\mathrm{div}, \Omega) \\
=& \{v_h \in \mathbb{P}^k(\mathcal{T}_h, \mathbb{R}^d) : [\![v_h \cdot n]\!] = 0 \text{ on all } F \in \mathcal{F}_h, v_h \cdot n = 0 \text{ on } \partial\Omega \ \}, \\
\hat{V}_h :=& \{\hat{v}_h \in \mathbb{P}^k(\mathcal{F}_h, \mathbb{R}^d) : \hat{v}_h|_F \cdot n = 0 \ \forall F \in \mathcal{F}_h, \hat{v}_h = 0 \text{ on } \partial\Omega\}, \\
Q_h :=& \mathbb{P}^{k-1}(\mathcal{T}_h, \mathbb{R}) \cap Q.
\end{aligned}$$

Hence, in contrast to before the velocity space V_h now is normal continuous and by that the weak divergence is well defined. Note, that the facet space $\hat{V}_h$ only consists of polynomials in tangential direction, thus for example in two dimensions we have on each $F \in \mathcal{F}_h$ and $\hat{v}_h \in \hat{V}_h$

$$\hat{v}_h|_F \in \{t\xi_h : \xi_h \in P^k(F, \mathbb{R})\},$$

where t is the tangential vector on F. Since the facet space is needed to incorporate H^1 conformity in a weak sense, and normal continuity is already considered in V_h, it makes sense that $\hat{V}_h$ only lies in the tangential plane. Next, let $\gamma_t(\cdot) = \cdot_t$ be the tangential projection on each facet, i.e. we have

$$\gamma_t \phi = \phi_t = \phi - (\phi \cdot n)n,$$

for all smooth enough functions ϕ, then we define for all $(u_h, \hat{u}_h), (v_h, \hat{v}_h) \in V_h \times \hat{V}_h$ and $q_h \in Q_h$ the bilinear forms

$$\begin{aligned}
a^{HDG}((u_h, \hat{u}_h), (v_h, \hat{v}_h)) :=& \sum_{T \in \mathcal{T}_h} \int_T \nu\varepsilon(u_h) : \varepsilon(v_h) dx - \int_{\partial T} \nu\varepsilon(u_h) \cdot n(v_h - \hat{v}_h)_t \, \mathrm{d}s \\
&- \int_{\partial T} \nu\varepsilon(v_h) \cdot n(u_h - \hat{u}_h)_t \, \mathrm{d}s + \frac{\nu\alpha k^2}{h} \int_{\partial T} (u_h - \hat{u}_h)_t (v_h - \hat{v}_h)_t \, \mathrm{d}s, \\
b^{HDG}(u_h, q_h) :=& \sum_{T \in \mathcal{T}_h} - \int_T \mathrm{div}(u_h) q_h \, \mathrm{d}x,
\end{aligned}$$

and the problem: Find $((u_h, \hat{u}_h), p_h) \in (V_h \times \hat{V}_h) \times Q_h$ such that

$$a^{HDG}((u_h, \hat{u}_h), (v_h, \hat{v}_h)) + b^{HDG}(v_h, p_h) = (f, v_h) \quad \forall (v_h, \hat{v}_h) \in V_h \times \hat{V}_h \tag{2.37a}$$

$$b^{HDG}(u_h, q_h) = 0 \qquad \forall q_h \in Q_h. \tag{2.37b}$$

Note that a^{HDG} now only includes tangential jumps. Further, since $V_h \subset H(\mathrm{div}, \Omega)$ we have that $b^{HDG}(u_h, q_h) = -(\mathrm{div}\, u_h, q_h)$. With the same techniques as before we can easily

proof that (2.37) is a consistent method. By defining the discrete velocity norms now as

$$\|(u,\hat{u})\|_{1,h}^2 := \sum_{T\in\mathcal{T}_h} \|\nabla u\|_T^2 + \frac{k^2}{h}\|(u-\hat{u})_t\|_{\partial T}^2$$

and again use the L^2-norm on the pressure space we further have the stability result

Lemma 17. *The bilinear form b^{HDG} is continuous on*

$$b^{HDG}(u_h,q_h) \lesssim (\sum_{T\in\mathcal{T}_h} \|\nabla u_h\|_T^2)^{1/2}\|q_h\|_0 \quad \forall (u_h,q_h)\in (V^{\mathrm{reg}}+V_h)\times Q.$$

Further there holds the discrete LBB condition

$$\sup_{(u_h,\hat{u}_h)\in V_h\times\hat{V}_h} \frac{b^{HDG}(u_h,q_h)}{\|u_h,\hat{u}_h\|_{1,h}} \gtrsim \|q_h\|_0 \quad \forall q_h\in Q_h.$$

Proof. The continuity follows with an element-wise Cauchy-Schwarz argument. For the ease we will again provide the proof only in two dimensions. The other case follows with the same steps. Let $I_{\mathrm{BDM}} : V \to V_h$ be the standard interpolation operator into the BDM space as presented in 2.5.6, or as in [8], and let $\Pi^t_{\mathcal{F}_h} : \hat{V} \to \hat{V}_h$ be the facet wise tangential L^2-projection, i.e. we have

$$\int_F \Pi^t_{\mathcal{F}_h}\hat{v}\cdot\hat{v}_h \,\mathrm{d}s = \int_F \hat{v}\cdot\hat{v}_h\,\mathrm{d}s \quad \forall F\in\mathcal{F}_h, \forall \hat{v}_h\in\hat{V}_h.$$

Then we define the Fortin operator $I_F := (I_{\mathrm{BDM}}, \Pi^t_{\mathcal{F}_h})$. By the functionals of the BDM interpolation operator we already have (using a restriction on V_h) for all $u\in V$ that

$$b^{HDG}(I_F u, q_h) = b^{HDG}(u,q_h) \quad \forall q_h\in Q_h.$$

It remains to prove stability $\|(I_{\mathrm{BDM}}u, \Pi^t_{\mathcal{F}_h}\hat{u})\|_{1,h} \lesssim \|u\|_1$, where $\hat{u} = (u|_F)_t$ on all facets $F\in\mathcal{F}_h$. On each element the H^1- stability of the BDM interpolator already gives $\|\nabla I_{\mathrm{BDM}}u\|_T \lesssim \|\nabla u\|_T$ thus using the triangle inequality we then have

$$\|(I_{\mathrm{BDM}}u, \Pi^t_{\mathcal{F}_h}\hat{u})\|_{1,h}^2 \lesssim \sum_{T\in\mathcal{T}_h} \|\nabla u\|_T^2 + \frac{k^2}{h}\|(I_{\mathrm{BDM}}u-u)_t\|_{\partial T}^2 + \frac{k^2}{h}\|(\hat{u}-\Pi^t_{\mathcal{F}_h}\hat{u})_t\|_{\partial T}^2.$$

We start with the last term. In section 2.5.6 we have proven that the Piola mapping is the er mapping for the normal component. Similarly, one shows that the covariant mapping is the proper transformation for the tangential components, i.e. it preserves the tangential component. For the ease of notation we use a tilde in this proof to denote quantities on

the reference element $\tilde{T}$ (instead of $\hat{T}$). Now let $F \subset \partial T$ with $F = \phi_T(\tilde{F})$ and the function $\tilde{u}$ such that $u = F_T^{-\mathrm{T}}\tilde{u}$ (i.e. covariant mapped). Next note, that the tangential L^2-projection is interpolation equivalent, i.e. we have $\Pi^t_{\mathcal{F}_h}\hat{u} = \Pi^t_{\mathcal{F}_h}(u|_F)_t = F_T^{-\mathrm{T}}\Pi^t_{\tilde{F}}(\tilde{u}|_{\tilde{F}})_{\tilde{t}}$, where $\Pi^t_{\tilde{F}}$ is the tangential L^2-projection on the reference facet $\tilde{F}$ and $\tilde{t}$ is the reference tangential vector. This gives

$$\|(\hat{u} - \Pi^t_{\mathcal{F}_h}\hat{u})_t\|^2_{\partial T} = \int_F (\hat{u} - \Pi^t_{\mathcal{F}_h}\hat{u})_t^2 \,\mathrm{d}s = h^{-1}\int_{\hat{F}} (\tilde{u} - \Pi^t_{\tilde{F}}\tilde{u})_{\tilde{t}}^2 \,\mathrm{d}\hat{s} = h^{-1}\|\tilde{u} - \Pi^t_{\tilde{F}}\tilde{u}\|^2_{\tilde{F}}.$$

On the reference element we use the continuity of the L^2-projection and the trace inequality to get

$$h^{-1}\|\tilde{u} - \Pi^t_{\tilde{F}}\tilde{u}\|^2_{\tilde{F}} \leq h^{-1}\|\tilde{u}\|^2_{\tilde{F}} \leq h^{-1}(\|\tilde{u}\|^2_{\tilde{T}} + \|\nabla\tilde{u}\|^2_{\tilde{T}}) \leq h^{-1}\|u\|^2_T + h\|\nabla u\|^2_T,$$

where we used a scaling argument (using the covariant mapping!) in the last step. In total this gives

$$\sum_{T\in\mathcal{T}_h} \frac{k^2}{h}\|(\hat{u} - \Pi^t_{\mathcal{F}_h}\hat{u})_t\|^2_{\partial T} \lesssim \sum_{T\in\mathcal{T}_h} h^{-2}\|u\|^2_T + \|\nabla u\|^2_T.$$

With the same technique we also prove the other boundary term. To this end let $w := (I_{\mathrm{BDM}}u - u)$, then as before we get the estimate

$$\|w_t\|^2_{\partial T} \lesssim h^{-2}\|w\|^2_T + \|\nabla w\|^2_T,$$

hence by the approximation properties (and the continuity) of I_{BDM} we then have $\|w\|_T \leq h\|\nabla u\|_T$ and thus in total finally get

$$\|(I_{\mathrm{BDM}}u, \Pi^t_{\mathcal{F}_h}\hat{u})\|^2_{1,h} \lesssim \sum_{T\in\mathcal{T}_h} \|\nabla u\|^2_T + \frac{1}{h^2}\|u\|^2_T \lesssim \|u\|^2_1,$$

and we can conclude with Theorem 14. □

Remark 12. Note that although above proof result does not provide robustness with respect to the polynomial order k one can indeed show that the inf-sup constant does not depend on k, see [28].

Remark 13. The continuity estimate can be trivially extended to $(V^{\mathrm{reg}} \times \hat{V}^{\mathrm{reg}}) + (V_h \times \hat{V}_h)$ using $\|\cdot\|_{1,h}$ on the right hand side. Further note that we do need increased regularity $q_h \in Q^{reg}$.

The remarkable property of the $H(\mathrm{div})$-conforming HDG approximation is that the dis-

crete velocity is also exactly divergence free, i.e. there holds a "non conforming" kernel inclusion property. To see this simply choose the test function $q_h = \operatorname{div}(u_h)$, then we have for the solution u_h by the second line of (2.37) that

$$0 = b^{HDG}(u_h, q_h) = -\int_\Omega |\operatorname{div}(u_h)|^2 \,\mathrm{d}x \Rightarrow \operatorname{div}(u_h) = 0.$$

This immediately shows that the method is also pressure robust and we can derive the following error estimate.

Theorem 20. *There exists a unique solution $(u_h, \hat{u}_h), p_h \in (V_h \times \hat{V}_h) \times Q_h$ of problem (2.37). Let $u \in H_0^1(\Omega, \mathbb{R}^2) \cap H^l(\mathcal{T}_h, \mathbb{R}^d)$ with $\hat{u} := u|_{\mathcal{F}_h}$ and $p \in L_0^2 \cap H^{l-1}(\mathcal{T}_h, \mathbb{R})$ be the exact solution of (2.7), there holds the approximation result*

$$\|(u - u_h, \hat{u} - \hat{u}_h)\|_{1,h,*} + \frac{1}{\nu}\|p - p_h\|_0 \lesssim h^s(\|u\|_{H^{s+1}(\mathcal{T}_h)} + \frac{1}{\nu}\|p\|_{H^s(\mathcal{T}_h)}).$$

where $s = \min(k, l-1)$. Further there holds the pressure robust error estimate

$$\|(u - u_h, \hat{u} - \hat{u}_h)\|_{1,h,*} \lesssim h^s \|u\|_{H^{s+1}(\mathcal{T}_h)}.$$

Proof. The existence follows by Theorem 10 and above stability results. The approximation results are derived with the same techniques as in Section 2.5.5. The pressure robustness follows with the error estimates as in section 2.5.6 and the exact divergence-free property of the discrete solution u_h. □

2.6.3 The MCS method

PL: Will be updated

3 The stationary Navier-Stokes equations

3.1 Variational formulation of the stationary Navier-Stokes equations

This chapter is dedicated to analyze and approximate the incompressible stationary Navier-Stokes equations (1.13), thus including homogeneous Dirichlet boundary conditions we aim to find a solution u, p such that

$$\begin{aligned} -\nu \operatorname{div}(\varepsilon(u)) + \operatorname{div}(u \otimes u) + \nabla p &= f \quad \text{in } \Omega, \\ \operatorname{div}(u) &= 0, \quad \text{in } \Omega, \\ u &= 0, \quad \text{on } \partial\Omega. \end{aligned}$$

In a first step we will derive the weak formulation of the above problem. Multiplying each equation with an appropriate test function and integrating by parts we derive the weak formulation: Find $(u, p) \in V \times Q$ such that

$$a(u, v) + c(u, u, v) + b(v, p) = f(v) \quad \forall v \in V, \tag{3.1a}$$

$$b(u, q) = 0 \qquad \forall q \in Q, \tag{3.1b}$$

where the bilinear forms a and b are defined as for the Stokes equations in (2.11), i.e.

$$a(u, v) = \int_\Omega \nu\varepsilon(u) : \varepsilon(v) \, \mathrm{d}x, \quad \text{and} \quad b(u, q) = -\int_\Omega \operatorname{div}(u) q \, \mathrm{d}x \,. \tag{3.2}$$

The convective trilinear form c can be defined in several different ways. For this first note that by the incompressibility constraint (3.3b) we can derive the following identities

$$\operatorname{div}(u \otimes u) = (u \cdot \nabla)u + \operatorname{div}(u)u = (u \cdot \nabla)u = \operatorname{curl}(u) \times u + \frac{1}{2}\nabla(u^2),$$

thus c is given by one of the following forms

$$\begin{aligned} c_\nabla(w,u,v) &:= \int_\Omega (w\cdot\nabla)u\cdot v\,\mathrm{d}x,\\ c_{\mathrm{div}}(w,u,v) &:= \int_\Omega ((w\cdot\nabla)u+\frac{1}{2}(\mathrm{div}(w)u))\cdot v\,\mathrm{d}x,\\ c_{\mathrm{curl}}(w,u,v) &:= \int_\Omega \mathrm{curl}(u)\times w\cdot v\,\mathrm{d}x\\ c_{\mathrm{skw}}(w,u,v) &:= \frac{1}{2}(c_\nabla(w,u,v)-c_\nabla(w,v,u)). \end{aligned}$$

For the definition of c_{div} we added the factor $1/2$ because this will give us skew symmetry, see lemma below. Further note, that in the case of $c(w,u,v)=c_{\mathrm{curl}}(w,u,v)$ the pressure in (3.1) is redefined to the so called Bernoulli pressure $p\to p+\frac{1}{2}u^2$. A crucial property for the stability analysis is the property of skew symmetry of the convective trilinear form.

Lemma 18. *Let $w,u\in H^1(\Omega,\mathbb{R}^d)$, then*

$$c_{\mathrm{curl}}(w,u,u)=c_{\mathrm{skw}}(w,u,u)=0.$$

If either $w\cdot n=0$ or if $u\in H^1_0(\Omega,\mathbb{R}^2)$ we further have

$$c_{\mathrm{div}}(w,u,u)=0.$$

If w is weakly divergence free and $w\cdot n=0$ on $\partial\Omega$ we further have

$$c_\nabla(w,u,u)=0.$$

Further, let $u,v,w\in V$, then there holds the continuity estimate

$$c_i(w,u,v)\lesssim \|w\|_1\|u\|_1\|v\|_1 \quad \textit{where} \quad i\in\{\nabla,\mathrm{skw},\mathrm{div},\mathrm{curl}\}.$$

Proof. The results for c_{curl} and c_{skw} follow from the definition. For the rest we use integration by parts and the assumptions stated in the lemma. The continuity follows by several applications of the Cauchy-Schwarz inequality. □

Remark 14. In the case of partial Dirichlet boundary conditions integration by parts shows that

$$c_\nabla(w,u,v)=-c(w,v,u)+\int_{\partial\Omega} w\cdot n(u\cdot v)\,\mathrm{d}s,$$

hence we then set

$$c_{\mathrm{skw}}(w,u,v) := \frac{1}{2}(c_\nabla(w,u,v) - c_\nabla(w,v,u)) + \frac{1}{2}\int_{\partial\Omega} w\cdot n(u\cdot v)\,\mathrm{d}s\,.$$

With respect to the discretization of the instationary Navier-Stokes equations we are particularly interested in the uniqueness and stability of the solution. If this is not the case, small fluctuations in the input data would produce very different solutions. In such a case the instationary Navier-Stokes equations should be considered. In order to derive the stability results we will first consider the linearized Oseen equation. To this end let $b \in V_0$ be a given fixed convection "wind", then we study the problem:Find $(u,p) \in V \times Q$ such that

$$a(u,v) + c_\nabla(b,u,v) + (\xi u, v) + b(v,p) = f(v) \quad \forall v \in V, \tag{3.3a}$$

$$b(u,q) = g(q) \quad \forall q \in Q. \tag{3.3b}$$

The additional reaction bilinear form $(\xi u, v)$ is included in order to make the analysis more general. Further, with respect to the instationary Navier-Stokes equations this term might correspond to the time derivative. In the following we will assume that $\xi \in L^\infty(\Omega)$ with $\xi(x) \geq 0$.

Lemma 19. *There exists a unique solution* $(u,p) \in V \times Q$ *of* (3.3) *such that*

$$\nu\|\nabla u\|_0^2 + \|\sqrt{\xi}u\|_0^2 \lesssim \frac{1}{\nu}\|f\|_{V^*} \quad \textit{and} \quad \|p\|_0 \lesssim \|f\|_{V^*} + c_p(\sqrt{\nu}\|\nabla u\|_0 + \|\sqrt{\xi}u\|_0),$$

with constant $c_p = (\sqrt{\nu} + \frac{\|b\|_\infty}{\sqrt{\nu}} + \|\xi\|_\infty^{1/2})$.

Proof. We aim to apply Brezzi's Theorem 10. First note, that lemma (18) and the Cauchy Schwarz inequality shows that the bilinear form

$$\tilde{a}(u,v) = a(u,v) + c_\nabla(b,u,v) + (\xi u, v)$$

is continuous. For the proof it remains to show that $\tilde{a}$ is coercive on the kernel V_0. Here, the crucial property is the skew symmetry of the convection bilinear form, i.e. we have $c_\nabla(b,u,u) = 0$ for all $u \in V$. This immediately gives

$$\tilde{a}(u,u) = \int_\Omega \nu\varepsilon(u) : \varepsilon(v)\,\mathrm{d}x + \int_\Omega \xi u\cdot v\,\mathrm{d}x \gtrsim \|u\|_1,$$

from which we conclude the existence. The stability results follow again by the Cauchy-Schwarz inequality and the positivity of ξ. A detailed proof is given in [25]. □

Using the results of the linearized equations we are now in the position of analyzing the non linear problem. To this end we define the constant

$$\mathcal{N}_0 := \sup_{w,u,v\in V_0} \frac{c_\nabla(w,u,v)}{\|w\|_V\|u\|_V\|v\|_V},$$

Theorem 21. *Assume that there holds the estimate*

$$\frac{\mathcal{N}_0\|f\|_{V^*}}{\nu^2 c_k} < 1,$$

(where c_k is the Korn inequality) then there exists a unique solution $(u,p) \in V \times Q$ of problem (3.1) with

$$\|\nabla u\|_0 \leq \frac{1}{\nu}\|f\|_{V^*} \quad \textit{and} \quad \|p\|_0 \lesssim \|f\|_{V^*} + \frac{1}{\nu^2}\|f\|_{V^*}^2.$$

Proof. For the existence of at least one solution of (3.1) we refer to [25, 16] since the proof is very technical and out of scope of this lecture. Nevertheless we prove uniqueness since it includes above assumption which might will also be essential for the discretization. To this end let $S : V_0 \to V_0$ be the solution operator that maps an arbitrary wind $b \in V_0$ to the solution of the Oseen problem (3.3) u_o. In the following we will show that S is a countinuous contradiction on V_0. The boundedness follows by

$$\|S\|_{V_0^*} = \sup_{b\in V_0, \|b\|_V=1} \|S(b)\|_V = \sup_{b\in V_0, \|b\|_V=1} \|u_o\|_V \leq \frac{1}{\nu}\|f\|_{V^*},$$

where we used the stability estimate of Lemma 19. Now let $b_1, b_2 \in V_0$ be arbitrary and let u_{o1} and u_{o2} be the corresponding solutions of (3.3) with the wind b_1 and b_2, respectively. Subtracting the equations (3.3) (with the same right hand side f) and testing with a divergence free test function gives

$$0 = a(u_{o1} - u_{o2}, v) + c_\nabla(b_1 - b_2, u_{o1}, v) + c_\nabla(b_2, u_{o1} - u_{o2}, v) \quad \forall v \in V_0.$$

Now choose $v = u_{o1} - u_{o2}$ to get (using again skew symmetry of c_∇)

$$\begin{aligned}\|u_{o1} - u_{o2}\|_1^2 &\leq \frac{1}{\nu c_k} c_\nabla(b_1 - b_2, u_{o1}, u_{o1} - u_{o2}) \\ &\leq \frac{\mathcal{N}_0}{\nu c_k} \|b_1 - b_2\|_1 \|u_{o1}\|_1 \|u_{o1} - u_{o2}\|_1 \\ &\leq \frac{\mathcal{N}_0 \|f\|_{V^*}}{c_k \nu^2} \|b_1 - b_2\|_1 \|u_{o1} - u_{o2}\|_1 \\ &< \|b_1 - b_2\|_1 \|u_{o1} - u_{o2}\|_V.\end{aligned}$$

Hence S is a contradiction and we conclude that there exists a unique solution of $u \in V_0$. The uniqueness of the pressure is now a consequence of the fact that V and Q satisfy the LBB-condition. □

Before we introduce finite element methods for the approximation of (3.1), we first discuss the approximation of a simplified set of equations in the next section.

3.2 Approximation of scalar convection-diffusion equations

In the previous section we saw that the existence proof of the stationary Navier-Stokes equations is based on the stability results of the linearized Oseen equations (3.3), which also motivates to first study approximation schemes for the latter one. Nevertheless, since these equations now include a transport term, we will first discuss the approximation of the much simpler scalar convection-diffusion equation to analyze the occurring difficulties resulting from the additional terms.

Let $b \in H(\mathrm{div}, \Omega) \cap L^\infty(\Omega, \mathbb{R}^d)$ be a divergence-free wind $\mathrm{div}(b) = 0$, then we consider the problem

$$\begin{aligned}-\nu \Delta u + b \cdot \nabla u &= f && \text{on } \Omega, \\ u &= u_D && \text{on } \Gamma_D, \\ \nabla u \cdot n &= g_N && \text{on } \Gamma_N,\end{aligned}$$

with a positive diffusion parameter $\nu > 0$. For the ease we used the same symbols as for the Navier-Stokes equations. The solution of the above equation will mainly be characterized by the wind b. According to the direction b we will now further split the boundary into

the following three parts

$$\begin{aligned}\Gamma_{in} &:= \{x \in \partial\Omega : b \cdot n < 0\},\\ \Gamma_{out} &:= \{x \in \partial\Omega : b \cdot n > 0\},\\ \Gamma_0 &:= \{x \in \partial\Omega : b \cdot n = 0\},\end{aligned}$$

representing the inflow, outflow and the so called characteristic boundary part. The additional convection term drastically changes the behaviour of the solution compared to the standard Poisson equation. Here, the crucial parameter will be the relation between the diffusive and the convective terms $\nu/|b|$ (considering a domain with diameter $\mathcal{O}(1)$). In the limiting case $\nu \to 0$ (without changing the wind b) the second order differential operator vanishes, hence we are not allowed to consider any boundary conditions anymore. The arising problem can be seen by considering the one dimensional problem $-\nu u'' + u' = 1$ on $\Omega = (0,1)$ with homogeneous Dirichlet boundary conditions on $\partial\Omega = \{0,1\}$. The exact solution is given by

$$u(x) = x(1 - e^{\frac{x-1}{\nu}})$$

If the diffusive parameter vanishes the exact solution is given by $u = x$. However, considering a small value $\nu \ll 1$, the homogeneous Dirichlet boundary conditions (in particular on the right side at the point 1) lead to a very thin boundary layer of size ν. Similarly one may also produce such **sharp gradients** inside of the domain if we consider for example a discontinuous boundary condition on the inflow boundary Γ_{in} which is transported by the wind into the inside. Although the tools developed from the functional analysis will prove solvability of the above problem in the continuous setting, these sharp gradients will play a crucial role when we aim to introduce a finite element approximation. The tools and techniques that we develop in this section can then also be applied the stationary and the instationary Navier-Stokes equations and will be particularly essential if we consider convection dominant flows, i.e. a high Reynolds number where turbulent flows will appear.

For the ease we only consider the case of homogeneous Dirichlet boundary conditions in the following. The general case follows as usual with a homogenization technique. Following the standard approach we can define the weak formulation: Find $u \in V := H^1_{0,\Gamma_D}(\Omega, \mathbb{R})$such that

$$a(u,v) + c(u,v) = f(v) \quad \forall v \in V \tag{3.4}$$

with

$$a(u,v) := \int_\Omega \nu \nabla u \cdot \nabla v \,\mathrm{d}x, \quad c(u,v) := \int_\Omega (b \cdot \nabla u) v \,\mathrm{d}x, \quad f(v) := \int_\Omega f v \,\mathrm{d}x + \int_{\Gamma_N} g v \,\mathrm{d}s\,.$$

There holds the following stability result.

Theorem 22. *Assume that $|\Gamma_D| > 0$ and that $b \cdot n \geq 0$ on Γ_N. There exists a unique solution of (3.4) and there holds the coercivity and continuity estimate*

$$a(u,v) + c(u,v) \leq \alpha_b \|\nabla u\|_0 \|\nabla v\|_0 \quad \textit{and} \quad a(u,u) + c(u,v) \geq \nu \|\nabla u\|_0^2,$$

where $\alpha_b = \nu + \|b\|_\infty c_F$ and c_F is the Friedrichs constant.

Proof. The continuity follows simply by using the Cauchy-Schwarz inequality and using that $b \in L^\infty$ and Friedrichs inequality Theorem to bound $\|v\|_0 \leq c_F \|\nabla v\|_0$. For the coercivity, integration by parts and $\mathrm{div}(b) = 0$ shows

$$\begin{aligned} c(u,v) &= \int_\Omega (b \cdot \nabla u) v \,\mathrm{d}x = \int_\Omega -(b \cdot \nabla v) u - (v \,\mathrm{div}(b)) u \,\mathrm{d}x + \int_{\Gamma_N} uvb \cdot n \,\mathrm{d}s \\ &= -c(v,u) + \int_{\Gamma_N} uvb \cdot n \,\mathrm{d}s, \end{aligned}$$

hence c is nearly skew symmetric. Using the assumption $b \cdot n \geq 0$ on Γ_N we then have

$$a(u,u) + c(u,u) = a(u,u) + \frac{1}{2} \int_{\Gamma_N} u^2 b \cdot n \,\mathrm{d}s \geq \nu \|\nabla u\|_0^2,$$

We conclude with the application of the Lax-Milgram theorem. □

Remark 15. In the above prove of the coercivity the boundary term on Γ_N is quadratic and has a positive sign which allows an estimate from below. A similar observation can be made for the skew symmetric trilinear form as discussed in remark 14. This allows to derive similar stability estimates for approximations of the Navier-Stokes or Oseen equations if the velocity (or wind) points in the proper direction ($u \cdot n \geq 0$ on $\Gamma_N \rightarrow$ outflow boundary).

Now let $V_h \subset V$ be a standard conforming discrete finite element space, then we have the problem: Find $u_h \in V_h$such that

$$a(u_h, v_h) + c(u_h, v_h) = f(v_h) \quad \forall v_h \in V_h. \tag{3.5}$$

Since we consider a conforming discretization, existence and uniqueness is inherited from

the continuous case. Applying Céa's lemma then gives the best approximation result

$$\|\nabla(u-u_h)\|_0 \leq (1+\mathcal{P}) \inf_{v_h \in V_h} \|\nabla(u-v_h)\|_0.$$

where we used that the Friedrichs constant scales like the size of the domain L and we defined the Péclet number by

$$\mathcal{P} := \frac{\|b\|_\infty L}{\nu}.$$

This shows, that best approximation with respect to the H^1-semi norm (or also the ν-weighted seminorm) might deteriorate when the Péclet number increases. Note, that we can still directly bound the error (using Galerkin orthogonality) by

$$\begin{aligned}
\nu\|\nabla(u-u_h)\|_0^2 &\leq a(u-u_h, u-u_h) + c(u-u_h, u-u_h) \\
&= a(u-u_h, u-I_h u) + c(u-u_h, u-I_h u) \\
&\leq \nu\|\nabla(u-u_h)\|_0\|\nabla(u-I_h u)\|_0 + \|b\|_\infty\|\nabla(u-u_h)\|_0\|u-I_h u\|_0,
\end{aligned}$$

where I_h is a standard conforming interpolation operator into V_h. Dividing by the ν-scaled error and using the approximation properties of I_h in the L^2 norm (assuming enough regularity of the solution) we get

$$\|\nabla(u-u_h)\|_0 \leq \|\nabla(u-I_h u)\|_0 + \frac{\|b\|_\infty}{\nu}\|u-I_h u\|_0 \lesssim (1+\frac{\|b\|_\infty h}{\nu})h\|u\|_{H^2}.$$

Hence, we still get optimal convergence rates if the so called mesh-Péclet number $\mathcal{P}_h := \frac{\|b\|_\infty h}{\nu}$ is smaller then 1. Considering the example from the beginning this shows that the mesh size h has to be so small such that the boundary layer of size ν is resolved appropriately. Since a global refinement might result in a high number of unknowns a local mesh refinement would be appreciable. Nevertheless, since in general one is not aware of the location of sharp gradients this approach is not useful in practice. In the following we aim to introduce stabilizing techniques that can be used for a more general approach and is based on the introduction of some artificial diffusion.

3.2.1 A streamline upwind Petrov Galerkin (SUPG) formulation

Several different approaches can be found in the literature to motivate the SUPG or as it is sometimes also called streamline diffusion method. As the latter name states, the main idea here is to add some diffusion in the direction of the stream lines of corresponding to

the wind b.

The first idea was presented by Brezzi and follows the approach of augmenting the lowest order linear finite element space such that the solution in the interior of elements is resolved more accurately. The resulting finite element space equals the scalar version of the velocity space of the MINI finite element method introduced for the approximation of the Stokes equations. Following exactly the same ideas as discussed in section 2.5.4 one can eliminate the local bubbles to define a stabilized method given by: Find $u_h \in V_h := \mathbb{P}^1(\mathcal{T}_h) \cap V$ such that

$$a(u_h, v_h) + c(u_h, v_h) + d(u_h, v_h) = f(v_h) + \sum_{T \in \mathcal{T}_h} \alpha \int_T f(b \cdot \nabla v_h) \quad \forall v_h \in V_h,$$

where

$$d(u_h, v_h) = \sum_{T \in \mathcal{T}_h} \alpha \int_T (b \cdot \nabla u_h)(b \cdot \nabla v_h) \, \mathrm{d}x,$$

and α is a stabilization parameter that needs to be chosen appropriately. An extensive study on this can be found in the literature and one may choose it on each element as

$$\alpha|_T = \begin{cases} \frac{h}{\|b\|_{\infty,T}} & \mathcal{P}_h \geq 1 \\ 0 & \text{else,} \end{cases}$$

where $\|b\|_{\infty,T}$ is the L^∞-norm on T. Above bilinear form d reads as a diffusion in the direction of b and hence motivates the name streamline diffusion. To generalize this method to high order cases we discuss the more traditional derivation in terms of a Petrov-Galerkin formulation. To this end consider a differential operator $\mathcal{L}$ and the problem statement $\mathcal{L}u = f$. We aim to find a solution in the trial space V such that (with an appropriate inner product) there holds

$$(\mathcal{L}u, v) = f(v) \quad v \in \tilde{V},$$

where $\tilde{V}$ is some (different!) test space. For the SUPG method we now use V_h as trial space and set

$$\tilde{V}_h := \{v_h + \alpha b \cdot \nabla v_h : v_h \in V_h\}.$$

The final method then reads as: Find $u_h \in V_h$ such that

$$a(u_h, v_h) + c(u_h, v_h) - \sum_{T \in \mathcal{T}_h} \alpha \int_T r_h(u_h)(b \cdot \nabla v_h) \, \mathrm{d}x = f(v_h) \quad \forall v_h \in V_h,$$

with the discrete residual defined on each element separately by

$$r_h(u_h) := f + \nu \Delta u_h - b \nabla u_h.$$

In the above derivation we replaced the integral on the domain Ω by the sum over all integrals on the elements $T \in \mathcal{T}_h$ because $\mathcal{L}$ includes the second order differential operator which is not well defined for functions in V_h. In the case of a linear approximation the second order operator vanishes resulting in above formulation. For the high order case we need to include the diffusive part from the residual such that the resulting method is still consistent.

For the analysis we now choose the norm

$$\|u_h\|_{SD}^2 := \nu \|\nabla u_h\|_0^2 + \|\alpha b \nabla u_h\|^2,$$

which naturally includes a scaling with respect to the Péclet number such that dominant diffusive or convective areas are measured appropriately. By defining the bilinear form

$$a_{SD}(u_h, v_h) := a(u_h, v_h) + c(u_h, v_h) + \sum_{T \in \mathcal{T}_h} \alpha \int_T (-\nu \Delta u_h + b \nabla u_h)(b \cdot \nabla v_h) \, \mathrm{d}x$$

we have the following stability result.

Lemma 20. *The bilinear form a_{SD} is elliptic with constant $c_{SD} = \mathcal{O}(1)$, i.e. there holds*

$$a_{SD}(u_h, u_h) \geq c_{SD} \|u_h\|_{SD}^2.$$

Proof. Follows with above definition of α, Young's inequality, and a scaling argument and the definition of the mesh Péclet number. □

The crucial point of this stability result is that the coercivity constant does not degrade in the limiting case $\nu \to 0$ and thus the method is also stable in the convection dominant case. Note however, that for the high order case the resulting method is not symmetric. Further note that in the instationary case the (local) residual also includes the time derivative (which makes the method not as practicable).

3.2.2 A Galerkin least-square stabilization

The least-square ansatz follows a very similar approach as the SUPG method, however in contrast to a Petrov-Galerkin approach one aims to stabilize the (original) Galerkin method my means of a local element by element weighted least squares approach. The resulting method then is simply given by: Find $u_h \in V_h$ such that

$$a(u_h, v_h) + c(u_h, v_h) + \sum_{T \in \mathcal{T}_h} \alpha \int_T r_h(u_h)(\nu \Delta v_h - b \cdot \nabla v_h) \, \mathrm{d}x = f(v_h) \quad \forall v_h \in V_h.$$

From a practical point of view there is no big advantage of the least squares method compared to the SUPG method. Note however, that the additional term $\nu \Delta v_h$ in the stabilizing bilinear form results in a symmetric formulation. For the stability analysis we choose the same norm as before to proof coercivity on the discrete level with a constant that is again robust for high Péclet numbers.

3.2.3 A discontinuous Galerkin method with upwinding

Although the least squares and the SUPG method have found a lot of attention in the literature (also due to the historical development) their main disadvantage is the rather difficult choice of the stabilization parameter which gets in particular more tricky in the case of the Navier-Stokes setting since then the wind equals the (maybe instationary) velocity. Further, the continuous finite element setting only allows to consider a local element wise stabilization neglecting any dominant transportation across interfaces.

A very elegant way of stabilization can be established if we consider a discontinuous approach. Note that DG methods actually have their origin in the work [40] where the authors considered a hyperbolic equation rather than an elliptic problem as discussed in section 2.6. To understand the stabilization technique in detail we first only consider the pure transport equation. To this end we assume that $\Gamma_D = \Gamma_{in}$, then we have the problem: Find $u \in V$ such that

$$\int_\Omega b \cdot \nabla u v \, \mathrm{d}x = \int_\Omega f v_h \, \mathrm{d}x \quad \forall v \in V. \tag{3.6}$$

For the derivation of the DG method let $v \in H^1(\mathcal{T}_h, \mathbb{R})$ be an element-wise smooth function, then we can apply locally integration by parts to get

$$\sum_{T \in \mathcal{T}_h} - \int_T u b \cdot \nabla v \, \mathrm{d}x + \int_{\partial T} b_n u v ds = \int_\Omega f v \, \mathrm{d}s \quad \forall v \in H^1(\mathcal{T}_h),$$

where we used that $\mathrm{div}(b) = 0$ and the abbreviation $b_n := b \cdot n$. Since the exact solution is continuous across element interfaces we can choose the trace of u on each facet F as the corresponding trace of one of the two adjacent elements. Whereas this choice equals for the exact solution, it might be different if we consider a discontinuous trial space later for the finite element method. We now aim to follow a similar approach as in the previous sections, hence incorporate the direction of the wind into our method. For this we define on each facet the so called upwind value by

$$u^{up}(x) := \lim_{\xi \to 0^+} u(x - \xi b) \quad \forall x \in F.$$

Now let $T \in \mathcal{T}_h$ be arbitrary with the normal vector $n = n_1$ and denote by T' all the neighbouring elements. The upwind value on ∂T equals the choice

$$u^{up} = \begin{cases} u|_T & \text{for } b_n > 0 \quad \text{outflow boundary} \\ u|_{T'} & \text{for } b_n \leq 0 \quad \text{inflow boundary} \end{cases}.$$

The upwind value is defined such that the approximate (discontinuous!) solution on interfaces is transport in the direction of b. In the case where $F \subset \partial T$ lies on the inflow boundary Γ_{in} we use the same idea and replace the upwind value by the Dirichlet value u_D. By this can rewrite above formulation as

$$\sum_{T \in \mathcal{T}_h} -\int_T ub \cdot \nabla v \,\mathrm{d}x + \int_{\partial T \setminus \Gamma_{in}} b_n u^{up} v ds = \int_\Omega fv \,\mathrm{d}s - \sum_{F \in \mathcal{F}_h \cap \Gamma_{in}} \int_F b_n u_D v \,\mathrm{d}s \tag{3.7}$$

Now let $V_h := \mathcal{P}^k(\mathcal{T}_h, \mathbb{R})$, then the DG method reads as: Find $u_h \in V_h$ such that

$$c^{DG}(u_h, v_h) = f^{DG}(v_h) \quad \forall v_h, \tag{3.8}$$

where we reformulated above equation to define the bilinear and linear form

$$\begin{aligned} c^{DG}(u_h, v_h) &:= \sum_{T \in \mathcal{T}_h} -\int_T u_h b \cdot \nabla v_h \,\mathrm{d}x + \int_{\partial T_{out}} b_n u_h [\![v_h]\!]^* ds \\ f^{DG}(v_h) &:= \int_\Omega f v_h \,\mathrm{d}x - \sum_{F \in \mathcal{F}_h \cap \Gamma_{in}} \int_F b_n u_D v_h \,\mathrm{d}s\,. \end{aligned} \tag{3.9}$$

In above definition we used the splitting

$$\partial T = \partial T_{in} \cup \partial T_{out} \quad \text{with} \quad \partial T_{in} := \{x \in \partial T : b_n \leq 0\}, T_{out} := \partial T \setminus \partial T_{in},$$

and that for $x \in \partial T_{out}$ we have

$$b_n u_h^{up} v_h|_T + b_{n'} u_h^{up} v_h|_{T'} = b_n u_h^{up} v_h|_T - b_n u_h^{up} v_h|_{T'} = b_n u_h^{up} [\![v_h]\!]^* = b_n u_h|_T [\![v_h]\!]^*,$$

where T' with normal vector n' is again a neighbouring element of an arbitrary $T \in \mathcal{T}_h$. Note that we can reformulate above bilinear form in various way. For this we first use again integration by parts for (3.7) to get

$$\begin{aligned} c^{DG}(u_h, v_h) &= \sum_{T \in \mathcal{T}_h} \int_T b \cdot \nabla u_h v_h \, \mathrm{d}x - \int_{\partial T} b_n uv \, \mathrm{d}s + \int_{\partial T \setminus \Gamma_{in}} b_n u_h^{up} v_h ds \\ &= \sum_{T \in \mathcal{T}_h} \int_T b \cdot \nabla u_h v_h \, \mathrm{d}x - \int_{\partial T_{in}} b_n [\![u_h]\!]^* v_h ds \end{aligned} \tag{3.10}$$

Lemma 21. *The upwind formulation* (3.8) *is consistent. Thus, let* $u \in H^1(\Omega)$ *be the exact solution of* (3.6)*, then*

$$c^{DG}(u, v_h) = (f, v_h) \quad \forall v_h \in V_h.$$

Proof. Follows by integration by parts and that $u^{up} = u$ for the exact (continuous) solution. □

Lemma 22. *There holds*

$$c^{DG}(u_h, u_h) = \frac{1}{2} |u_h|^2_{DG,2} := \frac{1}{2} \sum_{F \in \mathcal{F}_h} \int_F |b_n| ([\![u_h]\!]^*)^2 \, \mathrm{d}s \, .$$

Proof. We aim to combine formulations (3.9) and (3.10) similarly as in the definition of the skew symmetric convection bilinear form for the Navier-Stokes equations. For this let $F = T_1 \cap T_2$ be an arbitrary internal facet and assume that $b \cdot n_1 > 0$ thus $F \in (\partial T_1)_{out}$. With the notation $u_{hi} = (u_h)|_{T_i}$ and $v_{hi} = (v_h)|_{T_i}$ for $i \in 1, 2$ we get from (3.9) the contribution $b_{n_1} u_{h1}(v_{h1} - v_{h2})$. Similarly we have from (3.10) the contribution $b_{n_2}(u_{h1} - u_{h2}) v_{h2}$. Since $0 < b \cdot n_1 = -b \cdot n_2$ we get for the average and $u_h = v_h$

$$\frac{1}{2} b_{n_1}(u_{h1}(v_{h1} - v_{h2}) - (u_{h1} - u_{h2}) v_{h2}) = \frac{1}{2} |b_{n_1}| [\![u_h]\!]^* [\![u_h]\!]^*.$$

Using this relation on each internal facet, the average of (3.9) and (3.10) gives

$$\begin{aligned} c^{DG}(u_h, u_h) =& \frac{1}{2} \sum_{T \in \mathcal{T}_h} \int_T (b \cdot \nabla u_h u_h - u_h b \cdot \nabla u_h) \, \mathrm{d}x \\ &+ \frac{1}{2} \sum_{F \in \mathcal{F}_h^{\text{int}}} \int_F |b_n| ([\![u_h]\!]^*)^2 \, \mathrm{d}s + \frac{1}{2} \int_{\partial\Omega} |b_n| u_h^2 \, \mathrm{d}s \,. \end{aligned} \tag{3.11}$$

For the last integral we used that ob Γ_{in} we have $b \cdot n \leq 0$ and on Γ_{out} we have $b \cdot n > 0$ and thus

$$\begin{aligned} &\frac{1}{2} \sum_{T \in \mathcal{T}_h} \left(\int_{\partial T_{out} \cap \Gamma_{out}} b_n u_h [\![u_h]\!]^* ds - \int_{\partial T_{in} \cap \Gamma_{in}} b_n [\![u_h]\!]^* u_h ds \right) \\ =& \frac{1}{2} \left(\int_{\Gamma_{out}} b_n u_h u_h ds - \int_{\Gamma_{in}} b_n u_h u_h ds \right) = \frac{1}{2} \int_{\partial\Omega} |b_n| u_h u_h \, \mathrm{d}s \,. \end{aligned}$$

□

Above lemma shows, that in contrast to the SUPG stabilization, the upwinding does not lead to a coercive bilinear form because the $|u_h|_{DG,1}$ is only a semi norm on V_h. To this end we define the following norm

$$\|u_h\|_{DG}^2 := |u_h|_{DG,1}^2 + |u_h|_{DG,2}^2,$$

with

$$|u_h|_{DG,1}^2 := \sum_{T \in \mathcal{T}_h} \frac{h}{|b|_{\infty,T}} \|b \cdot \nabla u_h\|_T^2.$$

Theorem 23. *The bilinear form c^{DG} is continuous with respect to $\|u_h\|_{DG}$, and there holds the discrete inf-sup stability*

$$\inf_{u_h \in V_h} \sup_{v_h \in V_h} \frac{c^{DG}(u_h, v_h)}{\|u_h\|_{DG} \|v_h\|_{DG}} \geq \beta_{DG},$$

with a constant $\beta_{DG} > 0$ that only depends on the shape of the elements and the polynomial order k.

Proof. For simplicity we assume that b is piece-wise constant. A more general case can be found in the literature. Now let $u_h \in V_h$ be fixed. We aim to find a v_h such that $\|v_h\|_{DG} \lesssim \|u_h\|_{DG}$ and $c_{DG}(u_h, v_h) \gtrsim \|u_h\|_{DG}^2$. The main idea follows similar ideas as in the proof of stabilized methods for the Stokes problem, i.e. we split the test function into

to parts $v_h := \alpha u_h + v_h^2$, where $v_h^2 := \frac{h}{|b|_{\infty,T}} b \cdot \nabla u_h$. Here it is crucial that b is piece wise constant such that v_h^2 is still an element of V_h. This gives

$$c^{DG}(u_h, v_h) = \alpha \frac{1}{2} |u_h|_{DG,2}^2 + c^{DG}(u_h, v_h^2).$$

We continue to estimate the second term. Using representation (3.10) we get by the Cauchy-Schwarz, the Young inequality (Theorem 3)

$$\begin{aligned} c^{DG}(u_h, v_h^2) &= \sum_{T \in \mathcal{T}_h} \frac{h}{|b|_{\infty,T}} \|b \cdot \nabla u_h\|_T^2 - \int_{\partial T_{in}} b_n [\![u_h]\!]^* \frac{h}{|b|_{\infty,T}} b \cdot \nabla u_h \, \mathrm{d}s \\ &\geq \sum_{T \in \mathcal{T}_h} \frac{h}{|b|_{\infty,T}} \|b \cdot \nabla u_h\|_T^2 - |b_n| \frac{\varepsilon}{2} \|[\![u_h]\!]^*\|_{\partial T_{in}}^2 - \frac{h^2 |b_n|}{2\varepsilon |b|_{\infty,T}^2} \|b \cdot \nabla u_h\|_{\partial T_{in}}^2 \\ &\geq \sum_{T \in \mathcal{T}_h} \frac{h}{|b|_{\infty,T}} \|b \cdot \nabla u_h\|_T^2 - |b_n| \frac{\varepsilon}{2} \|[\![u_h]\!]^*\|_{\partial T_{in}}^2 - \frac{h}{2|b|_{\infty,T}} \|b \cdot \nabla u_h\|_T^2 \end{aligned}$$

where in the last step we used the inverse inequality for polynomials ($b \cdot \nabla u \in V_h$) (Theorem 1) with constant c_{inv} and set $\varepsilon = c_{inv}$ and that $|b_n| \leq |b|_{\infty,T}$. Now since

$$- \sum_{T \in \mathcal{T}_h} |b_n| \frac{\varepsilon}{2} \|[\![u_h]\!]^*\|_{\partial T_{in}}^2 \gtrsim -c_1 \frac{1}{2} \sum_F \int_F |b_n| ([\![u_h]\!]^*)^2,$$

we have in total

$$c^{DG}(u_h, v_h^2) \geq \frac{1}{2} |u_h|_{DG,1}^2 - c_1 |u_h|_{DG,2}^2.$$

Now let $\alpha = (2c_1 + 1)$ then we have

$$\begin{aligned} c^{DG}(u_h, v_h) &= \alpha \frac{1}{2} |u_h|_{DG,2}^2 + c^{DG}(u_h, v_h^2) \\ &\geq (2c_1 + 1) \frac{1}{2} |u_h|_{DG,2}^2 + \frac{1}{2} |u_h|_{DG,1}^2 - c_1 |u_h|_{DG,2}^2 \\ &\geq (2c_1 + 1) \frac{1}{2} |u_h|_{DG,2}^2 + \frac{1}{2} |u_h|_{DG,1}^2 - c_1 |u_h|_{DG,2}^2 \geq \frac{1}{2} \|u_h\|_{DG,2}^2. \end{aligned}$$

Again by the inverse inequality for polynomials and scaling arguments we further have

$$\begin{aligned}\|v_h^2\|_{DG}^2 &= \sum_{T\in\mathcal{T}_h}\frac{h}{|b|_{\infty,T}}\|b\cdot\nabla v_h^2\|_T^2+\sum_{F\in\mathcal{F}_h}|b_n|\|[\![v_h^2]\!]^*\|_F^2\\ &= \sum_{T\in\mathcal{T}_h}\frac{h}{|b|_{\infty,T}}\|b\cdot\nabla(\frac{h}{|b|_{\infty,T}}b\cdot\nabla u_h)\|_T^2+\sum_{F\in\mathcal{F}_h}|b_n|\|[\![\frac{h}{|b|_{\infty,T}}b\cdot\nabla u_h]\!]^*\|_F^2\\ &\lesssim \sum_{T\in\mathcal{T}_h}\frac{h}{|b|_{\infty,T}}\frac{|b|_{\infty,T}^2}{h^2}\|\frac{h}{|b|_{\infty,T}}b\cdot\nabla u_h\|_T^2+\sum_{T\in\mathcal{T}_h}|b_n|\frac{h^2}{|b|_{\infty,T}^2}\frac{1}{h}\|b\cdot\nabla u_h\|_T^2\lesssim\|u_h\|_{DG}^2,\end{aligned}$$

thus in total $\|v_h\|_{DG}^2\lesssim\|u_h\|_{DG}^2$. □

In order to prove that the bilinear form c^{DG} is continuous we introduce a second (stronger) norm by

$$\|u_h\|_{DG,*}^2:=\|u_h\|_{DG}^2+\sum_{T\in\mathcal{T}_h}\int_T\frac{|b|_{\infty,T}}{h}u_h^2\,\mathrm{d}x+\int_{\partial T}|b_n|u_h^2\,\mathrm{d}s\,.$$

Similarly as in the previous section we then have the continuity result not only on the discrete level but also in the continuous setting.

Lemma 23. *There holds*

$$c^{DG}(u,v)\lesssim\|u\|_{DG,*}\|v\|_{DG,*}\quad\forall u,v\in V_h+H^1(\Omega)\cap H^2(\mathcal{T}_h).$$

Proof. Follows with several applications of the Cuachy-Schwarz inequality. □

A very important feature of the DG method is that there holds a local discrete conservation property. To this end let $T\in\mathcal{T}_h$ be such that $\partial T\cap\Gamma=\emptyset$, and choose the characteristic test function $v_h=1$ on T and 0 on $\Omega\setminus T$. Then (3.8) reads as

$$\int_{\partial T}b_n u^{up}\,\mathrm{d}s=\int_T f\,\mathrm{d}x\,.$$

Hence, quantities that "enter" and "leave" the element T through the boundary ∂T are solely balanced by the local source $f|_T$.

3.2.4 A hybrid discontinuous Galerkin method for convection-diffusion problems

In the previous section we focused on the introduction of the upwinding technique for a pure hyperbolic convection problem. In the case of a discontinuous approximation of the

convection diffusion problem (3.4) we want to utilize the advantages of the hybrid approach introduced in section 2.6. To this end let $V_h := \mathbb{P}^k(\mathcal{T}_h, \mathbb{R})$ and $\hat{V}_h := \{\hat{v}_h \in \mathbb{P}^k(\mathcal{F}_h, \mathbb{R}) : \hat{v}_h = 0 \text{ on } \partial\Omega\}$ as in section 2.6.1. Further let a^{HDG} be the bilinear form as in (2.34), hence the weak formulation of the Laplacian in the HDG setting. Before we can combine the diffusive and the convective bilinear formulation from the previous section we have to reformulate in the setting of an HDG discretization. To this end we will first redefine the upwind value. Let $T \in \mathcal{T}_h$ be arbitrary with the normal vector $n = n_1$ and denote by T' a neighboring element and $F = T \cap T'$. Consider a given element wise (discontinuous) function $u_h \in V_h$ and a facet wise function $\hat{u}_h \in \hat{V}_h$. For a given wind b the upwind value on $F \subset \partial T$, hence seen from the direction of T, is then given by

$$u^{up} = \begin{cases} u_h|_T & \text{for } b_n > 0 \quad \text{outflow boundary} \\ \hat{u}_h|_F & \text{for } b_n \leq 0 \quad \text{inflow boundary} \end{cases}.$$

Following the same steps as before we then have (for the pure convection problem)

$$\sum_{T \in \mathcal{T}_h} -\int_T u_h b \cdot \nabla v_h \,\mathrm{d}x + \int_{\partial T} b_n u^{up} v_h ds = \int_\Omega f v \,\mathrm{d}s\,.$$

Considering and edge $F = T \cap T'$, we see that either the unknowns of u_h on T or on T' couple with the unknowns of $\hat{u}_h$ on F. However, in contrast to before, the volume unknowns do not couple at all. To fix this we add another stabilizing term on the outflow boundaries given by

$$\sum_{T \in \mathcal{T}_h} \int_{\partial T_{out}} b_n(\hat{u}_h - u_h)\hat{v}_h \,\mathrm{d}s\,. \tag{3.12}$$

Hence, in the case of an outflow boundary the values of $\hat{u}_h$ equal the values of u_h. This results in an "indirect" coupling of element unknowns via the facet variables. Next we define the bilinear form

$$\begin{aligned} &c^{HDG}((u_h, \hat{u}_h),(v_h, \hat{v}_h)) \\ &\qquad := \sum_{T \in \mathcal{T}_h} -\int_T u_h b \cdot \nabla v_h \,\mathrm{d}x + \int_{\partial T} b_n u^{up} v_h ds + \int_{\partial T_{out}} b_n(\hat{u}_h - u_h)\hat{v}_h \,\mathrm{d}s\,. \end{aligned}$$

Similarly as before, we can reformulate c^{HDG} as

$$\begin{aligned} &c^{HDG}((u_h,\hat{u}_h),(v_h,\hat{v}_h)) \\ &\qquad := \sum_{T\in\mathcal{T}_h} \int_T v_h b\cdot\nabla u_h \,\mathrm{d}x + \int_{\partial T_{in}} |b_n|(u-\hat{u})v_h ds + \int_{\partial T_{out}} |b_n|(\hat{u}_h-u_h)\hat{v}_h \,\mathrm{d}s, \\ &\qquad := \sum_{T\in\mathcal{T}_h} -\int_T u_h b\cdot\nabla v_h \,\mathrm{d}x + \int_{\partial T_{in}} b_n u^{up}(v_h-\hat{v}_h)\,\mathrm{d}s + \int_{\partial T_\Gamma} b_n\hat{u}_h\hat{v}_h \,\mathrm{d}s\,. \end{aligned}$$

Algebraically, the HDG bilinear form c^{HDG} results in the same solution as with the DG formulation. However we get the same nice advantages discussed in section 2.6.1 as element-wise assembly due to a decoupling of the element unknowns and that inner degrees of freedoms can be eliminated (static condensation). The new formulation now further lets us define a discrete method for the approximation of (3.4): Find $(u_h,\hat{u}_h)\in V_h\times\hat{V}_h$ such that

$$\nu a^{HDG}((u_h,\hat{u}_h),(v_h,\hat{v}_h)) + c^{HDG}((u_h,\hat{u}_h),(v_h,\hat{v}_h)) = (f,v_h) \quad \forall (v_h,\hat{v}_h)\in V_h\times\hat{V}_h.$$

The stability analysis follows similar techniques as introduced in this section and in section 2.6.1.

3.3 Finite element methods for the stationary Navier-Stokes equations

In this section we briefly discuss the solving algorithms of finite element methods of problem (3.1). We define the corresponding discrete problem: Find $(u_h,p_h)\in V_h\times Q_h$ such that

$$a(u_h,v_h)+c(u_h,u_h,v_h)+b(v_h,p_h) = (f,v_h) \quad \forall v_h\in V_h, \tag{3.13a}$$

$$b(u_h,q_h) = 0 \quad \forall q_h\in Q_h, \tag{3.13b}$$

Note that a crucial property in the proof of the uniqueness of the continuous stationary Navier Stokes equation is the skew symmetry of the convective bilinear form. For the gradient form $c=c_\nabla$ this is the case if the wind is exactly divergence, see Lemma 18. Particularly this holds true for the exact solution. However, since in general the solution of a finite element method is only weakly divergence free the bilinear form $c(u_h,\cdot,\cdot)$ might not be skew symmetric on the discrete level. To this end one often solves (3.13) by means of $c_{\mathrm{skw}}, c_{\mathrm{div}}$ or c_{curl}. For simplicity we fix now $c=c_{\mathrm{skw}}$, consider a conforming approximation

and that V_h and Q_h fulfills the Stokes inf-sup condition. We define

$$\mathcal{N}_{0,h} := \sup_{w,u,v \in V_{0,h}} \frac{c_{\text{skw}}(w_h, u_h, v_h)}{\|w_h\|_V \|u_h\|_V \|v_h\|_V},$$

Theorem 24. *Assume that there holds the estimate*

$$\frac{\mathcal{N}_{0,h}\|f\|_{V^*}}{\nu^2 c_k} < 1,$$

(where c_k is the Korn inequality) then there exists a unique solution $(u_h, p_h) \in V_h \times Q_h$ of problem (3.13) with

$$\|\nabla u_h\|_0 \leq \frac{1}{\nu}\|f\|_{V^*} \quad \textit{and} \quad \|p_h\|_0 \lesssim \|f\|_{V^*} + \frac{1}{\nu^2}\|f\|_{V^*}^2.$$

Proof. Follows with similar steps as in the continuous setting. A detailed proof is given in [25]. □

Note that similar results hold if problem (3.13) is enriched by certain stabilization bilinear forms in order to guarantee inf-sup solvability or sharp gradient for convection dominant flows. Further one can also consider a discontinuous approximation by exchanging the continuous bilinear forms with the corresponding forms defined in the previous sections.

3.3.1 Iterative schemes

We finish this chapter with the introduction of iteration schemes for solving the nonlinear problem (3.13). The most simple approach is given by a fixed point iteration that includes the solution of several Stokes problems. To this end let $\mathbb{S} : V' \times 0 \to V_h \times Q_h$ be the discrete Stokes operator that solves the discrete Stokes equation (2.13) with a given right hand side. Then the fixed point iteration is given by

$$u_h^{k+1} := \mathbb{S}(f(\cdot) - c(u_h^k, u_h^k, \cdot)).$$

Although this approach only requires to solve a Stokes problem in each iteration step, the convergence speed is very small if the viscosity is not sufficiently large. An alternative is given by the Newton method. To this end we write $u_h^{k+1} = u_h^k + \delta u_h$ and $p_h^{k+1} = p_h^k + \delta p_h$. We aim to find a linearized equation for the difference δu_h and δp_h. For this we first define

the nonlinear residual given by

$$r^k_{u,h}(v_h) := (f, v_h) - c(u^k_h, u^k_h, v_h) - a(u^k_h, v_h) - b(v_h, p^k_h),$$
$$r^k_{p,h}(q_h) := -b(u^k_h, q_h).$$

Assuming that (u^{k+1}_h, p^{k+1}_h) is the solution of (3.13), it is easy to see, that the corrections are fulfilling the equation

$$\begin{aligned} d(u^k_h, \delta u_h, v_h) + a(\delta u_h, v_h) + b(v_h, \delta p_h) &= r^k_{u,h}(v_h) \quad \forall v_h \in V_h, \\ b(\delta u_h, q_h) &= r^k_{p,h}(q_h) \quad \forall q_h \in Q_h, \end{aligned}$$

with the non linear difference

$$d(u^k_h, \delta u_h, v_h) := c(u^k_h, \delta u_h, v_h) + c(\delta u_h, u^k_h, v_h) + c(\delta u_h, \delta u_h, v_h).$$

If the corrections are small (i.e. we are “close” to the solution), we can linearize above equation by dropping the last term to get the symmetric linear problem: Find $\delta u_h, \delta p_h \in V_h \times Q_h$ such that

$$\begin{aligned} c(u^k_h, \delta u_h, v_h) + c(\delta u_h, u^k_h, v_h) + a(\delta u_h, v_h) + b(v_h, \delta p_h) &= r^k_{u,h}(v_h) \quad \forall v_h \in V_h, \\ b(\delta u_h, q_h) &= r^k_{p,h}(q_h) \quad \forall q_h \in Q_h. \end{aligned}$$

The Newton iteration calculates in each step the solution of the above problem and performs the corresponding update.

It is well known that the Newton method converges quadratically in the case where the current iterate is close to the fixed point. Although this seems to be very desirable, the convergence radius scales with the viscosity ν, hence convergence might not be guaranteed if the initial guess is not close enough. An alternative to the Newtons method is given by the so called Picard iteration. Beside dropping the quadratic term we also drop $c(\delta u_h, u_h, v_h)$ which results in the problem: Find $\delta u_h, \delta p_h \in V_h \times Q_h$ such that

$$\begin{aligned} c(u^k_h, \delta u_h, v_h) + a(\delta u_h, v_h) + b(v_h, \delta p_h) &= r^k_{u,h}(v_h) \quad \forall v_h \in V_h, \\ b(\delta u_h, q_h) &= r^k_{p,h}(q_h) \quad \forall q_h \in Q_h. \end{aligned}$$

In the case of $c = c_\nabla$ we see that the solution (u^{k+1}_h, p^{k+1}_h) of each step now solves the

problem

$$\begin{aligned} a(u_h^{k+1}, v_h) + c(u_h^k, u_h^{k+1}, v_h) + b(v_h, p_h^{k+1}) &= (f, v_h) \quad \forall v_h \in V_h, \\ b(u_h^{k+1}, q_h) &= 0 \qquad \forall q_h \in Q_h, \end{aligned}$$

which reads as an Oseen problem with the fixed convective wind u_h^k. We will use a similar approach in splitting methods when we consider the instationary Navier Stokes equations in the next section. The advantage of the Picard iteration is that, compared to the Newton method, it has a relatively large ball of convergence but has a smaller order of convergence.

4 The instationary Navier-Stokes equations

This chapter is dedicated to analyze and approximate the incompressible instationary Navier-Stokes equations (1.11). Including inflow Dirichlet boundary conditions u_{in} on Γ_{in} homogeneous Dirichlet boundary conditions on the walls Γ_w we might also consider homogeneous Neumann boundary conditions on the outflow boundary Γ_{out}. Let T be a fixed time, then we aim to find a solution u, p such that

$$\frac{\partial u}{\partial t} - \nu \operatorname{div}(\varepsilon(u)) + \operatorname{div}(u \otimes u) + \nabla p = f \qquad \text{in } \Omega \times (0,T], \tag{4.1}$$

$$\operatorname{div}(u) = 0, \qquad \text{in } \Omega \times (0,T], \tag{4.2}$$

$$u = u_{in}, \qquad \text{on } \Gamma_{in} \times (0,T] \tag{4.3}$$

$$u = 0, \qquad \text{on } \Gamma_w \times (0,T] \tag{4.4}$$

$$(-\nu\varepsilon(u) + p\mathrm{Id})n = 0, \qquad \text{on } \Gamma_{out} \times (0,T] \tag{4.5}$$

$$u = u_0 \qquad \text{on } \Omega \times 0. \tag{4.6}$$

4.1 Existence and uniqueness

PL: Will be updated

4.2 Method of lines and θ-schemes

A very traditional approach of solving the time-dependent Navier Stokes equations is the *method of lines*. Let $\mathcal{T}_h$ be a fixed triangulation of the spatial domain Ω. For the ease we consider an inf-sup stable finite element pair $V_h \times Q_h$ but we emphasize that stabilized methods can be used in a similar manner. The discrete spaces are chosen to fit the boundary conditions as in (4.1) where we assume that $\Gamma_{in} = \emptyset$. In the case of an inflow boundary condition (i.e. non-homogeneous Dirichlet boundary conditions) we use a standard homogenization process. In contrast to the stationary case we now assume that the coefficients of the finite element solutions are time dependent, i.e. we have the

semi-discrete approach

$$u_h(t,x) = \sum_{i \in N_u} u_i(t)\phi_i^u(x) \quad \text{and} \quad p_h(t,x) = \sum_{i \in N_p} p_i(t)\phi_i^p(x),$$

where ϕ_i^u and ϕ_i^p are the basis functions of the finite element spaces V_h and Q_h, respectively, with dimensions N_u, N_p. We derive a semi-discrete weak formulation of (4.1) as usual by multiplying with (time independent!) test functions and integrating by parts. The solution $(u_h, p_h) \in V_h \times Q_h$ must then satisfy for all $t \in (0,T]$

$$\begin{aligned}(\frac{\partial}{\partial t}u_h(t), v_h) + a(u_h(t), v_h) + c(u_h(t), u_h(t), v_h) + b(v_h, p_h(t)) = (f, v_h) \quad & \forall v_h \in V_h, \\ b(u_h(t), q_h) = 0 \quad & \forall q_h \in Q_h,\end{aligned}$$

and further $u_h(0) = u_0$. Next we introduce the matrices $M, A \in \mathbb{R}^{N_u \times N_u}$ and $B \in \mathbb{R}^{N_p \times N_u}$ by $M_{ij} := (\phi_i^u, \phi_j^u)$, $A_{ij} := a(\phi_i^u, \phi_j^u)$ and $B_{ij} = b(\phi_j^u, \phi_i^p)$. Further we define $F \in \mathbb{R}^{N_u}$ by $F_i := (f, \phi_i^u)$. Denoting by $\underline{u}(t) \in \mathbb{R}^{N_u}$ and $\underline{p}(t) \in \mathbb{R}^{N_p}$ with $\underline{u}(t)_i = u_i(t)$ and $\underline{p}(t)_i = p_i(t)$ the coefficient vectors of the finite element solutions we can reformulate equation (4.7) as

$$\begin{aligned}M\frac{d}{dt}\underline{u}(t) + A\underline{u}(t) + C(\underline{u}(t))\underline{u}(t) + B^T\underline{p}(t) = F, \\ B\underline{u}(t) = 0,\end{aligned}$$

where

$$C : \mathbb{R}^{N_u} \to \mathbb{R}^{N_u \times N_u}, \quad C(\underline{w}) := c(w_h, \phi_i^u, \phi_j^u) \quad \text{with} \quad w_h := \sum_{i \in N_u} \underline{w}_i \phi_i^u(x).$$

Above equation is a system of ordinary differential equations and can be solved by many different approaches.

Very frequently used schemes are so called one-step θ-schemes. To this end let τ be a fixed time step used for an equidistant mesh of the interval $[0,T]$ with N intervals. Let $t^n := \tau n$ with $0 \le n \le N$ and introduce the symbols $\underline{u}^n = \underline{u}(t^n)$ and $\underline{p}^n = \underline{p}(t^n)$. Further let $\theta \in [0,1]$ be fixed. We solve for each time step t^n the system

$$\begin{aligned}[M + \theta\tau[A + C(\underline{u}^{n+1})]]\underline{u}^{n+1} + \tau B^{\mathrm{T}}\underline{p}^{n+1} = [M - (1-\theta)\tau[A + C(\underline{u}^n)]]\underline{u}^n + \tau F \\ \tau B\underline{u}^{n+1} = 0.\end{aligned}$$

Here $\theta = 0$ gives the first order explicit Euler and $\theta = 1$ gives the first order A-stable implicit Euler method. The choice $\theta = 1/2$ results in the well known Crank-Nicolson method which

is of higher order but is not A-stable. A very popular method, which is no θ-scheme, is the backward difference formula of order two, called BDF2 scheme. Here, one solve for $n \geq 2$ the system

$$[\frac{3}{2}M + \tau[A + C(\underline{u}^{n+1})]]\underline{u}^{n+1} + \tau B^{\mathrm{T}}\underline{p}^{n+1} = [2M - \tau[A + C(\underline{u}^{n})]]\underline{u}^{n} - \frac{1}{2}M\underline{u}^{n-1} + \tau F$$
$$\tau B\underline{u}^{n+1} = 0,$$

which is a high order scheme in time and A-stable but one needs to store the additional vector $\underline{u}^{n-1}$. Another set of very popular methods are the so called fractional θ-schemes where additional intermediate steps at $t_n + \theta\tau$ and $t_{n+1} - \theta\tau$ are introduced. The three steps are given by

1. Step from $t^n \to t^{n+\theta}$:

$$[M + \alpha\theta\tau[A + C(\underline{u}^{n+\theta})]]\underline{u}^{n+\theta} + \theta\tau B^{\mathrm{T}}\underline{p}^{n+\theta} = [M - \beta\theta\tau[A + C(\underline{u}^{n})]]\underline{u}^{n} + \theta\tau F$$
$$\theta\tau B\underline{u}^{n+\theta} = 0.$$

2. Step from $t^{n+\theta} \to t^{n+1-\theta}$:

$$[M + \beta\theta'\tau[A + C(\underline{u}^{n+1-\theta})]]\underline{u}^{n+1-\theta} + \theta'\tau B^{\mathrm{T}}\underline{p}^{n+1-\theta}$$
$$= [M - \alpha\theta'\tau[A + C(\underline{u}^{n+\theta})]]\underline{u}^{n+\theta} + \theta'\tau F$$
$$\theta'\tau B\underline{u}^{n+1-\theta} = 0.$$

3. Step from $t^{n+1-\theta} \to t^{n+1}$:

$$[M + \alpha\theta\tau[A + C(\underline{u}^{n+1})]]\underline{u}^{n+1} + \theta\tau B^{\mathrm{T}}\underline{p}^{n+1}$$
$$= [M - \beta\theta\tau[A + C(\underline{u}^{n+1-\theta})]]\underline{u}^{n+1-\theta} + \theta\tau F$$
$$\theta\tau B\underline{u}^{n+1} = 0.$$

To retrieve a second order and A-stable method one chooses $\theta = 1 - \sqrt{2}/2$, $\theta' = 1 - 2\theta$, $\alpha \in (1/2, 1]$ and $\beta = 1 - \alpha$. Note that the choice $\alpha = \theta'/(1 - \theta)$ then further results in $\alpha\theta = \beta\theta'$ which helps in building the system matrices.

4.2.1 Splitting and projection schemes

Although above methods have very nice smoothing and convergence properties, the main two main difficulties given by the incompressibility constraint (resulting in a saddle point

problem) and the non-linearity due to the convection (demanding for an iterative method if treated implicitly) are still included in all intermediate steps. To solve this issue we introduce the splitting fractional θ-schemes by

1. Step from $t^n \to t^{n+\theta}$:

$$\begin{aligned}[M + \alpha\theta\tau A]]\underline{u}^{n+\theta} + \theta\tau B^{\mathrm{T}}\underline{p}^{n+\theta} &= [M - \beta\theta\tau A]\underline{u}^n - \theta\tau C(\underline{u}^n)\underline{u}^n + \theta\tau F \\ \theta\tau B\underline{u}^{n+\theta} &= 0.\end{aligned}$$

2. Step from $t^{n+\theta} \to t^{n+1-\theta}$:

$$\begin{aligned}&[M + \beta\theta'\tau[A+C(\underline{u}^{n+1-\theta})]]\underline{u}^{n+1-\theta} \\ &\qquad = [M - \alpha\theta'\tau[A + C(\underline{u}^{n+\theta})]]\underline{u}^{n+\theta} - \theta'\tau B^{\mathrm{T}}\underline{p}^{n+\theta} + \theta'\tau F\end{aligned}$$

3. Step from $t^{n+1-\theta} \to t^{n+1}$:

$$\begin{aligned}[M + \alpha\theta\tau A]\underline{u}^{n+1} + \theta\tau B^{\mathrm{T}}\underline{p}^{n+1} &\\ = [M - \beta\theta\tau A]\underline{u}^{n+1-\theta} - \tau C(\underline{u}^{n+1-\theta})\underline{u}^{n+1-\theta} + \theta\tau F &\\ \tau B\underline{u}^{n+1} = 0.&\end{aligned}$$

Note, that the first and the third step include solving a linear Stokes problem with an explicit convection in the right hand side, and the second step includes solving a non-linear convection diffusion equation without any incompressibility constraint. A simplified first order operator splitting scheme is given by the so called IMEX (implicit explicit Euler) where we solve

$$\begin{aligned}[M + \tau A]\underline{u}^{n+1} + \tau B^{\mathrm{T}}\underline{p}^{n+1} &= M\underline{u}^n - \tau C(\underline{u}^n)\underline{u}^n + \theta\tau F \\ \theta\tau B\underline{u}^{n+1} &= 0,\end{aligned} \tag{4.8}$$

hence we treat the incompressibility implicitly and the convection explicitly. This method can also be extended to high-order schemes resulting in so called diagonally implicit Runge-Kutta methods.

Remark 16. In section 3.2.4 we introduced how the upwind stabilization can be extended to the HDG setting. If one considers to use a splitting method for an HDG approximation one has to be careful if the convection is treated explicitly. After the implicit solve of (for example) (4.8) the trace variable $\hat{\underline{u}}^{n+1}$ on outflow boundaries ∂T_{out} does not equal the value of the corresponding element trace as it would be forced by the gluing term

introduced in equation (3.12), hence an application of the convection formulated in the HDG setting would not result in an upwind stabilization. Instead one simply considers a DG version of the convection and uses it as a driving force only seen by element variables.

Although the explicit treatment of the convection simplifies the solving routine tremendously one still has to solve a saddle point problem with the structure

$$\begin{pmatrix} M + \tau A & B^{\mathrm{T}} \\ B & 0 \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{p} \end{pmatrix} = \begin{pmatrix} G \\ 0 \end{pmatrix},$$

with some right hand side G including volume forces and the explicit convection terms. Since a direct solver is limited by the size of the problem, several different approaches using for example an iterative scheme with (for example) block-diagonal preconditioner. The main idea of this approach is to decouple the incompressibility constraint from the momentum equation and can be found in the literature under the terms “quasi-compressibility method”, “projection method”, “SIMPLE method” and more. We only discuss the very simple projection scheme proposed by Chorin. For simplicity we only consider the case of homogeneoues Dirichlet boundary conditions $\Gamma_w = \partial\Omega$. The projection then reads as

1. Perform an explicit (or implicit) nonlinear step for the pure convection diffusion step (also called a Burger’s step) to get an intermediate velocity $\tilde{\underline{u}}^{n+1}$

$$[M + \tau A]\tilde{\underline{u}}^{n+1} = M\underline{u}^n - \tau C(\underline{u}^n)\underline{u}^n + \tau F.$$

2.) Perform a L^2-projection of $\tilde{\underline{u}}^{n+1}$ into the manifold of divergence free velocities.

The projection scheme can be interpreted in various different ways. The most common is to perform the projection by solving a pressure Poisson problem, i.e. we solve the problem

$$\Delta\tilde{p}^{n+1} = \mathrm{div}(\tilde{u}^{n+1})$$

with homogeneous Neumann boundary conditions $\partial_n\tilde{p}^{n+1} = 0$. Then the projection is given by $u^{n+1} = \tilde{u}^{n+1} - \nabla\tilde{p}^{n+1}$. This immediately shows that $\mathrm{div}(u^{n+1}) = 0$. The unnatural boundary condition in above Poisson problem have caused a lot of discussion in the literature since it might result in oscillations in the pressure field close to the boundary. Further note that the finite element spaces have to be chosen appropriately. A very good overview of projection schemes can be found in [39].

Projection for the $H(\mathrm{div})$-conforming HDG method

In the following we show how the projection scheme can be applied to the $H(\mathrm{div})$-conforming HDG method introduced in section 2.6.2. To this end let

$$\begin{aligned} V_h :=& \mathrm{BDM}^k(\mathcal{T}_h, \Omega) \cap H_0(\mathrm{div}, \Omega) \\ =& \{v_h \in \mathbb{P}^k(\mathcal{T}_h, \mathbb{R}^d) : [\![v_h \cdot n]\!] = 0 \text{ on all } F \in \mathcal{F}_h, v_h \cdot n = 0 \text{ on } \partial\Omega \}, \\ \hat{V}_h :=& \{\hat{v}_h \in \mathbb{P}^k(\mathcal{F}_h, \mathbb{R}^d) : \hat{v}_h|_F \cdot n = 0 \; \forall F \in \mathcal{F}_h, \hat{v}_h = 0 \text{ on } \partial\Omega\}, \\ Q_h :=& \mathbb{P}^{k-1}(\mathcal{T}_h, \mathbb{R}) \cap Q. \end{aligned}$$

Assume that $(\tilde{u}_h, \hat{\tilde{u}}_h)$ is the solution of the pure convection step and that $\mathrm{div}\, \tilde{u}_h \neq 0$. Instead of solving a Poisson problem on the pressure space we reformulate it in a mixed setting, i.e. we have the problem: Find $(\delta u_h, \tilde{p}_h) \in V_h \times Q_h$ such that

$$\begin{aligned} \int_\Omega \delta u_h \cdot v_h + \int_\Omega \mathrm{div}\, v_h \tilde{p}_h = 0 \qquad & \forall v_h \in V_h \\ \int_\Omega \mathrm{div}\, \delta u_h q_h = \int_\Omega \mathrm{div}(\tilde{u}_h) q_h \qquad & \forall \tilde{q}_h \in Q_h. \end{aligned}$$

Note that we use the same spaces for the projection as used in the HDG method of the Navier-Stokes discretization. Further note that since $\mathrm{div}(V_h) = Q_h$ the solution of the above projection gives $\mathrm{div}\, \delta u_h = \mathrm{div}\, \tilde{u}_h$ in an exact manner, and hence $u_h = \tilde{u}_h - \delta u_h$ is exactly divergence-free.

Remark 17. If one considers to solve a big problem then the projection needs to be solved with an iterative method. In contrast to a Poisson problem the mixed formulation results in a saddle point problem which would demand to use a GMRES or MINRES solver including an $H(\mathrm{div})$ precondtioner. To this end one uses a hybridization of the normal-continuity of δu_h. After a static condensation the resulting system (for the facet Lagrange multiplier) is SPD and elliptic with respect to an H^1-like HDG norm, hence (more) standard preconditioners can be used.

4.2.2 Error analysis

PL: Will be updated

Bibliography

[1] R. Adams. *Sobolev Spaces*. Academic Press, 1970.

[2] Anderson, J. *Computational Fluid Dynamics*. Computational Fluid Dynamics: The Basics with Applications. McGraw-Hill Education, 1995.

[3] D. N. Arnold et al. “Unified analysis of discontinuous Galerkin methods for elliptic problems”. In: *SIAM journal on numerical analysis* 39.5 (2002), pp. 1749–1779.

[4] Babuška, I. and Aziz, A. K. “Survey lectures on the mathematical foundations of the finite element method”. In: *The mathematical foundations of the finite element method with Applications to Partial Differential Equations*. Ed. by A. K. Aziz. New Zork and London: Academic Press, 1972, pp. 3–359.

[5] Batchelor, G.K. *An Introduction to Fluid Dynamics*. Cambridge Mathematical Library. Cambridge University Press, 2000.

[6] C. Bernardi and Y. Maday. *Approximations spectrales de problèmes aux limites elliptiques*. Mathématiques et Applications. Springer Berlin Heidelberg, 1992.

[7] C. Bernardi and Y. Maday. “Spectral methods”. In: *Handbook of numerical analysis, Vol. V*. Ed. by P. G. Ciarlet and J. L. Lions. North-Holland, 1997, pp. 209–485.

[8] D. Boffi, F. Brezzi, and M. Fortin. *Mixed Finite Element Methods and Applications*. Springer Series in Computational Mathematics. Springer Berlin Heidelberg, 2013.

[9] J. H. Bramble. “A proof of the inf-sup condition for the Stokes equations on Lipschitz domains”. In: *Math. Models Methods Appl. Sci.* 13.3 (2003). Dedicated to Jim Douglas, Jr. on the occasion of his 75th birthday, pp. 361–371.

[10] C. Brennecke et al. “Optimal and pressure-independent L^2 velocity error estimates for a modified Crouzeix-Raviart Stokes element with BDM reconstructions”. In: *J. Comput. Math.* 33.2 (2015), pp. 191–208.

[11] G. Duvaut and J. Lions. *Inequalities in mechanics and physics*. Grundlehren der mathematischen Wissenschaften. Springer-Verlag, 1976.

[12] G. Duvaut and J. Lions. *Les inéquations en mécanique et en physique*. Travaux et Recherches Mathématiques. Dunod, 1972.

[13] A. Ern and J. L. Guermond. *Theory and Practice of Finite Elements*. 1st ed. Applied Mathematical Sciences 159. Springer-Verlag New York, 2004.

[14] L. C. Evans. *Partial differential equations*. Providence, R.I.: American Mathematical Society, 2010.

[15] N. R. Gauger, A. Linke, and P. W. Schroeder. "On high-order pressure-robust space discretisations, their advantages for incompressible high Reynolds number generalised Beltrami flows and beyond". en. In: *The SMAI journal of computational mathematics* 5 (2019), pp. 89–129.

[16] V. Girault and P.-A. Raviart. *Finite element methods for Navier-Stokes equations: theory and algorithms*. Vol. 5. Springer Science & Business Media, 2012.

[17] J. Gopalakrishnan, P. L. Lederer, and J. Schöberl. "A mass conserving mixed stress formulation for the Stokes equations". In: *IMA J. Numer. Anal.* 40.3 (2020), pp. 1838–1874.

[18] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1985.

[19] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1985.

[20] P. Grisvard. *Singularities in Boundary Value Problems*. Recherches en mathématiques appliquées. Masson, 1992.

[21] S. Gross and A. Reusken. *Numerical Methods for Two-phase Incompressible Flows*. Springer Series in Computational Mathematics. Springer Berlin Heidelberg, 2011.

[22] M. Gunzburger and R. Nicolaides. *Incompressible Computational Fluid Dynamics: Trends and Advances*. Cambridge University Press, 1993.

[23] R. Hiptmair, J. Li, and J. Zou. "Universal extension for sobolev spaces of differential forms and applications". In: *J. Func. Anal.* 263 (2012), pp. 364–382.

[24] V. John et al. "On the divergence constraint in mixed finite element methods for incompressible flows". In: *SIAM Review* 59 (2017), pp. 492–544.

[25] V. John. *Finite element methods for incompressible flow problems*. Vol. 51. Springer Series in Computational Mathematics. Springer, Cham, 2016, pp. xiii+812.

[26] C. Kreuzer, R. Verfürth, and P. Zanotti. *Quasi-optimal and pressure robust discretizations of the Stokes equations by moment- and divergence-preserving operators*. 2020.

[27] P. L. Lederer and S. Rhebergen. “A Pressure - Robust Embedded Discontinuous Galerkin Method for the Stokes Problem by Reconstruction Operators”. In: *SIAM J. Numer. Anal.* 58.5 (2020), pp. 2915–2933.

[28] P. L. Lederer and J. Schöberl. “Polynomial robust stability analysis for $H(\mathrm{div})$-conforming finite elements for the Stokes equations”. In: *IMA Journal of Numerical Analysis* (2017), drx051.

[29] P. L. Lederer et al. “Divergence-free Reconstruction Operators for Pressure-Robust Stokes Discretizations with Continuous Pressure Finite Elements”. In: *SIAM J. Numer. Anal.* 55.3 (2017), pp. 1291–1314.

[30] A. Linke, G. Matthies, and L. Tobiska. “Robust Arbitrary Order Mixed Finite Element Methods for the Incompressible Stokes Equations with pressure independent velocity errors”. In: *ESAIM: M2AN* 50.1 (2016), pp. 289–309.

[31] A. Linke. “On the role of the Helmholtz decomposition in mixed methods for incompressible flows and a new variational crime”. In: *Computer Methods in Applied Mechanics and Engineering* 268 (2014), pp. 782–800.

[32] J. Lions and E. Magenes. *Problèmes aux limites non homogènes et applications*. Problèmes aux limites non homogènes et applications Bd. 1. Dunod, 1968.

[33] K.-A. Mardal, J. Schöberl, and R. Winther. “A uniformly stable Fortin operator for the Taylor-Hood element”. In: *Numer. Math.* 123.3 (2013), pp. 537–551.

[34] N. Meyers and J. Serrin. “$H = W$”. In: *Proc* 8.R-2 (1974), pp. 129–151.

[35] J. Nečas. *Les Méthodes Directes en Théorie des Equations Elliptiques*. Masson, 1967.

[36] J. Nečas. “Sur une méthode pour résoudre les équations aux dérivées partielles du type elliptique, voisine de la variationnelle”. In: *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (3)* 16 (1962), pp. 305–326.

[37] L. Nirenberg. “Remarks on strongly elliptic partial differential equations”. In: *Comm. Pure Appl. Math.* 8 (1955), pp. 649–675.

[38] Panton, R.L. *Incompressible Flow*. Wiley, 2013.

[39] A. Prohl. *Projection and Quasi-Compressibility Methods for Solving the Incompressible Navier-Stokes Equations*. Vieweg+Teubner Verlag, 1997.

[40] W. H. Reed and T. Hill. “Triangular mesh methods for the neutron transport equation”. In: *Los Alamos Report LA-UR-73-479* (1973).

[41] P. W. Schroeder et al. “Towards computable flows and robust estimates for inf-sup stable FEM applied to the time-dependent incompressible Navier–Stokes equations”. In: *SeMA Journal* 75.4 (2018), 629–653.

[42] D. Tritton. *Physical Fluid Dynamics*. Oxford Science Publ. Clarendon Press, 1988.

[43] L. Zhao, E. Park, and E. Chung. *A pressure robust staggered discontinuous Galerkin method for the Stokes equations*. 2020.